Quantum Mechanical Characterization of Nucleic Acids in Solution: A Linear-Scaling Study of Charge Fluctuations in DNA and RNA

Jana Khandogin and Darrin M. York*

Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455

Received: December 31, 2001; In Final Form: April 9, 2002

Atom-centered point charges are a convenient and computationally efficient way to approximately represent the electrostatic properties of biological macromolecules. Atomic charges are routinely used in molecular modeling applications such as molecular simulations, molecular recognition, and ligand binding studies and for determining quantitative structure activity relationships. In the present paper a divide-and-conquer linear-scaling semiempirical method combined with a conductor-like screening model is applied to the calculation of charge distributions of solvated DNA and RNA duplexes in canonical A- and B-forms. The atomic charges on A-DNA, B-DNA, and A-RNA duplex decamers are analyzed to characterize the convergence of the linear-scaling method, and the effects of the charge model and semiempirical Hamiltonian. Furthermore, the inter- and intramolecular charge variations on DNA and RNA duplex 72-mers are investigated to gain insight into the influence of conformation, base stacking, and solvent polarization on the charge distributions. The charges derived from the linear-scaling semiempirical calculations reflect the electronic relaxation in the solvated macromolecular environment and therefore provide a better reference charge state for biomolecular modeling applications.

1. Introduction

For many biological applications, such as molecular recognition and metal ion binding, much information can be gained from examination of the electrostatic potential of a single native structure or a small number of representative conformers.[1−6] A particularly convenient computational model for the electrostatic potential involves the use of atom-centered point charges. Atomic charges are widely used in molecular mechanics force fields to model the electrostatic interactions[7] and as molecular descriptors in studies of quantitative structure−activity relationships (QSAR) for drug design problems.[8,9] Recently, the atomic charges of ligands have been used as variables to investigate ligand−receptor binding affinities and specificities.[6,10]

Under the most common approximation in conventional biomolecular modeling applications, the atom-centered point charges are *static*; i.e., they are held fixed, independent of conformation, chemical environment, and solvation. Recently, there has been much attention paid to the development and application of models that go beyond the static point charge description and can respond to their chemical environment.[11−15] These models attempt to include quantum mechanical "many-body" effects such as polarization and charge transfer in an empirical framework; applications of the methods have started to emerge (see references in ref 16). Nonetheless, for large biomolecules such as proteins and nucleic acids, the vast majority of present-day applications use the static atomic point-charge model.

The past several years have witnessed rapid advances in the development of linear-scaling electronic structure methods,[17−22] that can, in principle, offer a fully quantum mechanical description of the electron density (that includes quantum many-body effects) for very large systems. In fact, with such a description, many more *response properties*[23] would be accessible, contributing to the arsenal of chemical descriptors for QSAR applications. This is a topic that will be addressed in future work.[24] While a fully quantum description of all electrons in a biological macromolecule remains a daunting task at the ab initio level with quality basis sets, with the use of semiempirical model Hamiltonians,[25−27] linear-scaling approaches, such as divide-and-conquer[28−30] and pseudodiagonalization[31] algorithms, have found increasing applications.[23,32−36] On the condensed matter physics front, Fermi operator expansion[37,38] and density matrix minimization[39] based tight-binding calculations (see references in refs 19 and 20) have been widely applied to study crystalline and metal systems.

Once a relaxed macromolecular charge distribution has been calculated quantum mechanically for a representative set of structures, it can be analyzed and used to construct a highly accurate electrostatic potential map. However, to make the information accessible to programs that work with point charges, a model is required to reduce the full electron density and nuclear charges into an intuitively meaningful, transferable atomic charge set. The most common methods for the determination of atomic charges for small molecules can be divided into two main classes: methods based on (1) electrostatic potential (ESP) or electric field fitting and (2) electron density partitioning. Fitting methods, such as the Singh−Kollman−Besler−Merz scheme,[40,41] CHELP and CHELPG models,[42,43] RESP model,[44] and others[45−48] have the advantage of providing an accurate representation of the electrostatic potential (or electric field) in the region outside the molecular surface. These methods, however, frequently run into problems associated with ill-conditioning of the fitting procedure. These difficulties can be partially overcome using singular value decompositions;[48] however, some parameters may still be underdetermined, particularly for "buried" atoms, and lead to nontransferable unrealistic charges. Procedures that add a restraint on the

* Corresponding author: york@chem.umn.edu.

10.1021/jp0146667 CCC: $22.00
Published on Web 07/13/2002

magnitude of the atomic charges in the fitting have been introduced to help correct the problem and have been successfully applied in biological force field development.[44] Nonetheless, for macromolecules, the difficulties associated with conventional charge fitting are greatly exacerbated.

An alternative method to derive charges is through partitioning of the electron density or single-particle density matrix.[49−53] These methods treat buried atoms on more "equal footing" as compared to the ESP fitting procedures; however, they often depend strongly on the level of theory and basis set. Recently, efforts have been made to improve the quality of charges derived from density matrix partitioning methods by introduction of a model correction term that contains empirical parameters adjusted to reproduce experimental dipole moments.[54,55] Of particular interest for the purposes of this paper is the CM2 model,[55] based on an empirical mapping of Löwdin charges to reproduce experimental gas-phase dipole moments for a database of small molecules. The CM2 model provides a computationally economical way to generate high-quality charges for a variety of organic and biological molecules, including nucleic acid and protein fragments.[56] These charges have been recently used to describe electrostatic interactions in a QM/MM application[57] and to predict solvation polarization energies with the Generalized-Born (GB)[55,58] and Poisson−Boltzmann (PB) solvation models.[59]

In the present work, the CM2 charge model has been implemented in a linear-scaling semiempirical quantum program with the AM1[60] and PM3[61] Hamiltonians to study charge variations in solvated canonical forms of A- and B-DNA and RNA. Here a divide-and-conquer (D&C) approach[62,30] for the electronic structure is combined with a linear-scaling preconditioned conjugate-gradient/fast multipole method[30] for solvation calculation using the conductor-like screening model (COSMO).[63,64] Linear-scaling capability of the method has been demonstrated and discussed previously.[23,30,32,62,65,66] The purpose of the present work is to apply linear-scaling electronic structure methods to investigate the effects of conformation, solvation, and base stacking environment on the charge distributions of large nucleic acids.

The following section (section 2) outlines the computational details. Section 3 provides convergence results for calculations of decamer DNAs and RNAs. Section 4 gives a detailed analysis of intra- and intermolecular charge variations of 72-mer sequences of A-DNA, B-DNA, and A-RNA, discusses the effect of solvation on the charge distributions, and compares CM2 charges with Mulliken charges for both AM1 and PM3 semiempirical Hamiltonians. The last section draws conclusions from this work and outlines future research directions that have potential impact on pharmaceutical applications of computational chemistry.

2. Methods and Computational Details

2.1. Theory. In this study, a linear-scaling divide-and-conquer (D&C) semiempirical approach for electronic structure is applied to solvated macromolecules. A detailed description of the method and its implementation with a linear-scaling solvation model has been presented elsewhere.[30,62,64] The method is briefly outlined below, with emphasis on the essential features most relevant to the present work.

In the D&C approach, a molecule is divided into spatially localized groups of atoms termed *subsystems*. The goal of the method is to determine accurately the single-particle density matrix *in the region of a subsystem* without enforcement of a global orthogonalization or idempotency constraint that leads to the nonlinear scaling bottlenecks of conventional methods. If calculation of the single-particle density matrix in the region around any one subsystem could be achieved with a fixed amount of computational effort (independent of the total system size), then a linear-scaling algorithm could be attained by summation of the subsystem density matrix contributions to give the global density matrix. Here, it is assumed that the single-particle density matrix is inherently sparse, such as in the case of an insulating system where the off-diagonal matrix elements decay exponentially with distance, and that the subsystem density matrix contributions can be determined to a constant level of accuracy. For systems with a nonsparse density matrix (e.g., metallic systems) other linear-scaling methods such as those based on a Fermi operator expansion[37,38] offer significant advantages.

For quantum methods that use localized basis functions to expand the molecular orbitals, it is convenient to *partition* the global single-particle density matrix using basis functions centered on spatially localized sets of atoms. The basis functions associated with atoms within the same subsystem are termed the *subsystem basis*. The union of all the subsystem basis functions forms the global basis. To obtain an accurate representation of the density matrix in the region of a subsystem requires projection of the Fock operator in a basis space that extends beyond the subsystem. For this purpose, a set of neighboring *buffer atoms* are introduced for each subsystem, the associated basis functions of which (the *buffer space*) are used to extend the local basis space.

For an appropriately chosen subsystem and buffer region, the single-particle density matrix in the region of the subsystem can be accurately obtained by solving a set of local Hatree-Fock-like equations:

$$\mathbf{F}^{\alpha}\mathbf{C}^{\alpha} = \mathbf{S}^{\alpha}\mathbf{C}^{\alpha}\mathbf{E}^{\alpha} \tag{1}$$

where $\mathbf{F}^{\alpha}$ is a projection of the Fock matrix in the local basis formed from the subsystem α and its buffer, $\mathbf{S}^{\alpha}$ is the overlap matrix in the local basis, $\mathbf{C}^{\alpha}$ is the matrix of orbital coefficients, and $\mathbf{E}^{\alpha}$ is the corresponding diagonal matrix of local orbital eigenvalues.

Usually, buffer atoms are chosen to be those within a certain distance (buffer cutoff) from the subsystem atoms. The global density matrix is then constructed from the local subsystem density matrices, which for a closed shell calculation is written as

$$P_{\mu\nu} \approx \sum_{\alpha} P^{\alpha}_{\mu\nu} = \sum_{\alpha} W^{\alpha}_{\mu\nu} \cdot 2 \sum_{i} n_i^{\alpha} C^{\alpha}_{i\mu} C^{\alpha}_{i\nu} \tag{2}$$

where $C^{\alpha}_{i\mu}$ are the local orbital coefficients obtained from solving eq 1, n_i^{α} are occupation numbers, and $W^{\alpha}_{\mu\nu}$ are the local weight matrices defined as

$$W^{\alpha}_{\mu\nu} = \begin{cases} 1 & \text{if } \mu \text{ and } \nu \in \text{subsystem } \alpha \\ {}^1/_2 & \text{if } \mu \text{ or } \nu \in \text{subsystem } \alpha \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

This set of weight functions localize the single-particle density matrix and satisfy an equipartition of unity: $\sum_{\alpha} W^{\alpha}_{\mu\nu} = 1$. In eq 2, the occupation number n_i^{α} is determined by a Fermi function

$$n_i^{\alpha} = \frac{1}{1 + \mathrm{e}^{(\epsilon_i^{\alpha} - \mu)/k_{\mathrm{B}}T}} \tag{4}$$

TABLE 1: Convergence of the RMSD of Solution-Phase and Solvent-Induced Polarization Charge Vectors with Buffer and Matrix Cutoffs for a Canonical B-DNA Decamer[a]

AM1/CM2	4/5	6/7	8/9	10/11	12/13
4/5	–	0.444×10^{-2}	0.445×10^{-2}	0.445×10^{-2}	$\mathbf{0.445} \times 10^{-2}$
6/7	$\mathit{0.638} \times 10^{-2}$	–	0.538×10^{-4}	0.539×10^{-4}	$\mathbf{0.537} \times 10^{-4}$
8/9	$\mathit{0.638} \times 10^{-2}$	$\mathit{0.172} \times 10^{-3}$	–	0.331×10^{-5}	$\mathbf{0.351} \times 10^{-5}$
10/11	$\mathit{0.637} \times 10^{-2}$	$\mathit{0.177} \times 10^{-3}$	$\mathit{0.684} \times 10^{-4}$	–	$\mathbf{0.293} \times 10^{-5}$
12/13	$\boldsymbol{0.637} \times 10^{-2}$	$\boldsymbol{0.172} \times 10^{-3}$	$\boldsymbol{0.559} \times 10^{-4}$	$\boldsymbol{0.395} \times 10^{-4}$	–
AM1/Mul	**4/5**	**6/7**	**8/9**	**10/11**	**12/13**
4/5	–	0.443×10^{-2}	0.443×10^{-2}	0.443×10^{-2}	$\mathbf{0.443} \times 10^{-2}$
6/7	$\mathit{0.605} \times 10^{-2}$	–	0.534×10^{-4}	0.537×10^{-4}	$\mathbf{0.534} \times 10^{-4}$
8/9	$\mathit{0.605} \times 10^{-2}$	$\mathit{0.142} \times 10^{-3}$	–	0.326×10^{-5}	$\mathbf{0.401} \times 10^{-5}$
10/11	$\mathit{0.605} \times 10^{-2}$	$\mathit{0.148} \times 10^{-3}$	$\mathit{0.683} \times 10^{-4}$	–	$\mathbf{0.240} \times 10^{-5}$
12/13	$\boldsymbol{0.605} \times 10^{-2}$	$\boldsymbol{0.153} \times 10^{-3}$	$\boldsymbol{0.558} \times 10^{-4}$	$\boldsymbol{0.395} \times 10^{-4}$	–
PM3/CM2	**4/5**	**6/7**	**8/9**	**10/11**	**12/13**
4/5	–	0.695×10^{-2}	0.695×10^{-2}	0.697×10^{-2}	$\mathbf{0.697} \times 10^{-2}$
6/7	$\mathit{0.882} \times 10^{-2}$	–	0.486×10^{-4}	0.600×10^{-4}	$\mathbf{0.585} \times 10^{-4}$
8/9	$\mathit{0.883} \times 10^{-2}$	$\mathit{0.189} \times 10^{-3}$	–	0.290×10^{-4}	$\mathbf{0.261} \times 10^{-4}$
10/11	$\mathit{0.887} \times 10^{-2}$	$\mathit{0.189} \times 10^{-3}$	$\mathit{0.137} \times 10^{-3}$	–	$\mathbf{0.420} \times 10^{-5}$
12/13	$\boldsymbol{0.886} \times 10^{-2}$	$\boldsymbol{0.185} \times 10^{-3}$	$\boldsymbol{0.104} \times 10^{-3}$	$\boldsymbol{0.884} \times 10^{-4}$	–
PM3/Mul	**4/5**	**6/7**	**8/9**	**10/11**	**12/13**
4/5	–	0.721×10^{-2}	0.722×10^{-2}	0.724×10^{-2}	$\mathbf{0.724} \times 10^{-2}$
6/7	$\mathit{0.899} \times 10^{-2}$	–	0.532×10^{-4}	0.642×10^{-4}	$\mathbf{0.629} \times 10^{-4}$
8/9	$\mathit{0.902} \times 10^{-2}$	$\mathit{0.220} \times 10^{-3}$	–	0.292×10^{-4}	$\mathbf{0.264} \times 10^{-4}$
10/11	$\mathit{0.904} \times 10^{-2}$	$\mathit{0.216} \times 10^{-3}$	$\mathit{0.118} \times 10^{-3}$	–	$\mathbf{0.438} \times 10^{-5}$
12/13	$\boldsymbol{0.904} \times 10^{-2}$	$\boldsymbol{0.213} \times 10^{-3}$	$\boldsymbol{0.111} \times 10^{-3}$	$\boldsymbol{0.395} \times 10^{-4}$	–

[a] The table shows comparison matrices of the RMSD between charge vectors (in atomic units) calculated with different R_b/R_m cutoff schemes (shown as row and column headers in Å) in the divide-and-conquer method. The RMSD of a vector $\mathbf{\Delta x}$ is defined as $\sqrt{\langle \mathbf{\Delta x}^2 \rangle - \langle \mathbf{\Delta x} \rangle^2}$, where here $\mathbf{\Delta x}$ is the difference between two charge vectors. The RMSD of the solution charges $\mathbf{q}$ (normal font) and solvent-induced polarization charges $\boldsymbol{\delta q}$ (italics) are shown as the upper and lower triangles of each matrix, respectively. Rows and columns that compare to the highest level 12/13 Å R_b/R_m scheme are shown in bold and are reasonable indicators of the convergence level of the lower order schemes. All calculations were performed on the B-DNA duplex decamer.

where ϵ_i^α is the local orbital energy, k_B is the Boltzmann constant, T is the temperature, and μ is the chemical potential (the Fermi energy in the zero temperature limit) that is chosen to satisfy the normalization condition

$$N_e = \sum_{\mu\nu} P_{\mu\nu} S_{\nu\mu} \tag{5}$$

where N_e is the total number of electrons (or valence electrons in the case of the semiempirical methods applied here) in the system. Once the chemical potential is determined, the subsystem density matrix can be calculated via eq 2, which is coupled with eq 1. The calculation proceeds iteratively until self-consistency is achieved.

Linear-scaling electronic structure methods all contain "parameters" that control, for example, the threshold on the sparsity patterns of the matrices, and tolerances on the precision of the solution of the equations involved. In the present implementation of the D&C method, there are two main parameters, R_m and R_b: R_m, the "matrix cutoff" defining the matrix sparsity pattern, and R_b, the "buffer cutoff" that defines the basis functions in the buffer space.[62] The accuracy of the D&C method can be adjusted by adjusting R_b and R_m appropriately. Intuition suggests, and experience has born out, that R_m should be slightly larger than R_b for an optimal balance between accuracy and efficiency.

Applications to biological macromolecules require an efficient and accurate solvent model. The current linear-scaling implementation is based on the conductor-like screening model (COSMO).[63,64] In this model, a molecule can be considered to be situated in a cavity inside a dielectric continuum. The induced surface charge density responsible for the solvent reaction field is obtained by the minimization of the total electrostatic energy; i.e., the electrostatic interaction of the surface charge with the solute electron density plus the self-interaction of the surface charges themselves. The surface charge density is modeled by a set of surface elements on the solvent accessible surface of the macromolecule. To avoid a matrix inversion that scales as $O(M^3)$, where M is the number of the surface elements, a preconditioned conjugate gradient/recursive bisection fast multipole method[30] is used. The recursive bisection fast multipole technique[67,68] is applied to evaluate the Coulombic potential of surface charge vectors in $O[M \log(M)]$ effort.

The CM2 charge on atom k is defined as

$$q_k = q_k^0 + \sum_{k' \neq k} T_{kk'}(B_{kk'}) \tag{6}$$

where q_k^0 is the atomic charge obtained from Löwdin population analysis (same as Mulliken population analysis in the zero-differential overlap based semiempirical methods applied here), $T_{kk'}$ is a term that represents charge transfer from atom k' to atom k defined by

$$T_{kk'} = B_{kk'}(D_{kk'} + C_{kk'}B_{kk'}) \tag{7}$$

$B_{kk'}$ is the bond order[69]

$$B_{kk'} = \sum_{i \in k} \sum_{j \in k'} (\mathbf{PS})_{ij} (\mathbf{PS})_{ji} \tag{8}$$

and $D_{kk'}$ and $C_{kk'}$ are two parameters determined by fitting the charge-derived dipole moments to experimental (or ab initio) values.

TABLE 2: Comparison of Mulliken and CM2 Charges for Canonical A- and B-Form DNA and RNA Decamers[a]

	base mean (RMSD)	sugar mean (RMSD)	phosphate mean (RMSD)
		$\Delta\mathbf{x} = \mathbf{q}^{\mathrm{AM1}}_{\mathrm{CM2}} - \mathbf{q}^{\mathrm{AM1}}_{\mathrm{Mul}}$	
A-DNA	-0.518×10^{-2} (0.260)	0.110×10^{-1} (0.512×10^{-1})	-0.160×10^{-1} (0.113)
B-DNA	-0.545×10^{-2} (0.261)	0.111×10^{-1} (0.512×10^{-1})	-0.157×10^{-1} (0.111)
A-RNA	-0.539×10^{-2} (0.266)	0.102×10^{-1} (0.785×10^{-1})	-0.160×10^{-1} (0.114)
B-RNA	-0.570×10^{-2} (0.266)	0.103×10^{-1} (0.736×10^{-1})	-0.157×10^{-1} (0.112)
		$\Delta\mathbf{x} = \mathbf{q}^{\mathrm{PM3}}_{\mathrm{CM2}} - \mathbf{q}^{\mathrm{PM3}}_{\mathrm{Mul}}$	
A-DNA	-0.419×10^{-2} (0.225)	0.102×10^{-1} (0.354×10^{-1})	-0.167×10^{-1} (0.271)
B-DNA	-0.432×10^{-2} (0.226)	0.101×10^{-1} (0.351×10^{-1})	-0.165×10^{-1} (0.268)
A-RNA	-0.439×10^{-2} (0.230)	0.936×10^{-2} (0.703×10^{-1})	-0.166×10^{-1} (0.273)
B-RNA	-0.442×10^{-2} (0.231)	0.937×10^{-2} (0.622×10^{-1})	-0.163×10^{-1} (0.270)
		$\Delta\mathbf{x} = \delta\mathbf{q}^{\mathrm{AM1}}_{\mathrm{CM2}} - \delta\mathbf{q}^{\mathrm{AM1}}_{\mathrm{Mul}}$	
A-DNA	0.292×10^{-3} (0.528×10^{-2})	-0.727×10^{-3} (0.169×10^{-2})	0.160×10^{-2} (0.245×10^{-1})
B-DNA	-0.474×10^{-4} (0.371×10^{-2})	-0.381×10^{-3} (0.115×10^{-2})	0.116×10^{-2} (0.183×10^{-1})
A-RNA	0.198×10^{-3} (0.537×10^{-2})	-0.593×10^{-3} (0.243×10^{-2})	0.157×10^{-2} (0.235×10^{-1})
B-RNA	0.496×10^{-4} (0.355×10^{-2})	-0.314×10^{-3} (0.233×10^{-2})	0.113×10^{-2} (0.184×10^{-1})
		$\Delta\mathbf{x} = \delta\mathbf{q}^{\mathrm{PM3}}_{\mathrm{CM2}} - \delta\mathbf{q}^{\mathrm{PM3}}_{\mathrm{Mul}}$	
A-DNA	0.219×10^{-4} (0.463×10^{-2})	-0.288×10^{-3} (0.126×10^{-2})	0.124×10^{-2} (0.413×10^{-1})
B-DNA	-0.182×10^{-4} (0.390×10^{-2})	-0.208×10^{-3} (0.962×10^{-3})	0.750×10^{-3} (0.288×10^{-1})
A-RNA	0.225×10^{-4} (0.475×10^{-2})	-0.364×10^{-3} (0.235×10^{-2})	0.115×10^{-2} (0.392×10^{-1})
B-RNA	0.115×10^{-4} (0.371×10^{-2})	-0.171×10^{-3} (0.260×10^{-2})	0.750×10^{-3} (0.283×10^{-1})

[a] This table compares Mulliken and CM2 charges for DNA and RNA duplex decamers in different forms (A and B). Solution-phase charges **q** and solvent-induced polarization charges $\delta\mathbf{q}$ are compared for both AM1 and PM3 Hamiltonians. The mean and RMSD statistical quantities are calculated from the difference charge vector $\Delta\mathbf{x}$ defined in the table and partitioned into base, sugar, and phosphate components. The phosphate group is defined here as the PO_4 unit consisting of P, OP1, OP2, O3′, and O5′ atoms, while the rest of the atoms are grouped as base and sugar units according to the standard nomenclature.[77] The charges on the terminal group atoms (H5T, O3T, and H3T) do not contribute to the statistics in the table.

TABLE 3: Comparison of AM1 and PM3 Charges for Canonical A- and B-Form DNA and RNA Decamers[a]

	base mean (RMSD)	sugar mean (RMSD)	phosphate mean (RMSD)
		$\Delta\mathbf{x} = \mathbf{q}^{\mathrm{AM1}}_{\mathrm{CM2}} - \mathbf{q}^{\mathrm{PM3}}_{\mathrm{CM2}}$	
A-DNA	-0.948×10^{-2} (0.234)	0.175×10^{-1} (0.372×10^{-1})	-0.200×10^{-1} (0.421)
B-DNA	-0.947×10^{-2} (0.237)	0.175×10^{-1} (0.410×10^{-1})	-0.200×10^{-1} (0.421)
A-RNA	-0.975×10^{-2} (0.239)	0.166×10^{-1} (0.384×10^{-1})	-0.214×10^{-1} (0.418)
B-RNA	-0.984×10^{-2} (0.242)	0.168×10^{-1} (0.416×10^{-1})	-0.218×10^{-1} (0.418)
		$\Delta\mathbf{x} = \mathbf{q}^{\mathrm{AM1}}_{\mathrm{Mul}} - \mathbf{q}^{\mathrm{PM3}}_{\mathrm{Mul}}$	
A-DNA	-0.849×10^{-2} (0.205)	0.167×10^{-1} (0.477×10^{-1})	-0.207×10^{-1} (0.262)
B-DNA	-0.834×10^{-2} (0.209)	0.165×10^{-1} (0.495×10^{-1})	-0.209×10^{-1} (0.264)
A-RNA	-0.875×10^{-2} (0.210)	0.158×10^{-1} (0.467×10^{-1})	-0.220×10^{-1} (0.259)
B-RNA	-0.856×10^{-2} (0.213)	0.159×10^{-1} (0.491×10^{-1})	-0.225×10^{-1} (0.260)
		$\Delta\mathbf{x} = \delta\mathbf{q}^{\mathrm{AM1}}_{\mathrm{CM2}} - \delta\mathbf{q}^{\mathrm{PM3}}_{\mathrm{CM2}}$	
A-DNA	-0.485×10^{-3} (0.801×10^{-2})	0.175×10^{-2} (0.766×10^{-2})	-0.320×10^{-2} (0.408×10^{-1})
B-DNA	-0.186×10^{-3} (0.807×10^{-2})	0.108×10^{-2} (0.458×10^{-2})	-0.242×10^{-2} (0.300×10^{-1})
A-RNA	-0.481×10^{-3} (0.848×10^{-2})	0.154×10^{-2} (0.624×10^{-2})	-0.297×10^{-2} (0.380×10^{-1})
B-RNA	-0.111×10^{-3} (0.797×10^{-2})	0.108×10^{-2} (0.508×10^{-2})	-0.277×10^{-2} (0.289×10^{-1})
		$\Delta\mathbf{x} = \delta\mathbf{q}^{\mathrm{AM1}}_{\mathrm{CM2}} - \delta\mathbf{q}^{\mathrm{PM3}}_{\mathrm{CM2}}$	
A-DNA	-0.755×10^{-3} (0.802×10^{-2})	0.218×10^{-2} (0.715×10^{-2})	-0.356×10^{-2} (0.396×10^{-1})
B-DNA	-0.157×10^{-3} (0.844×10^{-2})	0.125×10^{-2} (0.440×10^{-2})	-0.283×10^{-2} (0.311×10^{-1})
A-RNA	-0.679×10^{-3} (0.862×10^{-2})	0.177×10^{-2} (0.589×10^{-2})	-0.339×10^{-2} (0.376×10^{-1})
B-RNA	-0.149×10^{-3} (0.835×10^{-2})	0.122×10^{-2} (0.483×10^{-2})	-0.315×10^{-2} (0.310×10^{-1})

[a] This table compares AM1 and PM3 charges for DNA and RNA duplex decamers in different forms (A and B). Solution-phase charges **q** and solvent-induced polarization charges $\delta\mathbf{q}$ are compared for both CM2 and Mulliken charge models. The mean and RMSD statistical quantities are calculated from the difference charge vector $\Delta\mathbf{x}$ defined in the table and partitioned into base, sugar, and phosphate components. The phosphate group is defined here as the PO_4 unit consisting of P, OP1, OP2, O3′, and O5′ atoms, while the rest of the atoms are grouped as base and sugar units according to the standard nomenclature.[77] The charges on the terminal group atoms (H5T, O3T, and H3T) do not contribute to the statistics in the table.

2.2. Computational Details. The nucleic acids used in this study include the A- and B-forms of the duplex DNA decamer d(CCAACGTTGG)$_2$ (640 atoms), and the corresponding duplex RNA decamer r(CCAACGUUGG)$_2$ (648 atoms), as well as a

TABLE 4: Comparison of Intermolecular (AM1/CM2) Charge Variations for Canonical B-DNA, A-DNA, and A-RNA 72-mers[a]

base mean (RMSD)	sugar mean (RMSD)	phosphate mean (RMSD)
$\Delta\mathbf{x} = \mathbf{q}^{AM1}_{CM2}(\text{B-DNA}_{72})$	$-\mathbf{q}^{AM1}_{CM2}(\text{A-DNA}_{72})$	
0.314×10^{-2} (0.323×10^{-1})	0.138×10^{-2} (0.186×10^{-1})	-0.602×10^{-2} (0.296×10^{-1})
$\Delta\mathbf{x} = \mathbf{q}^{AM1}_{CM2}(\text{B-DNA}_{72})$	$-\mathbf{q}^{AM1}_{CM2}(\text{A-RNA}_{72})$	
0.287×10^{-2} (0.372×10^{-1})	0.261×10^{-2} (0.978×10^{-1})	-0.794×10^{-2} (0.307×10^{-1})
$\Delta\mathbf{x} = \mathbf{q}^{AM1}_{CM2}(\text{A-DNA}_{72})$	$-\mathbf{q}^{AM1}_{CM2}(\text{A-RNA}_{72})$	
-0.457×10^{-3} (0.271×10^{-1})	0.122×10^{-2} (0.992×10^{-1})	-0.193×10^{-2} (0.266×10^{-1})
$\Delta\mathbf{x} = \mathbf{q}^{AM1}_{CM2}(\text{B-DNA}_{72})$	$-\boldsymbol{\delta}\mathbf{q}^{AM1}_{CM2}(\text{A-DNA}_{72})$	
0.407×10^{-3} (0.391×10^{-1})	-0.679×10^{-2} (0.315×10^{-1})	0.251×10^{-1} (0.534×10^{-1})
$\Delta\mathbf{x} = \mathbf{q}^{AM1}_{CM2}(\text{B-DNA}_{72})$	$-\boldsymbol{\delta}\mathbf{q}^{AM1}_{CM2}(\text{A-RNA}_{72})$	
0.156×10^{-2} (0.335×10^{-1})	-0.544×10^{-2} (0.286×10^{-1})	0.179×10^{-1} (0.427×10^{-1})
$\Delta\mathbf{x} = \mathbf{q}^{AM1}_{CM2}(\text{A-DNA}_{72})$	$-\boldsymbol{\delta}\mathbf{q}^{AM1}_{CM2}(\text{A-RNA}_{72})$	
0.113×10^{-2} (0.156×10^{-1})	0.135×10^{-2} (0.645×10^{-2})	-0.717×10^{-2} (0.676×10^{-1})

[a] This table compares AM1/CM2 charges between different DNA and RNA duplex 72-mers. Solution-phase charges **q** and solvent-induced polarization charges $\boldsymbol{\delta}\mathbf{q}$ are compared between B-DNA, A-DNA, and A-RNA. The mean and RMSD statistical quantities are calculated from the difference charge vector $\Delta\mathbf{x}$ defined in the table and partitioned into base, sugar, and phosphate components. The phosphate group is defined here as the PO_4-unit consisting of P, OP1, OP2, O3′, and O5′ atoms, while the rest of the atoms are grouped as base and sugar units according to the standard nomenclature.[77] For comparison between DNA and RNA that require charge vectors of identical lengths, the united charge (O2′ and H2″) of the RNA sugar is compared to that of H2′ of the DNA sugar, and the united charge of the C5 methyl group of thymine is compared to the H5 charge of uracil.

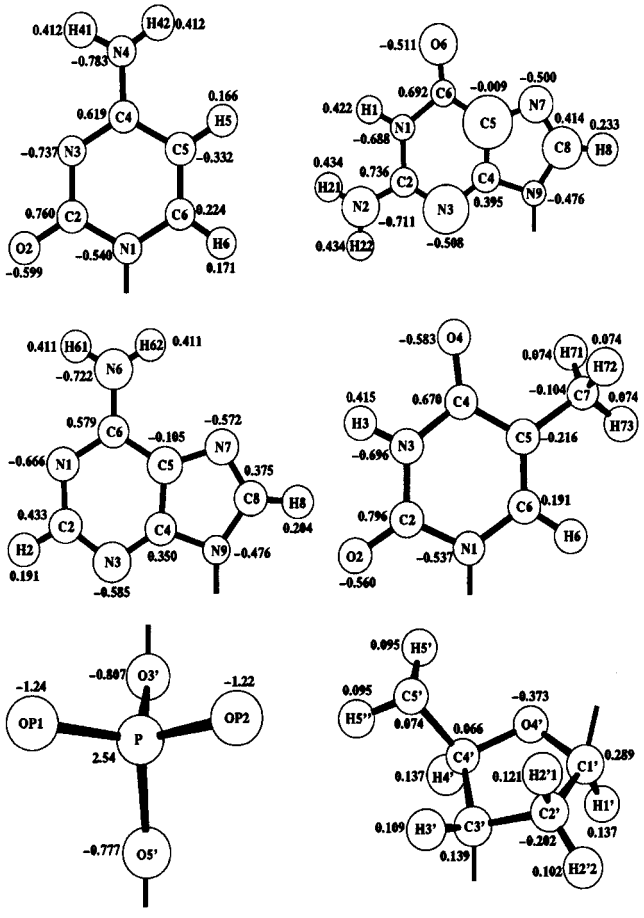

Figure 1. Intramolecular (AM1/CM2) solution charge variations in a canonical B-DNA duplex 72-mer. The average atomic charges are shown adjacent to the atoms. The atomic radii are proportional to the intramolecular RMSD (see text) plus a constant.

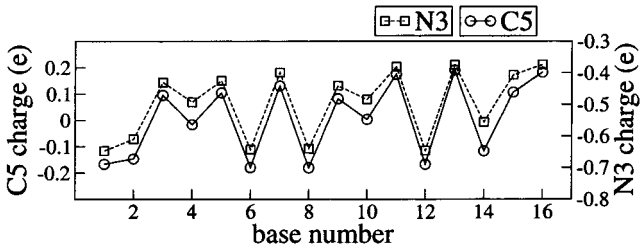

Figure 2. Variation of aqueous atomic charges (AM1/CM2) on C5 and N3 atoms in guanine bases of one strand of the B-DNA 72 mer.

duplex DNA 72-mer d(CGCG-TGTCCGAATACATTCCAG-GCATAAGTGCACCTTCTGAGGGACGACCATGTT ACAT-TGGCGCTGA-CGCG)$_2$ (4574 atoms), and the corresponding A-form of duplex RNA 72-mer (4622 atoms). Canonical A- and B-forms were generated from ideal monomer subunits obtained from fiber diffraction experiments[70−72] and optimized using CHARMM[73,74] with the CHARMM27 all-atom force field for nucleic acids[75] and a distance-dependent dielectric function ($\epsilon = r$) in evaluating nonbond interactions. Hydrogen positions were relaxed with 300 steps steepest descent and 600 steps adopted basis Newton Raphson minimization (keeping non-hydrogen positions fixed), followed by 1000 steps of unrestrained steepest descent minimization of all coordinates. The minimized structures were then used in the linear-scaling semiempirical program. In the linear-scaling semiempirical calculations, the subsystems were chosen to be single nucleotides. Unless otherwise specified, a buffer cutoff $R_b = 8$ Å and the matrix cutoff $R_m = 9$ Å were used. The SCF convergence criteria were set at 10^{-5} kcal/mol. The COSMO calculations were performed with a set of radii specifically parametrized for biomolecules.[30] All calculations were performed on a single processor SGI Origin200 machine with 1 GB memory and 270 MHz clock speed.

TABLE 5: Atom-Based Decomposition of Intramolecular (AM1/CM2) Solution Charge Variations for Canonical B-DNA, A-DNA, and A-RNA 72-mers[a]

	B-DNA mean (RMSD)	A-DNA mean (RMSD)	A-RNA mean (RMSD)
		GUA	
N9	−0.476 (0.333 × 10^{-2})	−0.494 (0.507 × 10^{-2})	−0.500 (0.466 × 10^{-2})
C4	0.395 (0.183 × 10^{-1})	0.391 (0.169 × 10^{-1})	0.394 (0.193 × 10^{-1})
N3	−0.508 (0.990 × 10^{-1})	−0.546 (0.101)	−0.544 (0.104)
C2	0.736 (0.154 × 10^{-1})	0.743 (0.151 × 10^{-1})	0.744 (0.152 × 10^{-1})
N1	−0.688 (0.545 × 10^{-2})	−0.691 (0.726 × 10^{-2})	−0.691 (0.591 × 10^{-2})
H1	0.422 (0.104 × 10^{-1})	0.421 (0.869 × 10^{-2})	0.421 (0.874 × 10^{-2})
N2	−0.711 (0.568 × 10^{-1})	−0.736 (0.560 × 10^{-1})	−0.732 (0.573 × 10^{-1})
H21/H22	0.434 (0.145 × 10^{-1})	0.432 (0.159 × 10^{-1})	0.434 (0.159 × 10^{-1})
C6	0.692 (0.689 × 10^{-2})	0.691 (0.528 × 10^{-2})	0.693 (0.124 × 10^{-1})
O6	−0.511 (0.611 × 10^{-1})	−0.520 (0.577 × 10^{-1})	−0.518 (0.598 × 10^{-1})
C5	−0.944 × 10^{-2} (0.129)	−0.344 × 10^{-1} (0.121)	−0.411 × 10^{-1} (0.124)
N7	−0.500 (0.483 × 10^{-1})	−0.494 (0.415 × 10^{-1})	−0.499 (0.484 × 10^{-1})
C8	0.414 (0.667 × 10^{-1})	0.392 (0.633 × 10^{-1})	0.387 (0.655 × 10^{-1})
H8	0.233 (0.285 × 10^{-1})	0.229 (0.264 × 10^{-1})	0.226 (0.238 × 10^{-1})
mean (RMSD)	*0.239 × 10^{-1} (0.514)*	*0.145 × 10^{-1} (0.520)*	*0.139 × 10^{-1} (0.520)*
		CYT	
N1	−0.540 (0.382 × 10^{-2})	−0.554 (0.347 × 10^{-2})	−0.560 (0.300 × 10^{-2})
C6	0.224 (0.587 × 10^{-2})	0.210 (0.703 × 10^{-2})	0.204 (0.523 × 10^{-2})
H6	0.171 (0.505 × 10^{-2})	0.176 (0.636 × 10^{-2})	0.176 (0.538 × 10^{-2})
C2	0.760 (0.840 × 10^{-2})	0.756 (0.444 × 10^{-2})	0.757 (0.429 × 10^{-2})
O2	−0.599 (0.819 × 10^{-2})	−0.611 (0.535 × 10^{-2})	−0.605 (0.524 × 10^{-2})
N3	−0.737 (0.166 × 10^{-1})	−0.733 (0.207 × 10^{-1})	−0.733 (0.206 × 10^{-1})
C4	0.619 (0.682 × 10^{-2})	0.615 (0.642 × 10^{-2})	0.620 (0.595 × 10^{-2})
N4	−0.783 (0.363 × 10^{-2})	−0.780 (0.462 × 10^{-2})	−0.781 (0.452 × 10^{-2})
H41/H42	0.412 (0.129 × 10^{-1})	0.411 (0.118 × 10^{-1})	0.412 (0.127 × 10^{-1})
C5	−0.332 (0.850 × 10^{-2})	−0.312 (0.837 × 10^{-2})	−0.316 (0.817 × 10^{-2})
H5	0.166 (0.681 × 10^{-2})	0.169 (0.672 × 10^{-2})	0.171 (0.576 × 10^{-2})
mean (RMSD)	*−0.190 × 10^{-1} (0.526)*	*−0.202 × 10^{-1} (0.525)*	*−0.202 × 10^{-1} (0.526)*
		ADE	
C5	−0.105 (0.254 × 10^{-1})	−0.093 (0.373 × 10^{-1})	−0.093 (0.448 × 10^{-1})
N7	−0.572 (0.204 × 10^{-1})	−0.555 (0.213 × 10^{-1})	−0.554 (0.269 × 10^{-1})
C8	0.375 (0.178 × 10^{-1})	0.375 (0.228 × 10^{-1})	0.374 (0.271 × 10^{-1})
H8	0.204 (0.128 × 10^{-1})	0.209 (0.123 × 10^{-1})	0.209 (0.128 × 10^{-1})
N9	−0.476 (0.254 × 10^{-2})	−0.493 (0.303 × 10^{-2})	−0.498 (0.287 × 10^{-2})
N1	−0.666 (0.157 × 10^{-1})	−0.663 (0.140 × 10^{-1})	−0.664 (0.166 × 10^{-1})
C2	0.433 (0.175 × 10^{-1})	0.439 (0.116 × 10^{-1})	0.440 (0.116 × 10^{-1})
H2	0.191 (0.133 × 10^{-1})	0.187 (0.141 × 10^{-1})	0.194 (0.165 × 10^{-1})
N3	−0.585 (0.332 × 10^{-1})	−0.603 (0.428 × 10^{-1})	−0.596 (0.524 × 10^{-1})
C4	0.350 (0.135 × 10^{-1})	0.341 (0.983 × 10^{-2})	0.343 (0.109 × 10^{-1})
C6	0.579 (0.120 × 10^{-1})	0.574 (0.831 × 10^{-2})	0.578 (0.988 × 10^{-2})
N6	−0.772 (0.339 × 10^{-1})	−0.763 (0.372 × 10^{-1})	−0.758 (0.481 × 10^{-1})
H61/H62	0.411 (0.970 × 10^{-2})	0.411 (0.105 × 10^{-1})	0.413 (0.121 × 10^{-1})
mean (RMSD)	*−0.158 × 10^{-1} (0.474)*	*−0.160 × 10^{-1} (0.474)*	*−0.142 × 10^{-1} (0.474)*
		THY/URA	
N1	−0.537 (0.340 × 10^{-2})	−0.553 (0.245 × 10^{-2})	−0.561 (0.213 × 10^{-2})
C6	0.191 (0.573 × 10^{-2})	0.176 (0.489 × 10^{-2})	0.193 (0.418 × 10^{-2})
H6	0.164 (0.493 × 10^{-2})	0.174 (0.563 × 10^{-2})	0.176 (0.433 × 10^{-2})
C2	0.796 (0.843 × 10^{-2})	0.794 (0.649 × 10^{-2})	0.796 (0.543 × 10^{-2})
O2	−0.560 (0.616 × 10^{-2})	−0.580 (0.634 × 10^{-2})	−0.570 (0.538 × 10^{-2})
N3	−0.696 (0.718 × 10^{-2})	−0.693 (0.785 × 10^{-2})	−0.695 (0.833 × 10^{-2})
H3	0.415 (0.443 × 10^{-2})	0.413 (0.518 × 10^{-2})	0.415 (0.566 × 10^{-2})
C4	0.670 (0.629 × 10^{-2})	0.667 (0.787 × 10^{-2})	0.669 (0.574 × 10^{-2})
O4	−0.583 (0.667 × 10^{-2})	−0.579 (0.695 × 10^{-2})	−0.579 (0.562 × 10^{-2})
C5	−0.261 (0.881 × 10^{-2})	−0.241 (0.105 × 10^{-1})	−0.312 (0.721 × 10^{-2})
C7	−0.104 (0.375 × 10^{-2})	−0.100 (0.577 × 10^{-2})	–
H71/H72/H73	0.735 × 10^{-1} (0.102 × 10^{-1})	0.738 × 10^{-1} (0.931 × 10^{-2})	–
H5	–	–	0.178 (0.555 × 10^{-2})
mean (RMSD)	*−0.203 × 10^{-1} (0.450)*	*−0.214 × 10^{-1} (0.451)*	*−0.264 × 10^{-1} (0.514)*
		Sugar	
C5′	0.732 × 10^{-1} (0.312 × 10^{-2})	0.918 × 10^{-1} (0.342 × 10^{-2})	0.943 × 10^{-1} (0.385 × 10^{-2})
H5′	0.947 × 10^{-1} (0.700 × 10^{-2})	0.890 × 10^{-1} (0.102 × 10^{-1})	0.928 × 10^{-1} (0.122 × 10^{-1})
C4′	0.665 × 10^{-1} (0.230 × 10^{-2})	0.589 × 10^{-1} (0.507 × 10^{-2})	0.544 × 10^{-1} (0.477 × 10^{-2})
H4′	0.137 (0.100 × 10^{-1})	0.114 (0.767 × 10^{-2})	0.127 (0.722 × 10^{-2})
O4′	−0.372 (0.632 × 10^{-2})	−0.362 (0.679 × 10^{-2})	−0.344 (0.680 × 10^{-2})
C1′	0.290 (0.974 × 10^{-2})	0.284 (0.112 × 10^{-1})	0.282 (0.104 × 10^{-1})
H1′	0.137 (0.654 × 10^{-2})	0.151 (0.148 × 10^{-1})	0.162 (0.141 × 10^{-1})
C2′	−0.202 (0.548 × 10^{-2})	−0.209 (0.442 × 10^{-2})	0.324 × 10^{-1} (0.392 × 10^{-2})
H2′1	0.121 (0.599 × 10^{-2})	0.116 (0.875 × 10^{-2})	–
H2′2	0.102 (0.703 × 10^{-2})	0.119 (0.109 × 10^{-1})	–
H2′	–	–	0.112 (0.969 × 10^{-2})
O2′	–	–	−0.516 (0.733 × 10^{-2})
O2′H	–	–	0.377 (0.291 × 10^{-2})
C3′	0.140 (0.231 × 10^{-2})	0.104 (0.717 × 10^{-2})	0.541 × 10^{-1} (0.686 × 10^{-2})
H3′	0.109 (0.663 × 10^{-2})	0.128 (0.385 × 10^{-2})	0.136 (0.299 × 10^{-2})
mean (RMSD)	*0.608 × 10^{-1} (0.161)*	*0.594 × 10^{-1} (0.159)*	*0.541 × 10^{-1} (0.219)*

TABLE 5: Continued

		Phosphate	
P	2.54 (0.466×10^{-2})	2.51 (0.488×10^{-2})	2.50 (0.238×10^{-2})
OP1	−1.24 (0.375×10^{-2})	−1.19 (0.856×10^{-2})	−1.22 (0.681×10^{-2})
OP2	−1.21 (0.520×10^{-2})	−1.23 (0.663×10^{-2})	−1.18 (0.413×10^{-2})
O5′	−0.777 (0.558×10^{-2})	−0.773 (0.470×10^{-2})	−0.780 (0.471×10^{-2})
O3′	−0.807 (0.305×10^{-2})	−0.778 (0.387×10^{-2})	−0.774 (0.403×10^{-2})
mean (RMSD)	*−0.299 (1.43)*	*−0.293 (1.42)*	*−0.291 (0.141)*

[a] This table lists the atom-based decomposition of intramolecular AM1/CM2 solution charge variations for DNA and RNA duplex 72-mers. The mean and RMSD are the *intramolecular mean and RMSD* for a particular atom in a nucleic acid subunit, taken over the distribution of all identical subunits. Atoms are named according to the standard nomenclature.[77]

3. Results and Discussion

3.1. Convergence of Atomic Charges with the D&C Method. It is important to verify with any new method the convergence behavior of the properties being calculated. Previous work has demonstrated that for solvated proteins and nucleic acids the electronic energy, solvation energy, and electronic density of states are tightly converged using a 8/9 Å R_b/R_m scheme.[65] The emphasis here is on the convergence of atomic charges both in the gas phase and in a solution. For the purposes of discussion, the following notation is introduced:

$$(\boldsymbol{\delta})\mathbf{q}^{\mathrm{AM1/PM3}}_{\mathrm{Mul/CM2}}(\mathrm{molecule}) \qquad (9)$$

where the superscript designates the semiempirical Hamiltonian (AM1 or PM3), the subscript designates the type of charges calculated (Mulliken or CM2), and the molecule is indicated in parentheses. A set of charges is designated by a vector (shown in boldface type): **q** for the *solution charge vector* and $\boldsymbol{\delta}\mathbf{q}$ for the *solvent-induced polarization charge vector* (defined as the difference between the solution-phase and gas-phase charge vectors). The former sum to the total molecular charge whereas the latter sum to zero.

Table 1 shows the convergence of the root-mean-square deviation (RMSD) of the atomic charge vectors for a canonical B-DNA (B-DNA_{10}) decamer using different R_b/R_m schemes. Both solution-phase and solvent-induced polarization charges are shown for the Mulliken and CM2 partitioning schemes as well as AM1 and PM3 semiempirical Hamiltonians. The charges show rapid convergence with increasing R_b/R_m schemes. With the 6/7 Å scheme, the solution and solvent-induced polarization charges converge to roughly 10^{-4} and 10^{-3} au, respectively. These values are well within an acceptable error range for most molecular modeling applications. The RMSD is reduced by over half an order of magnitude in going from the 6/7 Å to 8/9 Å scheme. The RMSD for solution charge vectors **q** is typically an order of magnitude smaller than for the solvent-induced polarization charge vectors $\boldsymbol{\delta}\mathbf{q}$. This is due to the inherent instability of the electronic structure of the highly charged polyanions in the gas phase. The convergence with AM1 and PM3 Hamiltonians is similar, although for the B-DNA_{10} molecule, the RMSD values of the AM1 charge vectors are around a factor of 2 smaller than those of the PM3 charge vectors. Very little difference was observed between the convergence behavior of the Mulliken and CM2 charge vectors with the same R_b/R_m cutoff schemes. In the subsequent part of this work, a 8/9 Å R_b/R_m scheme is used.

3.2. Effect of the Charge Model (CM2 vs Mulliken). It is expected that the CM2 model provides a more realistic description of the charge distribution than the Mulliken charge set does, and hence more suitable for modeling applications. Nonetheless, it is instructive to quantify the differences between the Mulliken and CM2 charge models for large solvated nucleic acids since this is one of the first applications of its kind to these systems.

Table 2 compares the CM2 and Mulliken solution charges and solvent-induced polarization charges for A- and B-form DNA and RNA decamers. The mean and RMSD of a charge difference vector Δx, defined as the difference between CM2 and Mulliken charges sets, are reported in terms of the base, sugar, and phosphate contributions. The value of the "Base mean", for example, is the average difference between the CM2 and Mulliken charges per base atom (so for B-DNA, a base atom will have, on average, a CM2 charge that is 0.00545 au *less* than the corresponding Mulliken charge using the AM1 Hamiltonian). The "base RMSD" is the corresponding root-mean-square deviation. Overall, the trends in the mean and RMSD between the CM2 and Mulliken solution charge sets are consistent for the different forms of DNA and RNA.

Since the CM2 model empirically corrects the bond dipole but leaves the total charge conserved, the *net* effect (reflected by the mean) is due primarily to the atoms at the boundary of the nucleic acid subunits. This provides insight, therefore, into the polarity of the covalent bonds between the sugar and base, and between the phosphate and sugars (these are explored in more detail below). The base and phosphate groups are observed to have slightly more negative charge and the sugars correspondingly less negative charge with the CM2 model for both AM1 and PM3 Hamiltonians. The CM2 charge model predicts more polar C5′−N9 (χ), C3′−O3′ (ϵ), and C5′−O5′ (β) bonds that favor charge transfer to the nitrogens and oxygens and accounts for the overall small charge transfer from sugar to base and sugar to phosphate.

Table 2 shows that the CM2 corrections to the aqueous Mulliken charges are the largest for the base and phosphate atoms: around 0.26 au (AM1) and 0.23 au (PM3) for base atoms and 0.11 au (AM1) and 0.27 au (PM3) for phosphate atoms. The atoms of the sugar subunits show the smallest RMSD between CM2 and Mulliken solution charges for both AM1 and PM3, falling in the range of 0.03−0.08 au. This is not surprising since the average bond order and bond polarity in a sugar subunit is smaller than that of a phosphate or base subunit, and therefore the charges on sugar atoms are least affected by the CM2 mapping procedure. For base and phosphate subunits, however, the differences are significant. Consequently, the application of CM2 corrections to the aqueous Mulliken charges may be quite important in modeling applications of nucleic acids. The observed deviations between the CM2 and Mulliken solution charge sets do not vary significantly between the different forms of DNA and RNA.

The mean and RMSD of the CM2 and Mulliken solvent-induced polarization charges are also shown in Table 2. The overall statistical quantities are roughly an order of magnitude smaller than the corresponding values for the solution charge sets. The observed trends are more strongly correlated between different structural forms (A and B) than between DNA and

TABLE 6: Decomposition of Intramolecular (AM1/CM2) Solvent-Induced Polarization Charge Variations for Canonical B-DNA, A-DNA, and A-RNA 72-mers[a]

	B-DNA mean (RMSD)	A-DNA mean (RMSD)	A-RNA mean (RMSD)
		GUA	
N9	-0.353×10^{-2} (0.421×10^{-2})	0.981×10^{-2} (0.248×10^{-2})	0.110×10^{-1} (0.340×10^{-2})
C4	-0.203×10^{-2} (0.176×10^{-1})	0.208×10^{-1} (0.454×10^{-1})	0.100×10^{-1} (0.384×10^{-1})
N3	-0.201×10^{-2} (0.114)	0.570×10^{-1} (0.763×10^{-1})	0.333×10^{-1} (0.815×10^{-1})
C2	-0.181×10^{-2} (0.867×10^{-2})	-0.237×10^{-2} (0.131×10^{-1})	-0.484×10^{-2} (0.108×10^{-1})
N1	-0.103×10^{-2} (0.463×10^{-2})	-0.850×10^{-2} (0.671×10^{-2})	-0.763×10^{-2} (0.701×10^{-2})
H1	-0.906×10^{-4} (0.840×10^{-2})	0.938×10^{-3} (0.268×10^{-2})	-0.188×10^{-3} (0.331×10^{-2})
N2	-0.431×10^{-2} (0.686×10^{-1})	0.478×10^{-1} (0.424×10^{-1})	0.174×10^{-1} (0.215×10^{-1})
H21/H22	0.583×10^{-2} (0.157×10^{-1})	0.187×10^{-1} (0.216×10^{-1})	0.174×10^{-1} (0.215×10^{-1})
C6	0.256×10^{-2} (0.448×10^{-2})	0.719×10^{-3} (0.478×10^{-2})	0.134×10^{-2} (0.316×10^{-2})
O6	-0.208×10^{-1} (0.659×10^{-1})	-0.424×10^{-1} (0.347×10^{-1})	-0.473×10^{-1} (0.404×10^{-1})
C5	-0.231×10^{-2} (0.659×10^{-1})	-0.262×10^{-1} (0.842×10^{-1})	-0.418×10^{-1} (0.906×10^{-1})
N7	-0.322×10^{-1} (0.498×10^{-1})	-0.739×10^{-1} (0.324×10^{-1})	-0.789×10^{-1} (0.379×10^{-1})
C8	0.844×10^{-2} (0.806×10^{-1})	-0.105×10^{-1} (0.500×10^{-1})	-0.194×10^{-1} (0.521×10^{-1})
H8	0.103×10^{-2} (0.254×10^{-1})	-0.947×10^{-2} (0.723×10^{-2})	-0.132×10^{-1} (0.803×10^{-2})
mean (RMSD)	*-0.575×10^{-2} (0.603×10^{-1})*	*0.896×10^{-4} (0.502×10^{-1})*	*-0.562×10^{-2} (0.509×10^{-1})*
		CYT	
N1	-0.581×10^{-2} (0.234×10^{-2})	0.300×10^{-2} (0.229×10^{-2})	0.166×10^{-2} (0.213×10^{-2})
C6	-0.166×10^{-2} (0.861×10^{-2})	0.397×10^{-2} (0.869×10^{-2})	-0.562×10^{-3} (0.847×10^{-2})
H6	-0.105×10^{-1} (0.854×10^{-2})	-0.118×10^{-1} (0.439×10^{-2})	-0.119×10^{-1} (0.399×10^{-2})
C2	0.134×10^{-2} (0.461×10^{-2})	0.219×10^{-2} (0.228×10^{-2})	0.169×10^{-2} (0.249×10^{-2})
O2	-0.133×10^{-1} (0.828×10^{-2})	0.288×10^{-1} (0.886×10^{-2})	0.217×10^{-1} (0.882×10^{-2})
N3	-0.241×10^{-2} (0.203×10^{-1})	0.291×10^{-2} (0.772×10^{-2})	0.591×10^{-2} (0.854×10^{-2})
C4	0.125×10^{-3} (0.581×10^{-2})	-0.553×10^{-2} (0.394×10^{-2})	-0.353×10^{-2} (0.368×10^{-2})
N4	0.125×10^{-1} (0.360×10^{-2})	-0.822×10^{-2} (0.544×10^{-2})	-0.387×10^{-2} (0.494×10^{-2})
H41	0.105×10^{-1} (0.153×10^{-1})	0.469×10^{-4} (0.363×10^{-2})	0.152×10^{-2} (0.443×10^{-2})
C5	0.105×10^{-1} (0.826×10^{-2})	-0.363×10^{-1} (0.968×10^{-2})	-0.298×10^{-1} (0.942×10^{-2})
H5	0.198×10^{-1} (0.758×10^{-2})	-0.130×10^{-1} (0.253×10^{-2})	-0.106×10^{-1} (0.255×10^{-2})
mean (RMSD)	*0.140×10^{-2} (0.152×10^{-1})*	*-0.283×10^{-2} (0.156×10^{-1})*	*-0.219×10^{-2} (0.131×10^{-1})*
		ADE	
C5	0.500×10^{-3} (0.196×10^{-1})	-0.140×10^{-1} (0.112×10^{-1})	-0.906×10^{-2} (0.184×10^{-1})
N7	-0.286×10^{-1} (0.160×10^{-1})	-0.632×10^{-1} (0.131×10^{-1})	-0.589×10^{-1} (0.165×10^{-1})
C8	-0.134×10^{-2} (0.166×10^{-1})	-0.110×10^{-1} (0.111×10^{-1})	-0.114×10^{-1} (0.146×10^{-1})
H8	-0.241×10^{-2} (0.906×10^{-2})	-0.100×10^{-1} (0.444×10^{-2})	-0.114×10^{-1} (0.502×10^{-2})
N9	-0.228×10^{-2} (0.385×10^{-2})	0.925×10^{-2} (0.168×10^{-2})	0.856×10^{-2} (0.206×10^{-2})
N1	0.331×10^{-2} (0.167×10^{-1})	-0.156×10^{-2} (0.723×10^{-2})	-0.625×10^{-4} (0.104×10^{-1})
C2	0.700×10^{-2} (0.135×10^{-1})	0.928×10^{-2} (0.383×10^{-2})	0.104×10^{-1} (0.562×10^{-2})
H2	-0.256×10^{-2} (0.101×10^{-1})	0.203×10^{-1} (0.554×10^{-2})	0.217×10^{-1} (0.724×10^{-2})
N3	-0.322×10^{-1} (0.348×10^{-1})	0.184×10^{-1} (0.137×10^{-1})	0.875×10^{-2} (0.272×10^{-1})
C4	0.394×10^{-2} (0.799×10^{-2})	0.247×10^{-1} (0.847×10^{-2})	0.212×10^{-1} (0.960×10^{-2})
C6	0.531×10^{-2} (0.839×10^{-2})	-0.813×10^{-2} (0.572×10^{-2})	-0.425×10^{-2} (0.569×10^{-2})
N6	0.328×10^{-2} (0.282×10^{-1})	-0.138×10^{-1} (0.945×10^{-2})	-0.909×10^{-2} (0.252×10^{-1})
H61/H62	0.600×10^{-2} (0.954×10^{-2})	-0.305×10^{-2} (0.377×10^{-2})	-0.175×10^{-2} (0.515×10^{-2})
mean (RMSD)	*-0.243×10^{-2} (0.205×10^{-1})*	*-0.329×10^{-2} (0.224×10^{-1})*	*-0.265×10^{-2} (0.233×10^{-1})*
		THY/URA	
N1	0.163×10^{-2} (0.416×10^{-2})	0.331×10^{-2} (0.199×10^{-2})	0.359×10^{-2} (0.291×10^{-2})
C6	-0.403×10^{-2} (0.797×10^{-2})	0.330×10^{-1} (0.862×10^{-2})	0.237×10^{-1} (0.975×10^{-2})
H6	-0.825×10^{-2} (0.705×10^{-2})	-0.878×10^{-2} (0.352×10^{-2})	-0.881×10^{-2} (0.357×10^{-2})
C2	0.138×10^{-1} (0.965×10^{-2})	0.101×10^{-1} (0.301×10^{-2})	0.106×10^{-1} (0.376×10^{-2})
O2	-0.210×10^{-1} (0.114×10^{-1})	0.194×10^{-1} (0.867×10^{-2})	0.137×10^{-1} (0.916×10^{-2})
N3	0.744×10^{-2} (0.926×10^{-2})	0.541×10^{-2} (0.450×10^{-2})	0.822×10^{-2} (0.547×10^{-2})
H3	0.175×10^{-2} (0.337×10^{-2})	0.297×10^{-2} (0.186×10^{-2})	0.384×10^{-2} (0.248×10^{-2})
C4	0.116×10^{-1} (0.807×10^{-2})	0.928×10^{-2} (0.283×10^{-2})	0.825×10^{-2} (0.396×10^{-2})
O4	-0.135×10^{-1} (0.121×10^{-1})	-0.383×10^{-1} (0.114×10^{-1})	-0.373×10^{-1} (0.108×10^{-1})
C5	0.797×10^{-2} (0.126×10^{-1})	-0.390×10^{-1} (0.812×10^{-2})	-0.415×10^{-1} (0.116×10^{-1})
C7	-0.609×10^{-2} (0.244×10^{-2})	0.453×10^{-2} (0.293×10^{-2})	–
H71/H72/H73	0.473×10^{-2} (0.108×10^{-1})	-0.145×10^{-1} (0.997×10^{-2})	–
H5	–	–	-0.156×10^{-1} (0.397×10^{-2})
mean (RMSD)	*0.397×10^{-3} (0.132×10^{-1})*	*-0.297×10^{-2} (0.207×10^{-1})*	*-0.285×10^{-2} (0.211×10^{-1})*
		Sugar	
C5′	-0.127×10^{-1} (0.217×10^{-2})	-0.151×10^{-1} (0.122×10^{-2})	-0.165×10^{-1} (0.952×10^{-3})
H5′	0.227×10^{-1} (0.763×10^{-2})	0.291×10^{-1} (0.326×10^{-1})	0.289×10^{-1} (0.333×10^{-1})
C4′	0.154×10^{-2} (0.117×10^{-2})	0.716×10^{-2} (0.182×10^{-2})	0.195×10^{-2} (0.139×10^{-2})
H4′	0.320×10^{-1} (0.666×10^{-2})	0.630×10^{-1} (0.647×10^{-2})	0.613×10^{-1} (0.621×10^{-2})
O4′	-0.995×10^{-2} (0.110×10^{-1})	0.113×10^{-1} (0.629×10^{-2})	0.924×10^{-2} (0.582×10^{-2})
C1′	0.544×10^{-2} (0.692×10^{-2})	-0.830×10^{-2} (0.390×10^{-2})	-0.129×10^{-1} (0.349×10^{-2})
H1′	0.898×10^{-3} (0.583×10^{-2})	0.481×10^{-1} (0.946×10^{-2})	0.406×10^{-1} (0.905×10^{-2})
C2′	-0.783×10^{-2} (0.303×10^{-2})	-0.105×10^{-2} (0.213×10^{-2})	0.107×10^{-1} (0.227×10^{-2})
H2′2	0.117×10^{-1} (0.471×10^{-2})	-0.166×10^{-1} (0.370×10^{-2})	–
H2′1	-0.465×10^{-2} (0.658×10^{-2})	0.526×10^{-1} (0.668×10^{-2})	–
H2′	–	–	-0.135×10^{-1} (0.465×10^{-2})
O2′	–	–	0.249×10^{-1} (0.710×10^{-2})
O2′H	–	–	0.117×10^{-1} (0.397×10^{-2})
C3′	-0.834×10^{-2} (0.223×10^{-2})	-0.115×10^{-1} (0.314×10^{-2})	-0.746×10^{-2} (0.293×10^{-2})
H3′	0.663×10^{-2} (0.652×10^{-2})	-0.414×10^{-1} (0.447×10^{-2})	-0.370×10^{-1} (0.440×10^{-2})
mean (RSMD)	*0.463×10^{-2} (0.149×10^{-1})*	*0.114×10^{-1} (0.329×10^{-1})*	*0.935×10^{-2} (0.285×10^{-1})*

TABLE 6: Continued

		Phosphate	
P	$0.176\ (0.621 \times 10^{-2})$	$0.154\ (0.102 \times 10^{-1})$	$0.155\ (0.876 \times 10^{-2})$
OP1	$-0.118\ (0.963 \times 10^{-2})$	$-0.231\ (0.533 \times 10^{-1})$	$-0.118\ (0.199 \times 10^{-1})$
OP2	$-0.118\ (0.614 \times 10^{-2})$	$-0.118\ (0.226 \times 10^{-1})$	$-0.200\ (0.454 \times 10^{-1})$
O5′	$-0.533 \times 10^{-2}\ (0.548 \times 10^{-2})$	$-0.222 \times 10^{-1}\ (0.589 \times 10^{-2})$	$-0.239 \times 10^{-1}\ (0.754 \times 10^{-2})$
O3′	$-0.666 \times 10^{-2}\ (0.246 \times 10^{-2})$	$0.197 \times 10^{-1}\ (0.519 \times 10^{-2})$	$0.226 \times 10^{-1}\ (0.480 \times 10^{-2})$
mean (RMSD)	*-0.145×10^{-1} (0.108)*	*-0.396×10^{-1} (0.132)*	*-0.324×10^{-1} (0.124)*

[a] This table lists the atom-based decomposition of intramolecular AM1/CM2 solvent-induced polarization charge variations for DNA and RNA duplex 72-mers. The mean and RMSD are the *intramolecular mean and RMSD* for a particular atom in a nucleic acid subunit, taken over the distribution of all identical subunits.

RNA. As shown in greater detail below, the charge distributions of DNA and RNA are in general similar, especially within the same form. The solvation effect, and therefore the solvent-induced polarization charges, are determined primarily from the structural forms that dictate, for example, which atoms are more exposed to solvent. The CM2 and Mulliken charge vectors show a high degree of correlation (linear correlation coefficients range from 0.94 to 0.99, data not shown here). The correlation is more pronounced with the AM1 Hamiltonian than with the PM3 Hamiltonian.

3.3. Effect of Hamiltonian (AM1 vs PM3). Table 3 compares the AM1 and PM3 solution charges and solvent-induced polarization charges for A- and B-form DNA and RNA decamers. The mean and RMSD of a charge difference vector Δx, defined as the difference between AM1 and PM3 charge sets, are reported in terms of the base, sugar, and phosphate contributions. For solution charges, RMSD values show that AM1 and PM3 charges are most different for phosphate and base atoms, which is mainly a result of the large difference between AM1 and PM3 charges on nitrogen and phosphorus atoms. The RMSD of the difference vectors in Table 3 indicate that AM1 and PM3 charges are similar for base and sugar atoms for both CM2 and Mulliken charge models. However, for phosphate atoms, the RMSD values are larger with the CM2 charges than with the Mulliken charges (RMSD of 0.42 and 0.26 au for CM2 and Mulliken charge difference vectors, respectively). This is a consequence of the much larger CM2 correction on the PM3 charge of the phosphorus atom compared to that on the AM1 charge, in agreement with the consensus that AM1 method is more reliable for modeling experimental dipole moments and ionization potentials for a variety of compounds.[76]

Overall, the trend in AM1 and PM3 difference charge vectors for base, sugar, and phosphate atoms resembles that of the CM2 and Mulliken difference vectors (see Table 2), implying some general statistical similarities between the difference in the model Hamiltonians and the CM2 charge correction for these systems. This is in accord with the fact that both the AM1 method and CM2 correction scheme better reproduce experimental dipole moments compared with the PM3 method and Mulliken scheme, respectively.

The RMSD values for the solvent-induced polarization charges $\mathbf{\Delta x} = \boldsymbol{\delta}\mathbf{q}^{\mathrm{AM1}} - \boldsymbol{\delta}\mathbf{q}^{\mathrm{PM3}}$ range from 0.004 to 0.04 au, with the largest RMSD values for phosphates, followed by those for the bases and sugars. The RMSD values for the charge difference vectors are mainly affected by (1) the difference in solvent reaction fields induced by the solution charges and (2) the difference in the atomic polarizabilities (or more accurately, the difference in the nonlinear response functions). In general, the differences between the charges derived from the AM1 and PM3 Hamiltonians show no strong influence by the form (A or B) of the DNA and RNA.

3.4. Inter- and Intramolecular Charge Variations. This section addresses questions as to (1) the degree to which atomic charges vary *between different molecules* of DNA and RNA (intermolecular) and (2) the degree to which atomic charges vary *within a given molecule* (intramolecular). To increase the statistical precision of the results, especially for the intramolecular charge variations, 72-mer sequences of canonical B-DNA, A-DNA, and A-RNA were constructed, and linear-scaling calculations were performed in the same manner as for the decamers. Note, the 72-mers have the form [CGCG-(semi-random 64-mer sequence)-CGCG]$_2$, where the "semirandom" sequence contains 16 of each of the 4 base types. The analysis has been performed only on these central 128 nucleotides, the 4 residues on the 3′ and 5′ ends of both strands of the duplex were discarded in the analysis in order to minimize the influence of end effects. For brevity, discussion is restricted to results using the AM1 Hamiltonian and the CM2 charge correction.

3.4.1. Intermolecular Charge Variations. Table 4 compares the AM1/CM2 solution charges ($\mathbf{q}_{\mathrm{CM2}}^{\mathrm{AM1}}$) and solvent-induced polarization charges ($\boldsymbol{\delta}\mathbf{q}_{\mathrm{CM2}}^{\mathrm{AM1}}$) between different nucleic acid forms. The average phosphate solution charge for B-DNA is slightly more negative than for A-DNA and A-RNA by -0.006 and -0.008 au, respectively. This is consistent with the fact that the B-DNA helix is narrower and therefore the negative charges tend to reside more on the backbone phosphate groups to minimize repulsion. This effect is even more pronounced in the gas phase. The RMSD for the phosphate solution charges is around 0.03 au between any of the forms. The largest difference in the solution charges occurs between DNA and RNA in the sugars. This is due to the presence of the 2′ OH group in the RNA that is not present in DNA. (For the purpose of analysis, the charges of the O2′ and H2′ atoms in A-RNA are added to form a "united atom" charge that is compared to the charge of the H2′ atom in the DNA.) The bases have the next largest RMSD between forms, particularly between B-DNA and A-RNA (0.04 au). The largest change in the solvent-induced polarization charges occurs at the phosphates. The overall polarization charge at the phosphates is less negative in B-DNA than in A-DNA and A-RNA, even though the solution charges show the opposite trend. This is because the gas phase phosphate charge in the more narrow B-DNA helix is more negatively charged than for the A helices (both in the gas phase and in solution). In solution, all of the phosphates become more negative due to solvent stabilization; however, it has the largest effect on the A helices since the charges are less negative in the gas phase. This effect is counterbalanced by the fact that the solvent reaction field becomes smaller in regions of less charge. The result is that the overall effect is moderate: the RMSD between different forms is in the range 0.04−0.07 au.

3.4.2. Intramolecular Charge Variations. *Solution Charges.* Table 5 displays the intramolecular mean and RMSD values

for $\mathbf{q}_{CM2}^{AM1}$ broken down into individual base, sugar, and phosphate atom contributions for the central 64 residues of the 72-mer duplex sequence of B-DNA, A-DNA, and A-RNA. In Figure 1 the intramolecular AM1/CM2 charge fluctuations are presented for B-DNA. For all three nucleic acids, the intramolecular RMSD values on sugar and phosphates are small (less than 0.02 au), suggesting that they are relatively insensitive to the base stacking environment. The same can be seen for the pyrimidine bases cytosine (CYT), thymine (THY), and uracil (URA), where the RMSD values are typically on the order of 10^{-3} au. In contrast, larger fluctuations are observed for the purine bases guanine (GUA) and adenine (ADE). In particular, for guanine N3 in B-DNA, A-DNA, and A-RNA, the RMSD is 0.099, 0.101, and 0.104 au, respectively, and for guanine C5 the RMSD is 0.129, 0.121, and 0.124 au, respectively. The charge variations of these atoms within the same base are highly correlated (Figure 2). This correlation was not observed in the gas phase, although it should be made clear that the gas phase is a highly artificial state for nucleic acids. The N3 and C5 atoms are separated by 2 bonds in the conjugated π system of the guanine base. The AM1 model predicts that guanine has the highest occupied molecular orbital (HOMO) energy level among the bases and also has the smallest energy gap between the HOMO and lowest unoccupied molecular orbital (LUMO) energy level. Consequently, it is the guanine energy states that lie closest to the Fermi level and are the most sensitive to variations in charge state (similar to saying that they are chemically the "softest") as a function of the local solvation and base stacking environment.

Solvent-Induced Polarization Charges. Previous work has shown that solvation has a great impact on dipole moments[56] and electrostatic potential distribution on biological molecules.[23] Table 6 displays the mean and RMSD values of $\boldsymbol{\delta}\mathbf{q}_{CM2}^{AM1}$ for individual base, sugar, and phosphate atoms in a 72-mer duplex sequence of B-DNA, A-DNA, and A-RNA. The magnitude of mean and RMSD for base and sugar groups for A- and B-DNA and A-RNA are in the range $10^{-2}-10^{-3}$ au and are typically lower than 1% of the corresponding solution atomic charges (compare with Table 5), suggesting solvent polarization has a fairly small effect on the charges. The largest polarization occurs at the phosphate P, OP1, and OP2 atoms where the changes in atomic charges are around 0.15 au. This might be expected since the solvent reaction field is largest at the phosphates.

4. Conclusion

This paper describes a systematic study of the charge distributions of solvated DNA and RNA duplexes in canonical A- and B-forms using recently developed linear-scaling electronic structure methods. The charges derived from the linear-scaling electronic structure method rapidly converge with matrix and buffer cutoffs and provide high accuracy (better than 10^{-5} au), well beyond that needed for most biological modeling applications (around 10^{-3} au). The calculation of the CM2 charge correction to the semiempirical Mulliken (Löwdin) charges provides a tractable computational prescription for deriving high-quality macromolecular charges. These charges reflect the electronic relaxation due to the solvated macromolecular environment and may be a better starting point for modeling applications such as QSAR studies, electrostatic potential characterization, and the prediction of pK_a shifts and ligand binding free energies. Intermolecular charge variations (between A- and B-form DNA and RNA) are observed to be fairly small; i.e., the conformational dependence in general does not have a large effect on the charges. Not surprisingly, the 2′ OH functional group substitution in RNA had a larger effect. Intramolecular charge variations are observed to be small for most of the atoms on sugar, phosphate, and pyrimidine bases. The largest charge fluctuations occur in guanine bases, particularly at the N3 and C5 positions, and are related to the high-lying HOMO energy and small HOMO−LUMO gap in guanine that make it particularly sensitive to base stacking and the solvent environment. Solvation has a significant effect, reflected by the solvent-induced polarization charges, in regions of large charge concentration (such as the phosphates) where the solvent reaction field is large. It is expected that charge variation will become more significant in more complex heterogeneous environments such as that of DNA and RNA-binding proteins where polarization and charge transfer occur upon binding. This is a topic of future work.

Acknowledgment. D.Y. is grateful for financial support provided by the National Institutes of Health (Grant 1R01-GM62248-01A1), and the donors of the Petroleum Research Fund, administered by the American Chemical Society. J.K. was partially supported by the Louise T. Dosdall Fellowship from the University of Minnesota. Computational resources were provided by the Minnesota Supercomputing Institute.

References and Notes

(1) Gilson, M. K. *Curr. Opin. Struct. Biol.* **1995**, *5*, 216−223.
(2) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(3) Gunner, M. R.; Nicholls, A.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 4277−4291.
(4) Ullmann, G. M.; Hauswald, M.; Jensen, A.; Kostic, N. M.; Knapp, E.-W. *Biochemistry* **1997**, *36*, 16187−16196.
(5) Chin, K.; Sharp, K. A.; Honig, B.; Pyle, A. M. *Nature Struct. Biol.* **1999**, *6*, 1055−1061.
(6) Kangas, E.; Tidor, B. *J. Chem. Phys.* **2000**, *112*, 9120−9131.
(7) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
(8) Ghafourian, T.; Dearden, J. *J. Pharmacy Pharmacol.* **2000**, *52*, 603−610.
(9) Gancia, E.; Bravi, G.; Mascagni, P.; Zaliani, A. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 293−306.
(10) Lee, L.-P.; Tidor, B. *Nature Struct. Biol.* **2001**, *8*, 73−76.
(11) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047−1064.
(12) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730−4737.
(13) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141−6156.
(14) York, D.; Yang, W. *J. Chem. Phys.* **1996**, *104*, 159−172.
(15) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741−754.
(16) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.
(17) Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213−13225.
(18) Yang, W.; Pérez-Jordá, J. M. Linear scaling methods for electronic structure calculation. In *Encyclopedia of Computational Chemistry*; von Schleyer, P., Ed.; John Wiley and Sons: New York, 1998.
(19) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085−1123.
(20) Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*, 4782−4790.
(21) Gogonea, V.; Suárez, D.; van der Vaart, A.; Merz Jr, K. M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 217−213.
(22) Galli, G. *Phys. Status Solidi B* **2000**, *217*, 231−249.
(23) Khandogin, J.; Hu, A.; York, D. M. *J. Comput. Chem.* **2000**, *21*, 1562−1571.
(24) Khandogin, J.; York, D. M. Manuscript in preparation.
(25) Stewart, J. J. P. *Rev. Comput. Chem.* **1990**, *1*, 45−81.
(26) Zerner, M. C. *Rev. Comput. Chem.* **1991**, *2*, 313−365.
(27) Thiel, W. Perspectives on semiempirical molecular orbital theory. In *Advances in Chemistry and Physics*, Vol. 93; Prigogine, I., Rice, S. A., Eds.; John Wiley and Sons: New York, 1996.
(28) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1996**, *104*, 6643−6649.
(29) Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674−5678.
(30) York, D. M.; Lee, T.-S.; Yang, W. *Chem. Phys. Lett.* **1996**, *263*, 297−306.
(31) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133−146.
(32) York, D.; Lee, T.-S.; Yang, W. *J. Am. Chem. Soc.* **1996**, *118*, 10940−10941.

(33) Lewis, J.; Liu, S.; Lee, T.-S.; Yang, W. *J. Comput. Phys.* **1999**, *151*, 242−263.
(34) Titmuss, S. J.; Cummins, P. L.; Bliznyuk, A. A.; Rendell, A. P.; Gready, J. E. *Chem. Phys. Lett.* **2000**, *320*, 169−176.
(35) van der Vaart, A.; Valentin, G.; Dixon, S. I.; Merz, K. M., Jr. *J. Comput. Chem.* **2000**, *21*, 1494−1504.
(36) Liu, H.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. *Proteins* **2001**, *44*, 484−489.
(37) Goedecker, S.; Colombo, L. *Phys. Rev. Lett.* **1994**, *73*, 122−125.
(38) Goedecker, S.; Teter, M. *Phys. Rev. B* **1995**, *51*, 9455−9464.
(39) Li, X.-P.; Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 10891−10894.
(40) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129−145.
(41) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431−439.
(42) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894−905.
(43) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(44) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(45) Luque, F.; Orozco, M. *J. Comput. Chem.* **1990**, *11*, 909−914.
(46) Ferenczy, G. G.; Reynolds, C. A.; Richards, W. G. *J. Comput. Chem.* **1990**, *11*, 159−164.
(47) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1995**, *102*, 7549−7556.
(48) Francl, M. M.; Carey, C.; E., C. L.; Gange, D. M. *J. Comput. Chem.* **1996**, *17*, 367−383.
(49) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(50) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1841−1846.
(51) Löwdin, P. O. *J. Chem. Phys.* **1950**, *18*, 365−375.
(52) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735−746.
(53) Bader, R. F. W. *Acc. Chem. Res.* **1985**, *18*, 9−15.
(54) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87−95.
(55) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.
(56) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chim. Phys.* **1997**, *94*, 1448−1481.
(57) Kaminski, G. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787−1796.
(58) Li, J.; Cramer, C. J.; Truhlar, D. G. *Biophys. Chem.* **1999**, *78*, 147−155.
(59) Gogonea, V.; Merz, K., Jr. *J. Phys. Chem. A* **1999**, *103*, 5171−5188.
(60) Dewar, M. J. S.; Zoebisch, E.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(61) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.
(62) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744−2750.
(63) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799−805.
(64) York, D. M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 11060−11079.
(65) York, D. M.; Lee, T.-S.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 5011−5014.
(66) York, D. M. Application of linear-scaling electronic structure methods to the study of polarization of proteins and DNA in solution. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M., Eds.; ACS Symposium Series 712; Oxford University Press: New York, 1998.
(67) Pérez-Jordá, J. M.; Yang, W. *Chem. Phys. Lett.* **1995**, *247*, 484−490.
(68) Pérez-Jordá, J.; Yang, W. *J. Chem. Phys.* **1996**, *104*, 8003−8006.
(69) Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons: Chichester, England, 1999.
(70) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504−1509.
(71) Arnott, S.; Hukins, D. W. L.; Dover: S. D. *Biochem. Biophys. Res. Commun.* **1972**, *48*, 1392−1399.
(72) Arnott, S.; Selsing, E. *J. Mol. Biol.* **1974**, *88*, 509−521.
(73) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(74) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parametrization with an Overview of the Program. In *Encyclopedia of Computational Chemistry*, Vol. 1; Schleyer, P., Allinger, N., Clark, T., Gasteiger, J., Kollman, P., Schaefer, H., III, Schreiner, P., Eds.; John Wiley & Sons: Chichester, England, 1998.
(75) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(76) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221−264.
(77) Liebécq, C., Ed. *Biochemical Nomenclature and Related Documents: a compendium prepared for the Committee of Editors of Biochemical Journals*, 2nd ed.; Portland Press: London, 1992.