Published in final edited form as:
J Phys Chem B. 2011 January 27; 115(3): 580–596. doi:10.1021/jp1092338.

Development of CHARMM Polarizable Force Field for Nucleic Acid Bases Based on the Classical Drude Oscillator Model

Christopher M. Baker, **Victor M. Anisimov**[†], and **Alexander D. MacKerell Jr**[*]
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn Street, Baltimore, MD 21201

Abstract

A polarizable force field for nucleic acid bases based on the classical Drude oscillator model is presented. Parameter optimization was performed to reproduce crystallographic geometries, crystal unit cell parameters, heats of sublimation, vibrational frequencies and assignments, dipole moments, molecular polarizabilities and quantum mechanical base-base and base-water interaction energies. The training and validation data included crystals of unsubstituted and alkyl-substituted adenine, guanine, cytosine, uracil, and thymine bases, hydrated crystals, and hydrogen bonded base pairs. Across all compounds, the RMSD in the calculated heats of sublimation is 4.1%. This equates to an improvement of more than 2.5 kcal/mol in accuracy compared to the non-polarizable CHARMM27 force field. However, the level of agreement with experimental molecular volume decreased from 1.7% to 2.1% upon moving from the non-polarizable to the polarizable model. The representation of dipole moments is significantly improved with the Drude polarizable force field. Unlike in additive force fields, there is no requirement for the gas-phase dipole moments to be overestimated, illustrating the ability of the Drude polarizable force field to treat accurately differently dielectric environments and indicating the improvements in the electrostatic model. Validation of the model was performed based on the calculation of the gas phase binding enthalpies of base pairs obtained via potential of mean force calculations; the additive and polarizable models both performed satisfactorily with average differences of 0.2 and 0.9 kcal/mol, respectively, and RMS differences of 1.3 and 1.7 kcal/mol, respectively. Overall, considering the number of significant improvements versus the additive CHARMM force field, the incorporation of explicit polarizability into the force field for nucleic acid bases represents an additional step toward accurate computational modeling of biological systems.

Keywords

Molecular modeling; empirical force field; molecular dynamics simulation; potential of mean force; biomolecules; DNA; RNA

Introduction

Studies of nucleic acid structure and function have been facilitated by the application of theoretical methods based on empirical force fields.[1,2,3] Force field based techniques enhance the interpretation of a wide variety of biochemical and biophysical data and provide

alex@outerbanks.umaryland.edu, Corresponding author phone: (410) 706-7442; fax: (410) 706-5017.
[†]Current address: School of Biomedical Informatics, University of Texas Health Science Center at Houston, 7000 Fannin St., Houston, TX 77030.

Supporting Information Available. Full force field parameters for the nucleic acid bases; detailed results for interactions with sodium cations, internal geometries, vibrational spectra, crystal cell parameters and interactions with water molecules; further details on the calculation of gas-phase PMFs. This information is available free of charge via the Internet at http://pubs.acs.org

insights that are difficult or impossible to obtain from experiment.[4,5] In addition, empirical force fields are increasingly utilized in experimental structure refinement in conjunction with X-ray or NMR data.[6,7,8]

There are a number of empirical force fields for nucleic acids available at present, including AMBER,[9,10] CHARMM27,[11] Bristol-Myers Squibb (BMS),[12] GROMOS[13] and OPLS-AA.[14] While the utility of these models has been discussed extensively in the literature,[15,16,17,18,19] they are all limited to the use of fixed atomic charges. As well as being computationally cost effective, such a treatment of electrostatic interactions has proven to be remarkably useful for a range of systems. However, its underlying approximation prevents an accurate estimation of electrostatic interactions as a function of the polarity of the surrounding environment.[20] In response to this limitation, a number of groups have undertaken efforts to develop polarizable force fields. To date, the majority of these efforts have focused on proteins,[21,22,23,24,25,26,27] with only a small number of studies focused on nucleic acids. Nakagawa *et al.*[28] developed a polarizable force field for nucleic acid bases based on the point dipole model using gas-phase quantum mechanical (QM) data. This model was limited to only the electrostatic portion of the force field with the remainder of the base parameters obtained from the CHARMM and AMBER non-polarizable force fields: it was not extended to oligonucleotides. In 2004 a polarizable point dipole model simulation of the DNA duplex $d(CCAACGTTGG)_2$ was reported[29] using AMBER.[30] The simulation showed the system to be stable and a small improvement in the reproduction of the reference crystallographic geometry over a non-polarizable force field was reported. A subsequent study tested the ability of the same model to reproduce the ideal B-DNA for the decamer, finding it to perform well, and again slightly better than a non-polarizable model,[31] though little information on the parameters used was given in either study. Also, no electrostatic 1–2 or 1–3 interactions were included and, as the authors note, this “results in molecular polarizabilities smaller than the real ones” meaning that the “force field might be slightly underpolarized.”[31] Work from this laboratory reported a polarizable simulation of $(GAGTACTC)_2$ based on a classical Drude oscillator.[32] The simulation was shown to be stable, validating the use of the Drude model for polarizable simulations of macromolecules in the condensed phase; however, the structure deviated significantly from the expected B form geometry. Such behavior was anticipated as only the electrostatic aspect of the model was carefully optimized, with the majority of the parameters transferred directly from the additive CHARMM27 model.

Due to the labor-intensive nature of parameter optimization, the practice of transferring the majority of parameters for a new model directly from the previous generation of a force field is common within force field development publications. And although the validity of such an approach has never been thoroughly investigated, all of the previously published polarizable force fields for nucleic acids used van der Waals and bonding parameters transferred from available additive force fields. While this assumption of transferability of some parameters is convenient, it inherently limits the quality of the resulting force field. Results reported to date for the optimization of a CHARMM Drude polarizable force field, encompassing water,[33,34] alkanes,[35] ethers,[36,37] alcohols,[38] aromatics,[39] heterocycles,[40] sulfur containing compounds[41] and ions,[42] show that the introduction of electronic polarization into an empirical force field requires complete reoptimization of all bonding and nonbonding parameters in order to attain a fully consistent polarizable force field where the different terms in the model are properly balanced.[19] Motivated by this need, the first fully optimized polarizable force field for nucleic acid bases has been developed, and is presented here. Starting from the previously published parameters for heterocycles,[40] the model was optimized by fitting to a wide variety of experimental and QM data. Electrostatic parameters were explicitly optimized for the NA bases, with Lennard-Jones (LJ), bond, angle, and torsion parameters taken directly from the CHARMM Drude polarizable force field for

heterocycles, and modified only where necessary. The final force field obtained in this work represents an extension of the available CHARMM Drude polarizable force field, representing an additional step in the ongoing efforts to develop a comprehensive polarizable empirical force field for biological molecules.

Methods

The parameter optimization performed in this work targeted the nucleobases shown in Figure 1: cytosine; 1-methyl-cytosine; thymine; 1-methyl-thymine; uracil; 1-methyl-uracil; adenine; 9-methyl-adenine; guanine, and 9-methyl-guanine. The overall parametrization scheme is presented in Figure 2. The induced electrostatic polarization approach employed in this work is based on the classical Drude oscillator model as previously described.[34] According to this model, the polarizability is introduced via the inclusion of a single auxiliary charged (Drude) particle attached to each non-hydrogen atom by a harmonic spring. The Drude particle does not carry LJ parameters, although LJ parameters could, in principle, be assigned to the Drude particles if required. Placement of a polarizable atom in an external electric field causes a displacement of the charged Drude particle in response to that field, giving rise to an induced dipole. Intramolecular polarization of 1,2 and 1,3 bonded atoms is treated via an induced dipole-dipole formalism. These interactions are damped using the method of Thole,[43] which is extended to include atom-based Thole scaling factors.[44] In addition, for atoms that act as hydrogen bond acceptors, lone pairs are introduced and the atomic polarizability is described by a polarizability tensor, as required to reproduce the anisotropic polarizability: the polarization response as a function of orientation around the acceptor.[45] Electrostatic parameter optimization was initiated using MP2/6-31G* optimized geometries of methylated DNA bases. Charges, polarizabilities and atom-specific Thole factors were determined based on electrostatic fitting. The fitting was performed using the FITCHARGE module in the CHARMM program.[46] For this purpose, QM electrostatic potential (ESP) calculations were performed on methyl-substituted bases in the presence of perturbation charges: the methyl-derivatives were used to allow for accurate reproduction of the polarity of the bonds linking the bases to the sugar. The electrostatic parameters obtained for the methylated bases were then directly transferred to the non-methylated bases, preserving the nitrogen electrostatic parameters and assigning the positive net charge of the methyl group to the hydrogen atom it is replaced by. Correspondingly, the polarizability of the methyl group was nullified by hydrogen atom replacement, following the decision to maintain zero polarizability on hydrogen atoms.[47] This protocol for replacing the methyl groups by hydrogen atoms was validated by calculation of gas-phase dipole moments of the isolated bases and by crystal calculations of the unsubstituted bases. Throughout this work the non-methylated bases served as a test set whereas all parameter optimization was performed on Me-substituted bases. Virtual charged particles carrying no LJ parameters were placed on the hydrogen bond acceptor nitrogen and oxygen atoms at the corresponding lone-pair (LP) positions. The partial atomic charge of the host atom was moved entirely to the corresponding LP sites (two LPs per oxygen atom, one LP per nitrogen atom), while the polarizability was retained on the atomic center. The optimization of the charges, polarizabilities and Thole factors was then achieved by fitting to the QM ESP maps. The initial guess charges for ESP fitting were obtained from the CHARMM27 force field.[11] Initial values for the polarizabilities were taken from the additive atomic polarizabilities of Miller,[48] modified for the present non-hydrogen polarizability model, as described previously,[47] and initial Thole factors were set to a value of 1.3. Electrostatic fitting was performed on monomer geometries optimized at the MP2/6-31G* level of theory using the Gaussian 03 package[49] with ESPs calculated using the B3LYP/aug-cc-pVDZ level of theory. The atomic polarizabilities fitted to QM perturbed ESPs were then scaled by a factor of 0.85. Scaling was performed adopting the procedure applied to the SWM4-NDP water model,[34] with the value of the scaling factor, 0.85, determined based on the

reproduction of the dielectric constants of pyridine and pyrole.[40] Such scaling of the gas phase polarizability values has been shown to be an important factor in the accurate reproduction of condensed phase properties.[37] And while the specific reason that scaling is required remains a point for debate, it is thought to be indicative of underlying physical phenomena; suggested explanations have been based on the Pauli exclusion principle[50] and inhomogeneities in the electric field within the exclusion volume of the molecule.[51]

While the electrostatic fitting procedure described above involves fitting to the total dipole moment of the molecule, it does not explicitly consider the individual components (and therefore orientation) of the dipole moment and polarizability tensor. To incorporate this information, the final electrostatic parameters were subjected to a "parameter minimization" procedure in which the steepest descent algorithm was used to minimize a target function as a function of the electrostatic parameters (charges, polarizabilities and Thole factors). The target function, in this case, was the sum of the percentage errors in the calculated values of the individual components of the polarizability tensor and dipole moment vector. This process also ensured that the final polarizability values were as close as possible to the QM values, scaled by a factor of 0.85.

During the charge fitting procedure described above, the polarizabilities of all atoms were assumed to be isotropic. Thereafter, the polarizability tensors around hydrogen bond accepting atoms were initially set to be the same as those previously obtained for N-methylacetamide (carbonyl groups) and pyrimidine/imidazole (ring N atoms). The polarizability anisotropies were then optimized by considering the calculated ESP around the molecule as a function of orientation, in the presence and absence of a perturbing ion.[45] Perpendicular arcs of interaction positions were constructed in-plane and out-of plane around all H bond acceptor atoms (Figure 3); a single sodium cation was then placed at each interaction point, and the total interaction energy evaluated. The components of the anisotropy tensors were then varied until optimum agreement with equivalent QM calculations, performed at the MP2/6-311G** level of theory with counterpoise corrections,[52] was achieved.

Due to their structural similarity, the majority of the bonding (bond, angle and dihedral) and LJ parameters for the NA bases were taken directly from the recently published Drude polarizable model for nitrogen-containing heterocycles.[40] In practice, the work on nucleic acid bases and heterocyclic compounds was conducted in concert, so as to minimize simultaneously any errors in the reproduction of condensed phase target data for both groups of compounds. The bonding and LJ parameters unique to the nucleic acid bases were initially set to the corresponding CHARMM27 nucleic acid force field values, and subsequently optimized.

Equilibrium bond length and angle parameters were optimized to reproduce target crystallographic geometries from a survey[53] of the Cambridge Structural Database.[54] A small number of bonds that initially produced unacceptably large deviations from the target crystallographic geometries were fixed via the assignment of new atom types different to those employed in the heterocyclic compounds.[40] The new atom types introduced are noted in Table S1 of the Supplementary Information. The target values for bonds and angles involving hydrogen atoms were obtained from MP2/6-31G* optimized geometries.

LJ parameters were then optimized to reproduce interactions of hydrogen-bonded dimers as well as condensed phase properties in the form of crystal heats of sublimation, volumes and lattice geometries. The logic behind this approach is that the structures and energies of nucleic acid base complexes are determined by a subtle balance of hydrogen bonding and stacking interactions. Stacking interactions, however, are notoriously difficult to evaluate by

either QM[55] or MM[56] methods, and therefore cannot be considered as reliable target data in the optimization process. To address this, emphasis is initially placed on the ability of the force field to accurately reproduce QM data for hydrogen-bonded dimer structures. Crystal simulations are then performed. As the crystals involve both hydrogen-bonding and stacking interactions, if the crystal data is accurately reproduced, it is assumed that the balance between hydrogen-bonding and stacking interactions has been captured. In effect, the stacking interactions have been implicitly included in the parametrization procedure.

The optimization of the LJ parameters represents the most intensive aspect of the project, involving manual iterative adjustment of selected LJ parameters to reproduce more accurately the target data. During this process the derived LJ parameters were also periodically tested on the parent N-containing heterocycles (e.g. pyridine, pyrimidine and purine) to ensure that consistent nonbond parameters were obtained for both classes of molecules. As above, a small number of problematic interactions were fixed by the assignment of new atom types.

As reference data for the hydrogen-bonding interactions, results were taken directly from the work of Jurečka *et al.*[57] In that study, interaction energies were calculated for numerous hydrogen-bonded base pairs at the CCSD(T) complete basis set limit. Of the hydrogen-bonded dimers considered by Jurečka *et al.*, 18 were used for LJ parameter optimization, as presented in Figure 4. For the molecular mechanics (MM) calculations, the individual monomers were minimized for 5000 steps using the adopted basis Newton Raphson (ABNR) method. Base pair structures were then subjected to 500 steps of minimization using the steepest descent method, followed by 5000 steps of minimization using the ABNR method. Interaction energies were then determined as the difference between the dimer energy minimum and the sum of the isolated monomer energy minima.

Polarizable molecular dynamics (MD) simulations of crystals were performed at a constant pressure of 1 atm with periodic boundary conditions and using the velocity Verlet integrator[58] that includes treatment of Drude particles via an extended Lagrangian double thermostat formalism.[59] For the simulations, a mass of 0.4 amu was transferred from real atoms to the corresponding Drude particles and the amplitude of Drude oscillations was controlled with a separate low-temperature thermostat at 1 K to simulate near-SCF conditions. The integration timestep was 1 fs for both polarizable and additive simulations using the Nose-Hoover thermostat[60,61] with a relaxation time of 0.1 ps applied to all real atoms. A modified Andersen-Hoover barostat[62] with a relaxation time of 0.1 ps was used to maintain the system at constant pressure. The SHAKE algorithm was used to constrain covalent bonds involving hydrogen.[63] LJ interactions were treated explicitly out to 12 Å with switch smoothing applied over the range of 10–12 Å. Both polarizable and additive force field simulations utilized an atom based switching algorithm.[64] Non-bond pair lists were maintained out to 16 Å, and a long-range correction for LJ interactions was applied in the condensed-phase simulations, to account for errors introduced by the truncation of the LJ interactions.[65] Electrostatic interactions were treated using particle mesh Ewald (PME) summation[66] with a coupling parameter of 0.34 and a sixth-order spline for mesh interpolation. Crystal simulations were performed with starting coordinates obtained from the Cambridge Structural Database.[54] In all cases, except 9-methyl-adenine, the final simulation cell was created by extending the unit cell edges by a factor of two in each direction, leading to 8-unit cells per simulation box. With 9-methyl adenine, only a single unit cell was used, because an 8-unit box was not supported by CHARMM for this system. To obtain convergent results from the crystal calculations, 10 independent MD simulations were run, each for 500 ps with initial velocities assigned using a random number seed. The first 100 ps of the simulations were treated as equilibration, and the final 400 ps were used for the analysis. Averages were obtained from the 10 independent simulations, with errors

calculated as the standard deviations over the 10 simulations. Heats of sublimation, ΔH_{sub}, and molecular volume, V_m, were determined following the standard procedure.[40] Gas-phase simulations required to calculate the heats of sublimation were performed using Langevin dynamics in the SCF regimen with infinite cutoffs for nonbond interactions. The friction coefficient was set to 5 ps^{-1} for all atoms except Drude particles. To prevent problems with overpolarization, in all simulations, gas and crystal, "an additional anharmonic restoring force" was included "to prevent excessively large excursions of the Drude particle away from the atom".[42] This "anharmonic term" has the form shown in Equation 1:

$$U_{hyp}=K_{hyp}(\Delta R - \Delta R_{cut})^n \quad (1)$$

In Equation 1, the exponent n represents the "order" of the correction. In this case, we use n = 4 and, as a second order correction to the polarizability, Equation 1 can be considered to represent the hyperpolarizability of the atom. K_{hyp} is then a force constant and ΔR_{cut} represents the Drude-atom separation at which the anharmonic potential is switched on. The values of K_{hyp} and ΔR_{cut} used in this work were 40,000 kcal/mol/Å^4 and 0.2 Å, respectively, values previously determined to be appropriate for the study of atomic ions.[42] As a final test of the optimized LJ parameters, energies were also evaluated for a set of stacked structures. As with the hydrogen-bonded structures, reference structures and energies were taken directly from the literature.[67] For the force field based calculations, these structures were minimized, first with the steepest descent method (2000 steps) and then the ABNR method (2000 steps). The resulting energies were then compared to the reference values.

Following optimization of the LJ parameters, force constants were adjusted based on potential energy distributions of calculated vibrational spectra performed using the MOLVIB utility[68] in CHARMM, with reference infrared spectra calculated using QM calculations at the MP2/6-31G* level of theory. The internal coordinate assignment was performed according to the method of Pulay *et al.*[69] A scale factor of 0.9434 was applied to the computed QM normal modes to account for the limitation in the level of theory.[70] The optimized empirical force constant parameters were obtained by matching the empirical vibrational modes to the corresponding QM data. The procedure for the optimization of equilibrium bond lengths and angles, as well as force constant parameter values was repeated each time a new set of LJ parameters became available. An iterative procedure such as this is necessary due to the interdependence of the bonded and nonbonded parameters. The equilibrium geometries, and hence vibrational spectra, depend upon both sets of parameters, meaning that they must be considered together to ensure a good balance that results in an internally consistent force field.

As another test of the optimized parameters, heterodimeric base-water interactions were considered. QM calculations on base-water monohydrates were performed with the bases fixed in the geometries obtained from optimization at the MP2/6-31G* level of theory and the water molecules fixed in the geometry of the SWM4-NDP model.[34] The minimum interaction energy distances between the individual water molecules and the bases were optimized at the MP2/6-31G* level of theory; in all cases only the interaction distance was optimized. Interaction energies were then determined using single point RI-MP2/cc-pVQZ calculations on the MP2/6-31G* minimum energy geometries calculated with the program Q-Chem,[71] applying counterpoise correction[52] to account for basis set superposition error (BSSE).[72] In the corresponding empirical calculations, the water geometry was fixed to the SWM4-NDP geometry while the bases were optimized in the gas phase prior to starting the monohydrate calculations. The base geometries were held rigid during the base-water distance scans. Previously, it has been observed that LJ parameters optimized to reproduce

condensed phase thermodynamic data tend to yield hydration free energies that are slightly too favorable.[73] It has been suggested that this problem relates to the use of non-optimal combining rules for the LJ parameters, and work to test this hypothesis is currently ongoing. In the absence of alternative combining rules, this error has been corrected through the use of "pair-specific LJ parameters" that override the standard combining rules, and specify directly the LJ interaction parameters between a given heavy atom and the O atom of the SWM4-NDP water model.[73] Typically, these pair-specific LJ parameters are optimized to reproduce experimental hydration free energies.[73] While such data is not available for the nucleic acid bases, pair-specific parameters have previously been optimized by fitting to hydration free energies for a number compounds that share atom types with the NA bases. Specifically, pair-specific LJ parameters have previously been optimized for atom types found in benzene,[39] pyridine,[40] pyrimidine,[40] acetamide and N-methylacetamide. The pair-specific LJ parameters employed in this work are taken directly from this previous work,[40] with parameter values listed in the supplementary material.

Experimental values for the gas phase binding enthalpies of a number of homo- and heterodimeric base-base pairs have been reported.[74] To evaluate the ability of the CHARMM27 and Drude polarizable force field models to reproduce these data, calculations were performed to determine potentials of mean force (PMFs) for the binding of 6 base-base pairs. These PMFs yield directly the free energies of binding for each base pair, when the known experimental value is the enthalpy of binding. To overcome this problem, PMFs can be calculated at three different temperatures, and a finite difference method used to calculate the enthalpic and entropic contributions at the temperature of interest.[75] Specifically, the entropic contribution to the binding free energy, $T\Delta S$, is determined via equation 2, and the enthalpic contribution, ΔH, is determined via equation 3.

$$T\Delta S_{(T)} = -T\frac{\Delta G(T+\Delta T) - \Delta G(T-\Delta T)}{2\Delta T} \quad (2)$$

$$\Delta H_{(T)} = \Delta G_{(T)} + T\Delta S_{(T)} \quad (3)$$

A total of six base-base dimers were considered: Me-Ade/Me-Ura; Me-Ura/Me-Ura; Me-Ade/Me-Thy; Me-Thy/Me-Thy; Me-Gua/Me-Cyt, and Me-Cyt/Me-Cyt. In all cases, and with both the CHARMM27 and Drude polarizable force fields, the PMF calculations were performed using the same procedure based on umbrella sampling along a reaction coordinate defined as the separation between the base centers of mass, R_{COM}. In all cases except Me-Thy/Me-Thy, molecular dynamics simulations were performed in the presence of a biasing potential running from R_{COM} = 4.5 Å to R_{COM} = 19.5 Å, in increments of 1 Å. For Me-Thy/Me-Thy the simulations began at R_{COM} = 3.5 Å but were otherwise identical to those performed for the other base-base pairs. In all cases, the dimers were simulated for 60 ns per window, which was found to be sufficient for convergence to have occurred (see Supplementary Material, Section S7 for details). The center of mass biasing potentials were implemented using the miscellaneous mean field potential (MMFP)[76] module within the CHARMM program. In all cases, a force constant of 1 kcal/mol/Å^2 was used. Following the molecular dynamics simulations, the weighted histogram analysis method (WHAM)[77] was used to unbias the umbrella sampling simulations and calculate the full PMF. Final ΔG_{bind} values were then calculated at each temperature by integrating over the bound and unbound states of the PMFs.[78] In general, ΔG_{bind} can be obtained from equation 4:[79]

$$\Delta G_{bind} = -kT \ln \left[\frac{P_B}{P_U} \right] \tag{4}$$

Where k is the Boltzmann Constant, T is the temperature and P_B and P_U are the probabilities of finding the dimer in the bound and unbound states, respectively. The normalized probabilities of finding the dimers in each of the bound and unbound states are then described by equations 5 and 6.[80]

$$P_B = \frac{\int_{r_{min}}^{r^*} r^2 exp\left(-\frac{W(r)}{kT}\right) dr}{\int_{r_{min}}^{r_{max}} r^2 exp\left(-\frac{W(r)}{kT}\right) dr} \tag{5}$$

$$P_U = \frac{\int_{r^*}^{r_{max}} r^2 exp\left(\frac{-W(r)}{kT}\right) dr}{\int_{r_{min}}^{r_{max}} r^2 exp\left(-\frac{W(r)}{kT}\right) dr} \tag{6}$$

Where $W(r)$ is the PMF, r_{min} and r_{max} represent the smallest and largest values of R_{COM} considered, and r^* represents the value of R_{COM} at which the dimer moves from being bound to being unbound. The factor of r^2 is included in equations 5 and 6 to account for the fact that we are integrating over a spherical volume.[81] Combining equations 5 and 6, we obtain equation 7, which allows for the direct calculation of ΔG_{bind} via equation 4.

$$\frac{P_B}{P_U} = \frac{\int_{r_{min}}^{r^*} r^2 exp\left(-\frac{W(r)}{kT}\right) dr}{\int_{r^*}^{r_{max}} r^2 exp\left(-\frac{W(r)}{kT}\right) dr} \tag{7}$$

Having established the method of calculation, the most important part of this procedure becomes identifying the point at which the bases cease to be bound, and therefore the most appropriate value of r^* to use in equation 7. To do this, ΔG_{bind} was calculated as a function of r^* for each PMF and plotted against R_{COM} (Figure S1, supplementary material). When ΔG_{bind} showed a sharp change in gradient, the two bases were no longer considered to be bound, and the value of R_{COM} at which the change in gradient occurred was adopted as r^*. With ΔG_{bind} values in place, ΔH_{bind} and $T\Delta S_{bind}$ values were calculated using equations 2 and 3. As mentioned above, PMFs were required at 3 different temperatures to enable the calculation of ΔH_{bind} values for comparison to experiment. In all cases, PMFs were calculated at 283.15 K, 298.15 K and 313.15 K. While the general procedure employed was the same for both the CHARMM27 and Drude polarizable force fields, the molecular dynamics simulation protocols used for the umbrella sampling differed between the two force fields. With CHARMM27, Langevin dynamics was used with a friction coefficient of 0.5 ps^{-1} applied to all atoms. A cutoff of 999.0 Å was used for LJ interactions, with the SHAKE algorithm[63] applied to constrain bonds to hydrogen, and a timestep of 1 fs. With the CHARMM Drude polarizable force field, the VV2 integrator was used and the amplitude of Drude oscillations was controlled using a separate low-temperature thermostat at 1 K to simulate near-SCF conditions. The integration timestep was 1 fs, with the SHAKE algorithm again used to constrain covalent bonds involving hydrogen.

All of the empirical force field calculations described above were performed using the program CHARMM.[46]

Results and Discussion

Parametrization of the Drude polarizable force field involves an iterative approach accounting for the correlation between the different terms in the potential energy function, as shown schematically in Figure 2. In the remainder of this manuscript, for the sake of clarity, results will be presented only for the final set of optimized parameters. Those final parameter values are presented in Section S1 of the supporting information.

1. Electrostatic Parameters

The quality of the developed electrostatic model is illustrated by the calculated polarizability tensors and dipole moments. Reproduction of the scaled QM components of the polarizability tensors (Table 1 for purine bases; Table 2 for pyrimidine bases) is uniformly good. A comparison of QM calculated dipole moments with those obtained from the force fields is shown in Table 3. As the QM data illustrate, the change in the dipole moment upon replacement of a methyl group by a H atom is non-trivial. The QM calculations show that the dipole moment increases upon moving from Me- to H-Cyt, while for the remaining bases the dipole moment decreases. The polarizable Drude model correctly reproduces this trend for all bases except Ura, despite the unsubstituted bases not being explicitly subjected to electrostatic parameter fitting. In contrast, the additive CHARMM27 model fails this test, showing the opposite trend for all of the bases. The root mean square deviation (RMSD) between empirical and QM dipole moments dropped from 1.24 D to 0.34 D for methyl-substituted bases and from 0.89 D to 0.41 D for the unsubstituted bases on moving from the CHARMM27 model to the polarizable model. The relatively poor agreement between the CHARMM27 dipole moments and the QM dipole moments should not, however, be seen as a sign of a low quality force field. It is well known that additive force fields must overestimate gas phase dipole moments in order to accurately reproduce liquid phase properties.[19] The improvement seen for the Drude model does, however, indicate an improvement in the electrostatic model, and illustrates its ability to function in a range of dielectric environments. The individual components of the dipole moment, and therefore its orientation, are also much more accurately described by the Drude polarizable model than by the CHARMM27 model: this is in large part attributable to the use of the “parameter minimization” procedure to fine tune these properties after the initial ESP fitting. Interactions with sodium cations on in-plane and out-of-plane arcs around the hydrogen-bond acceptor atoms provide another measure of the quality of the electrostatic interactions, and specifically their anisotropy. These results are summarized in Table 4 and Figure 5, with the complete results available in Tables S7–S11 of the supporting information. In all cases, the Drude polarizable force field provides a better representation of the interactions with Na^+ than does the CHARMM27 model. The improvement is particularly dramatic for C and U, where the RMSDs relative to QM results drop from 2.91 kcal/mol and 3.22 kcal/mol to 0.53 kcal/mol and 0.68 kcal/mol, respectively, when considering all orientations. The significant improvements in the ability of the polarizable model to reproduce electrostatic properties are not surprising: it is explicitly designed to describe more accurately the electrostatic properties of molecules in contrasting environments. This represents the major advantage of the polarizable model over the additive approach. With the polarizable model it is possible to reproduce the dipole moments in the gas phase, nonpolar environment as well as in polar, condensed phases, as emphasized by the ability of the model to reproduce crystal geometries and energies, discussed below. With the additive model it is necessary to overestimate the gas phase dipole moments, as is evident from the data in Table 3, to reproduce the condensed phase properties. Accordingly, it is anticipated that the polarizable

model will have a wider range of applicability with respect to environments of varying polarity.

2. Bond and Angle Parameters

The equilibrium bond length and angle parameters were optimized to reproduce experimental geometries of the bases obtained from a crystallographic survey.[82] RMSDs for the empirical geometries, including the CHARMM27 additive and the Drude polarizable model, with respect to the target data are shown in Table 5; data on the individual bonds and valence angles are shown in Tables S12–S16 of the supporting information. The overall RMSDs for bond distances are 0.011 Å and 0.023 Å for the CHARMM27 and Drude models, respectively. The corresponding values for valence angles are 1.4 ° and 2.5 °. The smaller RMSD values obtained for the CHARMM27 model are due to the larger number of adjustable parameters that were available in the non-polarizable model. In the additive CHARMM27 model, the nucleic acid base parameters were optimized totally independently of the remainder of the force field such that all bonded parameters could be optimized to maximize reproduction of the target data. With the present polarizable model, the majority of the bonded parameters were first optimized based on simpler N-containing heterocycles, including imidazole, pyridine, pyrmidine and purine.[40] This limited the number of adjustable parameters in the Drude model, thereby limiting the extent of agreement with the target data. However, RMSDs for the Drude model are close to the target values of 0.02 Å for bond lengths and 2 ° for angles. One may also add that a smaller RMSD is not necessarily indicative of a more robust model because using a large number of adjustable parameters may lead to overparametrization. This is one of the reasons for adopting the hierarchical approach outlined above: to develop a polarizable force field that is both robust and maintains a reasonable level of transferability. By sharing a large number of parameters between structurally similar compounds (e.g., nucleic acid bases and nitrogen-containing heterocycles), the common (shared) parameters are implicitly optimized for the extended data sets of the underlying hierarchy.

The optimization of force constants was performed to reproduce QM vibrational spectra including both frequencies and assignments of the methylated bases, based on potential energy decomposition analysis computed by the MOLVIB utility of CHARMM. The vibrational frequencies and assignments for CHARMM27, the Drude model and the QM model are shown in Tables S17–S21 of the supporting information and the percentage RMSDs for all modes below 2000 cm^{-1} are shown in Table 6. High energy stretching modes involving hydrogen atoms were excluded from the RMSD calculations because such modes are effectively excluded from MD simulations by using the SHAKE algorithm. The overall percentage RMSDs for modes below 2000 cm^{-1} over all the methylated bases were 24.3% and 10.1% for the CHARMM27 and Drude models, respectively. The improvement in the polarizable model is largely due to the CHARMM27 parametrization targeting scaled HF/6-31G* QM vibrational data and partially due to improvements in methyl torsions, which were not explicitly considered during optimization of the additive model. However, it is clear that the Drude model satisfactorily reproduces the vibrational modes below 2000 cm^{-1}, which dominate structural motions occurring in the bases during MD simulations.

3. Lennard-Jones Parameters

An essential feature of a force field for nucleic acid bases is the ability to reproduce base-base hydrogen-bond and stacking interactions, including Watson-Crick (WC)[83] and Hoogsteen[84] base pairing. Accordingly, analysis of 18 hydrogen-bonded base pairs defined by Jurečka *et al.*[57] was undertaken. The 18 interaction orientations are illustrated in Figure 4. The results, including the minimum interaction energies and distances along with the differences between the empirical and QM results, are presented in Table 7. An important

point to note here is that the CHARMM27 energies presented here differ from those presented in the original CHARMM27 paper[11] for equivalent structures. This situation arises because of a slight difference in the method used for the calculation of interaction energies. In the original CHARMM27 paper, interaction energies were calculated as the energy of the minimized dimer, minus the sum of the energies of monomers in the geometries they adopt *within the minimized dimer*. In this work, the interaction energies have been calculated as the energy of the minimized dimer, minus the sum of the energies of monomers in the geometries they adopt *when minimized in isolation*. For the interaction energies, the Drude model has an RMSD of 1.86 kcal/mol compared with a value of 2.75 kcal/mol for CHARMM27. The CHARMM27 model systematically underestimated the interaction energy, giving an average error of 2.27 kcal/mol; in the Drude model, this systematic error was reduced to 0.70 kcal/mol. While it was possible to eliminate this systematic error without adversely affecting the overall RMSD, such an approach resulted in interaction distances that were much too small, such that this step was not included in the present optimization. With the selected polarizable model, the base-base interaction distances are almost equivalent to those in CHARMM27, with the RMS differences being 0.06 Å and 0.16 Å for the CHARMM27 and Drude models, respectively. Based on the interaction energy RMS and average differences it is clear that the polarizable model is in better agreement with the target QM data than the CHARMM27 additive model. While hydrogen-bonding interactions were considered as target data for the optimization of CHARMM27 parameters,[11] the QM data employed was obtained at the MP2/6-31G(d) level with BSSE correction,[85] which itself provides a relatively poor agreement with the higher level QM data employed in this work. Of particular note in this case is the quality of the agreement for the WC hydrogen-bonding interactions, which are important given their central role in biological processes and the nucleic acid crystals (see below).

Aromatic stacking interactions are thought to be important in a number of biological situations,[86,87,88] including the determination of nucleic acid structures[89] and their interactions with proteins[90] and small molecules.[91] In spite of this importance, however, aromatic stacking interaction energies are notoriously difficult to calculate, via either MM[56] or QM[55] methods. For this reason, stacked complexes have not been considered as target data in the optimization of parameters for the nucleic acid bases. Instead, the strategy adopted targets the accurate reproduction of both hydrogen-bonded dimer energies and crystal properties (where the crystals include both hydrogen-bonding and stacking interactions). It is then assumed that if both of these sets of data are accurately reproduced, then the correct balance between hydrogen-bonding and stacking interactions must have been achieved; in effect, the stacking interactions have been implicitly included in the parametrization process. As a test of this strategy, the energies of a set of 10 stacked intrastrand base pairs[67] were evaluated. The results can be seen in Table 8. Comparing the force field results to the MP2 calculated results reveals a tendency for both force field models to predict the interaction energies to be less favorable than the QM calculated energies, by an average of 0.50 kcal/mol for the CHARMM27 and 2.02 kcal/mol for the Drude polarizable model. Work performed by Jurečka *et al.*[57] to evaluate base interaction energies at the CCSD(T) complete basis set limit, however, suggests that the MP2 method overestimates the favorability of base-stacking interactions. For the Drude model, the structure in worst agreement is CG-GC, which is less favorable by 4.8 kcal/mol than the MP2 result. Jurečka *et al.*[57] estimate that moving from MP2 to CCSD(T) makes the GC stacked interaction less favorable by around 2 kcal/mol. With the CG-GC stack containing two such interactions, the Drude calculated value does not appear to be unreasonable. In short, while significant uncertainty will remain in stacking interaction energies until extremely high level QM calculations can be performed, it appears that neither the CHARMM27 nor the Drude model is unreasonable.

The primary target data for the LJ parameter optimization and validation were condensed phase properties of the bases. In the parametrization process, the emphasis was placed on accurate fitting to both the experimental heats of sublimation and the lattice parameters as these observables are directly related to the strength of intermolecular interactions of the bases. In addition, the molecular packing in the crystals, including specific interactions between functional groups, may be used to directly tune LJ parameters for specific atom types, an option that is not available when LJ parameters are optimized for neat liquids. For these reasons, crystals represent near ideal target data for the optimization of the base LJ parameters.

In the present work, the condensed-phase target data consisted of twelve base crystals. Among these, the heats of sublimation were available for seven crystals at different temperatures. This represents a significant improvement relative to the training set data used in the optimization of the CHARMM27 force field, where only two heats of sublimation, for Ura and 9-methyl-Ade, were considered. Among the seven crystals for which heats of sublimation are available, six are pyrimidine bases and only one is a purine base. While this imbalance makes accurate optimization of purine base parameters more difficult, it should be noted that the unmodified purine parameters used as the starting point for the parametrization were previously optimized to reproduce crystal properties as well as interactions with rare gases, water and base dimers.[40] The use of such high quality initial parameters provides a safety net to overcome the limited amount of target experimental data on heats of sublimation and the large number of adjustable parameters that may lead to non-physical parameters (i.e. the parameter correlation problem).[19] In addition, simultaneous optimization of the heats of sublimation and the lattice parameters facilitated the identification of physically meaningful LJ parameters.

Results for the final Drude polarizable parameter set along with results from the CHARMM27 force field and the differences with respect to the experimental data are shown in Tables 9 and 10 for the unit cell molecular volumes and the heats of sublimation, respectively. For the molecular volumes, the two empirical models give similar results, with the Drude results being slightly worse than the additive model. For the additive model, the overall RMSD of 1.7% is perfectly within the 2% target limit; the corresponding RMSD for the Drude model is 2.1%, which is still considered acceptable. This larger difference is largely dominated by one of the hydrated crystals, H-Gua:water, included in the training set. For this crystal, the Drude model overpredicts the volume by 5.5%. However, the Drude model gives a good account of the H-Ade:water complex, with an error of −1.3%. While the relatively large error in the H-Gua:water crystal volume is not ideal, it is reassuring that there is not a systematic error in the volumes of the hydrated crystals. As noted above, for the purine bases there is only one crystal structure, 9-Me-adenine, for which both the experimental molecular volume and heat of sublimation are known, compared to six for the pyrimidine bases. This limited number of high quality crystal structures makes it more difficult to get the Ade and Gua parameters right, and this is a possible source of the error in the H-Gua:wat crystal. However, the molecular volumes of Me-Ade, Et-Gua, MT:MA, and EG:MC crystals calculated using the polarizable force field are also all in good agreement with the experimental results and H-Gua:water is the only purine base crystal with a calculated volume error outside of the 2% target region. We can, therefore, conclude that the purine base parameters are reasonable.

When the heats of sublimation are considered (Table 10), the polarizable model shows a significant improvement over the additive model. With the additive model the RMSD is 3.85 kcal/mol, compared to 1.32 kcal/mol for the Drude model. Even the two crystals used in the optimization of the additive model, uracil and 9-methyl-adenine, are better predicted by the polarizable model. While several of the calculated heats of sublimation are outside the target

of “within 2% of experimental data” only one, Me-Cyt, is worse than the equivalent CHARMM27 value, and there is no evidence of any systematic error. Using CHARMM27, the largest error was 29.0 % for Me-Ade, which is reduced to 2.3 % with the polarizable model. The significant improvement in the crystal energies observed for the polarizable model speaks of its ability to treat more accurately the intermolecular interactions occurring between the different molecules in the crystal.

The analysis of condensed phase properties is concluded by reviewing the lattice parameters optimized by the CHARMM27 and Drude polarizable force fields in comparison to experimental X-ray data. Calculated results are presented in Tables S22 and S23 for the pyrimidine and purine bases, respectively. Looking at the percentage differences in lattice parameters for the pyrimidine bases, one can see that CHARMM27 has noticeable problems with thymine bases (H-Thy and Me-Thy). These compounds display differences of 30–40% in the values of the *a* and *c* cell edges and about 9% in the monoclinic angle β. This problem is corrected in the Drude model and the differences for the thymine crystals are reduced to a maximum of 8% and 8% for the cell edges and monoclinic angle, respectively. For the other pyrimidine bases, both the non-polarizable and polarizable models performed equally well. For the entire set of six pyrimidine crystals the RMSDs in unit cell edges improved for A and C from 3.55 Å and 3.02 Å to 2.96 Å and 0.66 Å upon going from the CHARMM27 to Drude force field, respectively. The RMSD for cell edge B, however, went from 0.61 Å to 2.09 Å upon adoption of the polarizable force field. The quality in description of unit cell angles is roughly unchanged with the largest RMSD changing from 6.5 ° to 7.9 ° upon going from the non-polarizable to polarizable force field.

For the purine bases, (Table 13) the CHARMM27 force field showed the largest deviations from experiment for the H-Ade hydrated crystal, with a 12% difference for the unit cell edge *a* and an 11% difference for the unit cell angle γ. In the Drude model the largest deviations are observed for the 9-ethyl-guanine–1-methyl-cytosine base pair crystal where there is a difference of 16% in the unit cell edge *c*, 13% in the unit cell edge *b* and 16% in the unit cell angle α. Interestingly, however, the Drude model gives a better reproduction of the Watson-Crick hydrogen bonding interaction within the EGMC crystal than does CHARMM27 (Figure 6). While CHARMM27 predicts the optimal hydrogen bonding interaction distance to be 2.96 Å (measured from N1-N3), the Drude model predicts 2.91 Å, compared to an experimental value of 2.92 Å. Overall, looking at the summary of RMSDs for the purine bases, it can be seen that both the CHARMM27 and Drude force field performed similarly against the experimental X-ray data, with the CHARMM 27 model having a slight advantage.

4. Validation: Base-Water Heterodimer Interactions

Traditionally, gas-phase interactions with water have been used in the non-polarizable CHARMM force field to derive partial atomic charges based on the reproduction of the QM interaction energies and distances.[11,92] This represents a computationally economical model to derive charges which, in some form, reflects the presence of environmental effects explicitly incorporated into the parametrization, an approach originally pioneered by Jorgensen[93] and Lifson.[94] Although this practice is replaced in the polarizable model by ESP fitting,[47] it is reasonable to assume that the derived electrostatic model in combination with the optimized LJ parameters should closely reproduce gas-phase interactions with water. Satisfying this test ensures that the empirical electrostatic model can reproduce the anisotropy of hydrogen-bond interactions as judged by the reproduction of the minimum interaction distances and energies with water as a function of orientation. This is assessed by placing individual water molecules at locations where they interact directly with various atoms of the bases, including both in and out-of-plane interaction orientations, and orientations where the water molecules probe lone pairs at the hydrogen bond acceptor sites;

an example of the orientations of water molecules around cytosine is shown in Figure 7. The target data for nucleic acid base - water interactions were from MP2/6-31G(d) optimizations to obtain the minimum interaction distances and from single point RI-MP2/cc-pVQZ calculations (including counterpoise correction to account for basis set superposition error) to obtain interaction energies. The other methylated nucleic acid bases were treated in a similar manner. A statistical summary of the differences in the interactions of the methylated bases with water as compared to the QM results is presented in Table 11. Details for the individual water orientations are available in the Supporting Information, Tables S24–S28.

An important point to note is that the CHARMM27 model fails to predict some interaction orientations as minima. These minima are correctly identified by the Drude polarizable force field, indicating an improvement in the representation of interactions with water. However, for the sake of a fair comparison, the orientations for which the additive model predicts no minimum are excluded from the statistics shown in Table 11 and discussed below. For the distances, with both models the RMSDs are consistently around 0.2 Å. For CHARMM27, the average differences over all conformations are slightly negative at −0.05 Å; for the Drude model the average error over all orientations is −0.02 Å, with the average error over in-plane interactions being 0.00 Å and the average error over out-of-plane interactions being −0.05 Å. Essentially, interaction distances are basically of equivalent quality for both the CHARMM27 and Drude models, although the Drude model shows a slightly smaller systematic underestimation than does the CHARMM27 model. In the reproduction of interaction energies, the use of the Drude polarizable model yields a small improvement over the CHARMM27 model. The RMSD in energies calculated over all orientations drops from 1.32 kcal/mol to 1.10 kcal/mol on moving from the CHARMM27 to Drude model, with the average error decreasing from −0.49 kcal/mol to −0.29 kcal/mol. It is also useful to consider separately the in-plane and out-of-plane water interactions. As mentioned above, for the non-polarizable CHARMM27 mode, there were several orientations at which no minimum could be identified: these all corresponded to out-of-place interactions. Even with these interactions omitted, when moving from the non-polarizable to polarizable force field, the RMSD in out-of-plane interaction energies decreases from 1.50 kcal/mol to 0.85 kcal/mol, with the average error decreasing from −0.54 kcal/mol to −0.39 kcal/mol. For the out-of-plane orientations the improvement is significant enough that it must be attributed to the explicit consideration of electronic polarizability. For in-plane interactions, the RMSD for energies is smaller for the non-polarizable model than for the polarizable model, although the average error is still of larger magnitude for the non-polarizable model. These results are in part attributable to the fact that, in the original CHARMM27 force field,[11] the charges were optimized based only upon in-plane orientations. However, the level of agreement with CHARMM27 in the present context is impressive given that scaled HF/6-31G* QM results were the target data in that optimization. In contrast, water-base interactions were not considered during parameter optimization for the polarizable model: therefore, the observed improvement in water interactions should be considered a byproduct of explicit incorporation of polarizability in the Drude model. This mostly pertains to the reduced relative differences in interaction energies, which are more uniform in the polarizable model than in the non-polarizable model.

5. Validation: Gas Phase Enthalpies of Binding

As a final validation test of the newly developed CHARMM Drude polarizable force field models, PMF calculations were used to calculate the gas phase dimerization energies of a series of homo- and heterodimeric base-base pairs for comparison to known experimental values.[74] Final PMFs calculated at 298 K are shown in Figure 8, and the enthalpies of binding obtained via the finite difference method described above are shown in Table 12. Both the CHARMM27 and Drude polarizable force fields give an acceptable reproduction

of the experimental binding enthalpies (Table 12). For the CHARMM27 force field, two of the six calculated binding energies deviate from the experimental values by more than 1 kcal/mol. For the Drude polarizable force field, three binding energies deviate by more than 1 kcal/mol from the experimental values. With the polarizable force field, the worst reproduction of the experimental binding energy is for mC:mC, where the error in the binding energy is 3.7 kcal/mol compared to experiment. The CHARMM27 force field also has its largest error for this base pair, at 2.9 kcal/mol. The mC:mC dimer is also one of only two cases in which the CHARMM27 PMF at 298 K differs significantly from the Drude polarizable PMF (Figure 8). The other case is mA:mU, and in both instances the result is that the Drude polarizable force predicts significantly stronger binding than does CHARMM27. For mA:mU the result is an error of unchanged magnitude but opposite sign, while for mC:mC the result is a worsening in the reproduction of the binding free energy by 1.0 kcal/mol.

The PMF calculations can also be used to obtain information on the conformational preferences in the regions around the energy minima of the PMFs. Such analysis is of interest as it indicates whether the additive and polarizable force fields are sampling similar or different relative orientations in the bound state. The analysis was performed by first noting the distance of the energy minimum for every PMF calculated. The trajectory from the MD simulation corresponding to the window with the center-of-mass restraint closest to the distance of the energy minimum was then analyzed to determine the orientations being sampled by the interacting pairs. This analysis involved arbitrarily choosing one of the two bases as the reference, and then calculating the frequency with which the O and N atoms of its partner base occurred in 0.2 Å cubic volume elements around the reference molecule. The resulting information was then visualized as a series of density plots,[95] shown in Figure 9. From Figure 9, it can be seen that the biggest differences in population preferences occur for the mA:mU and mA:mT base pairs (Figures 9a and 9c, respectively). In both cases, CHARMM27 predicts that the most populated conformation is a Hoogsteen type hydrogen-bonded structure. For mA:mT, this is the only significantly populated conformation, while for mA:mU, the Watson-Crick hydrogen-bonding conformation is also observed. With the Drude polarizable force field, the Hoogsteen structures are only sparsely populated in both base pairs. Preferred instead are structures involving H bonding interactions with the N3 of adenine. *Ab initio* calculations performed by Kratochvíl *et al.* have also identified this structure as the global energy minimum for the A:T base pair.[96] It is important to note here that while Figure 9 shows, for example, no population of Hoogsteen structures for mA:mU with the Drude polarizable force field, this does not mean that they do not occur at all, but rather that they do not occur in the region around the free energy minimum. The mA:mU minimum for the Drude polarizable force field occurs at a center of mass separation of 6.15 Å, meaning that Figure 9 shows the population from the simulation with the center of mass separation restrained to 6.5 Å. With CHARMM27, the mA:mU minimum occurs at 5.75 Å and Figure 9 shows the population from the MD simulation with the center of mass restrained to 5.5 Å. Significant differences in the populations are also observed for mU:mU and mT:mT (Figures 9b and 9d, respectively). For mU:mU, the CHARMM27 force field predicts two approximately equally populated minima, related by a flipping of the partner base. For the Drude polarizable model, the atomic density is much less well defined, suggesting that a number of conformations are being sampled. This is consistent with the corresponding PMF (Figure 8), which is characterized by a broad low-energy region around the global minimum. For mT:mT, both models predict that the preferred conformation of the molecules is stacked, rather than H bonded. Both models also, however, predict a second minimum in the PMF at around 6 Å, which corresponds to H bonding structures. The preference for stacked structures relative to mU:mU where none occur, must be attributable to the presence of the additional methyl group in Thy. In both the gas phase[97] and the liquid phase,[98,99] the addition of a methyl group to benzene, making toluene, significantly

increases the population of such stacked conformations by increasing the magnitude of the dispersion interaction. It is possible that such an effect is also present here. The final two base pairs, mG:mC and mC:mC, both give very similar results for the two force fields. For mG:mC, the population around the minimum energy conformation is exclusively Watson-Crick hydrogen-bonded. For mC:mC, a hydrogen-bonded dimer is also strongly preferred, although a little more conformation flexibility is observed with the CHARMM27 force field than with the Drude polarizable force field.

Conclusions

A CHARMM Drude polarizable force field model for the nucleic acid bases has been developed, with parameters optimized based on the methylated bases Me-Cyt, Me-Thy, Me-Ura, Me-Ade and Me-Gua. Electrostatic parameters have been shown to give an excellent reproduction of polarizability tensors and dipole moments, performing significantly better than the corresponding CHARMM27 additive model. This is perhaps unsurprising: the rationale for moving from fixed-charge to polarizable force fields is that an improved representation of electrostatic interactions should be obtained. Lennard-Jones and bonding parameters have been built upon those previously optimized for CHARMM Drude polarizable force field models for N-containing aromatic heterocycles. Adopting this strategy reduces the risk of overfitting and increases the confidence in the applicability of the model in situations beyond the scope of the initial tests. Where additional optimization of LJ parameters was required, it was performed in an iterative fashion to reproduce accurately both crystal phase thermodynamic and structural data as well as structural and energetic properties of a range of hydrogen-bonded dimers. While stacked structures were not considered explicitly, it is expected that this approach will have led to the implicit optimization of these interactions: the balance between hydrogen-bonding and stacking interactions plays an important role in determining the base crystal structures. With the final set of parameters in place, good reproductions of internal geometries, vibrational spectra, interactions with water molecules, interactions with sodium cations and gas phase binding energies were also achieved.

Overall, the development of reliable force field parameters for the nucleic acid bases represents a significant step towards the completion of a CHARMM Drude polarizable force field for the simulation of nucleic acids, including oligonucleotides. CHARMM Drude polarizable force field parameters for the sugar ring have previously been developed, and can now be combined with those for the bases and the phosphate group to yield models of the full nucleic acids. Such a combining process will not be trivial: optimization of bond, angle and dihedral parameters associated with the connections between the different moieties will be a significant task, as will the validation of the final models against condensed phase data for the full oligonucleotides.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge financial support from the NIH (GM051501) and computational support from the DoD High Performance Computing, the Pittsburgh Supercomputing Center and the NSF/TeraGrid computational resources.

References

1. Orozco M, Pérez A, Noy A, Luque FJ. Chem. Soc. Rev. 2003; 32:350–364. [PubMed: 14671790]

2. Cheatham, TE, III. Molecular Modeling and Atomistic Simulation of Nucleic Acids. In: Spellmeyer, DC., editor. Annual Reports in Computational Chemistry. 1st ed.. Vol. Vol. 1. Amsterdam, The Netherlands: Elsevier; 2005. p. 75-89.
3. MacKerell AD Jr, Nilsson L. Curr. Opin. Struc. Biol. 2008; 18:194–199.
4. Cheatham TE III. Curr. Opin. Struc. Biol. 2004; 14:360–367.
5. Norberg J, Nilsson L. Q. Rev. Biophys. 2003; 36:257–306. [PubMed: 15029826]
6. Apostolakis J, Hofmann DWM, Lengauer T. Acta Crystallogr. A. 2001; 57:442–450. [PubMed: 11418755]
7. Žídek L, Štefl R, Sklenář V. Curr. Opin. Struc. Biol. 2001; 11:275–281.
8. Kuszewski J, Schwieters C, Clore GM. J. Am. Chem. Soc. 2001; 123:3903–3918. [PubMed: 11457140]
9. Cheatham TE III, Cieplak P, Kollman PA. J. Biomol. Struct. Dyn. 1999; 16:845–862. [PubMed: 10217454]
10. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE III, Laughton CA, Orozco M. Biophys. J. 2007; 92:3817–3829. [PubMed: 17351000]
11. Foloppe N, MacKerell AD Jr. J. Comput. Chem. 2000; 21:86–104.
12. Langley DR. J. Biomol. Struct. Dyn. 1998; 16:487–509. [PubMed: 10052609]
13. Soares TA, Hünenberger PH, Kastenholz MA, Kräutler V, Lenz T, Lins RD, Oostenbrink C, van Gunsteren WF. J. Comput. Chem. 2005; 26:725–737. [PubMed: 15770662]
14. Pranata J, Wierschke SG, Jorgensen WL. J. Am. Chem. Soc. 1991; 113:2810–2819.
15. Reddy SY, Leclerc F, Karplus M. Biophys. J. 2003; 84:1421–1449. [PubMed: 12609851]
16. Giudice E, Lavery R. Acc. Chem. Res. 2002; 35:350–357. [PubMed: 12069619]
17. Bosch D, Foloppe N, Pastor N, Pardo L, Campillo M. J. Mol. Struct. Theochem. 2001; 537:283–305.
18. Beveridge DL, McConnell KJ. Curr. Opin, Struc. Biol. 2000; 10:182–196.
19. MacKerell AD Jr. J. Comput. Chem. 2004; 25:1584–1604. [PubMed: 15264253]
20. Khandogin J, York DM. J. Phys. Chem. B. 2002; 106:7693–7703.
21. Kaminski GA, Stern HA, Berne BJ, Friesner RA, Cao YX, Murphy RB, Zhou R, Halgren TA. J. Comput. Chem. 2002; 23:1515–1531. [PubMed: 12395421]
22. Ogawa T, Kurita N, Sekino H, Kitao O, Tanaka S. Chem. Phys. Lett. 2004; 397:382–387.
23. Friesner RA. Adv. Prot. Chem. 2005; 72:79–104.
24. Kim B, Young T, Harder E, Friesner RA, Berne BJ. J. Phys. Chem. B. 2005; 109:16529–16538. [PubMed: 16853101]
25. Harder E, Kim B, Friesner RA, Berne BJ. J. Chem. Theory Comput. 2005; 1:169–180.
26. Wang Z-X, Zhang W, Wu C, Lei H, Cieplak P, Duan Y. J. Comput. Chem. 2006; 27:781–790. [PubMed: 16526038]
27. Isegawa M, Kato S. J. Chem. Theory Comput. 2009; 5:2809–2821.
28. Nakagawa S. J. Comput. Chem. 2007; 28:1538–1550. [PubMed: 17342710]
29. Baucom J, Transue T, Fuentes-Cabrera M, Krahn JM, Darden TA, Sagui C. J. Chem. Phys. 2004; 121:6998–7008. [PubMed: 15473761]
30. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. J. Comput. Chem. 2005; 26:1668–1688. [PubMed: 16200636]
31. Babin V, Baucom J, Darden TA, Sagui C. J. Phys. Chem. B. 2006; 110:11571–11581. [PubMed: 16771434]
32. Anisimov VM, Vorobyov IV, Lamoureux G, Noskov S, Roux B, MacKerell AD Jr. Biophys. J. 2004; 86:415A.
33. Lamoureux G, MacKerell AD Jr, Roux B. J. Chem. Phys. 2003; 119:5185–5197.
34. Lamoureux G, Harder E, Vorobyov IV, Roux B, MacKerell AD Jr. Chem. Phys. Lett. 2006; 418:245–249.
35. Vorobyov IV, Anisimov VM, MacKerell AD Jr. J. Phys. Chem. B. 2005; 109:18988–18999. [PubMed: 16853445]

36. Vorobyov I, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, MacKerell AD Jr. J. Chem. Theory Comput. 2007; 3:1120–1133.
37. Baker CM, MacKerell AD Jr. J. Mol. Model. 2010; 16:567–576. [PubMed: 19705172]
38. Anisimov VM, Vorobyov IV, Roux B, MacKerell AD Jr. J. Chem. Theory Comput. 2007; 3:1927–1946. [PubMed: 18802495]
39. Lopes PEM, Lamoureux G, Roux B, MacKerell AD Jr. J. Phys. Chem. B. 2007; 111:2873–2885. [PubMed: 17388420]
40. Lopes PEM, Lamoureux G, MacKerell AD Jr. J. Comput. Chem. 2009; 30:1821–1838. [PubMed: 19090564]
41. Zhu X, MacKerell AD Jr. J. Comput. Chem. 2010; 31:2330–2341. [PubMed: 20575015]
42. Yu H, Whitfield TW, Harder E, Lamoureux G, Vorobyov I, Anisimov VM, MacKerell AD Jr, Roux B. J. Chem. Theory Comput. 2010; 6:774–786. [PubMed: 20300554]
43. Thole BT. Chem. Phys. 1981; 59:341–350.
44. Harder E, Anisimov VM, Whitfield T, MacKerell AD Jr, Roux B. J. Phys. Chem. B. 2008; 112:3509–3521. [PubMed: 18302362]
45. Harder E, Anisimov VM, Vorobyov IV, Lopes PEM, Noskov SY, MacKerell AD Jr, Roux B. J. Chem. Theory Comput. 2006; 2:1587–1597.
46. Brooks BR, Brooks CL III, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. J. Comput. Chem. 2009; 30:1545–1614. [PubMed: 19444816]
47. Anisimov VM, Lamoureux G, Vorobyov IV, Huang N, Roux B, MacKerell AD Jr. J. Chem. Theory Comput. 2005; 1:153–168.
48. Miller KJ. J. Am. Chem. Soc. 1990; 112:8533–8542.
49. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA., Jr; Vreven, T.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, JE.; Hratchian, HP.; Cross, JB.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Gaussian 03, Revision D.01. Wallingford, CT: Gaussian Inc; 2004.
50. Kaminski GA, Stern HA, Berne BJ, Friesner RA. J. Phys. Chem. A. 2004; 108:621–627.
51. Schropp B, Tavan P. J. Phys. Chem. B. 2008; 112:6233–6240. [PubMed: 18198859]
52. Boys SF, Bernardi F. Mol. Phys. 1970; 19:553–566.
53. Clowney L, Jain SC, Srinivasan AR, Westbrook J, Olson WK, Berman HM. J. Am. Chem. Soc. 1996; 118:509–518.
54. Allen FH. Acta Crystallogr. B. 2002; 58:380–388. [PubMed: 12037359]
55. Sinnokrot MO, Sherrill CD. J. Phys. Chem. A. 2004; 108:10200–10207.
56. Hunter CA, Sanders JKM. J. Am. Chem. Soc. 1990; 112:5525–5534.
57. Jurečka P, Šponer J, Černý J, Hobza P. Phys. Chem. Chem. Phys. 2006; 8:1985–1993. [PubMed: 16633685]
58. Swope WC, Anderson HC, Berens PH, Wilson KR. J. Chem. Phys. 1982; 76:637–649.
59. Lamoureux G, Roux B. J. Chem. Phys. 2003; 119:3025–3039.
60. Nosé S. Mol. Phys. 1984; 52:255–268.
61. Hoover WG. Phys. Rev. A. 1985; 31:1695–1697. [PubMed: 9895674]
62. Martyna GJ, Tobias DJ, Klein ML. J. Chem. Phys. 1994; 101:4177–4189.

63. Ryckaert J-P, Ciccotti G, Berendsen HJC. J. Comput. Phys. 1977; 23:327–341.
64. Steinbach PJ, Brooks BR. J. Comput. Chem. 1994; 15:667–683.
65. Lagüe P, Pastor RW, Brooks BR. J. Phys. Chem. B. 2004; 108:363–368.
66. Darden T, York D, Pedersen L. J. Chem. Phys. 1993; 98:10089–10092.
67. Fiethen A, Jansen G, Hesselmann A, Schütz M. J. Am. Chem. Soc. 2008; 130:1802–1803. [PubMed: 18201088]
68. Wiorkiewicz-Kuczera, J.; Kuczera, K.; Karplus, M. MOLVIB. Cambridge, MA: Harvard University; 1989.
69. Pulay P, Fogarasi G, Pang F, Boggs JE. J. Am. Chem. Soc. 1979; 101:2550–2560.
70. Scott AP, Radom L. J. Phys. Chem. 1996; 100:16502–16513.
71. Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio RA Jr, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Lin CY, van Voorhis T, Chien SH, Sodt A, Steele RP, Rassolov VA, Maslen PE, Korambath PP, Adamson RD, Austin B, Baker J, Byrd EFC, Dachsel H, Doerksen RJ, Dreuw A, Dunietz BD, Dutoi AD, Furlani TR, Gwaltney SR, Heyden A, Hirata S, Hsu C-P, Kedziora G, Khalliulin RZ, Klunzinger P, Lee AM, Lee MS, Liang WZ, Lotan I, Nair N, Peters B, Proynov EI, Pieniazek PA, Rhee YM, Ritchie J, Rosta E, Sherrill CD, Simmonett AC, Subotnik JE, Woodcock HL III, Zhang W, Bell AT, Chakraborty AK, Chipman DM, Keil FJ, Warshel A, Hehre WJ, Schaefer HF III, Kong J, Krylov AI, Gill PMW, Head-Gordon M. Phys. Chem. Chem. Phys. 2006; 8:3172–3191. [PubMed: 16902710]
72. Ransil BJ. J. Chem. Phys. 1961; 34:2109–2118.
73. Baker CM, Lopes PEM, Zhu X, Roux B, MacKerell AD Jr. J. Chem. Theory Comput. 2010; 6:1181–1198. [PubMed: 20401166]
74. Yanson IK, Teplitsky AB, Sukhodub LF. Biopolymers. 1979; 18:1149–1170. [PubMed: 435611]
75. Koller AN, Božilović J, Engels JW, Gohlke H. Nucl. Acids Res. 2010; 38:3133–3146. [PubMed: 20081201]
76. Beglov D, Roux B. J. Chem. Phys. 1994; 100:9050–9063.
77. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. J. Comput. Chem. 1992; 13:1011–1021.
78. Jorgensen WL. Acc. Chem. Res. 1989; 22:184–189.
79. Deng N-J, Cieplak P. Biophys. J. 2010; 98:627–636. [PubMed: 20159159]
80. Bursulaya BD, Brooks CL III. J. Phys. Chem. B. 2000; 104:12378–12383.
81. Fukunishi Y, Mitomo D, Nakamura H. J. Chem. Inf. Model. 2009; 49:1944–1951. [PubMed: 19807195]
82. Clowney L, Jain SC, Srinivasan AR, Westbrook J, Olson WK, Berman HM. J. Am. Chem. Soc. 1996; 118:509–518.
83. Watson JD, Crick FHC. Nature. 1953; 171:737–738. [PubMed: 13054692]
84. Hoogsteen K. Acta. Crystallogr. 1963; 16:907–916.
85. Hobza P, Kabeláč M, Šponer J, Mejzkík P, Vondrášek J. J. Comput. Chem. 1997; 18:1136–1150.
86. Butterfield SM, Patel PR, Waters ML. J. Am. Chem. Soc. 2002; 124:9751–9755. [PubMed: 12175233]
87. Tatko CD, Waters ML. J. Am. Chem. Soc. 2002; 124:9372–9373. [PubMed: 12167022]
88. Burley SK, Petsko GA. Science. 1985; 229:23–28. [PubMed: 3892686]
89. Hunter CA. Philos. Trans. R. Soc. London, Ser. A. 1993; 345:77–85.
90. Baker CM, Grant GH. Biopolymers. 2007; 85:456–470. [PubMed: 17219397]
91. Řeha D, Kabeláč M, Ryjáček F, Šponer J, Šponer JE, Elstner M, Suhai S, Hobza P. J. Am. Chem. Soc. 2002; 124:3366–3376. [PubMed: 11916422]
92. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. J. Phys. Chem. B. 1998; 102:3586–3616.

93. Jorgensen WL, Maxwell DS, Tirado-Rives J. J. Am. Chem. Soc. 1996; 118:11225–11236.
94. Hagler AT, Huler E, Lifson S. J. Am. Chem. Soc. 1974; 96:5319–5327. [PubMed: 4851860]
95. Baker CM, Grant GH. J. Phys. Chem. B. 2007; 111:9940–9954. [PubMed: 17672488]
96. Kratochvíl M, Šponer J, Hobza P. J. Am. Chem. Soc. 2000; 122:3495–3499.
97. Sinnokrot MO, Sherrill CD. J. Am. Chem. Soc. 2004; 126:7690–7697. [PubMed: 15198617]
98. Baker CM, Grant GH. J. Chem. Theory Comput. 2006; 2:947–955.
99. Baker CM, Grant GH. J. Chem. Theory Comput. 2007; 3:530–548.
100. McClure RJ, Craven BM. Acta. Crystallogr. B. 1973; 29:1234–1238.
101. Portalone G, Bencivenni L, Colapietro M, Pieretti A, Ramondo F. Acta. Chem. Scand. 1999; 53:57–68.
102. Stewart RF, Jensen LH. Acta. Crystallogr. 1967; 23:1102–1105.
103. Rossi M, Kistenmacher TJ. Acta. Crystallogr. B. 1977; 33:3962–3965.
104. Kvick Å, Koetzle TF, Thomas R. J. Chem. Phys. 1974; 61:2711–2719.
105. McMullan RK, Craven BM. Acta. Crystallogr. B. 1989; 45:270–276. [PubMed: 2619960]
106. McMullan RK, Benci P, Craven BM. Acta. Crystallogr. B. 1980; 36:1424–1430.
107. Destro R, Kistenmacher TJ, Marsh RE. Acta. Crystallogr. 1974; 39:79–85.
108. O'Brien EJ. Acta. Crystallogr. 1967; 23:92–106. [PubMed: 6072867]
109. Tret'yak SM, Mitkevich VV, Sukhodub LF. Crystallogr. Rep. 1987; 32:1268.
110. Thewalt U, Bugg CE, Marsh RE. Acta. Cryst. B. 1971; 27:2358–2363.
111. Chickos JS, Acree WE. J. Phys. Chem. Ref. Data. 2002; 31:537–698.

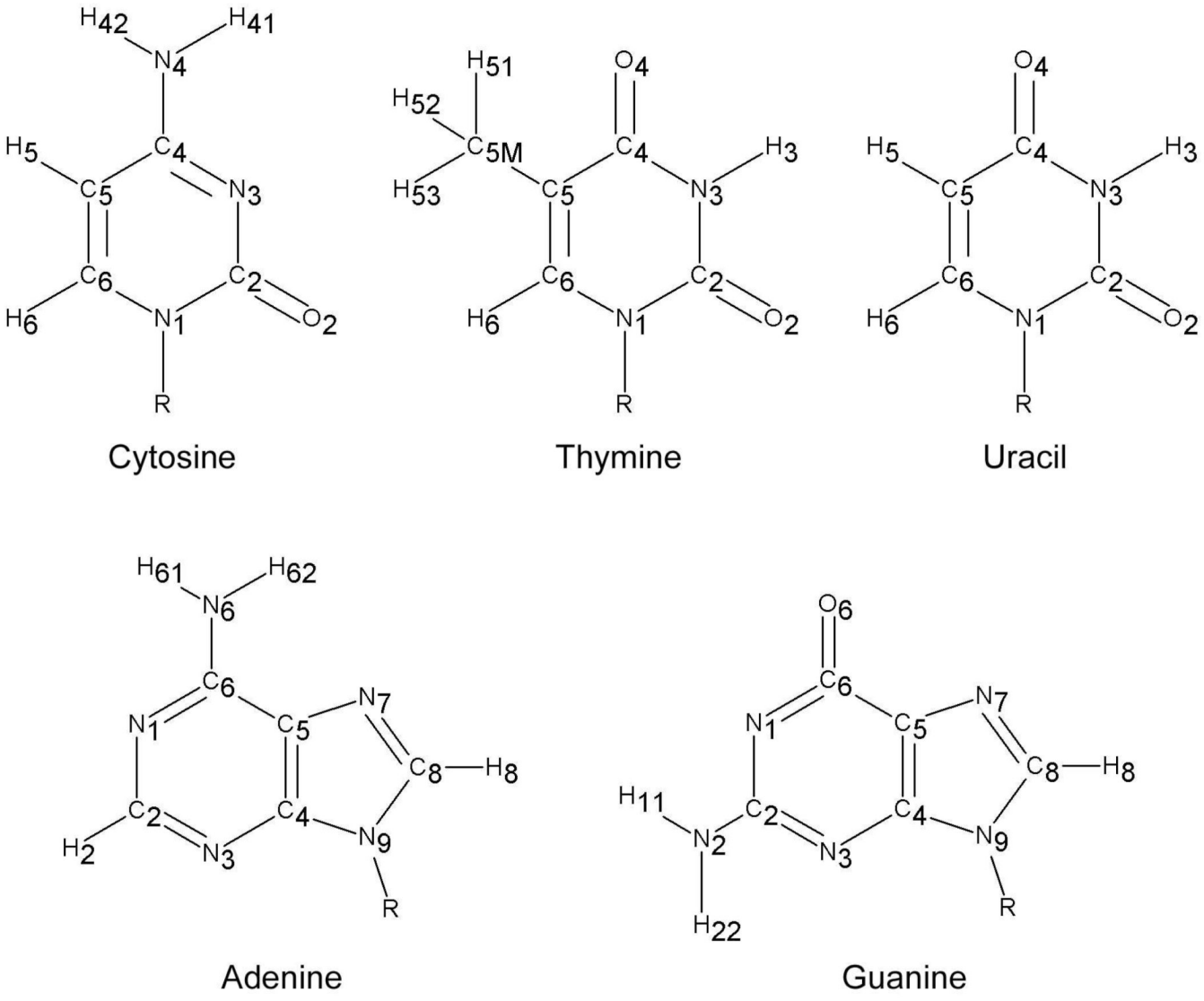

Figure 1.
Nucleic acid bases used as model compounds, R = H or $–CH_3$.

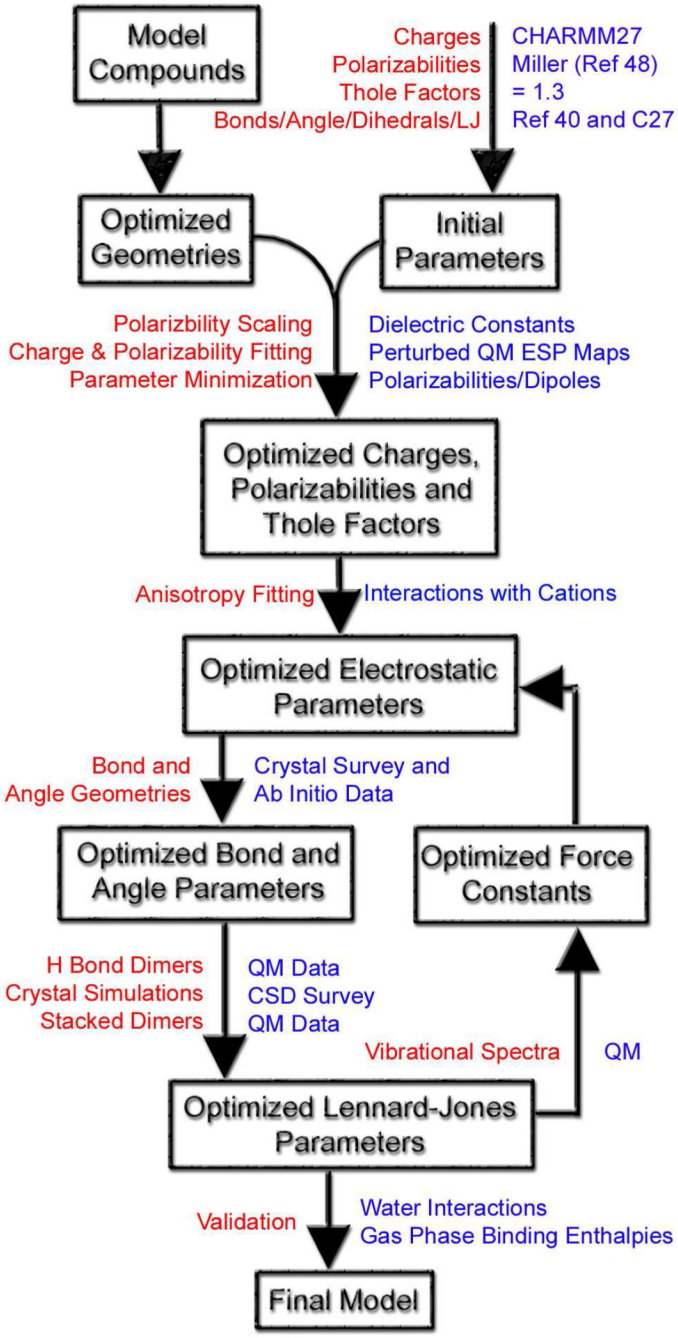

Figure 2.
Schematic description of the parameter optimization procedure employed in this work.

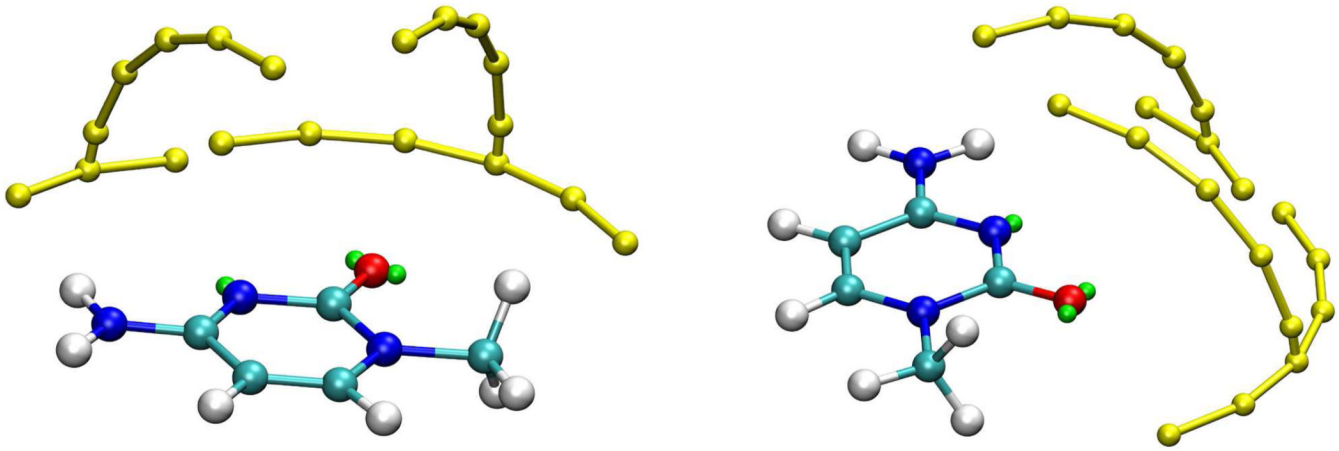

Figure 3.
Interaction orientations of Me-Cytosine base with sodium cations (yellow) located on in-plane and out-of-plane arcs placed in the vicinity of lone-pair (green) carrying sites.

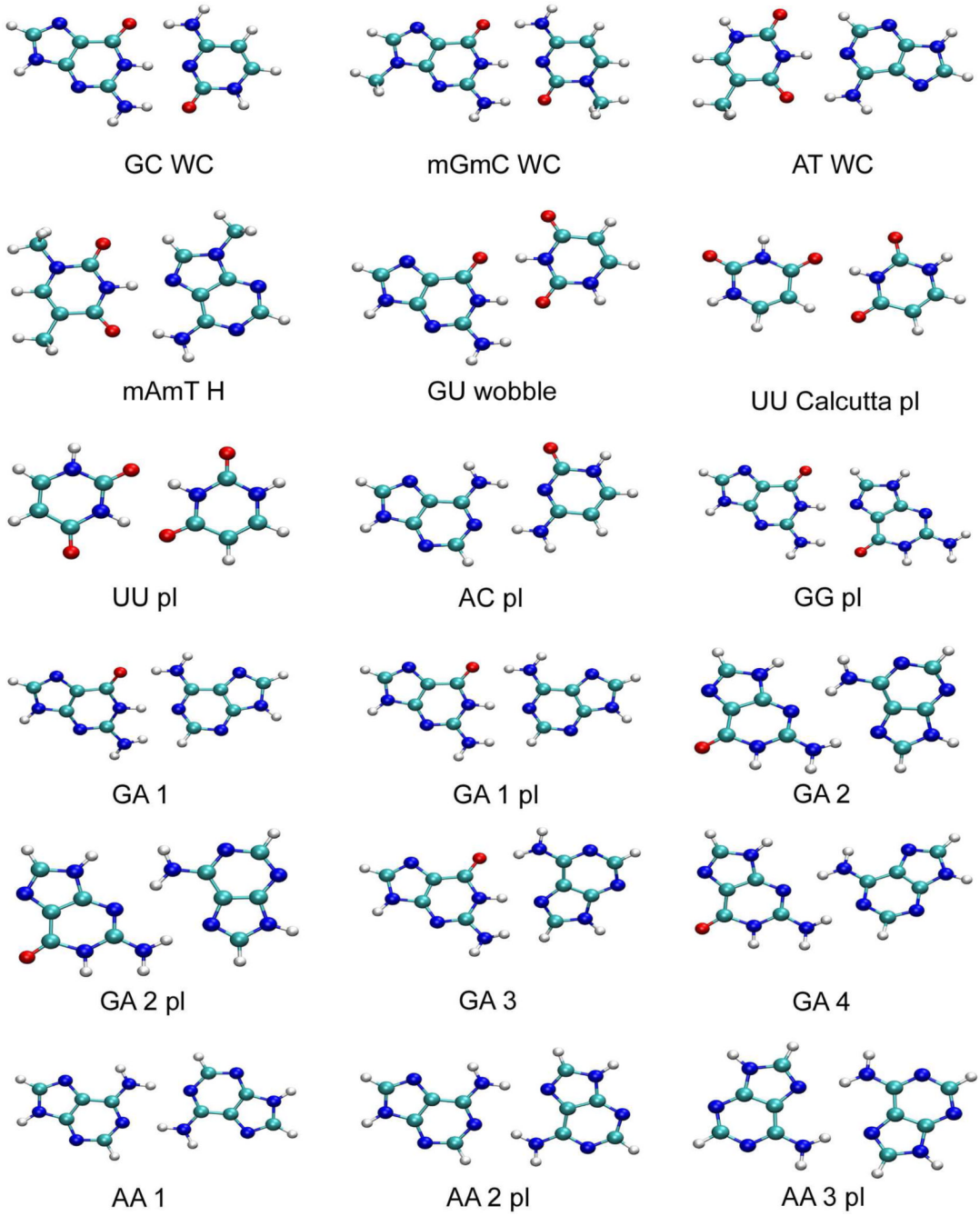

Figure 4.
The 18 hydrogen bonded base-pair structures used in the optimization and evaluation of LJ parameters.

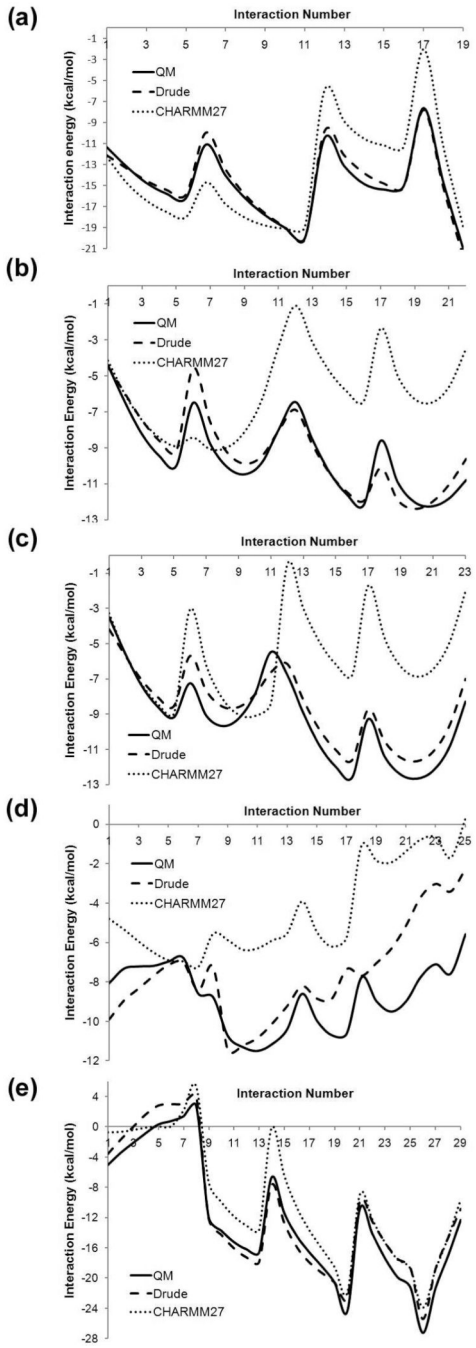

Figure 5.

Interaction energies for sodium cations on arcs around base H bond acceptors. (a) Me-Cyt (b) Me-Thy (c) Me-Ura (d) Me-Ade (e) Me-Gua. The orientations corresponding to each of the interaction numbers are listed in Tables S17–S21.

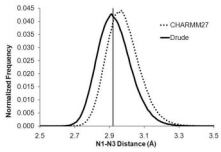

Figure 6.

N1–N3 distance distributions taken from simulations of the 9-ethyl-guanine – 1-methyl-cytosine base pair crystal performed using the CHARMM27 and Drude polarizable force fields. The vertical line on the graph represents the experimental value taken from the crystal structure, 2.91 Å.

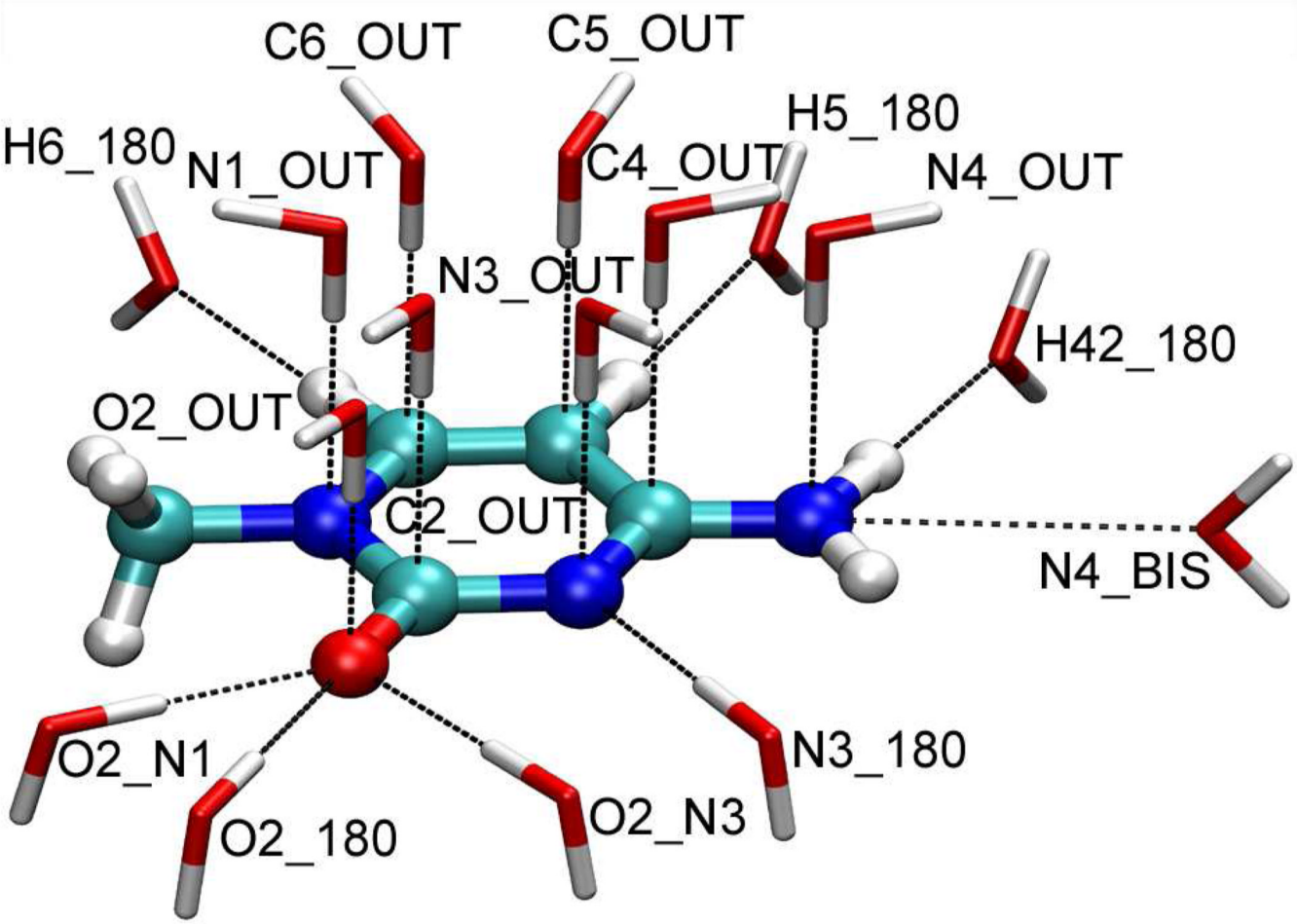

Figure 7.
Orientations of interactions between Methyl-Cytosine and water molecules. Analysis was performed on base-water monohydrates; for the purpose of visualization, in this figure all water interaction orientations are displayed simultaneously.

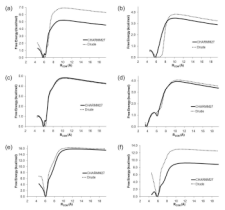

Figure 8.
Potentials of mean force (PMFs) calculated for a series of base-base dimers at 298 K using both the CHARMM27 and Drude polarizable force fields. (a) Me-Ade/Me-Ura (b) Me-Ura/Me-Ura (c) Me-Ade/Me-Thy (d) Me-Thy/Me-Thy (e) Me-Gua/Me-Cyt (f) Me-Cyt/Me-Cyt.

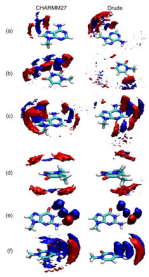

Figure 9.
Density plots for nucleic acid base pairs, calculated from gas phase MD simulations in the regions around the PMF minima. In all cases, one molecule is arbitrarily treated as fixed, with atomic densities of its partner molecule calculated relative to it. O atom density is shown in red; N atom density is shown in blue. Densities were calculated based on 0.2 Å cubic. (a) Me-Ade/Me-Ura; (b) Me-Ura/Me-Ura; (c) Me-Ade/Me-Thy; (d) Me-Thy/Me-Thy; (e) Me-Gua/Me-Cyt; (f) Me-Cyt/Me-Cyt.

Table 1

Components of molecular gas-phase polarizability tensor and isotropic polarizability, Å^3, of methylated pyrimidine bases. QM results calculated at the B3LYP/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ level of theory.

	Me-Cyt			Me-Thy			Me-Ura		
Comp.	**QM**	**QM*0.85**	**Drude**	**QM**	**QM*0.85**	**Drude**	**QM**	**QM*0.85**	**Drude**
XX	18.93	16.09	16.21	19.67	16.72	15.66	17.15	14.58	15.97
YY	13.94	11.85	11.78	14.94	12.70	13.15	12.80	10.88	11.55
ZZ	8.22	6.99	6.40	8.89	7.56	6.71	7.54	6.41	6.18
Isotr	13.70	11.65	11.47	14.50	12.33	11.84	12.50	10.63	11.23

J Phys Chem B. Author manuscript; available in PMC 2012 January 27.

Table 2

Components of molecular gas-phase polarizability tensor and isotropic polarizability, $Å^3$, of methylated pyrimidine bases. QM results calculated at the B3LYIP/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ level of theory.

	Me-Ade			Me-Gua		
Comp	**QM**	**QM*0.85**	**Drude**	**QM**	**QM*0.85**	**Drude**
XX	22.17	18.84	18.90	23.04	19.58	21.88
YY	17.99	15.29	15.12	19.52	16.59	15.81
ZZ	9.44	8.02	8.02	9.72	8.26	8.12
Isotr	16.53	14.05	14.01	17.43	14.82	15.27

J Phys Chem B. Author manuscript; available in PMC 2012 January 27.

Table 3

Gas phase dipole moments of methylated and unsubstituted bases, in Debye. QM values calculated at B3LYP/aug-cc-pVDZ//MP2/6-31G* level of theory.

	QM	CH27	Drude	QM	CH27	Drude
	Me-Cyt			**H-Cyt**		
X	−1.95	−4.73	−2.14	−4.13	−5.26	−4.28
Y	5.77	6.97	5.78	5.03	5.87	5.10
Z	0.76	−0.03	−0.01	0.73	−0.02	0.18
Total	6.14	8.42	6.17	6.55	7.88	6.67
	Me-Thy			H-Thy		
X	−1.20	1.71	−1.47	4.48	4.51	4.64
Y	4.71	3.71	4.79	−0.72	0.02	−1.12
Z	0.00	0.06	0.04	−0.06	−0.15	−0.13
Total	4.86	4.08	5.01	4.54	4.51	4.78
	Me-Ura			H-Ura		
X	−3.42	−0.09	−2.32	−1.31	−0.48	−1.34
Y	3.70	3.70	3.66	4.41	4.27	4.72
Z	0.00	0.00	0.00	0.00	0.00	0.00
Total	5.04	3.70	4.33	4.60	4.30	4.91
	Me-Ade			H-Ade		
X	−2.43	−1.42	−2.43	−1.91	−2.29	−1.80
Y	1.30	2.38	1.39	1.60	1.84	1.78
Z	0.63	−0.12	0.57	0.71	0.00	0.62
Total	2.83	2.77	2.86	2.59	2.94	2.61
	Me-Gua			H-Gua		
X	−0.45	2.13	−0.82	2.88	4.59	2.33
Y	−7.02	−6.31	−6.80	−5.90	−5.94	−5.89
Z	0.95	0.99	0.90	0.84	0.95	0.89
Total	7.10	6.73	6.91	6.62	7.56	6.40

J Phys Chem B. Author manuscript; available in PMC 2012 January 27.

Table 4

Differences in interactions of bases with sodium cations placed on arcs, in kcal/mol.

	Me-Cyt		Me-Thy		Me-Ura		Me-Ade		Me-Gua	
	Drude	**CH27**	**Drude**	**CH27**	**Drude**	**CH27**	**Drude**	**CH27**	**Drude**	**CH27**
RMS in[a]	0.54	2.83	0.75	2.83	0.63	3.11	1.83	2.29	1.45	1.45
AVE in[a]	0.21	0.32	0.29	3.87	0.67	3.45	1.19	4.39	0.84	2.64
RMS out[b]	0.53	2.89	0.44	2.44	0.49	2.90	1.81	2.46	1.39	1.24
AVE out[b]	0.32	1.34	0.23	3.14	0.50	3.04	0.68	4.57	1.04	2.53
RMS all[c]	0.53	2.91	0.93	2.69	0.68	3.22	1.56	1.99	1.49	1.63
AVE all[c]	0.09	0.86	0.33	3.54	0.81	3.76	1.97	4.13	0.62	2.76

[a] In-plane position;

[b] Out-of-plane position;

[c] Sum of in-plane and out-of-plane positions.

Table 5

RMSDs of the internal geometries of the nucleic acid bases calculated via force field simulations versus crystallographic survey data from ref 53.

Base	Bonds, Å		Angles, degrees	
	CHARMM27	Drude	CHARMM27	Drude
Me-Cyt	0.008	0.019	2.1	2.3
Me-Thy	0.012	0.022	1.1	1.7
Me-Ura	0.011	0.024	1.1	2.1
Me-Ade	0.010	0.024	1.5	2.9
Me-Gua	0.012	0.025	2.7	3.8
Me-Gua[a]			0.8	3.1
RMS[a]	0.011	0.023	1.4	2.5

[a] Neglecting differences in angles related to geometry of NH_2 group

Table 6

Percentage RMSDs for low energy vibrational modes versus MP2/6-31G* data.

Base	% RMS	
	CHARMM27	Drude
Me-Cyt	6.1	11.5
Me-Thy	14.7	13.6
Me-Ura	12.7	9.3
Me-Ade	36.1	7.6
Me-Gua	35.3	6.6
RMS	24.3	10.1

Table 7

Base-base hydrogen bond minimum interaction energies (kcal/mol) and distances (Å). QM values are CCSD(T) complete basis set limit, taken from Ref 57. Geometries are shown in Figure 4.

Basepair	Distance	QM		CHARMM27		Drude	
		E	R	E	R	E	R
GC WC	H1-N3	−28.80	1.91	−23.79	1.90	−27.99	1.85
mGmC WC	H1-N3	−28.50	1.87	−23.95	1.92	−28.18	1.86
AT WC	H3-N1	−15.43	1.82	−12.34	1.87	−15.46	1.79
mAmT H	H3-N7	−16.27	1.75	−13.22	1.86	−13.12	1.68
GU wobble	H1-O2	−16.10	1.74	−12.67	1.83	−12.03	1.83
UU Calcutta	O4-H3	−9.80	1.85	−7.81	1.80	−7.54	1.90
UU pl	H3-O4	−12.60	1.81	−10.45	1.83	−9.20	1.88
AC pl	H61-N3	−15.90	1.89	−11.48	1.94	−17.85	1.74
GG pl	H1-N7	−18.40	1.85	−18.67	1.87	−19.14	1.70
GA 1	H21-N1	−17.50	1.82	−14.19	1.87	−17.18	1.70
GA 1 pl	H1-N1	−16.10	1.94	−13.97	2.00	−15.13	1.92
GA 2	H21-N7	−10.90	1.96	−10.78	1.98	−11.23	1.67
GA 2 pl	H21-N7	−10.50	1.93	−9.95	1.98	−10.99	1.64
GA 3	N7-H21	−16.80	1.82	−13.72	1.98	−13.94	1.72
GA 4	N1-H21	−12.10	1.92	−11.03	1.98	−13.84	1.75
AA 1 pl	N1-H61	−13.10	1.92	−11.33	1.96	−14.00	1.78
AA 2 pl	N1-H62	−12.30	1.95	−11.07	1.98	−11.48	1.77
AA 3 pl	H62-N7	−10.90	1.98	−10.76	1.97	−11.03	1.63
RMSD				2.75	0.06	1.86	0.16
AVE Err				2.27	0.04	0.70	−0.10

Table 8

Intra-strand base stacking interaction energies, kcal/mol

Bases	MP2[a]	SCS[b]	DFT[c]	CH27	Drude
AT-AT	−16.44	−11.10	−11.38	−15.37	−13.23
AT-CG	−14.76	−9.32	−9.81	−14.22	−14.02
AT-GC	−14.90	−9.34	−10.02	−15.82	−14.73
AT-TA	−12.21	−6.62	−7.32	−9.36	−12.30
CG-AT	−15.55	−10.33	−11.31	−14.92	−13.15
CG-CG	−13.54	−7.98	−8.57	−12.74	−13.33
CG-GC	−17.50	−11.82	−13.13	−16.12	−12.75
GC-AT	−15.09	−9.82	−9.96	−15.90	−12.91
GC-CG	−16.53	−11.15	−11.35	−17.46	−13.42
TA-AT	−14.81	−9.96	−10.53	−14.41	−11.30
Ave Err v MP2				0.50	2.02
RMSD v MP2				1.23	2.58

[a] MP2 energies;

[b] spin-component-scaled MP2;

[c] DFT-SAPT.67

Table 9

Molecular volumes[a] of base crystals (Å^3) along with percent deviation from the corresponding experimental values.

Base	Temp, K	CH27		Drude		Exp.
		V_m	% dev.	V_m	% dev.	V_m
Cyt	298	116.16 (0.29)	−1.6	115.70 (0.06)	−2.0	118.1[b]
Thy	298	139.60 (0.05)	−3.5	145.25 (0.25)	0.4	144.7[c]
Ura	298	113.81 (0.53)	−1.7	117.36 (0.10)	1.4	115.8[d]
Me-Cyt	298	141.10 (0.08)	−1.4	144.09 (0.05)	0.7	143.1[e]
Me-Thy	298	169.48 (0.09)	0.3	172.75 (0.13)	2.3	168.9[f]
Me-Ura	15	135.93 (0.01)	1.4	135.34 (0.01)	0.9	134.1[g]
Me-Ade	126	163.27 (0.33)	−0.6	160.97 (0.08)	−2.0	164.3[h]
Et-Gua	298	212.79 (0.65)	−2.6	217.27 (0.50)	−0.5	218.4[i]
MT:MA[j]	298	336.92 (1.04)	0.8	333.34 (3.07)	−0.3	334.3[k]
EG:MC[l]	298	343.90 (0.14)	−0.3	348.77 (0.22)	1.1	344.9[m]
H-Ade:w[n]	298	221.38 (1.19)	1.1	216.10 (3.09)	−1.3	218.9[o]
H-Gua:w[n]	298	170.85 (0.11)	1.4	177.77 (0.03)	5.5	168.5[p]
RMSD		2.8	1.7%	3.5	2.1 %	
Ave Error			−0.6%		0.5%	

[a] standard deviation is shown in brackets;

[b] Experimental datum from Ref 100;

[c] Experimental datum from Ref 101;

[d] Experimental datum from Ref 102;

[e] Experimental data from Ref 103;

[f] Experimental datum from Ref 104;

[g] Experimental datum from Ref 105;

[h] Experimental datum from Ref 106;

[i] Experimental datum from Ref 107;

[j] 1-Methyl-Thymine : 9-Methyl-Adenine;

[k] Experimental data from Ref 84;

[l] 9-Ethyl-Guanine : 1-Methyl-Cytosine;

[m] Experimental data from Ref 108;

[n] Hydrated Crystal;

[o] Experimental datum from Ref 109;

[p] Experimental datum from Ref 110.

Table 10

Heats of sublimation[a] of bases (kcal/mol) along with percent deviation from the corresponding experimental value.

Base	Temp,	CH27		Drude		Exp.[b]
	K	ΔH_{sub}	% dev.	ΔH_{sub}	% dev.	ΔH_{sub}
Cyt	453	34.10 (0.20)	−3.1	34.98 (0.04)	−0.6	35.18
Thy	403	27.52 (0.07)	−7.4	28.88 (0.02)	−2.9	29.73
Ura	439	27.75 (0.08)	−8.6	28.67 (0.03)	−5.5	30.35
Me-Cyt	298	36.58 (0.18)	2.6	38.14 (0.01)	7.0	35.64
Me-Thy	398	25.80 (0.10)	−13.2	28.79 (0.42)	−3.2	29.73
Me-Ura	398	24.98 (0.19)	−7.1	27.92 (0.06)	3.8	26.89
Me-Ade	428	37.52 (7.76)	29.0	29.75 (0.44)	2.3	29.09
RMSD		3.85	13.2%	1.32	4.1%	
			−1.1%		0.1%	

[a] standard deviation is shown in brackets.

[b] Experimental data from Ref 111.

Table 11

Statistical analysis of the differences in the minimum interaction energies and geometries of the Methylated bases with water between QM MP2/6-31G(d)//RI-MP2/cc-pVQZ data and the CHARMM27 and polarizable force fields.

	CH27		Drude	
Me-Cyt	**ΔR**	**ΔE**	**ΔR**	**ΔE**
RMS all	0.17	1.75	0.10	1.09
AVE all	−0.03	−0.79	−0.02	−0.33
RMS out of plane	0.20	2.32	0.09	0.99
AVE out of plane	−0.01	−0.58	−0.05	−0.22
RMS in plane	0.15	1.17	0.11	1.16
AVE in plane	−0.04	−0.95	0.01	−0.42
Me-Thy				
RMS all	0.14	1.20	0.09	1.12
AVE all	−0.08	−0.23	0.02	−0.16
RMS out of plane	0.16	1.12	0.10	0.63
AVE out of plane	−0.07	−0.62	−0.06	−0.46
RMS in plane	0.14	1.25	0.08	1.38
AVE in plane	−0.09	0.06	0.01	0.07
Me-Ura				
RMS all	0.17	1.22	0.12	1.10
AVE all	−0.07	−0.34	0.01	−0.08
RMS out of plane	0.19	1.22	0.13	0.81
AVE out of plane	−0.07	−0.64	−0.04	−0.39
RMS in plane	0.15	1.22	0.11	1.26
AVE in plane	−0.08	−0.14	0.05	0.12
Me-Ade				
RMS all	0.20	1.05	0.19	1.43
AVE all	0.01	−0.21	−0.05	−0.59
RMS out of plane	0.25	1.33	0.17	1.05
AVE out of plane	0.01	−0.12	−0.04	−0.59
RMS in plane	0.11	0.50	0.22	1.80
AVE in plane	0.00	−0.34	−0.05	−0.59
Me-Gua				
RMS all	0.16	1.31	0.17	0.66
AVE all	−0.07	−0.86	−0.04	−0.26
RMS out of plane	0.18	1.33	0.15	0.63
AVE out of plane	−0.07	−0.88	−0.07	−0.23
RMS in plane	0.14	1.29	0.18	0.69
AVE in plane	−0.07	−0.85	−0.01	−0.27

	CH27		Drude	
Me-Cyt	**ΔR**	**ΔE**	**ΔR**	**ΔE**
Overall				
RMS all	0.17	1.32	0.14	1.10
Ave all	−0.05	−0.49	−0.02	−0.29
RMS out of plane	0.20	1.50	0.14	0.85
Ave out of plane	−0.04	−0.54	−0.05	−0.39
RMS in plane	0.14	1.02	0.15	1.48
Ave in plane	−0.05	−0.34	0.00	−0.26

Table 12

Thermodynamic properties of gas phase base-base dimerization, in kcal/mol, obtained from PMF calculations.

Base Pair	Exp[a]	CHARMM27				Drude			
	ΔH_{bind}	ΔH_{bind}	Diff	ΔG_{bind}	$T\Delta S_{bind}$	ΔH_{bind}	Diff	ΔG_{bind}	$T\Delta S_{bind}$
mA:mU	14.5	13.3	−1.2	4.9	8.4	15.7	1.2	7.3	8.4
mU:mU	9.5	10.1	0.6	3.2	6.9	9.7	0.2	4.2	5.5
mA:mT	13.0	12.9	−0.1	4.4	8.5	12.1	−0.9	4.4	7.7
mT:mT	9.0	9.0	0.0	2.6	6.4	9.1	0.1	2.9	6.2
mG:mC	21.0	20.2	−0.8	16.5	3.7	22.8	1.8	20.4	2.4
mC:mC	16.0	18.9	2.9	8.1	6.2	19.7	3.7	15.7	4.1
Ave Diff			0.2				0.9		
RMSD			1.3				1.7		

[a]Experimental values are from Ref 74.