

# An $n \log n$ Generalized Born Approximation

Ramu Anandakrishnan,<sup>†</sup> Mayank Daga,<sup>†</sup> and Alexey V. Onufriev<sup>\*,‡</sup>

<sup>†</sup>Department of Computer Science and <sup>‡</sup>Department of Computer Science and Physics, Virginia Tech, Blacksburg, Virginia 24061, United States

**ABSTRACT:** Molecular dynamics (MD) simulations based on the generalized Born (GB) model of implicit solvation offer a number of important advantages over the traditional explicit solvent based simulations. Yet, in MD simulations, the GB model has not been able to reach its full potential partly due to its computational cost, which scales as  $\sim n^2$ , where  $n$  is the number of solute atoms. We present here an  $\sim n \log n$  approximation for the generalized Born (GB) implicit solvent model. The approximation is based on the hierarchical charge partitioning (HCP) method (Anandakrishnan and Onufriev *J. Comput. Chem.* **2010**, *31*, 691–706) previously developed and tested for electrostatic computations in gas-phase and distant dependent dielectric models. The HCP uses the natural organization of biomolecular structures to partition the structures into multiple hierarchical levels of components. The charge distribution for each of these components is approximated by a much smaller number of charges. The approximate charges are then used for computing electrostatic interactions with distant components, while the full set of atomic charges are used for nearby components. To apply the HCP concept to the GB model, we define the equivalent of the effective Born radius for components. The component effective Born radius is then used in GB computations for points that are distant from the component. This HCP approximation for GB (HCP-GB) is implemented in the open source MD software, NAB in AmberTools, and tested on a set of representative biomolecular structures ranging in size from 632 atoms to  $\sim 3$  million atoms. For this set of test structures, the HCP-GB method is 1.1–390 times faster than the GB computation without additional approximations (the reference GB computation), depending on the size of the structure. Similar to the spherical cutoff method with GB (cutoff-GB), which also scales as  $\sim n \log n$ , the HCP-GB is relatively simple. However, for the structures considered here, we show that the HCP-GB method is more accurate than the cutoff-GB method as measured by relative RMS error in electrostatic force compared to the reference (no cutoff) GB computation. MD simulations of four biomolecular structures on 50 ns time scales show that the backbone RMS deviation for the HCP-GB method is in reasonable agreement with the reference GB simulation. A critical difference between the cutoff-GB and HCP-GB methods is that the cutoff-GB method completely ignores interactions due to atoms beyond the cutoff distance, whereas the HCP-GB method uses an approximation for interactions due to distant atoms. Our testing suggests that completely ignoring distant interactions, as the cutoff-GB does, can lead to qualitatively incorrect results. In general, we found that the HCP-GB method reproduces key characteristics of dynamics, such as residue fluctuation,  $\chi_1/\chi_2$  flips, and DNA flexibility, more accurately than the cutoff-GB method. As a practical demonstration, the HCP-GB simulation of a 348 000 atom chromatin fiber was used to refine the starting structure. Our findings suggest that the HCP-GB method is preferable to the cutoff-GB method for molecular dynamics based on pairwise implicit solvent GB models.

## 1. INTRODUCTION

Atomistic molecular dynamics (MD) simulations can be used to study biomolecules where experimental investigation is expensive or infeasible.<sup>1–6</sup> However, the duration and system size for such simulations are limited by the computational cost of long-range electrostatic interactions, which, without further approximations, scales as  $\sim N^2$  where  $N$  is the total number of atoms.<sup>7–10</sup> We do not consider coarse-grained approximations here, which are in general less accurate than atomic-level approximations.<sup>11</sup>

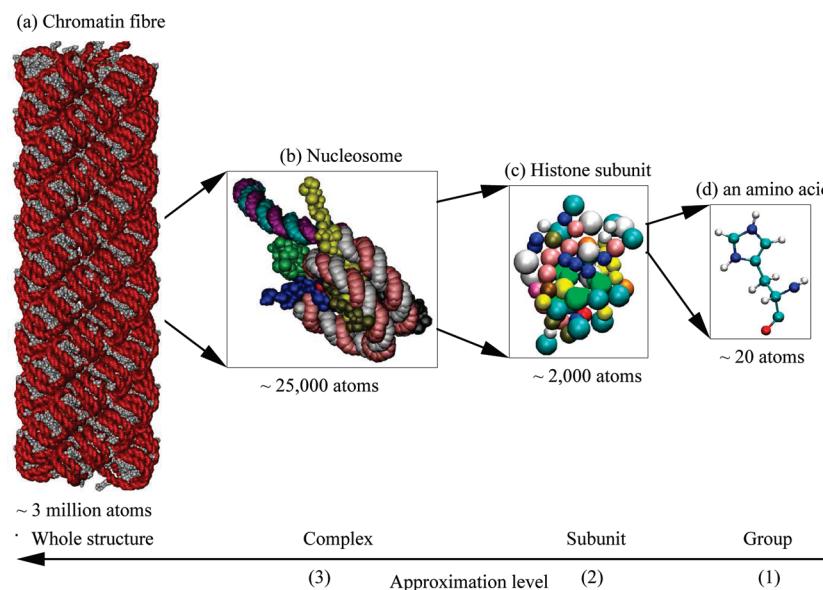
Historically, the first approximation widely used in the context of molecular dynamics (MD) to speedup the computation of long-range interactions was the spherical cutoff method. The simple spherical cutoff method treats atoms within a cutoff distance exactly while ignoring all other atoms.<sup>12,13</sup> The spherical cutoff method can however produce many artifacts such as spurious forces or artificial structures around the cutoff distance.<sup>14–16</sup> The particle mesh Ewald (PME) method was developed to address the shortcomings of the spherical cutoff method and has become the de facto “industry standard” for

explicit solvent MD. The PME imposes an artificial periodic boundary condition where a central cell containing the molecule of interest is assumed to be surrounded by an infinite array of images of the central cell. With this assumption, the long-range interaction, which decays slowly with distance, can be represented as the sum of two fast converging series—one in real space and the other in Fourier space.<sup>17–20</sup> Another explicit solvent method, the fast multipole method,<sup>21–23</sup> was tested for biomolecular simulations but has not been widely adopted, most likely due to its algorithmic complexity and instabilities caused by discontinuities inherent in the method.<sup>24</sup> In general, all three of these methods scale as  $\sim N \log N$ , where  $N$  is the total number of atoms in the system including the solvent atoms.

Realistic simulations require that the biomolecular structure be immersed in a solvent, typically water with ions. Implicit solvent models, such as the generalized Born (GB) approximation,

Received: July 13, 2010

Published: January 27, 2011

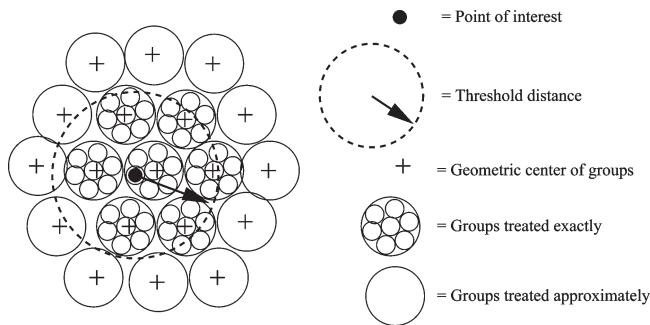


**Figure 1.** Example of the natural hierarchical partitioning of a chromatin fiber. (a) The fiber is made up of 100 nucleosome complexes. The individual nucleotide groups in the fiber are represented as red beads and amino acid groups as gray beads. (b) Each complex (level 3) is made up of 13 subunits with the segments of DNA linking nucleosome complexes being treated as separate subunits. A complex is shown here with each subunit represented in a different color. (c) Each subunit (level 2) is made up of 49–142 groups. The linker histone subunit is shown here with the groups colored by the type of amino acid. (d) Each group (level 1) is made up of 7–32 atoms (level 0). A histidine amino acid group is shown here with atoms represented as small spheres and covalent bonds between the atoms represented as links. The atoms are colored by the type of atom. The total fiber consists of approximately 3 million atoms. The fiber was constructed as described in Wong et al.<sup>59</sup> The images were rendered using VMD.<sup>60</sup> For clarity, only 10 of the 13 subunits are shown in a and b.

analytically represent the solvent as a continuum.<sup>25–50</sup> An important benefit of implicit solvent simulations is that conformational space is sampled faster due to the reduction of solvent viscosity.<sup>51,52</sup> Other benefits include instantaneous dielectric response from the solvent due to changes in solute charge state, and the elimination of “noise” in the energy landscape due to small variations in solvent structure.<sup>53</sup> Consequently, implicit solvent models are often used for applications where it is important to explore a large number of conformational states, such as for protein folding,<sup>54</sup> replica exchange,<sup>55</sup> and docking simulations.<sup>56</sup> However, the functional form for the most widely used practical implicit solvent model for MD, the GB model, scales as  $\sim n^2$ , where  $n$  represents the number of solute atoms only. ( $n < N$  where  $N$  refers to the total number of atoms, including solute and solvent atoms, used for explicit solvent computations, while  $n$  refers to the number of solute atoms only, used in the implicit solvent computations.) One approach for reducing computational cost is to apply the spherical cutoff concept to the GB implicit solvent model, i.e., ignore interactions and computations involving atoms beyond a cutoff distance. We refer to this approach as the cutoff-GB method. Such an approach can reduce computational cost to  $\sim n \log n$ . However, the cutoff-GB may suffer from the same shortcomings as the spherical cutoff method, such as spurious forces and artificial structures around the cutoff distance. Although there are studies based on the successful use of the cutoff-GB method,<sup>34,57</sup> we are not aware of a large scale systematic study that examines the effect of the cutoff on the accuracy of the GB model. To the best of our knowledge, the GB model has not been used with the PME or the fast multipole methods, most likely because the functional form of the GB model does not easily lend itself to the Ewald transformation used by the PME method or the multipole expansion used by the fast multipole method.

We present here an  $\sim n \log n$  GB approximation that retains the simplicity of the cutoff-GB approximation, while in most cases being more accurate for the set of test structures considered here. Moreover, our testing demonstrates that the method presented here more accurately reproduces important characteristics of dynamics compared to the cutoff-GB method. Our approach is based on the hierarchical charge partitioning (HCP) approximation developed by us previously.<sup>58</sup> To approximate long-range electrostatic interactions, the HCP uses the natural organization of biomolecules into multiple hierarchical levels of components, as illustrated in Figure 1—atoms (level 0); nucleic and amino acid groups (level 1); protein, DNA, and RNA subunits (level 2); complexes of multiple subunits (level 3); and higher level structures such as fibres and membranes. The charge distribution for components above the atomic level are approximated by a much smaller number of charges. For components that are distant from the point of interest, these approximate charges are used in the computation of electrostatic interaction, while the atomic charges (level 0) are used for nearby components (Figure 2). The greater the distance from the point of interest, the larger (higher level) is the component used in the approximation of electrostatic interactions. In our previous study, we have shown that this approximation scales as  $\sim n \log n$  for biomolecular structures. The HCP concept is used here to reduce the computational cost of each of the three  $\sim n^2$  computations in the GB model—the computation of electrostatic vacuum energy, solvation energy, and the so-called effective Born radii—to  $\sim n \log n$ .

The remainder of the paper is organized as follows. In the Methods section, we briefly review the GB implicit solvent model and describe how the HCP concept is applied to the implicit solvent GB model (HCP-GB). The HCP-GB method was tested using a set of representative biomolecular structures ranging in



**Figure 2.** The HCP threshold distance. For the first level of approximation shown here, groups within the threshold distance from the point of interest are treated exactly using atomic charges (level 0), while groups beyond the threshold distance are approximated by a small number of charges (level 1). The distance to a group is computed from the point of interest to the geometric center of the group.

size from 632 atoms to 3 016 000 atoms with absolute total charge ranging from 1 to 21 424 e. The accuracy, speedup, dynamics, and conservation of momentum and energy are discussed in the Results section. In the Conclusion section, we summarize our finding and discuss the applicability of HCP-GB to practical MD problems.

## 2. METHODS

A short description of the GB implicit solvent model is included below, followed by a detailed description of how the HCP approximation is used to reduce the computational cost of the GB model from  $\sim n^2$  to  $\sim n \log n$ . Also included in this section is a description of the structures and protocols used for testing the HCP-GB method.

**2.1. Generalized Born (GB) Implicit Solvent Model.** The electrostatic energy of a system,  $E^{\text{elec}}$ , in the presence of a solvent can be approximated by the GB implicit solvent model<sup>61</sup> as

$$E^{\text{elec}} = E^{\text{vac}} + E^{\text{solv}} \quad (1)$$

$$E^{\text{vac}} = \sum_i^n \sum_{j>i}^n \frac{q_i q_j}{r_{ij}} \quad (2)$$

$$E^{\text{solv}} \approx -\frac{1}{2} \left( 1 - \frac{1}{\varepsilon_w} \right) \sum_i^n \sum_j^n \frac{q_i q_j}{[r_{ij}^2 + B_i B_j e^{(-r_{ij}^2/4B_i B_j)}]^{1/2}} \quad (3)$$

where  $E^{\text{vac}}$  and  $E^{\text{solv}}$  are the electrostatic vacuum and solvation energy,  $\varepsilon_w$  is the dielectric constant of the solvent,  $q_i$  and  $q_j$  are the charges of atoms  $i$  and  $j$ ,  $r_{ij}$  is the distance between the atoms, and  $B_i$  and  $B_j$  are their effective Born radii. For the purpose of this work, we consider the following Coulomb field approximation for the effective Born radius,  $B_i$ , as implemented in NAB (or Amber):<sup>62</sup>

$$\frac{1}{B_i} \approx \frac{1}{R_i} - \frac{1}{4\pi} \int_{|r_{ik}| > R_i}^{\text{solute}} \frac{1}{|r_{ik}|^4} dV \quad (4)$$

where  $R_i$  is the intrinsic radius of charge  $i$ ,  $r_{ik}$  is the distance from  $i$  to any point  $k$  in the solute volume, and  $\int_{|r_{ik}| > R_i}^{\text{solute}} dV$  is the volume integral over the volume occupied by the solute (the cavity formed in the solvent by the solute) excluding the volume of the atom  $i$  itself.

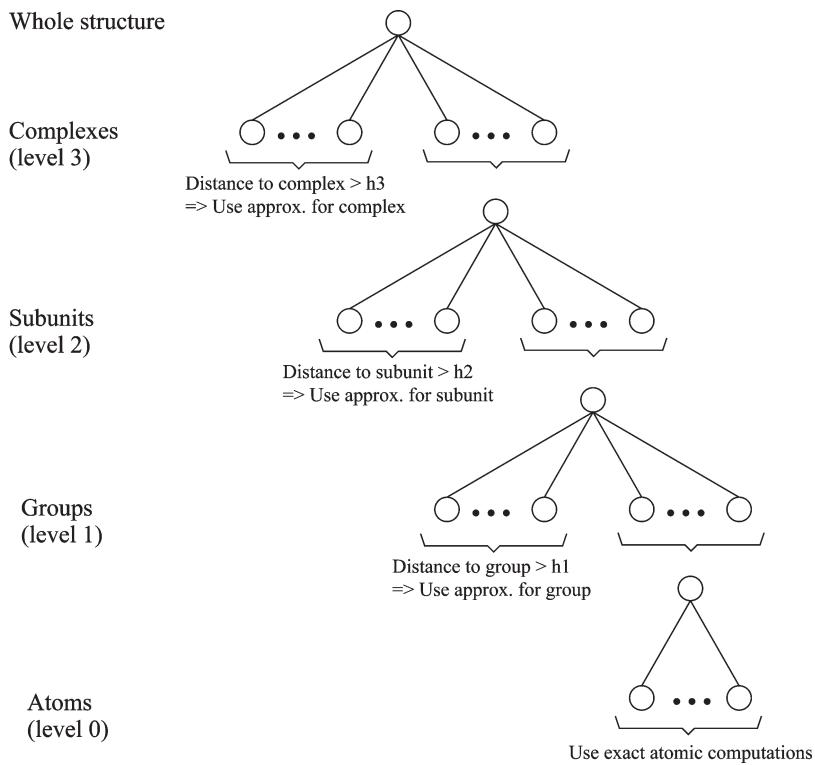
**2.2. HCP-GB—An  $n \log n$  GB Approximation.** Note that the computation of electrostatic vacuum energy  $E^{\text{vac}}$  in eq 2 and solvation energy  $E^{\text{solv}}$  in eq 3 both scale as  $\sim n^2$ . In MD software

that implements the GB implicit solvent model, such as Amber,<sup>62</sup> analytical pairwise approximations for computing the effective Born radii  $B_i$  in eq 4 also scale as  $\sim n^2$ , unless further approximations are made. The HCP concept is used to reduce the computational cost for each of these computations to  $\sim n \log n$ , as described below.

**2.2.1.  $n \log n$  Approximation for Electrostatic Vacuum Energy.** The previous HCP study<sup>58</sup> describes in detail the  $\sim n \log n$  approximation for computing electrostatic vacuum energy  $E^{\text{vac}}$ . The key concepts from the study are summarized here. Biomolecular structures are naturally organized into multiple hierarchical levels as illustrated in Figure 1 for a chromatin fiber. Atoms are at the lowest level (level 0); groups of atoms form amino and nucleic acids (level 1); protein, DNA, and RNA chains made up of these groups form subunits (level 2); multiple subunits form complexes (level 3); and multiple complexes join together to form larger structures such as fibers and membranes. The HCP approximates atomic charges within each of the components above level 0, by a much smaller number of charges (1 or 2). For the one-charge approximation for a component, the approximate charge is placed at the “center of charge” for the component with a charge value equal to the net charge of the component. For the two-charge approximation, the two approximate charges are placed at the “center of charge” of the positive and negative charges with charge values equal to the total positive and negative charges, respectively. The center of charge is calculated in a manner similar to the center of mass when the total charge is nonzero.<sup>58</sup> When the total charge is zero, the component does not contribute to the approximate computation and is ignored.

The HCP then uses these approximate charges for computing electrostatic interactions beyond predefined threshold distances (Figure 2). For example, consider a structure consisting of four levels, 0–3, see Figure 3. A separate threshold distance,  $h_1$ ,  $h_2$ , and  $h_3$ , is defined for levels 1, 2, and 3, respectively. For complexes (level 3) farther than  $h_3$  from the point of interest, the approximate charges for the complex are used in the computation. Otherwise, for subunits (level 2) within the complex that are farther than  $h_2$ , the approximate charges for the subunit are used in the computation. Otherwise, for groups (level 1) within the subunit that are farther than  $h_1$ , the approximate charges for the group are used in the computation. Finally, individual atomic charges are used in the computations for charges within the level 1 threshold distance  $h_1$ . This top-down algorithm results in  $\sim n \log n$  scaling based on assumptions generally consistent with realistic biomolecular systems. Consider a hypothetical structure consisting of  $n$  atoms such that, for any given atom, there are  $k$  atoms (level 0) within the level 1 threshold distance  $h_1$ ,  $k$  groups (level 1) between  $h_1$  and the level 2 threshold distance  $h_2$ ,  $k$  subunits (level 2) between  $h_2$  and the level 3 threshold distance  $h_3$ , and so on. Such a structure can be represented as a hierarchical tree with each internal node representing a component with  $k$  nodes immediately below each internal node, and with a total of  $n$  leaf nodes representing the atoms. The computational cost of the HCP algorithm for such a structure scales as  $\sim n \log n$ . For a more detailed description, refer to the previous HCP study.<sup>58</sup> This previous study also showed that for the computation of electrostatic vacuum energy,  $E^{\text{vac}}$ , in eq 2, the relatively simple HCP approximation can be comparable in accuracy to the more complex particle mesh Ewald (PME) method and more accurate than the simple spherical cutoff method.

**2.2.2.  $n \log n$  Approximation for Solvation Energy.** To reduce the computational cost of solvation energy,  $E^{\text{solv}}$ , in eq 3 from  $\sim n^2$  to  $\sim n \log n$ , we first define a component effective



**Figure 3.** Illustration of the HCP approximation. Biomolecular structures are naturally organized into multiple hierarchical levels of components—complexes, subunits, groups, and atoms—as represented by the tree structure shown here. Approximations are used for computations involving distant components, while exact atomic computations are used for atoms within nearby groups. The HCP algorithm proceeds from the top level down to the lowest level to determine the level of approximation to use. The level of approximation used is determined by the distance of a component from the point of interest compared to the threshold distance for the level of the component— $h_1$ ,  $h_2$ , and  $h_3$  for levels 1, 2, and 3, respectively.

Born radius,  $B_c$  for components above the atomic level. The component effective Born radii,  $B_c$  are then used to approximate the contribution of distant components to the solvation energy of atom  $i$  instead of the effective Born radii,  $B_j$ , of the individual atoms within these distant components.

To derive a simple functional form for the component effective Born radius,  $B_c$ , we consider the limit of  $r_{ij}, r_{ic} \rightarrow 0$ , where  $r_{ij}$  is the distance from atom  $i$  to atom  $j \in c$  and  $r_{ic}$  is the distance from atom  $i$  to component  $c$ . Let  $E_{ic}^{\text{solv}}$  represent the contribution of component  $c$  to the solvation energy of atom  $i$ .

$$E_{ic}^{\text{solv}} = -\frac{1}{2} \left(1 - \frac{1}{\varepsilon_w}\right) \sum_{j \in c} \frac{q_i q_j}{[r_{ij}^2 + B_i B_j e^{(-r_{ij}^2/4B_i B_j)}]^{1/2}} \quad (5)$$

$$\approx -\frac{1}{2} \left(1 - \frac{1}{\varepsilon_w}\right) \sum_{j \in c} \frac{q_i q_j}{[B_i B_j]^{1/2}} \quad (r_{ij} \rightarrow 0) \quad (6)$$

$$\text{Also, } E_{ic}^{\text{solv}} = -\frac{1}{2} \left(1 - \frac{1}{\varepsilon_w}\right) \frac{q_i q_c}{[r_{ic}^2 + B_i B_c e^{(-r_{ic}^2/4B_i B_c)}]^{1/2}} \quad (7)$$

$$\approx -\frac{1}{2} \left(1 - \frac{1}{\varepsilon_w}\right) \frac{q_i q_c}{[B_i B_c]^{1/2}} \quad (r_{ic} \rightarrow 0) \quad (8)$$

$$\Rightarrow \frac{1}{B_c} \approx \left[ \frac{1}{q_c} \sum_{j \in c} \frac{q_j}{B_j^{1/2}} \right]^2 \quad (9)$$

For the one-charge HCP approximation,  $q_c$  is the net charge of the component, and the sum in the above equations is over all the atoms in the component. For the two-charge approximation, two separate component effective Born radii are computed, one for each of the two approximate charges, i.e., one for the positively charged atoms and another for the negatively charged atoms. In this case,  $q_c$  represents the total positive or negative charge and the sum is over the positively or negatively charged atoms, respectively.

In this work, we also considered two alternatives to eq 9 for component effective Born radii. One approximation, by Archontis and Simonson<sup>39</sup> developed in the context of a coarse grain model, defines the equivalent of the component effective Born radius as the harmonic average of its constituent atomic Born radii weighted by the square of atomic charges. The resulting expression is similar to eq 9 except that the constituent atomic Born radii are weighted by atomic charges. Another approach is to use the analytical approximation for effective Born radii defined by eq 10 described below, with  $i$  representing a component  $c$  instead of an atom, and  $j \neq i$  replaced by  $j \notin c$ . We examined these alternatives (results included in Appendix A.1) and found that on average the approach described above by eq 9 is more accurate, although in some specific instances, one of the other alternatives can be more accurate. We have therefore chosen to base all further analysis on eq 9 but note that future work may lead to better approximations.

**2.2.3.  $n \log n$  Approximation for Effective Born Radii  $B_i$ .** To compute the integral in eq 4, the Coulomb field approximation for effective Born radii  $B_i$  we will consider here one commonly

used approximation to eq 4. However, the main idea can be applied to any volume based approximation for computing effective Born radii. In the specific approximation considered here, the integral in eq 4 is computed over the volume occupied by individual atoms, ignoring overlaps between atoms and spaces between atoms that are inaccessible to the solvent.<sup>44</sup> For the case where the atoms  $i$  and  $j$  do not overlap, this analytical approximation computes the effective Born radius  $B_i$  as

$$\frac{1}{B_i} \approx \frac{1}{R_i} - \sum_{j \neq i} \left[ \frac{R_j}{2[r_{ij}^2 - R_j^2]} - \frac{1}{4r_{ij}} \log \frac{r_{ij} + R_j}{r_{ij} - R_j} \right] \quad (10)$$

The computation of effective Born radii using the above approximation scales as  $\sim n^2$ . The HCP approximation can be used to reduce the computational cost of the above equation to  $\sim n \log n$ , as follows. We define a component radius  $R_c$  for a component  $c$  which can be used to approximate the contribution of distant components to the effective Born radius of an atom, replacing the computations involving the individual atoms within the component. Then, using the HCP approach described in section 2.2.1 above, the effective Born radius  $B_i$  for atom  $i$  can be approximated in  $\sim n \log n$  computations.

To derive a simple expression for component radius  $R_c$ , we consider the limit of  $r_{ij} \rightarrow r_{ic}$ , where  $r_{ij}$  is the distance from atom  $i$  to atom  $j \in c$ , and  $r_{ic}$  is the distance from atom  $i$  to component  $r_{ij}$ . For distant components  $c$ , let  $B_{ic}$  be the contribution of the atoms  $j \in c$ , to the effective Born radius of  $i$ .  $B_{ic}$  can be approximated by a truncated Taylor series expansion of eq 10 as

$$\frac{1}{B_{ic}} = C \sum_{j \in c} \frac{R_j^3}{r_{ij}^4} + \text{higher order terms} \quad (11)$$

$$\approx C \sum_{j \in c} \frac{R_j^3}{r_{ic}^4} (r_{ij} \rightarrow r_{ic}) \quad (12)$$

$$\text{Also, } \frac{1}{B_{ic}} = C \frac{R_c^3}{r_{ic}^4} \quad (13)$$

$$\Rightarrow R_c^3 \approx \sum_{j \in c} R_j^3 \quad (14)$$

where  $C$  is a constant. Interestingly,  $R_c$  is the radius of a sphere with the same volume as the sum of volumes of its constituent atoms, which is what one may intuitively expect. On the basis of the simple and intuitive nature of the expression for  $R_c$  we conjecture that the form of eq 14 is independent of the specifics of eq 10. For a distant component  $c$ , the component radius  $R_c$  is used in eq 10 in place of  $R_j$  and  $r_{ic}$  in place of  $r_{ij}$  for atoms  $j \in c$ . Higher level components are used in the computation of effective Born radii  $B_i$  for more distant components such that computational cost scales as  $\sim n \log n$ , as described in section 2.2.1 above.

**2.2.4.  $n \log n$  Approximation for Solvation Forces.** The solvation force on an atom  $i$ ,  $F_i^{\text{solv}}$ , is computed as the derivative of the solvation potential  $\phi_i^{\text{solv}}$  using the chain rule, as follows:

$$\phi_i^{\text{solv}} = \sum_{j \neq i} \phi_{ij}^{\text{solv}} \quad (15)$$

$$\phi_{ij}^{\text{solv}} = -\frac{1}{2} \left( 1 - \frac{1}{\varepsilon_w} \right) \frac{q_j}{[r_{ij}^2 + B_i B_j e^{(-r_{ij}^2/4B_i B_j)}]^{1/2}} \quad (16)$$

$$F_i^{\text{solv}} = \sum_{j \neq i} F_{ij}^{\text{solv}} \quad (17)$$

$$F_{ij}^{\text{solv}} = -q_i \partial \phi_{ij}^{\text{solv}} / \partial r_{ij} \quad (18)$$

$$= -\frac{1}{4} \left( 1 - \frac{1}{\varepsilon_w} \right) \frac{q_i q_j}{[r_{ij}^2 + B_i B_j e^{-r_{ij}^2/4B_i B_j}]^{3/2}} \times \quad (19)$$

$$\left[ 2r_{ij} - \frac{1}{2} r_{ij} e^{-r_{ij}^2/4B_i B_j} + e^{-r_{ij}^2/4B_i B_j} \left( 1 + \frac{r_{ij}^2}{4B_i B_j} \right) \left( B_j \frac{\partial B_i}{\partial r_{ij}} + B_i \frac{\partial B_j}{\partial r_{ij}} \right) \right] \quad (20)$$

$$\frac{\partial B_i}{\partial r_{ij}} = B_i^2 \sum_{j \neq i} \left[ \frac{-R_j r_{ij}}{(r_{ij}^2 - R_j^2)^2} + \frac{1}{4r_{ij}^2} \log \frac{R_j + r_{ij}}{R_j - r_{ij}} + \frac{R_j}{2r_{ij}(R_j^2 - r_{ij}^2)} \right] \quad (21)$$

where  $\phi_{ij}^{\text{solv}}$  is the solvation potential contribution of atom  $j$  at atom  $i$ ,  $F_{ij}^{\text{solv}}$  is the corresponding force contribution,  $\varepsilon_w$  is the dielectric constant of the solvent,  $r_{ij}$  is the distance between atoms  $i$  and  $j$ ,  $q_i$  and  $q_j$  are the atomic charges, and  $R_i$  and  $R_j$  are the intrinsic radii. Here,  $\partial B_i / \partial r_{ij}$  is the derivative of the effective Born radii (eq 10). For distant components, component intrinsic radii and component effective Born radii are used in the above equations instead of the atomic intrinsic radii and atomic effective Born radii. Using the HCP approach described in section 2.2.1 above, solvation forces can then be approximated in  $\sim n \log n$  computations.

**2.3. Test Structures and Protocols.** To assess performance of the HCP-GB method in the context of molecular dynamics, we implemented the method in NAB, the open source molecular dynamics (MD) software in AmberTools v1.3.<sup>63</sup> The HCP implementation in NAB is scheduled to be released with AmberTools v1.5, for general use. In some sense, NAB is a minimal version of the production Amber MD software<sup>62</sup> and is particularly well suited for experimentation unlike the highly optimized but also more complex production version. NAB however does use the same force fields and implements the same GB implicit solvent methods and options as the production Amber code. For the purpose of this study, we used the commonly used OBC GB model (IGB = 5 in Amber<sup>65</sup>).

Performance in both accuracy and speed was evaluated relative to the reference GB computation without any additional approximations (reference GB). We also compared the HCP-GB method to the same GB implicit solvent model with a spherical cutoff (cutoff-GB). The cutoff-GB method ignores all interactions beyond a cutoff distance for the computation of electrostatic energy and effective Born radii in eqs 1–3 and 10. Our previous study<sup>58</sup> had compared the electrostatic vacuum energy and forces computed by the HCP method to the particle mesh Ewald (PME) explicit solvent method. However, to the best of our knowledge, the GB implicit solvent model has not been implemented for the PME method in readily available molecular dynamics software. Therefore, a similar comparison for the HCP-GB method was not performed here.

The HCP-GB method was tested on a set of eight representative biomolecular structures ranging in size from 632 atoms to 3 016 000 atoms with absolute total charge ranging from 1 to 21 424e (Table 1). The H++ server (<https://biophysics.cs.vt.edu/H++>) was used to add missing hydrogens to these structures.<sup>64</sup>

**Table 1.** List of Representative Structures Used for Testing<sup>a</sup>

structure	PDB ID	size (atoms)	charge  (e)	cutoff dist (Å)	threshold dist (Å)		
					<i>h</i> <sub>1</sub>	<i>h</i> <sub>2</sub>	<i>h</i> <sub>3</sub>
10 bp B-DNA fragment	2BNA	632	18	21	21	n/a	n/a
immunoglobulin binding domain	1BDD	726	2	15	15	n/a	n/a
ubiquitin	1UBQ	1231	1	15	15	n/a	n/a
thioredoxin	2TRX	1654	5	15	15	n/a	n/a
nucleosome core particle	1KX5	25101	133	21	21	90	n/a
microtubule sheet	<i>b</i>	158016	360	15	15	48	n/a
virus capsid	1A6C	475500	120	15	15	66	n/a
chromatin fiber	<i>c</i>	3016000	21424	21	21	90	169

<sup>a</sup> Unless stated otherwise, the cutoff and threshold distances listed here were used for all testing. <sup>b</sup> The microtubule sheet was constructed as described in Wang and Nogales.<sup>66</sup> <sup>c</sup> The chromatin fiber was constructed as described in Wong et al.<sup>59</sup>

The HCP threshold distances were chosen such that, for a given atom within a given test structure, the exact atomic computation (level 0) is used for interactions with other atoms within its own and nearest neighboring groups (level 1), as illustrated in Figure 2. To satisfy this condition, threshold distances *h<sub>l</sub>* are calculated as *h<sub>l</sub>* = *R<sub>l</sub><sup>max</sup>* + 2 × *R<sub>1</sub><sup>max</sup>* where *l* is the HCP level, *R<sub>l</sub><sup>max</sup>* is the maximum component radius at level *l*, and *R<sub>1</sub><sup>max</sup>* is the maximum group (level 1) radius, for a given structure. The HCP threshold distances thus calculated for each of the test structures are shown in Table 1. These are the suggested conservative defaults for these and other similar structures. The HCP-GB level 1 threshold distance for a given structure is also used as the cutoff distance for the cutoff-GB computations. Unless stated otherwise, these threshold and cutoff distances were used for all of the testing described in the Results section.

Four metrics were used to measure the accuracy of the approximate methods: relative error in electrostatic energy (relative energy error) Err<sup>E</sup>, for vacuum, solvation, and net electrostatic energy, and relative RMS error in electrostatic force (relative force error) Err<sup>F</sup>, calculated as

$$\text{Err}^E = |E^{\text{approx}} - E^{\text{ref}}|/E^{\text{ref}} \quad (22)$$

$$\text{Err}^F = \text{Err}^{\text{rms}}/F^{\text{avg}} \quad (23)$$

$$\text{Err}^{\text{rms}} = [\sum_{i=1}^n |F_i^{\text{approx}} - F_i^{\text{ref}}|^2/n]^{1/2} \quad (24)$$

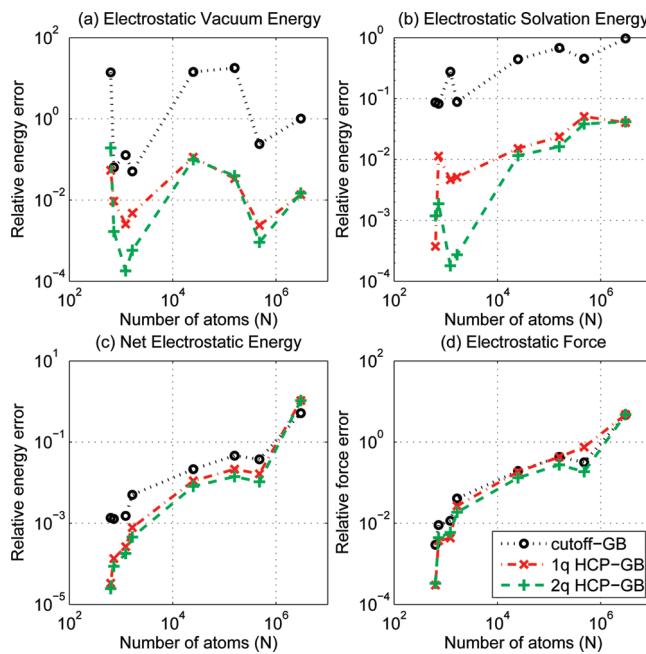
$$F^{\text{avg}} = \sum_{i=1}^n |F_i^{\text{ref}}|/n \quad (25)$$

where *E<sup>approx</sup>* is the energy calculated using an approximation, *E<sup>ref</sup>* is the energy calculated using the reference GB computation without cutoffs or the use of HCP, Err<sup>rms</sup> is the root-mean-square (RMS) error in force for the atoms in a given structure, *F<sup>avg</sup>* is the average force, and *F<sub>i</sub><sup>approx</sup>* and *F<sub>i</sub><sup>ref</sup>* are the force on atom *i* calculated using the approximate and reference GB computations, respectively.

Speedup was measured as CPU time for the reference (no cutoff) GB computation divided by the CPU time for the approximation tested.<sup>a</sup> All testing was conducted on Virginia Tech's System X computer cluster (<http://www.arc.vt.edu>) consisting of 1100 dual core 2.5 GHz PowerPC 970FX processors with 4

GB of RAM, running the Apple Mac OS X 10.3.9, and connected by 10 Gbps InfiniBand switches. Where possible, testing was performed using a single CPU (a single core of the dual core processor) to reduce the potential variability due to interprocessor communication. However, due to the large memory requirements for the neighbor list used by the cutoff-GB method, it was not possible to run the cutoff-GB computation for structures larger than 200 000 atoms using a single CPU in the test environment described above. Therefore, 16 CPUs were used for the 475 500 atom virus capsid and 128 CPUs for the 3 016 000 atom chromatin fiber. For comparison on an equal footing, the reference GB computations and the HCP-GB computations were also performed with the same number of CPUs. When multiple CPUs were used, the CPU time for the longest running CPU was used to calculate speedup. To limit the run time for the reference GB computation to a few days, speedup was calculated for 1000 iterations of MD for structures with <10 000 atoms, 100 iterations for structures with 10 000–1 000 000 atoms, and 10 iterations for the structure with >1 000 000 atoms. To make the results representative of typical simulations involving much larger numbers of iterations, the CPU time excludes the time for loading the data and initialization prior to starting the simulation. Note that the speedup may vary with the computing system characteristics, such as interprocessor communication network, number of processors used, processor architecture, memory configuration, etc. A detailed analysis of the effect of these characteristics on speedup is beyond the scope of this study, which focuses on the algorithm.

The following parameters and protocol were used for the simulations, unless otherwise stated. The threshold distances used are listed in Table 1. 6–12 van der Waals interactions for the HCP-GB were computed using only the atoms that are within the level 1 threshold distance, i.e., atoms that are treated exactly. The simulations used the Amber ff99SB force field.<sup>67</sup> Langevin dynamics with a collision frequency of 50 ps<sup>-1</sup> (appropriate for water) was used for temperature control, a surface-area dependent energy of 0.005 kcal/mol/Å<sup>2</sup> was added, and an inverse Debye–Hückel length of 0.125 Å<sup>-1</sup> was used to represent a 0.145 M salt concentration. A 1 fs time step was used for the simulation with the nonbonded neighbor list being updated after every step. Note that updating the nonbonded neighbor list less frequently will improve the speedup of the cutoff-GB method; however, the speedup of the HCP-GB method can also be improved similarly by updating component radii and charges less frequently. For simplicity and for



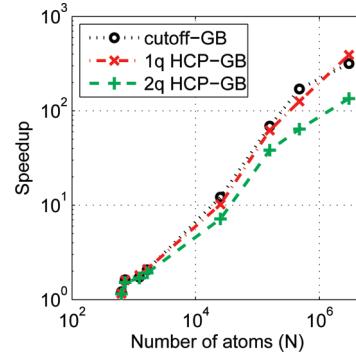
**Figure 4.** Accuracy of the HCP-GB and cutoff-GB methods relative to the reference GB computation without cutoffs. Accuracy is computed as the relative error in (a) vacuum, (b) solvation, and (c) net electrostatic energy, and (d) relative RMS error in electrostatic force, for the one-charge and two-charge HCP-GB and the cutoff-GB method. Connecting lines are shown to guide the eye.

comparison on an equal footing, the nonbonded neighbor list and component radii and charges are updated after every step. Default values were used for all other parameters. The simulation protocol consisted of five stages. First, the starting structure was minimized using the conjugate gradient method with a restraint weight of 5.0 kcal/mol/Å<sup>2</sup>. Next, the system was heated to 300 K over 10 ps with a restraint weight of 1.0 kcal/mol/Å<sup>2</sup>. The system was then equilibrated for 10 ps at 300 K with a restraint weight of 0.1 kcal/mol/Å<sup>2</sup>, and then for another 10 ps with a restraint weight of 0.01 kcal/mol/Å<sup>2</sup>. Finally, all restraints were removed for the production stage.

### 3. RESULTS AND DISCUSSION

We examined a number of characteristics of the HCP-GB method that are important for molecular dynamics—accuracy, speed, dynamics, and conservation of energy and momentum, which are discussed below.

**3.1. Accuracy.** Figure 4 shows the accuracy for the one- and two-charge HCP-GB methods compared to the cutoff-GB method. For the test structures considered here, the values of the two components of electrostatic interactions—vacuum and solvation energies—as calculated by the HCP-GB method are significantly more accurate than that of the cutoff-GB method (Figure 4a and b). For the net electrostatic energy and force, the relative improvement provided by the HCP-GB is significant for the smaller structures and decreases with structure size (Figure 4c and d). For the largest structure considered here—the 3 million atom chromatin fiber—the relative energy error for the HCP-GB method is slightly higher than that of the cutoff-GB method. Preliminary analysis suggests that the larger HCP-GB error for the chromatin fiber may be due to the negligible contribution of very distant components to the net energy



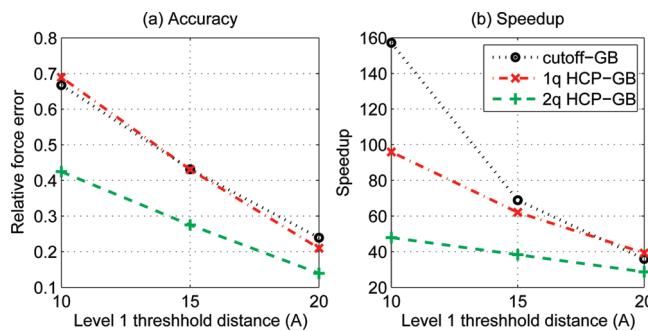
**Figure 5.** Speedup for the HCP-GB and cutoff-GB methods relative to the reference GB computation without cutoffs. Threshold and cutoff distances used for the different structures are listed in Table 1. Connecting lines are shown to guide the eye.

computation, even though the individual vacuum and solvation components may be large. Small errors in the estimation of these individual components can result in a large relative error in net electrostatic energy. Thus, ignoring the contribution of these distant components, as the cutoff-GB method does, may actually decrease the error in total energy as defined by the above metrics. However, as our examination of key characteristics of dynamics in section 3.6 below shows, the single point net force and net energy error metrics presented above are too crude to unambiguously differentiate between the expected performance of the cutoff-GB and HCP-GB methods in the context of molecular dynamics. For example, one can expect the cutoff scheme to neglect a roughly equal number of pairwise interactions of roughly equal magnitude but of opposite sign. The resulting cancellation of error in total electrostatic energy can be deceptive. As we shall see later in section 3.6, neglect of charge–charge interactions clearly manifests itself by producing artifacts in dynamics.

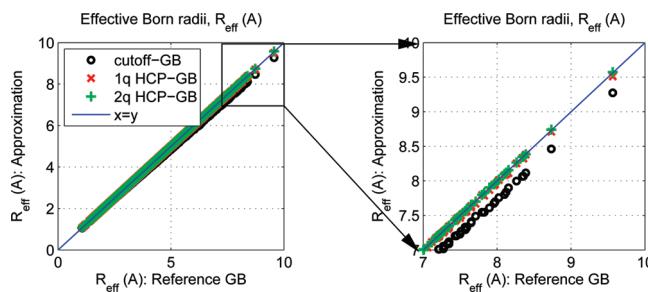
**3.2. Speedup.** Figure 5 shows that the speedup for the one-charge HCP-GB and cutoff-GB methods are comparable,<sup>b</sup> while the two-charge HCP-GB is slower. Surprisingly, the cutoff-GB method is slower than the one-charge HCP-GB method for the three million atom chromatin fiber. We speculate that this is because the NAB implementation of the cutoff method does not scale well with system size due to the additional memory access required for the large neighbor list used by the method.

As noted earlier, unlike the production pmemd module of Amber 8, NAB is not highly optimized. However, on the basis of the run times for a 0.1 ns simulation of the nucleosome core particle (1KX5), compared to an equivalent simulation by Ruscio and Onufriev,<sup>68</sup> we estimate that NAB v1.3 is only about 1.5 times slower than the production pmemd module of Amber 8 on Virginia Tech's System X computer cluster described above.

**3.3. Tradeoff between Speed and Accuracy.** For a given structure, the speed and accuracy of the HCP-GB method depends primarily on two parameters: the number of charges used to approximate the components and the threshold distances. As seen in Figures 4 and 5, on the basis of net energy and force metrics, the two-charge approximation is more accurate but slower than the one-charge approximation. Figure 6 shows that increasing the threshold distance improves accuracy but reduces speed. However, as our analysis of key characteristics of dynamics (section 3.6) shows, the single-point error metrics, based on net energy or force, used above may not provide a complete measure



**Figure 6.** Tradeoff between (a) accuracy and (b) speed for the 158 016 atom microtubule structure. Cutoff and level 1 threshold distances are varied from 10 Å to 20 Å. Level 2 threshold distance is 48 Å. Connecting lines are shown to guide the eye.



**Figure 7.** Accuracy of effective Born radii approximations. Effective Born radii for the 1653 atoms of thioredoxin (2TRX) are calculated using the one-charge HCP-GB, the two-charge HCP-GB, and the cutoff based (cutoff-GB) approximations and plotted against the reference GB computation without cutoffs.

of correctness in the context of molecular dynamics. The optimal choice of parameters depends on the structure and problem under consideration. For the purpose of this study, we have chosen conservative threshold distances (Table 1) such that for any given atom, atoms within its own group and immediately neighboring groups (level 1) are treated exactly, as described in section 2.3. It is possible that shorter than default threshold distances may be acceptable for specific applications, but we suggest that the decision to use shorter threshold distances be made on a case-by-case basis. For example, the two-charge HCP-GB simulation of immunoglobulin binding domain (1BDD) remains stable when the level 1 threshold distance is reduced from the recommended 15 Å to 10 Å, but the protein quickly unfolds when the threshold distance is further reduced to 5 Å.

**3.4. Accuracy of the HCP Approximation for Effective Born Radii.** We tested the HCP based approximation for effective Born radii on a typical structure used in this context, thioredoxin (2TRX). Figure 7 shows that for this structure, the HCP-GB approximation with a threshold distance of 15 Å is slightly more accurate than a cutoff based approximation with a 15 Å cutoff distance. The overall RMS error in effective Born radii relative to the reference GB computation without cutoffs is 0.0058 Å for the one-charge HCP-GB, 0.0017 Å for the two-charge HCP-GB, and 0.0557 Å for the cutoff-GB method.

Both the HCP-GB and the cutoff-GB introduce two sources of error into the total electrostatic energy, relative to the no-cutoff reference. One is the approximations to effective Born radii, and the other is the approximations to the electrostatic interactions.

**Table 2. Relative RMS Error in Total Electrostatic Energy for Thioredoxin (2TRX) Due to Different Approximations of Effective Born Radii<sup>a</sup>**

effective Born radii approximated using	relative RMS error in total electrostatic energy	
	energy calculated with	
	cutoff	no cutoff
cutoff	0.50%	0.49%
1-q HCP	0.05%	<0.01%

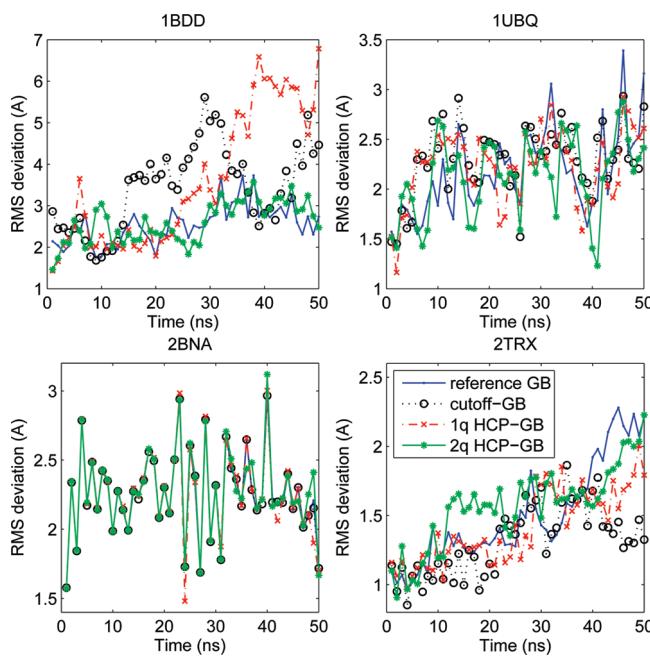
<sup>a</sup> RMS error is calculated relative to the reference GB computation. Cutoff and level 1 threshold distances of 15 Å were used for these computations.

The relative impact of these two sources is shown in Table 2. Clearly, for the spherical cutoff, the error in effective Born radii is the dominant source of error in the total electrostatic energy. The use of the HCP-GB approximation for effective Born radii can reduce this error by an order of magnitude. Whether these errors in effective Born radii will have a material impact on dynamics depends on, among other factors, the relative magnitude of the errors inherent in the approximation used to compute effective Born radii in the reference model. Nevertheless, the improvement in the accuracy of effective Born radii using the HCP approximation, compared to the cutoff approximation, comes at little or no additional cost and should therefore be used instead of the cutoff approximation.

**3.5. Stability in MD Simulations.** To test the stability of the HCP-GB algorithm, we ran 50 ns MD simulations of the immunoglobulin binding domain (1BDD), ubiquitin (1UBQ), a 10 base-pair fragment of B-DNA (2BNA), and thioredoxin (2TRX). Figure 8 shows the backbone RMS deviation from the crystal structure for the simulations, which are summarized in Table 3. These results suggest that the trajectory for the cutoff-GB and HCP-GB methods are generally in reasonable agreement with the reference GB simulation. For 1BDD, the one-charge HCP-GB trajectory shows RMS deviations similar to the cutoff-GB trajectory but substantially larger than the two-charge HCP-GB or the reference GB trajectories. This example emphasizes how subtle errors in charge–charge interactions can result in qualitatively different conformational dynamics. On a practical level, it suggests that the one-charge HCP-GB may not be appropriate for the simulation of small flexible structures, such as 1BDD, where small inaccuracies in the potential can lead to large structural deviations over the course of the trajectory. For such structures, we recommend the two-charge HCP-GB.

The above simulations were run with a Langevin collision frequency of 50 ps<sup>-1</sup> for thermal coupling. We also performed, for the same set of structures, 10 ns simulations with the thermal coupling reduced to 0.01 ps<sup>-1</sup> (results not shown). As expected, these simulations resulted in an enhanced sampling of conformational space, as was seen by more frequent excursions in RMS space. The weak Langevin coupling simulations were in general agreement with the simulations that used strong Langevin coupling. For example, for 1BDD, the two-charge HCP-GB and the reference GB simulations exhibited similar RMS deviations from the starting structure, while the one-charge HCP-GB and cutoff-GB resulted in much higher RMS deviations toward the end of the respective trajectories.

**3.6. Detailed Characteristics of the Simulation Dynamics.** An important qualitative difference between the HCP-GB



**Figure 8.** RMS deviation from the starting structure for 50 ns MD simulations of immunoglobulin binding domain (1BDD), ubiquitin (1UBQ), B-DNA (2BNA), and thioredoxin (2TRX) using the reference GB, cutoff-GB, and HCP-GB methods. RMS deviation is calculated for backbone heavy atoms. The trajectory is sampled every 1 ns. Connecting lines are shown to guide the eye.

**Table 3. RMS Deviation from the Starting Structure for 50 ns Simulations of Immunoglobulin Binding Domain (1BDD), Ubiquitin (1UBQ), B-DNA (2BNA), and Thioredoxin (2TRX)<sup>a</sup>**

PDB ID	average RMS deviation $\pm$ standard deviation (Å)				
	reference GB	cutoff-GB	one-charge HCP-GB	two-charge HCP-GB	
1BDD	2.64 $\pm$ 0.45	3.69 $\pm$ 0.88	3.92 $\pm$ 1.66	2.72 $\pm$ 0.48	
1UBQ	2.29 $\pm$ 0.37	2.39 $\pm$ 0.29	2.32 $\pm$ 0.33	2.22 $\pm$ 0.36	
2BNA	2.24 $\pm$ 0.30	2.24 $\pm$ 0.30	2.23 $\pm$ 0.31	2.25 $\pm$ 0.31	
2TRX	1.60 $\pm$ 0.33	1.41 $\pm$ 0.23	1.50 $\pm$ 0.23	1.67 $\pm$ 0.19	

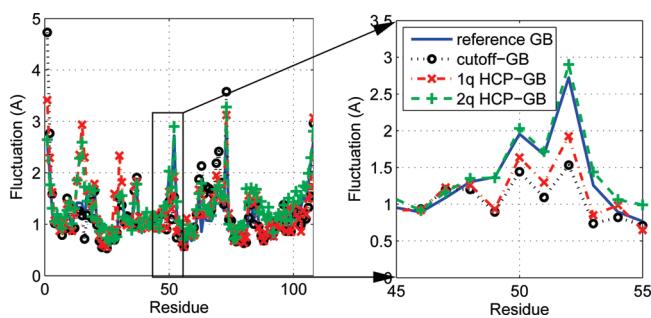
<sup>a</sup> RMS deviation is calculated for backbone heavy atoms. The trajectory is sampled every 10 ps. Averages are for the last 40 ns of the 50 ns simulations. Standard deviation is computed as  $\sqrt{[\sum_i (\text{RMS}_i - \mu)^2]/s}$ , where RMS<sub>i</sub> is the RMS deviation for the i<sup>th</sup> sample,  $\mu$  is the average RMS deviation, and s is the number of samples.

method and the cutoff-GB method is that the cutoff-GB method completely ignores the effect of all charges beyond the cutoff distance, while the HCP-GB method approximates the effect of distant charges. We believe that ignoring these distant charges can, under many circumstances, lead to qualitatively different, and incorrect, results. Consider for example the RMS fluctuation in the position of residues—a characteristic of internal dynamics of the structure. To quantify the overall difference in fluctuation for all residues compared to the reference GB simulation, we compute the RMS difference in RMS fluctuation for the 50 ns simulation of the four structures described in the Stability section above. For the structures tested here, the RMS difference in fluctuation (Table 4) indicates that on average both the one-charge

**Table 4. Detailed Characteristics of Simulation Dynamics from 50 ns Simulations of Immunoglobulin Binding Domain (1BDD), Ubiquitin (1UBQ), B-DNA (2BNA), and Thioredoxin (2TRX)<sup>a</sup>**

PDB ID	cutoff-GB	one-charge HCP-GB	two-charge HCP-GB
RMS difference in RMS residue fluctuations (Å)			
1BDD	0.51	0.54	0.37
1UBQ	0.24	0.20	0.15
2BNA	0.003	0.01	0.03
2TRX	0.37	0.32	0.22
average	0.28	0.26	0.19
RMS difference in distribution of $\chi_1$ angles (% occurrence)			
1BDD	3.26	2.19	2.49
1UBQ	3.15	3.26	2.81
2TRX	3.17	3.48	3.16
average	3.19	2.98	2.82
RMS difference in distribution of $\chi_2$ angles (% occurrence)			
1BDD	2.79	2.01	2.21
1UBQ	2.66	2.63	2.37
2TRX	2.87	2.91	3.01
average	2.77	2.52	2.53

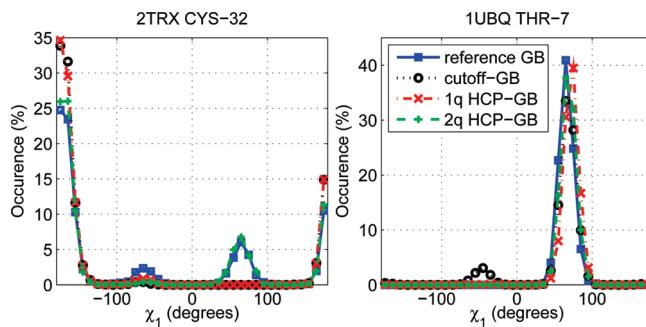
<sup>a</sup> RMS difference was calculated relative to the reference GB simulation. The trajectory was sampled every 10 ps.  $\chi$  angles do not apply to the DNA strand 2BNA. A bin size of 10° was used for calculating the distribution of  $\chi$  angles.



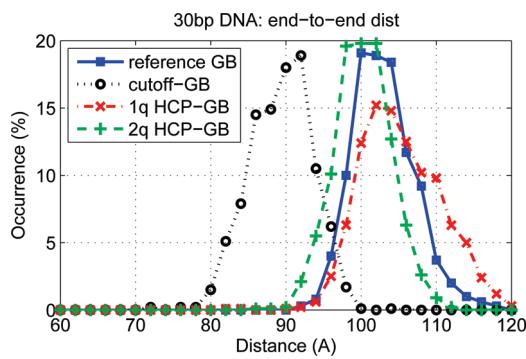
**Figure 9.** RMS fluctuation in residue positions for a 50 ns simulation of thioredoxin. The trajectory was sampled every 10 ps. Connecting lines are shown to guide the eye.

and two-charge HCP-GB simulations are in better agreement with the reference GB simulation than the cutoff-GB method. The differences in RMS fluctuation from the 50 ns simulation of thioredoxin are highlighted in Figure 9.

Similarly, consider the  $\chi_1$  angles for the functionally important CYS-32 of thioredoxin and THR-7 of ubiquitin (Figure 10). The  $\chi_1$  angle for CYS-32 flips between approximately  $-180^\circ$  and  $+60^\circ$  during the two-charge HCP-GB simulations as does the “correct” reference GB simulation. Whereas, for the one-charge HCP-GB and cutoff-GB methods, the angle stays at approximately  $-180^\circ$ . And the  $\chi_1$  angle for THR-7 stays around approximately  $60^\circ$  during the reference GB and the HCP-GB simulations, whereas for the cutoff-GB simulation, the angle flips briefly between approximately  $-60^\circ$  and  $+60^\circ$ . To quantify the overall difference in the distribution of  $\chi_1$  and  $\chi_2$  angles, we



**Figure 10.** Distribution of  $\chi_1$  angles for the functionally important CYS-32 of thioredoxin (2TRX) and THR-7 of ubiquitin (1UBQ) from 50 ns simulations. The trajectory was sampled every 10 ps. A bin size of  $10^\circ$  was used for calculating the distribution of  $\chi_1$  angles. Connecting lines are shown to guide the eye.

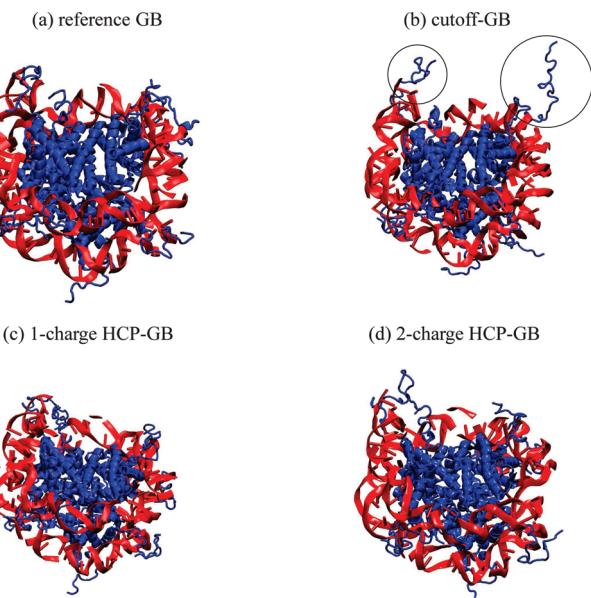


**Figure 11.** Distribution of distances between terminal base pairs for a 30 bp DNA strand from a 10 ns MD simulation. Trajectories were sampled every 10 ps, and a bin size of  $2 \text{ \AA}$  is used for calculating the distribution of distances between terminal base pairs. Connecting lines are shown to guide the eye.

computed the RMS difference in the distribution compared to the reference GB simulation. The RMS difference in the distribution of  $\chi_1$  and  $\chi_2$  angles for 50 ns simulations of the four structures described in the “Stability” section above indicates that on average the one- and two-charge HCP-GB simulations are in better agreement with the reference GB simulation than the cutoff-GB simulations (Table 4).

To further examine the effect of ignoring distant charges we ran a 10 ns simulation of a 30 base-pair DNA strand with the same setup as described in section 2.3 above, but without the salt. Figure 11 shows the distribution of distances between terminal base pairs. The distribution shows that for the cutoff-GB method, with an average end-to-end distance of 89 Å, the structure is more flexible than for the reference GB or the one- and two-charge HCP-GB methods, with average end-to-end distances of 103, 105, and 100 Å, respectively. This difference in flexibility is most likely due to the fact that the cutoff-GB method completely ignores distant charges which contribute to the bending rigidity of the DNA chain.

We also ran a 0.3 ns simulation of the nucleosome core particle, with the same setup as described in section 2.3 above, but without the salt. When using the reference GB and the HCP-GB methods, all of the histone tails collapse onto the DNA chain within 0.1 ns, consistent with experimental observations at low salt concentrations,<sup>69</sup> whereas when the cutoff-GB method is used, two of the positively charged tails fail to collapse onto the



**Figure 12.** Nucleosome core particle after 0.1 ns simulation using (a) reference GB, (b) cutoff-GB, (c) one-charge HCP-GB, and (d) two-charge HCP-GB computations. Positively charged histone tails have collapsed onto the DNA for the reference GB and HCP-GB, whereas for the cutoff-GB, two of the tails fail to collapse.

negatively charged DNA, Figure 12. Again, this is most likely because, in the case of the cutoff-GB method, the positive charges at the ends of the histone tail do not “feel” the attraction from the highly charged DNA chain.

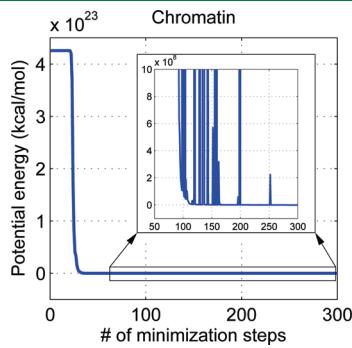
The above results suggest that in general the HCP-GB reproduces the dynamics of the reference GB simulation more accurately than the cutoff-GB method.

**3.7. A Practical Application: Chromatin Fiber.** We expect the HCP-GB to be indispensable in the modeling of large structures where the pairwise GB without further approximation is impractical. One such example is the chromatin fiber where a 348 000 atom (12 nucleosome) structure is needed at a minimum to study its functional characteristics. Such a structure can be constructed using the crystal structure for the nucleosome (1KX5) as a starting point. Multiple copies of the nucleosome can then be combined to construct the chromatin fiber, using a set of coordinate transformations described by Wong et al.<sup>59</sup> The coordinate transformations result in a number of severe steric clashes. A 15 ps simulation of the fiber using the two-charge HCP-GB significantly reduces the steric clashes, as seen by the large reduction in the potential energy (Figure 13). To reduce run time for this simulation, the protocol described in section 2.3 was modified to reduce the heating and equilibration stages from 10 to 2 ps.

**3.8. Mitigating the Effect of Violating Newton’s Third Law.** Although the HCP uses the same all-atom force field as the reference GB computation, the HCP is a multiscale model in that different levels of approximations are used for the same set of atoms depending on their distance from the point of interest. The asymmetric interactions due to the multiscale approximations can violate Newton’s third law, resulting in a residual force on the system.<sup>58,70</sup> This residual force can produce an artificial center of mass motion and an overall rotation of the structure. A net residual force within a closed system causes the system as a whole to accelerate, even though there is no external force, resulting in the nonconservation of energy. Table 5 shows the net

force and torque due to the violation of Newton's third law for the HCP-GB method, on the set of test structures considered here, along with estimated center of mass displacement, rotation, and kinetic energy after 10 steps of a typical molecular dynamics simulation (10 fs). To estimate the kinetic energies, we treat the structures as rigid bodies and assume that the principal axis of rotation passes through the center of mass.

For the test structures considered here, the 3 million atom chromatin fiber represents the worst case with a linear displacement of  $1 \times 10^{-6}$  Å, rotation of  $1 \times 10^{-8}$  radians and kinetic energy of 0.09 kT after 10 steps of MD. These spurious motions are small compared to the stochastic collisions used in constant temperature simulations, which are on the order of 1 kT, and may not materially affect the dynamics of the simulation if a strong enough coupling to a thermal bath is used. However, implicit solvent MD simulations often use minimal or no viscosity to increase the sampling of conformation space. These regimes can result in a large center of mass drift, which can be inconvenient when visualizing or analyzing the trajectory. For example, for a 1 ns simulation of thioredoxin



**Figure 13.** HCP-GB simulation of a 348 000 atom chromatin fiber. Potential energy drops rapidly as steric clashes are resolved. Connecting lines are shown to guide the eye.

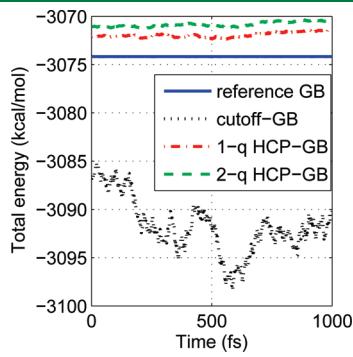
using the two-charge HCP-GB method and Langevin dynamics with a collision frequency of  $50 \text{ ps}^{-1}$ , the center of mass drift is 2.24 Å, similar to the 1.81 Å drift for the reference GB simulation. Whereas with a collision frequency of  $1 \text{ ps}^{-1}$ , the center of mass drift in the HCP-GB simulation is 30.96 Å, and as much as 183.54 Å with a collision frequency of  $0.1 \text{ ps}^{-1}$ . A commonly used approach for removing center of mass drift and rotation during the course of molecular dynamics is to employ a velocity correction algorithm, e.g., the NSCM option in Amber<sup>62</sup> which specifies the frequency at which center-of-mass motion is removed. We have implemented the same option in NAB. The velocity correction approach does not however correct the source of the problem—the net residual force. Moreover, a velocity correction not only eliminates the artificial motion caused by the violation of Newton's third law but also affects the random motion due to Langevin dynamics which may not be desirable in some situations. Therefore, we considered applying a force correction aimed at mitigating the effects of the third law violation. Appendix A.2 describes the two force correction approaches we considered—a molecular level and a component level force correction. By neutralizing the net residual forces, the force correction eliminates the systematic drift in the center of mass position caused by the violation of Newton's third law. For example, for the 1 ns simulation of thioredoxin using the two-charge HCP-GB method with a Langevin dynamics collision frequency of  $0.1 \text{ ps}^{-1}$ , the molecular level force correction reduces the drift from 183.54 Å to 29.49 Å, which is similar to the 29.47 Å drift for the reference GB simulation, which, of course, does not violate the third law within numerical precision of the integrator. The force correction however has several shortcomings compared to the velocity correction often used by existing MD algorithms. It causes an increase in the force error as described in Appendix A.2, while velocity correction does not affect the forces. Unlike velocity correction, the force correction only eliminates drift, not rotation. And, the force correction must be applied at every step of the simulation since it eliminates the center of

**Table 5. Center-of-Mass Motion Due to Violation of Newton's Third Law for HCP-GB**

structure PDB ID	residual		after 10 iterations of dynamics		
	force (kcal/mol/Å)	torque (kcal/mol)	displacement (Å)	rotation (radians)	kinetic energy (kT)
2BNA	0.003	0.38	$1 \times 10^{-8}$	$1 \times 10^{-8}$	$6 \times 10^{-9}$
1BDD	0.22	0.73	$9 \times 10^{-7}$	$3 \times 10^{-8}$	$4 \times 10^{-7}$
1UBQ	0.24	0.64	$6 \times 10^{-7}$	$1 \times 10^{-8}$	$2 \times 10^{-7}$
2TRX	0.97	3.81	$2 \times 10^{-6}$	$4 \times 10^{-8}$	$3 \times 10^{-6}$
1KX5	1.36	51.40	$1 \times 10^{-7}$	$3 \times 10^{-9}$	$6 \times 10^{-7}$
microtubule	179.24	2603.63	$3 \times 10^{-6}$	$5 \times 10^{-9}$	0.001
1A6C	$2 \times 10^{-11}$	$6 \times 10^{-10}$	$1 \times 10^{-19}$	$2 \times 10^{-22}$	$4 \times 10^{-30}$
chromatin	6804.00	1198829.76	$6 \times 10^{-6}$	$1 \times 10^{-8}$	0.09
2-q HCP-GB					
2BNA	0.015	0.07	$5 \times 10^{-8}$	$2 \times 10^{-9}$	$2 \times 10^{-9}$
1BDD	0.18	0.44	$7 \times 10^{-7}$	$2 \times 10^{-8}$	$2 \times 10^{-7}$
1UBQ	0.23	0.58	$6 \times 10^{-7}$	$1 \times 10^{-8}$	$2 \times 10^{-7}$
2TRX	0.98	4.64	$2 \times 10^{-6}$	$5 \times 10^{-8}$	$3 \times 10^{-6}$
1KX5	2.94	57.98	$3 \times 10^{-7}$	$4 \times 10^{-9}$	$2 \times 10^{-6}$
microtubule	59.76	598.27	$1 \times 10^{-6}$	$1 \times 10^{-9}$	0.0001
1A6C	$2 \times 10^{-11}$	$4 \times 10^{-10}$	$1 \times 10^{-19}$	$1 \times 10^{-22}$	$3 \times 10^{-30}$
chromatin	3483.09	836570.85	$3 \times 10^{-6}$	$7 \times 10^{-9}$	0.03

mass acceleration (change in velocity), whereas a velocity correction can be applied less frequently since it eliminates center of mass velocity itself. For these reasons, a velocity correction, its drawbacks notwithstanding, may be preferable to the force correction for eliminating the drift and rotation caused by the violation of Newton's third law. Very preliminary testing suggests that the velocity correction may improve the stability of HCPGB simulations for small structures. For example, the RMS deviation for the one-charge HCP-GB simulation of 1BDD (Figure 8a), is in closer agreement with the reference GB simulation when the velocity correction is used (results not shown).

**3.9. Mitigating the Effect of the Discontinuity at Threshold Boundaries.** During the course of molecular dynamics, atoms may cross threshold boundaries that determine the level of approximation used in the computation of potentials and forces. These discontinuous changes in the level of approximation result in changes in potential that are inconsistent with the forces acting on individual atoms; i.e., the force is not equal to the derivative of the potential with respect to distance and can result in the nonconservation of energy and instability in the simulation.<sup>71</sup> For example, 1000 steps of constant energy simulation for thioredoxin, without Langevin dynamics or surface-area dependent energy (Figure 14), show that energy is not conserved for the cutoff-GB and HCP-GB methods. However, the nonconservation of energy is much larger in the case of the cutoff-GB method. For the cutoff-GB method, the energy contribution of an atom abruptly drops to zero when the atom moves beyond the cutoff boundary, resulting in a larger change in energy compared



**Figure 14.** Total energy for 1000 steps of constant energy MD simulation for thioredoxin. The figure shows that energy is not conserved for the cutoff-GB and to a lesser extent the HCP-GB methods. Connecting lines are shown to guide the eye.

**Table 6. Effect of the Use of the Smoothing Function on Energy Conservation and Discontinuity for 1000 steps of MD for Thioredoxin<sup>a</sup>**

standard deviation (kcal/mol)	reference GB	cutoff GB	1-q HCP-GB	2-q HCP-GB
total energy (no smoothing)	0.0012	2.8213	0.2483	0.2476
(HCP-GB with smoothing)			0.0803	0.0992
(HCP-GB with $h_1 = 18 \text{ \AA}$ )			0.0942	0.1013
$\Delta$ total energy (no smoothing)	0.0007	0.3316	0.0228	0.0203
(HCP-GB with smoothing)			0.0033	0.0025
(HCP-GB with $h_1 = 18 \text{ \AA}$ )			0.0062	0.0051

<sup>a</sup> Degree of energy conservation is measured as the standard deviation in total energy and degree of discontinuity as the standard deviation in change in total energy between consecutive steps of MD simulation. The default level 1 threshold distance,  $h_1$ , for the HCP-GB method is 15 Å. For comparison, we include the reference GB computation, the cutoff-GB method with a 15 Å cutoff distance, and the HCP-GB method with a 18 Å level 1 threshold distance. The small fluctuation in total energy for the reference GB computation is due to the finite integrator time step.

to the HCP-GB method, where the energy contribution of the atom is replaced by an approximation when the component containing the atom moves beyond the threshold boundary.

The discontinuity at threshold and cutoff boundaries can be eliminated by the use of a smoothing function.<sup>72,73</sup> The smoothing function eliminates the discontinuity by gradually switching from one level of approximation to another over a short switching distance. Delle Site<sup>70</sup> has however shown that the smoothing function cannot in general restore the conservation of energy for multiscale methods. To eliminate the discontinuity at threshold boundaries, we adapted the smoothing function described by Loncharich and Brooks.<sup>15</sup> The smoothing function is used to calculate the force  $f(r)$  at a distance  $r$  from the point of interest inside the switching region  $h < r < h + s$ , where  $h$  is the threshold distance and  $s$  is the switching distance, as follows:

$$f(r) = S(r)f_{h+s} + (1 - S(r))f_h \quad (h \leq r \leq h + s) \quad (26)$$

$$S(r) = \frac{(h^2 - r^2)^2(h^2 + 2r^2 - 3(h+s)^2)}{(h^2 - (h+s)^2)^3} \quad (27)$$

where  $f_h$  and  $f_{h+s}$  are the forces due to a component computed by the HCP-GB at  $h$  and  $h + s$ , respectively.

The nonconservation of energy can be measured as the standard deviation in total energy, and the discontinuity in computed energy as the standard deviation in the change in total energy between consecutive steps in the MD simulation. Table 6 summarizes these metrics for the HCP-GB method with and without smoothing, the cutoff-GB method, and the reference GB computation. The table shows that, although the HCP-GB method does not conserve energy, it represents a significant improvement over the cutoff-GB method in that respect. On the other hand, although smoothing does improve energy conservation, it is comparable to extending the level 1 threshold distance to the end of the smoothing region, 18 Å in this case, as measured by the standard deviation in total energy (Table 6). The HCP-GB with smoothing does show less discontinuity than the HCP-GB with the extended threshold distance; however, the difference may not be sufficient to justify the higher computational cost of smoothing compared to simply extending the threshold distance. Note that the smoothing function can also improve the accuracy of the spherical cutoff method and has been studied previously.<sup>15,72,73</sup> The results shown in the preceding subsections do not include the smoothing function for either the cutoff-GB or the HCP-GB methods.

We stress that in the case of multiscale approximations based on pairwise potentials, such as the HCP, exact energy

conservation cannot be achieved due to the violation of Newton's third law.

#### 4. CONCLUSION

Implicit solvent models are routinely used where it is important to sample a large conformation space, such as for protein folding, replica exchange, and docking simulations. However, the implicit solvent model employed most extensively in molecular dynamics—the generalized Born (GB) model—scales poorly as  $\sim n^2$ , where  $n$  is the number of solute atoms, limiting their usefulness for long time-scale simulations or the simulation of large structures. We have presented here an  $\sim n \log n$  implementation of the implicit solvent GB model based on the hierarchical charge partitioning (HCP) approximation previously developed by us. The HCP method uses the natural organization of biomolecular structures to partition the structures into multiple hierarchical levels of components such as atoms, groups (residues), subunits (chains), and complexes. The charge distribution for each of these components other than the atoms are approximated by a small (one or two) number of charges. For the computation of electrostatic interactions with distant components, the HCP uses the approximate charges while using the atomic charges for nearby components. The greater the distance from the point of interest, the larger (higher level) is the component used in the approximation. We have previously described a top-down algorithm for HCP that scales as  $\sim n \log n$  for biomolecular structures.

This study extends the HCP approximation to the GB model (HCP-GB) such that both the computation of pairwise interactions and the effective Born radii scale as  $\sim n \log n$ .

The HCP-GB method is implemented in the open source molecular dynamics software, NAB, in AmberTools v1.3. The HCP implementation in NAB is scheduled to be released with AmberTools v1.5, for general use. The accuracy, speed, and stability of the method were then evaluated on a set of representative biomolecular structures ranging in size from 632 to  $\sim 3$  million atoms. The performance of the HCP-GB method was compared to the spherical cutoff method with GB (cutoff-GB) where all computations, including the computation of the effective Born radii, ignore all atoms beyond a specified cutoff distance. Our results show that the HCP-GB method is more accurate, as measured by the relative RMS error in electrostatic force, than the cutoff-GB method for the structures tested. Depending on the size of the structure, the HCP-GB method was also 1.1 to 390 times faster than the reference GB computation. An analysis of 50 ns simulations of four structures—B-DNA, immunoglobulin binding domain, ubiquitin, and thioredoxin—shows that the results for the HCP-GB simulation are in reasonable agreement with the reference GB simulation without cutoffs. For the very small (726 atom) immunoglobulin binding domain protein (1BDD), the one-charge HCP-GB method exhibited RMS deviations from the crystal structure similar to the cutoff-GB and larger than the reference-GB and two-charge HCP-GB simulations. Therefore, we do not recommend the use of one-charge HCP-GB for the simulation of such small structures. However, very preliminary testing suggests that the velocity correction, described below, may improve the stability of HCP-GB simulations for small structures.

There is also an important qualitative difference between the HCP-GB method and the cutoff-GB method. The cutoff-GB ignores charges beyond the cutoff distance while the HCP-GB

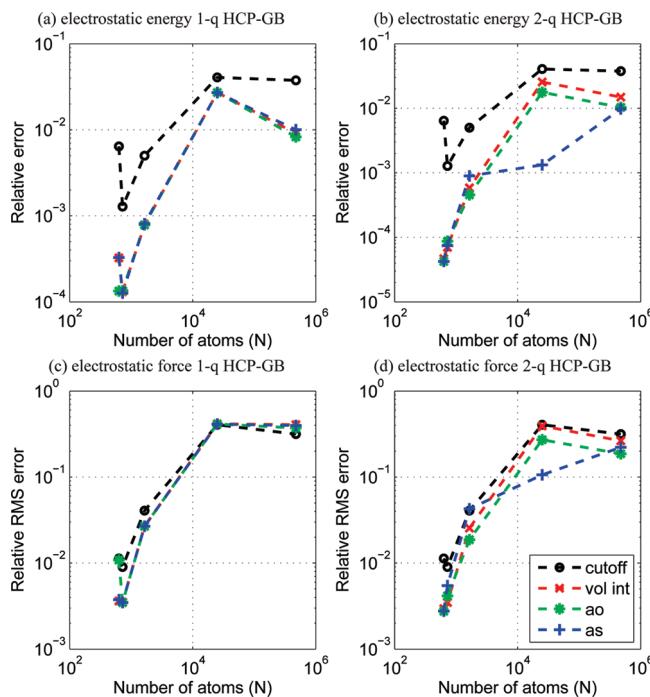
method approximates the influence of distance charges. Our testing suggests that this difference can have a significant impact on details of the dynamics. For example, for the 50 ns simulations of four structures, the residue flexibility and  $\chi_1$  and  $\chi_2$  angles for the cutoff-GB simulations show larger deviations from the reference GB simulation than the HCP-GB simulations. Similarly, a 10 ns simulation of a 30 base-pair DNA strand showed that the flexibility of the molecule, as measured by end-to-end distance, using the HCP-GB method was similar to that of the reference GB simulation, whereas the cutoff-GB method results showed a more flexible molecule. And a series of simulations of the nucleosome core particle showed that with the reference GB and HCP-GB methods all of the positively charged tails of the histone chains collapsed onto the negatively charged DNA, whereas two of the histone tails failed to do so with the cutoff-GB method.

Due to its multiscale nature, the HCP-GB method can violate Newton's third law, resulting in a residual center of mass force and torque. For the structures tested here, the effect of the residual force and torque is much smaller than the "noise" due to stochastic collisions used in constant temperature simulations with strong coupling to a thermal bath. However, when a weak coupling is used to increase the sampling of conformational space, the residual force and torque may cause the structure to drift and rotate, making it inconvenient for visualization and analysis. For simulations with weak coupling to a thermal bath, the center of mass motion and rotation can be eliminated by using a velocity correction. The multiscale nature of HCP-GB can also result in discontinuities at threshold boundaries, which can cause energy not to be conserved. The discontinuity and the resultant nonconservation of energy for the HCP-GB method is however much smaller than that of the cutoff-GB method. Smoothing functions can be used to reduce the discontinuities and the nonconservation of energy. However, we found that increasing the threshold distance may be a more effective way of achieving the same result.

To demonstrate a practical application of the HCP-GB method, we used it to refine a 348 000 atom chromatin fiber. The 15 ns all-atom simulation successfully resolved numerous severe steric clashes, significantly improving the quality of the starting structure.

In conclusion, the  $\sim n \log n$  HCP-GB method is always faster than the  $\sim n^2$  reference GB computation without additional approximations. Although the speed of the HCP-GB method is comparable to using a spherical cutoff for GB computations, which also scales as  $\sim n \log n$ , the HCP-GB method on average more closely reproduces key characteristics of the dynamics of the reference GB simulations. Our testing suggests that this may be because the HCP-GB method approximates the influence of distant charges, unlike the cutoff-GB method, which completely ignores them. In general, our findings suggest that compared to the cutoff-GB, the HCP-GB method may always be the preferable approach for speeding up pairwise GB computations for molecular dynamics. Where speed is critical, one can consider using the one-charge HCP-GB instead of the two-charge HCP-GB or reducing threshold distances from the recommended conservative threshold distances.

This study was intended to be a proof-of-concept of a novel method, and a number of potential improvements and optimizations remain to be studied, in particular, further optimization of the placement of approximate charges, comparison of alternate approximations for component effective Born radii, choice of parameters, comparison of velocity vs force correction, and the treatment of very distant components. Most importantly, more extensive testing is required to further define the applicability and limitations of the proposed  $\sim n \log n$  GB method.



**Figure 15.** Comparison of four alternative methods for computing component effective Born radii showing relative error in electrostatic energy for (a) one-charge HCP-GB, (b) two-charge HCP-GB, and relative RMS error in electrostatic force for (c) one-charge HCP-GB and (d) two-charge HCP-GB. The four alternative methods are the spherical cutoff method, the volume integral (vol int) based on eq 10, Anandakrishnan–Onufriev (ao) defined by eq 9, and Archontis–Simonson (as) defined by eq 28. Cutoff distance = 15 Å, HCP threshold distance  $h_1$  = 15 Å,  $h_2$  = 80 Å, and  $h_3$  = 175 Å. Connecting lines are shown to guide the eye.

## A. APPENDIX

**A.1. Component Effective Born Radii.** The component effective Born radii are used to approximate the contribution of distant components to the solvation energy, as described in the Methods section. We examined three different alternatives for this approximation: the approximation defined by eq 9, the approximation defined by Archontis and Simonson,<sup>39</sup> and the volume integral approximation based on eq 10. Archontis and Simonson approximate the Born radius  $B_c$  for a component  $c$  as

$$\frac{1}{B_c} \approx \frac{1}{\sum_{i \in c} q_i^2} \sum_{i \in c} \frac{q_i^2}{B_i} \quad (28)$$

where  $q_i$  and  $B_i$  are the charges and effective Born radii, respectively, for the atoms  $i$  belonging to component  $c$ . On average, we found that the first alternative (eq 9) was most accurate, as shown in Figure 15.

**A.2. Force Corrections for Neutralizing Net Residual Force.** The violation of Newton's third law by the HCP-GB method, as described in section 3.8, results in a net residual force. We considered two approaches for neutralizing the net residual force—a molecular level and a component level force correction.

The molecular level approach applies a mass weighted force correction to each atom to neutralize the total residual force on the whole structure. This force correction is computed as

$$\mathbf{f}_i^{\text{corr}} = \mathbf{f}^{\text{res}} m_i / M \quad (29)$$

where  $\mathbf{f}_i^{\text{corr}}$  is the force correction subtracted from the force on atom  $i$ ,  $\mathbf{f}^{\text{res}}$  is the total residual force,  $m_i$  is the mass of atom  $i$ , and  $M$  is the total mass of the structure.

The component level approach is more complex. It aims to eliminate not only the net residual force but also the net residual force for each component, where the residual force for a component is the difference between the total force on a component due to all other atoms and the total force on all other atoms due to the component. In other words, it aims to restore Newton's third law at the component level. For the first level of HCP approximation, the force correction  $\mathbf{f}_i^{\text{corr}}$  for an atom  $i$  belonging to component  $c$  is calculated as

$$\mathbf{f}_i^{\text{corr}} = \mathbf{f}_c^{\text{diff}} / n_c \quad (i \in c) \quad (30)$$

$$\mathbf{f}_c^{\text{diff}} = [\sum_{j \notin c} \sum_{i \in c} \mathbf{f}_{ji} + \sum_{k \neq c} \sum_{i \in c} \mathbf{f}_{ki}] + [\sum_{i \in c} \sum_{j \notin c} \mathbf{f}_{ij} + \sum_{j \notin c} \sum_{i \in c} \mathbf{f}_{cj}] \quad (31)$$

where  $\mathbf{f}_c^{\text{diff}}$  is the difference between the total force on a component due to all other atoms and the total force on all other atoms due to the component. The first term on the right hand side of eq 31 is the force  $\mathbf{f}_{ji}$  on the atoms  $i$  belonging to component  $c$  due to other atoms  $j$  not belonging to  $c$ , computed at the atomic level (level 0). The second term is the force  $\mathbf{f}_{ki}$  on atoms  $i$  belonging to  $c$  due to other components  $k$  treated at the component level (level 1). The third term is the force  $\mathbf{f}_{ij}$  due to  $c$  on atoms  $j$  not belonging to  $c$  where the atoms  $i$  within  $c$  are treated at the atomic level (level 0). And the last term is the force  $\mathbf{f}_{cj}$  due to  $c$  on atoms  $j$  not belonging to  $c$  where  $c$  is treated at the component level (level 1). This force correction is generalized for higher levels of HCP by including terms in eq 31 for higher level HCP components. Since the individual terms on the right hand side of eq 31 are already being computed, the incremental cost of both force correction approaches scales as  $\sim n$ , which is  $< n \log n$ .

Although the above force corrections neutralize the net force due to the violation of Newton's third law, they also cause an increase in the force error. To see why, consider the case where the force error compared to the reference GB computation is approximately zero. In this case, any net force correction will result in an increase in the force error. In general, when the force error is less than half the net force correction, the net force correction will cause an increase in force error. Thus, on average, where the force error is randomly distributed, the net force correction will result in an increase in force error.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: alexey@cs.vt.edu.

## ■ ADDITIONAL NOTE

<sup>a</sup>NAB, and the production Amber software, do not explicitly include a “no-cutoff” option; instead, the no-cutoff computation is performed by using the cutoff method with the cutoff distance set to a value greater than the structure size (large-cutoff), e.g., 999. This approach requires a large amount of memory for the neighbor list used by the cutoff method, even though the list is unnecessary in this case, since it always contains all of the atoms. Due to the large memory requirement, structures larger than 200 000 atoms require a larger number of processors (>128)

than was readily available in the system used. Therefore, we implemented a no-cutoff option in NAB that does not use a neighbor list. Since a neighbor list does not need to be computed, the no-cutoff option is faster than the large-cutoff approach, and the speedup results reported here are somewhat lower than what would have been obtained using the large-cutoff approach available in NAB.

<sup>b</sup> The average speedup for the seven structures tested here was 82× for the cutoff-GB, 85× for the one-charge HCP-GB, and 36× for the two-charge HCP-GB methods.

## ■ REFERENCES

- (1) Dodson, G. G.; Lane, D. P.; Verma, C. S. *EMBO Rep.* **2008**, *9*, 144–150.
- (2) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. *Structure* **2006**, *14*, 437–449.
- (3) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (4) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6679–6685.
- (5) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211–243.
- (6) Venkateswarlu, D. *BMC Struct. Biol.* **2010**, *10*, 7.
- (7) Kumar, S.; Huang, C.; Zheng, G.; Bohm, E.; Bhatele, A.; Phillips, J. C.; Yu, H.; Kalé, L. V. *IBM J. Res. Dev.* **2008**, *52*, 177–187.
- (8) Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 9204–9209.
- (9) Shaw, D. E.; et al. *Comm. ACM* **2008**, *51*, 91–97.
- (10) Zhou, R.; Eleftheriou, M.; Hon, C. C.; Germain, R. S.; Royyuru, A. K.; Berne, B. J. *IBM J. Res. Dev.* **2008**, *52*, 19.
- (11) Ejtehadi, M. R.; Avall, S. P.; Plotkin, S. S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15088–15093.
- (12) Beck, D. A. C.; Armen, R. S.; Daggett, V. *Biochemistry* **2005**, *44*, 609–616.
- (13) Ruvinsky, A. M.; Vakser, I. A. *Proteins* **2008**, *70*, 1498–1505.
- (14) Mark, P.; Nilsson, L. *J. Comput. Chem.* **2002**, *23*, 1211–1219.
- (15) Loncharich, R. J.; Brooks, B. R. *Proteins* **1989**, *6*, 32–45.
- (16) Schreiber, H.; Steinhauser, O. *Chem. Phys.* **1992**, *168*, 75–89.
- (17) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (18) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (19) Toukmaji, A. Y.; Board, J. A. *Comput. Phys. Commun.* **1996**, *95*, 73–92.
- (20) York, D.; Yang, W. *J. Chem. Phys.* **1994**, *101*, 3298–3300.
- (21) Carrier, J.; Greengard, L.; Rokhlin, V. *SIAM J. Sci. Stat. Comp.* **1988**, *9*, 669–686.
- (22) Cai, W.; Deng, S.; Jacobs, D. *J. Chem. Phys.* **2007**, *223*, 846–864.
- (23) Lambert, C. G.; Darden, T. A.; Board, J. A., Jr. *J. Chem. Phys.* **1996**, *126*, 274–285.
- (24) Bishop, T. C.; Skeel, R. D.; Schulten, K. *J. Comput. Chem.* **1997**, *18*, 1785–1791.
- (25) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (26) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (27) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (28) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824–19836.
- (29) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.
- (30) Lee, M. S.; Salsbury, J.; Brooks, C. L., III. *J. Chem. Phys.* **2002**, *116*, 10606–10614.
- (31) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.
- (32) Tsui, V.; Case, D. J. *Am. Chem. Soc.* **2000**, *122*, 2489–2498.
- (33) Cramer, C.; Truhlar, D. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (34) David, L.; Luo, R.; Gilson, M. K. *J. Comput. Chem.* **2000**, *21*, 295–309.
- (35) Im, W.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.
- (36) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578–1599.
- (37) Calmet, N.; Schaefer, M.; Simonson, T. *Proteins* **2001**, *45*, 144–158.
- (38) Feig, M.; Im, W.; Brooks, C. L. *J. Chem. Phys.* **2004**, *120*, 903–911.
- (39) Archontis, G.; Simonson, T. *J. Phys. Chem. B* **2005**, *109*, 22667–22673.
- (40) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.
- (41) Nymeyer, H.; Garcia, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13934–13939.
- (42) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. A* **1997**, *101*, 8098–8106.
- (43) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765–3773.
- (44) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479–499.
- (45) Grant, J. A.; Pickup, B. T.; Sykes, M. J.; Kitchen, C. A.; Nicholls, A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4913–4922.
- (46) Haberthür, U.; Caflisch, A. *J. Comput. Chem.* **2007**, *29*, 701–715.
- (47) Spassov, V. Z.; Yan, L.; Szalma, S. *J. Phys. Chem. B* **2002**, *106*, 8726–8738.
- (48) Ulmschneider, M. B.; Ulmschneider, J. P.; Sansom, M. S.; Di Nola, A. *Biophys. J.* **2007**, *92*, 2338–2349.
- (49) Tanizaki, S.; Feig, M. *J. Chem. Phys.* **2005**, *122*, 124706.
- (50) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591–607.
- (51) Shen My, M. Y.; Freed, K. F. *Biophys. J.* **2002**, *82*, 1791–1808.
- (52) Rhee, Y. M.; Pande, V. S. *J. Phys. Chem. B* **2008**, *112*, 6221–6227.
- (53) Onufriev, A. V. In *Modeling Solvent Environments*; Feig, M., Ed.; Wiley-VCH: Weinheim, Germany, 2010; Chapter 6. Continuum Electrostatic Solvent Modeling with the Generalized Born Model, pp 127–165.
- (54) Zagrovic, B.; Pande, V. *J. Comput. Chem.* **2003**, *24*, 1432–1436.
- (55) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7587–7592.
- (56) Guimaraes, C. R.; Cardozo, M. *J. Chem. Inf. Model.* **2008**, *48*, 958–970.
- (57) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275–291.
- (58) Anandakrishnan, R.; Onufriev, A. V. *J. Comput. Chem.* **2010**, *31*, 691–706.
- (59) Wong, H.; Victor, J.-M.; Mozziconacci, J. *PLOS One* **2007**, *436*, e877.
- (60) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (61) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.
- (62) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (63) Macke, T.; Case, D. In *Molecular Modeling of Nucleic Acids*; Leontes, N. B., SantaLucia, J., Eds.; Americal Chemical Society: Washington, DC, 1998; Chapter: Modeling Unusual Nucleic Acid Structures, pp 379–393.
- (64) Gordon, J.; Myers, J.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. *Nucleic Acids Res.* **2005**, *33*, 68–71.
- (65) Onufriev, A.; Bashford, D.; Case, D. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (66) Wang, H.-W. W.; Nogales, E. *Nature* **2005**, *435*, 911–915.

- (67) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
- (68) Ruscio, J. Z.; Onufriev, A. *Biophys. J.* **2006**, *91*, 4121–4132.
- (69) Bertin, A.; Leforestier, A.; Durand, D.; Livolant, F. *Biochemistry* **2004**, *43*, 4773–4780.
- (70) Delle Site, L. *Phys. Rev. E* **2007**, *76*, 047701.
- (71) Ensing, B.; Nielsen, S. O.; Moore, P. B.; Klein, M. L.; Parrinello, M. *J. Chem. Theory Comput.* **2007**, *3*, 1100–1105.
- (72) Leach, A. In *Molecular Modeling: Principles and Applications*; Prentice Hall: New York, 2001; Chapter 6. Computer Simulation Methods, pp 303–352.
- (73) Schlick, T. *Molecular Modeling and Simulation, an Interdisciplinary Guide*; Springer-Verlag: New York, 2002.