# Atomic Radii for Continuum Electrostatics Calculations on Nucleic Acids

**Nilesh K. Banavali and Benoıt Roux***

*Department of Biochemistry and Structural Biology, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021*

*Received: March 28, 2002; In Final Form: August 9, 2002*

The accuracy of continuum electrostatic models, in which the solvent is represent as a featureless dielectric medium, depends sensitively on the choice of the atomic radii used for setting the dielectric boundary between the solute and the solvent. Here, a set of optimal atomic radii for performing accurate continuum electrostatic calculations on naturally occurring nucleic acids is parametrized using molecular dynamics simulations and free energy perturbation calculations with explicit water molecules. Small model compounds constituting the building blocks of nucleic acids (base, sugar, phosphate) as well as combinations of these units (nucleoside, diphosphate sugar) are chosen to get proper representation of all the molecular moieties found in nucleic acids. Different conformations of the model compounds representative of both A-form and B-form conformations of DNA are considered. An initial set of atomic radii is first determined using a statistical mechanical analysis based on the average radial solvent charge density calculated from molecular dynamics simulations with explicit water molecules. Minor adjustments are then made to refine the radii to reproduce quantitatively the electrostatic contribution to the solvation free energy calculated by free energy perturbation techniques. The ability of the continuum dielectric approximation to accurately represent the free energy associated with base pairing is also considered by examining all possible normal and mismatched base pairs. In all cases, the agreement between molecular dynamics free energy perturbations with explicit water molecules and the continuum electrostatic calculations with optimized atomic radii is excellent. As a final illustration, the free energy profile is calculated for AT and GC base pairs as a function of base pair separation. Continuum electrostatic calculations with the optimized atomic radii provides a computationally inexpensive approach to study nucleic acids and their complexes with proteins.

## I. Introduction

The structural and biophysical properties of nucleic acids and their complexes with proteins are greatly dependent on the nature of the solvent environment.[1,2] Because of the repeated occurrence of the negatively charged phosphate moiety in their backbone the interaction of nucleic acids with the surrounding solvent is dominated by electrostatic effects. One of the most ambitious and detailed approaches for studying the effect of solvation on the conformation of biological macromolecules is to carry out all-atom computer simulations in which a large number of solvent molecules are included explicitly.[3] Recent advances in nucleic acid force fields[4−8] and methodologies to treat long-range interactions such as particle mesh Ewald (PME)[9] have made it possible to perform accurate and stable large-scale molecular dynamics simulations on nucleic acids.[10−12] All-atom simulations can be used in conjunction with thermodynamic perturbation (TP) or thermodynamic integration (TI) methods to calculate solvation free energies of molecules.[13] Such calculations are, however, computationally expensive and often exhibit statistical uncertainties caused by limited sampling.[14] In addition, further difficulties arise in free energy calculations in the case of processes involving a change in the total charge when the long-range electrostatic interactions are truncated or treated through Ewald summations.[15−18]

An alternative approach is incorporating the influence of the solvent implicitly on the basis on macroscopic continuum electrostatics.[19−22] This approximation, in which the polar solvent is represented as a structureless continuum dielectric medium, was originally pioneered by Born in 1920 to calculate the hydration free energy of spherical ions.[23] According to this model, the electrostatic free energy is dependent on the ionic radius $R_{ion}$, the dielectric constant of the solvent $\epsilon_{bulk}$, and the ionic charge $Q_{ion}$:

$$\Delta G_{elec} = \frac{{Q_{ion}}^2}{2R_{ion}}\left(\frac{1}{\epsilon_{bulk}} - 1\right) \qquad (1)$$

With the use of present age computers, the solvation properties of solutes of arbitrary shape can be studied routinely by solving the Poisson−Boltzmann (PB) equation numerically with finite-difference algorithms,[24]

$$\nabla\cdot[\epsilon(\mathbf{r})\,\nabla\phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r})\,\phi(\mathbf{r}) = -4\pi\rho^{(solute)}(\mathbf{r}) \qquad (2)$$

where $\phi(\mathbf{r})$ is the total electrostatic potential, $\epsilon(\mathbf{r})$ is the position-dependent dielectric constant, $\bar{\kappa}^2(\mathbf{r})$ is the position-dependent ionic screening factor, and $\rho^{solute}(\mathbf{r})$ is the charge density due to the solute. One attractive feature of the continuum electrostatic approximation is that it is computationally much less expensive than all-atom simulations. For example, important quantities such as the solvation free energy of a macromolecular system can be readily evaluated.[25] Furthermore, approaches exploiting the combined stengths of both all-atom simulations and continuum electrostatic calculations have recently become increasingly useful.[26−28] For example, it is possible to estimate the

* Corresponding author. Tel: 212-746-6018. Fax: 212-746-4843. E-mail: Benoit.Roux@med.cornell.edu. Website: http://thallium.med.cornell.edu/RouxLab/.

10.1021/jp025852v CCC: $22.00 
Published on Web 10/01/2002

binding affinity of ligands to macromolecules by averaging a series of continuum electrostatic calculations over a large number of configurations generated by simulating the system with explicit solvent molecules.[27] Despite their approximate nature, continuum dielectric continuum models are thus exceedingly useful in studies of biomolecular systems.

The electrostatic free energy calculated on the basis of a continuum electrostatic model depends sensitively on the location of the solute−solvent dielectric boundary. This boundary is generally constructed by assigning specific "Born-like" radii to all the solute atoms. In trying to develop a meaningful continuum electrostatic model for biomolecules, it is important to understand the significance of the solute−solvent dielectric boundary at the atomic level. Theoretical studies aimed at clarifying the microscopic basis of the Born model shed some light on the simplest case of a monatomic ion.[29−31] In particular, it has been shown that the radius $R_{ion}$ as it appears in eq 1 is related to the first peak in the average radial charge distribution function from the solvent around the ion.[31,32] The optimal value of the Born radius for an atom is therefore dependent on the characteristics of both the individual atom and the solvent molecules.[31] For example, a different behavior is expected in the case of cations and anions of similar sizes because of the markedly asymmetric character of the charge distribution of the water molecule. Generalization of these concepts to biomolecular solutes to develop a set of atomic Born radii for biomolecular solutes is, however, not straighforward. Various strategies for the parametrization of a set of radii aimed at reproducing experimental solvation free energies have been proposed.[14,30,33,34] Such approaches are, nonetheless, limited due to scarcity of experimental data for biomolecular systems. In addition, it is generally difficult (if not impossible) to unambiguously separate the experimental solvation free energies into nonpolar and electrostatic contributions. In this context, molecular dynamics simulations and free energy calculations with explicit solvent molecules arguably provide the most complete source of information for characterizing the solvation free energy of biomolecules. Previously, we have used this strategy for determining an optimal set of atomic radii for continuum electrostatic calculations for all twenty amino acids.[32] Rough estimates are first obtained for the Born radii using the average radial solvent charge density calculated from explicit solvent molecular dynamics simulations; minor adjustments are then made to improve the accuracy of the calculated electrostatic solvation free energies. Our goal with the current effort is to obtain a set of atomic Born radii for nucleic acids that is consistent and compatible with the atomic radii for proteins.

In section II, the relationship between the average solvent charge distribution obtained from explicit solvent molecular dynamics calculations and the Born radii used in continuum electrostatics calculations is briefly reviewed. In section III, the model compounds used to represent nucleic acids and their different conformations studied are described along with the computational details of the radial charge distribution, free energy perturbation (FEP) and continuum electrostatics calculations. In section IV, the results for the determination and validation of the atomic radii of nucleic acids are discussed, followed by a brief summary and conclusions in section V.

## II. Theory

A nonpolarizable solute molecule immersed in a fixed configuration in uniform bulk solvent is considered. By introducing a scaling parameter, $\lambda$, we can express the electrostatic contribution to the solvation free energy as the total reversible work for charging the solute ($\lambda = 0$ and $\lambda = 1$ corresponds to uncharged and fully charged state of the solute)[35]

$$\Delta G_{elec} = \int_0^1 d\lambda \left\langle \sum_{i,j} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \right\rangle_{(\lambda)} \quad (3)$$

where $q_i$ and $q_j$ are the charges corresponding to solute atom $i$ and solvent atom $j$, respectively, and $\mathbf{r}_i$ and $\mathbf{r}_j$ are their respective positions. The symbol $\langle ... \rangle_{(\lambda)}$ represents an average over all coordinates of the solvent, for a fixed solute with charges scaled by $\lambda$. The electrostatic free energy change or the charging free energy of the solute can be related to the average solvent reaction field at each solute atom $i$ as follows:

$$\Delta G_{elec} = \int_0^1 d\lambda \sum_i q_i \langle \phi_{rf}(\mathbf{r}_i) \rangle_{(\lambda)} \quad (4)$$

where $\phi_{rf}(\mathbf{r}_i)\rangle_{(\lambda)}$ is the average solvent reaction field acting at the position of the $i$th solute atom,

$$\langle \phi_{rf}(\mathbf{r}_i) \rangle_{(\lambda)} = \int d\mathbf{r} \frac{1}{|\mathbf{r}_i - \mathbf{r}|} \langle \rho_{elec}(\mathbf{r}) \rangle_{(\lambda)} \quad (5)$$

where $\langle \rho_{elec}(\mathbf{r}) \rangle_{(\lambda)}$ is the average solvent charge density. Even though the integration is taken over the volume of the entire system, the average solvent charge density that gives rise to the reaction field is nonzero only within a finite region in the neighborhood of the solute−solvent boundary. It decays rapidly to zero away from the solute, at a distance that is large compared to the solvent correlation length.[32] When the solvent is approximated by a continuum dielectric medium, the solvent charge density $\langle \rho_{elec}(\mathbf{r}) \rangle_{(\lambda)}$ corresponds to an infinitesimal layer at the solute−solvent dielectric boundary and eq 5 is reduced to a surface integral. An optimal set of atomic Born-like radii, used to set the position of the solute−solvent dielectric boundary, should therefore be chosen to yield accurate estimates of the charging free energy of the solute while retaining a meaningful correspondence to the microscopic structure of the average solvent charge density as much as possible. In practice, it is expected that some empirical adjustments in the value of the radii (much smaller than the correlation length of the solvent) will be needed to yield a final model of optimal accuracy because the magnitude of the solvent reaction field is highly sensitive to the exact position of the solute−solvent dielectric boundary.

## III. Computational Methods

**(a) Model Compounds.** The model compounds used to represent nucleic acids in the present study were divided into two categories, those representing the basic components of nucleic acids and those representing specific combinations of two or more of these basic components. The first category consisted of the five base analogues, deoxyribose and ribose sugars, and the phosphate moiety; the second category consisted of the nucleoside analogues (base + sugar) and the sugar−diphosphate (sugar + 5′ and 3′ phosphates) moiety. All these model compounds are illustrated in Figure 1. The conformations of the model compounds were obtained by harmonically constraining the individual torsion angles to their modal values from crystal structure distributions of A- and B-form DNA structures[7] followed by 200 steps of minimization using the adopted basis Newton Raphson (ABNR)[36] method of the program CHARMM.[37] The all-atom CHARMM27 force field for nucleic acids[7,8] was used in all the calculations. The torsion angles used for the A-form and B-form DNA backbone are

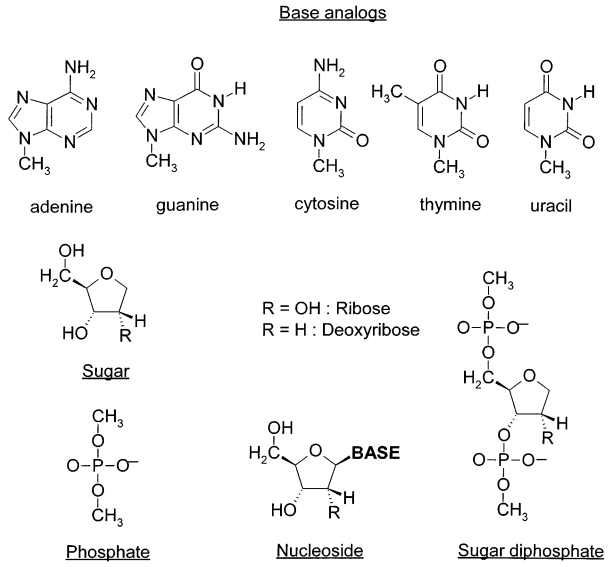


**Figure 1.** Model compounds representative of nucleic acids used to obtain electrostatic solvation free energies using explicit solvent FEP calculations. Base analogues have an extra methyl group representing the C1′ atom in sugar. Basic units are the 5 bases, ribo and deoxyribo furanose sugars, and dimethyl phosphate. Combinations of these basic units yield larger model compounds such as nucleosides (base + sugar) and sugar−diphosphate (sugar + 5′ and 3′ phosphates)

**TABLE 1: Modal Values of Specific Nucleic Acid Dihedrals Obtained from DNA Crystal Structure Distributions[7] Used To Generate the A-Form and B-Form Local Structures of Model Compounds**

| DNA form | sugar conformation | α | *β* | *γ* | *ϵ* | *ζ* |
|---|---|---|---|---|---|---|
| A-DNA | C3′-endo | 291 | 175 | 57 | 205 | 287 |
| B-DNA | C2′-endo | 298 | 168 | 51 | 187 | 262 |

shown in Table 1. For all model compounds that contain a sugar moiety there were two final fixed conformations, one corresponding to the A-form and the other corresponding to the B-form. The sugar type and sugar conformation are denoted in short by the following convention: the first letter represents the sugar type with "d" denoting deoxyribose and "r" denoting ribose; the number following represents the sugar conformation with "2" representing C2′-endo conformation and "3" representing C3′-endo conformation. For the phosphate moiety, both experimental[7] and theoretical[38,39] data show that three minimum energy conformations are possible: the gauche,gauche (gg), gauche,trans (gt), and trans,trans (tt) conformations. These three conformations were studied separately for the phosphate moiety whereas only the global minimum gg conformation was used for the sugar−diphosphate model compound.

**(b) FEP Calculations.** The electrostatic solvation free energy of the model compounds was calculated using the PERT module of CHARMM.[37] Each model compound was solvated in a 12 Å radius sphere of explicit TIP3P water molecules,[40] sufficient to provide approximately 3−4 hydration shells around each solute atom. The spherical solvent boundary potential (SSBP)[41] including Kirkwood's multipolar expansion reaction field[42] was used to approximate the influence of bulk water beyond the explicit water sphere. All of the nonbonded interactions between the explicit atoms in the simulation region were included (no truncation or cutoff scheme was used). For each model compound, the solvent molecules were first equilibrated for 100 ps. All trajectories were generated with Langevin dynamics simulation with a relatively weak friction constant corresponding to a relaxation time of 5 ps applied to the water oxygens to ensure that the configurations were sampled from a Boltzman distribution at 300 K. The water geometry was kept fixed using SHAKE,[43] and the integration time step was 0.002 ps. All further calculations were performed on these preequilibrated systems. After the initial equilibration of 100 ps, the perturbations were generated in the forward and backward directions with values of the thermodynamic coupling parameter $\lambda$ varying by $\pm 0.1$ between 0 and 1 for a total of 20 windows per FEP calculation. The trajectory for each window was started from the last configuration of the previous window with 10 ps of equilibration followed by 30 ps of Langevin dynamics. The charging calculation for each compound includes a total of 880 ps, corresponding to about 60 CPU hours on a 700 MHz Pentium. The results from the individual windows were combined using the weighted histogram analysis method (WHAM).[32,44,45] WHAM generates free energy constants $F_k$ for each window using a self-consistent iterative procedure based on the equation

$$\exp\left[\frac{-F_k}{k_B T}\right] = \sum_{i=1}^{N_w}\sum_{t=1}^{T_i} \exp\left[\frac{-\lambda_k \Delta U_i(t)}{k_B T}\right]\left[\sum_{j=1}^{N_w} T_j \exp\left[\frac{F_j - \lambda_j \Delta U_i(t)}{k_B T}\right]\right]^{-1} \quad (6)$$

where $N_w$ is the number of windows, $T_i$ is the simulation length of the $i$th window, $\lambda_k$ is the coupling parameter for the $k$th window, and $\Delta U_i(t)$ is the potential energy difference ($U_1 - U_0$) for time $t$ from the $i$th window. In practice, the WHAM equations are iterated until self-consistency of the constant $F_k$ is achieved. The electrostatic contribution to the solvation free energy of the solute is given by $F_1 - F_0$.

**(c) Continuum Electrostatic Calculations.** The electrostatic solvation free energies were calculated by solving the PB equation (eq 2) using a finite difference algorithm.[46] No salt concentration was included and $\bar{\kappa}^2(\mathbf{r})$ was set to zero. In this case, eq 2 is reduced to the Poisson equation. The dielectric constant of the solvent was set to 80. The dielectric constant of the solute was set to 1 to be consistent with the nonpolarizable force field and fixed solute conformation. The solvent reaction field $\phi_{rf}$ required for computing the charging free energy was obtained from the difference between the electrostatic potential computed in solvent ($\epsilon_{bulk} = 80$) and in a vacuum ($\epsilon_{bulk} = 1$). Two different methods were used for setting the dielectric boundary used in the Poisson calculations: as the surface formed by overlapping solvent-excluding atomic spheres,[32] and as the solvent-accessible molecular surface formed.[47,48] The latter method corresponds to the solvent-inaccessible volume,[49] which includes the reentrants of the molecular surface. In particular, the small voids between solute atoms that are not accessible to a solvent sphere are assimilated to the low dielectric solute region. A three-dimensional grid extending at least 6 Å from any solute atom was used; results from three grid spacings of 0.1, 0.2, and 0.5 Å were examined. All continuum electrostatic calculations were performed using the finite-difference algorithm[24,46] implemented in the PBEQ[50] module of CHARMM.[37]

## IV. Results and Discussion

**(a) Radial Charge Distributions and Atomic Radii.** The radial charge distributions for individual non-hydrogen atoms were determined from 100 ps Langevin dynamics trajectories for all the model compounds shown in Figure 1. The initial value of the atomic radii is based on the smallest distance the individual solute atoms and the average solvent charge density.[32] Variations in the initial values are observed for a given atom

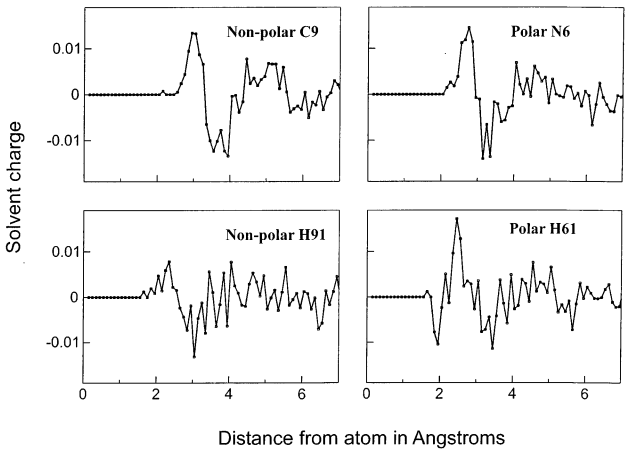


**Figure 2.** Representative solvent charge radial distribution functions for polar and nonpolar atoms and their attached hydrogen atoms in 100 ps explicit solvent Langevin simulation of adenine base analogue. The polar hydrogen atom shown is one of the hydrogen atoms attached to the N6 atom in adenine; the nonpolar hydrogen atom shown is one of the hydrogen atoms in the methyl group attached to the N9 atom in adenine

type in the different model compounds as well as different conformations of the same model compound, reflecting the fact that the Born atomic radius assigned to an atom depends on the partial charge of the atom as well as its structural environment. Atoms with the same atom type (i.e., same Lennard-Jones parameters) can have different assigned partial charges depending on the structural context. For example, the N7 atoms of the adenine base and the guanine base both have the atom type NN4 in the CHARMM force field but differ in partial charge by 0.11 units (−0.60 vs −0.71, respectively). The van der Waals radius assigned to the atom thus does not necessarily correspond to the radial solvent charge distribution characteristics of the atom and therefore cannot be used directly as the Born radius of the atom.

Because variations in the solute−solvent boundary occur as a result of conformation and neighboring groups, the final values of the Born radii must be empirically optimized to get best possible agreement with all model compounds and conformations studied. A natural consequence of including a large goal data set is that compromises have to be made in trying to reproduce the free energy of individual model compounds or conformations. The final values of the Born atomic radii were therefore adjusted empirically to quantitatively reproduce, as best possible, the electrostatic free energy for different relevant conformations.

An interesting problem was encountered when we tried to assign Born radii for hydrogen atoms. A few characteristic solvent charge distributions for the hydrogen atoms are shown in Figure 2. The radial solvent charge distributions for the atoms are dependent on the partial charge and structural environment. The polar hydrogen atom attached to the electronegative N6 atom therefore shows greater proximity to negative solvent charge, whereas the nonpolar hydrogen atom attached to the methyl carbon does not show any preference for negative solvent charges. In all cases, however, the minimum distance between the hydrogen atoms and the solvent atomic charges is much greater than 1 Å. Upon further consideration, this is not surprising. The partial atomic charges assigned to the hydrogen atoms are mostly positive, and their van der Waals radii mostly have small values close to 1 Å. Because the atomic radii of heavy atoms to which these hydrogens are attached tend to be larger (around 2 Å), and the bond length between the heavy

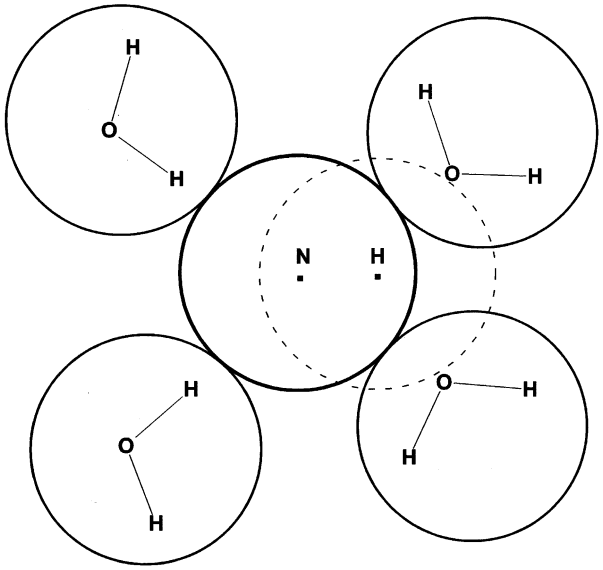


**Figure 3.** Representation of the ideal solute−solvent boundary for the N−H group. Reorientation of explicit water molecules and the larger radius of water oxygen atoms carrying negative partial charges causes the water oxygen atoms carrying the negative solvent charge to be located at a finite distance (approximately 2 Å) from the hydrogen atom, resulting in a larger Born radii for the hydrogen atom indicated by the dotted sphere. The overall solvent charge distribution around the entire N−H group is located around the overlapping boundary of the nitrogen atomic sphere (bold) and the apparent hydrogen atomic sphere (dotted). The solute−solvent boundary for the same N−H group is, however, spherical if the Born radius of hydrogen is set to 0.0 Å.

atom and the hydrogen is usually around 1 Å, any change in the hydrogen radius of up to 1 Å does not change the solute−solvent boundary significantly. The choice of the Born radius of each atom in the present study is, however, based on the radial charge distribution around that atom. The larger radius of the solvent oxygen atoms carrying negative partial charges prevents them from approaching the hydrogen atoms with greater proximity. This fact is responsible for the position of the solvent charge rather than the hydrogen atom radius itself. Ultimately, each hydrogen atom might have a different Born radius depending on its charge and structural context. This would complicate the assignment of the radii by greatly increasing the number of individual radii to be assigned. A schematic description of this phenomenon is shown in Figure 3 where the oxygen atoms in the solvent approach the polar hydrogens and the hydrogen atoms approach the attached electronegative nitrogen. A simplification is to assign one value for all hydrogen radii. Apparently, setting the hydrogen radius to 0.0 Å does not allow a very accurate representation of the solute−solvent dielectric boundary. However, in an earlier study for determining Born radii for protein atoms, accurate results were obtained by assigning a radius of 0.0 Å to the hydrogens.[32] It is possible that these considerations are more important in the case of nucleic acid structures due to abundance of polar hydrogens. The ulimate test for the radii, however, is not exact reproduction of the solute−solvent boundary but good reproduction of electrostatic free energies. It may be possible to achieve the latter goal even with the simplification of zero hydrogen radius. To test whether using the approximation of a zero hydrogen radius is valid, two values of the hydrogen radius, 1.5 and 0.0 Å, were tested in conjunction with the continuum model of solvent.

**(b) Electrostatic Solvation Free Energies with Explicit Solvent.** A set of nucleic acid Born radii could, in principle, be

**TABLE 2: Electrostatic Contribution to the Solvation Free Energy of Basic Units Calculated Using Explicit and Continuum Models of Solvent with a Reentrant Molecular Surface**

| basic unit | name | FEP | continuum[a] | continuum[b] |
|---|---|---|---|---|
| base | Ade | −12.9 | −11.9 | −12.6 |
| | Cyt | −17.1 | −16.5 | −17.4 |
| | Gua | −21.4 | −20.5 | −21.7 |
| | Thy | −9.7 | −9.1 | −9.5 |
| | Ura | −10.8 | −10.4 | −10.4 |
| sugar | d2 | −15.1 | −15.7 | −13.8 |
| | d3 | −14.9 | −15.8 | −14.1 |
| | r2 | −20.5 | −20.3 | −19.8 |
| | r3 | −18.1 | −19.3 | −17.8 |
| phosphate | gg | −72.4 | −71.9 | −71.7 |
| | gt | −72.3 | −71.7 | −71.6 |
| | tt | −72.3 | −72.1 | −71.9 |

[a] Set of radii for hydrogen radius 1.5 Å. [b] Set of radii for hydrogen radius 0.0 Å, energy values in kcal/mol units.

obtained by considering only the small model compounds (base, sugar, and phosphate). The problem with this approach, which is common to most parametrization strategies, is that accuracy in reproducing the solvation free energies of the smaller model compounds does not ensure the ability to accurately treat combinations of these units into larger model compounds or macromolecules. Because the primary goal of developing these radii is to be able to study macromolecular properties in a computationally efficient manner, it is essential that the set of radii properly represent combinations of the basic units. Nucleic acids are not rigid molecules and their primary role of propagation of information requires large changes in conformation during replication and transcription.[51] Nucleic acids bound to proteins, in some cases, show significantly different conformations than the canonical B-DNA structures.[52] Large global changes in DNA structure can be achieved by a series of local minimum energy conformation changes, as seen in B-form to A-form transformation primarily being associated with sugar pucker changes.[53] In the present study, different local minimum energy conformations of the sugar and phosphate moieties were considered to ensure that all forms of nucleic acids were represented.

The solvation free energy contributions for the basic units are shown in Table 2. The calculated values for all FEP calculations obtained with 15 ps sampling per window and those obtained with 30 ps sampling per window did not differ significantly, indicating convergence of the results. The average error in the FEP calculations is estimated to be about 0.5 kcal/mol. Among the bases, the largest solvation free energy is seen for the guanine base (−21.4 kcal/mol) and the smallest is seen for the thymine base (−9.7 kcal/mol). Because the bases are relatively rigid and planar in conformation, it was not necessary to consider multiple conformations. The values obtained here, despite not including nonpolar contributions to the solvation free energy are remarkably consistent with both previously calculated and experimental solvation free energies for the base analogues.[54] The sugar moiety is, however, quite flexible and can exist in two prominently populated conformations: the C2′-endo (south) conformation characteristic of the B-form and the C3′-endo (north) conformation characteristic of the A-form.[55] Both "2" and "3" conformations were studied for the deoxyribofuranose (d) and ribofuranose (r) sugar moieties (by convention, the various cases are labeled with the first letter representing the form of the sugar and the following number representing the sugar pucker). The C2′-endo form of the ribofuranose sugar (r2) shows the largest solvation free energy contribution (−20.5 kcal/mol) and the C3′-endo conformation of the deoxyribofuranose sugar (d3) shows the smallest solvation free energy contribution (−14.9 kcal/mol). The conformational dependence of the solvation free energy is clearly illustrated by the difference of 2.4 kcal/mol between the values of −20.5 and −18.1 kcal/mol for r2 and r3, respectively. Expectedly, the negatively charged phosphate moiety represented by dimethyl phosphate shows large solvation free energy contributions (around −72 kcal/mol), which is comparable to a value obtained previously for the analogous $H_2PO_4^-$ moiety.[56] The phosphodiester moiety has three possible minimum energy conformations, but the change in solvation free energy due to conformation is a relatively small percentage (0.1%) of its actual magnitude.

**TABLE 3: Electrostatic Contribution to Solvation Free Energy of Basic Unit Combinations Calculated Using Explicit and Continuum Models of Solvent with a Reentrant Molecular Surface**

| basic unit | name | FEP | continuum[a] | continuum[b] |
|---|---|---|---|---|
| nucleoside | | | | |
| Ade | d2 | −22.2 | −21.8 | −22.4 |
| | d3 | −19.7 | −19.7 | −20.5 |
| | r2 | −28.7 | −27.4 | −28.8 |
| | r3 | −23.9 | −23.6 | −25.0 |
| Cyt | d2 | −26.2 | −25.5 | −25.9 |
| | d3 | −24.0 | −23.3 | −24.3 |
| | r2 | −34.8 | −33.5 | −35.1 |
| | r3 | −28.9 | −28.1 | −29.7 |
| Gua | d2 | −31.6 | −31.1 | −31.9 |
| | d3 | −29.9 | −29.5 | −30.2 |
| | r2 | −36.9 | −36.2 | −38.1 |
| | r3 | −32.6 | −32.7 | −34.0 |
| Thy | d2 | −20.2 | −19.8 | −19.5 |
| | d3 | −17.8 | −17.2 | −17.7 |
| Ura | r2 | −28.4 | −27.2 | −27.6 |
| | r3 | −23.7 | −23.4 | −23.8 |
| sugar−diphosphate | | | | |
| | d2 | −180.2 | −182.6 | −179.3 |
| | d3 | −202.8 | −204.4 | −201.5 |
| | r2 | −177.4 | −182.2 | −178.0 |
| | r3 | −188.9 | −191.3 | −187.2 |

[a] Set of radii for hydrogen radius 1.5 Å. [b] Set of radii for hydrogen radius 0.0 Å, energy values in kcal/mol units.

The solvation free energy contributions of the combinations of basic units (nucleosides, sugar−diphosphate) are listed in Table 3. In the nucleoside compounds, there are four values (d2, d3, r2, r3) corresponding to each base because each base is studied in attachment to deoxyribose and ribose sugars in both C2′-endo and C3′-endo conformations. The only exceptions are uracil and thymine, with uracil being studied only in attachment to the ribose sugar and thymine being studied only in attachment to the deoxyribose sugar. This reduction in the number of compounds studied is possible because uracil is exclusively present in RNA whereas thymine is exclusively present in DNA. With the increase in size of the compounds, the differences between the different conformations and sugar types also increase. The largest solvation free energy contribution is seen for the r2 guanine nucleoside (−36.9 kcal/mol), the smallest is seen for the d3 thymine nucleoside (−17.8 kcal/mol). The largest difference between two conformations of the same compound occurs for the r2 and r3 cytosine nucleoside (a difference of 5.9 kcal/mol).

The sugar−diphosphate model compounds represent a sugar with its two flanking phosphates on the 3′ and 5′ side of the sugar. The presence of two negatively charged groups results in a large solvation free energy contribution (the largest is d3 with −202.8 kcal/mol; the smallest is r2 with −177.4 kcal/mol). Because introducing different conformations for the phosphodiester moieties creates a large number of combinations that differ

**TABLE 4: Comparison of Electrostatic Contribution to Solvation Free Energy of Basic Units Calculated Using the Continuum Solvent Model with a Reentrant Sphere Solute−Solvent Interface and a Reentrant Solute−Solvent Interface**

| | | | continuum[a] | | continuum[b] | |
|---|---|---|---|---|---|---|
| basic unit | name | FEP | reentrant | nonreentrant | reentrant | nonreentrant |
| base | Ade | −12.9 | −11.9 | −12.8 | −12.6 | −13.3 |
| | Cyt | −17.1 | −16.5 | −17.2 | −17.4 | −17.8 |
| | Gua | −21.4 | −20.5 | −21.7 | −21.7 | −22.5 |
| | Thy | −9.7 | −9.1 | −9.8 | −9.5 | −9.8 |
| | Ura | −10.8 | −10.4 | −11.0 | −10.4 | −10.6 |
| sugar | d2 | −15.1 | −15.7 | −19.6 | −13.8 | −14.4 |
| | d3 | −14.9 | −15.8 | −19.0 | −14.1 | −14.9 |
| | r2 | −20.5 | −20.3 | −24.1 | −19.8 | −20.5 |
| | r3 | −18.1 | −19.3 | −22.7 | −17.8 | −18.9 |
| phosphate | gg | −72.4 | −71.9 | −73.5 | −71.7 | −73.1 |
| | gt | −72.3 | −71.7 | −73.5 | −71.6 | −73.2 |
| | tt | −72.3 | −72.1 | −73.8 | −71.9 | −73.4 |

[a] Set of radii for hydrogen radius 1.5 Å. [b] Set of radii for hydrogen radius 0.0 Å, energy values in kcal/mol units.

very little in their solvation free energy, only the gg conformation of the phosphodiester moiety was considered. A change in sugar conformation, however, greatly affects the relative orientation and distance between the two negatively charged phosphodiester moieties. The C3′-endo conformation results in decreased distance between the two negatively charged phosphate groups, resulting in a decrease in the electrostatic solvation free energy. The influence of conformation on the solvation free energy is clearly illustrated by the difference of 22.6 kcal/mol between the d2 and d3 forms of the sugar−diphosphate moiety. The overall differences between the solvation free energies of the A- and B-forms of DNA are well-characterized.[57,58] The present calculations show that the large differences in electrostatic solvation free energy between A- and B-forms of DNA can be traced back to the local structural level and are probably caused primarily by changes in distances between negatively charged phosphate moieties due to change in sugar conformation.

**(c) Electrostatic Solvation Free Energies with Continuum Model.** Tables 2 and 3 show the results from the continuum electrostatics calculations using the two sets of optimized radii. It is to be noted that the optimization of the radii was done iteratively by optimizing the radii to fit the goal FEP data in Table 2 and then further optimizing this set of radii to get good agreement with the goal FEP data in Table 3. It is possible to get much better agreement with the basic unit model compound FEP results than shown here, but it is usually balanced out by reducing the agreement for the larger model compounds. A compromise therefore had to be made during the optimization where some of the accuracy for the smaller model compounds was sacrificed to get better accuracy for the larger model compounds. The optimization was made such that, for the 0.2 Å grid spacing and reentrant molecular surface, the difference between explicit solvent and implicit solvent results for both the basic unit model compounds and the nucleoside model compounds was as small as possible.

The comparison between the values obtained using the overlapping spheres representation and the reentrant molecular surface representation are shown in Tables 4 and 5. For most macromolecules, it is important to use a reentrant molecular surface when dealing with macromolecules to prevent a high dielectric value being assigned to a region inside the solute that cannot possibly be occupied by explicit solvent molecules. On the other hand, using overlapping solvent-excluded spheres is computationally simpler and makes the calculation of analytitc gradient of the electrostatic free energy possible.[50] In the present work, the optimization of the radii was carried out with a reentrant molecular surface that best describes the real solute−solvent boundary (the entire solvent-inaccessible region is included as part of the low-dielectric solute interior). The results obtained with no reentrant, shown in Tables 4 and 5, are typically more negative (more favorable electrostatic solvation free energies) than the results obtained with reentrant. In practice, a good reproduction of the electrostatic solvation free energies could be regained by scaling up the radii by 3% uniformly for the nonreentrant surface calculation (data not shown).

Two independent sets of radii for the heavy atoms were parametrized. The two sets differ because, in one set, the

**TABLE 5: Comparison of Electrostatic Contribution to Solvation Free Energy of Basic Unit Combinations Calculated Using Continuum Solvent Model with an Overlapping Sphere Solute−Solvent Interface and a Reentrant Solute−Solvent Interface**

| | | | continuum[a] | | continuum[b] | |
|---|---|---|---|---|---|---|
| basic unit | name | FEP | reentrant | nonreentrant | reentrant | nonreentrant |
| nucleoside | | | | | | |
| Ade | d2 | −22.2 | −21.8 | −26.3 | −22.4 | −23.8 |
| | d3 | −19.7 | −19.7 | −24.3 | −20.5 | −22.1 |
| | r2 | −28.7 | −27.4 | −32.1 | −28.8 | −30.3 |
| | r3 | −23.9 | −23.6 | −28.5 | −25.0 | −26.5 |
| Cyt | d2 | −26.2 | −25.5 | −30.0 | −25.9 | −27.3 |
| | d3 | −24.0 | −23.3 | −27.1 | −24.3 | −25.3 |
| | r2 | −34.8 | −33.5 | −38.4 | −35.1 | −36.5 |
| | r3 | −28.9 | −28.1 | −32.4 | −29.7 | −30.9 |
| Gua | d2 | −31.6 | −31.1 | −36.0 | −31.9 | −33.6 |
| | d3 | −29.9 | −29.5 | −34.0 | −30.2 | −32.0 |
| | r2 | −36.9 | −36.2 | −41.4 | −38.1 | −39.6 |
| | r3 | −32.6 | −32.7 | −37.6 | −34.0 | −35.7 |
| Thy | d2 | −20.2 | −19.8 | −23.7 | −19.5 | −20.5 |
| | d3 | −17.8 | −17.2 | −21.1 | −17.7 | −18.8 |
| Ura | r2 | −28.4 | −27.2 | −31.9 | −27.6 | −28.9 |
| | r3 | −23.7 | −23.4 | −27.6 | −23.8 | −25.1 |
| sugar−diphosphate | | | | | | |
| | d2 | −180.2 | −182.6 | −190.4 | −179.3 | −184.1 |
| | d3 | −202.8 | −204.4 | −211.7 | −201.5 | −205.8 |
| | r2 | −177.4 | −182.1 | −190.4 | −178.0 | −183.1 |
| | r3 | −188.9 | −191.3 | −199.0 | −187.2 | −191.8 |

[a] Set of radii for hydrogen radius 1.5 Å. [b] Set of radii for hydrogen radius 0.0 Å, energy values in kcal/mol units.

**TABLE 6: Optimized Born Atomic Radii (Å) for Nucleic Acids**

| moiety | identity | atom type | radius[a] | radius[b] |
|---|---|---|---|---|
| base | Ade | N1 | 1.75 | 1.75 |
| | Gua | N1 | 1.50 | 2.15 |
| | Ade/Gua | C2 | 2.15 | 2.15 |
| | Ade/Gua | N3 | 1.69 | 1.69 |
| | Ade/Gua | C4 | 2.12 | 2.12 |
| | Ade/Gua | C5 | 2.12 | 2.12 |
| | Ade/Gua | C6 | 2.12 | 2.12 |
| | Ade/Gua | N7 | 1.69 | 1.69 |
| | Ade/Gua | C8 | 1.60 | 2.12 |
| | Ade/Gua | N9 | 2.13 | 2.13 |
| | Ade | N6 | 1.50 | 2.17 |
| | Gua | O6 | 1.55 | 1.55 |
| | Gua | N2 | 1.50 | 2.12 |
| | Cyt/Thy/Ura | N1 | 2.20 | 2.20 |
| | Cyt/Thy/Ura | C2 | 2.04 | 2.04 |
| | Cyt | N3 | 1.68 | 1.68 |
| | Thy/Ura | N3 | 1.50 | 2.20 |
| | Cyt/Thy/Ura | C4 | 2.12 | 2.12 |
| | Cyt/Thy/Ura | C5 | 1.50 | 2.25 |
| | Cyt/Thy/Ura | C6 | 1.50 | 2.25 |
| | Cyt/Thy/Ura | O2 | 1.60 | 1.60 |
| | Cyt | N4 | 1.40 | 2.08 |
| | Thy/Ura | O4 | 1.60 | 1.60 |
| | Thy | C5M | 1.50 | 2.30 |
| sugar | | C1′ | 1.71 | 2.57 |
| | | C2′ | 1.84 | 2.70 |
| | | C3′ | 1.87 | 2.73 |
| | | C4′ | 1.64 | 2.50 |
| | | O4′ | 1.55 | 1.55 |
| | | C5′ | 1.71 | 2.57 |
| | | O3′ | 1.65 | 1.65 |
| | | O5′ | 1.65 | 1.65 |
| | | O2′ | 1.50 | 1.75 |
| phosphate | | P | 2.35 | 2.35 |
| | | O1P | 1.49 | 1.49 |
| | | O2P | 1.49 | 1.49 |

[a] Set of radii for hydrogen radius 1.5 Å. [b] Set of radii for hydrogen radius 0.0 Å. The blocking groups, −CH3 in the case of the base and phosphate analogues, and −OH in the case of sugar and nucleoside analogues, have different radii in the model compounds than those shown here.

hydrogen radii has to be included within the radii of the heavy atom. This ignores some of the structural details of the solute−solvent interface, which depends on the exact orientation of the hydrogen. During the optimization of the radii for both sets, it became clear that it was possible to get comparable accuracy with both approaches. The variation in the results due to other factors such as grid spacing was much greater than variation caused by using a hydrogen radius of 1.5 or 0.0 Å. Although it is clear that the solute−solvent interface is better represented with finite values of hydrogen radii, setting the hydrogen radii to zero is a much simpler approximation. The approximation of setting hydrogen radii to zero is especially desirable because this makes the nucleic acid radii compatible with atomic radii for protein atoms determined in an earlier study.[32] It may be useful to explicitly test the differences between both sets of radii during the determination of the electrostatic properties of large nucleic acids containing macromolecular complexes to understand the influence of having a finite hydrogen radius in the implicit solvent representation of macromolecules. The final optimized Born radii for both zero and nonzero hydrogen radii are shown in Table 6.

Computational details of the continuum model solvent calculations involve the choice of certain parameters that govern both the accuracy and the cost of the calculation. The tests quantifying changes in these parameters were done on the

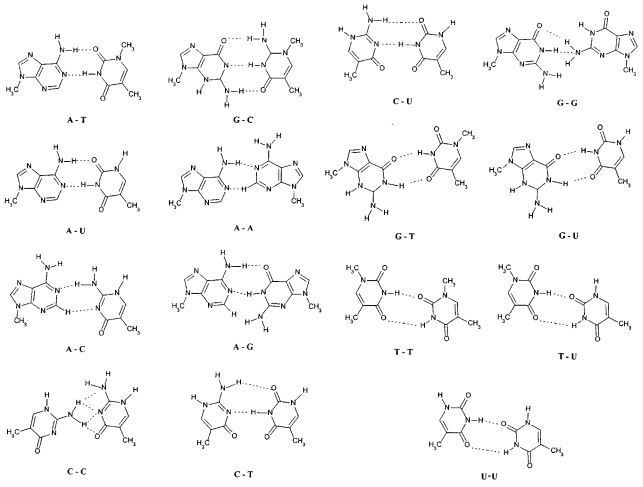


**Figure 4.** Configurations of regular and mismatched base pair combinations studied. All base pairs are coplanar. Except for the regular base pairs (GC, AT, AU), all base pairs have an arbitrary hydrogen bonding pattern with a minimum of one clear hydrogen bond.

nucleoside model compounds. For example, changing the grid spacing for the continuum electrostatic calculation can affect the computational time as well as the value of the calculated solvation free energy. Although decreasing the grid spacing from 0.2 to 0.1 Å resulted in a change of less than 1 kcal/mol in the solvation free energy, it required 15 times as much computational time. On the other hand, increasing the grid spacing from 0.2 to 0.5 Å can lead to change in the free energy of up to 4 kcal/mol and a decrease in the computational time by a factor of 14. The size of the grid is another important parameter that affects the distance from the outer atoms of the molecule to the edge of the grid. Tests were done with distances of 6 and 12 Å from the edge atoms of the molecules to the edge of the grid box. The increase in grid size did not change the results for solvation free energy for the 0.2 Å grid size. The positioning of the grid with respect to the molecule is another possible factor affecting the solvation free energy. This was tested by translating the grid by half the grid spacing in all directions and calculating the change in the calculated solvation free energy. For the 0.2 Å grid spacing and grid size extending 6 Å beyond the furthest atom, the largest difference in results was 0.2 kcal/mol, illustrating that the positioning of the grid does not play a significant role in determining the solvation free energy results.

**(d) Base Pairing.** The excellent agreement between the continuum electrostatics results with the optimized radii and the explicit solvent FEP calculations is encouraging. It is of interest to see, however, if the radii obtained for individual atoms can also reproduce the electrostatic solvation free energies for associated complexes of these compounds. For nucleic acids, the most interesting complexes are the hydrogen-bonded base pairs. To do this comparison, specific conformations of all possible base pair combinations were generated. Figure 4 shows the orientations used for all the base pair combinations. The normal base pairs (GC, AT, and AU) were generated by taking base pair structures from canonical B-DNA[55] and carrying out 500 ABNR minimization steps with the final structures being coplanar. Because there is very little structural information on mismatched base pairs, they were generated in arbitrary orientations with the only restrictions on the structures being that the two bases have at least one clear interbase hydrogen bond and that they are coplanar. The geometry of these mismatched structures was optimized using 500 steps of ABNR energy minimization. It is possible that the mismatched base pair global

**TABLE 7: Electrostatic Contribution to Solvation Free Energy of Base Pair Combinations Calculated Using Explicit and Continuum Models of Solvent with Overlapping Spheres and Reentrant Molecular Surfaces**

| | | continuum[a] | | continuum[b] | |
|---|---|---|---|---|---|
| base pair | FEP | reentrant | nonreentrant | reentrant | nonreentrant |
| AA | −21.8 | −20.5 | −22.6 | −21.2 | −23.1 |
| AC | −26.1 | −24.3 | −26.4 | −24.9 | −26.8 |
| AG | −24.1 | −23.9 | −26.0 | −24.5 | −26.5 |
| AT | −15.2 | −15.2 | −16.9 | −15.6 | −16.8 |
| AU | −15.6 | −16.2 | −17.6 | −16.1 | −17.2 |
| CC | −33.8 | −31.6 | −33.7 | −32.4 | −35.4 |
| CT | −20.8 | −20.7 | −22.5 | −20.8 | −22.2 |
| CU | −20.7 | −20.7 | −22.5 | −20.5 | −21.9 |
| GC | −21.9 | −21.2 | −22.8 | −21.8 | −23.1 |
| GG | −39.4 | −36.6 | −39.8 | −37.6 | −40.8 |
| GT | −21.9 | −21.5 | −23.2 | −21.7 | −23.2 |
| GU | −22.4 | −22.6 | −24.3 | −22.6 | −24.2 |
| TT | −12.2 | −12.8 | −13.9 | −12.6 | −13.3 |
| TU | −13.7 | −14.5 | −15.7 | −14.0 | −14.7 |
| UU | −14.7 | −15.5 | −16.7 | −14.6 | −15.2 |

[a] Set of radii for hydrogen radius 1.5 Å. [b] Set of radii for hydrogen radius 0.0 Å, energy values in kcal/mol units.

**TABLE 8: Free Energy of Base Pairing Estimated Using the Molecular Mechanics Vacuum Energy Difference and the FEP Results for the Electrostatic Solvation Free Energy of Base Pair Combinations and Individual Bases**

| Base pair, *ij* | $\Delta U_{MM}(ij)$ | $\Delta G_{elec}(ij)$ | $\Delta G_{elec}(i)$ | $\Delta G_{elec}(j)$ | $\Delta G(ij)$ |
|---|---|---|---|---|---|
| AA | −6.9 | −21.8 | −12.9 | −12.9 | −2.9 |
| AC | −7.5 | −26.1 | −12.9 | −17.1 | −3.6 |
| AG | −13.3 | −24.1 | −12.9 | −21.4 | −3.1 |
| AT | −12.3 | −15.2 | −12.9 | −9.7 | −4.9 |
| AU | −12.8 | −15.6 | −12.9 | −10.8 | −4.7 |
| CC | −7.1 | −33.8 | −17.1 | −17.1 | −6.7 |
| CT | −10.7 | −20.8 | −17.1 | −9.7 | −4.7 |
| CU | −11.2 | −20.7 | −17.1 | −10.8 | −4.0 |
| GC | −25.7 | −21.9 | −21.4 | −17.1 | −9.1 |
| GG | −7.1 | −39.4 | −21.4 | −21.4 | −3.7 |
| GT | −13.3 | −21.9 | −21.4 | −9.7 | −4.1 |
| GU | −14.1 | −22.4 | −21.4 | −10.8 | −4.3 |
| TT | −9.7 | −12.2 | −9.7 | −9.7 | −2.5 |
| TU | −10.2 | −13.7 | −9.7 | −10.8 | −3.4 |
| UU | −10.6 | −14.7 | −10.8 | −10.8 | −3.7 |

[a] Energy values in kcal/mol. The base pairing free energy, $\Delta G(ij)$, is calculated as $\Delta U_{MM}(ij) + \Delta G_{elec}(ij) - \Delta G_{elec}(i) - \Delta G_{elec}(j)$, where $\Delta U_{MM}(ij) = U_{MM}(ij) - U_{MM}(i) - U_{MM}(j)$.

minimum structures are different than the structures used here. In fact, the canonical structures show deviations from coplanarity even for the normal base pairs due to phenomena such as propeller twisting. The present structures are, therefore, idealized versions of base pair complexes. However, they can serve as useful approximations for judging of the accuracy of the continuum electrostatics model.

The base pair combinations yielded electrostatic solvation free energy contributions ranging from −12.2 kcal/mol for the TT complex to −39.4 kcal/mol for the GG complex. This wide range of values is to be expected given the differences in the electrostatic solvation free energy contributions of the isolated bases themselves and the differences in the hydrogen-bonding patterns of the various complexes. It should be noted that these FEP results for the base pair combinations were not used as goal data in the atomic radii optimization. Therefore, the good agreement between the continuum model and FEP is solely due to the quality of the optimized radii.

Table 7 shows the results for the overlapping spheres solute surface and the reentrant solute surface for the 0.2 Å grid spacing. Surprisingly, for the set with hydrogen radii set to 0.0, the differences were less than or equal to 1 kcal/mol for most base pair combinations in Table 7 with the reentrant surface. The only exceptions were the AC, CC, and GG mismatches with differences on the order of 1−2 kcal/mol. For the set with hydrogen radii set to 1.5 Å, the agreement was not as good, with the largest difference, 2.7 kcal/mol, being seen for the GG mismatch. Expectedly, the results from the nonreentrant surface are more negative than the explicit solvent FEP results and can be adjusted by scaling up the radii by 3%. All base pair combinations thus show relatively good agreement with the FEP results. For the regular base pairs (GC, AT, and AU), the results for the 0.2 Å grid spacing with hydrogen radii set to 0.0 Å show better agreement with the FEP results with the largest difference being 0.6 kcal/mol for the AU base pair. For the set with hydrogen radii set to 1.5 Å, the largest difference of 0.7 kcal/mol among the regular base pairs occurs in the GC base pair. The ability to reproduce a wide range of FEP values highlights the ability of the continuum solvent representation with optimized atomic radii to accurately treat the diverse electrostatic properties of the nucleic acid bases.

Using the molecular mechanics vacuum energy difference between the assembled base pair and the individual bases, and the electrostatic solvation free energy of the associated base pair and the individual bases, it is possible to estimate the free energy difference between the assembled base pair and infinitely separated bases. Table 8 shows the base pairing free energy estimates for all base pair combinations studied. It is generally true that the regular base pairs (AT, AU, and GC) have more favorable pairing free energies (greater than 4.5 kcal/mol) as compared to most of the mismatched base pairs. The CC and CT mismatches, however, show favorable pairing free energies of −6.7 and −4.7 kcal/mol, respectively. This might indicate that they possess comparable or greater binding affinity than the regular base pairs. The base pairing in nucleic acids, is, however, a function of not just the electrostatic interaction between the bases. A range of other factors such as base stacking and backbone conformational preferences play also an important role in the base pairing stability of each base pair.[55] Thus, in nucleic acid polymers, a pyrimidine−pyrimidine base pair such as CC or CT would probably have decreased base stacking and a strained backbone structure opposing the pairing. A look at the first column in Table 8 also indicates that estimation of the binding affinities purely based on vacuum interaction energies is not informative, because the contribution from solvation has a large impact on the final base pairing free energies. For example, the TT base pair vacuum interaction energy is −9.7 kcal/mol, but after consideration of electrostatic solvation effects, the base pairing free energy estimate is merely −2.5 kcal/mol. Despite the approximation of fixed solute structures, these results indicate that most of the mismatched base pairs (in the simplified geometries studied here) have smaller intrinsic pairing ability as compared to the regular base pairs due in part to solvation effects.

The separation of paired bases is an integral activity in the biological functions of nucleic acids such as replication and transcription.[51] Though this process certainly involves complex conformational transitions,[59] it is of interest to examine the influence of solvation on the separation of two bases along a highly idealized pathway. The free energy profile $\mathcal{W}(R)$ corresponding to base pair separation was calculated as

$$\mathcal{W}(R) = \Delta U_{MM}(R;ij) + \Delta G_{elec}(R;ij) - \Delta G_{elec}(i) - \Delta G_{elec}(j) \quad (7)$$

where $R$ is the distance between the center-of-mass of the two

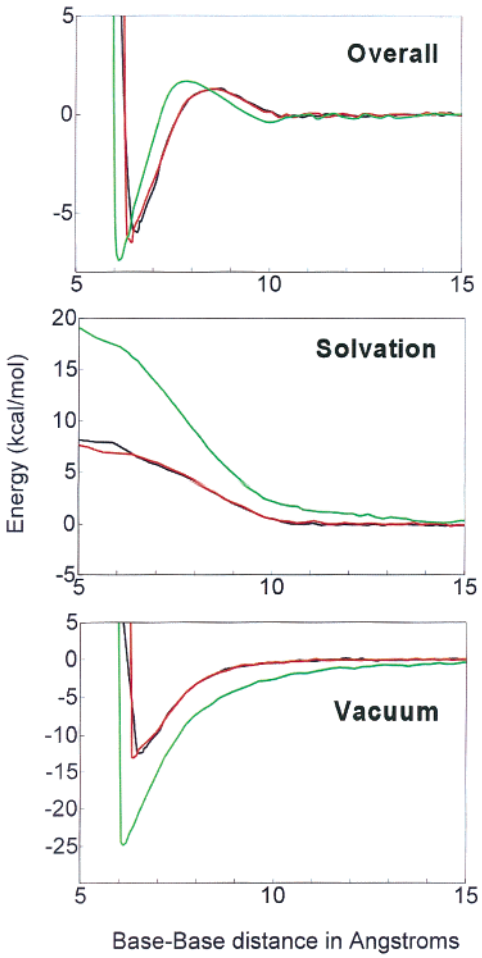


**Figure 5.** Change in energy associated with base pair separation in one dimension calculated by combining potential energy change in a vacuum with change in electrostatic solvation free energy for the naturally occurring AT, AU, and GC base pairs. Each configuration was generated by translating one base by a certain distance along an axis followed by 100 steps of ABNR energy minimization while the distance between the two bases was constrained using a 10 kcal/mol center-of-mass constraint. The electrostatic solvation free energy for all structures was calculated by solving the PB equation with a 0.2 Å grid spacing with a reentrant molecular surface and the set of radii with hydrogen radius set to 0.0 Å. The AT base pair is represented by the black line, the AU base pair by the red line, and the GC base pair by the green line

bases $i$ and $j$, $\Delta U_{MM}(R;ij)$ and $\Delta G_{elec}(R;ij)$ are the molecular mechanics potential energy and the electrostatic solvation free energy calculated for the base pair, respectively, and $\Delta G_{elec}(i)$ and $\Delta G_{elec}(j)$ are the electrostatic solvation free energies of isolated base pairs $i$ and $j$. Nonpolar contribution to the free energy profile are ignored in the present treatment. By convention, $\mathcal{W}(R)$ goes to zero at infinite separation.

The free energy profiles obtained for the AT, C−G, and AU base pairs are shown in Figure 5. The present calculation does not take into account the fluctuations in the base pair structures that contribute to deviations from the planar structure, i.e., propellar twisting or angular base pair opening. The calculations, however, roughly illustrate how the change in a vacuum interaction energy is balanced by the change in electrostatic solvation free energy. Figure 5 clearly show that the AT, AU, and GC base pairs in the gas phase have much greater interaction energies in a vacuum. The presence of the high-dielectric solvent reduces this interaction because of solvent screening and the fact that the separated individual bases are better solvated. The free energy of base pairing is directly related to the balance between increase in base−base interaction and desolvation penalty. Previous studies using both explicit[60] and implicit[61] solvent models have shown that the interaction energy per hydrogen bond is 2−3 kcal/mol between the paired bases. The present calculations agree well with this estimate. The overall order of base pair strengths is reproduced (AT ≈ AU < GC); however the difference between the GC and AT/AU base pair strengths is small, which is probably due to the inherent approximations mentioned above. Interestingly, an energy barrier opposing the base pair separation process, similar in magnitude to the results with explicit solvent simulations,[60] is present when the reentrant molecular surface is used in the PB calculations (Figure 5). This small barrier (less than 2 kcal/mol) is absent when the nonreentrant surface is used (data not shown).

## V. Summary

FEP calculations using molecular dynamics with explicit solvent were used to obtain electrostatic solvation free energy estimates for a range of model compounds representative of nucleic acids. The effect of conformational variation in these model compounds was also considered by looking at a few representative conformers known to be populated experimentally. Atomic radii obtained from radial charge solvent distributions in the explicit solvent simulations were used as initial starting points for the optimizaiton. Two sets of heavy atom Born radii were developed with hydrogen atomic radii set to be 1.5 and 0.0 Å, respectively. Either set of Born atomic radii reproduced the FEP results with satisfactory accuracy. The inclusion of a large number of model compounds with different conformations in the test set, however, limits the possibility of error when macromolecules are studied.

The properties of large macromolecular complexes can be explored with ease using the continuum solvent representation, providing insight into effects of change in structure, charge characteristics, and association of the molecules. Nonetheless, caution must be exerted when subtle quantitative comparisons are made with these methods because the results can be significantly influenced by the computational procedure. The agreement of the continuum model and explicit solvent results is not exact, which is to be expected given the approximations inherent in the continuum solvent representation. A continuum model is undoubtedly limited. The strongly charged character of nucleic acids results in an accumulation of positively charged ions in their vicinity, an effect called the counterion effect.[62,63] Solution properties of nucleic acids therefore exhibit a strong dependence on salt concentration.[1] In addition, an increased density of water molecules called a "spine" of hydration is also detectable in the groove regions of the nucleic acids.[64] These locally concentrated counterions and water molecules are believed to form an inherent part of nucleic acid structure.[65] Although the distribution of counterions in the vicinity of the nucleic acid could be represented in future studies by using properly parametrized radii for the counterions, further work will be required to characterize the importance of such counterion interactions and structured solvent molecules and, ultimately, assess the validity of the continuum model. It is our hope that the current radii can be used in conjunction with the previously determined set of radii obtained for protein atoms to study protein−nucleic acid complexes as well as large nucleic acid structures such as the ribosome.

**Acknowledgment.** This work was funded by grant MCB-0110847 from the National Science Foundation (NSF). We thank Wonpil Im for helpful discussions regarding the PB calculations.

## References and Notes


(1) McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2000**, *304*, 803.
(2) Hard, T. *Q. Rev. Biophys.* **1999**, *32*, 57.
(3) Tobias, D. J.; Mertz, J. E.; Brooks, C. L., III. *Biochemistry* **1991**, *30*, 6054.
(4) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(5) Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487.
(6) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
(7) Foloppe, N.; MacKerell, A. D., Jr., *J. Comput. Chem.* **2000**, *21*, 86.
(8) MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105.
(9) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155.
(10) Cheatham, T. E., III; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *15*, 435.
(11) York, D. M.; Yang, W. T.; Lee, H.; Darden, T.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001.
(12) MacKerell, A. D., Jr.; Lee, G. U. *Eur. Biophys. J. Biophys.* **1999**, *28*, 415.
(13) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.
(14) Jean-Charles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 1454.
(15) Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1988**, *89*, 5876.
(16) Bader, J. S.; Chandler, D. *J. Phys. Chem.* **1992**, *96*, 6423.
(17) Hunenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856.
(18) Hunenberger, P. H.; McCammon, J. A. *Biophys. Chem.* **1999**, *78*, 69.
(19) Sharp, K. A.; Friedman, R. A.; Misra, V.; Hecht, J.; Honig, B. *Biopolymers* **1995**, *36*, 245.
(20) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101.
(21) Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211.
(22) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1.
(23) Born, M. *Z. Phys.* **1920**, *1*, 45.
(24) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671.
(25) Honig, B.; Nicholls, A. *Science.* **1995**, *268*, 1144.
(26) Vorobjev, Y. N.; Hermans, J. *Biophys. Chem.* **1999**, *78*, 195.
(27) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889.
(28) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591.
(29) Hirata, F.; Redfern, P.; Levy, R. M. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1988**, *15*, 179.
(30) Jayaram, B.; Fine, R.; Honig, B. *J. Phys. Chem.* **1989**, *93*, 4320.
(31) Roux, B.; Yu, H.-A.; Karplus, M. *J. Phys. Chem.* **1990**, *94*, 4683.
(32) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem.* **1997**, *101*, 5239.
(33) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588.
(34) Marrone, T. J.; Gilson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1439.
(35) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.
(36) Schlick, T.; Overton, M. *J. Comput. Chem.* **1987**, *8*, 1025.
(37) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(38) Florian, J.; Strajbl, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 7959.
(39) Banavali, N. K.; MacKerell, A. D., Jr. *J. Am. Chem. Soc.* **2001**, *123*, 6747.
(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(41) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050.
(42) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351.
(43) Ryckaert, P. J.; Ciccotti, G.; Berendsen, J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(44) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(45) Souaille, M.; Roux, B. *Comput. Phys. Comm.* **2001**, *135*, 40.
(46) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct. Genet.* **1986**, *1*, 47.
(47) Davis, M. E.; Madura, J. D.; Luty B.; McCammon, J. A. *Comput. Phys. Comm.* **1991**, *62*, 187.
(48) Gilson, M. K.; Honig, B. *Proteins: Struct., Funct. Genet.* **1988**, *3*, 32.
(49) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(50) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59.
(51) Barton, M. C.; Crowe, A. J. *Oncogene* **2001**, *20*, 3094.
(52) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163.
(53) Dickerson, R. E.; Ng, H. L. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6986.
(54) Miller, J. L.; Kollman, P. A. *J. Phys. Chem. B* **1996**, *100*, 8587.
(55) Saenger, W. *TITLE*; Springer-Verlag: New York, 1984.
(56) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.
(57) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Am. Chem. Soc.* **1998**, *120*, 10629.
(58) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.
(59) Banavali, N. K.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2002**, *in press*.
(60) Stofer, E.; Chipot, C.; Lavery, R. *J. Am. Chem. Soc.* **1999**, *121*, 9503.
(61) Arora, N.; Jayaram, B. *J. Phys. Chem. B* **1998**, *102*, 6139.
(62) Manning, G. S.; Ray, J. *J. Biomol. Struct. Dyn.* **1998**, *16*, 461.
(63) Volker, J.; Klump, H. H.; Manning, G. S.; Breslauer, K. J. *J. Mol. Biol.* **2001**, *310*, 1011.
(64) Soler-Lopez, M.; Malinina, L.; Liu, J.; Huynh-Dinh, T.; Subirana, J. A. *J. Biol. Chem.* **1999**, *274*, 23683.
(65) Feig, M.; Pettitt, B. M. *Biophys. J.* **1999**, *77*, 1769.