

# A Charge-Scaling Implementation of the Variational Electrostatic Projection Method

BRENT A. GREGERSEN, DARRIN M. YORK

Department of Chemistry, University of Minnesota, 207 Pleasant St. SE,  
Minneapolis, MN 55455-0431

Received 1 June 2005; Accepted 26 July 2005

DOI 10.1002/jcc.20318

Published online in Wiley InterScience (www.interscience.wiley.com).

**Abstract:** Two new charge-scaling methods for efficient modeling of the solvated macromolecular environment in hybrid QM/MM calculations of biological reactions are presented. The methods are extensions of the variational electrostatic projection (VEP) method, and allows a subset of atomic charges in the external environment to be adjusted to mimic, in the active dynamical region, the electrostatic potential and field due to the large surrounding macromolecule and solvent. The method has the advantages that it offers improved accuracy, does not require the use of a three-dimensional grid or auxiliary set of fitting points, and requires only minor molecular simulation code modifications. The VEP-cs and VEP-RVM+cs methods are able to attain very high accuracy (relative force errors of  $10^{-7}$  or better with appropriate choice of control parameters), and take advantage of a recently introduced set of high-order discretization schemes and Gaussian exponents for boundary element solvation and VEP methods. The methods developed here serve as potentially powerful tools in the arsenal of computational techniques used in multiscale computational modeling problems.

© 2005 Wiley Periodicals, Inc. J Comput Chem 27: 103–115, 2006

**Key words:** charge-scaling implementation; electrostatic projection method

## Introduction

The efficient treatment of solvation in hybrid quantum mechanical/molecular mechanical (QM/MM) simulations of biological reactions<sup>1</sup> in enzymes and ribozymes is an area of considerable current interest and concentration.<sup>2,3</sup> For biological macromolecules, it is important to capture the electrostatic effects due to the solvated macromolecular environment. As has been pointed out in detail elsewhere,<sup>1</sup> use of electrostatic cutoffs<sup>4</sup> or treatment of an isolated macromolecule in the absence of solvation can lead to severe artifacts in QM/MM simulations. Several methods have been suggested to circumvent this problem and facilitate more accurate and efficient QM/MM simulations in macromolecular environments.<sup>1,5–8</sup> One method, the surface charge representation of the electrostatic embedding potential (SCREEP),<sup>9</sup> has been utilized to study reactivity at solid–liquid surfaces, and involves combining continuum dielectric methods with an embedded cluster approach incorporating the Madelung potential. For biological macromolecules, an alternate method has been applied to QM/MM simulations<sup>6</sup> based on a charge-scaling procedure originally developed for molecular-mechanical free-energy simulations.<sup>10</sup> In this method, scaled charges are derived for ionic amino acid side chains using a potential-based scheme<sup>10</sup> by fitting to the electrostatic

potential derived from finite-difference Poisson or linearized Poisson-Boltzmann calculation. The electrostatic potential is evaluated at grid points located within the van der Waals radii of the QM atoms in a representative conformation or set of conformations. The charge-scaling procedure can be used in conjunction with continuum electrostatic methods to correct the free energy values. This results in a more realistic and computationally tractable model for the solvated macromolecular environment. The method has been applied in a study of the hydrolysis reaction of uracil–DNA glycosylase complexed with double-stranded DNA.<sup>6</sup> A similar nonuniform charge-scaling method has also been investigated to account for solvent screening in molecular mechanics calculations and applied to the motor protein myosin.<sup>7</sup>

**Correspondence to:** D. M. York; e-mail: york@chem.umn.edu

Contract/grant sponsor: National Institutes of Health; contract/grant number: GM 62248.

Contract/grant sponsor: Army High Performance Computing Research Center (AHPCRC) under the Department of the Army, Army Research Laboratory (ARL); contract/grant number: DAAD19-01-2-0014

This article contains Supplementary Material available at <http://www.interscience.wiley.com/jpages/0192-8651/suppmat>

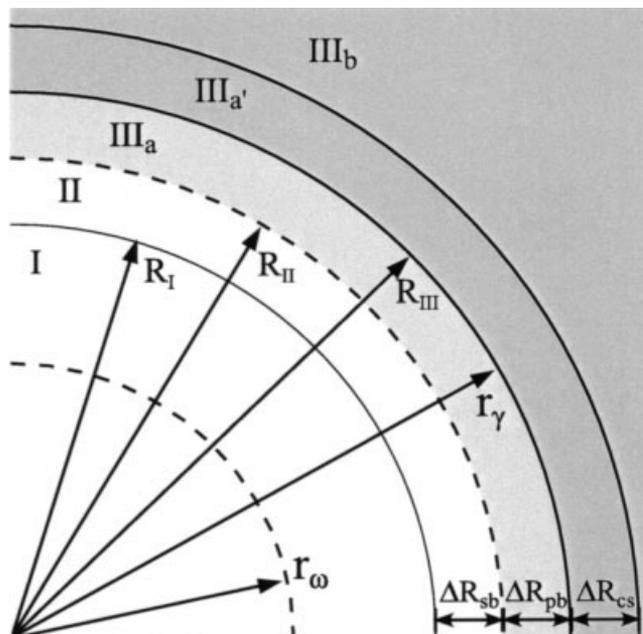
Recently, another approach has been introduced to model the complex solvated macromolecular environment by a discretized surface that encompasses the active dynamical simulation region. In this method, a variational electrostatic projection (VEP) technique<sup>8</sup> is employed to determine a set of charges on a discretized surface surrounding, for example, a ribozyme active site, such that these charges accurately reproduce the electrostatic potential and field inside the surface. The auxiliary set of surface elements and their interactions must be integrated into molecular dynamics codes as a separate set of routines. However, the procedure can also be easily modified to map the charges onto a suitable set of existing atomic positions such that the set of charge corrections serve to reproduce the electrostatic potential and field inside the active site due to the entire solvated macromolecular environment.

The present work presents a new charge-scaling implementation of the VEP method for modeling solvation effects in hybrid QM/MM calculations. The method has the advantage that the solvent effect due to the entire macromolecular charge distribution is considered, and the procedure does not require the specification of a three-dimensional grid or auxiliary set of fitting points. Moreover, the use of environment charge corrections would require little or no code modifications to existing molecular simulation software. Once the charge corrections have been calculated, they can be applied through the definition of new atom types, updated atom charges or dummy atoms. The external atoms from which these charge corrections were obtained can then be deleted from the molecular system, or removed from any electrostatic evaluation lists. The method is demonstrated to be very accurate for a highly charged ribozyme system.

## Theory

Previously, a VEP method and related VEP reverse variational mapping (VEP-RVM) procedure have been described in detail.<sup>8</sup> The method is based on an electrostatic variational principle, related to that used in the smooth conductor-like screening solvation model,<sup>11</sup> whereby a spatially extensive electrostatic problem on one side of a bounding surface can be transformed into a boundary value problem and solved efficiently for the electrostatic potential and field on the other side of the surface. This is exactly the situation encountered in many activated dynamics simulations of biological reactions performed with combined QM/MM potentials. In the previous work, the surface charge distribution was obtained either directly in the "direct" VEP method, or via an even more accurate VEP-RVM procedure that involved reverse variational mapping of the surface charge onto an external surface.

In the present work, this procedure is generalized to reverse map the VEP charge distribution onto a set of charges, such as a subset of existing atomic positions or a supplementary set of discretized surface points for increased accuracy. This extension is designated the variational electrostatic projection with charge scaling (VEP-cs). Alternately, the original projection surface of the VEP-RVM method is retained, but in addition, an auxiliary subset of atomic positions is used to augment the projection surface to form the generalized VEP-RVM+cs method. These methods, as will be demonstrated below, have several important advantages: (1) the molecular simulation code modifications are simple, (2) the



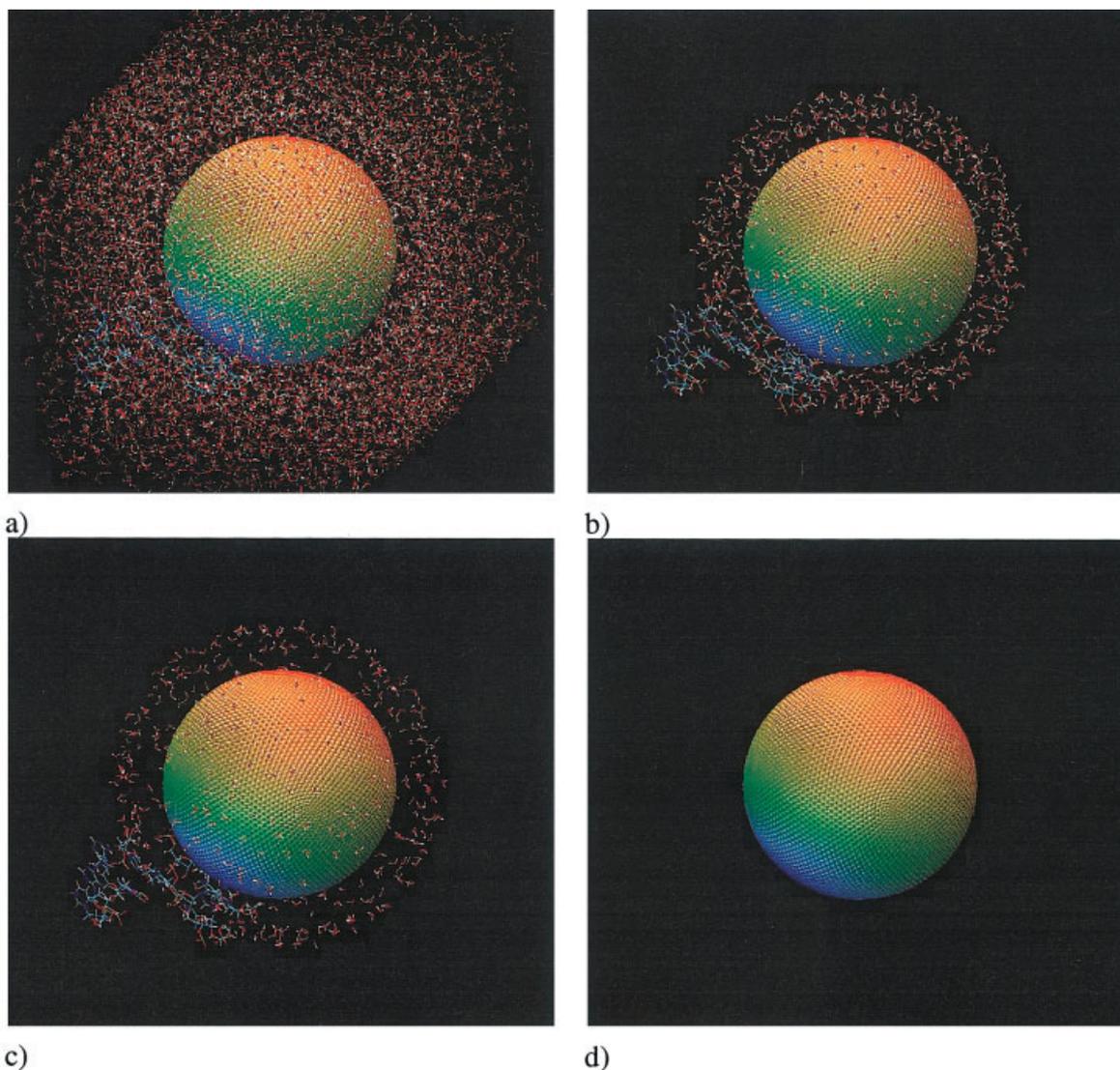
**Figure 1.** Regions defined VEP charge-scaling procedure. Regions I and II are the active dynamical regions, with the atoms of region I propagated using Newtonian dynamics, and those of the surrounding stochastic buffer (region II) of thickness  $\Delta R_{sb}$  propagated using Langevin dynamics. Region III (subdivided into IIIa, IIIa' and IIIb) comprises the external macromolecular and solvent environment that, in the present work, are held fixed. The electrostatic potential due to atoms of region IIIb, excluding those of the *projection buffer* (region IIIa) of thickness  $\Delta R_{pb}$  and *charge-scaling buffer* (region IIIa') of thickness  $\Delta R_{cs}$ , is replaced by the electrostatic potential of auxiliary charges placed at the centers of the atoms in the charge-scaling region. The  $\omega$  surface is used as an intermediate surface for variational reverse mapping in the VEP-RVM, VEP-cs and VEP-RVM+cs methods.

accuracy is considerably improved relative to the VEP and VEP-RVM methods, and (3) the methods afford more facile extension to macromolecular and solvent linear response.

With these extensions, some new control parameters are introduced that can be adjusted to balance the accuracy and computational efficiency afforded by the different methods. In the following sections, the new procedure is developed, with emphasis on the novel aspects not previously described. The first subsection outlines the setup for the different spatial regions of interest. The second subsection briefly outlines the essential background for the original VEP and VEP-RVM methods. The third subsection describes the new VEP-cs and VEP-RVM+cs methods.

### Setup for the VEP and Related Methods

The regions defined in the VEP and VEP-RVM methods, and new VEP-cs and VEP-RVM+cs methods are shown in Figure 1. Three variables,  $R_{II}$ ,  $\Delta R_{pb}$  and  $\Delta R_{cs}$  completely determine the regions or sets of atoms used in the VEP-cs and VEP-RVM+cs methods. First,  $R_{II}$  delineates the cutoff between active and frozen atoms. Second, a projection buffer of thickness  $\Delta R_{pb}$  (i.e.,  $R_{III} -$



**Figure 2.** The exact electrostatic potential due to atoms in region IIIb and implicit solvent surrounding the system displayed on a discretized sphere of radius 16.0 Å with 5810 surface points. For clarity, the implicit solvent surface layer surrounding the system is not shown. Potentials on the surface for the VEP, VEP-RVM, VEP-cs, and VEP-RVM+cs methods are indistinguishable from the exact result displayed. Also shown are (a) explicit solvent and atoms in the projection buffer ( $\Delta R_{pb} = 2.0$  Å) and the charge-scaling buffer ( $\Delta R_{cs} = 2.0$  Å), (b) explicit atoms in the projection buffer ( $\Delta R_{pb} = 2.0$  Å) and the charge-scaling buffer ( $\Delta R_{cs} = 2.0$  Å) only, (c) explicit atoms in the charge-scaling buffer ( $\Delta R_{cs} = 2.0$  Å) only, (d) the surface potential only.

$R_{II}$  is introduced that has been shown<sup>8</sup> to improve the force errors in region  $R_I$ . Atoms contained in  $\Delta R_{pb}$  are not used in the initial variational electrostatic projection, nor are their charges modified in the reverse-mapping process. They are simply held fixed and evaluated explicitly. Last, an additional layer of thickness  $\Delta R_{cs}$  is introduced external to  $\Delta R_{pb}$ . Atoms in this region have their charges augmented to best reproduce the field due to all the atoms outside of  $R_{III} + \Delta R_{cs}$ . The charge-scaling region can be used in addition to a discretized  $\gamma$  surface at  $r_\gamma$  in the combined VEP-RVM+cs approach. Singular value decomposition techniques are employed to calculate a set of additional charges for these fixed

atoms that best reproduce the potential due to the external environment. For the series of tests presented here, the external environment consists of an extensive layer of explicit solvent, surrounded by implicit solvent (see Methods).

Figure 2 shows the exact electrostatic potential generated by the external environment (region IIIb and implicit solvent) and illustrates the hierarchy of different regions considered in the charge-scaling *methods* for control parameter values of  $R_I = 16$  Å,  $R_{II} = 20$  Å,  $\Delta R_{pb} = 2.0$  Å, and  $\Delta R_{cs} = 2.0$  Å. The electrostatic potential of the VEP based methods are indistinguishable from the exact potential displayed.

### VEP and VEP-RVM Methods

In the direct VEP method, a single discretized projection surface is used as a basis for the variational projection procedure. This method's accuracy is limited partially by the projection of atomic charges that reside close to the discretized surface, especially when the distance of the atoms becomes comparable to the effective distance between discretized surface elements. Improved accuracy can be obtained using the variational electrostatic projection with reverse variational mapping (VEP-RVM) procedure. In the VEP-RVM procedure, an intermediate surface that resides interior to the VEP surface is used to *forward* project the atomic charges of the external environment (thus eliminating the issues related to proximity in the projection). Reverse variational mapping back to the VEP surface outside the active dynamical region is then used to derive the electrostatic potential and fields inside. The dual-surface approach provides considerably improved accuracy (for the same surface discretization level) relative to the direct VEP method.

To develop the present VEP-cs method, the key elements of the VEP and VEP-RVM methods are first reviewed. The VEP surface is designated the  $\gamma$  surface in Figure 1, and the intermediate surface used in the VEP-RVM method is designated the  $\omega$  surface. The discretization of the  $\gamma$  and  $\omega$  surfaces were 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,<sup>12</sup> first pioneered by Lebedev<sup>13</sup> and later extended<sup>14,15</sup> to very high order by Lebedev and Laikov.<sup>16</sup> Surface element interactions are modeled by Coulomb integrals between spherical Gaussian functions with exponents chosen to reproduce the exact variational energy and Gauss' law for a point charge in a spherical cavity. This procedure has recently been extended to very high order, and is described in detail elsewhere.<sup>17</sup>

The matrix notation used in the previous development<sup>8</sup> is continued here. The electrostatic interaction between discretized Gaussian surface elements of either the  $\gamma$  or  $\omega$  projection surfaces are designated by  $\mathbf{P}_{\alpha\beta}$ , where the subscripts indicate the two interacting projection spheres. Following the convention for the  $\mathbf{P}$  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  $\mathbf{B}$  and  $\mathbf{C}$ , respectively. For specific formulas for the matrix elements, see ref. 8.

The  $\omega$  surface (which lies interior to the  $\gamma$  surface) is an intermediate construct for obtaining a charge distribution on the  $\gamma$  surface that best reproduces the potential and forces due to the charge distribution of the external environment. The charges on the gamma surface are determined from

$$\gamma_{\mathbf{x}}(0) = [\mathbf{P}_{\omega\omega}^{-1} \cdot \mathbf{B}_{\omega\gamma}(0)]^{-1} \cdot \omega_{\mathbf{x}}(0), \quad (1)$$

where  $\omega_{\mathbf{x}}(0)$  is the surface charges on the  $\omega$  surface obtained from direct solution of the VEP equations, that is,

$$\omega_{\mathbf{x}}(0) = \mathbf{P}_{\omega\omega}^{-1} \cdot \mathbf{B}_{\omega\mathbf{x}}(0) \cdot \mathbf{X}. \quad (2)$$

In eqs. (1) and (2) it is assumed that the dimensions of the  $\mathbf{P}_{\omega\omega}$  and  $\mathbf{B}_{\omega\gamma}$  matrices are the same; that is, that the number of discretized surface elements of the  $\gamma$  and  $\omega$  surfaces are identical ( $N_{\gamma} = N_{\omega}$ ). If this is not the case, it still may be possible to find a satisfactory solution for  $\gamma_{\mathbf{x}}(0)$  if the equations are properly conditioned (see below). Note that eqs. (1) and (2) correspond to the *unconstrained* forms of the  $\gamma$  and  $\omega$  surfaces, indicated by the zero argument in parentheses. Unlike with the direct VEP method, negligible improvement of the accuracy was obtained through the use of the constrained formulation of the VEP-RVM method.<sup>8</sup>

The location of the temporary  $\omega$  surface used in the VEP-RVM method is given by the equation

$$\frac{r_{\omega}}{r_{\gamma}} = 1 - \frac{c_1 - c_2 \cdot \exp[-c_3 \cdot N]}{\sqrt{N} + c_0}, \quad (3)$$

where  $c_0$ ,  $c_1$ ,  $c_2$ , and  $c_3$  are unitless empirical parameters,  $N$  is the number of points used to discretize the  $\omega$  and  $\gamma$  surfaces ( $N = N_{\gamma} = N_{\omega}$ ), and  $r_{\omega}$  and  $r_{\gamma}$  are the radial distances to the  $\omega$  and  $\gamma$  surfaces, respectively. This functional form is also observed to work well for the VEP-cs method (see below).

### Charge-Scaling VEP-cs and VEP-RVM+cs Methods

The VEP-cs method is a generalization of the VEP-RVM method whereby an auxiliary set of points outside the  $\gamma$  surface can be used to model the electrostatic potential and fields instead of the  $\gamma$  surface elements themselves. As will be discussed in more detail below, this leads to improved accuracy and ease of code modification, as well as allowing more facile extension to the linear response of the solvated macromolecule (which will be addressed in future work).

The idea behind the VEP-RVM method is the following: first, the charge distribution of the external atoms are variationally projected onto the  $\omega$  surface. However, the  $\omega$  surface cannot itself be used as a basis for representing the electrostatic potential and forces because it resides interior to the  $\gamma$  surface; that is, inside the active dynamical region. Clearly, some of the atoms in the active dynamical region can lie on top of or even exterior to the  $\omega$  surface. A set of charges that reside outside the active dynamical region is desired that produces the same electrostatic potential on the  $\omega$  surface as did the entire set of external atoms, including that of the explicit and/or implicit solvent. If such a set of auxiliary charges can be found that number much less than the number of explicit external solute and solvent atoms, then considerable computational savings can be realized through use of this set of charges. In the VEP-RVM method, the discretized set of surface elements of the  $\gamma$  surface served as the auxiliary set of charges. However, there is no reason that an alternate set of charges could be used in place of, or in addition to, the  $\gamma$  surface elements. A natural choice for an alternate or augmented set of charges would be to consider a subset of the external atoms themselves.

This, in essence, is precisely the idea behind the VEP-cs method: a charge-scaling buffer (region IIIa' in Fig. 1) of thickness  $\Delta R_{cs}$  is chosen where atoms are selected by residue and used as a basis for the charge-scaling procedure. If these positions are used to entirely replace the  $\gamma$  surface elements as the basis for varia-

tional projection, the method is denoted VEP-cs. The advantage of the VEP-cs method is that, once the scaled charges are determined, essentially no code modifications (besides modification of the atomic charges) are required. If, on the other hand, the atoms in the charge-scaling buffer are to augment the  $\gamma$  surface elements, this method is designated the VEP-RVM+cs method. This method would allow higher accuracy to be achieved beyond the VEP-RVM method alone. The increase in accuracy anticipated with the VEP-RVM+cs method vs. the VEP-RVM method alone for a constant number of  $\gamma$  surface points is the result of two factors: (1) increased size of the basis set used to model the external potential, and (2) more atoms being treated explicitly (and therefore exactly) by not being variationally projected. It is anticipated that the  $\gamma$  surface constitutes a better general projection basis (in terms of number of degrees of freedom) by construction than does the atom positions of the charge-scaling buffer, but this is to some degree counterbalanced by the fact that the atoms of the charge-scaling buffer, which are the external atoms closest to the active dynamical region need not be explicitly projects (or, alternately stated, the atom positions of the charge scaling buffer constitute an exact basis for projection of the atoms in that region). A thorough direct comparison of the VEP-RVM and VEP-RVM+cs methods with the same number of degrees of freedom is difficult to do systematically and consistently because atoms in the charge-scaling buffer are chosen using a radial group-based cutoff.

The VEP-cs and VEP-RVM+cs methods both utilize a reverse variational mapping procedure that has been demonstrated previously to considerably improve the accuracy of the original direct VEP method.<sup>8</sup> In previous work the discretization levels of the  $\gamma$  surface and the intermediate  $\omega$  surfaces were chosen to be the same to ensure stable conditioning of the linear algebraic equations and balanced accuracy of the forward projection and reverse-mapping procedures. In the present work, the basis for reverse variational mapping is increased by inclusion of explicit atomic positions in the charge-scaling buffer. With this increase in basis, if the discretization of the  $\omega$  surface was unchanged, the forward mapping procedure would emerge as the obstacle to higher accuracy. The present work takes advantage of a recently introduced set of high-order discretization schemes and Gaussian exponents for boundary element solvation and VEP methods<sup>16</sup> that allows the  $\omega$  surface to be constructed in an accurate, systematic, numerically stable fashion. To circumvent problems associated with ill-conditioning and solution of underdetermined sets of linear equations in the reverse-mapping procedure, singular value decomposition methods<sup>18</sup> were used.

## Methods

### Activated Dynamics Simulation Model

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. Due to 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, stochastic boundary simulations are frequently performed.<sup>19,20</sup> The present VEP methods are intended to be used with stochastic boundary molecular simulations to decrease the computational cost of evaluating the long-ranged electrostatic interactions due to the external solvated macromolecular environment. In stochastic boundary calculations, typically a set 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 1:

- *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. In addition to the fixed solute and explicit solvent contained in this region, implicit solvation of the complete simulation system has also been added for the calculations in this work (see later)

The charge-scaling VEP method introduced in the present work is aimed at simplifying the problem of evaluating the electrostatic potential in a localized region (the active site of an enzyme, zone I) due to a large external charge distribution (macromolecule and solvent, zone III) surrounding the active site.

### Linear-Scaling Solvation

The effect of a static solvation reaction field can easily be included within the VEP, VEP-RVM, or charge-scaling methodologies. For the tests of the methods included here, a discretized surface using 50 points per atomic center was constructed 1.4 Å away from each atom in the model system. This resulted in a boundary containing 10,793 surface elements for the hammerhead system with which to approximate the dielectric response of embedding the system in a dielectric of 80.

Normally, the surface charges are directly obtained through inversion of the surface element interaction matrix. However, storage of this matrix rapidly becomes limiting as the number of surface element increases (the  $10,793 \times 10,793$  surface element interaction matrix requires nearly 890 MB when using 8-byte reals, and does not include any additional temporary storage required for actually performing the inverse). Using packed storage for the matrix essentially reduces the storage requirement in half, but requires more complicated code for performing matrix-matrix and matrix-vector operations. An alternative to the matrix inversion with or without packed storage is to use the preconditioned conjugate gradient minimization technique<sup>18</sup> to solve for the reaction-field surface charge  $\sigma$  through minimization of the function  $E_{\text{pol}}(\sigma) = \frac{1}{2}\sigma\mathbf{A}\sigma + \sigma\mathbf{B}\rho$  (with  $\mathbf{A}$  and  $\mathbf{B}$  defined as in ref. 10). Recursive bisection fast multipole methods<sup>21,22</sup> can also be applied when computing the electrostatic potentials at the surface positions ( $\mathbf{A}\sigma$  and  $\mathbf{B}\rho$ ). Preconditioned conjugate gradient minimization was used to obtain the reaction-field surface charge vector. A tolerance

of  $10^{-10}$  on the quantity  $|\mathbf{A}\sigma + \mathbf{B}\rho|/|\mathbf{B}\rho|$  was used to indicate a converged solution to the surface charges.

### Analysis of Force Errors in Hammerhead Ribozyme System

The hammerhead ribozyme system was examined to assess the accuracy of the charge-scaling method presented in the present work. The solvated system used for testing was constructed based on the 299D crystal structure of Scott et al.<sup>23</sup> This structure was solvated in a rhombododecahedral unit cell ( $a = b = c = 68.5$  Å,  $\alpha = \gamma = 60^\circ$ ,  $\beta = 90^\circ$ ) with the scissile phosphate centered at the origin. The final system contained 1329 solute atoms, 7187 TIP3P<sup>24</sup> water molecules, 130 sodium ions, and 89 chlorine ions. Equilibration of the solvent and ions was performed under periodic boundary conditions for 1 ns at 300 K using CHARMM<sup>25,26</sup> c28a1 program with the CHARMM27 All-Hydrogen Nucleic Acid<sup>27</sup> parameters. SHAKE<sup>28</sup> was used to constrain bonds containing hydrogen to a tolerance of  $10^{-7}$ . The solute was restrained to its crystallographic coordinates. Instead of variationally projecting the complete periodic potential, the final set of coordinates was placed within implicit solvent (to solvate the unit crystal cell). The final potential that is reproduced in this work is therefore that of the nonperiodic potential due to the set of wrapped coordinates (inside the unit cell) obtained from the end of the equilibration of the periodic explicit-solvent simulation, in addition to that of the implicit solvent surrounding the unit cell.

For the tests of the VEP based methods, the reaction zone was defined by a 16.0-Å sphere around the scissile phosphate ( $R_I = 16.0$  Å containing 1875 solute and solvent atoms) with a 4.0-Å stochastic buffer ( $\Delta R_{s,b} = 4.0$  Å) leading to an overall 20.0-Å region of active dynamics ( $R_{II} = 20.0$  Å containing 3638 total solute and solvent atoms). The active dynamics region was surrounded by projection buffers of 0.0, 2.0, and 4.0 Å containing 0, 1189, and 2592 fixed solute and solvent atoms, respectively.

The accuracy of the forces on the atoms in the active region due to modeled external environment were used to access the usefulness of the charge-scaling method. 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

$$\mathbf{f}_i^{\text{exact}} = \sum_j^{\text{zone III}} -\nabla_i \left( \frac{Q_i Q_j}{r_{ij}} \right) = \sum_j^{\text{zone III}} \frac{Q_i Q_j}{r_{ij}^3} \mathbf{r}_{ij}, \quad (4)$$

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  $r_{ij} = |\mathbf{r}_{ij}|$  is their internuclear separation. The sum involving index  $j$  runs over atoms of the external environment (i.e., of zone III). The corresponding errors ( $\text{ERR}_i$ ) in the electrostatic force on atom  $i$  are given by

$$\text{ERR}_i = |\mathbf{f}_i - \mathbf{f}_i^{\text{exact}}|. \quad (5)$$

The average error over the entire reaction zone, or over radial shells of the reaction zone, was obtained by straight averaging of the corresponding  $\text{ERR}_i$  values. For the purposes of error analysis,

it is more useful to consider an average *relative error* ( $\langle \text{RELE} \rangle$ ) in the external electrostatic force as defined by

$$\langle \text{RELE} \rangle = \langle \text{ERR} \rangle / \langle \mathbf{F}^{\text{exact}} \rangle, \quad (6)$$

where  $\langle \text{ERR} \rangle$  is the average  $\text{ERR}_i$  value and  $\langle \mathbf{F}^{\text{exact}} \rangle$  is the average  $|\mathbf{f}_i^{\text{exact}}|$  value and the average is taken over a set of atoms. In the tables, this set of atoms includes all atoms in the entire reaction zone (zone I), and for the figures that show a radial distribution of average relative errors, the sets of atoms are those of evenly spaced radial shells. It is generally acceptable for molecular simulations when the relative force errors due to long-range electrostatics fall below  $10^{-4}$ – $10^{-5}$ .<sup>29,30</sup>

## Results and Discussion

It is a purpose of the present work to systematically explore the accuracy and convergence properties of the VEP–cs and VEP–RVM+cs methods to establish a recommendation for performing activated dynamical simulations with QM/MM methods. It has been established that a generally acceptable level of relative force error for molecular simulations ranges between  $10^{-4}$ – $10^{-5}$ ,<sup>29,30</sup> for which the present methods (demonstrated below) are shown to easily obtain throughout zone I of the active dynamical region. Nonetheless, it remains to further test the method with molecular simulation for a host of properties and over a diverse array of heterogeneous systems to fully establish the limits of reliability of the various VEP methods. This section presents relative force error results as a function of the control parameters in the model, in particular, the effect of:

- variation of the discretization level of the  $\gamma$  surface ( $N_\gamma$ )
- variation of the discretization level of the  $\omega$  surface ( $N_\omega$ )
- variation of the projection buffer ( $\Delta R_{pb}$ )
- variation of the charge-scaling buffer ( $\Delta R_{cs}$ )

Tables 1–3 compare the average relative force errors of the atoms in the active dynamical region (zone I) for several different variational projection methods (VEP, VEP–RVM, VEP–cs, and VEP–RVM+cs) as a function of  $N_\omega$ ,  $N_\gamma$ , and  $\Delta R_{cs}$  for projection buffer values of  $\Delta R_{pb} = 0.0$  Å (Table 1),  $\Delta R_{pb} = 2.0$  Å (Table 2), and  $\Delta R_{pb} = 4.0$  Å (Table 3).

The radial distribution of average force errors are show in Figures 3–5 at various discretization levels for the VEP–RVM and VEP–RVM+cs methods using a 2.0 Å projection buffer,  $N_\gamma = 1202$  points for the  $\gamma$  surface, and charge scaling buffer zones of  $\Delta R_{cs} = 0.0$  Å (Fig. 3), 2.0 Å (Fig. 4), and 4.0 Å (Fig. 5). The solute and solvent atoms are assigned to the different regions on a residue basis: if any atom of a residue is within a more interior region, the whole residue is defined to be within that region. In the above, the charge-scaling buffer ( $\Delta R_{cs}$ ) implies the region used for augmentation of the *solvent atom* charges, because the external hammerhead ribozyme solute atoms (not already included in the projection buffer) are considered to be within the charge-scaling region. Note that any atoms in the charge-scaling region (the solute atoms external to the projection buffer or the solvent atoms in  $\Delta R_{cs}$ ) are *not* variationally projected, but instead treated explicitly.

**Table 1.** Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using the VEP, VEP-RVM, VEP-cs, and VEP-RVM+cs Methods with  $\Delta R_{pb} = 0.0$  Å.

Method	$N_\omega$	$N_\gamma = 0$	$\langle \text{RELE} \rangle$		
			$N_\gamma = 302$	$N_\gamma = 590$	$N_\gamma = 1202$
$\Delta R_{cs} = 0.0$ Å (680 atoms)					
VEP	—		1.14E-01	7.43E-02	5.74E-02
VEP-RVM	1202		1.47E-02	3.45E-03	3.87E-04
	3890		1.41E-02	3.26E-03	3.89E-04
	5810		1.41E-02	3.21E-03	3.75E-04
VEP-cs	1202	1.41E+03			
	3890	1.37E-01			
	5810	6.43E-02			
VEP-RVM+cs	1202		1.96E+02	2.07E+02	3.96E-04
	3890		1.64E-02	5.16E-03	2.62E-03
	5810		9.78E-03	2.58E-03	3.20E-04
$\Delta R_{cs} = 2.0$ Å (1690 atoms)					
VEP	—		1.55E-02	3.93E-03	8.15E-04
VEP-RVM	1202		4.05E-03	5.18E-04	1.97E-05
	3890		3.91E-03	4.87E-04	1.97E-05
	5810		3.89E-03	4.79E-04	1.93E-05
VEP-cs	1202	4.05E-03			
	3890	7.12E-05			
	5810	4.64E-05			
VEP-RVM+cs	1202		1.87E-04	1.79E-04	2.33E-05
	3890		3.32E-05	2.45E-05	2.62E-05
	5810		1.85E-05	9.88E-06	2.67E-06
$\Delta R_{cs} = 4.0$ Å (2950 atoms)					
VEP	—		3.44E-03	1.08E-03	3.66E-04
VEP-RVM	1202		1.51E-03	1.01E-04	1.95E-06
	3890		1.47E-03	9.84E-05	1.95E-06
	5810		1.47E-03	9.77E-05	1.93E-06
VEP-cs	1202	7.06E-06			
	3890	5.10E-08			
	5810	1.83E-08			
VEP-RVM+cs	1202		9.82E-06	3.69E-05	6.49E-06
	3890		3.32E-08	3.71E-08	2.69E-07
	5810		1.13E-08	9.17E-09	1.96E-09

This table compares the accuracy at various discretization levels of the  $\gamma$  and  $\omega$  surface for the VEP, VEP-RVM, VEP-cs, and VEP-RVM+cs methods with a projection buffer of  $\Delta R_{pb} = 0.0$  Å and charge-scaling buffers of  $\Delta R_{cs} = 0.0, 2.0$  and  $4.0$  Å. For the direct VEP method, no  $\omega$  surface is used and  $r_\gamma$  is set at  $R_{II} = 20.0$  Å. For the VEP-cs method, no  $\gamma$  surface is used. For the VEP-RVM and VEP-RVM+cs methods  $r_\gamma$  was set at  $R_{III} = R_{II} + \Delta R_{pb}$ . See text for further details.

Hence, for consistency with the charge-scaling methods, the average force errors for the VEP and VEP-RVM methods will depend on  $\Delta R_{cs}$ , because for larger values of  $\Delta R_{cs}$ , less atoms are actually being projected and more are being treated explicitly (and therefore exactly). Note also that the placement of the  $\gamma$  surface in the VEP-RVM+cs method is between the projection and charge-scaling buffers (zones IIIa and IIIa', respectively). This placement is identical to that used in the VEP-RVM method. For the direct VEP method, the  $\gamma$  surface is located between zones II and IIIa.

#### Comparison of Force Errors in the Direct VEP and VEP-RVM Methods

The force errors in the direct VEP method are in general the poorest, as has been discussed in detail elsewhere.<sup>8</sup> Reasonably high accuracy

can be obtained with the direct VEP method using symmetric (equal)  $\Delta R_{pb}$  and  $\Delta R_{cs}$  buffers: the average force error values with the direct VEP method (with  $N_\gamma = 1202$ ) are  $5.7 \times 10^{-2}$ ,  $5.6 \times 10^{-4}$  and  $1.5 \times 10^{-4}$  for  $\Delta R_{pb} = \Delta R_{cs} = 0.0, 2.0$  and  $4.0$  Å, respectively. The VEP-RVM method performs considerably better with corresponding average force error values (with  $N_\gamma = 1202$ ) of  $3.9 \times 10^{-4}$ ,  $1.2 \times 10^{-6}$ , and  $1.039 \times 10^{-8}$  for  $\Delta R_{pb} = \Delta R_{cs} = 0.0, 2.0,$  and  $4.0$  Å, respectively. Note that increasing the number of surface elements on the  $\omega$  surface, ( $N_\omega$ ), has negligible effect on the average force errors. In the original VEP paper, the number of surface elements for the  $\omega$  and  $\gamma$  surfaces were taken to be the same (i.e.,  $N_\omega = N_\gamma$ ). This was reasonable because this ensured that certain connection matrices were square and nonsingular for all the cases presented. In the current work, the value of  $N_\omega$  can be set to larger values for the forward projection

**Table 2.** Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using the VEP, VEP-RVM, VEP-cs, and VEP-RVM+cs Methods with  $\Delta R_{pb} = 2.0$  Å.

Method	$N_\omega$	$N_\gamma = 0$	$\langle RELE \rangle$ $N_\gamma = 302$	$N_\gamma = 590$	$N_\gamma = 1202$
$\Delta R_{cs} = 0.0$ Å (501 atoms)					
VEP	—		8.30E-02	7.30E-02	3.55E-02
VEP-RVM	1202		3.05E-03	4.01E-04	1.48E-05
	3890		2.96E-03	3.79E-04	1.49E-05
	5810		3.04E-03	3.82E-04	1.49E-05
VEP-cs	1202	1.84E+02			
	3890	3.80E-02			
	5810	4.35E-02			
VEP-RVM+cs	1202		5.71E+00	2.78E+00	1.55E-05
	3890		1.71E-03	2.56E-04	1.68E-05
	5810		1.74E-03	2.59E-04	1.72E-05
$\Delta R_{cs} = 2.0$ Å (1761 atoms)					
VEP	—		1.07E-02	2.96E-03	5.56E-04
VEP-RVM	1202		8.67E-04	6.04E-05	1.20E-06
	3890		8.41E-04	5.75E-05	1.20E-06
	5810		8.55E-04	5.78E-05	1.21E-06
VEP-cs	1202	1.41E-05			
	3890	1.45E-06			
	5810	2.16E-06			
VEP-RVM+cs	1202		6.38E-06	9.06E-06	1.27E-06
	3890		6.30E-07	2.90E-07	3.21E-07
	5810		1.03E-06	4.25E-07	7.44E-08
$\Delta R_{cs} = 4.0$ Å (3362 atoms)					
VEP	—		2.08E-03	6.43E-04	2.17E-04
VEP-RVM	1202		2.46E-04	1.12E-05	9.62E-08
	3890		2.39E-04	1.07E-05	9.61E-08
	5810		2.44E-04	1.07E-05	9.65E-08
VEP-cs	1202	3.92E-07			
	3890	1.96E-10			
	5810	2.96E-10			
VEP-RVM+cs	1202		4.42E-07	1.49E-06	2.94E-07
	3890		3.30E-10	9.87E-10	1.52E-10
	5810		1.68E-10	9.05E-11	5.87E-11

This table compares the accuracy at various discretization levels of the  $\gamma$  and  $\omega$  surface for the VEP, VEP-RVM, VEP-cs and VEP-RVM+cs methods with a projection buffer of  $\Delta R_{pb} = 2.0$  Å and charge-scaling buffers of  $\Delta R_{cs} = 0.0, 2.0$  and  $4.0$  Å. For the direct VEP method, no  $\omega$  surface is used and  $r_\gamma$  is set at  $R_{II} = 20.0$  Å. For the VEP-cs method, no  $\gamma$  surface is used. For the VEP-RVM and VEP-RVM+cs methods  $r_\gamma$  was set at  $R_{III} = R_{II} + \Delta R_{pb}$ . See text for further details.

in some cases because, in the charge-scaling methods, there can be an effectively larger basis for the reverse projection procedure that would benefit from a higher discretization of the  $\omega$  surface.

#### Comparison of Force Errors in the VEP-cs Method

The VEP-cs method is one that is particularly simplistic in that, once the set of scaled charges have been determined, activated dynamics simulations can be performed without further code modifications: only the modified set of atomic partial charges that mimic the full external electrostatic environment are needed.

The number of charges that are used as a basis for the reverse-mapping procedure are controlled by the size of the charge-scaling buffer ( $\Delta R_{cs}$ ). The VEP-cs method utilizes an intermediate  $\omega$

surface for the forward projection, but the reverse-mapping procedure is onto a set of atomic positions alone; that is, there is no  $\gamma$  surface in the VEP-cs method (and, hence, no code modifications are required to include the  $\gamma$  surface). Consequently, the accuracy of the VEP-cs method will depend ultimately on the degree to which the atoms of the charge-scaling buffer are able to act as a basis for the electrostatic environment of the fully solvated system. This aspect is controlled by the  $\Delta R_{cs}$  parameter. However, to achieve the best possible reverse-mapping result for a given  $\Delta R_{cs}$ , the discretization of the  $\omega$  surface must also be optimized. The accuracy of the forward projection is controlled by  $\Delta R_{pb}$  and  $N_\omega$ .  $\Delta R_{pb}$  provides an extra buffer between region II and the atoms in region III, in particular, those that are to be variationally

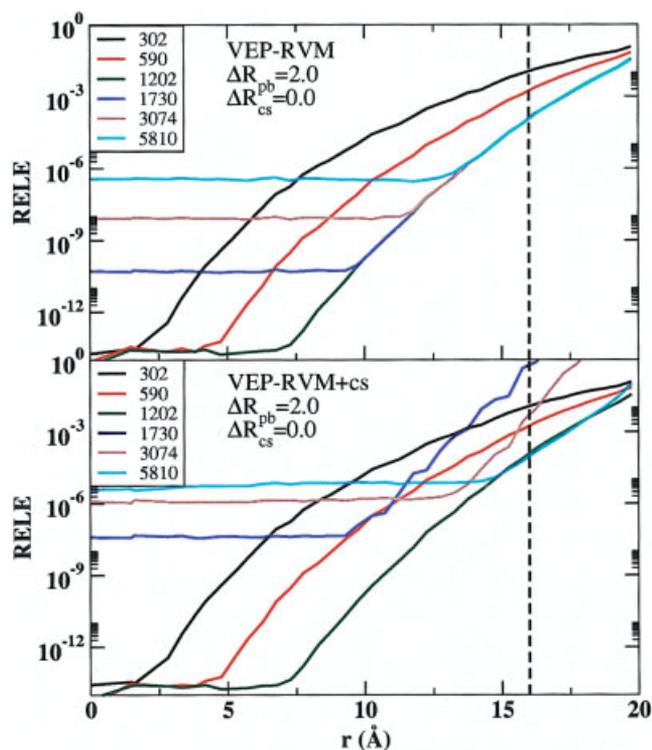
**Table 3.** Average Relative Force Errors for Atoms Inside Reaction Zone I (16.0 Å) of the Hammerhead Ribozyme System Using the VEP, VEP-RVM, VEP-cs, and VEP-RVM+cs Methods with  $\Delta R_{pb} = 4.0$  Å.

Method	$N$	$N_\gamma = 0$	$\langle \text{RELE} \rangle$		
			$N_\gamma = 302$	$N_\gamma = 590$	$N_\gamma = 1202$
$\Delta R_{cs} = 0.0$ Å (358 atoms)					
VEP	—		4.52E-02	4.25E-02	3.28E-02
VEP-RVM	1202		7.27E-04	5.16E-05	1.03E-06
	3890		7.24E-04	5.11E-05	1.06E-06
	5810		7.73E-04	5.65E-05	1.26E-06
VEP-cs	1202	1.12E+00			
	3890	3.56E-02			
	5810	4.36E-02			
VEP-RVM+cs	1202		6.86E-02	3.76E-02	1.04E-06
	3890		5.12E-04	3.85E-05	1.24E-06
	5810		5.68E-04	4.34E-05	1.68E-06
$\Delta R_{cs} = 2.0$ Å (1959 atoms)					
VEP	—		8.18E-03	3.21E-03	5.64E-04
VEP-RVM	1202		1.88E-04	9.09E-06	7.24E-08
	3890		1.88E-04	8.82E-06	7.33E-08
	5810		2.04E-04	9.38E-06	8.19E-08
VEP-cs	1202	6.22E-07			
	3890	4.32E-08			
	5810	1.35E-07			
VEP-RVM+cs	1202		3.05E-07	4.19E-07	8.08E-08
	3890		2.17E-08	1.06E-08	2.98E-09
	5810		1.57E-07	4.79E-08	1.17E-08
$\Delta R_{cs} = 4.0$ Å (3794 atoms)					
VEP	—		1.53E-03	4.47E-04	1.47E-04
VEP-RVM	1202		8.99E-05	2.32E-06	1.01E-08
	3890		8.97E-05	2.28E-06	1.03E-08
	5810		9.67E-05	2.46E-06	1.14E-08
VEP-cs	1202	3.00E-08			
	3890	1.49E-11			
	5810	4.68E-11			
VEP-RVM+cs	1202		3.96E-08	9.93E-08	2.11E-08
	3890		1.83E-12	7.82E-13	9.16E-13
	5810		2.77E-11	3.34E-11	1.24E-11

This table compares the accuracy at various discretization levels of the  $\gamma$  and  $\omega$  surface for the VEP, VEP-RVM, VEP-cs, and VEP-RVM+cs methods with a projection buffer of  $\Delta R_{pb} = 4.0$  Å and charge-scaling buffers of  $\Delta R_{cs} = 0.0, 2.0$  and  $4.0$  Å. For the direct VEP method, no  $\omega$  surface is used and  $r_\gamma$  is set at  $R_{II} = 20.0$  Å. For the VEP-cs method, no  $\gamma$  surface is used. For the VEP-RVM and VEP-RVM+cs methods  $r_\gamma$  was set at  $R_{III} = R_{II} + \Delta R_{pb}$ . See text for further details.

projected (region IIIb). The discretization level of the  $\omega$  surface,  $N_\omega$ , further controls the accuracy of the forward projection of the atoms in region IIIb. The  $N_\omega$  value, however, cannot be infinite or the reverse-mapping procedure becomes ill-conditioned. Even with singular value decomposition,<sup>18</sup> the methods become numerically less stable and very high accuracy is not easily achieved. Hence, the optimal value of  $N_\omega$  involves a delicate balance between being as accurate as possible for the forward projection, without corrupting the conditioning of the reverse-mapping procedure. Algorithms that implement the VEP-cs method should take care to check explicitly the conditioning and null space of the matrices associated with the reverse-mapping procedure in determining if the discretization level of the  $\omega$  surface is appropriate (see Supplementary Information).

The VEP-cs method is not particularly sensitive to the  $\Delta R_{pb}$  value, because this parameter does not affect the size of the basis used for the charge-scaling procedure. Instead, the  $\Delta R_{cs}$  controls the number of solvent molecules that are used, in addition to the entire set of solute atom positions, as the basis for the charge-scaling procedure. Hence, for values of  $\Delta R_{cs} = 0.0$  Å, the charge-scaling basis consists solely of the solute atom positions outside of the projection buffer (although note that as  $\Delta R_{pb}$  increases, the atoms of region IIIb that are variationally projected are pushed farther away from the active dynamical region). The errors for the VEP-cs method with no charge-scaling buffer ( $\Delta R_{cs} = 0.0$  Å), for example, decrease less than an order of magnitude (from  $6.4 \times 10^{-2}$  to  $3.6 \times 10^{-2}$ ) when the  $\Delta R_{pb}$  value increased from 0 to  $4.0$  Å. However, the VEP-cs errors become more



**Figure 3.** Figures showing the radial force errors associated with a 2.0 Å projection buffer (1189 atoms), a 0.0 Å charge-scaling region (501 atoms), and a 1202 point  $\gamma$  surface located at  $r_\gamma = R_{\text{III}} = 22.0$  Å for different discretization levels of the  $\omega$  surface. The optimal location of the gamma surface is determined using eq. (3). (Top) Results using the VEP-RVM method. (Bottom) Results for the VEP-RVM method including charge-scaling. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

sensitive to  $\Delta R_{pb}$  as  $\Delta R_{cs}$  increases, for example, with a  $\Delta R_{cs}$  value of 4.0 Å, the VEP-cs errors decrease from  $1.8 \times 10^{-8}$  with  $\Delta R_{pb} = 0.0$  Å to  $4.7 \times 10^{-11}$  with  $\Delta R_{pb} = 4.0$  Å. The VEP-cs errors are clearly the most sensitive to  $\Delta R_{cs}$  that adjusts the size of the basis in the reverse-mapping procedure, and is ultimately used in the calculation of the forces in the active dynamical region.

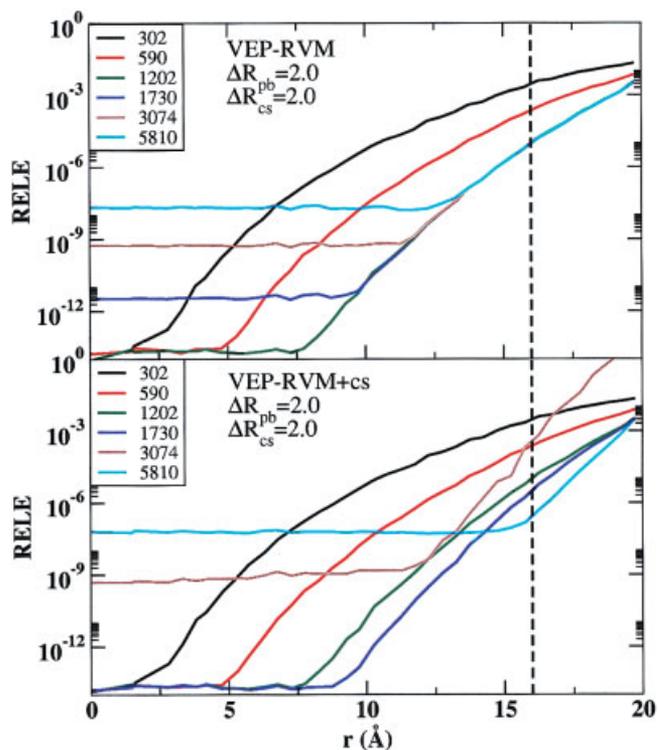
#### Comparison of Force Errors in the VEP-RVM+cs Method

The ultimate goal in the development of the new VEP charge-scaling methods presented in the current work is to provide tools that allow the best overall accuracy for the least computational expense. As has already been seen with the VEP-RVM method, the use of a  $\gamma$  surface provides a robust, accurate basis for the reverse variational mapping procedure with a relatively small number of points, while the VEP-cs method affords perhaps a less optimal (in terms of efficiency) but easily implemented alternative. Future work will extend the use of the  $\gamma$  surface to serve as a basis for expansion of a Green's function to include the electrostatic linear-response of the solute and solvent. In this way, the  $\gamma$  surface will play an important role in future extensions of variational electrostatic projection methods for improved modeling of the

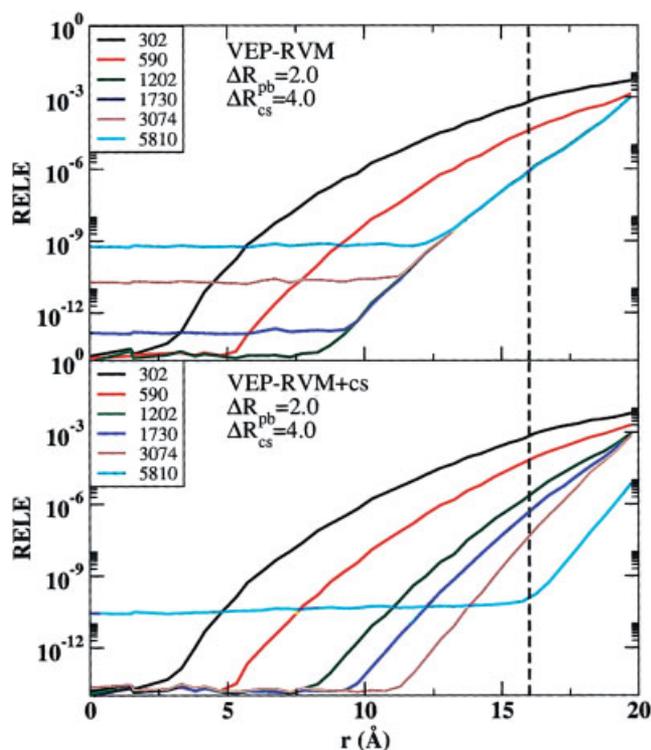
solvated macromolecular environment in activated dynamics simulations of biocatalytic processes. However, it might be the case that a level of accuracy beyond that provided by the  $\gamma$  surface alone is desired for projection of the electrostatic potential of the environment. In this case, one might wish to augment the variational projection basis of the  $\gamma$  surface with a charge-scaling buffer to create a combined VEP-RVM and VEP-cs method, designated VEP-RVM+cs.

The VEP-RVM+cs method affords the highest accuracy of all the VEP methods. The use of an explicit set of atomic positions to augment the projection basis used in the reverse variational mapping procedure significantly improves the relative force errors in the active dynamical region. For example, in Table 2, with only a relatively small charge-scaling buffer ( $\Delta R_{cs} = 2.0$  Å), the lowest VEP-RVM errors are  $8.4 \times 10^{-4}$ ,  $5.8 \times 10^{-5}$ , and  $1.2 \times 10^{-6}$  for  $N_\gamma$  values of 302, 590, and 1202, respectively, whereas the force errors in the VEP-RVM+cs method are reduced to  $6.3 \times 10^{-6}$ ,  $2.9 \times 10^{-7}$ , and  $7.4 \times 10^{-8}$ , respectively.

As might be expected, the VEP-RVM+cs method is sensitive to both the discretization level of the  $\gamma$  surface ( $N_\gamma$ ), and the size of the charge-scaling buffer ( $\Delta R_{cs}$ ). Together, these parameters control the size of the basis used in the reverse variational mapping



**Figure 4.** Figures showing the radial force errors associated with a 2.0 Å projection buffer (1189 atoms), a 2.0 Å charge-scaling region (1761 atoms), and a 1202 point  $\gamma$  surface located at  $r_\gamma = R_{\text{III}} = 22.0$  Å for different discretization levels of the  $\omega$  surface. The optimal location of the gamma surface is determined using eq. (3). (Top) Results using the VEP-RVM method. (Bottom) Results for the VEP-RVM method including charge-scaling. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]



**Figure 5.** Figures showing the radial force errors associated with a 2.0 Å projection buffer (1189 atoms), a 4.0 Å charge-scaling region (3362 atoms), and a 1202 point  $\gamma$  surface located at  $r_\gamma = R_{III} = 22.0$  Å for different discretization levels of the  $\omega$  surface. The optimal location of the gamma surface is determined using eq. (3). (Top) Results using the VEP-RVM method. (Bottom) Results for the VEP-RVM method including charge-scaling. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

procedure. The augmented charge-scaling basis in the VEP-RVM+cs method raises an important issue with regard to the discretization level of the intermediate  $\omega$  surface. In the VEP-RVM method, if the discretization level of the  $\omega$  surface (used for the forward projection) greatly exceeds that of the  $\gamma$  surface (used in the reverse mapping), ill-conditioning of the transformation matrix can lead to numerical inaccuracies. A balanced basis for the forward projection and reverse variational mapping can be achieved by choosing the discretization levels of the  $\omega$  and  $\gamma$  surfaces to be identical ( $N_\omega = N_\gamma$ ).

However, with the augmented charge-scaling buffer, the intermediate  $\omega$  surface may remain well-conditioned at discretization levels significantly exceeding that of the  $\gamma$  surface. In this case, increasing  $N_\omega$  appropriately will increase the accuracy of the forward projection, which otherwise would be a bottleneck to the overall accuracy. This is illustrated in Tables 1–3 by examination of the reduction of relative error values as  $N_\omega$  is increased. For example, in Table 1, the values of the relative error for the VEP-RVM+cs method ( $N_\gamma = 1202$ ,  $\Delta R_{cs} = 4.0$  Å) are  $6.5 \times 10^{-6}$ ,  $2.7 \times 10^{-7}$ , and  $2.0 \times 10^{-9}$  for  $N_\omega$  values of 1202, 3890, and 5810, respectively. Note that the discretization of the intermediate  $\omega$  surface requires inversion and linear algebraic manipula-

tion of matrices with larger dimensions (larger  $N_\omega$  values), but once these manipulations are mapped down to the dimension of the  $\gamma$  surface, the computation of the energy and forces no longer depends on  $N_\omega$ . Hence, there is no real impetus for use of anything but the optimal  $N_\omega$  value. It should be emphasized, as with the VEP-cs method, that the algorithm used to choose the proper discretization of the  $\omega$  surface should take into consideration the conditioning and null space of the matrices associated with the reverse-mapping procedure (see Supplementary Information).

#### Comparison of Radial Force Errors with the in VEP-RVM and VEP-RVM+cs Methods

It is important not only to assess the average relative errors with the VEP methods, but also the distribution of errors within the active dynamical regions. Toward this end, the radial distribution of average relative force errors are shown in Figures 3–5 at various discretization levels of the  $\omega$  surface ( $N_\omega$  values) and charge-scaling buffers ( $\Delta R_{cs}$  values) for the VEP-RVM and VEP-RVM+cs methods using a 2.0 Å projection buffer ( $\Delta R_{pb} = 2.0$  Å) and discretization level of 1202 points ( $N_\gamma = 1202$ ).

Figure 3 compares result with a 0.0 Å charge-scaling buffer for which 501 solute atoms outside the projection buffer, in addition to the 1202 points of the  $\gamma$  surface, were used as the basis for the reverse variational mapping procedure. With this small set of atomic positions used to augment the basis in the reverse-mapping procedure, little improvement of performance of the VEP-RVM+cs method relative to the VEP-RVM method is observed. The errors for different  $N_\omega$  values up to 1202 show systematically improved radial distribution of relative errors. Very high accuracy (essentially machine precision) is achieved near the center of the active dynamical region. As the  $N_\omega$  values increase to 1202, the range of very high accuracy is pushed toward larger distances from the center. However, for  $N_\omega$  value greater than 1202, the accuracy of the reverse-mapping procedure is retarded by ill-conditioning (recall that  $N_\gamma = 1202$ ).

Figure 4 compares result with a 2.0-Å charge-scaling buffer for which 1761 solute atoms outside the projection buffer, in addition to the 1202 points of the  $\gamma$  surface, were used as the basis for the variational reverse-mapping procedure. Whereas the errors for the VEP-RVM method show little improvement with respect to those of Figure 3, the increased  $\Delta R_{cs} = 2.0$  Å value allows the VEP-RVM+cs method to have significantly improved radial errors. With the increased reverse-mapping basis for the VEP-RVM+cs method, the discretization of the  $\omega$  surface used in the forward mapping can be increased to  $N_\omega = 1730$ , and allows the range of very high-precision to be extended to larger radial distance, as well as shifting all the radial relative errors values in the active dynamical region to smaller values.

Figure 5 compares result with a 4.0-Å charge-scaling buffer for which 3362 solute atoms outside the projection buffer were used, in addition to the 1202 points of the  $\gamma$  surface, as the basis for the variational reverse-mapping procedure. For  $\Delta R_{cs} = 4.0$  Å, considerable improvement of the radial distribution of relative force errors are observed for the VEP-RVM+cs method, shifting the range of very high accuracy out beyond 12.0 Å, and increasing the optimal resolution of the  $\omega$  surface to  $N_\omega = 3074$ . It is of interest to note that with  $N_\omega = 5810$ , the relative errors remain below

$10^{-10}$  all the way out to 16.0 Å—the full extent of region I of the active dynamical region that undergoes Newtonian dynamics without the constraints or stochastic forces present in the stochastic buffer zone (region II).

The radial distribution of average errors indicate that the largest errors occur for atoms nearest the  $\gamma$  surface; however, for moderate discretization levels and charge-scaling buffer sizes, the error at the boundary of the stochastic region remains below an acceptable level of  $10^{-4}$ .

## Conclusion

New charge-scaling methods for efficient modeling of the solvated macromolecular environment in hybrid QM/MM calculations of biological reactions are presented. The methods are extensions of the variational electrostatic projection method that allow atomic charges in the external environment to be adjusted so as to mimic, in the active dynamical region, the electrostatic potential and field due to the large surrounding macromolecule and solvent. The method has the advantages that it offers improved accuracy, does not require the use of a three-dimensional grid or auxiliary set of fitting points, and requires little or no molecular simulation code modifications.

Two variants of the charge-scaling method are described: the VEP-cs method and the VEP-RVM+cs method. Each method has advantages and are able to attain very high accuracy (below  $10^{-4}$ – $10^{-5}$  relative error in the expected force) with proper choice of control parameters ( $R_{II}$ ,  $\Delta R_{pb}$ ,  $\Delta R_{cs}$ ,  $N_\omega$  and  $N_\gamma$ ). Increasing  $N_\omega$  and  $N_\gamma$  results in an immediate increase in accuracy at the expense of additional computational cost. Treating more atoms explicitly by not variationally projecting them (increasing the sum of  $R_{II}$ ,  $\Delta R_{pb}$ , and  $\Delta R_{cs}$ ) also allows for increased accuracy. Finally, increasing the separation distance between the active atoms and the projected atoms (larger  $\Delta R_{pb}$ ) also results in increased accuracy. It is the purpose of this work to characterize in detail the force accuracy of each method as a function of the control parameters to determine the limitations of the methods and outline practical prescriptions for use in activated dynamics simulations.

The VEP-cs method is a straightforward prescription for determining a variational set of charge adjustments that best mimic the electrostatics of the surrounding environment. The accuracy of the VEP-cs method is determined by the subset of atomic positions used as a basis for the determination of the electrostatic potential and fields. Relative force errors on the order of  $10^{-8}$  or better are achieved with a charge scaling buffer ( $\Delta R_{cs}$  value) of 4.0 Å surrounding the active dynamical region. The VEP-cs method requires no molecular simulation code modifications. It is simply a well-defined means of rescaling charges of a set of atoms immediately surrounding the active dynamical region so as to closely reproduce, inside the active dynamical region, the electrostatics of the full external environment that might contain orders of magnitude more particles. For simulations using only atomic positions as a basis (the VEP-cs method) moderate accuracy on the order of  $10^{-4}$ , can be obtained using  $\Delta R_{pb} = 0$  and  $\Delta R_{cs} = 2.0$  (which also includes the entire solute. For high accuracy ( $10^{-6}$ – $10^{-7}$  relative force error),  $\Delta R_{pb} = 0$  and  $\Delta R_{cs} = 4.0$  is recom-

mended. Finally, for benchmark accuracy calculations ( $10^{-10}$ ),  $\Delta R_{pb} = 4.0$  and  $\Delta R_{cs} = 4.0$  are recommended. For reduced computational cost, and/or if the VEP method is to be used with solute and solvent linear response, then the VEP-RVM or VEP-RVM+cs approaches can be used. The effect of these combinations in molecular simulations that include solute and solvent linear response is forthcoming. However, the present work explores the possibility of using atomic positions as a basis for variational projection of the solvated macromolecular electrostatic environment.

The VEP-RVM+cs method is an extension of the VEP-RVM method, which offers improved accuracy and affords a mechanism for further generalization to treat the environmental linear response, which is a topic of future work. In this method, the projection surface (the  $\gamma$  surface) is augmented with an explicit set of atomic positions to use as a basis for the VEP-RVM procedure. Very high accuracy is attainable with this method with proper choice of control parameters. Relative force errors on the order of  $10^{-7}$  are possible with a 590-point  $\gamma$  surface and 2.0-Å charge-scaling buffer, and relative force errors better than  $10^{-10}$  are possible with a 1202-point  $\gamma$  surface and 4.0-Å charge-scaling buffer. Comparison of the radial distribution of relative force errors indicate that the VEP-RVM+cs method can considerably extend the region of very high accuracy within the active dynamical region.

The VEP-cs and VEP-RVM+cs methods are both methods that utilize an intermediate  $\omega$  surface and reverse variational mapping procedure. This technique provides considerably improved accuracy relative to the original direct VEP method.<sup>8</sup> However, in the present work, the discretization level of the  $\omega$  surface was required to have considerably higher accuracy than in the previous work, because with the charge-scaling methods, the basis for reverse variational mapping is considerably increased. With this increase in basis size, if the discretization of the  $\omega$  surface was unchanged, the forward mapping procedure would be an obstacle to higher accuracy. To remedy this problem, the intermediate  $\omega$  surfaces were constructed using a recently introduced set of high-order discretization schemes and Gaussian exponents for boundary element solvation and VEP methods.<sup>16</sup>

The VEP-cs and VEP-RVM+cs methods serve as potentially powerful tools in the arsenal of computational techniques used in multiscale computational modeling problems. It is the hope that these methods will improve the accuracy and computational efficiency of modeling the complex chemical environments encountered in hybrid QM/MM activated dynamics simulations of biocatalytic processes, and provide deeper insight into important related biomedical applications.

## Acknowledgments

Computational resources were provided by the Minnesota Supercomputing Institute.

## References

1. Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley and Sons: New York, 1991.

2. Warshel, A. *Acc Chem Res* 2002, 35, 385.
3. Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* 2004, 303, 186.
4. Nam, K.; Gao, J.; York, D. M. *J Chem Theory Comput* 2005, 1, 2.
5. Stefanovich, E. V.; Truong, T. N. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M., Eds.; ACS Symposium Series 712; Oxford University Press: New York, 1998; chap 6, 92.
6. Dinner, A. R.; Lopez, X.; Karplus, M. *Theor Chem Acc* 2003, 109, 118.
7. Schwarzl, S. M.; Huang, D.; Smith, J. C.; Fischer, S. In *Silico Biol* 2003, 3, 187.
8. Gregersen, B. A.; York, D. M. *J Phys Chem B* 2005, 109, 536.
9. Vollmer, J. M.; Stefanovich, E. V.; Truong, T. N. *J Phys Chem B* 1999, 103, 9415.
10. Simonson, T.; Archontis, G.; Karplus, M. *J Phys Chem B* 1997, 1001, 8349.
11. York, D. M.; Karplus, M. *J Phys Chem A* 1999, 103, 1106.
12. Stroud, A. H. *Approximate Calculation of Multiple Integrals*; Prentice Hall: Englewood Cliffs, NJ, 1971.
13. Lebedev, V. I. *Sibirsk Mater Ž* 1977, 18, 132.
14. Lebedev, V. I. *Russ Acad Sci Dokl Math* 1995, 50, 283.
15. Delley, B. *J Comput Chem* 1996, 17, 1152.
16. Lebedev, V. I.; Laikov, D. N. *Russ Acad Sci Dokl Math* 1999, 59, 477.
17. Gregersen, B. A.; York, D. M. *J Chem Phys* 2005, 122, 194110.
18. Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, W. P. *Numerical Recipes in Fortran*; Cambridge University Press: Cambridge, 1992, 2nd ed.
19. Berkowitz, M.; McCammon, J. A. *Chem Phys Lett* 1982, 90, 215.
20. Brooks, C. L., III; Brunger, A.; Karplus, M. *Biopolymers* 1985, 24, 843.
21. Pérez-Jordá, J. M.; Yang, W. *Chem Phys Lett* 1995, 247, 484.
22. Pérez-Jordá, J.; Yang, W. *J Chem Phys* 1996, 104, 8003.
23. Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science* 1996, 274, 2065.
24. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* 1983, 79, 926.
25. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J Comput Chem* 1983, 4, 187.
26. MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *Encyclopedia of Computational Chemistry*; v. R. Schleyer, P.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F., III; Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998, 271, vol. 1.
27. Foloppe, N.; MacKerell, A. D., Jr. *J Comput Chem* 2000, 21, 86.
28. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J Comput Phys* 1977, 23, 327.
29. Allen, M.; Tildesley, D. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
30. Darden, T.; York, D.; Pedersen, L. *J Chem Phys* 1993, 98, 10089.