Variational Electrostatic Projection (VEP) Methods for Efficient Modeling of the Macromolecular Electrostatic and Solvation Environment in Activated Dynamics Simulations

Brent A. Gregersen and Darrin M. York*

Department of Chemistry, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455-0431

Received: July 7, 2004; In Final Form: September 30, 2004

New methods for the calculation of electrostatic interactions between the active dynamical region and surrounding external solvated macromolecular environment in hybrid quantum mechanical/molecular mechanical (QM/MM) simulations are presented. The variational electrostatic projection (VEP) method, and related variational reverse-mapping procedure (VEP−RVM) utilize an expansion in Gaussian surface elements for representation of electrostatic interactions. The use of a discretized surface that surrounds the active dynamical region greatly reduces the number of interactions with the particles of the external environment. The methods are tested on two catalytic RNA systems: the hammerhead and the hairpin ribozymes. It is shown that with surface elements numbering from 302 to 1202 points the direct VEP and VEP−RVM methods are able to obtain relative force errors in the range of 0.5−0.05% and 0.09−0.0001%, respectively, using a 4.0 Å projection buffer. These results are encouraging and provide an essential step in the development of new variational macromolecular solvent-boundary methods for QM/MM calculations of enzyme reactions.

1. Introduction

Molecular simulation of reactions catalyzed by enzymes and ribozymes using hybrid quantum mechanical/molecular mechanical methods offer a potentially powerful theoretical tool to study biological reactions.[1−4] These methods continue to undergo improvement, including the reliability of fast quantum models,[5−7] methods for enhanced sampling of reaction paths,[8] molecular simulation force fields[9−13] including new-generation polarizable force fields,[14−16] and methods for efficiently treating the macromolecular electrostatic fields.[17,18]

Proper modeling of the solvated macromolecular environment[19] is critical for reliable simulation of catalytic events, especially when considering highly charged systems such as those involved in RNA catalysis.[20,21] Electrostatic preorganizational effects, that involve both the short and long-range interactions, play a key role in enzyme catalysis and are often attributed as being a main source of catalytic activity.[19,22] It has long been established that neglect of long-range electrostatic interactions in nucleic acid systems leads to severe artifacts in molecular simulations.[18,23,24] Nonetheless, in many hybrid quantum mechanical/molecular mechanical (QM/MM) simulations reported in the modern literature (for a recent review, see ref 4), long-ranged electrostatic interactions are still often treated using an electrostatic cutoff, and results derived from such simulations should perhaps be viewed with considerable skepticism.

In the case of periodic boundary conditions, proper treatment of long-range electrostatics can be realized through the use of linear-scaling Ewald methods.[18] In some instances, however, full periodic boundary simulation of a highly solvated macromolecule with a QM/MM potential may be too computationally demanding, especially when combined with the extensive amount of sampling required to obtain converged free energy profiles along multiple reaction paths. Hybrid QM/MM simulations are typically considerably more costly than purely MM simulations, in part due to the QM/MM electrostatic interaction, and this increase in computational cost is magnified for explicit solvent periodic boundary simulations since (1) introduction of large amounts of explicit solvent requires many more QM/MM interactions to be calculated, (2) for highly charged systems, the presence of mobile counterions requires especially long equilibration times, and (3) the increase in explicit degrees of freedom necessitates much longer simulation to properly sample low-frequency vibrational modes of the system, especially for conformationally dynamic systems. Together, these factors impose tremendous computational demands on the calculation of complex biological reaction mechanisms, and they are particularly severe for RNA enzymes (ribozymes), such as the hammerhead and hairpin ribozymes, that are both highly charged and conformationally dynamic.

An alternative to performing full periodic boundary simulations for QM/MM simulations is to use a stochastic boundary method[25−27] whereby an "active" region that undergoes normal Hamiltonian dynamics is surrounded by a "stochastic" region whereby atoms undergo Langevin dynamics. The atoms in the stochastic region are thermalized by random forces and confined by a solvent boundary potential, and in macromolecular simulations, the solute-atom positions are usually also restrained by harmonic forces.

The motivation for use of solvent boundary potential and stochastic simulation for calculations of QM/MM free energy profiles of enzymes is to avoid the tremendous cost associated with a large scale QM/MM calculation and converged sampling at every step along one or more 1-dimensional or even 2-dimensional reaction coordinates. A key step in this process is to accurately evaluate the long-ranged electrostatic forces in the active region due to all the external atoms in order to properly establish the solvated macromolecular environment. Here it is critical to emphasize that not only the macromolecular solute electrostatics but also the electrostatics due to the polarized solvent need to be considered. In fact, inclusion of

* Corresponding author. Telephone: (612) 624−8042. Fax: (612) 626−7541. E-mail: york@chem.umn.edu. Url: http://riesling.chem.umn.edu.

10.1021/jp0469968 CCC: $30.25
Published on Web 12/14/2004

the former without consideration of the latter is typically much *worse* than completely neglecting the external electrostatic environment entirely (which is of course disastrous), since the effect of solvent is to greatly damp the exposed solute charges.

A few methods have been reported that address the problem of accurately modeling the solvated macromolecular environment in the active region and have offered considerable improvement. One method that has been applied to QM/MM simulations[28] is based on a charge-scaling procedure originally developed for molecular-mechanical free-energy simulations.[29] In this method, a thermodynamic cycle is used to obtain QM/MM reaction profiles from explicit simulation using scaled charges and a minimal amount of explicit solvent. These profiles are then corrected with finite-difference Poisson or linearized Poisson−Boltzmann calculations. An alternate approach is to employ a generalized solvent boundary potential[30−32] method based on a grid representation for the static potential and a matrix representation for the Green's function that represents the interaction between the atoms in the active region with the generalized multipoles of the surrounding solvated macromolecular environment. This method appears very promising, but has not yet been applied with QM/MM simulations.

In the present paper two new methods are presented, the variational electrostatic projection (VEP) method and related VEP with reverse variational mapping procedure (VEP−RVM) for accurate and efficient calculation of the interaction between two charge distributions separated by a surface. The methods introduce a smooth Gaussian basis set for expansion of the multipolar charge distribution of a solvated macromolecule on a discretized surface that surrounds the active/stochastic region of interest, and a variational mapping procedure to obtain the electrostatic interaction between particle distributions on opposite sides of the surface. The theoretical background is described and the details of the method are presented along with tests of the force accuracy using different discretization and constraint protocols. The methods presented here have several potential advantages over other methods that have been previously proposed, and they provide a foundation for further development of generalized variational macromolecular-solvent boundary Green's function techniques for QM/MM simulations.

The paper is organized as follows: the following (second) section of the paper outlines the basic theoretical background and key equations for the VEP and VEP−RVM methods. The third section presents a numerical implementation and describes how the error analysis was performed. The fourth section presents the results of numerical tests of force accuracy for RNA systems. The fifth section discusses the method in relation to other proposed methods in the literature. The sixth section summarizes the main conclusions and identifies directions for future work.

2. Theory

This section describes the theoretical formulation for the VEP method from classical electrostatics, beginning with a variational principle for a conductor. The appendices provide a detailed theoretical formulation of the VEP and related VEP−RVM methods, variational principles and Green's function expansion methods, both with and without generalized constraints. This section provides a brief illustration of the basic principles that serve as the foundation of the VEP method, and presents only the key equations that are described in full detail in the appendices.

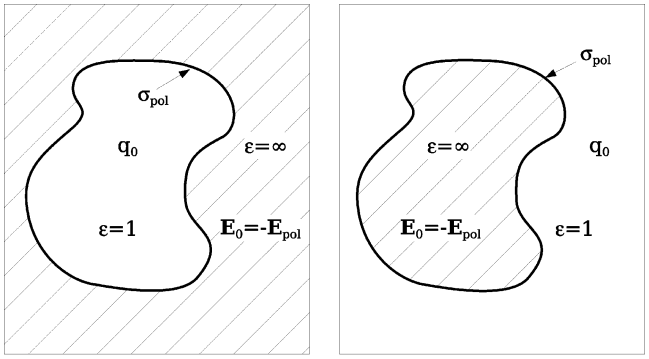
2.1. Conductors and Polarizable Media. Polarization is the response of a charge distribution to a perturbing electric field. The response can be, in general, a nonlocal, nonlinear, and frequency dependent function of the electric field perturbation. Polarization can arise from induced electronic polarization, nuclear relaxation, and molecular reorientation or conformational change.[33]

The *molecular polarizability*, in general a tensor quantity, is defined as the second derivative of the energy with respect to the applied electric field components, and is related to the linear-response function.[34] In the linear-response regime, the induced dipole moment of a molecule is proportional to the magnitude of the applied field, the proportionality constant for which is the molecular polarizability. The inverse of the polarizability plays the role of a quadratic energetic penalty the molecule must pay in order that the charge distribution can respond from its unperturbed equilibrium state to interact favorably with the applied electric field perturbation.

In the theory of continuum electrostatics, for a linear isotropic polarizable medium, the dielectric function is simply related to the *electric susceptibility*, which is the macroscopic analogue of to the molecular polarizability.[35] The polarizable medium responds to (polarizes in) the presence of electric fields ($\mathbf{E}_0$) such as those arising from a static charge distribution ($\mathbf{q}_0$). The polarization response of the medium can be described in terms of a *polarization density*, σ_{pol}, that gives rise to the potential, ϕ_{pol}, and its associated electric field, $\mathbf{E}_{\mathrm{pol}} = -\nabla\phi_{\mathrm{pol}}$, sometimes referred to as the *reaction field*. The total electric field, in the presence of the dielectric, is the sum $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_{\mathrm{pol}}$. A vacuum is a nonpolarizable medium (because nothing is there to be polarized) and is characterized by a unit dielectric ($\epsilon = 1$). A conductor, on the other hand, has an infinite dielectric ($\epsilon = \infty$). The electrons of a conductor, such as a metal, can move freely without intrinsic energy penalty beyond the electric self-energy of the polarization density itself. As a result, a conducting medium *cannot support an electric field*.[35,36] In other words, if a charge distribution is surrounded by a conductor, the electric field due to the dielectric response ($\mathbf{E}_{\mathrm{pol}}$) must *exactly cancel* the field of the static charge distribution itself ($\mathbf{E}_0$) in the region where the dielectric is infinite. Thus, solution of the conductor dielectric problem produces a polarization density with associated electric field that is exactly minus the electric field of the static charge distribution *in the region where* $\epsilon = \infty$. This result is exploited in the case of the VEP and VEP−RVM methods where the goal is to model the electric field due to a large spatially delocalized static charge distribution inside a relatively small dynamical region (see below).

Relation between the COSMO Solvation Model and the VEP Method. The solution to the conductor problem where the charge distribution is contained in a cavity of unit dielectric and is surrounded by a conductor gives rise to the conventional conductor-like screening model (COSMO) for solvation[37,38] (Figure 1a). In the present VEP and VEP−RVM methods, the situation is reversed: the dielectric problem is solved for a cavity filled with a conducting medium surrounded by a unit dielectric that contains a large spatially delocalized charge distribution *outside of the cavity* (Figure 1b). Both of the dielectric models of Figure 1 can be formulated as boundary value problems that can be solved efficiently with boundary element methods. In the case of the VEP method, a Green's function formulation is presented that allows extension to macromolecular and solvent linear response (see the appendices for full details).

The spatially delocalized charge distribution, in the present development of the VEP and VEP−RVM methods, is a large solvated biological macromolecule (a ribozyme), and the small dynamical region is the active site where a chemical reaction



a) COSMO Method b) VEP Method

Figure 1. Schematic showing the dielectric problems that are solved in (a) the conductor-like screening model for solvation, and (b) the variational electrostatic projection methods of the present work. For the VEP methods, it should be noted that treatment of the active dynamical region as a conductor is done solely to obtain the polarization density, the negative value for which reproduces the electric field (and potential) of the static external charge distribution at significantly reduced computational cost (see text). During dynamics, the active region is of course not treated as a conductor.

occurs. The goal is to derive a variational method to calculate a simplified model charge distribution that, in the dynamical active site region, reproduces the electric field (and potential) due to the surrounding solvated macromolecule to very high accuracy. Great computational advantage can be realized in molecular dynamics (or Monte Carlo) simulations if the number of points used to describe the model charge distribution is much less than the number of points used to describe the solvated macromolecule. As is discussed in the next section, the VEP and VEP−RVM methods achieve this goal by using as the model charge distribution minus the polarization density of the conductor solution to the dielectric problem described in Figure 1b.

2.2. The VEP and VEP−RVM Equations. This section presents only the key equations associated with the VEP and VEP−RVM methods. The full development of these equations are provided in the appendices.

Consider the total electrostatic energy of a system of charged particles subdivided into two regions, designated the *active region* and the *environment*, and separated by a bounding surface. Let the reference charge distribution inside the active region be designated $\boldsymbol{\rho}_0$, and let the reference static charge distribution of the environment consist of two parts: one part arising from the macromolecular solute, designated $\mathbf{q}_0$, and another arising from the surrounding solvent, designated $\boldsymbol{\sigma}_0$. Here, the charge distributions $\boldsymbol{\rho}_0$, $\mathbf{q}_0$ and $\boldsymbol{\sigma}_0$ are written as matrix (column vector) quantities with dimensions $N_\rho \times 1$, $N_\mathbf{q} \times 1$ and $N_\sigma \times 1$, respectively. The elements of the vectors represent expansion coefficients for the charge distributions in a representative basis. The specific basis used to represent the vector and matrix quantities introduced in this section are discussed in detail in section 4.2.

The total electrostatic energy of the system is given by

$$E = \frac{1}{2}\boldsymbol{\rho}_0{}^{\mathrm{T}}\cdot\mathbf{C}_{\rho\rho}\cdot\boldsymbol{\rho}_0 + \frac{1}{2}\mathbf{q}_0{}^{\mathrm{T}}\cdot\mathbf{C}_{\mathbf{qq}}\cdot\mathbf{q}_0 + \boldsymbol{\rho}_0{}^{\mathrm{T}}\cdot\mathbf{C}_{\rho\mathbf{q}}\cdot\mathbf{q}_0 + \frac{1}{2}\boldsymbol{\sigma}_0{}^{\mathrm{T}}\cdot\mathbf{A}_{\sigma\sigma}\cdot\boldsymbol{\sigma}_0 + \boldsymbol{\rho}_0{}^{\mathrm{T}}\cdot\mathbf{B}_{\rho\sigma}\cdot\boldsymbol{\sigma}_0 + \mathbf{q}_0{}^{\mathrm{T}}\cdot\mathbf{B}_{\mathbf{q}\sigma}\cdot\boldsymbol{\sigma}_0$$

where the matrices **A**, **B**, and **C** represent Coulomb integrals between basis functions and are subscripted by the specific charge distributions that they connect. The convention is used that a reversal of the matrix subscripts indicates the matrix transpose operation, e.g., $\mathbf{A}_{\beta\alpha} \equiv (\mathbf{A}_{\alpha\beta})^{\mathrm{T}}$.

The VEP method provides a mechanism for determination of a projection surface charge vector $\boldsymbol{\gamma}_\mathbf{X}(\lambda)$ that resides on the surface surrounding the active region (here the subscript **X** refers to the charge distribution vector that describes the combined external charge distributions $\mathbf{q}_0$ and $\boldsymbol{\sigma}_0$). The direct VEP method for determination of $\boldsymbol{\gamma}_\mathbf{X}(\lambda)$ is derived from a variational principle and is given by

$$\boldsymbol{\gamma}_\mathbf{X}(\lambda) = \mathbf{P}_{\gamma\gamma}{}^{-1}\cdot\mathbf{B}_{\gamma\mathbf{X}}(\lambda)\cdot\mathbf{X}$$

where $\mathbf{P}_{\gamma\gamma}$ is the projection surface element interaction matrix, $\mathbf{B}_{\gamma\mathbf{X}}(\lambda)$ is the interaction matrix between the external charge distribution **X** and the projection surface elements, and the λ indicates that an optional set of constraints can be imposed in the variational procedure, for example, inclusion of generalized far field Taylor multipole moments (see Appendix A.2).

An improvement on the direct VEP method is the modified VEP−RVM method that uses an intermediate projection surface charge vector ($\boldsymbol{\omega}_\mathbf{X}(\lambda)$) that resides interior to the actual projection surface.

$$\boldsymbol{\gamma}_\mathbf{X}(\lambda) = [\,\mathbf{P}_{\omega\omega}{}^{-1}\cdot\mathbf{B}_{\omega\gamma}(\lambda)]^{-1}\cdot\boldsymbol{\omega}_\mathbf{X}(\lambda) = \mathbf{B}_{\omega\gamma}(\lambda)^{-1}\cdot\mathbf{B}_{\omega\mathbf{X}}(\lambda)\cdot\mathbf{X}$$

where $\mathbf{P}_{\omega\omega}$ is the intermediate projection surface element interaction matrix, and $\mathbf{B}_{\omega\gamma}(\lambda)$ and $\mathbf{B}_{\omega\mathbf{X}}(\lambda)$ are analogous to $\mathbf{B}_{\gamma\mathbf{X}}(\lambda)$ above.

A point that needs to be emphasized is that the VEP formulation has tremendous potential application in molecular dynamics simulations, such as hybrid QM/MM simulations in the active site of an enzyme or ribozyme, where accurate and efficient modeling of the full solvated macromolecular electrostatic environment is critical. If the number of points required to discretize the projection surface at an acceptable level of accuracy is much less than the number of explicit particles outside of the active dynamical region, then tremendous computational savings would be realized. Moreover, if demonstrated to be sufficiently accurate, the method would provide the basis for other useful procedures such as charge rescaling and inclusion of the solvent and macromolecular linear response. In effect, new methods for rigorous treatment of a solvated macromolecular environment within a linear-response framework could be afforded for QM/MM simulations at a fraction of the computational requirement for full, explicit simulation. Moreover, the use of linear-scaling electrostatic methods and preconditioned conjugate gradient minimization techniques, as has been applied previously with the COSMO method,[39] would allow the VEP and VEP−RVM methods to scale linearly with number of particles. For active dynamical regions of size typical to most modern QM/MM simulations, only a fairly modest number of projection surface elements are required (302−1202), and all the associated vectors (and matrices in the case of extension to a linear-response framework) can be easily precomputed and stored. The key step toward the development of all of these methods is presented in the present work: the derivation, implementation and validation of a VEP and VEP−RVM methods for biological macromolecules.

3. Methods

3.1. Numerical Implementation. The technical details of the implementation of the discretized surfaces used in the variational

TABLE 1: Gaussian Exponents for a Unit Sphere at Different Discretization Levels[a]

N_γ	l_{max}	$2l_{max}+1$	ζ	γ_{relrms}
14	2	5	4.865	8.9×10^{-4}
26	3	7	4.855	6.9×10^{-3}
50	5	11	4.893	3.5×10^{-3}
110	8	17	4.901	3.8×10^{-3}
194	11	23	4.903	2.4×10^{-3}
302	14	29	4.905	1.9×10^{-3}
434	17	35	4.906	1.2×10^{-3}
590	20	41	4.905	2.7×10^{-3}
770	23	47	4.899	5.6×10^{-3}
974	26	53	4.907	5.9×10^{-3}
1202	29	59	4.907	4.5×10^{-4}

[a] Optimized values of ζ defined in eq 4 that give the exact Born ion energy are shown for angular quadrature schemes designated by l_{max} defined in eqs 1 and 2. The relative deviation γ_{relrms} is defined as $\gamma_{relrms} = \sqrt{\langle(\boldsymbol{\gamma}_{calc}-\boldsymbol{\gamma}_{exact})^2\rangle}/\sqrt{\langle\boldsymbol{\gamma}_{exact}{}^2\rangle}$ where $\boldsymbol{\gamma}_{calc}$ is the surface charge vector calculated with the (unconstrained) VEP method, $\boldsymbol{\gamma}_{exact}$ is the "exact" uniform surface charge vector defined as $\boldsymbol{\gamma}_{exact} = -\mathbf{w}/4\pi$, and $\mathbf{w}$ is the vector of angular quadrature weights for a given discretization level. Note that as the number of points increases, the denominator $\langle\boldsymbol{\gamma}_{exact}{}^2\rangle$ decreases such that the relative deviation is fairly constant. The minimum linear correlation coefficient between $\boldsymbol{\gamma}_{calc}$ and $\boldsymbol{\gamma}_{exact}$ is 0.99998 for the 26 point scheme.

electrostatic projection and reverse mapping procedures parallel very closely those described for the construction of the discretized surfaces used in a smooth boundary element solvation method, based on a conductor-like screening model that has been presented elsewhere.[38] In the smooth solvation model, the discretized surface that is used as a basis to represent the solvent reaction field is constructed as a superposition of individually discretized atom-centered spheres. The discretization of each atomic sphere was derived from sets of points and weights used in high-order numerical angular quadrature schemes with octahedral symmetry adapted for integration of spherical harmonic functions,[40] first pioneered by Lebedev[41] and extended to high order by Delley.[42] The atom-centered spheres were all of atomic dimensions and could be individually discretized to fairly low order in practice (typically 26−110 points). In the present method, the variational projection surface must encapsulate a spherical region the size of a typical active site that would undergo activated dynamics simulation − typically of radius 20−25 Å. This requires a single large spherical cavity that is discretized to very high order (with respect to points per sphere).

The high order angular integration quadrature schemes provide a set of M points $\{\hat{\mathbf{r}}_k\}$ and weights $\{w_k\}$, $k = 1, ..., M$, on a unit sphere such that for $0 \le l \le 2l_{max}+1$, the L_1 norm conditions for spherical harmonics $Y_{l,m}(\hat{\mathbf{r}})$ are satisfied:

$$\int Y_{l,m}(\hat{\mathbf{r}})\,d\Omega = \sum_k^M Y_{l,m}(\hat{\mathbf{r}}_k)w_k = \sqrt{4\pi}\delta_{l,0} \quad (1)$$

For $0 \le l, l' \le l_{max}$, the related L_2 orthonormality conditions are satisfied:

$$\int Y^*_{l,m}(\hat{\mathbf{r}})Y_{l',m'}(\hat{\mathbf{r}})\,d\Omega = \sum_k^M Y^*_{l,m}(\hat{\mathbf{r}}_k)Y_{l',m'}(\hat{\mathbf{r}}_k)w_k = \delta_{l,l'}\delta_{m,m'} \quad (2)$$

The numerical accuracy of the various octahedrally symmetric angular quadrature schemes are shown in Table 1. Surface elements are modeled by L_1-normalized spherical Gaussian functions located at the quadrature points, i.e.,

$$g_k(\mathbf{r}) = \left(\frac{\zeta_k^2}{\pi}\right)^{3/2} \exp[\zeta_k^2|\mathbf{r}-\mathbf{r}_k|^2] \quad (3)$$

where ζ_k satisfies a simple relation with the angular quadrature weight w_k:

$$\zeta_k = \zeta/\sqrt{w_k} \quad (4)$$

where the constant ζ is chosen for each quadrature scheme to reproduce the exact projection of a charge Q at the center of the unit sphere. Owing to the relation between the Gaussian exponents and angular quadrature weights (eq 4), the solution of the variational electrostatic projection very closely reproduces a uniform projected charge distribution γ_k for a point charge Q inside a spherical cavity

$$\gamma_k = Qw_k/4\pi \quad (5)$$

Table 1 lists the optimized values of the scaling parameter ζ of eq 4 for each discretization level, along with the relative root-mean-square error of the calculated VEP surface charge coefficients from the ideal uniform values given in eq 5. Also shown are the number of discretization points and the l_{max} values (eqs 1 and 2) corresponding to the angular quadrature schemes from which the points were derived. It is interesting to note that with only 26 points, exact orthonormality relations are satisfied for multipole orders up to octupole ($l_{max} = 3$), with 302, 590, and 1202 points, these relations are extended to orders $l_{max} = 14$, 20 and 29, respectively.

To obtain discretized spheres of different radii R, the discretization points, weights and Gaussian exponents obey the exact scaling relations

$$\mathbf{r}_k(R) = \mathbf{r}_k \cdot R \quad (6)$$

$$w_k(R) = w_k \cdot R^2 \quad (7)$$

$$\zeta_k(R) = \zeta_k/R \quad (8)$$

It remains to describe the explicit functional forms used to construct the various matrix elements used in the linear algebraic equations of the previous sections that describe the VEP and VEP−RVM procedures. The convention that is used in subsections 6.3 and 6.4 is that matrices that represent the electrostatic interaction between discretized Gaussian surface elements of either the γ or ω projection surfaces are designated by $\mathbf{P}_{\alpha\beta}$, where the subscripts indicate the two interacting projection spheres (i.e., $\mathbf{P}_{\gamma\gamma}$, $\mathbf{P}_{\gamma\omega}$, or $\mathbf{P}_{\omega\omega}$). The matrix element $(\mathbf{P}_{\alpha\beta})_{ij}$ represents the electrostatic interactions between the ith surface element of projection sphere α with the jth surface element of projection sphere β (where α and β generically represent γ or ω). This interaction takes the form of a Gaussian−Gaussian repulsion integral between the surface elements:

$$(\mathbf{P}_{\alpha\beta})_{ij} = \frac{\mathrm{erf}[\zeta_{\alpha i,\beta j}|\mathbf{r}_{\alpha i}-\mathbf{r}_{\beta j}|]}{|\mathbf{r}_{\alpha i}-\mathbf{r}_{\beta j}|} \quad (9)$$

$$\zeta_{\alpha i,\beta j} = \frac{\zeta_{\alpha i}\zeta_{\beta j}}{\sqrt{\zeta_{\alpha i}{}^2+\zeta_{\beta j}{}^2}} \quad (10)$$

$$\zeta_{\alpha i} = \frac{\zeta}{\sqrt{w_i}\cdot R_\alpha} \quad (11)$$

where $\boldsymbol{r}_{\alpha i}$ and $\boldsymbol{r}_{\beta j}$ are the positions of the ith Gaussian on the α projection surface and jth Gaussian on the β projection surface, respectively, and $\zeta_{\alpha i}$ and $\zeta_{\beta j}$ are the corresponding Gaussian exponents (eq 3). Note: the projection surfaces can be the same, i.e., $\alpha = \beta$, in which case the diagonal element is obtained as the limit as $|\boldsymbol{r}_{\alpha i} - \boldsymbol{r}_{\beta j}| \rightarrow 0$

$$(\mathbf{P}_{\alpha\alpha})_{ii} = \frac{2}{\sqrt{\pi}}\zeta_{\alpha i,\alpha i} = \sqrt{\frac{2}{\pi}}\zeta_{\alpha_i} \tag{12}$$

Note: there is a misprint in eq 67 of ref 38 involving the neglect of a multiplicative factor of 2.

Following the convention for the **P** matrices, the matrices that represent electrostatic interaction between the discretized Gaussian surface elements of the projection surfaces and the molecular charge distributions, and between the molecular charge distributions themselves are designated **B** and **C**, respectively. The **B** matrices are subscripted with two indices $\alpha\beta$ that indicate the interacting projection surface and molecular charge distribution, and the **C** matrices are subscripted with two indices $\alpha\beta$ that indicate the two interacting molecular charge distributions.

Specific evaluation of the elements of the **B** and **C** matrices, therefore, requires specification of a basis for expansion of molecular charge distributions. In the present work, all molecular charge distributions are represented in the basis of point charges, as is typical in conventional force field models. For this choice of basis, the **B** matrix takes the form

$$(\mathbf{B}_{\alpha\beta})_{ij} = \frac{\mathrm{erf}[\zeta_{\alpha i}|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|]}{|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|} \tag{13}$$

where $\mathbf{r}_{\alpha i}$ and $\mathbf{r}_{\beta j}$ are the positions of the ith Gaussian on the α projection surface and jth point charge of the β molecular charge distribution, respectively, and $\zeta_{\alpha i}$ is the Gaussian exponent of the former (eq 3). A drawback of this form for the **B** matrix, however, is that as a point charge becomes superimposed directly on top of a Gaussian surface element on the projection sphere, the VEP method will slightly overestimate the corresponding projected surface charge coefficient since a Gaussian-point charge interaction is greater than a Gaussian−Gaussian interaction (with the L_1 norm condition applied here). To produce the correct limit, an alternate form of the **B** matrix was tested whereby the Gaussian-point charge interaction was replaced by a Gaussian−Gaussian interaction of the form

$$(\mathbf{B}'_{\alpha\beta})_{ij} = \frac{\mathrm{erf}[\zeta_{\alpha i,\alpha i}|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|]}{|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|} \tag{14}$$

where $\zeta_{\alpha i,\alpha i} = \zeta_{\alpha i}/\sqrt{2}$. The **B′** form of the matrix was generally observed to give better overall results than the **B** matrix, the difference quickly becomes insignificant as the discretization level is increased and as the distance between the atoms of the molecular charge distribution and the projection surface increases (see section 4.1 for further details).

Evaluation of the elements of the **C** matrix for a point-charge basis expansion of the molecular charge distributions follow Coulomb's law, with the standard convention that the infinite self-energy of a point charge is neglected

$$(\mathbf{C}_{\alpha\beta})_{ij} = \frac{(1 - \delta_{\alpha i,\beta j})}{|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|} \tag{15}$$

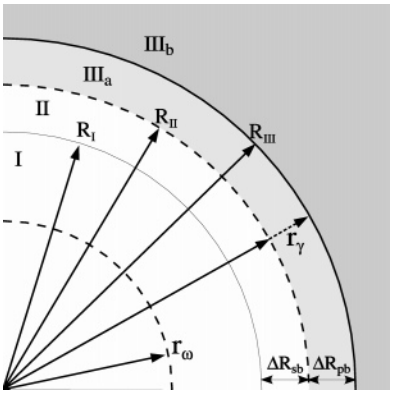


Figure 2. Schematic showing the regions defined for the VEP and VEP−RVM procedures. Regions I and II are the active dynamical regions, with the atoms of region I propagated using Newtonian dynamics and those of region II propagated using Langevin dynamics. Region III (subdivided into IIIa and IIIb) comprises the external macromolecular and solvent environment that, in the present work are held fixed. In the direct VEP method and VEP−RVM methods, the electrostatic potential due to atoms of external environment (region IIIb), excluding those of the *projection buffer* (region IIIa), are replaced by the electrostatic potential derived from a surface charge distribution (γ) located at a position r_γ from the center of the dynamical zone. In this paper, the γ surface is located at $r_\gamma = R_{II}$ for the VEP method, and at $r_\gamma = R_{III}$ for the VEP−RVM method. The ω surface is an intermediate surface used for the VEP−RVM method, and its location is given by eq 24.

where $\delta_{\alpha i,\beta j}$ is the Kronecker δ function that has a value of 1 if $\alpha = \beta$ *and* $i = j$, and a value of 0 otherwise.

3.2. Activated Dynamics Simulation Model. The VEP and VEP−RVM methods introduced in the present work are aimed at the problem of evaluating the electrostatic potential in a localized spatial region (e.g., an enzyme active site) due to an expansive spatially delocalized charge distribution (e.g., the surrounding macromolecule and solvent). The problem is equivalent to the transformation of a direct Green's function solution of the electrostatic problem (that involves integration over the spatial extent of the delocalized charge distribution) to a boundary value problem that involves integration over only the surface that encompasses the localized region (the bounding surface). The latter approach can lead to a tremendous computational advantage so long as knowledge of the electrostatic potential and associated fields are required only on the inside of the bounding surface.

In hybrid QM/MM calculations, it is important to establish a realistic model for the dynamical environment in the region where the key residues involved in the chemical reaction occur. Because of the expense of the QM/MM calculations, combined with large system sizes and need for extensive configurational sampling, full molecular dynamics simulation of all of the enzyme and solvent degrees of freedom is often not feasible. To circumvent this problem, so-called stochastic boundary simulations are frequently performed.[25,27] In these calculations, a hierarchy of concentric zones are built up, and atoms within each zone are subject to different restraint forces and propagated by different dynamical methods. Most commonly there are three spherical zones, illustrated in Figure 2:

Zone I (the reaction zone) contains the key residues and solvent in the active site. The atoms in this region undergo Hamiltonian dynamics.

Zone II (the stochastic buffer zone) forms a layer around zone I. The atoms in this region undergo stochastic (Langevin) dynamics.

Zone III (the external environment) contains the macromolecular and solvent environment. The atoms in this region are held fixed.

The goal is to accurately model the energetics and dynamics of the atoms in zone I where the chemical reaction takes place without explicitly simulating the very large number of atoms that make up the solvated macromolecular environment (zone III). To accomplish this, a stochastic buffer zone (zone II) is introduced.

Zones I and II are the active dynamical regions that require evaluation of atomic forces at each time step of the simulation. The atoms of zone I are propagated with Hamiltonian dynamics and it is from these atoms that the free energy profiles and other properties of interest are derived. The atoms of zone II are propagated with Langevin dynamics, and hence this zone acts as a thermal reservoir to maintain constant simulation temperature. The solute atoms in zone II are typically restrained by harmonic potentials to their equilibrium positions in order to maintain the structural integrity of the enzyme environment immediately surrounding the active site, and solvent molecules are contained by a deformable solvent boundary potential[27] to prevent diffusion into zone III. Hence, zone II acts as a stochastic buffer region between the reaction zone (zone I) and the surrounding external environment (zone III). Zone III consists of the atoms of the macromolecular and solvent environment that have an important electrostatic preorganizational effect[22] but otherwise do not participate directly in the chemical reaction of interest. The atoms of zone III are most commonly held fixed—although it is possible to treat them in a linear-response framework [The VEP and VEP−RVM methods presented can, in principle, be extended to treat macromolecular and solvent linear-response. This is a focus of future work.] *It is the goal of the present work to introduce a method to rapidly evaluate accurate electrostatic forces on the atoms in zone I due to the atoms in zone III, and establish a reduced-dimensional basis for representation of the interactions between these zones.*

3.3. Analysis of Force Errors in Ribozyme Systems. To assess the accuracy of the VEP and VEP−RVM methods presented in the present work, the active sites of two prototype ribozyme systems[43] were considered as test systems: the hammerhead ribozyme[44,45] and the hairpin ribozyme.[46−48] These systems are of tremendous current interest from the viewpoint of gaining a fundamental understanding of mechanisms of RNA catalysis, and they provide the motivation for the present work. The highly charged nature of these systems combined with a need for inclusion and equilibration of extensive amounts of water and counterions makes these systems particularly challenging for hybrid QM/MM simulations.

In the case of the hammerhead ribozyme, the smallest known ribozyme, the solvated system used for testing was constructed based on the 299D crystal structure of Scott et al.[44] This structure was solvated in a rhombododecahedral unit cell ($a = b = c =$ 68.5 Å, $\alpha = \gamma = 60°$, $\beta = 90°$) with the scissile phosphate centered at the origin. The final system contained 1329 solute atoms, 7187 water molecules, 130 sodium ions and 89 chlorine ions and was equilibrated for 1 ns at 300 K. For the tests of the VEP and VEP−RVM methods, the reaction zone was defined by a 16 Å sphere around the scissile phosphate ($R_I = 16$ Å) with a 4 Å stochastic buffer ($\Delta R_{sb} = 4$ Å) leading to an overall 20 Å region of active dynamics ($R_{II} = 20$ Å). In the case of the hairpin ribozyme, the solvated system used for testing was constructed based on the 1HP6 crystal structure of Rupert et al.[46] This structure was solvated in a rhombododecahedral unit cell ($a = b = c = 75.1$ Å, $\alpha = \gamma = 60°$, $\beta = 90°$) with the scissile phosphate centered at the origin. The final system contained 2459 solute atoms, 9641 water molecules, 99 sodium ions, and 23 chloride ions, and it was equilibrated for 1 ns at 300 K. For the tests of the VEP and VEP−RVM methods, the reaction zone was defined by a 20 Å sphere centered on the scissile phosphate ($R_I = 20$ Å) with a 5 Å stochastic buffer ($\Delta R_{sb} = 5$ Å) leading to an overall 25 Å region of active dynamics ($R_{II} = 25$ Å).

The accuracy of the VEP and VEP−RVM methods for these systems is characterized by the force errors due to electrostatic interactions of the external environment (zone III) with the atoms of the reaction zone (zone I). As will be discussed in more detail below, the accuracy of the VEP and VEP−RVM methods is poorest for atoms that lie almost on top of the projection surfaces. Hence, improved accuracy can be obtained by treating atoms within some small distance of the projection surface explicitly and projecting only the electrostatics due to atoms outside this distance. The external environment (zone III) can hence be subdivided into two regions, designated zone IIIa and zone IIIb. Zone IIIa (termed the *projection buffer*) consists of atoms within a certain distance ($\Delta R_{pb} = R_{IIIa} - R_{II}$), of the active region contained within R_{II}. Inclusion of a projection buffer slightly increases the computational cost of the method since now the electrostatics of the atoms within the projection buffer must be calculated explicitly.

The exact force on atom i of the reaction zone due electrostatic interactions with atoms from the external environment is designated $\mathbf{f}_i^{\text{exact}}$ and given by

$$\mathbf{f}_i^{\text{exact}} = \sum_j^{\text{zone III}} -\nabla_i\left(\frac{Q_iQ_j}{r_{ij}}\right) = \sum_j^{\text{zone III}} \frac{Q_iQ_j}{r_{ij}^3}\mathbf{r}_{ij} \qquad (16)$$

where Q_i and Q_j are the partial atomic charges of atoms i and j, respectively, $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is the vector of the relative atomic positions, and r_{ij} is their internuclear separation. The sum involving index j runs over atoms of the external environment (i.e., of zone III). If the atoms of the projection buffer (zone IIIa) are neglected and only the forces arising from the atoms of the external environment that are variationally projected are considered (i.e., zone IIIb), the exact force on atom i, designated $\mathbf{f'}_i^{\text{exact}}$, is given by

$$\mathbf{f'}_i^{\text{exact}} = \sum_j^{\text{zone IIIb}} -\nabla_i\left(\frac{Q_iQ_j}{r_{ij}}\right) = \sum_j^{\text{zone IIIb}} \frac{Q_iQ_j}{r_{ij}^3}\mathbf{r}_{ij} \qquad (17)$$

where the sum involving index j runs over atoms of the external environment (zone IIIb) *not including those in the projection buffer*.

Let the external forces calculated with the VEP or VEP−RVM methods corresponding to the exact forces $\mathbf{f}_i^{\text{exact}}$ and $\mathbf{f'}_i^{\text{exact}}$ be designated as $\mathbf{f}_i$ and $\mathbf{f'}_i$, respectively. The corresponding errors (ERR_i and ERR'_i) in the electrostatic force on atom i are given by

$$\text{ERR}_i = |\mathbf{f}_i - \mathbf{f}_i^{\text{exact}}| \qquad (18)$$

$$\text{ERR'}_i = |\mathbf{f'}_i - \mathbf{f'}_i^{\text{exact}}| \qquad (19)$$

The average error over the entire reaction zone, or over radial shells of the reaction zone, is obtained by straight averaging of the corresponding ERR_i or ERR'_i values.

The magnitude of the force errors depicted by the ERR_i or ERR'_i values may appear smaller in comparison with the

TABLE 2: Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using the Unconstrained Direct VEP Method[a]

	$\Delta R_{pb} = 0.0$ Å		$\Delta R_{pb} = 2.0$ Å		$\Delta R_{pb} = 4.0$ Å	
N_γ	Pt−Gau	Gau−Gau	Pt−Gau	Gau−Gau	Pt−Gau	Gau−Gau
14	4.89E−01	5.01E−01	4.07E−01	3.96E−01	2.87E−01	2.71E−01
	4.89E−01	*5.01E−01*	*4.99E−01*	*4.85E−01*	*3.79E−01*	*3.57E−01*
26	4.07E−01	4.17E−01	2.85E−01	2.59E−01	2.16E−01	2.06E−01
	4.07E−01	*4.17E−01*	*3.50E−01*	*3.18E−01*	*2.85E−01*	*2.72E−01*
50	3.21E−01	3.18E−01	2.10E−01	1.85E−01	9.92E−02	9.54E−02
	3.21E−01	*3.18E−01*	*2.58E−01*	*2.27E−01*	*1.31E−01*	*1.26E−01*
110	2.14E−01	2.06E−01	8.79E−02	7.66E−02	3.07E−02	3.06E−02
	2.14E−01	*2.06E−01*	*1.08E−01*	*9.39E−02*	*4.06E−02*	*4.04E−02*
194	1.85E−01	1.63E−01	4.22E−02	3.90E−02	9.85E−03	9.85E−03
	1.85E−01	*1.63E−01*	*5.17E−02*	*4.78E−02*	*1.30E−02*	*1.30E−02*
302	1.56E−01	1.28E−01	1.85E−02	1.80E−02	3.78E−03	3.78E−03
	1.56E−01	*1.28E−01*	*2.26E−02*	*2.21E−02*	*4.98E−03*	*4.98E−03*
434	1.37E−01	1.15E−01	8.62E−03	8.56E−03	1.77E−03	1.77E−03
	1.37E−01	*1.15E−01*	*1.06E−02*	*1.05E−02*	*2.33E−03*	*2.33E−03*
590	1.21E−01	9.58E−02	4.77E−03	4.76E−03	1.11E−03	1.11E−03
	1.21E−01	*9.58E−02*	*5.85E−03*	*5.84E−03*	*1.46E−03*	*1.46E−03*
770	9.70E−02	8.35E−02	2.65E−03	2.65E−03	7.91E−04	7.91E−04
	9.70E−02	*8.35E−02*	*3.25E−03*	*3.25E−03*	*1.04E−03*	*1.04E−03*
974	8.06E−02	6.74E−02	1.57E−03	1.57E−03	5.16E−04	5.16E−04
	8.06E−02	*6.74E−02*	*1.92E−03*	*1.92E−03*	*6.81E−04*	*6.81E−04*
1202	7.54E−02	6.50E−02	9.88E−04	9.88E−04	3.77E−04	3.77E−04
	7.54E−02	*6.50E−02*	*1.21E−03*	*1.21E−03*	*4.98E−04*	*4.98E−04*

[a] This table compares the accuracy at various discretization levels (N_γ) of using two different representations for the matrix elements involved in the VEP method: a point charge−Gaussian interaction (Pt−Gau, see eq 13) and a Gaussian−Gaussian interaction (Gau−Gau, see eq 14). Shown are the average relative force errors (see text) for the atoms of the reaction zone (zone I) due to the atoms of the external environment (zones IIIa and IIIb) for projection buffers of $\Delta R_{pb} = 0.0$, 2.0, and 4.0 Å. Also shown in *italics* are the relative force errors due only to the atoms that are projected (zone IIIb), i.e., neglecting the atoms in the projection buffer (zone IIIa). For an illustration of the different zones of the activated dynamics simulation model, see Figure 2.

magnitude of the *total* electrostatic force on any particular atom of the reaction zone since these forces are dominated by short-ranged interactions that do not involve atoms of the external environment [For the test systems examined in the present work, the long-range force accounts for approximately 20% of the total *electrostatic* force.] However, in molecular simulations, these short-ranged forces are largely counterbalanced by short-ranged repulsive forces that are high frequency in nature and quickly average to much smaller values. The forces due to the external environment, on the other hand, are very low frequency (or zero frequency) in nature and are ever-present throughout the simulation. Consequently, for the purposes of error analysis, it is more fitting to consider the *relative error* (RELE) in the external electrostatic force as defined by

$$\mathrm{RELE}_i = \mathrm{ERR}_i / \langle \mathrm{ERR} \rangle \tag{20}$$

$$\mathrm{RELE}'_i = \mathrm{ERR}'_i / \langle \mathrm{ERR}' \rangle \tag{21}$$

where $\langle \mathrm{ERR} \rangle$ and $\langle \mathrm{ERR}' \rangle$ are the average ERR_i and ERR'_i values, respectively, over the entire reaction zone (zone I).

4. Results

4.1. Discretization of the Projection Surfaces. The VEP and VEP−RVM procedures introduced in the present work rely critically on the ability to discretize the projection surfaces in a robust, systematically improvable fashion, such that a high degree of accuracy is obtained with a relatively small number of surface elements. Toward this end, a series of discretized spherical cavities have been designed based on the points and weights of high-order angular quadrature rules for spherical harmonic functions under octahedral symmetry[42] (see section 3.1 for further details). The accuracy of the angular integration and the projected surface charge for a unit charge at the origin are shown in Table 1. Note that for discretization levels of 302, 590, and 1202 points/sphere, the angular quadrature rules from which the points were based reproduce the L_2 orthonormality conditions (eq 2) up to angular momentum values $l = 14$, 20 and 29, respectively. The relative root-mean-square error in the surface charge remains fairly constant as a function of the discretization level ranging from 6.9×10^{-3} to 4.5×10^{-4}. Each surface element is modeled as a spherical Gaussian function with exponent related to the angular quadrature weight according to eq 4. The same type of Gaussian surface elements have also been applied to discretize molecular surfaces in a smooth analytic solvation model[38] based on an electrostatic conductor variational principle.

4.2. Matrix Elements between Atoms and the VEP Surface. The VEP method and VEP−RVM procedures require that matrix elements between the Gaussian surface elements and the atomic point charges be evaluated, as discussed in detail in section 3.1. This interaction takes the form of a Coulomb-like potential modified by an error function according to eq 13, which represents the electrostatic interaction between a point charge and a normalized Gaussian (eq 3). As mentioned previously, application of this form to the VEP and VEP−RVM procedures results in a slight overestimation of the projected surface charge in the limit that an atom is directly superimposed on top of a Gaussian surface element. Ideally, a method that produces a unit coefficient at that surface element is preferred. This limit can be realized if the point charge−Gaussian (pt−Gau) interaction is replaced by a Gaussian−Gaussian (Gau−Gau) interaction according to eq 14.

A comparison of results for the direct VEP method using the point charge−Gaussian (eq 13) and Gaussian−Gaussian (eq 14) interaction models is shown in Table 2. In all cases, with the exception of the two lowest discretization levels with no projection buffer (i.e., $N_\gamma = 14$, 20 and $\Delta R_{pb} = 0.0$ Å), the Gau−Gau interaction potential leads to slightly improved accuracy (lower relative force error). The difference between

TABLE 3: Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using the Unconstrained and Constrained Direct VEP Method with $\Delta R_{pb} = 0.0$ Å[a]

	constraint (l value)						
N_γ	none	0	2	4	6	8	10
14	5.01E−01	5.01E−01	4.57E−01				
	5.01E−01	*5.01E−01*	*4.57E−01*				
26	4.17E−01	4.17E−01	3.82E−01				
	4.17E−01	*4.17E−01*	*3.82E−01*				
50	3.18E−01	3.18E−01	2.68E−01	2.24E−01			
	3.18E−01	*3.18E−01*	*2.68E−01*	*2.24E−01*			
110	2.06E−01	2.06E−01	1.81E−01	1.46E−01	1.12E−01	1.05E−01	
	2.06E−01	*2.06E−01*	*1.81E−01*	*1.46E−01*	*1.12E−01*	*1.05E−01*	
194	1.63E−01	1.63E−01	1.37E−01	1.13E−01	8.85E−02	8.12E−02	8.04E−02
	1.63E−01	*1.63E−01*	*1.37E−01*	*1.13E−01*	*8.85E−02*	*8.12E−02*	*8.04E−02*
302	1.28E−01	1.28E−01	1.06E−01	8.80E−02	6.97E−02	6.46E−02	6.42E−02
	1.28E−01	*1.28E−01*	*1.06E−01*	*8.80E−02*	*6.97E−02*	*6.46E−02*	*6.42E−02*
434	1.15E−01	1.15E−01	1.05E−01	7.49E−02	5.78E−02	5.37E−02	5.32E−02
	1.15E−01	*1.15E−01*	*1.05E−01*	*7.49E−02*	*5.78E−02*	*5.37E−02*	*5.32E−02*
590	9.58E−02	9.58E−02	8.25E−02	6.51E−02	5.03E−02	4.19E−02	4.17E−02
	9.58E−02	*9.58E−02*	*8.25E−02*	*6.51E−02*	*5.03E−02*	*4.19E−02*	*4.17E−02*
770	8.35E−02	8.35E−02	7.34E−02	5.08E−02	4.26E−02	4.21E−02	4.14E−02
	8.35E−02	*8.35E−02*	*7.34E−02*	*5.08E−02*	*4.26E−02*	*4.21E−02*	*4.14E−02*
974	6.74E−02	6.74E−02	6.03E−02	5.10E−02	3.92E−02	3.67E−02	3.64E−02
	6.74E−02	*6.74E−02*	*6.03E−02*	*5.10E−02*	*3.92E−02*	*3.67E−02*	*3.64E−02*
1202	6.50E−02	6.50E−02	5.92E−02	4.22E−02	3.28E−02	2.68E−02	2.66E−02
	6.50E−02	*6.50E−02*	*5.92E−02*	*4.22E−02*	*3.28E−02*	*2.68E−02*	*2.66E−02*

[a] This table compares the accuracy at various discretization levels (N_γ) of the direct VEP method with no projection buffer ($\Delta R_{pb} = 0.0$ Å). Relative force errors for the direct VEP method in the absence of constraints ("none") and with constraints on the far-field Taylor moments up to order l are compared. Also shown in *italics* are the relative force errors due only to the atoms that are projected (zone IIIb)—which for this case the values are identical since there is no projection buffer. See text for further details.

the errors of the pt−Gau and Gau−Gau protocols becomes less as the level of discretization increases. This results because as the discretization level increases, the Gaussian exponents of the individual surface elements become larger, and the spatial range where the pt−Gau and Gau−Gau interactions differ becomes smaller. As soon as even a small projection buffer is introduced, providing a small space between the projection surface and the atoms that are being projected, the relative force errors rapidly converge. With a 2.0 Å projection buffer ($\Delta R_{pb} = 2.0$ Å), there is negligible difference between protocols for discretization levels greater than 590 points, and with a 4.0 Å projection buffer ($\Delta R_{pb} = 4.0$ Å), there is negligible difference for discretization levels greater than 110 points. As a result of the general slight improvement obtained using the Gau−Gau interaction model, this protocol was adopted for the VEP and VEP−RVM methods and applied in numerical results for the remainder of the manuscript.

4.3. The Direct VEP Method. The direct VEP method uses a single-step variational electrostatic projection of an external charge distribution onto the projection surface to obtain a set of coefficients for the expansion of the Gaussian surface elements from which the projected electrostatic potential inside the surface can be derived. In the present work, the projection surface was chosen, after testing, to be placed at the outer boundary of the stochastic buffer zone (i.e., $r_\gamma = R_{II}$ in Figure 2). Since the γ surface is discretized by a finite number of elements (N_γ), at a given discretization level, the average distance between surface elements varies linearly with the radius r_γ. Consequently, the greatest errors occur in the projection or evaluation of the potential (or force) of atoms that lie closest to the discretized surface (especially in the vicinity between surface elements). The accuracy of the forces in the stochastic buffer region are of minor importance due to the combination of random forces and restraining potentials used in this region during simulation. Hence, the stochastic region also serves as a buffer zone that leads to improved accuracy for the evaluation of the electrostatic potential from the discretized projection surface in the reaction zone.

With placement of the projection surface immediately outside of the stochastic buffer (zone II), the atoms in the external environment may lie very close to the surface and result in errors in their variational projection at low discretization levels. To balance the projection errors with computational effort, a projection buffer is introduced whereby atoms within ΔR_{pb} of the projection surface are treated explicitly and only atoms of the external environment outside this distance enter the variational electrostatic projection (zone IIIb in Figure 2).

The direct VEP method can be applied with a set of linear variational constraints. In particular, constrained variational electrostatic projection can be performed whereby the far-field (Taylor) multipole moments of the charge distribution of the external environment are preserved up to a certain order of angular momentum l, leading to a total of $(l + 1)^2$ linear constraints.

Tables 3−5 compare relative force errors for the unconstrained and constrained VEP method at different surface discretization levels for projection buffer values of $\Delta R_{pb} = 0.0$ Å (i.e., no projection buffer, Table 3), $\Delta R_{pb} = 2.0$ Å (Table 4) and $\Delta R_{pb} = 4.0$ Å (Table 5). At low discretization levels, the number of variational degrees of freedom preclude application of high-order variational constraints. For discretization levels of 194 points or greater, constraints up to $l = 10$ can easily be accommodated.

In the absence of a projection buffer, the RELE and RELE′ values are identical (Table 3). At fairly low discretization (302 points), the RELE for the unconstrained VEP method is 1.28×10^{-1}, or about 13%. At constant 302-point discretization, the RELE steadily decreases as higher order constraints are applied reaching a value of around 6.4% error for $l = 10$. The unconstrained VEP error values decrease steadily with increasing discretization going from 13% (302 points), to 9.6% (590

TABLE 4: Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using the Unconstrained and Constrained Direct VEP Method with $\Delta R_{pb} = 2.0$ Å[a]

	constraint (l value)						
N_γ	none	0	2	4	6	8	10
14	3.96E−01	3.95E−01	3.93E−01				
	4.85E−01	*4.85E−01*	*4.82E−01*				
26	2.59E−01	2.59E−01	2.44E−01				
	3.18E−01	*3.18E−01*	*2.99E−01*				
50	1.85E−01	1.85E−01	1.65E−01	1.33E−01			
	2.27E−01	*2.27E−01*	*2.02E−01*	*1.63E−01*			
110	7.66E−02	7.66E−02	7.63E−02	6.64E−02	5.69E−02	5.00E−02	
	9.39E−02	*9.39E−02*	*9.35E−02*	*8.14E−02*	*6.98E−02*	*6.13E−02*	
194	3.90E−02	3.90E−02	3.68E−02	3.14E−02	2.46E−02	2.31E−02	2.28E−02
	4.78E−02	*4.78E−02*	*4.51E−02*	*3.85E−02*	*3.02E−02*	*2.83E−02*	*2.79E−02*
302	1.80E−02	1.80E−02	1.66E−02	1.43E−02	1.16E−02	1.06E−02	1.04E−02
	2.21E−02	*2.21E−02*	*2.04E−02*	*1.76E−02*	*1.42E−02*	*1.30E−02*	*1.28E−02*
434	8.56E−03	8.56E−03	7.59E−03	6.22E−03	5.75E−03	5.47E−03	5.32E−03
	1.05E−02	*1.05E−02*	*9.30E−03*	*7.62E−03*	*7.05E−03*	*6.71E−03*	*6.53E−03*
590	4.76E−03	4.76E−03	4.37E−03	3.77E−03	3.44E−03	3.11E−03	3.10E−03
	5.84E−03	*5.84E−03*	*5.35E−03*	*4.62E−03*	*4.22E−03*	*3.82E−03*	*3.80E−03*
770	2.65E−03	2.65E−03	2.47E−03	2.27E−03	2.13E−03	1.90E−03	1.89E−03
	3.25E−03	*3.25E−03*	*3.03E−03*	*2.78E−03*	*2.61E−03*	*2.33E−03*	*2.32E−03*
974	1.57E−03	1.57E−03	1.49E−03	1.27E−03	1.15E−03	1.07E−03	1.06E−03
	1.92E−03	*1.92E−03*	*1.83E−03*	*1.55E−03*	*1.41E−03*	*1.31E−03*	*1.30E−03*
1202	9.88E−04	9.88E−04	9.66E−04	8.59E−04	7.97E−04	7.34E−04	7.28E−04
	1.21E−03	*1.21E−03*	*1.18E−03*	*1.05E−03*	*9.78E−04*	*9.00E−04*	*8.93E−04*

[a] This table compares the accuracy at various discretization levels (N_γ) of the direct VEP method with projection buffer of $\Delta R_{pb} = 2.0$ Å. Relative force errors for the direct VEP method in the absence of constraints ("none") and with constraints on the far-field Taylor moments up to order l are compared. Also shown in *italics* are the relative force errors due only to the atoms that are projected (zone IIIb). See text for further details.

TABLE 5: Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using the Unconstrained and Constrained Direct VEP Method with $\Delta R_{pb} = 4.0$ Å[a]

	constraint (l value)						
N_γ	none	0	2	4	6	8	10
14	2.71E−01	2.71E−01	2.67E−01				
	3.57E−01	*3.57E−01*	*3.52E−01*				
26	2.06E−01	2.06E−01	2.00E−01				
	2.72E−01	*2.72E−01*	*2.65E−01*				
50	9.54E−02	9.54E−02	9.03E−02	8.19E−02			
	1.26E−01	*1.26E−01*	*1.19E−01*	*1.08E−01*			
110	3.06E−02	3.06E−02	2.93E−02	2.65E−02	2.34E−02	2.24E−02	
	4.04E−02	*4.04E−02*	*3.87E−02*	*3.49E−02*	*3.09E−02*	*2.96E−02*	
194	9.85E−03	9.85E−03	9.56E−03	8.82E−03	7.36E−03	7.18E−03	6.97E−03
	1.30E−02	*1.30E−02*	*1.26E−02*	*1.16E−02*	*9.72E−03*	*9.47E−03*	*9.20E−03*
302	3.78E−03	3.78E−03	3.73E−03	3.46E−03	2.92E−03	2.68E−03	2.50E−03
	4.98E−03	*4.98E−03*	*4.93E−03*	*4.57E−03*	*3.85E−03*	*3.53E−03*	*3.30E−03*
434	1.77E−03	1.77E−03	1.75E−03	1.60E−03	1.33E−03	1.14E−03	1.04E−03
	2.33E−03	*2.33E−03*	*2.32E−03*	*2.11E−03*	*1.75E−03*	*1.51E−03*	*1.37E−03*
590	1.11E−03	1.11E−03	1.09E−03	9.86E−04	8.19E−04	6.88E−04	6.10E−04
	1.46E−03	*1.46E−03*	*1.43E−03*	*1.30E−03*	*1.08E−03*	*9.08E−04*	*8.06E−04*
770	7.91E−04	7.91E−04	7.61E−04	6.84E−04	5.81E−04	4.74E−04	4.18E−04
	1.04E−03	*1.04E−03*	*1.01E−03*	*9.03E−04*	*7.67E−04*	*6.26E−04*	*5.52E−04*
974	5.16E−04	5.16E−04	5.06E−04	4.61E−04	3.78E−04	3.18E−04	2.82E−04
	6.81E−04	*6.81E−04*	*6.68E−04*	*6.09E−04*	*5.00E−04*	*4.19E−04*	*3.72E−04*
1202	3.77E−04	3.77E−04	3.71E−04	3.37E−04	2.77E−04	2.32E−04	2.06E−04
	4.98E−04	*4.98E−04*	*4.89E−04*	*4.45E−04*	*3.66E−04*	*3.06E−04*	*2.72E−04*

[a] This table compares the accuracy at various discretization levels (N_γ) of the direct VEP method with projection buffer of $\Delta R_{pb} = 4.0$ Å. Relative force errors for the direct VEP method in the absence of constraints ("none") and with constraints on the far-field Taylor moments up to order l are compared. Also shown in *italics* are the relative force errors due only to the atoms that are projected (zone IIIb). See text for further details.

points), to 6.5% (1202 points). The corresponding constrained values ($l = 10$) similarly decrease, having RELE values of 6.4, 4.2, and 2.7 for 302, 590, and 1202 points, respectively. Considering that the projection surface at the highest discretization level contains less than 10% of the total number of points of the external charge distribution, these errors are fairly good—much better that those of any cutoff method (see section 5.1). Nonetheless, it would be preferable to have relative errors below 1% for accurate simulations.

Table 4 shows the corresponding RELE and RELE′ values using a 2 Å projection buffer. The unconstrained VEP RELE values are 1.8%, 0.48%, and 0.099% for 302, 590, and 1202 points, respectively. If the contribution of the atoms in the projection buffer region are excluded (since they are not variationally projected, but instead treated explicitly and hence exactly), the RELE′ values are slightly larger with values of 2.2%, 0.58%, and 0.12% for 302, 590, and 1202 points, respectively. These values decrease slightly (less than a factor

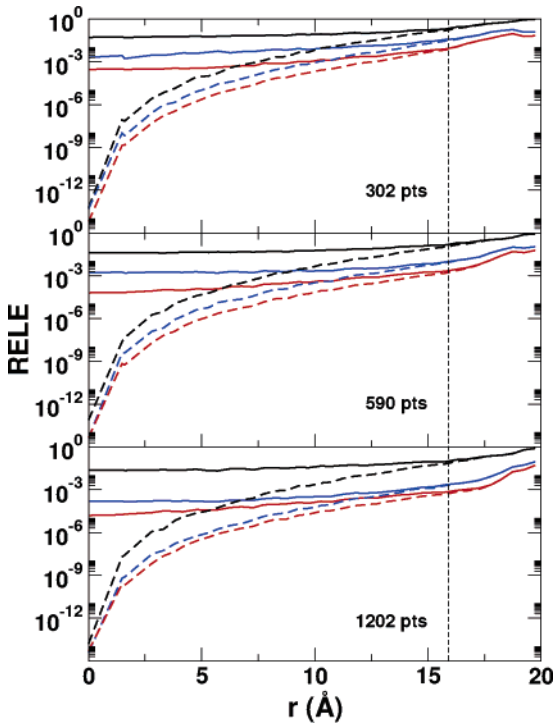


Figure 3. Distribution of the average RELE values as a function of radial distance r from the origin of the reaction zone for 302 (top), 590 (middle) and 1202 (bottom) points for the unconstrained direct VEP method (solid lines) and constrained direct VEP method (dashed lines). Results for projection buffer values $\Delta R_{pb} = 0.0$ Å (black), 2.0 Å (blue) and 4.0 Å (red) are shown. A vertical dashed line indicates the edge of the reaction zone (R_I).

of 2) when variational constraints up to $l = 10$ are applied—a notable feature being that with constraints the RELE at 302 points (1.04×10^{-2}) is already at approximately the 1% relative error level. These relative force error values represent a considerable improvement over the VEP method without projection buffer.

Table 5 shows the corresponding RELE and RELE′ values using a 4 Å projection buffer. The unconstrained VEP RELE values are 0.38%, 0.11% and 0.038% for 302, 590, and 1202 points, respectively. Application of constraints up to $l = 10$ lead to a decrease in relative errors by about 30% (RELE values of 0.25%, 0.061%, and 0.021% for 302, 590, and 1202 points, respectively). The RELE′ values arising from the projection only are increased from the corresponding RELE values by a factor of approximately 1.3 for both the unconstrained and constrained VEP methods with $\Delta R_{pb} = 4.0$ Å. These errors at the lower discretization level (302 points) are below 1%, and at medium discretization level (590 points) are around 0.1%, and at the high discretization level (1202 points) are below 0.1%. This suggests that the direct VEP method, with or without constraints, can be made considerably accurate with increasing discretization and a fairly minimal projection buffer.

Figure 3 shows a distribution of the average RELE values as a function of radial distance from the origin of the reaction zone for 302, 590, and 1202 points for the unconstrained direct and constrained direct methods. As expected, increasing the projection buffer ΔR_{pb} from a value of 0.0 Å (black) to 4.0 Å (red) shifts the curves to lower RELE values. A notable feature is that inclusion of constraints greatly reduces the error at the center of the reaction zone by several orders of magnitude, but does little to improve errors toward the boundary at R_I.

4.4. The VEP−RVM Method. In both the direct VEP and VEP−RVM procedures, the electrostatic potential and forces due to the charge distribution of the external environment (zone IIIb in Figure 2) are determined in the dynamical region of the system (zones I and II in Figure 2) from the γ charge distribution. The discretized γ surface is therefore used as a basis for evaluation of the energy and forces due to atoms outside of the dynamical region. As pointed out earlier, high accuracy is particularly important in the reaction zone (zone I in Figure 2) that undergoes unconstrained Hamiltonian dynamics and is used to construct free energy profiles and analyze the equilibrium or dynamical properties associated with the reaction of interest.

The direct VEP method becomes exact in the limit that the projection surface discretization forms a complete basis for the variational procedure. This limit is never realized, however, and for the finite discretization levels tested in the present work, error arises from to two main sources: (1) projection of external charges of the environment that reside close to the outside of the projection surface and (2) evaluation of the potential and forces at atomic positions close to the inside of the projection surface. These errors are directly related to atoms located at distances from the projection surface (either on the inside or the outside) that are comparable to or less than the average spacing between surface elements.

In accord with the variational principle, the location of the projection surface is constrained to lie within the region between the external charges that are to be projected and the atomic positions where electrostatic potential and forces are to be evaluated. In the direct VEP method, this implies the γ surface must reside within the inner and outer boundary of the projection buffer (i.e., within zone IIIa of Figure 2). The convention that was adopted for the direct VEP method was to place the γ surface at the boundary directly between the stochastic and projection buffer zones (i.e., $r_\gamma = R_{II}$). As illustrated in the previous section, relatively high accuracy was attained (less than 0.04% RELE in the force) when the stochastic and projection buffer distances were set equal at 4.0 Å.

The analysis of the errors in the direct VEP method lead to the idea that one might achieve higher accuracy with a dual-surface approach: one that simultaneously utilized an intermediate forward projection surface (designated the ω surface in Figure 2) that was farther interior to the external charge distribution of the environment and a γ surface that was buffered further from the reaction zone for evaluation of the potential and forces. This idea would necessitate an additional procedure whereby the forward projected surface charge on the ω surface was variationally reversed-mapped onto the γ surface. This dual-surface approach is designated the VEP−RVM method. When the ω and γ surfaces coincide and are identically discretized, the reverse variational mapping is just an identity operation, (i.e., the direct VEP method is a particular case of the VEP−RVM method, when $r_\omega = r_\gamma$).

4.4.1. Optimal Placement of the ω Surface in the VEP−RVM Method. The convention that is used in the present work is that the γ surface in the VEP−RVM procedure is taken to be at the outer edge of the projection buffer (zone IIIa in Figure 2), i.e., as far away from the reaction zone as possible without penetrating the external environment (zone IIIb). Since the forward projection ω surface is not used directly for evaluation of the potential or forces, but rather as an intermediate for determination of γ surface charge, there is no restriction that it needs to reside outside of the reaction zone (zone I in Figure 2). Consequently, there remains the question as to where to best place the ω surface to affect the maximum accuracy for the VEP−RVM procedure. The optimal placement of the ω surface

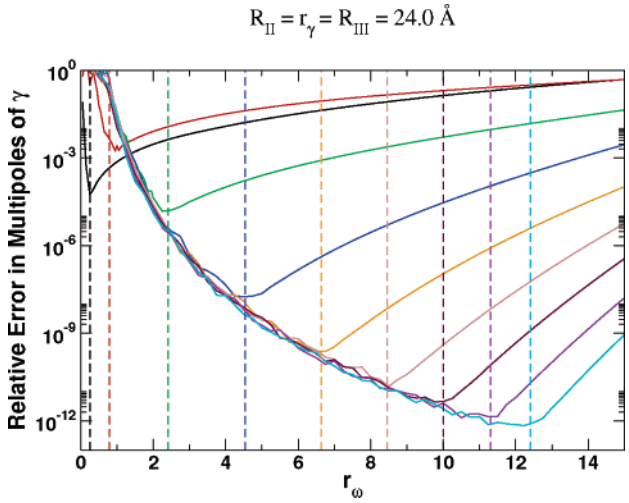


Figure 4. Relative errors in the charge weighted Taylor moments of the γ surface with respect to the location of the ω surface for $N = 50$, 110, 194, 302, 434, 590, 770, 974, and 1202. Dashed vertical lines indicate the optimal calculated locations based on eq 24.

is expected to be a function of discretization level and distance from the γ surface. In the present work, the convention is adopted that the discretization level (number of surface elements) of the ω surface is always taken to be the same as for the γ surface ($N_\omega = N_\gamma = N$). Moreover, as the number of surface elements becomes infinite (i.e., as the basis for expansion of the variational electrostatic projection becomes complete), the positions of the ω and γ surfaces should coincide.

To access the optimal placement of the ω surface, exact relations on the VEP−RVM method were considered. In particular, the final projected γ surface charge should have the same far-field expansion coefficients (Taylor moments, see eq 51) about the origin of the reaction zone as those of the external charge distribution of the environment. These far-field expansion coefficients were previously applied as constraints in the direct VEP approach and shown in Tables 3−5.

In principle, an exact variational procedure would reproduce these coefficients exactly in the absence of explicit constraints. Figure 4 shows the relative error in the charge weighted Taylor moments of the γ surface (located at $r_\gamma = 24.0$ Å) at different discretization levels as a function of the placement of the ω surface. Dotted lines in the figure mark the position (r_ω) of the optimal ω surface that resulted in the smallest relative error in the theoretical Taylor moments (up to order $l = 10$). Clear patterns are evident as the number of surface elements increases.

4.4.2. Fitting a Functional Form for Estimation of the Optimal r_ω Positions. In this subsection, an empirical functional form is presented to estimate the optimal position, r_ω, for the ω surface in the VEP−RVM method. Consider first the average spacing between discretized surface elements on a sphere of radius R, with surface area A. In the limit that the number of discretized surface elements, N, becomes infinite, the average spacing between points uniformly distributed on the sphere is equal to $(A/N)^{1/2} = (4\pi R^2/N)^{1/2} \sim R/\sqrt{N}$. Since in practice, the number of surface elements is finite and optimal discretization requires the elements to be nonuniform (meaning not all surface elements are identical), one can instead consider an *effective spacing* between surface elements that is proportional to $R/\sqrt{N + c_0}$ where c_0 is a unitless empirical parameter that corrects for nonuniform behavior, especially at low discretization levels. This parameter will be part of the fitting procedure described below.

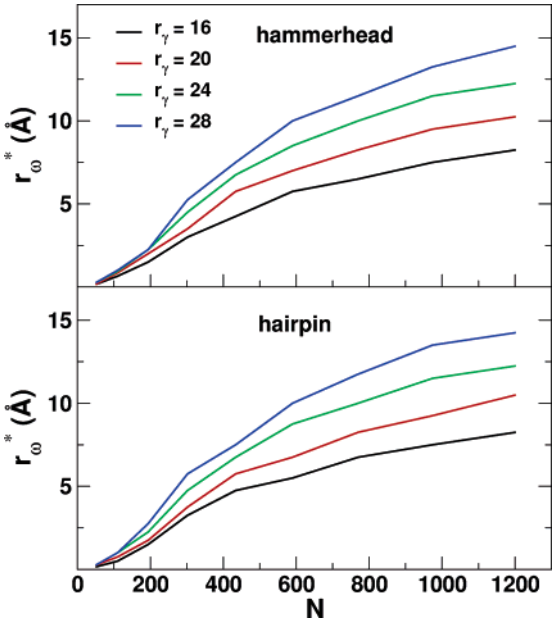


Figure 5. Optimal location r_ω^* of the ω surface with respect to N for $R_{II} = R_{III} = r_\gamma = [16, 20, 24, 28$ Å] for solvated hammerhead (top) and hairpin (bottom) ribozyme systems.

As mentioned previously, a significant component of the error in the VEP-type procedures involves a balance between the effective spacing between surface elements on the projection surface and the distance from the projection surface to the points where the electrostatic potential and forces are desired. In the VEP−RVM method, the ω surface serves as an intermediate to obtain the final projected surface charge. Hence, it is not unreasonable to expect that the spacing between the γ surface and optimally placed ω surface (r_ω) should be proportional to the effective average spacing between surface elements

$$r_\gamma - r_\omega = \frac{c_1 r_\gamma}{\sqrt{N + c_0}} \tag{22}$$

where c_1 is the proportionality constant. This leads to an empirical relation between the *scaled* position of the optimal ω surface, r_ω/r_γ:

$$\frac{r_\omega}{r_\gamma} = 1 - \frac{c_1}{\sqrt{N + c_0}} \tag{23}$$

The values of the optimal r_ω locations shown in Figure 5 for different values of r_γ (16, 20, 24, and 28 Å) are shown in their average scaled form in Figure 6 with bars indicating the standard deviation. The standard deviation values are all quite small and range from 0.001 to 0.01 Å. This suggests that a transferable empirical function might be derived to predict the optimal scaled r_ω/r_γ values for general applications of the VEP−RVM method. Fitting of the parameters c_0 and c_1 in eq 23 gives reasonable overall agreement for N values greater than 194, but begins to break down for smaller values (dashed lines, Figure 6). This observation lead to the introduction of an empirical correction term that modifies the predicted optimal scaled r_ω/r_γ values at lower discretization levels. Several forms were tried and compared, and excellent agreement was obtained using a simple 2-parameter exponential form for the correction:

$$\frac{r_\omega}{r_\gamma} = 1 - \frac{c_1 - c_2 \times \exp[-c_3 N]}{\sqrt{N + c_0}} \tag{24}$$

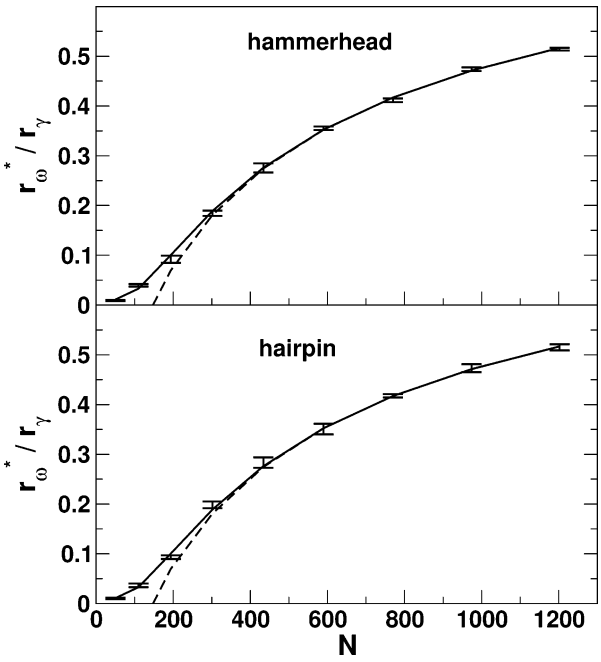


Figure 6. Average value of r^*_ω/r_γ vs N with $r_\gamma = R_{II} = R_{III}$ for hammerhead (top) and hairpin (bottom) ribozyme systems (standard deviations are indicated with error bars). Simultaneous least-squares fit to top and bottom graphs according to the functional form of eq 24 (solid line) and Eq 23 (dotted line). The unitless parameters ($c_0 - c_3$) resulting from the least-squares fit are $c_0 = 178.459$, $c_1 = 17.9528$, $c_2 = 5.27423$, and $c_3 = 0.011264$.

Fits using both eqs 23 and 24 are shown in Figure 6. The agreement between the exact and fitted optimal scaled r_ω/r_γ is excellent, and provides a general prescription for obtaining the optimal r_ω position over a wide range of surface discretization levels.

4.4.3. Comparison of the Force Errors with the VEP−RVM Method. Table 6 compares the average relative force errors for the (unconstrained) direct VEP and VEP−RVM methods for projection buffer values of $\Delta R_{pb} = 0.0$, 2.0, and 4.0 Å. The optimal r_ω values for placement of the ω surface were determined from the empirical relation given in eq 24 using the optimized parameters given in the legend of Figure 6. The relative force errors in the VEP−RVM method show marked improvement over the direct VEP method for all cases. Without projection buffer ($\Delta R_{pb} = 0.0$), the RELE values for 302, 590, and 1202 points with the VEP−RVM method (2.1, 0.5, and 0.04%, respectively) are reduced considerably from those of the direct VEP method (13, 9.6, and 6.5%, respectively). As with the direct VEP method, errors significantly improve when increasing the size of the projection buffer. With a 2 Å projection buffer, the RELE values for 302, 590, and 1202 points are 0.4%, 0.04%, and 2×10^{-3}%, respectively, and with a 4 Å projection buffer, the RELE values for 302, 590, and 1202 points are 0.09%, 5×10^{-3}%, and 1×10^{-4}%, respectively.

Figure 7 compares the convergence behavior with respect to discretization level of the direct VEP and VEP−RVM methods. It is clear from the figure that the VEP−RVM method provides considerably greater accuracy than the direct VEP method, especially at higher discretization levels, and more uniform linear error convergence as a function of the square root of the number of surface elements ($N_\gamma^{1/2}$), which is proportional to the average distance between surface elements. The linear error convergence with respect to $N_\gamma^{1/2}$ is more rapid with increased projection buffer size (ΔR_{pb}): the slopes of the best-fit lines to the VEP−RVM method (dashed lines in Figure 7) are −0.976, −0.141, and −0.175 for ΔR_{pb} values of 0.0, 2.0, and 4.0 Å, respectively. Convergence of the direct VEP method deviates from linearity at higher discretization levels, thus preventing the very high accuracy obtainable with the VEP−RVM method. For example, at the highest discretization level (1202 points), the error for the direct VEP method with a 4 Å projection buffer (0.04%) is nearly identical to the error for the VEP−RVM method without any projection buffer.

Figure 8 compares the distribution of errors as a function of radial distance from the origin for the unconstrained direct VEP and VEP−RVM methods. Whereas the errors of the direct VEP method are fairly uniform except toward in the region near to the projection surface, the errors in the VEP−RVM method are not uniform. For the VEP−RVM method, the errors within a finite region around the center of the reaction zone (zone I) are essentially exact to numerical machine precision, and then slope toward lower accuracy toward the projection surface. The errors for the VEP−RVM method are always considerably less than those of the direct VEP method over all radial distances in the reaction zone, and only converge to the direct VEP values from below very near to the projection surface (the VEP−RVM method was never observed to be less accurate than the direct VEP method). The finite region of very high accuracy with the VEP−RVM method begins at the origin and increases in extent as a function of the discretization level. Overall, the VEP−RVM method provides high accuracy over the entire range of radial points and extremely high accuracy in the most important area: the center of the reaction zone. Moreover, the unconstrained VEP−RVM method, with optimal placement of the ω surface, satisfies the constraint conditions to very high accuracy (obviating the need to employ the constraints explicitly). These observations demonstrate that the VEP−RVM method will have considerable promise in forming the basis of new generalized solvent boundary methods for QM/MM simulations.

5. Discussion

The current work focuses on the introduction of two new approaches, the direct VEP and VEP−RVM methods, for calculating the electrostatic interaction energy and forces between atoms in a dynamical region (zones I and II of Figure 2) and those of an external environment (zone III in Figure 2). These techniques represent novel computational algorithms for the calculation of electrostatic interactions at significantly reduced cost, and as such offer an important potential advancement for activated dynamics simulations. However, the methods have even more far-reaching implication. The transformation of the electrostatic interaction problem between two spatial charge distributions separated by a surface into a variational boundary value problem offers a potentially powerful tool toward the development of new macromolecular solvent-boundary potentials. The use of a *discretized surface* as a basis for a variational electrostatic projection procedure is potentially more accurate than methods based on expansions about a central point (such as generalized multipoles[31,49]), and it is more easily generalizable to arbitrary cavity shapes. Moreover, the small number of surface elements required to provide high accuracy is well within the capability of constructing more complicated linear-response functions for isotropic and anisotropic polarizability of the solvated macromolecular environment. These response functions would require matrices to be computed that have dimensions of the number of discretized surface elements, and as demonstrated in the present work, would provide very high accuracy for matrices of 302×302 to 1202×1202, that are well within the memory and computational bounds of a modern personal computer. The further development of these

TABLE 6: Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using the Unconstrained Direct VEP and VEP−RVM Methods with ΔR_{pb} = 0.0, 2.0, and 4.0 Å[a]

	ΔR_{pb} = 0.0 Å		ΔR_{pb} = 2.0 Å		ΔR_{pb} = 4.0 Å	
N	VEP	VEP−RVM	VEP	VEP−RVM	VEP	VEP−RVM
14	5.01E−01	4.56E−01	3.96E−01	3.76E−01	2.71E−01	2.21E−01
	5.01E−01	*4.56E−01*	*4.85E−01*	*4.61E−01*	*3.57E−01*	*2.92E−01*
26	4.17E−01	3.66E−01	2.59E−01	2.12E−01	2.06E−01	1.46E−01
	4.17E−01	*3.66E−01*	*3.18E−01*	*2.60E−01*	*2.72E−01*	*1.93E−01*
50	3.18E−01	2.14E−01	1.85E−01	1.43E−01	9.54E−02	5.60E−02
	3.18E−01	*2.14E−01*	*2.27E−01*	*1.76E−01*	*1.26E−01*	*7.39E−02*
110	2.06E−01	1.04E−01	7.66E−02	4.06E−02	3.06E−02	1.38E−02
	2.06E−01	*1.04E−01*	*9.39E−02*	*4.98E−02*	*4.04E−02*	*1.83E−02*
194	1.63E−01	5.66E−02	3.90E−02	1.26E−02	9.85E−03	3.27E−03
	1.63E−01	*5.66E−02*	*4.78E−02*	*1.55E−02*	*1.30E−02*	*4.32E−03*
302	1.28E−01	2.09E−02	1.80E−02	3.58E−03	3.78E−03	8.51E−04
	1.28E−01	*2.09E−02*	*2.21E−02*	*4.39E−03*	*4.98E−03*	*1.12E−03*
434	1.15E−01	9.60E−03	8.56E−03	1.36E−03	1.77E−03	2.15E−04
	1.15E−01	*9.60E−03*	*1.05E−02*	*1.67E−03*	*2.33E−03*	*2.84E−04*
590	9.58E−02	4.57E−03	4.76E−03	4.38E−04	1.11E−03	4.99E−05
	9.58E−02	*4.57E−03*	*5.84E−03*	*5.38E−04*	*1.46E−03*	*6.58E−05*
770	8.35E−02	2.09E−03	2.65E−03	1.41E−04	7.91E−04	1.22E−05
	8.35E−02	*2.09E−03*	*3.25E−03*	*1.73E−04*	*1.04E−03*	*1.60E−05*
974	6.74E−02	9.68E−04	1.57E−03	5.22E−05	5.16E−04	3.30E−06
	6.74E−02	*9.68E−04*	*1.92E−03*	*6.41E−05*	*6.81E−04*	*4.36E−06*
1202	6.50E−02	4.27E−04	9.88E−04	1.69E−05	3.77E−04	9.61E−07
	6.50E−02	*4.27E−04*	*1.21E−03*	*2.07E−05*	*4.98E−04*	*1.27E−06*

[a] This table compares the accuracy at various discretization levels (*N*) of the unconstrained direct VEP and VEP−RVM methods with projection buffers of ΔR_{pb} = 0.0, 2.0, and 4.0 Å. For the direct VEP method, r_γ was set at R_{II} = 20 Å, and for the VEP−RVM method, r_γ was set at $R_{III} = R_{II} + \Delta R_{pb}$. The optimal r_ω^* values for the VEP−RVM method were obtained from fitting to eq 23 (see Figure 6). Also shown in *italics* are the relative force errors due only to the atoms that are projected (zone IIIb). See text for further details.

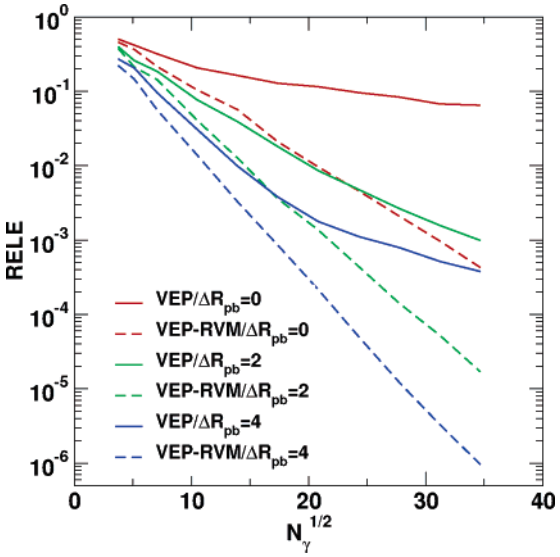


Figure 7. Comparison of average relative RMS force errors (RELE) as a function of the square root of the number of surface elements $(N_\gamma)^{1/2}$ for the unconstrained VEP (solid lines) and VEP−RVM (dashed lines) methods with projection buffer distances (ΔR_{pb}) values of 0.0 (black), 2.0 (blue), and 4.0 Å (red).

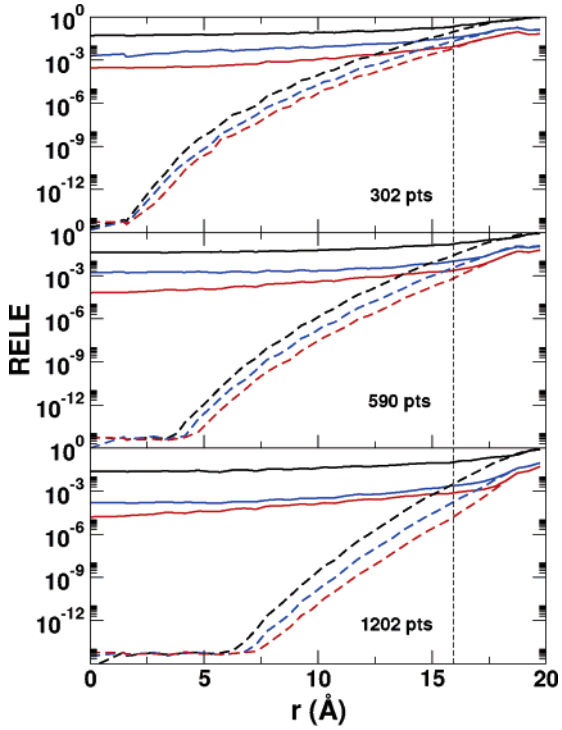


Figure 8. Distribution of the average RELE values as a function of radial distance *r* from the origin of the reaction zone for 302 (top), 590 (middle), and 1202 (bottom) points for the unconstrained direct VEP method (solid lines) and unconstrained VEP−RVM method (dashed lines). Results for projection buffer values ΔR_{pb} = 0.0 (black), 2.0 (blue), and 4.0 Å (red) are shown. Vertical dashed line indicates the edge of the reaction zone (R_I).

methods is forthcoming; however, these methods all depend critically on the demonstration and characterization of the accuracy of the fundamental variational electrostatic projection techniques that form their basis. Consequently the present work is of fundamental importance and an essential first step toward the design of new variational macromolecular solvent boundary methods for hybrid QM/MM simulations.

5.1. Comparison with Cutoff Methods. It is important to establish within the context of more conventional treatments of electrostatics the degree of accuracy that can be obtained with a new method. It has long been established, especially for simulations of highly charged systems, that the use of electrostatic cutoffs introduces severe errors in dynamical simulations and free energies of biomolecular systems.[18,23,50] Despite these clear demonstrations, a great number of present-day QM/MM calculations of enzyme reactions still utilize electrostatic cutoffs. In any event, the design and implementation of new methods

TABLE 7: Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using Different Electrostatic Cutoffs

Cutoff Distance (Å)								
8.0	12.0	16.0	20.0	24.0	28.0	32.0	36.0	40.0
Group-Based Cutoffs: No Switching Region								
9.35E−01	7.94E−01	6.60E−01	5.21E−01	4.13E−01	2.86E−01	2.09E−01	1.49E−01	9.44E−02
Group-Based Cutoffs: 4 Å Switching Region								
9.68E−01	8.36E−01	6.95E−01	5.60E−01	4.29E−01	3.10E−01	2.19E−01	1.48E−01	9.90E−02
Atom-Based Cutoffs: No Switching Region								
1.03E+00	9.48E−01	8.36E−01	7.17E−01	6.09E−01	4.87E−01	3.97E−01	3.02E−01	2.22E−01
Atom-Based Cutoffs: 4 Å Switching Region								
9.53E−01	8.03E−01	6.40E−01	5.08E−01	3.87E−01	2.78E−01	1.80E−01	1.10E−01	6.90E−02

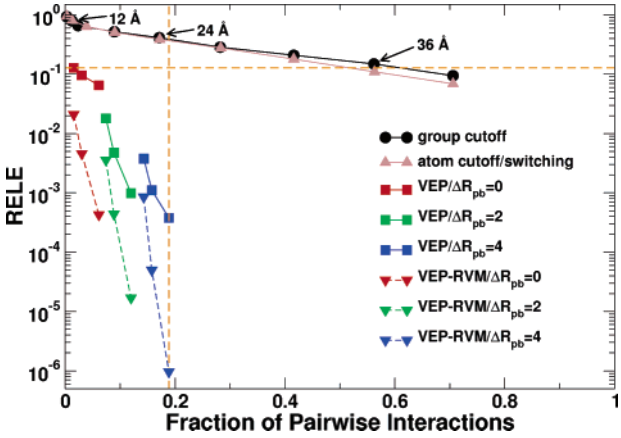


Figure 9. Average RELE vs fraction of evaluated pairwise interactions. Complete evaluation of the force on all atoms in region I due to all atoms in region III corresponds to pairwise evaluation fraction of 1.0. RELE using group and atom based cutoffs, as well as the VEP and VEP−RVM methods with 302, 590, and 1202 surface points is shown.

that offer efficient alternatives to this part of the computational chemistry community represents and important advance.

The relative force error that arises from using a group-based electrostatic cutoff[51] for the interactions between the atoms of the reaction zone and those of the environment (and neglecting the even more severe cutoff errors that arise from the interactions within the reaction zone) are listed in Table 7. With a 12.0 Å group-based cutoff, the average relative force error of the atoms in the reaction zone due to the atoms of the environment is around 80%. Doubling and tripling this cutoff value to 24.0 and 36.0 Å decreases the average relative error to 41% and 15% error, respectively. Hence, a 36.0 Å group-based cutoff provides an error that is greater than the error for the lowest level direct VEP method with 302 points and no projection buffer (13%).

The CHARMM c27 force field for nucleic acids[12] utilizes neutral (or integer) charged groups to represent electrostatic interactions. The relative force errors correspond to the standard group-based cutoff convention employed by CHARMM,[51] and atom-based cutoff method. The group-based and atom-based cutoff methods were performed with both hard truncation and with the force switch method used in CHARMM applied over the last 4.0 Å.

A comparison of the average relative force errors for the direct VEP, VEP−RVM and group-based cutoff methods as a function of the number of pairwise interactions (either atom−surface element or atom−atom) are shown in Figure 9. With smaller cutoffs, the number of pairwise interactions with the atoms of the *external* environment (zone III) are small since many of the atoms in the reaction zone do not interact with the external atoms at all. With a 24 Å group-based cutoff, all the atoms of the reaction zone have at least some interaction with atoms of the environment (RELE 41%). The most accurate method that was presented, the VEP−RVM method with 1202 points and a 4 Å projection buffer, has an RELE value of less than 1×10^{-4} %, and the number of pairwise interactions is just slightly more than that of a 24 Å group-based or atom-based cutoff. The VEP−RVM method with 590 points and no projection buffer provides an accuracy of less than 1% and has a number of pairwise interactions less than a 16 Å atom-based cutoff. These results help to put into more tangible perspective the degree of accuracy and computational savings that can be achieved by the methods presented in the current work.

5.2. Relation of the Current Work with Other Methods. In the present method, the discretized surface is a single large sphere that encompasses the stochastic region used for explicit dynamical simulation with a solvent boundary potential. The choice of discretization scheme is based on spherical harmonic angular quadrature rules and, therefore, is particularly relevant since the surface itself must be able to capture the far-field (Taylor) and near-field (multipole) moments expanded about the center of the sphere (eqs 50 and 51). These moments are directly related to spherical harmonic functions, which are eigenfunctions of the angular part of the Laplacian operator, and hence they are ideally suited for expansion of Green's function associated with the solution of the Poisson equation.

It should be noted that the present method is very similar in spirit to the generalized multipoles proposed by Roux et al.[30−32] that are used as a basis for expansion of a Green's function solvent boundary potential. The expansion in generalized multipoles is a method of expanding about a point in the center of the dynamical simulation region that uses projection of moments into a basis that is consistent with the shape of the cavity: for example, spherical harmonic functions in the case of a spherical cavity and Cartesian Legendre polynomials functions in the case of an orthorhombic cavity. These functions are eigenfunctions of the Laplacian operator under these boundary conditions, and hence they are well-suited for a Green's function expansion. An advantage of this type of expansion is that, like in the methods presented here, the number of basis functions, N_l, required to represent the Green's function is sufficiently small and can be precomputed and stored, making subsequent simulations extremely efficient. For example, the total number of basis functions required to expand in spherical harmonic multipoles up to order l is $N_l = (l + 1)^2$.

There are two possible advantages of the current approach over the method of generalized multipoles used in the generalized solvent boundary potential[30−32] although a direct comparison has yet to be made (see below). Both of them are related to the use of a basis expansion about a discretized surface as opposed to about a central point. First, methods based on an expansion about a central point are likely to be less accurate near the boundary surface than methods based on an expansion

about the surface itself. Second, an expansion about a discretized surface is straightforward to generalize to arbitrary-shaped cavities since the surface elements themselves form a natural basis intimately tied to the shape of the boundary. The details regarding the degree to which these perceived advantages may be realized in practice have yet to be characterized; but what is clear from the results of the present work is that the VEP and VEP−RVM methods provide high accuracy for a small number of discretized points (basis functions), and have considerable promise toward the development of improved solvent boundary potentials.

It should also be pointed out that the authors of the generalized solvent boundary potential do not employ the generalized multipoles for representation of the *static* potential of the external environment as is done in the present work, but rather represent this potential on a 3-dimensional grid. This approach is favorable in terms of computational efficiency for the static potential since the potential and forces could easily be interpolated from the grid. In this sense, the basis for the static potential is that of a discrete 3-dimensional grid. This grid, however, contains many points (a typical grid spacing is such that the volume associated with each grid point is on the order of 1 Å^3), and hence would be on the order of 33000−65000 points for the 20−25 Å spherical regions considered in the present work. Such a grid is fairly easily precomputed and stored as a vector. However, in the corresponding matrix required for a Green's function expansion, the number of elements needed would be the number of grid points squared (ignoring symmetry), and construction of the Green's function would require cubic-scaling matrix inversion operations. To circumvent this problem for the Green's function, the generalized multipoles are employed in the generalized solvent boundary potential[30,31] that has already been successfully applied.[32,49] It would be useful in future work to make explicit comparisons of accuracy and efficiency with the methods introduced here.

Recently there has been introduced a charge-scaling method for the treatment of a solvated macromolecular environment in hybrid QM/MM simulations[28] based on a method previously developed for correcting free-energy simulations with continuum electrostatic methods.[29] The basic idea is to generate a QM/MM free energy profile for a reaction with a limited amount of explicit solvent and a set of scaled charges on the ionizable groups that are able reproduce a reasonable set of representative configurations in the course of the simulation. Corrections to the free energy profile can then be made with continuum electrostatic methods to first recover the free energy of the *unscaled* charges with a unit dielectric followed by the free energy associated with the introduction of a dielectric medium of the fully solvated environment. An important step in this process involves the assignment of charge scaling factors to the ionizable groups of the form where $\bar{\Phi}_v(i \rightarrow \mathrm{QM})$ is the

$$q'_i = \frac{\bar{\Phi}_s(i \rightarrow \mathrm{QM})}{\bar{\Phi}_v(i \rightarrow \mathrm{QM})} q_i = \frac{q_i}{\alpha_i} \qquad (25)$$

average electrostatic potential in vacuo due to the ionizable group i with the QM atoms and $\bar{\Phi}_s(i \rightarrow \mathrm{QM})$ is the corresponding average electrostatic potential in solution calculated using the finite-difference Poisson equation. The charge scaling procedure is performed with a limited amount of explicit solvent and the charge scale factors are calculated only for ionizable groups. The average vacuum and solution electrostatic potential of each ionizable group is determined at the positions of the QM atoms based on some representative conformation, or else on a 3-dimensional grid in the active site region. The method has been demonstrated to give reasonable results in QM/MM simulations of hydrolysis reactions of highly charged systems.[28] A rigorous assessment of the force errors on atoms on the dynamical region, however, was not provided.

The method introduced in the present work could afford an alternative procedure for electrostatic representation of the solvated macromolecular environment that could be used in the hybrid QM/MM step of the method. The VEP and VEP−RVM methods would allow an accurate representation of the static field of the solvated macromolecular environment without recourse into empirical charge scaling procedures. Further tests are needed to determine the degree to which the VEP and VEP−RVM methods would influence the reaction free energies.

6. Conclusion

The present paper presents new methods for the efficient calculation of electrostatic interactions between the active dynamical region and surrounding solvated macromolecular environment in hybrid QM/MM calculations. The methods utilize a variational electrostatic projection technique that transforms the 3-dimensional electrostatic problem into a 2-dimensional boundary value problem and is rigorous in the limit such that the boundary surface is represented exactly. Discretized spherical surfaces were generated based on numerical angular quadrature methods[42] for integration of spherical harmonic functions and smooth Gaussian surface elements analogous to those used in a recently developed smooth analytic solvation method based on a conductor-like screening model.[38]

Direct implementation of the variational principle within the numerical framework described leads to the so-called direct VEP method. Analysis of the data leads to the design of an improved method, termed VEP−RVM, that utilizes an intermediate internal projection surface that can be reverse variationally mapped onto an external projection surface used for energy and force evaluation. Different variations of the methods were tested on two highly charged catalytic RNA systems of interest for QM/MM simulations: the hammerhead[44] and hairpin[48] ribozymes.

Variations of the method include enforcement of theoretical constraints on the far-field Taylor moments of the external charge distribution, and inclusion of a "projection buffer" region where atoms are included explicitly instead of being projected. With the direct VEP method, medium accuracy (less that 1% relative force error) was observed at a moderate discretization level (302 points) with a 4 Å projection buffer. Inclusion of constraints improved results typically by a factor of 2. The VEP−RVM method was observed to be considerably more accurate for discretization levels of 302 or greater. Moreover, with appropriate choice of the intermediate projection surface and reverse variational mapping, the resulting projected surface charge showed remarkable agreement with the far-field Taylor moments of the external charge distribution in the absence of any explicit constraints. The VEP−RVM method, using a 4 Å projection buffer, provides improved accuracy (less than 0.1% relative force error) with 302 points, and is able to achieve high accuracy (less than 10^{-4} of a percent) with 1202 points. Errors calculated with radial cutoff methods did not converge; e.g., relative force errors with radial cutoffs of 12.0 and 24.0 Å into the external environment resulted in errors in excess of 6% relative force error.

A key step in the VEP−RVM method involved the position of the intermediate projection surface (the ω surface). From simple scaling arguments, an empirical form for the position of the ω surface was introduced and tested against the agreement with the far-field Taylor moments. The empirical form provided an outstanding fit as a function of distance of the outer projection surface and number of discretized surface elements for both the hammerhead and hairpin systems. The resulting parameters thus form a simple prescription for determination of the ω surface that is automated in the VEP−RVM procedure. Analysis of the radial distribution of atomic force errors indicated that the VEP−RVM method is able to obtain relative force errors less than 10^{-12} near the center of the active dynamical region, and errors that are considerably better than the direct VEP method over all radial distances.

The VEP and VEP−RVM methods are complementary to previously proposed methods for treating the solvated macromolecular environment[28,29,31,32] and may have potential advantages, although these have yet to be tested. The main point to realize is that the VEP and VEP−RVM methods offer a very powerful tool: they allow the electrostatic interactions between an active dynamical region (typically less than 5000 atoms) and a fixed external environment (typically several tens of thousands of atoms) to be calculated to high accuracy via a surface of only 302−1202 points. This forms a useful basis for the design of new approaches that use this basis for the representation of the static potential, and more importantly as a linear-response matrix (i.e., a Green's function) of reduced dimensions for the complex anisotropic response of the macromolecule and solvent that can be precomputed and stored for very efficient QM/MM simulations. Hence, the VEP and VEP−RVM methods provide the foundation for the development of new-generation generalized macromolecular-solvent boundary methods for hybrid QM/MM calculations of biological reactions with improved accuracy. The present work provides a comprehensive account of the implementation of the VEP and VEP−RVM methods and an assessment of their accuracy and performance.

Acknowledgment. D.M.Y. is grateful for financial support provided by the National Institutes of Health (Grant 1R01-GM62248-01A1), and the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement number DAAD19-01-2-0014. Computational resources were provided by the Minnesota Supercomputing Institute.

A. Appendices

A.1. Electrostatic Variational Principles. This section establishes the basic variational principles from the theory of electrostatics that form the basis of the variational electrostatic projection and reverse variational mapping methods. Consider the Maxwell relation[35]

$$\nabla\cdot\mathbf{D} = 4\pi\rho_0 \tag{26}$$

where ρ_0 is an electrostatic charge distribution and $\mathbf{D}$ is the dielectric displacement. The dielectric displacement can be written as

$$\begin{aligned}\mathbf{D} &= (1 + 4\pi\chi_e)\mathbf{E}\\ &= \epsilon\mathbf{E}\end{aligned} \tag{27}$$

where χ_e is the electrostatic susceptibility, here assumed to be a local linear, isotropic dipole response function, ϵ is the corresponding dielectric function related to χ_e by $\epsilon = 1 + 4\pi\chi_e$, and $\mathbf{E}$ is the electrostatic field defined as the gradient of the electrostatic potential ϕ (i.e., $\mathbf{E} = -\nabla\phi$).

Rewriting the Maxwell equation for the dielectric displacement in terms of the electrostatic potential and dielectric function, the familiar form of the Poisson equation for a linear isotropic polarizable medium is obtained

$$\nabla\cdot[\epsilon\nabla\phi] = -4\pi\rho_0 \tag{28}$$

where ϕ is the total electrostatic potential arising from the static charge distribution ρ_0 in the presence of a dielectric medium ϵ.

The solution of the Poisson equation for the electrostatic potential obeys a variational condition on the functional

$$W[\phi;\rho_0,\epsilon] = \int \rho_0\phi \, d^3r - \frac{1}{8\pi}\int \nabla\phi\cdot\epsilon\cdot\nabla\phi \, d^3r \tag{29}$$

such that solution of the stationary condition

$$\begin{aligned}\frac{\delta W[\phi;\,\rho_0,\,\epsilon]}{\delta\phi} &= \rho_0 + \frac{1}{4\pi}\nabla\cdot[\epsilon\nabla\phi]\\ &= 0\end{aligned} \tag{30}$$

provides a solution for the Poisson equation for the total electrostatic potential. It can be shown that this functional is a maximum at the extremal value of ϕ, and the value of the functional $W[\phi;\,\rho_0,\,\epsilon]$ is the total electrostatic energy. Hence maximization of the functional in eq 29 with respect to variations in ϕ provides a variational approximation to the true electrostatic potential and energy that becomes exact as the space of the variation becomes complete. This method has been used as the basis for the numerical solution of the Poisson equation in the presence of a dielectric,[52] and in the case of a constant unit dielectric, it forms the variational procedure proposed by Dunlap for fitting of electron densities with an auxiliary basis for rapid evaluation of the electrostatic energy and potential in density-functional calculations.[53] The same variational principle has also been used as a basis for the calculation of atomic charges in molecules.[54]

It can be shown[38] that maximization of $W[\phi;\,\rho_0,\,\epsilon]$ is equivalent to minimization of the functional $\chi_{\mathbf{E}}{}^2[\phi;\,\phi_0,\epsilon]$ defined as

$$\chi_{\mathbf{E}}{}^2[\phi;\,\phi_0,\,\epsilon] = \frac{1}{2}\int \epsilon(\mathbf{E} - \epsilon^{-1}\mathbf{E}_0)^2 \, d^3r \tag{31}$$

At this point, it is useful to introduce a decomposition of the total electrostatic potential into two components:

$$\phi = \phi_0 + \phi_{pol} \tag{32}$$

where ϕ_0 is the electrostatic potential due to the charge distribution ρ_0 in vacuo; i.e., a solution of

$$\nabla^2\phi_0 = -4\pi\rho_0 \tag{33}$$

and ϕ_{pol} is the component of the electrostatic potential arising from the polarization of the dielectric medium. Under real-space boundary conditions in 3-dimensions, the Green's function for the Poisson equation for a unit dielectric is simply $|\mathbf{r} - \mathbf{r}'|^{-1}$, and the solution for the potential can be written in closed form as

$$\phi_0(\mathbf{r}) = \int \frac{\rho_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}^3 r' \tag{34}$$

Substitution of eq 32 for the decomposition of ϕ into eq 28 leads to the Poisson-like expression for ϕ_{pol}

$$\begin{aligned} \nabla^2 \phi_{\mathrm{pol}} &= -4\pi\rho_0\left(\frac{1-\epsilon}{\epsilon}\right) - \frac{\nabla\epsilon}{\epsilon}\cdot\nabla\phi \\ &= -4\pi\sigma_{\mathrm{pol}} \end{aligned} \tag{35}$$

where σ_{pol} is the dipole polarization density defined by

$$\sigma_{\mathrm{pol}} = \rho_0\left(\frac{1-\epsilon}{\epsilon}\right) + \frac{1}{4\pi}\frac{\nabla\epsilon}{\epsilon}\cdot\nabla\phi \tag{36}$$

Note that the equation relating ϕ_{pol} and σ_{pol} is very similar to that of the Poisson equation with a unit dielectric constant (eq 33); however, unlike the Poisson equation, the σ_{pol} depends on ϕ_{pol} through the term involving $\nabla\phi = \nabla(\phi_0 + \phi_{\mathrm{pol}})$. Nonetheless, once σ_{pol} is determined, the term ϕ_{pol} can be calculated in closed form analogously to eq 34 as

$$\phi_{\mathrm{pol}}(\mathbf{r}) = \int \frac{\sigma_{\mathrm{pol}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}^3 r' \tag{37}$$

Alternately, one can use a finite-difference procedure to solve numerically for $\phi_{\mathrm{pol}}(\mathbf{r})$ directly and eliminate the "self-energy" term that arises in the solution of the Poisson and Poisson−Boltzmann equations that employ a discrete grid representation for distributions of point charges.[55]

The variational principles on the functionals of eqs 29 and 31 can be extended to involve directly ϕ_{pol}, the component of the total electrostatic potential arising from the response of the medium. As will be seen shortly, this extension has direct bearing on the variational electrostatic projection method. Of particular interest to the present work is the manipulation of eq 31 to give a new functional with ϕ_{pol} as the variational argument

$$\chi_{\mathbf{E}_{\mathrm{pol}}}{}^2[\phi_{\mathrm{pol}};\, \phi_0,\, \epsilon] = \frac{1}{2}\int \epsilon\left[\mathbf{E}_{\mathrm{pol}} + \left(\frac{\epsilon - 1}{\epsilon}\right)\mathbf{E}_0\right]^2 \mathrm{d}^3 r \tag{38}$$

where $\mathbf{E}_{\mathrm{pol}} \equiv -\nabla\phi_{\mathrm{pol}}$.

Consider now the specific case of two spatial regions, designated regions 1 and 2, separated by a bounding surface, *S*. Let region 1 contain a static charge distribution ρ_0 and have a unit dielectric constant $\epsilon_1 = 1$. Let region 2 contain no static charge density (i.e., $\rho_0(\boldsymbol{r}) = 0\ \forall\ \mathbf{r} \in$ region 2) and have a dielectric constant ϵ_2. For the purposes of the present general discussion, region 1 may reside inside the surface *S* with region 2 on the outside, or visa versa. With this model, the polarization density of eq 36 reduces to

$$\sigma_{\mathrm{pol}}(\mathbf{r}) = -\frac{1}{4\pi}\left(\frac{\epsilon_2 - 1}{\epsilon_2}\right)\mathbf{D}\cdot\mathbf{n}_{21} \tag{39}$$

where $\mathbf{D}$ is the dielectric displacement of eq 26 and $\mathbf{n}_{21}$ is the unit vector normal to the bounding surface *S* and directed into region 2. Hence, $\sigma_{\mathrm{pol}}(\mathbf{r})$ vanishes at all points not contained on *S*, as is also apparent from eq 35 for the present model system. These conditions lead to a variational principle on the total electrostatic energy that gives rise to the conductor-like screening model.[37,38] The total electrostatic energy (under real-space boundary conditions) for a conductor can be written as

$$\begin{aligned} W[\sigma_{\mathrm{pol}};\rho_0,\epsilon = \infty] &= \frac{1}{2}\int\int (\sigma_{\mathrm{pol}}(\mathbf{r}) + \\ &\qquad \rho_0(\mathbf{r}))\frac{1}{|\mathbf{r} - \mathbf{r}'|}(\sigma_{\mathrm{pol}}(\mathbf{r}') + \rho_0(\mathbf{r}'))\, \mathrm{d}^3 r\, \mathrm{d}^3 r' \\ &= E_0 + E_{\mathrm{pol}} \end{aligned} \tag{40}$$

where

$$E_0 = \frac{1}{2}\int\int \frac{\rho_0(\mathbf{r})\rho_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}^3 r\, \mathrm{d}^3 r' \tag{41}$$

and

$$\begin{aligned} E_{\mathrm{pol}} = \frac{1}{2}\int\int \frac{\sigma_{\mathrm{pol}}(\mathbf{r})\sigma_{\mathrm{pol}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r\, d^3 r' + \\ \int\int \frac{\sigma_{\mathrm{pol}}(\mathbf{r})\rho_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}^3 r\, \mathrm{d}^3 r' \end{aligned} \tag{42}$$

where the conductor surface charge is a solution of the constrained variation

$$\frac{\delta}{\delta\sigma_{\mathrm{pol}}(\mathbf{r})}\{W[\sigma_{\mathrm{pol}};\rho_0] - \int_{\mathbf{r}' \neq S} \lambda(\mathbf{r}')\sigma_{\mathrm{pol}}(\mathbf{r}')\, \mathrm{d}^3 r'\} = 0 \tag{43}$$

where $\lambda(\mathbf{r})$ is a Lagrange multiplier that constrains σ_{pol} to vanish at all points not contained on the bounding surface *S*. Numerically, this constraint need not be explicitly employed as a Lagrange multiplier if the basis for σ_{pol} is only given variational freedom at the bounding surface.

The variational condition on the total electrostatic energy with respect to variations of σ_{pol} *at the bounding surface S* constitute a prescription for the determination of σ_{pol} that solves the dielectric problem for a charge distribution ρ_0 in a region of unit dielectric that is separated from a conducting region by a bounding surface *S*. If the charge distribution is contained within the bounding surface, this model is essentially the conductor-like screening model formulated previously[38] (Figure 1a). However, the variational principles derived here are general, and do not restrict the charge density from being outside the bounding surface which encapsulates the conducting medium. Consideration of the latter gives rise to the variational electrostatic projection method (Figure 1b).

A key property of the solution for σ_{pol} (and hence ϕ_{pol} and $\mathbf{E}_{\mathrm{pol}}$) for the conductor problem is made apparent from examination of the variational functional $\chi_{\mathbf{E}_{\mathrm{pol}}}{}^2$ of eq 38. In the present case, the functional takes the form of a weighted least squares (χ^2) function, where the weight function is the dielectric ϵ. In the region where $\epsilon = \infty$, the term $(\epsilon - 1)/\epsilon = 1$, and minimization of the functional reduces to the requirement that $\mathbf{E}_0 = -\mathbf{E}_{\mathrm{pol}}$ in the conductor region (region 2 in the present model). It is a well-known result of classical electrostatics[35] that a conducting medium cannot support an electric field, and hence the conductor solution cancels exactly the electrostatic potential and field due to ρ_0 everywhere in region 2. This is the key result that forms the basis of the variational electrostatic projection method: *the electrostatic potential and field in region 2 due to a charge density ρ_0 located in region 1 can be obtained exactly from a surface charge density $\gamma_0(\boldsymbol{r})$ on the bounding surface S that is minus the polarization density of the conductor dielectric problem $\sigma_{pol}(\boldsymbol{r})$,* i.e.,

$$\phi_0(\mathbf{r}) = \int \frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^3r' \;\forall\; \mathbf{r}$$
$$= \int_s \frac{\gamma_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^2r' \;\forall\; \mathbf{r}\in \text{region 2} \quad (44)$$

where

$$\gamma_0(\mathbf{r}) = \frac{1}{4\pi}\mathbf{D}\cdot\mathbf{n}_{21}$$
$$= -\sigma_{\mathrm{pol}}(\mathbf{r};\,\epsilon_1=1,\,\epsilon_2=\infty)\;\forall\;\mathbf{r}\in S \quad (45)$$

The variational procedure that projects ρ_0 onto the bounding surface in order to obtain the electrostatic potential in region 2 is termed the *variational electrostatic projection* (VEP) method.

A.2. Green's Function Expansions. The present section briefly discusses Green's function expansions of the Coulomb potential within the context of variational electrostatic projection. The Green's function associated with the Laplacian operator in real space is simply $|\mathbf{r}_1-\mathbf{r}_2|^{-1}$, and can be expanded in terms of solid harmonic functions[56] as

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(M^{+}_{l,|m|}(\mathbf{r}_>)N^{+}_{l,|m|}(\mathbf{r}_<) + M^{-}_{l,|m|}(\mathbf{r}_>)N^{-}_{l,|m|}(\mathbf{r}_<)\right)$$
$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{\pm} M^{\pm}_{l,|m|}(\mathbf{r}_>)N^{\pm}_{l,|m|}(\mathbf{r}_<) \quad (46)$$

where $\mathbf{r}_>$ and $\mathbf{r}_<$ indicate the vectors $\mathbf{r}$ and $\mathbf{r}'$ with the larger and smaller magnitude, respectively, and the functions $M^{\pm}_{l,|m|}(\mathbf{r})$ and $N^{\pm}_{l,|m|}(\mathbf{r})$ are given by

$$M^{\pm}_{l,|m|}(\mathbf{r}) = i^{(1\mp1)/2}a^{+}_{l,|m|}\cdot\frac{Y_l^{-|m|}(\theta,\phi)\pm(-1)^mY_l^{|m|}(\theta,\phi)}{r^{l+1}} \quad (47)$$

$$N^{\pm}_{l,|m|}(\mathbf{r}) = i^{(1\mp1)/2}a^{-}_{l,|m|}\cdot r^l[Y_l^{-|m|}(\theta,\phi)\pm(-1)^mY_l^{|m|}(\theta,\phi)] \quad (48)$$

where

$$a^{\pm}_{l,|m|} = \sqrt{\frac{\pi}{2l+1}}((l+|m|)!(l-|m|)!)^{\pm1/2} \quad (49)$$

Using the above definitions, the charge weighted multipole ($Q^{\pm}_{lm}$) and Taylor ($T^{\pm}_{lm}$) moments about the origin for a system of N_A charged particles q_i with positions $\mathbf{r}_i$, $i=1,\ldots,N_A$, can be calculated respectively as

$$Q^{\pm}_{l,m} = \sum_i^{N_A} q_i N^{\pm}_{l,m}(\mathbf{r}_i) \quad (50)$$

$$T^{\pm}_{l,m} = \sum_i^{N_A} q_i M^{\pm}_{l,m}(\mathbf{r}_i) \quad (51)$$

The VEP method that was introduced in section A.1, and is discussed in detail in section A.3, can be thought of in terms of a reduced dimensional mapping of the Green's function elements associated with electrostatic interactions that occur on different sides of a bounding surface. For the electrostatic potential due to a charge distribution located on one side of a spherical surface evaluated at points located on the other side of the surface, a reduced dimensional mapping is possible via eq 44. In effect, the solution to the Poisson equation to determine the electrostatic potential on the opposite side of the surface can be transformed from the direct solution that involves 3-dimensional spatial integration to a boundary value problem that involves 2-dimensional surface integration. The VEP method provides a solution to the electrostatic boundary value problem.

If the boundary value problem is solved exactly, the electrostatic potential and forces are also exact. Consequently, the variationally projected surface charge will satisfy conditions on the near-field and far-field multipole and Taylor moments. If the spatial charge distribution is enclosed by the surface, then the VEP surface charge distribution will have the same near-field multipole moments (eq 50) as the spatial charge distribution. If instead, the spatial charge distribution lies on the outside of the projection surface, the VEP surface charge distribution will have the same far-field (Taylor) multipole moments (eq 51) as the spatial charge distribution.

These conditions are related to Gauss' law for the integrated value of the polarization density $\sigma_{\mathrm{pol}}(\mathbf{r})$

$$\int_s \sigma_{\mathrm{pol}}(\mathbf{r})\,\mathrm{d}a = -\left(\frac{\epsilon_2-1}{\epsilon_2}\right)\int_V \rho_0(\mathbf{r})\,\mathrm{d}^3r \quad (52)$$

where V is the volume that contains the charge distribution. In the limit that $\epsilon_2 \to \infty$, the prefactor becomes unity and a conductor solution is obtained. Hence, the lowest order multipole moment of the VEP surface charge (in the case the charge distribution resides inside the surface) is analogous to the Gauss law for the conductor dielectric problem (with a change in sign).

A.3. Variational Electrostatic Projection. Consider the total electrostatic energy of a system of charged particles subdivided into two regions, designated the *active region* and the *environment*, and separated by a bounding surface. Let the reference charge distribution inside the active region be designated $\boldsymbol{\rho}_0$, and let the reference static charge distribution of the environment consist of two parts: one part arising from the macromolecular solute, designated $\mathbf{q}_0$, and another arising from the surrounding solvent, designated $\boldsymbol{\sigma}_0$. Here, the charge distributions $\boldsymbol{\rho}_0$, $\mathbf{q}_0$, and $\boldsymbol{\sigma}_0$ are written as matrix (column vector) quantities with dimensions $N_{\boldsymbol{\rho}} \times 1$, $N_{\mathbf{q}} \times 1$ and $N_{\boldsymbol{\sigma}} \times 1$, respectively. The elements of the vectors represent expansion coefficients for the charge distributions in a representative basis. The specific basis used to represent the vector and matrix quantities introduced in this section are discussed in detail in section 4.2.

The total electrostatic energy of the system is given by

$$E = \frac{1}{2}\boldsymbol{\rho}_0^{\mathrm{T}}\cdot\mathbf{C}_{\boldsymbol{\rho}\boldsymbol{\rho}}\cdot\boldsymbol{\rho}_0 + \frac{1}{2}\mathbf{q}_0^{\mathrm{T}}\cdot\mathbf{C}_{\mathbf{q}\mathbf{q}}\cdot\mathbf{q}_0 + \boldsymbol{\rho}_0^{\mathrm{T}}\cdot\mathbf{C}_{\boldsymbol{\rho}\mathbf{q}}\cdot\mathbf{q}_0 +$$
$$\frac{1}{2}\boldsymbol{\sigma}_0^{\mathrm{T}}\cdot\mathbf{A}_{\boldsymbol{\sigma}\boldsymbol{\sigma}}\cdot\boldsymbol{\sigma}_0 + \boldsymbol{\rho}_0^{\mathrm{T}}\cdot\mathbf{B}_{\boldsymbol{\rho}\boldsymbol{\sigma}}\cdot\boldsymbol{\sigma}_0 + \mathbf{q}_0^{\mathrm{T}}\cdot\mathbf{B}_{\mathbf{q}\boldsymbol{\sigma}}\cdot\boldsymbol{\sigma}_0 \quad (53)$$

where the matrices **A**, **B**, and **C** represent Coulomb integrals between basis functions and are subscripted by the specific charge distributions that they connect. The convention is used that a reversal of the matrix subscripts indicates the matrix transpose operation, e.g., $\mathbf{A}_{\beta\alpha} \equiv (\mathbf{A}_{\alpha\beta})^{\mathrm{T}}$.

If the reference solvent charge distribution $\boldsymbol{\sigma}_0$ comes from an equilibrated simulation, then it can be represented explicitly. If the reference charge distribution comes from an implicit continuum model, in particular if it is derived from a smooth conductor-like screening model,[38] it can be calculated as

$$\boldsymbol{\sigma}_0 = -f(\epsilon_1,\epsilon_2)\cdot\mathbf{A}_{\boldsymbol{\sigma}\boldsymbol{\sigma}}^{-1}\cdot[\mathbf{B}_{\boldsymbol{\sigma}\boldsymbol{\rho}_0}\cdot\boldsymbol{\rho}_0 + \mathbf{B}_{\sigma\mathbf{q}}\cdot\mathbf{q}_0] \quad (54)$$

where $\boldsymbol{\rho}_0$ and $\mathbf{q}_0$ indicate the reference solute charge distributions

in region 1 and 2, respectively, ϵ_1 and ϵ_2 are the dielectric constants inside and outside the macromolecular cavity, respectively, and $f(\epsilon_1,\epsilon_2)$ is defined by

$$f(\epsilon_1,\epsilon_2) = \left(\frac{\epsilon_2 - \epsilon_1}{\epsilon_1 \epsilon_2}\right) \tag{55}$$

The variationally projected surface charge distribution $\boldsymbol{\gamma}_\mathbf{X}$ used to model the electrostatic potential and forces due to an external charge distribution $\mathbf{X}$ (for example $\mathbf{q}_0$ and/or $\boldsymbol{\sigma}_0$) can be obtained from variation of the functional

$$W_\mathbf{X}[\boldsymbol{\gamma}_\mathbf{X}] = \frac{1}{2}\boldsymbol{\gamma}_\mathbf{X}^{\mathrm{T}}\cdot\mathbf{P}_{\gamma\gamma}\cdot\boldsymbol{\gamma}_\mathbf{X} - \boldsymbol{\gamma}_\mathbf{X}^{\mathrm{T}}\cdot\mathbf{B}_{\gamma\mathbf{X}}\cdot\mathbf{X} \tag{56}$$

with respect to $\boldsymbol{\gamma}_\mathbf{X}$, where the basis functions for expansion of $\boldsymbol{\gamma}_\mathbf{X}$ are located on the bounding surface that divides the active region from the environment (i.e., the projection surface), and $\mathbf{P}_{\gamma\gamma}$ is the square matrix of Coulomb integrals between these basis functions. The formal solution of the unconstrained variation gives

$$\boldsymbol{\gamma}_\mathbf{X}(0) = \mathbf{P}_{\gamma\gamma}^{-1}\cdot\mathbf{B}_{\gamma\mathbf{X}}(0)\cdot\mathbf{X} \tag{57}$$

where $\mathbf{B}_{\gamma\mathbf{X}}(0) \equiv \mathbf{B}_{\gamma\mathbf{X}}$, with the zero in parentheses indicating that the projection has been done in the absence of constraints.

In some cases, it may be desirable that the projection result satisfy a set of linear constraints (see below). Suppose that the surface charge $\boldsymbol{\gamma}_\mathbf{X}$ were constrained to obey a set of N_λ conditions of the form

$$\mathbf{D}_{\lambda\gamma}\cdot\boldsymbol{\gamma}_\mathbf{X} = \mathbf{Z}_{\gamma\mathbf{X}}\cdot\mathbf{X} \tag{58}$$

where $\mathbf{D}_{\lambda\gamma}$ and $\mathbf{Z}_{\gamma\mathbf{X}}$ are $N_\lambda \times N_\gamma$ and $N_\lambda \times N_\mathbf{X}$ matrices, respectively, that define the linear constraint conditions. For example, the matrix operations could represent a set of generalized far field Taylor multipole moments that represent theoretical conditions the exact solution for $\boldsymbol{\gamma}_\mathbf{X}$ would obey, but that an approximate numerical solution might not. The constrained variation is given by

$$\frac{\delta}{\delta\,\boldsymbol{\gamma}_\mathbf{X}}\{W_\mathbf{X}[\boldsymbol{\gamma}_\mathbf{X}] - \boldsymbol{\lambda}_\mathbf{X}^{\mathrm{T}}\cdot(\mathbf{D}_{\lambda\gamma}\cdot\boldsymbol{\gamma}_\mathbf{X} - \mathbf{Z}_{\gamma\mathbf{X}}\cdot\mathbf{X})\} = 0 \tag{59}$$

where λ is a $N_\lambda \times 1$ vector of Lagrange multipliers on the N_λ linear constraint conditions (eq 58). The constrained solution for $\boldsymbol{\gamma}_\mathbf{X}$ is given by

$$\begin{aligned}\boldsymbol{\gamma}_\mathbf{X}(\lambda) &= \mathbf{P}_{\gamma\gamma}^{-1}\cdot(\mathbf{B}_{\gamma\mathbf{X}}\cdot\mathbf{X} + \mathbf{D}_{\gamma\lambda}\cdot\boldsymbol{\lambda}_\mathbf{X}) \\ &= \boldsymbol{\gamma}_\mathbf{X}(0) + \delta\boldsymbol{\gamma}_\mathbf{X}(\lambda)\end{aligned} \tag{60}$$

where $\boldsymbol{\gamma}_\mathbf{X}(0)$ is given in eq 57. Here the λ in parentheses indicates that the projection has arisen from a constrained variation. The vector of Lagrange multipliers $\boldsymbol{\lambda}_\mathbf{X}$ in eq 60 is given by

$$\boldsymbol{\lambda}_\mathbf{X} = \mathbf{Q}_{\lambda\lambda}^{-1}\cdot\mathbf{R}_{\gamma\mathbf{X}}\cdot\mathbf{X} \tag{61}$$

$$\mathbf{Q}_{\lambda\lambda} = \mathbf{D}_{\lambda\gamma}\cdot\mathbf{P}_{\gamma\gamma}^{-1}\cdot\mathbf{D}_{\gamma\lambda} \tag{62}$$

$$\mathbf{R}_{\gamma\mathbf{X}} = (\mathbf{Z}_{\gamma\mathbf{X}} - \mathbf{D}_{\lambda\gamma}\cdot\mathbf{P}_{\gamma\gamma}^{-1}\cdot\mathbf{D}_{\gamma\lambda}) \tag{63}$$

(recall the convention that $\mathbf{D}_{\lambda\gamma} \equiv \mathbf{D}_{\gamma\lambda}^{\mathrm{T}}$) and the constraint correction $\delta\boldsymbol{\gamma}_\mathbf{X}(\lambda)$ is given by

$$\begin{aligned}\delta\boldsymbol{\gamma}_\mathbf{X}(\lambda) &= \mathbf{P}_{\gamma\gamma}^{-1}\cdot\mathbf{D}_{\gamma\lambda}\cdot\boldsymbol{\lambda}_\mathbf{X} \\ &= \mathbf{P}_{\gamma\gamma}^{-1}\cdot\mathbf{D}_{\gamma\lambda}\cdot\mathbf{Q}_{\lambda\lambda}^{-1}\cdot\mathbf{R}_{\gamma\mathbf{X}}\cdot\mathbf{X} \\ &= \mathbf{P}_{\gamma\gamma}^{-1}\cdot\delta\mathbf{B}_{\gamma\mathbf{X}}(\lambda)\cdot\mathbf{X}\end{aligned} \tag{64}$$

where

$$\delta\mathbf{B}_{\gamma\mathbf{X}}(\lambda) = \mathbf{D}_{\gamma\lambda}\cdot\mathbf{Q}_{\lambda\lambda}^{-1}\cdot\mathbf{R}_{\gamma\mathbf{X}} \tag{65}$$

These relations lead to a concise and general formula for the variational electrostatic projection of a charge distribution $\mathbf{X}$ onto the projection surface to give the set of projection coefficients $\boldsymbol{\gamma}_\mathbf{X}$ as

$$\boldsymbol{\gamma}_\mathbf{X}(\lambda) = \mathbf{P}_{\gamma\gamma}^{-1}\cdot\mathbf{B}_{\gamma\mathbf{X}}(\lambda)\cdot\mathbf{X} \tag{66}$$

$$\mathbf{B}_{\gamma\mathbf{X}}(\lambda) = \mathbf{B}_{\gamma\mathbf{X}}(0) + \delta\mathbf{B}_{\gamma\mathbf{X}}(\lambda) \tag{67}$$

Equations 66 and 67 form the basis of the direct variational electrostatic projection (VEP) method, and are the main results of this subsection.

Application of the VEP method (eq 66) to the macromolecular solute and solvent charge distributions of the environment provides the following relations for the electrostatic potential at the particle positions ρ in the active region

$$\mathbf{C}_{\rho\mathbf{q}}\cdot\mathbf{q}_0 = \mathbf{B}_{\rho\gamma}\cdot\mathbf{P}_{\gamma\gamma}^{-1}\cdot\mathbf{B}_{\gamma\mathbf{X}}(\lambda)\cdot\mathbf{q}_0 = \mathbf{B}_{\rho\gamma}\cdot\boldsymbol{\gamma}_{\mathbf{q}_0}(\lambda) \tag{68}$$

$$\mathbf{B}_{\rho\sigma}\cdot\boldsymbol{\sigma}_0 = \mathbf{B}_{\rho\gamma}\cdot\mathbf{P}_{\gamma\gamma}^{-1}\cdot\mathbf{B}_{\gamma\sigma}(\lambda)\cdot\boldsymbol{\sigma}_0 = \mathbf{B}_{\rho\gamma}\cdot\boldsymbol{\gamma}_{\sigma_0}(\lambda) \tag{69}$$

Let the $\boldsymbol{\gamma}_0(\lambda)$ be defined as the projected surface charge arising from both the external solute ($\mathbf{q}_0$) and solvent ($\boldsymbol{\sigma}_0$) charge distributions, i.e., $\boldsymbol{\gamma}_0(\lambda) = \boldsymbol{\gamma}_{\mathbf{q}_0}(\lambda) + \boldsymbol{\gamma}_{\sigma_0}(\lambda)$, then substitution into the energy equation gives

$$\begin{aligned}E &= \frac{1}{2}\boldsymbol{\rho}^{\mathrm{T}}\cdot\mathbf{C}_{\rho\rho}\cdot\boldsymbol{\rho} + \boldsymbol{\rho}^{\mathrm{T}}\cdot\mathbf{B}_{\rho\gamma}\cdot\boldsymbol{\gamma}_0 + \\ &\qquad \frac{1}{2}\mathbf{q}_0^{\mathrm{T}}\cdot\mathbf{C}_{\mathbf{qq}}\cdot\mathbf{q}_0 + \mathbf{q}_0^{\mathrm{T}}\cdot\mathbf{B}_{\mathbf{q}\sigma}\cdot\boldsymbol{\sigma}_0 + \frac{1}{2}\boldsymbol{\sigma}_0^{\mathrm{T}}\cdot\mathbf{A}_{\sigma\sigma}\cdot\boldsymbol{\sigma}_0 \\ &= \frac{1}{2}\boldsymbol{\rho}^{\mathrm{T}}\cdot\mathbf{C}_{\rho\rho}\cdot\boldsymbol{\rho} + \boldsymbol{\rho}^{\mathrm{T}}\cdot\mathbf{B}_{\rho\gamma}\cdot\boldsymbol{\gamma}_0(\lambda) + C_0\end{aligned} \tag{70}$$

where the constant C_0 is independent of the instantaneous positions of the particles in the active region and is given by

$$C_0 = \frac{1}{2}\mathbf{q}_0^{\mathrm{T}}\cdot\mathbf{C}_{\mathbf{qq}}\cdot\mathbf{q}_0 + \mathbf{q}_0^{\mathrm{T}}\cdot\mathbf{B}_{\mathbf{q}\sigma}\cdot\boldsymbol{\sigma}_0 + \frac{1}{2}\boldsymbol{\sigma}_0^{\mathrm{T}}\cdot\mathbf{A}_{\sigma\sigma}\cdot\boldsymbol{\sigma}_0 \tag{71}$$

The final energy expression of eq 70 (ignoring the constant C_0) requires the calculation of the electrostatic potential due to ρ, the charge distribution of the particles of the active region that will be changing in the dynamics, only at (1) the positions of the active particles themselves and (2) the positions of the N_γ surface elements that make up the bounding γ surface. Tremendous computational savings, therefore, may be realized if the number of surface elements is significantly smaller than the number of macromolecular solute and solvent particles in the environment; i.e., if $N_\gamma \ll (N_\mathbf{q} + N_\sigma)$.

A.4. Reverse Variational Mapping. The variational electrostatic projection method can be improved with the use of two surfaces, ω and γ, located at r_ω and r_γ, respectively (Figure 2). The γ surface is the surface that is used to derive the electrostatic potential and forces in the dynamical region (zones I and II in Figure 2). The ω surface, which lies interior to the

γ surface, is used as an intermediate projection surface that allows projection onto the final γ surface and subsequent evaluation of the potential and forces to be more accurate than in the direct VEP method. The motivation behind this new procedure is to increase the accuracy of the VEP-modeled potential and forces in the dynamical region. This subsection outlines the formalism associated with this dual-surface approach, which is designated *reverse variational mapping* (RVM). A detailed discussion of the rationale and testing of the VEP−RVM method is deferred to section 4.4.

The first step of the VEP−RVM method is to perform a VEP of the external charge distribution on the ω surface. As will be discussed in section 4.4, atoms very close to the projection surface are projected less accurately than those farther away. Hence, use of the ω surface, that lies interior to R_{III} (the boundary that defines the external atoms that are to be variationally projected) allows for a more accurate projection than use of a surface that is right at the boundary as in the direct VEP method. This choice for the ω surface position places it within the dynamical region itself, and hence, the ω surface cannot be used directly for evaluation of the potential and forces throughout the dynamical region. Instead, the γ surface that is placed at or even beyond R_{III} is used for this purpose. This requires one to establish a procedure whereby the projected charge distribution on the ω surface is *reverse mapped* onto the final γ surface.

It is required, then, that the final γ charge distribution is the one that upon direct VEP projection onto the ω surface reproduces as closely as possible the same surface charge distribution as did direct VEP of the atoms of the external environment (region IIIb). Equating these conditions, with the charge distribution designated generically as **X**, the relation is obtained:

$$\begin{aligned}\boldsymbol{\omega}_{\mathbf{X}}(0) &= \mathbf{P}_{\omega\omega}{}^{-1}\cdot\mathbf{B}_{\omega\mathbf{X}}(0)\cdot\mathbf{X} \\ &= \mathbf{P}_{\omega\omega}{}^{-1}\cdot\mathbf{B}_{\omega\gamma}(0)\cdot\boldsymbol{\gamma}_{\mathbf{X}}\end{aligned} \tag{72}$$

The $\boldsymbol{\gamma}_{\mathbf{X}}(0)$ surface therefore has the solution

$$\boldsymbol{\gamma}_{\mathbf{X}}(0) = [\,\mathbf{P}_{\omega\omega}{}^{-1}\cdot\mathbf{B}_{\omega\gamma}(0)]^{-1}\cdot\boldsymbol{\omega}_{\mathbf{X}}(0) \tag{73}$$

Note: this expression for $\boldsymbol{\gamma}_{\mathbf{X}}$ assumes that the matrix in brackets is invertible. In principle, this will be the case if the discretization level of the γ surface is equal to that of the ω surface (i.e., if $N_\gamma = N_\omega$), although the solution might also be obtained through use of singular value decompositions in other cases. In the present work the discretization levels of the ω and γ surfaces in the VEP−RVM method are the same, and $\mathbf{B}_{\omega\gamma}(0)$ is a square (but not symmetric) matrix. In this case, the expression for $\boldsymbol{\gamma}_{\mathbf{X}}(0)$ can be simplified as

$$\boldsymbol{\gamma}_{\mathbf{X}}(0) = \mathbf{B}_{\omega\gamma}(0)^{-1}\cdot\mathbf{B}_{\omega\mathbf{X}}(0)\cdot\mathbf{X} \tag{74}$$

As in the direct VEP method, it may be desirable that certain conditions on the reverse variational mapping be preserved. For example, suppose that it is desired that the projection result obey a set of linear constraint conditions of the form

$$\mathbf{D}_{\lambda\omega}\cdot\boldsymbol{\omega}_{\mathbf{X}} = \mathbf{Z}_{\lambda\gamma}\cdot\boldsymbol{\gamma}_{\mathbf{X}} \tag{75}$$

A similar set of constraints could be applied to the direct VEP mapping of the external charge distribution **X** onto the ω surface as in eq 58. Analogous to eq 72, the relation between **X**, $\boldsymbol{\omega}_{\mathbf{X}}(\lambda)$, and $\boldsymbol{\gamma}_{\mathbf{X}}(\lambda)$ (under the above-mentioned constraints) is given by

$$\begin{aligned}\boldsymbol{\omega}_{\mathbf{X}}(\lambda) &= \mathbf{P}_{\omega\omega}{}^{-1}\cdot\mathbf{B}_{\omega\mathbf{X}}(\lambda)\cdot\mathbf{X} \\ &= \mathbf{P}_{\omega\omega}{}^{-1}\cdot\mathbf{B}_{\omega\gamma}(\lambda)\cdot\boldsymbol{\gamma}_{\mathbf{X}}\end{aligned} \tag{76}$$

where $\mathbf{B}_{\omega\gamma}(\lambda)$ is obtained from the general VEP equation (eq 66 and 67) with the projection surface charge distribution $\boldsymbol{\gamma}_{\mathbf{X}}$ and external charge vector **X** being replaced by $\boldsymbol{\omega}_{\mathbf{X}}$ and $\boldsymbol{\gamma}_{\mathbf{X}}$, respectively. Solution of the set of equations in eq 76 leads to the following expression for $\boldsymbol{\gamma}_{\mathbf{X}}$:

$$\begin{aligned}\boldsymbol{\gamma}_{\mathbf{X}}(\lambda) &= [\,\mathbf{P}_{\omega\omega}{}^{-1}\cdot\mathbf{B}_{\omega\gamma}(\lambda)]^{-1}\cdot\boldsymbol{\omega}_{\mathbf{X}}(\lambda) \\ &= \mathbf{B}_{\omega\gamma}(\lambda)^{-1}\cdot\mathbf{B}_{\omega\mathbf{X}}(\lambda)\cdot\mathbf{X}\end{aligned} \tag{77}$$

where again, it is noted that the matrix $\mathbf{B}_{\omega\gamma}$ is a nonsymmetric square matrix. *Equation 77 forms the basis of the VEP−RVM method analogous to eq 66 for the VEP method, and is the main result of this subsection.*

References and Notes

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(2) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523−2544.
(3) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.
(4) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.
(5) Thiel, W. Perspectives on Semiempirical Molecular Orbital Theory. In *Advances in Chemical Physics*; Prigogine, I.; Rice, S. A., Eds.; John Wiley and Sons: New York, 1996; Vol. 93, pp 703−757.
(6) Clark, T. *J. Mol. Struct. (THEOCHEM)* **2000**, *530*, 1−10.
(7) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. Status Solidi B* **2000**, *217*, 357−376.
(8) Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(9) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(10) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(11) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(12) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(13) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(14) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.
(15) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.
(16) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621−627.
(17) Grengard, L. *Science* **1994**, *265*, 909−904.
(18) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155−179.
(19) Warshel, A. *Acc. Chem. Res.* **2002**, *35*, 385−395.
(20) Auffinger, P.; Westhof, E. *Nucleic Acid Sci.* **2001**, *56*, 266−274.
(21) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182−196.
(22) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.
(23) York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001−5002.
(24) Cheatham, T. E., III; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435−471.
(25) Berkowitz, M.; McCammon, J. A. *Chem. Phys. Lett.* **1982**, *90*, 215−217.
(26) Brunger, A.; Brooks, C. L., III; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495−500.

(27) Brooks, C. L., III; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843−865.
(28) Dinner, A. R.; Lopez, X.; Karplus, M. *Theor. Chem. Acc.* **2003**, *109*, 118−124.
(29) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 8349−8362.
(30) Roux, B.; Beglov, D.; Im, W. Simulation and Theory of Electrostatic Interations in Solution. In *Proceedings of the Santa Fe Workshop on Treatment of Electrostatic Interactions in Computer Simulations of Condensed Media*; Pratt, L. R.; Hummer, G., Eds.; AIP Conference Proceedings 492; AIP: Melville, New York, 1999.
(31) Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924−2937.
(32) Banavali, N. K.; Im, W.; Roux, B. *J. Chem. Phys.* **2002**, *117*, 7381−7388.
(33) Stone, A. *The Theory of Intermolecular Forces*; International Series of Monographs in Chemistry 32; Clarendon Press: Oxford, England, 1996.
(34) Böttcher, C. J. F. *Theory of Electric Polarisation;* Elsevier Publishing Company: Amsterdam, 1952.
(35) Jackson, J. D. *Classical Electrodynamics*, 3rd ed.; Wiley: New York, 1999.
(36) Landau, L. D.; Lifshitz, E. M.; Pitaevskiı, L. P. *Electrodynamics of Continuous Media*, 2nd ed.; Butterworth-Heinenann, Oxford, England, 1996; Vol. 8.
(37) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.
(38) York, D. M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 11060−11079.
(39) York, D. M.; Lee, T.-S.; Yang, W. *Chem. Phys. Lett.* **1996**, *263*, 297−306.
(40) Stroud, A. H. *Approximate Calculation of Multiple Integrals.* Prentice Hall, Englewood Cliffs, NJ, 1971.
(41) Lebedev, V. I. *Sibirsk. Mater. Ž.* **1977**, *18*, 132−142.
(42) Delley, B. *J. Comput. Chem.* **1996**, *17*, 1152−1155.
(43) Doherty, E. A.; Doudna, J. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 457−475.
(44) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065−2069.
(45) Scott, W. G. *Q. Rev. Biophys.* **1999**, *32*, 241−294.
(46) Rupert, P. B.; Ferré-D'Amaré, A. R. *Nature (London)* **2001**, *410*, 780−786.
(47) Burke, J. M. *Nat. Struct. Biol.* **2001**, *8*, 382−384.
(48) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferré-D'Amaré, A. R. *Science* **2002**, *298*, 1421−1424.
(49) Wonpil, I.; Roux, B. *J. Chem. Phys.* **2001**, *115*, 4850−4861.
(50) York, D. M.; Darden, T.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345−8348.
(51) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(52) Luty, B. A.; Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1992**, *13*, 1114−1118.
(53) Dunlap, B. I. *J. Chem. Phys.* **1983**, *78*, 3140−3142.
(54) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1995**, *102*, 7549−7556.
(55) Zhou, Z.; Payne, P.; Vasquez, M.; Kuhn, N.; Levitt, M. *J. Comput. Chem.* **1996**, *11*, 1344−1351.
(56) Pérez-Jordá, J.; Yang, W. *J. Chem. Phys.* **1996**, *104*, 8003−8006.