ARMY RESEARCH LABORATORY

A Hybrid Explicit/Implicit Solvent Technique for Biomolecular Simulations

by Michael S. Lee and Mark A. Olson

ARL-TR-3268 August 2004

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5067

ARL-TR-3268 **August 2004**

A Hybrid Explicit/Implicit Solvent Technique for Biomolecular Simulations

Michael S. Lee
Computational and Information Sciences Directorate, ARL

Mark A. Olson
U.S. Army Medical Research Institute of Infectious Diseases

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
August 2004	Final	October 2002–August 2003

4. TITLE AND SUBTITLE
A Hybrid Explicit/Implicit Solvent Technique for Biomolecular Simulations

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michael S. Lee* and Mark A. Olson[†]

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: AMSRD-ARL-CI-HC
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-3268

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
[†]U.S. Army Medical Research Institute of Infectious Diseases
ATTN: MCMR-UIT
Ft. Detrick, MD 21702-5011

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
*Michael S. Lee is employed by the U.S. Army Research Laboratory but located at the U.S. Army Medical Research Institute of Infectious Diseases.

14. ABSTRACT
In this work, we design and use a new biomolecular simulation technique that combines explicit solvent molecules and implicit solvent theory. We employ a multigrid approach to speed up the computationally expensive long-range electrostatic energy terms ubiquitous to the classical simulation of biomolecules. It is shown that the multigrid technique affords an order-of-magnitude speedup over conventional cutoff approaches. Our algorithm makes possible a new class of biomolecular simulations that was previously untenable (i.e., the calculation of solvation energies of entire proteins using explicit water molecules). We present comparisons of protein solvation energies obtained from our method and a popular implicit solvent model.

15. SUBJECT TERMS
biomolecular simulation, multigrid, solvation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Raju Namburu
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL	24	19b. TELEPHONE NUMBER (*Include area code*) 410-278-0274

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

Biomolecular simulations can play a pivotal role in the development of new vaccines and therapeutics for the war fighter. For instance, in vaccine development for toxins, the protein structure of the toxin needs to be modified to reduce its toxicity while retaining the shape characteristics and solubility of the native fold. Since there are an overwhelming number of ways to mutate the sequence of a protein, a computational biologist can offer key insights into which residues are most advantageous to modify, thereby greatly simplifying this combinatorial task (*1*).

An essential tool for the computational biologist is a biomolecular simulation that includes the explicit solvent. However, this method is often computationally intensive. In recent years, two different solutions have emerged to deal with this issue. One solution is to remove the explicit solvent entirely and simulate the biomolecule with an implicit solvent approach such as a continuum dielectric treatment of bulk water (*2*, *3*). This approach is very fast; however, the extent to which accuracy is sacrificed in completely ignoring local water-solute interactions such as hydrogen bonding has yet to be determined. The alternative solution is to employ periodic boundary conditions in the explicit solvent simulation (*4*). This greatly reduces the computational effort of evaluating the long-range electrostatic interactions between atoms. The main drawback to this approach, however, is that the solute sees infinite periodic images of itself, and hence certain electrostatically-derived thermodynamic properties are difficult or nearly impossible to evaluate (*5*).

Given that the implicit solvent approach is fast and the explicit solvent approach has the correct local interactions, it begs the question whether it is possible to combine the salient features of each in a hybrid method. This idea, known most generally as the finite cluster method, is by no means new, having originated in the literature over a decade ago (*6*). Nonetheless, the widespread usage of such a hybrid method has been held back for a variety of reasons. First, finite cluster methods are often limited to a spherical simulation volume because the reaction field Poisson equation is only analytically known for a spherical model (*7*). This means that a large number of water molecules are necessary to hydrate a system when the solute is highly nonspherical. Second, these methods offer no benefits regarding effort reduction in evaluating long-range electrostatic interactions. A seemingly clever way around both of these issues is the generalized reaction field (GRF) approach where the simulation system is arbitrary but the long-range interactions are truncated, beyond which a dielectric treatment is assumed (*8*). This approach can be fast; however, it is misleading because the assumption of high dielectric outside the cutoff may be wrong if there are still solute atoms beyond the cutoff.

In section 2 of this report, we develop a new hybrid solvent methodology that successfully deals with the two aforementioned issues concerning finite-cluster methods: the spherical restriction and computational inefficiency. Our hybrid approach uses the Generalized Born (GB) theory (*3*), which is an accurate approximation to the more intensive Poisson equation that can readily handle nonspherical simulation volumes. Furthermore, our approach uses the recent advances in multigrid technology (*9*, *10*) to greatly reduce the computational effort in evaluating the long-range electrostatic interactions between atomic charges. In section 3, we evaluate the computational gain from using a multigrid approach as compared to conventional cutoff schemes. We then show how the combined technologies in our approach make the calculation of protein solvation energies possible in an explicit solvent environment, which to our knowledge has only rarely been attempted before, due in part to the technical limitations of prior explicit solvent approaches. Given solvation energies of protein conformations obtained by the hybrid method, the accuracies of several Poisson implicit solvent models are evaluated.

2. Methods

Classical biomolecular simulations involve the motion integration of an empirical molecular mechanics (MM) potential in the form

$$E_{MM} = E_{int} + E_{elec} + E_{vdW} + E_{solv}, \tag{1}$$

where E_{int} is a parameterized energy function of the internal degrees of freedom (i.e., bond/angle stretching and torsional rotations). The term E_{elec} is the Coulomb electrostatic interaction energy between all pairs of partial atomic charges q_i and q_j:

$$E_{elec} = k\sum_{i<j}\frac{q_i q_j}{r_{ij}}, \tag{2}$$

where $k \sim -332$ (kcal/mol)*Å/ecu[2*] and r_{ij} is the distance between atoms i and j. The term E_{vdW} is the sum of an attractive inverse sixth-power distance function and a repulsive inverse twelfth-power distance function that simulates the dispersion and Pauli exclusion interactions, respectively, between atoms. Finally, E_{solv} is an optional term representing the system interactions to a continuum reaction field mimicking bulk solvent. For instance, in the GB model, E_{solv} has the following definition:

$$E_{solv} = -k\sum_{ij}\frac{q_i q_j}{r_{ij} + \frac{1}{2}\left(\alpha_i + \alpha_j\right)\exp\left[-2r_{ij}/\left(\alpha_i + \alpha_j\right)\right]}, \tag{3}$$

[*] ecu = electronic charge unit.

where α_i is the Born radius of atom *i* (*3*, *11*). The Born radius is approximately the spherically averaged distance between the atom and the bulk solvent continuum. It is exactly defined in terms of the Born self-polarization of an atom E_i^{pol}:

$$\alpha_i = \frac{-k\left(1-\varepsilon^{-1}\right)q^2 i}{E_i^{pol}}. \tag{4}$$

There is no exact definition for the self-polarization energy or Born radius of a point charge placed in an arbitrary dielectric volume. However, we devised an accurate approximation given an arbitrary dielectric volume function $V(\vec{r})$ (*12*):

$$\begin{aligned}\alpha_i^{-1} = &\left(1-\frac{1}{\sqrt{2}}\right)\left(R_i^{-1} - \frac{1}{4\pi}\int\int_{R_i}^{\infty}\frac{V(\vec{r})}{\left|\vec{r}-\vec{x}_i\right|^4}drd\Omega\right) \\ &+\left(\frac{1}{4R_i^4} - \frac{1}{4\pi}\int\int_{R_i}^{\infty}\frac{V(\vec{r})}{\left|\vec{r}-\vec{x}_i\right|^7}drd\Omega\right)^{1/4},\end{aligned} \tag{5}$$

where R_i is the atomic radius of atom *i*.

2.1 Simulation Volume Definition

In our simulation protocol, the collection of solute and explicit solvent molecules is encapsulated in a predefined simulation volume. Our simulation volume is defined as interlocking spheres resembling the shape of the initial solute conformation. Building a sphere around each heavy atom turns out to be identical to the criteria that a water molecule should deviate no more than a prescribed width *w* from at least one solute atom. To ensure a smooth transition between spheres of different atoms, we use a volume function similar to that proposed by Im et al. (*13*). We define a volume function $V(\vec{r})$ such that $V = 1$ signifies the internal region and $V = 0$ signifies the external region. The volume function $V(\vec{r})$, resulting from superposition of spheres around each atom *i*, has the following mathematical form:

$$V(\vec{r}) = 1 - \prod_i \left[1 - V_i(\vec{r}-\vec{x}_i)\right], \tag{6}$$

where

$$V_i(r_i) = \begin{cases} 1, & s \le 0 \\ 1 - 6s^5 + 15s^4 - 10s^3, & 0 < s < 1 \\ 0, & s \ge 1 \end{cases} \tag{7}$$

and

$$s = \frac{r_i - w_i^0}{w_i^1 - w_i^0},\ r_i = \left|\vec{r}-\vec{x}_i\right|. \tag{8}$$

The width parameters w_i^0 and w_i^1 define the endpoints of the contour of sphere *i*. The polynomial in equation 7 has continuity up to second derivatives at its piecewise endpoints.

The wall potential and dielectric volume are both defined in terms of this simulation volume. The wall potential is a repulsive boundary that confines the solute and the explicit water molecules to the simulation volume and is defined as a function of the simulation volume:

$$E_{wall}(\vec{r}) = \frac{1}{E_{wall}^{\max -1} + V(\vec{r})} - \frac{1}{E_{wall}^{\max -1} + 1}, \tag{9}$$

such that $V = 1$ maps to $E_{wall}(\vec{r}) = 0$ and $V = 0$ maps to $E_{wall}(\vec{r}) \approx E_{wall}^{\max} - 1$, where $E_{wall}^{\max}$ = 30 kcal/mol in this work. Because we use a finite value for $E_{wall}^{\max}$, the wall potential is actually penetrable. This turns out to be useful because a system that is started with too many water molecules will actually spill the excess water molecules beyond the simulation volume, thus relieving excess pressure on the system. For this study, the width parameters for the wall potential were derived empirically: $w_i^0 = w - 0.6$ Å and $w_i^1 = w + 0.4$ Å. Strictly speaking, a wall potential, analogous to the vdW interaction energy, should also have an attractive component. However, we did not incorporate an attractive wall potential because some of our initial tests indicated that such a potential would have only a relatively small impact.

The dielectric volume is the same form as the simulation volume with empirically derived width parameters of $w_i^0 = w + 1.8$ Å and $w_i^1 = w + 2.2$ Å. In this work, we assumed a fixed simulation volume for the entire molecular dynamics run. This simplification implies a fixed dielectric volume and allows us to precalculate Born radii on a grid. The Born radii grid has a cubic cell size of 2 Å extending over a rectangular region that encompasses the entire simulation volume. With this grid, the Born radius of each atom is obtained via cubic interpolation from nearest neighbor grid points using the same interpolation basis functions that were employed in the multigrid procedure (see section 2.2). The actual numerical integration procedure to obtain Born radii from equation 5 has been described elsewhere (*12*, *14*).

2.2 Multigrid Algorithm

The internal energy term E_{int} in equation 1 requires very little computational effort and has $O(N)$ scaling with system size, where N is the number of atoms. On the other hand, the electrostatic, dispersion, and solvation terms are costly because they require $O(N^2)$ computational effort. Nonetheless, to reduce the cost of these terms, a cutoff is often used, which smoothly zeroes out these terms as the interacting atoms reach a certain threshold distance apart. This works well for the dispersion term, since the function has such a sharp decay (i.e., r^{-6}). However, the electrostatic and solvation terms require a much larger cutoff because their functions decay very slowly (i.e., r^{-1}).

There are two standard alternative means to reducing the computational cost of the electrostatic term. First, the simulation volume can be assumed to be periodic; therefore, an Ewald summation in a dual real/Fourier space can be used to compute all of the long-range interactions. The main drawback of this approach is that the solute sees images of itself, which can cause major artifacts when the solute or solute fragment has a net charge. Second, the fast multipole approximation can be used to accurately estimate long-range electrostatic interactions. One major drawback of this approach, for our purposes, is that it is defined only for the simple Coulomb functional form and cannot be used with GB theory.

Our solution for reducing the computational cost of the long-range interactions is the multigrid method. Multigrid approaches (*15*) break a problem into local and nonlocal components, and as applied to finite difference equations, involve numerical solutions at different grid resolutions. The multigrid method for three-dimensional (3-D) Coulomb electrostatics has been developed by Skeel et al. (*10*). We adapted their method to evaluate the GB electrostatic terms. The basic principle of a multigrid electrostatics procedure is to cast pairwise charge interactions onto grids and treat various interaction ranges with different degrees of approximation (indexed as L). The multigrid algorithm is sketched as follows:

- Perform explicit local interactions of atomic charges, when $r < a$, where a is a user-defined local cutoff ($L = 0$).
- Anterpolate atomic charges onto the finest resolution charge grid $(L = 0 \rightarrow L = 1)$.
- Create a hierarchy of grids by anterpolating each finer resolution charge grid to its next coarser resolution charge grid $(L = n \rightarrow L = n + 1)$.
- For each individual charge grid L, build a potential grid such that each grid point sees charges less than a distance $2^L a$ away.
- Interpolate potential grids from coarse to fine resolution $(L = n + 1 \rightarrow L = n)$.
- Contract atomic charges with finest resolution potential grid to obtain energy and forces $(L = 1 \rightarrow L = 0)$.

To separate the ranges precisely, the interaction kernel for each level $K^L(r)$ is split into two components: a soft function and a local function.

$$K^L(r) = K^L_{soft}(r) + K^L_{local}(r). \tag{10}$$

The soft function for the $L = 0$ grid has to be slowly varying because it is the basis of interaction for the $L = 1$ grid. For a Coulombic potential $K^0(r) = 1/r$, the soft function is chosen such that $K^0_{local}(r)$ goes to zero beyond some cutoff a. In the original article of Skeel et al. (*10*), several

soft functions were presented and tested. Nonetheless, we wanted to reduce the variance of the soft function completely in the domain $r < a/2$ so that certain self-interaction and exclusion interaction artifacts could be reduced or eliminated. Therefore, we created a new piecewise soft potential that is continuous up to second derivatives at $r = a$:

$$K_{soft}^{0}(r) = \begin{bmatrix} \frac{18}{3a^5}\left(r-\frac{a}{2}\right)^4 - \frac{16}{3a^4}\left(r-\frac{a}{2}\right)^3 + \frac{31}{24a}, & \frac{a}{2} < r < a \\ \frac{31}{24a}, & r < \frac{a}{2} \end{bmatrix}. \tag{11}$$

The local function for $L = 0$, by definition, is the difference of the original kernel and the soft function:

$$K_{local}^{0}(r) = \frac{1}{r} - K_{soft}^{0}(r). \tag{12}$$

The interaction kernels for $L = 0$ are depicted in figure 1. Analogously, for each grid level L, except the coarsest $L = L_{max}$, the interaction kernel is composed of a soft function, where a in equation 11 is replaced by $2^{L-1}a$, minus a softer function, where a is substituted with $2^{L}a$. For the coarsest grid L_{max}, the interaction kernel is simply the softest function corresponding to a in equation 11 being replaced by $2^{L_{\max}-1}a$.

Multigrid steps 2, 3, 5, and 6, involving interpolation or anterpolation, use 3-D basis functions of the form (*10*)

$$\phi(x,y,z) = \Phi\left(\frac{x-x_{ij}}{h}\right)\Phi\left(\frac{y-y_{ij}}{h}\right)\Phi\left(\frac{z-z_{ij}}{h}\right), \tag{13}$$

where

$$\Phi(p) = \begin{cases} (1-|p|)\left(1+|p|-\frac{3}{2}p^2\right), & |p| \le 1 \\ -\frac{1}{2}(|p|-1)(2-|p|)^2, & 1 \le |p| \le 2 \\ 0, & |p| \ge 2 \end{cases} \tag{14}$$

and (x, y, z) are the coordinates of the charge, (x_{ij}, y_{ij}, z_{ij}) are the coordinates of a grid point, and h is the side length of a grid cube.

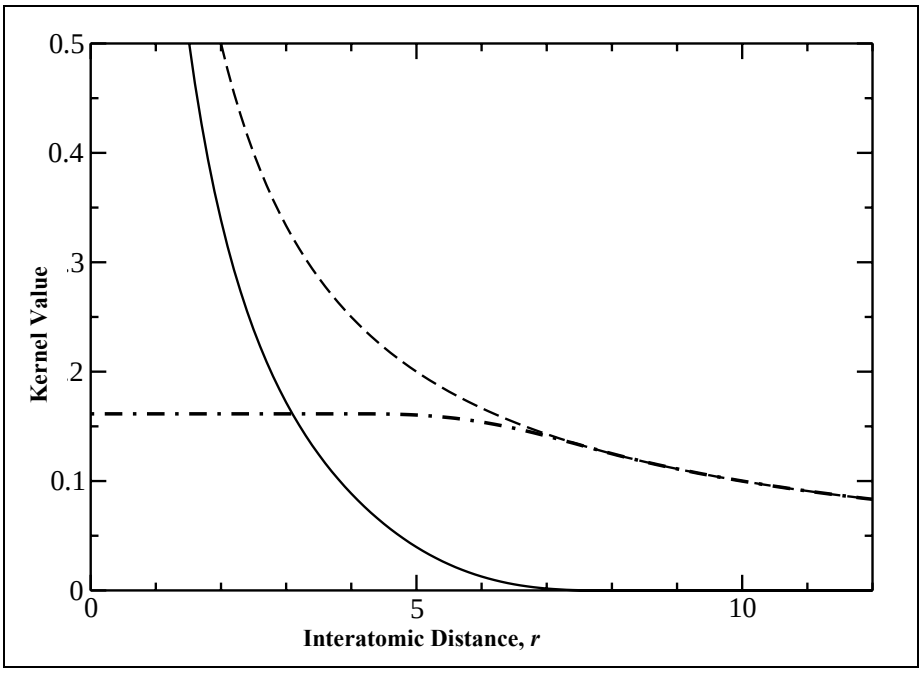

Figure 1. The interaction kernels used in a multigrid implementation of the Coulomb electrostatic term. The cutoff value *a*, in this example, is 8 Å. The dashed line corresponds to the standard Coulomb potential. The solid line indicates the local kernel $K_{local}(r)$, and the dot-dashed line indicates the soft kernel $K_{soft}(r)$.

To extend the multigrid procedure to handle the GB theory, the GB kernel $J = \left\{ r_{ij} + \frac{1}{2}(\alpha_i + \alpha_j)\exp\left[-2r_{ij}/(\alpha_i + \alpha_j)\right]\right\}^{-1}$ also needs to be split into local and soft terms J_{local} and J_{soft}:

$$J^L(r_{ij}, \alpha_i, \alpha_j) = J^L_{local}(r_{ij}, \alpha_i, \alpha_j) + J^L_{soft}(r_{ij}, \alpha_i, \alpha_j). \tag{15}$$

Experimentation determined that a reasonable splitting involves simply setting r_{ij} equal to the corresponding inverse of the respective Coulomb kernels K^L_{local} and K^L_{soft}:

$$J^L_{local}(r_{ij}, \alpha_i, \alpha_j) = J\left(\left[K^L_{local}\right]^{-1}, \alpha_i, \alpha_j\right) \text{ and } J^L_{soft}(r_{ij}, \alpha_i, \alpha_j) = J\left(\left[K^L_{soft}\right]^{-1}, \alpha_i, \alpha_j\right). \tag{16}$$

Furthermore, Born radii need to be evaluated at every multigrid point. Born radii at the lowest level multigrid ($L = 0$) are obtained via cubic interpolation of the same Born radii grid used to determine atomic Born radii. Coarser grids ($L>0$) are obtained by cubic anterpolation of their next finest Born radii grid (L–1).

2.3 Simulation Protocol

All of the molecular dynamics simulations in this work were performed with the CHARMM program (*16*) using the PARAM22 empirical force field potential (*17*). The equations of motion were integrated with a Langevin dynamics algorithm at constant temperature (298 K) using a friction constant of 5 ps^{-1} and an integration time step of 2 fs. Covalent bonds between heavy atoms and hydrogens were constrained using the SHAKE algorithm (*18*).

The general procedure for adding explicit water molecules to a solute involves two steps: determining the number of water molecules needed to fill the simulation volume and carving out these water molecules from a large block of water. The number of desired water molecules is calculated by first determining the water simulation volume (WSV), which is equal to the expanded solute volume, where each atomic radius is augmented by the specified width *w* minus the standard solute volume. The desired number of water molecules then equals the WSV multiplied by the standard bulk water density, 0.0334 $Å^{-3}$. Then, a large cubic box of water molecules is overlaid onto the solute. The water molecules that are less than 2.1 Å from a heavy atom are deleted. Next, an iterative procedure is performed in which water molecules beyond a certain cutoff distance from all heavy atoms of the solute are deleted. The procedure repeats deletion with decreasing cutoffs until the number of water molecules remaining is less than or equal to the desired number of water molecules. A large spherical boundary potential is placed around the system to prevent the drift of water molecules that escape from the smaller finite boundary of equation 9.

2.4 Calculation of the Solvation Energy Using Free Energy Methods

A relevant thermodynamic quantity for this study is the electrostatic free energy of solvation ΔG_{solv}^{elec} of an arbitrary solute in a fixed conformation. To obtain this quantity, we calculate the free energy necessary to charge the solute from zero (state $\lambda = 0$) to fully charged (state $\lambda = 1$) using the thermodynamic integration formula (*19*)

$$\Delta G_{\lambda=0\to\lambda=1}^{elec} = \int_0^1 \left\langle \frac{\partial E}{\partial \lambda} \right\rangle d\lambda \,. \tag{17}$$

In our approach, *n* simulation windows are run corresponding to values $\lambda = 1/2n, 3/2n, ...,$ $(2n-1)/2n$. In each window, the force term $\partial E/\partial\lambda$ is statistically averaged. Then, a straight line is fit through the force terms $y = a_0 + a_1\lambda$ to arrive at an electrostatic charging free energy $\Delta G_{\lambda=0\to\lambda=1}^{elec} = a_0 + a_1/2$. Finally, the electrostatic free energy of solvation ΔG_{solv}^{elec} is obtained from the formula $\Delta G_{\lambda=0\to\lambda=1}^{elec} = \Delta G_{solv}^{elec} + E_{elec}^{solute}$, where E_{elec}^{solute} is the total Coulomb electrostatic energy of the solute interacting with itself, which is independent of the solvent environment.

2.5 Implicit Solvent Protocol

In this work, we compare the solvation energies obtained from the hybrid method to a benchmark implicit solvation scheme, the Poisson equation. Calculation of the solvation energy of a molecule using the Poisson equation can be performed in a variety of software packages. We use the Poisson-Boltzmann equation (PBEQ) module in CHARMM (*16*) for this study. The protocol for Poisson solvation energy involves iterative solution of the Poisson equation on a cubic grid to obtain a potential map $\psi(\vec{r})$.

$$\nabla \bullet [\varepsilon(\vec{r})\nabla\psi(\vec{r})] = -4\pi\rho(\vec{r}), \tag{18}$$

where a dielectric boundary $\varepsilon(\vec{r})$ is defined such that $\varepsilon = 1$ signifies the solute and $\varepsilon = 80$ designates the bulk solvent. A common dielectric boundary is the one of Lee and Richards (*20*), also known as the molecular surface (MS). The MS is defined by first superimposing spheres for each atom with radius $R_i + r_{probe}$. Then, a water probe of radius r_{probe} is rolled across this boundary to carve out regions where the water probe can reenter. The resultant molecular surface is the vdW surface of the solute plus regions where a water probe cannot access. The partial atomic charges, which make up the charge density $\rho(\vec{r})$ are placed on a grid using trilinear interpolation. Details on the numerical solution of the differential equation (equation 18) are available elsewhere (*2*).

3. Results and Discussion

In tables 1 and 2, we evaluate the accuracy of the multigrid approximation compared to standard cutoff approaches for two model systems: a sodium ion embedded in a 17-Å sphere of water molecules and a protein surrounded by a 10-Å layer of water molecules. Three accuracy measures were used to evaluate the multigrid and cutoff approximations to the combined electrostatic and GB terms.

Energy error:

$$\%EE = 100\% * \frac{\left|E - E_{exact}\right|}{\left|E_{exact}\right|}. \tag{19}$$

Average force error:

$$\%AFE = 100\% * \frac{\sum_i m_i^{-1/2}\left|F_i^{exact} - F_i\right|}{\sum_i m_i^{-1/2}\left|F_i^{exact}\right|}. \tag{20}$$

Table 1. Accuracy of multigrid and standard cutoff approaches for the model system of a sodium ion embedded in a 17-Å sphere.

Method	Cutoff Start (Å)	Cutoff Stop (Å)	Energy Error (%)	Average Force Error (%)	Maximum Force Error (%)
MG	8	8	0.22	3.5	13
MG	10	10	0.068	1.7	6.9
MG	12	12	0.015	0.97	3.8
STD	8	10	0.45	54	250
STD	10	12	0.031	39	170
STD	14	16	3.6	22	98
STD	18	20	0.85	10	42
STD	22	24	0.65	3.9	19
STD	26	28	0.013	1.1	5.8

Notes: MG = multigrid, and STD = standard cutoff approaches.
Definitions of error measures are described in the text. Energies consist of the sum of the electrostatic and GB energies.

Table 2. Accuracy of multigrid and standard cutoff approaches for two proteins embedded in a 10-Å layer: turkey ovomucoid receptor (PDB identifier: 1OMT) and trypsin (PDB identifier: 1TNJ).

Method	Cutoff Start (Å)	Cutoff Stop (Å)	Energy Error (%)	Average Force Error (%)	Maximum Force Error (%)	Time (s)
Turkey Ovomucoid Receptor (PDB Identifier: 1OMT)						
MG	8	8	0.49	4.4	19	35.7
STD	26	28	0.22	3.3	19	223.7
Trypsin (PDB identifier: 1TNJ)						
MG	8	8	0.23	4.2	21.6	88.05
STD	26	28	0.042	6.2	44.7	992.7

Notes: MG = multigrid, and STD = standard cutoff approaches.
Definitions of error measures are described in the text. Energies consist of the sum of the electrostatic and GB energies. Last column indicates central processing unit times necessary to run 100 steps of geometry optimization.

Maximum force error:

$$\%MFE = 100\% * \frac{\max\left[m_i^{-1/2}\left|F_i^{exact} - F_i\right|\right]}{N^{-1}\sum_i m_i^{-1/2}\left|F_i^{exact}\right|}, \quad (21)$$

where m_i is the mass of atom i, F_i is the total electrostatic + GB force acting on atom i, and N is the total number of atoms. As can been seen in tables 1 and 2, a cutoff >20 Å is required to match the results of the multigrid approach that uses a local cutoff of 8 Å. Furthermore, when

compared to the central processing unit times required to perform 100 steps of geometry optimization, the multigrid method is approximately an order of magnitude faster than the standard cutoff approach.

In table 3, we evaluate the accuracy of the hybrid method at different simulation widths for calculating the electrostatic charging free energy of four small molecules. It appears that at about a 10- to 12-Å width, the free energies converge more or less to a bulk limit. This width range corresponds to approximately three shells of water molecules surrounding a solute. This is reasonably consistent with experimental studies, which indicate that the first two layers of water molecules surrounding a protein surface have properties deviating from bulk solvent (*21*).

Table 3. Charging free energies of small solutes in various layers of solvent.

Width (Å)	Na^+ (kcal/mol)	Cl (kcal/mol)	H_2O (kcal/mol)	Ethanol (kcal/mol)
6	–104.4	–82.6	–8.74	–12.54
8	–104.5	–82.6	–8.44	–12.27
10	–105.3	–83.0	–8.68	–12.34
12	–105.6	–83.9	–8.80	–12.32
24	–105.0	–84.0	–8.41	–12.12

Note: thermodynamic integration was performed over 10 equally spaced 100-ps windows.

With a sense of what simulation protocol is reasonable to use, we evaluated the solvation free energies of several model compounds. In table 4, solvation energies of the 20 capped amino acids [$CH_3C(=O)$-X-NCH_3] are compared between hybrid and fully implicit solvent schemes. One can see that for the uncharged species, the implicit solvent scheme has an excellent correspondence to the explicit approach. However, for the charged groups (Asp^-, Glu^-, Lys^+, and Arg^+), there are significant deviations. The standard protocol for alleviating this situation is to modify some of the atomic radii of the side chain atoms to bring the implicit approach in better agreement. We found, for instance, that the following scheme, which we term S1, improves results for the charged groups: reducing the Asp and Glu carboxylate oxygen radii by 0.3 Å, augmenting the amino nitrogen radii of Lys by 0.26 Å, and increasing the side chain nitrogen radii of Arg by 0.2 Å.

While comparisons of solvation energies for small model compounds, such as capped amino acids, have been previously presented in the literature (*22*), little work has been done in comparing solvation energies of proteins (*23*). In tables 5 and 6, comparisons of charging free energies in solvent for several protein conformations are made using various metrics such as root mean squared deviation (RMSD) and correlation coefficient (R). We look at charging free energy, which includes the Coulomb energy, rather than just solvation energy, since the former is a more relevant measure of the conformational free energy. The thermodynamic free energies for the hybrid method (10-Å layer) entail five 300-ps windows. First, it can be seen that there is

Table 4. Solvation free energies using hybrid and Poisson methods.

Name	Hybrid (kcal/mol)	Poisson (kcal/mol)
Nonpolar Groups		
Gly	–12.0	–12.2
Ala	–13.9	–14.3
Val	–17.0	–16.6
Leu	–17.3	–16.6
Ile	–16.9	–15.2
Pro	–11.9	–11.0
Phe	–19.9	–19.4
Trp	–21.3	–22.5
Met	–16.9	–16.8
Polar Groups		
Ser	–23.7	–24.1
Thr	–24.3	–24.5
Cys	–19.1	–19.4
Tyr	–24.6	–25.1
Asn	–21.1	–21.6
Gln	–26.3	–26.7
His	–29.9	–29.9
Charged Groups		
Asp^-	–100.3	–88.5
Glu^-	–97.5	–86.4
Lys^+	–86.3	–93.8
Arg^+	–75.6	–84.7

Note: thermodynamic integration for the hybrid method (12-Å layer) was performed over five equally spaced 20-ps windows.

a noticeable improvement in accuracy progressing from a cell size of 0.5 Å to 0.25 Å. It should be noted, however, that the 0.25-Å calculations are approximately eight times slower than the 0.5-Å calculations. Second, it appears that a nonstandard water probe radius of 1.0 Å outperforms the conventional probe radius of 1.4 Å. This is also illustrated in figure 2. This result suggests that water molecules are able to access deeper into the crevices between atoms than was previously thought.

Some researchers have suggested that if many atomic radii are scaled appropriately to obtain accurate small-molecule solvation energies, one would obtain a more accurate solvent description for proteins. However, as one can see in tables 5 and 6 and figure 3, the elaborate prescription of Nina et al. (*22*), which involves the modification of over 20 atomic radii, appears

Table 5. A comparison of various Poisson solvation schemes vs. the hybrid electrostatic scheme for charging free energies of 120 conformations of the villin headpiece.

Method	Cell Size (Å)	Probe Radius (Å)	RMSD (kcal/mol)	Shifted RMSD[a] (kcal/mol)	Correlation Coefficient	Slope[b]
MS[c]	0.5	1.4	19.0	8.3	0.953	0.96
MS	0.5	1.0	8.4	7.6	0.961	0.87
MS	0.25	1.4	38.7	7.3	0.973	1.08
MS	0.25	1.0	19.2	5.3	0.981	0.99
MS w/S1[d]	0.25	1.4	31.5	6.5	0.978	1.08
MS w/S1	0.25	1.0	11.6	4.9	0.984	0.99
MS w/NBR[e]	0.25	1.4	74.5	16.7	0.957	1.44
Null[f]	NA	NA	1318.0	27.0	0	NA

Notes: RMSD = root mean square deviation, NBR = Nina, Beglov, and Roux, and NA = not applicable.
[a] The RMSD calculated after the two sets being compared are each subtracted by their mean.
[b] The slope obtained from a linear least-squares fit.
[c] The Poisson method with a molecular surface dielectric boundary.
[d] A small set of scaled atomic radii (see text).
[e] The radii set of Nina et al. (*22*).
[f] A zero charging free energy.

Table 6. Comparison of various Poisson solvation schemes vs. the hybrid electrostatic scheme for charging free energies of 39 conformations of the B1 domain of protein L.

Method	Cell Size (Å)	Probe Radius (Å)	RMSD (kcal/mol)	Shifted RMSD[a] (kcal/mol)	Correlation Coefficient	Slope[b]
MS[c]	0.5	1.4	104.3	16.4	0.688	0.916
MS	0.5	1.0	31.7	10.7	0.807	0.832
MS	0.25	1.4	155.0	13.7	0.731	0.852
MS	0.25	1.0	77.8	8.3	0.872	0.763
MS w/S1[d]	0.25	1.4	100.5	12.7	0.752	0.837
MS w/S1	0.25	1.0	30.7	9.2	0.839	0.712
MS w/NBR[e]	0.25	1.4	207.0	18.2	0.646	0.908
Null[f]	NA	NA	1802.8	16.9	0	NA

Notes: RMSD = root mean square deviation, NBR = Nina, Beglov, and Roux, and NA = not applicable.
[a] The RMSD calculated after the two sets being compared are each subtracted by their mean.
[b] The slope obtained from a linear least-squares fit.
[c] The Poisson method with a molecular surface dielectric boundary.
[d] A small set of scaled atomic radii (see text).
[e] The radii set of Nina et al. (*22*).
[f] A zero charging free energy.

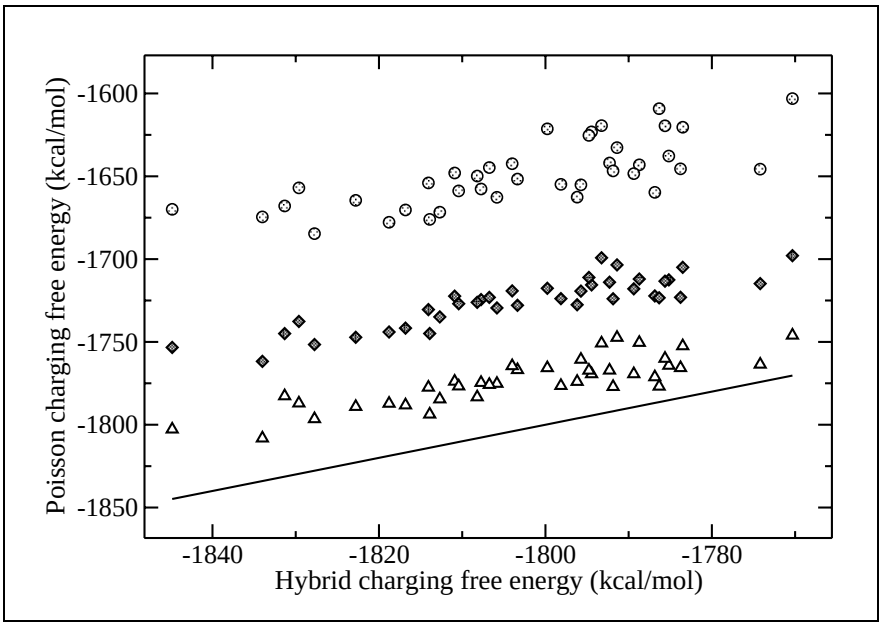

Figure 2. A comparison of molecular surface Poisson with probe radii of 1.0 and 1.4 Å vs. the hybrid method for charging free energies of 39 conformations of the B1 domain of protein L. The circles indicate the 1.4-Å results, the diamonds signify the 1.0-Å results, and the triangles indicate the 1.0-Å results with a modified radii set S1 (see text). The

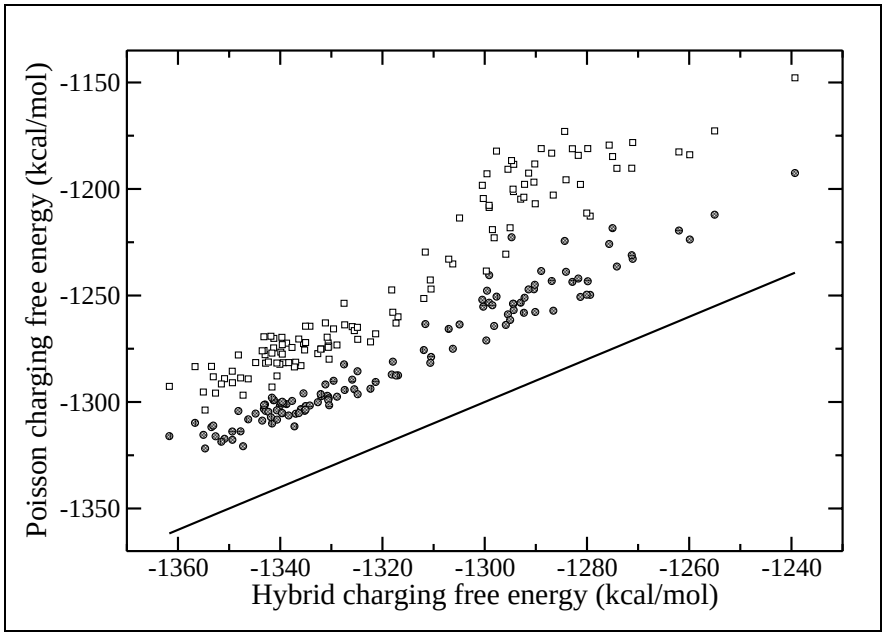

Figure 3. A comparison of molecular surface Poisson with standard PARAM22 and modified atomic radii vs. the hybrid method for charging free energies of 120 conformations of the villin headpiece. The filled circles indicate the standard PARAM22 radii results, and the open squares signify the modified atomic radii of Nina et al. (*22*). The straight line indicates $y = x$.

to have a deleterious effect on electrostatic charging free energies for both sets of protein conformations. For example, in the villin set (table 5), the shifted RMSD using the radii of Nina et al. (*22*) is twice as poor as the protocol that involves no radii scaling. In fact, figure 3 illustrates that the native states of villin, which are lower in energy, become artificially more favored compared to the nonnative states. This may be appropriate for native protein structure prediction. Nonetheless, the NBR scheme would be completely incorrect for modeling the transition between folded and unfolded states. Our radii modification scheme S1, which only involves four atom types, does not generally affect the shifted RMSDs or correlation coefficients. However, the S1 protocol does improve absolute solvation energies significantly. This could prove useful in simulations of ligand-binding interactions.

4. Conclusion

In this report, we have presented a new algorithm for performing biomolecular simulations in irregularly shaped volumes. Proper electrostatic treatment is achieved by encapsulating the explicit simulation volume in an implicit solvent described by the GB theory. Significant computational enhancement of this approach is achieved through the use of a pairwise multigrid technique, which has been extended to incorporate the GB model.

It has been shown that force field-consistent bulk limit properties such as charging free energies can be achieved with an explicit water layer of 10–12 Å. This permits a rather thrift number of water molecules compared to alternative approaches such as periodic boundary conditions and spherical clusters. The computational advantage of the multigrid technique and the reduction in number of water molecules allow one to compute electrostatic solvation energies of small proteins in a reasonable amount of time. We have illustrated how such explicit solvation energies provide a benchmark for implicit solvation schemes and indicate possible strategies for improvement of implicit solvent approaches.

The methodologies that have been designed in this report can also be used to study dynamical solutes. For instance, we have applied this method to accurately model protein-ligand binding and calculate binding free energies. These simulations are often very time consuming because of the thousands of explicit water molecules that must be incorporated in a water box. Interestingly, in our studies of ligands binding to trypsin, implicit solvent methods such as GB often produce unstable trajectories. Therefore, the hybrid method has provided a practical alternative that now makes these studies possible.

Given this new technique, we are now better equipped to look at the effects of protein toxin mutation for potential vaccines and/or emerging threats. Furthermore, this approach will be applied towards the development of improved computational chemistry methodologies for drug design and validation.

5. References

1. Ulrich, R.; Olson, M.; Bavari, S. Development of Engineered Vaccines Effective Against Structurally Related Bacterial Superantigens. *Vaccine* **1998**, *16* (19), 1857.

2. Gilson, M. K.; Honig, B. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis. *Proteins: Struct., Funct., Genet*. **1988**, *4*, 7.

3. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society* **1990**, *112* (16), 6127.

4. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys*. **1995**, *103*, 8577.

5. Bogusz, S.; Cheatham, T. E. I.; Brooks, B. R. Removal of Pressure and Free Energy Artifacts in Charged Periodic Systems Via Net Charge Corrections to the Ewald Potential. *J. Chem. Phys.* **1998**, *108* (17), 7070.

6. King, G.; Warshel, A. A Surface-Constrained All-Atom Solvent Model for Effective Simulations of Polar Solutions. *J. Chem. Phys.* **1989**, *91* (6), 3647.

7. Beglov, D.; Roux, B. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. *J. Chem. Phys.* **1994**, *100* (12), 9050.

8. Neumann, M.; Steinhauser, O. The Influence of Boundary Conditions Used in Machine Simulations on the Structure of Polar Systems. *Mol. Phys*. **1980**, *39*, 437.

9. Sandak, B. Multiscale Fast Summation of Long-Range Charge and Dipolar Interactions. *J. Chem. Phys.* **2001**, *22* (7), 717.

10. Skeel, R. D.; Tezcan, I.; Hardy, D. J. Multiple Grid Methods for Classical Molecular Dynamics. *J. Chem. Phys.* **2002**, *23* (6), 673.

11. Salsbury, F. R., Jr.; Lee, M. S.; Feig, M.; Brooks, C. L., III. To be submitted for publication.

12. Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. New Analytic Approximation of the Standard Molecular Volume Model: Application to Generalized Born Calculations. *J. Comput. Chem.* **2003**, *24* (11), 1348.

13. Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: Computation of Electrostatic Forces From Numerical Solutions to the Poisson-Boltzmann Equation. *Comput. Phys. Commun*. **1998**, *111*, 59.

14. Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. Novel Generalized Born Methods. *J. Chem. Phys.* **2002**, *116* (24), 10606.

15. Brandt, A. *Guide to Multigrid Development.* Hackbusch, W., Trottenberg, V., Eds.; Springer-Verlag: New York, 1982.

16. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminatham, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem*. **1983**, *4*, 187.

17. 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-McCarthey, 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.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586.

18. Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System With Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327.

19. Kirkwood, J. G. *Theory of Liquids*; Alder, B. J., Ed.; Gordon and Breach: New York, 1968.

20. Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibilities. *J. Mol. Bio.* **1971**, *55*, 151.

21. Pal, S. K.; Peon, J.; Zewail, A. H. Biological Water at the Protein Surface: Dynamical Solvation Probed Directly With Femtosecond Resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (4), 1763.

22. Nina, M.; Beglov, D.; Roux, B. Atomic Radii for Continuum Electrostatics Based on Molecular Dynamics Free Energy Simulations. *J. Phys. Chem. B* **1997**, *101*, 5239.

23. Zhang, L. Y.; Gallichio, E.; Friesner, R. A.; Levy, R. M. Solvent Models for Protein-Ligand Binding: Comparison of Implicit Solvent Poisson and Surface-Generalized Born Models With Explicit Solvent Simulations. *J. Comput. Chem*. **2001**, *22* (6), 591.

NO. OF COPIES	ORGANIZATION
1 (PDF ONLY)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	COMMANDING GENERAL US ARMY MATERIEL CMD AMCRDA TF 5001 EISENHOWER AVE ALEXANDRIA VA 22333-0001
1	INST FOR ADVNCD TCHNLGY THE UNIV OF TEXAS AT AUSTIN 3925 W BRAKER LN STE 400 AUSTIN TX 78759-5316
1	US MILITARY ACADEMY MATH SCI CTR EXCELLENCE MADN MATH THAYER HALL WEST POINT NY 10996-1786
1	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CS IS R 2800 POWDER MILL RD ADELPHI MD 20783-1197
3	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL RD ADELPHI MD 20783-1197
3	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CS IS T 2800 POWDER MILL RD ADELPHI MD 20783-1197

NO. OF COPIES	ORGANIZATION
	ABERDEEN PROVING GROUND
1	DIR USARL AMSRD ARL CI OK TP (BLDG 4600)