An All-Atom Empirical Energy Function for the Simulation of Nucleic Acids†

Alexander D. MacKerell, Jr.,*,‡,⊥ Joanna Wiórkiewicz-Kuczera,‡ and Martin Karplus*,‡

Contribution from the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, and Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland at Baltimore, Baltimore, Maryland 21201

Received June 13, 1994⊗

Abstract: Nucleic acid parameters are developed for the all-atom empirical energy function used in the CHARMM program. The parameters were determined by use of results for model compounds, including the nucleic acid bases, dimethyl phosphate and anionic and dianionic methyl phosphate, ribose, and deoxyribose. Internal parameters (bond length, bond angle, Urey−Bradley, dihedral, and improper dihedral terms) were chosen to reproduce geometries and vibrational spectra from experimental crystal structures, infrared and Raman spectroscopic data, and *ab initio* calculations. Interaction parameters (electrostatic and van der Waals terms) were derived from 6-31G* *ab initio* interaction energies and geometries for water molecules bonded to polar sites of the model compounds and from the experimentally measured gas phase Watson−Crick base pair energies and geometries, base heats of sublimation, and experimental and 6-31G* *ab initio* dipole moments. Emphasis was placed on a proper balance between solvent−solvent, solvent−solute, and solute−solute interactions with reference to the TIP3P water model. Tests on nucleic acid base crystals showed satisfactory agreement between calculated and experimental values for the lattice parameters, nonbonded interactions, and heats of sublimation. Base pair variations in stacking energies are consistent with experiment and *ab initio* calculations. Further testing was performed on GpC and B-DNA dodecamer crystal structures, including water molecules and counterions. Simulations of these systems revealed the parameters to accurately reproduce Watson−Crick base pairing, internal geometries including the backbone dihedrals, sugar puckering and glycosidic linkages, and the hydration of the nucleic acids. The present parameters should be useful for modeling and simulation studies of nucleic acids, including both structural and energetic analysis. Further, they provide a parameter set that is consistent with the protein parameters in CHARMM so that complexes between proteins and nucleic acids can be modeled. A list of the parameter values is included in an Appendix (supporting information). In the test simulations, a number of interesting results were obtained. Significant anharmonicity is present due to nonbonded terms in certain bond angles (P−O−C) of the phosphate group that may play a role in the observed variation of these angles. There is a wide range of stacking energies of the bases of B-DNA due to repulsive electrostatic interactions. The motion of the two strands appears to be correlated for bases involved in Watson−Crick interactions while there is a lack of correlation for the sugar and phosphate moieties. The importance of conformational substates in the B-DNA dodecamer is pointed out. Comparisons with nucleic acid simulations made with other energy functions show the importance of the parameters in determining the internal geometry and the interactions with the surroundings. A summary of these results is given in the concluding section.

I. Introduction

Theoretical treatments of macromolecules of biological interest are being widely used to obtain an increased understanding of their properties.[1,2] Most calculations are based on empirical energy functions,[3] which have been employed in energy minimization and molecular dynamics simulations for the analysis of the structure, energetics, dynamics, and thermodynamics of such systems. Empirical approaches make possible the simulation of systems containing thousands of atoms, such as proteins or nucleic acids in aqueous solution. However, the approximations inherent in the empirical energy functions limit the accuracy of the results. In this paper we describe the development of a refined all-atom empirical energy function for nucleic acids. Early empirical energy functions treated hydrogen atoms as part of the heavy atom to which they were bonded and had explicit hydrogen bonding terms.[3−5] More recently, polar hydrogen sets were developed; they include polar hydrogens (those bonded to O, N, and S) as explicit atoms, but use extended atoms for nonpolar hydrogens (those bonded to carbon).[3,4,6] All-atom models exist[7,8] but the emphasis and details of the approaches used for the parametrization is different from that employed in the present work. In developing the parameters an effort was made to maintain consistency with protein, lipid, and sugar parameters, to balance

† Supported in part by a grant from the National Science Foundation and a gift from Molecular Simulations, Inc.

‡ Harvard University.

⊥ University of Maryland at Baltimore.

⊗ Abstract published in *Advance ACS Abstracts*, November 1, 1995.

(1) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.

(2) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*; Wiley Series on Advances in Chemical Physics; Prigogine, I., Rice, S., Eds.; John Wiley & Sons: New York, 1988; Vol. LXXI.

(3) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187−217.

(4) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(5) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230−252.

(6) Nilsson, L.; Karplus, M. *J. Comp. Chem.* **1986**, *7*, 591−616.

(7) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(8) MacKerell et al. Manuscript in preparation.

0002-7863/95/1517-11946$09.00/0

the intramolecular and intermolecular portions of the energy function, and to correctly represent the solvent–solvent, solvent–solute, and solute–solute contributions to the energy function. The latter, which may be referred to as the "interaction triad", is essential for accurate calculations of the properties of macromolecules in aqueous solution and in crystals. The approach described in this paper is easily extended to related molecules, such as nucleic acid analogs.

Numerous studies based on empirical energy functions have been performed on DNA and RNA,[9-12] including the use of simulated annealling for the determination of structures from NMR data.[13,14] These calculations have been made in vacuum,[15-18] with explicit solvent,[19,20] and with the inclusion of cations.[21-23] Due to the highly charged nature of DNA and RNA, it is necessary to minimize electrostatic effects in vacuum simulations.[17,24-26] Approaches toward this end have relied on the use of scaled charges on the phosphate moieties,[17] distance-dependent dielectric constants[17,23,24] and/or the use of hydration spheres or solvatons.[26] Problems, including helix unwinding and unphysical distortions, have been observed in some cases.[16] This emphasizes the importance of solvation in DNA and RNA simulations.

Molecular dynamics simulations of nucleic acid duplexes in solution using a full charge representation have difficulty in producing stable helical structures.[27] In the study of McConnell et al. the charges on the phosphates were reduced to −0.24 eu to implicitly treat counterion contributions. The simulation yielded stable helical structures for the d(CGCGAATTCGCG) dodecamer and it was shown that simulations of a nanosecond or more are required to adequately sample conformational space. Based on the problems encountered in a large number of nucleic acid duplex simulations performed to date the need for empirical force field parameters allowing for stable simulations of DNA and RNA in solution is evident. Additionally, the requirement that simulations involving on the order of 10 000 atoms be carried into the nanosecond time regime makes the inclusion of the explicit treatment of electronic polarization in the force field prohibitive due to the approximately 2- to 4-fold increase in CPU requirements.[28]

The present paper describes the development of an all-atom representation of nucleic acids based on the form of the empirical energy implemented in the CHARMM program.[3] The parameters are designed to be compatible with the recently developed all-atom representation for proteins.[8] Neither set includes an explicit hydrogen bonding term in the force field. Such a term was included in the earlier united atom representation of nucleic acids[6] but was shown to be unnecessary when the polar hydrogen model was introduced for proteins.[29] Emphasis has been placed on determining parameters that yield agreement with both experimental and *ab initio* data for model compounds representing the base, sugar, and phosphate moieties of nucleic acids, while at the same time providing accurate results for nucleic acids and their building blocks in solution and crystal simulations. An iterative refinement of the internal parameters for bonds, bond angles, dihedral angles, and the external parameters for electrostatic and van der Waals interactions has been employed to obtain consistent and accurate results. The internal parameters for the structure of molecular fragments were optimized to yield agreement with average geometries collected from surveys of the Cambridge Crystal Databank (CCDB[30]), as previously published,[31] or, for thymine, done as part of the present study. Experimentally determined vibrational spectra, supplemented as necessary by *ab initio* calculations, were used in the optimization of the force constants. Optimization was based on both the frequencies and the normal mode assignments of the internal coordinates. Reoptimization of the internal parameters was done following adjustment of the interaction parameters to maintain consistency between the internal and external contributions to the potential energy function.

Interaction parameters were optimized to maintain a proper balance between solvent–solvent, solvent–solute, and solute–solute interactions. The van der Waals parameters were obtained from previously reported values, from experimental base pairing geometries, from *ab initio* interaction energies and geometries between water and the model compound, and from heats of sublimation of some base analogs. For consistency certain van der Waals parameters were set equal to those for related systems. These include the polar hydrogen and hydroxyl and ester oxygens, for which the CHARMM modified TIP3P model parameters[32,33] were used, and the aliphatic carbon and hydrogen, for which those developed for small aliphatic molecules were employed.[34] Charges were based on reproducing scaled *ab initio* 6-31G* interaction energies between model compounds and water, dipole moments of the model compounds, and, where available, heats of sublimation of the model compounds. The present approach is different from one where the charges are obtained directly from the *ab initio* molecular orbitals via Mulliken population analysis or by fitting the point charges to the *ab initio* derived electrostatic potential.[35-37] It is based on earlier work on the development of the polar hydrogen

(9) Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. *J. Biomol. Struct. Dyn.* **1983**, *1*, 231–252.
(10) Srinivasan, J.; Withka, J. M.; Beveridge, D. L. *Biophys. J.* **1990**, *58*, 533–547.
(11) Rao, S. N.; Kollman, P. *Biopolymers* **1990**, *29*, 517–532.
(12) Fritsch, V.; Westhof, E. *J. Comp. Chem.* **1991**, *12*, 147–166.
(13) Nilsson, L.; Clore, G. M.; Gronenborn, A. M.; Brunger, A. T.; Karplus, M. *J. Mol. Biol.* **1986**, *188*, 455–475.
(14) Nikonowicz, E. P.; Meadows, R. P.; Fagan, P.; Gorenstein, D. G. *Biochemistry* **1991**, *30*, 1323–1334.
(15) Keepers, J. W.; Kollman, P. A.; Weiner, P.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5537–5541.
(16) Levitt, M. *Cold Spring Harbor Symp. Quant. Biol.* **1983**, *47*, 251–275.
(17) Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. *J. Biomol. Struct. Dyn.* **1983**, *1*, 231–252.
(18) Westhof, E.; Chevrier, B.; Gallion, S. L.; Weiner, P. K.; Levy, R. M. *J. Mol. Biol.* **1986**, *190*, 699–712.
(19) Seibel, G. L.; Singh, U. C.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6537–6540.
(20) Swamy, K. N.; Clementi, E. *Biopolymers* **1987**, *26*, 1901–1927.
(21) Laaksonen, A.; Nilsson, L.; Schneider, C.; Guschlbauer *Nucleic Acid Res.* **1989**, *6*, 2831–2837.
(22) Jung, S.-H. Simulation of DNA and its Interaction with Ligands, Ph.D. Thesis, Harvard University, Department of Chemistry, 1989.
(23) Veal, J. M.; Gao, X.; Brown, F. K. *J. Am. Chem. Soc.* **1993**, *115*, 7139–7145.
(24) Srinivasan, J.; Withka, J. M.; Beveridge, D. L. *Biophys. J.* **1990**, *58*, 533–547.
(25) Fritsch, V.; Westhof, E. *J. Am. Chem. Soc.* **1991**, *113*, 8271–8277.
(26) Ravishanker, G.; Swaminathan, S.; Beveridge, D. L.; Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 669–699.
(27) McConnell, K. J.; Nirmala, R.; Young, M. A.; Ravishankar, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1994**, *116*, 4461–4462.
(28) Smith, P. E.; Pettitt, B. M. *J. Phys. Chem.* **1994**, *98*, 9700–9711.
(29) Reiher, W.; Karplus, M. Unpublished (see also ref 33).
(30) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr. B35*, 2331–2339.
(31) Taylor, R.; Kennard, O. *J. Mol. Struct.* **1982**, *78*, 1–28.
(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(33) Reiher, W. E., III Ph.D. Thesis, Department of Chemistry, Harvard University, 1985.
(34) MacKerell, A. D., Jr.; Fischer, S.; Field, M. J.; Watanabe, M.; Karplus, M. Manuscript in preparation.
(35) Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1984**, *5*, 129–145.
(36) Chirlian, L. E.; Francl, M. M. *J. Comp. Chem.* **1987**, *8*, 894–905.
(37) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

model (CHARMM 19) for proteins.[29,33] The fitted 6-31G* interaction energies between the model compounds and water were scaled by 1.16[29,33] for polar species, while the 6-31G* interaction energies were used directly for charged species. The 1.16 scale factor, which is equal to the ratio of the water dimer interaction energy from the TIP3P model[32] and that from a 6-31G* calculation, is applied to compensate for the absence of explicit polarization effects in the potential energy function. The value of 1.16 assumes that the model compounds have polarizabilities similar to that of water. This is, of course, an oversimplification, but it is better than neglecting the average polarization effect altogether. In addition to fitting the scaled 6-31G* interaction energies, the charges (and van der Waals parameters) are chosen to yield optimized minimum energy geometries that are approximately 0.1 to 0.2 Å shorter than the 6-31G* value. Such reduced distances are used to account for the lack of polarization effects and for the absence of dispersion ($1/R^6$) attraction in the Hartree−Fock approximation. This was suggested by Jorgensen and co-workers,[39,40] who have adopted the Reiher/Karplus approach[29,33] of using the water interaction energies for determining the charges; they do not introduce a scale factor. The Reiher/Karplus approach produces pure solvent properties in good agreement with experiment for a number of polar compounds.[38] In the determination of the charges for the nucleic acid bases the experimental base pairing energies, dipole moment, and heats of sublimation[41] were used as additional constraints. The fitting of the charges was done so as to maintain neutral or unit charge groups of between 2 and 10 atoms. Such groups are used in the group list option of the CHARMM program[3] and simplify the extrapolation of charges from the model compounds to the full nucleic acid moieties. Transfer of charges from the model compounds to nucleic acid fragments was performed by taking the charge of the hydrogen atom being removed and adding the charge to the attached heavy atom.

Section II describes the methodolgy used in the calculations. Section III is concerned with the internal and external parameters of the nucleic acid bases and also includes results on crystal minimzations and molecular dynamics simulations of several base analogs. Section IV presents the internal and external parametrization for the phosphate group and a summary of the sugar parametrization results; details of the sugar parametrization will be presented elsewhere.[42] Section V gives tests of the parameters by crystal simulations of nucleic acids, including GpC[43] and a B-DNA dodecamer[44] and a summary of results on a Z-DNA hexamer.[45] A concluding discussion is presented in Section VI. The parameters are listed in the Appendix, which is included with the supporting information.

II. Formulations and Methodology

Molecular mechanics calculations were performed using the program CHARMM,[3] for which the empirical energy function was designed. The energy function includes internal (bonding) terms for bond stretching, angle bending, Urey−Bradley 1,3 interaction, torsional rotation, and out-of-plane (improper) motion while the external (nonbonded) interactions are represented by a Lennard-Jones 6-12 term for the van der Waals repulsion and dispersion interaction and a Coulomb term for the charge−charge interactions; in the latter both partial and full charges are included. The potential energy function has the form

$$U(\vec{R}) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angle}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{UB}} K_{\text{UB}}(S - S_0)^2 + \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} K_{\text{imp}}(\varphi - \varphi_0)^2 + \sum_{\text{nonbond}} \epsilon\left[\left(\frac{R_{\text{min},ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\text{min},ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi e r_{ij}} \quad (1)$$

where K_b, K_θ, K_{UB}, K_χ, and K_{imp} are the bond, bond angle, Urey−Bradley, dihedral angle, and improper dihedral angle force constants, respectively; b, θ, S, χ, and φ are the bond length, bond angle, Urey−Bradley 1,3 distance, dihedral torsion angle, and improper dihedral angle, respectively, with the subscript zero representing the equilibrium values for the individual terms. Coulomb and Lennard-Jones 6-12 terms contribute to the external or nonbonded interactions; ϵ is the Lennard-Jones well-depth and R_{min} is the distance at the Lennard-Jones minimum, q_i is the partial atomic charge, and e is the dielectric "constant" which can depend on r_{ij}; r_{ij} is the distance between atoms i and j, respectively. Given $\vec{R}$, the vector of the coordinates of the atoms, the various distances and angles required to evaluate $U(\vec{R})$ in eq 1 are readily determined. All possible bonds, bond angles, and dihedral angles are included in $U(\vec{R})$, while a limited number of Urey−Bradley terms and improper dihedral angles are used to optimize the fit to the vibrational spectra; a complete list of the parameters is given in the Appendix (see the supporting information). As can be seen from eq 1 only the quadratic term is included in the Urey−Bradley function; this is in accord with an analysis[46] that shows the linear term can be omitted when Cartesian coordinates are used and the minimum energy structure is employed for determining the vibrational frequencies. For the model compounds all possible nonbonded interactions were included with no cutoff and a dielectric constant of unity was used, unless noted. Intramolecular nonbonded interactions between atoms attached by three (1−4 interactions) or more covalent bonds were included in the energy calculations and were not scaled.[47] The Lennard-Jones parameters between different atom types are derived by use of the widely employed Lorentz−Berthelodt combination rule.[48]

Vacuum minimizations of model compounds were performed starting with 100 steps of the ABNR minimizer[3] followed by minimization with the Newton−Raphson minimizer to a gradient of 10^{-6} (kcal mol^{-1})/Å. Analysis of the normal modes was performed via the MOLVIB program that provides the potential energy distribution of each mode.[49] Crystal minimizations and dynamics were performed in three stages with a switch truncation for the van der Waals interaction and a shift truncation for the electrostatic interaction;[3] the truncation distances used are listed with the results. Initially, the hydrogen positions were optimized for 50 steps with the ABNR minimizer with all non-hydrogen atoms and the lattice parameters fixed. This was followed by 200 ABNR steps with all atoms free but with the lattice parameters fixed; finally all atoms and the lattice parameters were allowed to vary. The final stage was performed with 1000 ABNR steps or until the rms gradiant of the energy was below 10^{-3} (kcal mol^{-1})/Å. Heats of sublimation, ΔH, were determined via

$$\Delta H = E_{\text{vac}} - E_{\text{xtal}} + RT \quad (2)$$

Here E_{vac} is the potential energy of the isolated molecule in the gas

(38) MacKerell, A. D., Jr.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 10559−10560.
(39) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569−587.
(40) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276−1284.
(41) Warshel, A.; Lifson, S. *J. Chem. Phys.* **1970**, *53*, 582−594.
(42) MacKerell, A. D., Jr.; Karplus, M. Manuscript in preparation.
(43) Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145−167.
(44) Privé, G. G.; Yanagi, R.; Dickerson, R. E. J. *Mol. Biol.* **1991**, *217*, 177−199.
(45) Gessnet, R. V.; Quigley, G. J.; Wang, A. H.-J.; van der Marel, G.; van Boom, J. H.; Rich, A. *Biochemistry* **1985**, *24*, 237−240.
(46) Pettitt, B. M.; Karplus, M. *J. Am. Chem. Soc.* **1985**, *107*, 1166−1173.
(47) Smith, J. C.; Karplus, M. *J. Am. Chem. Soc.* **1992**, *114*, 801−812.
(48) Fischer, W.; Brickmann, J. *Ber. Bunsenges. Phys. Chem.* **1982**, *86*, 650
(49) Kuczera, K.; Wiórkiewicz-Kuczera, J. The MOLVIB Program.

phase and E_{xtal} is the potential energy of the molecule in the crystal. The RT term corresponds to the difference between the translational and rotational energy of the gas ($3RT$) plus the $PV = RT$ term and the classical (Dulong–Petit) energy ($3RT$) of the vibrational motion of the molecules in the crystal.[50] Lattice energies were calculated following the approach of Warshel and Lifson.[41] Changes in the internal vibrational motion in going from the gas to the crystal are neglected in eq 2; they are, however, included in some calculations of interaction energies.

Crystal molecular dynamics simulations of the base analogs and the GpC crystal were performed with the crystal facility of the CHARMM program. The leap-frog algorithm in the NVT ensemble[1] was used with a time step of 1 fs and a temperature coupling constant of 0.1 ps^{-1}.[51] Simulations were carried out for 20 ps with the 200 step ABNR minimized structure (see above) used as the starting point. Molecular dynamics simulations of the B-DNA dodecamer crystals were performed in the NVE ensemble with the leap algorithm and a time step of 1 fs. SHAKE[52] was invoked to constrain the length of covalent bonds involving hydrogens. In these simulations a heating period of 5 ps was used followed by a 5 ps equilibration period during which velocities were scaled if necessary; production runs were performed without velocity scaling in the NVE simulations. Analysis was performed using coordinate sets at 0.1-ps intervals except in the case of hydration properties where only coordinate sets from every 0.5-ps interval were used because of the large number of water molecules.

The B-DNA dodecamer duplex structure[44] was generated from a single strand, four octahedrally water coordinated Mg^{2+} ions and 49 additional water molecules, which comprises the asymmetric unit in the X-ray results. A Mg^{2+} water cluster and a water molecule, each on one of the crystallographic 2-fold axes, was removed to avoid having duplicate species in the same place when the IMAGE facility of CHARMM is applied to generate the crystal. This left 7 Mg^{2+} water clusters plus 97 additional waters in the primary unit cell. The total charge of the system is −4, suggesting that there are additional positive ions in the unit cell; they were neglected. To fill the remaining space between the primary DNA duplex and the surrounding images,[3] the entire system, including the images, was overlayed with a 29.3 × 26.5 × 39.1 Å water box which was centered on the primary atoms. Dimensions of the water box were selected to cover all heavy atoms in the dimer. All overlayed water molecules whose oxygen atoms were within 2.8 Å of any of the X-ray primary or image heavy atoms were removed. A total of 107 water molecules were added in this way and hydrogen atoms were generated using HBUILD.[53] All hydrogen atoms were then minimized for 50 steepest descent (SD) steps with fixed heavy atoms using a switch truncation between 10 and 12 Å for the van der Waals interactions and a shift truncation of 12 Å for the electrostatic interactions.[3] This was followed by minimization of the 107 added water molecules for 100 SD steps. Finally a series of constrained minimizations were performed with harmonic constraints on all heavy atoms. The system was subjected to 50 ABNR steps followed by a decrease in the harmonic constraints and reminimization. This was repeated with harmonic constraints of 100, 50, 10, 5, and 1 (kcal mol^{-1})/$Å^2$. The final structure, which was close to the X-ray structure (root mean square (rms) deviation of 0.2 Å for the BDNA heavy atoms) but with essentially all strain removed, was used to initiate the simulation.

Analysis of the parameters in terms of hydration of nucleic acids was performed via comparison of the average water oxygen to nucleic acid distances and hydration numbers of selected nucleic acid heteroatoms. Due to the crystal structures having only a limited number of solvent–nucleic acid interactions for the individual hydrogen bonding sites, the determination of radial distribution functions (RDF) is generally not feasible. RDFs between solvent and the phosphate anionic and ester oxygens of B-DNA (see Figure 17, below) were calculated because of the relatively large number of solvent–solute interactions. However, even here, the presence of discrete points in the experimental RDF emphasizes limitations in this type of analysis. Alternatively, hydration properties may be analyzed in terms of solvent–solute interactions occurring within a cutoff distance. In the present study a cutoff distance of 3.5 Å was employed. This distance corresponds to the first minimum in the TIP3P water model O to O RDF[32] and, therefore, is assumed in the present study to represent the outer limit of the first hydration shell. Applying this cutoff distance allows for the average water O to nucleic acid atom distance and hydration number to be obtained from the experimental X-ray structures, even though only a limited number of such interactions may be present. A similar methodology for the analysis of solvent nucleic acid interactions has been reported.[54] For consistency, the same approach was used for analysis of the dynamics simulations. Average distances and hydration numbers were obtained over the individual time frames in the trajectory and normalized for the particular types of nucleic acid hydrogen bonding sites used in the analysis. Some problems in comparing the X-ray and dynamics results can arise with this approach. Since the X-ray results often yield idealized solvent–solute geometries there is often an integer number of waters at a specific site. In the dynamics simulations the motion of the water molecules in specific hydration sites and, more importantly, the exchange of water molecules with the bulk solvent will introduce contributions to the average distance and hydration numbers from water molecules not in idealized solvent–solute geometries. In addition, water molecules not in the first hydration shell may transiently move to within 3.5 Å of the particular heteroatom. Such effects will tend to increase both the average distance and hydration number. To account for such contributions the RDFs for the various solvent–nucleic acid interactions were calculated from the dynamics trajectories and the positions of the first peaks from those RDFs presented in addition to the average distances. All analyses are based on the water oxygen to nucleic acid heteroatoms interactions. In specific cases (see below) where no solvent–nucleic acid interactions occurred within 3.5 Å in the X-ray results, the cutoff distance was increased to 5 or 8 Å to ensure inclusion of all the interactions in the crystal structure. This allows for a qualitative comparison of the experimental and calculated results and is especially useful when the number of interactions in the experimental structure does not allow for an accurate determination of a RDF.

Ab initio Hartree–Fock calculations to determine the minimum energy geometry and interaction energy between the model compounds and water were performed for the orientations shown in Figures 1 and 6. The monomer geometries were fixed and only the distance and, in some instances, the angles were varied. The Gaussian program[55] with default basis sets was used for the calculations. HF/3-21G optimized geometries were used for the bases, HF/6-31G* optimized geometries were used for the anionic and neutral phosphate model compounds, a HF/6-31+G* geometry was used for the dianionic methyl phosphate, and the experimental gas phase geometry was used for water.[56] All interaction energies and geometries were obtained at the HF/6-31G* level of theory. Along with the CHARMM program, the OPLS35 program,[57] with combination rules from the CHARMM energy function, was used in the optimization of the model compound nonbonded parameters.

III. Nucleic Acid Bases

(a) Internal Parameters. Model compounds used for the optimization of the nucleic acid bases included guanine, adenine, uracil, thymine, and cytosine along with certain methylated analogs (see Figure 1); the latter are used to aid in the

(50) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319–5327.

(51) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(52) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Chem. Phys.* **1977**, *23*, 327–341.

(53) Brünger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148–156.

(54) Seibel, G. L.; Singh, U. C.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6537–6540.

(55) Binkley, J. S.; Frisch, M.; Krishnam, R.; DeFrees, D. J.; Schlegel, H. B.; Whiteside, R. A.; Fluder, E.; Seeger, R.; Pople, J. A. GAUSSIAN 82, Releases A and H; Carnegie-Mellon University; Frisch, M. J.; Binkley, J. S.; Schlegel, H. B.; Raghavachari, K.; Melius, C. F.; Martin, R. L.; Stewart, J. J. P.; Bobrowicz, F. W.; Rohlfing, C. M.; Kahn, L. R.; Defrees, D. J.; Seeger, R.; Whitesides, R. A.; Fox, D. J.; Fleuder, E. M.; Pople, J. A. Gaussian 86, Carnegie-Mellon Quantum Chemistry Publishing Unit: Pittsburgh, PA, 1984.

(56) Hirato, E.; Sugisake, R.; Nielsen, C. J.; Sórensen, G. O. *J. Mol. Spectrosc.* **1974**, *53*, 62–76.

(57) Gao, J.; Jorgensen, W. L. Personal communication.

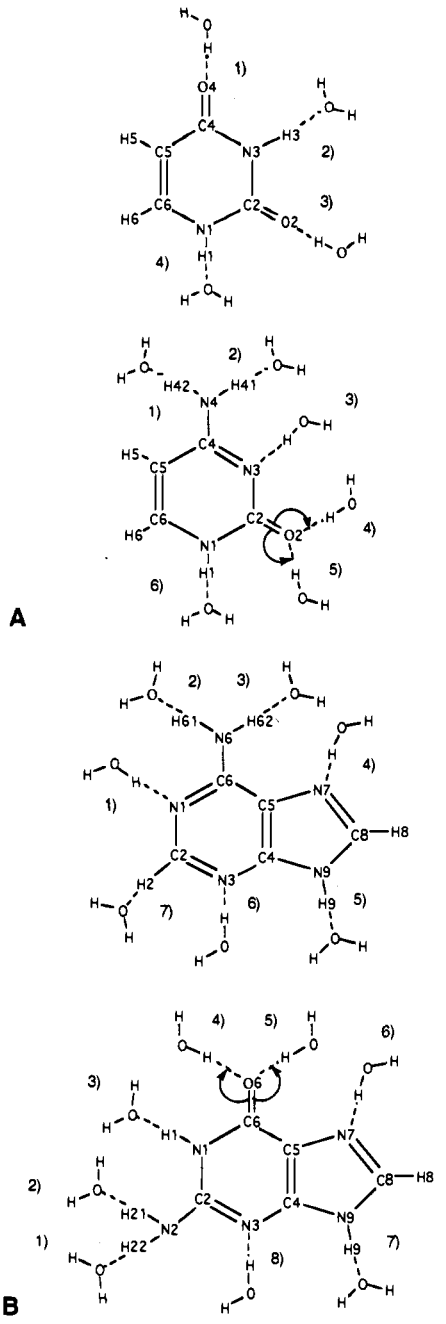

Figure 1. Water–nucleic acid base interaction geometries for the (A) pyrimidines and (B) purines. Atom numbering in the figure is that referred to in the text.

comparison of base pairing energies with experiment and in the crystal calculations. Initially the internal geometries were optimized. The equilibrium values of bond lengths and bond angles (b_0 and θ_0 in eq 1) in uracil, thymine, cytosine, adenine, and guanine were first chosen to be equal to the mean values obtained from CCDB survey results for the bases. The values for cytosine, uracil, adenine, and guanine were those given in ref 31, while those for thymine were determined here. The atom numbering is given in Figure 1. The θ_0 values were constrained in each system so that the sum of intraring angles was equal to 720° in six-membered rings and to 540° in five-membered rings, and to 360° for the sum of apex angles. This resulted in molecules that remained planar during energy minimization. The bond and angle parameters were optimized to obtain agreement with the CCDB results following adjustment of the interaction parameters (see below). This was necessary to compensate for the influence of the external terms in the force field on the minimized structures.

Data used for the optimization of the equilibrium bond lengths and angles along with the CHARMM optimized structures are presented as Tables 1–5 of the supporting information. In addition to the bond lengths and angles from the CCDB surveys, hydrogen distances and angles were obtained from 3-21G optimized base geometries.[58] The rms differences in the bond lengths and angles between the CHARMM and survey values are also included in the tables. Due to the relatively low level of the *ab initio* calculations used in geometry optimization, the values were used only as a guideline for the optimization of the hydrogen positions. Rigorous optimizations of the angles was not feasible because the non-hydrogen values were from the survey and the hydrogen values were from the *ab initio* study. Thus, the sums of the apex angles did not equal 360°. Angles involving heavy atoms were adjusted to yield agreement with the survey data; the angles involving hydrogens were then chosen to yield agreement with the *ab initio* results. In the CCDB survey results, the standard deviations in the bond lengths varied between 0.005 and 0.016 Å, while the angle values varied between 0.3 and 1.3°. The standard deviations of the CCDB results for different compounds are used as a limit on the accuracy expected from the optimization of the empirical parameters.

In Figure 2 the atom types and charges are presented for the individual bases. The atom types are unique to nucleic acids; they include an extra N to indicate that they are associated with nucleic acid. This allows the nucleic acid parameters to be readily combined with the protein parameters, for which standard labels are used. Several of the atom types are common to the two purines and the three pyrimidines. To improve both the empirical geometries and vibrational spectra (see below), additional atom types were introduced where necessary. In total, 8 nitrogen, 11 carbon, 7 oxygen, and 7 hydrogen atom types were used (see Appendix). For example, initially, there were 3 nitrogen ring atom types (NN2, NN3, NN4), in addition to the NH_2 nitrogen NN1; NN2 was the protonated ring nitrogen, NN3 was the 6-membered unprotonated nitrogen, and NN4 was the 5-membered-ring unprotonated nitrogen. However, the presence of corresponding bonds, including the NN2–CN1 bond in the purines and guanine, led to the addition of the NN2G atom type at N1 in guanine and the NN2U atom type for N1 in uracil and thymine. Similarly, the NN3–CN2 bond in cytosine and adenine required the addition of the NN3A atom type for positions N1 and N3 in adenine. Of the exocyclic atoms, the only additional atom type required was ON1C for the O2 oxygen in cytosine, due to a significant difference with the ON1–CN1–NN2 angle in uracil and thymine. The additional atom types allowed for the further refinement of the parameters to better fit the CCDB survey data. The different carbon atom types in the 5- versus 6-membered rings appeared to allow enough flexibility for the parametrization of both ring types. Thus, for example, atom type CN4 occurs in both the 5- and 6-membered rings of adenine. In this case, the adjacent atom types allow for the associated bond and angle parameters to be adjusted to yield good agreement with the CCDB survey data.

The rms differences with the CCDB results are 0.008 Å or less for the bond lengths and 0.5° or less for the bond angles; the ranges are from 0.0 to 0.013 Å for the bond length differences and from 0.0 to 1.1° for the angles. Thus, the empirical values fall within the criteria for agreement suggested by the standard deviations in the CCDB survey values. The largest discrepancies in the bond lengths occur for the endocyclic

(58) Aida, M. *J. Comp. Chem.* **1988**, *9*, 362–368.

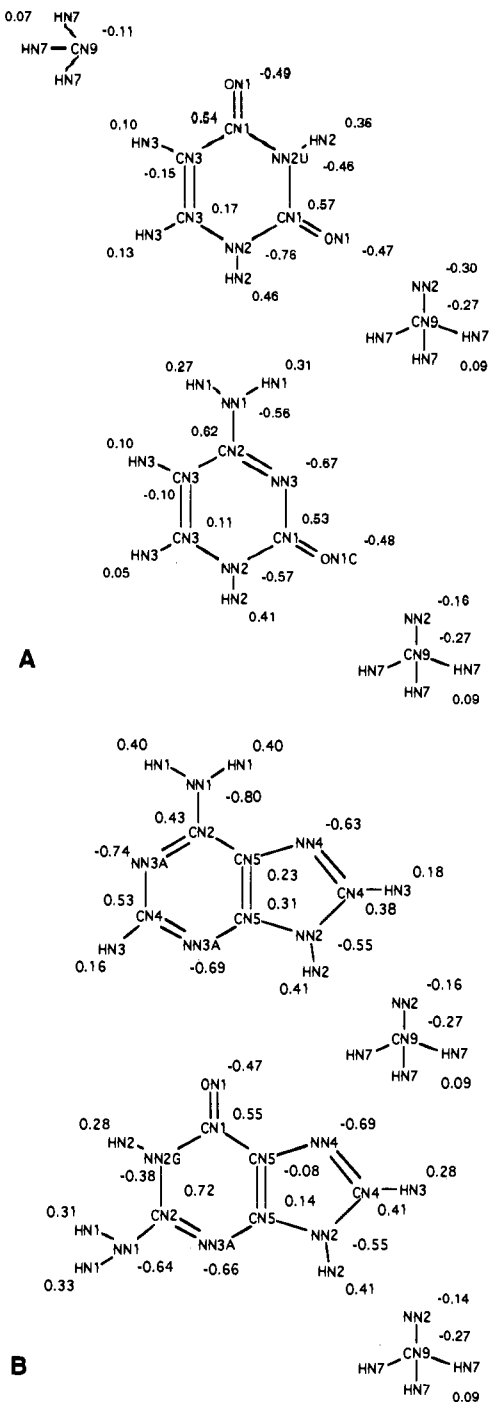

Figure 2. Diagrams of the (A) pyrimidine and (B) purine nucleic acid bases, including the atom types and partial atomic charges. The methyl group on the upper-left corner of (A) contains the charges used in the 5-methyl group of thymine. The fragments on the right-hand side of the figure are the charges used on the (A) 1-methylated pyrimidines, including the adjusted N1 charge, and on the (B) 9-methylated purines, including the adjusted N9 charge.

bonds of thymine and guanine. This is due to the parametrization being performed initially on uracil and adenine with the parameters then being transferred to thymine and guanine, respectively. The resulting differences, however, are still within the stated criteria so that no additional atom types were introduced. For the valence angles the largest differences occur for the exocyclic substituents, i.e., the C4−C5−C5m and C6−C5−C5m angles of the C5 methyl group in thymine, the N6 amine in adenine, and the O6 in guanine. The differences are again within the criteria derived from the CCDB results so that no additional atom types were introduced. In the base parametrization only a single Urey−Bradley term was used in the case of the N2−C2−N1 angle in guanine. This parameter allowed for the optimization of the geometry and gave better agreement for the vibrational frequency of the βC_2N bending mode of the N2 nitrogen with an empirical frequency of 305 cm^{-1} (see Table 10 of the supporting information).

The force constants K_b, K_θ, K_{U-B}, K_φ, and K_{imp} for the bond, angle, Urey−Bradley, dihedral, and improper dihedral deformation terms in eq 1 were determined by fitting the vibrational spectra of the five bases. Low-temperature rare-gas matrix spectra were used for the experimental frequencies. Such matrix isolation spectroscopy yields relatively sharp and well-defined bands for monomeric molecules with minimal influence from intermolecular interactions and, in the absence of gas-phase measurements, provides the best data for the internal parametrization of an empirical energy function. In the very low frequency region (below 400 cm^{-1}) for which the least data are available, mainly *ab initio* results had to be used for the parametrization. These vibrations make the dominant contribution to the larger scale motions of the bases, although the rigidity is such that only relatively small displacements are possible. Also, in assigning normal modes, *ab initio* results were used extensively. The *ab initio* calculations are very important for the assignment of molecules of high complexity, such as the nucleic acid bases, for which experimental assignments are not unique unless many isotopic species have been studied; only limited isotopic data are available. However, the published *ab initio* results are not sufficiently accurate to permit unique normal mode assignment in crowded regions of the vibrational spectrum. This is particularly true below 1000 cm^{-1}. For this reason possible assignments for the bases are not unequivocal for many of the bands. Moreover, the CHARMM energy function (eq 1) is not expected to reproduce the vibrational spectra of all the nucleic acid bases with the accuracy of the force fields used for individual molecules in vibrational spectroscopy.[59] The limitations are two-fold: first, one seeks to minimize the number of atom types in the parametrization; second, cross-terms between internal coordinates used in most vibrational force fields are not included in the CHARMM energy function. Both of these limitations are introduced to obtain a simple and widely applicable potential energy function. We require that the calculated frequencies differ from the reference values by no more than 10%. Detailed comparison of the CHARMM calculated frequencies and potential energy distributions with experimental and *ab initio* data for the five bases is included in Tables 6−10 of the supporting information. The following discussion will focus on selected aspects of the optimization process and the quality of agreement of the empirical and goal data.

Optimization of the force constants for uracil was based on experimental data from two low-temperature matrices[60] and in the solid phase,[61] as well as on HF/4-21G *ab initio* calculations.[62] Uracil has the highest symmetry among nucleic acid bases and its vibrational modes are thus the easiest to assign. For the modes above 1000 cm^{-1}, the agreement with experimental and *ab initio* results is very good for the ring and exocyclic deformations and the various stretching modes. In the region below 1000 cm^{-1}, the agreement is reasonable for both the in-

(59) Derreamaux, P.; Vergoten, G.; Lagant, P. *J. Comp. Chem.* **1990**, *11*, 560−568.
(60) Barnes, A. J.; Stuckey, M. A.; LeGall, L. *Spectrochim. Acta* **1984**, *40A*, 419−431.
(61) Harsanyi, L.; Csaszar, P.; Csaszar, A.; Boggs, J. E. *Int. J. Quantum Chem.* **1986**, *29*, 799−815.
(62) Szczesniak, M.; Nowak, M. J.; Rostkowska, H.; Szczepaniak, K.; Person, W. B.; Shugar, D. *J. Am. Chem. Soc.* **1983**, *105*, 5969−5976.

plane and out-of-plane modes. Examples include the ring stretching modes at 1447, 1523, and 1629 cm^{-1}, the ring deformations at 522, 593, and 615 cm^{-1}, and the carbonyl bend at 378 and 615 cm^{-1}. The NH wags at 563 and 683 cm^{-1} are well represented in terms of both frequencies and assignments. Somewhat larger differences occur in the C−H deformations at 1079 and 1265 cm^{-1}. The C_2O and C_5H wags at 853 and 1048 cm^{-1}, respectively, are calculated to be higher than the C_4O and C_6H wags at 803 and 683 cm^{-1} contrary to the experimental and *ab initio* data; however, those assignments are not unambiguous. The low-frequency out-of-plane ring torsional modes at 173, 203, and 410 cm^{-1} nicely reproduce the scaled *ab initio* values.

Spectra obtained in low-temperature matrices[63,64] and in the crystalline solid[65] were used as the goal data for the optimization of thymine. No *ab initio* normal mode calculations are available for thymine; the experimental assignments are thus approximate, based on comparison with the spectrum of uracil and studies of the influence of selective deuteration on the absorption band frequencies. In the region above 1000 cm^{-1} the agreement is again good; the in-plane deformations of N_1H and N_3H in the range of 1246 to 1401 cm^{-1} are somewhat lower than those observed experimentally. In the region below 1000 cm^{-1} the C5 methyl group stretching frequency at 475 and 747 cm^{-1} is somewhat underestimated while the C_2O and C_4O in-plane deformations at 378 and 592 cm^{-1} are well represented. As to the out-of-plane modes, the C_6H wag at 654 and 787 cm^{-1} is somewhat underestimated, while the carbonyl C_2O, C_4O (725 and 890 cm^{-1}) and the N_1H and N_3H wags (557 and 654 cm^{-1}) are fairly well represented. No experimental data are available for the lowest 5 modes involving the C5 methyl and ring torsions. Further optimization of the thymine force constants would require *ab initio* results.

Force constants associated with cytosine were optimized using experimental spectra in Ar and N_2 low-temperature matrices and in KBr pellets[66] and with *ab initio* 3-21G results.[67] As shown in Figure 2, the atom types in cytosine are identical with those in uracil, except for the N3 and C4 positions; thus, most of the force constants are the same. The high frequencies are again well reproduced by the empirical force field. The C_2O stretch at 1711 and 1772 cm^{-1} is in good agreement with both the experimental and *ab initio* results. This fit was aided by the addition of a new atom type, ON1C, for the O2 atom. The C_4N stretch and the N_1H deformation at 1402 and 1394 cm^{-1}, respectively, are switched as compared to experiment, such that the N_1H deformation is too low; however, agreement with the *ab initio* results is good. These discrepancies between experiment and *ab initio* values suggest that matrix effects may be influencing the experimental results. In the lower frequency region of the spectrum between 332 and 948 cm^{-1} the in-plane modes involving amino bending, ring stretches and deformations, and carbonyl bending modes are well reproduced; an exception occurs with the NH_2 rock at 1017 cm^{-1}, which is somewhat underestimated as compared with both the experimental and *ab initio* values. With this mode a compromise had to be made to allow for better agreement of the NH_2 wag and torsional modes (446 and 550 cm^{-1}). The latter are important for maintaining a balance between proper Watson−Crick hydrogen bonding and base fluctuations. If the geometry of the bases does not distort properly as the relative orientation of the individual bases changes, Watson−Crick hydrogen bonds could be broken. For the remaining out-of-plane modes the agreement is satisfactory.

Optimization of the purine associated force constants was initiated using adenine. Data used in the adjustment of the force constants included HF/4-21G *ab initio* results,[68,69] low-temperature matrix spectra,[70] and the spectrum in the crystalline solid.[71] An accurate fit of the vibrational spectra of purine bases is difficult, especially for modes below 1000 cm^{-1}. Purine bases are molecules of high complexity, with a large number of normal modes (39 for adenine, 42 for guanine), which makes the assignments uncertain. Also, as may be seen in Figure 2, atom types in purine bases were in some instances chosen to be identical with those in the pyrimidine bases, limiting the flexibility of the parametrization. Additional atom types unique to purine bases were introduced where necessary, as outlined above. Overall, the high-frequency region of the spectrum is well modeled by the empirical energy function. Some of the ring stretching and deformation modes (1434−1737 cm^{-1}) appear to be somewhat high, while the N_9H bend at 1242 and 1346 cm^{-1} is too low. In the lower region of the spectrum, the NH_2 rock at 1097 and 919 cm^{-1} is somewhat too low with respect to the *ab initio* and polycrystalline solid results. The NH_2 deformation at 1641 cm^{-1} is in good agreement with the *ab initio* and experimental data. The empirical results adequately model the low region of the in-plane modes including the βC_6N deformation (316 cm^{-1}) and the ring stretches and deformations in the region of 488−1057 cm^{-1}. The out-of-plane NH_2 modes at 494, 562, and 584 cm^{-1} are fitted in a satisfactory fashion. The remaining out-of-plane modes are also well reproduced. The "butterfly" motion of the two rings, which is included in the ring out-of-plane vibrations, makes a significant contribution, 69%, to the empirical mode at 269 cm^{-1}.

Optimization of the vibrational frequencies of guanine used the experimental spectrum in an Ar matrix[72] and results of a 3-21G *ab initio* calculation.[73] The interpretation of the spectrum is even more difficult for guanine than for adenine, not only because of the greater complexity of the system, but also due to the presence of more than one tautomer in the Ar matrix.[72] Thus, the assignment of experimental bands to normal modes is not unequivocal. This makes the task of fitting the CHARMM results particularly difficult, although the more detailed understanding of the adenine spectrum was helpful. The CHARMM frequencies for guanine in the high-frequency region compare well with the experimental bands in the Ar spectrum. The C_6O stretch at 1840 cm^{-1} is somewhat overestimated relative to the the empirical value. The NH_2 deformation (1681 cm^{-1}) and high-frequency ring stretching (1330 and 1474 cm^{-1}) are adequately modeled. The region of the spectrum below 1000 cm^{-1} is reproduced in accord with the experimental results, especially for the ring deformation and stretching modes between 459 and 1194 cm^{-1}, also the C_6O and C_2N bending modes at 375 and 305 cm^{-1}, respectively, are well represented.

(63) Graindourze, M.; Smets, J.; Zeegers-Huyskens, Th.; Maes, G. *J. Mol. Struct.* **1990**, *222*, 345−364.
(64) Nowak, M. J. *J. Mol. Struct.* **1989**, *193*, 35−49.
(65) Mathlouthi, M.; Seuvre, A.-M.; Koenig, J. L. *Carbohydr. Res.* **1984**, *134*, 23−38.
(66) Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; KuBulat, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, *110*, 8319−8330.
(67) Radchenko, E. D.; Sheina, G. G.; Smorygo, N. A.; Blagoi, Yu. P. *J. Mol. Struct.* **1984**, *116*, 387−396.
(68) Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 5324−5340.
(69) Nowak, M. J.; Lapinski, L.; Kwiatkowski, J. S.; Leszczynski, J. *Spectrochim. Acta* **1991**, *47A*, 87−103. Leszczynski, J. *Int. J. Quantum Chem. Quantum Biol. Symp.* **1992**, *19*, 43.
(70) Stepanian, S. G.; Sheina, G. G.; Radchenko, E. D.; Blagoi, Yu. P. *J. Mol. Struct.* **1985**, *131*, 333.
(71) Majoube, M. *J. Raman Spectrosc.* **1985**, *16*, 98.
(72) Szczepaniak, K.; Szczesniak, M. *J. Mol. Struct.* **1987**, *156*, 29−42.
(73) Latajka, Z.; Person, W. B.; Morokuma, K. *J. Mol. Struct. (THEOCHEM)* **1986**, *135*, 253−266.

Table 1. Base Pair Interaction Geometries and Energies, Including Experimental Values for the Methylated Bases[a]

Methylated Watson–Crick Base Pairing

pair		interaction energy, kcal/mol cdiel	rdiel	exp		distance, Å cdiel	rdiel	exp
G-C	total	−24.8	−21.7	−21.0[b]	N2–O2	2.84	2.76	2.86[c]
	elec	−23.8	−23.3		N1–N3	2.93	2.87	2.95[c]
	vdw	0.0	2.0		O6–N4	2.85	2.80	2.91[c]
	internal	−1.0	−0.4					
A-T	total	−14.0	−16.1	−13.0[b]	N6–O4	2.87	2.73	2.93[d]
	elec	−13.5	−18.5		N3–N1	2.86	2.79	2.85[d]
	vdw	0.2	2.6					
	internal	−0.7	−0.3					
A-U	total	−14.0	−16.1	−14.5[b]	N6–O4	2.88	2.73	2.93[d]
	elec	−13.6	−18.5		N3–N1	2.86	2.79	2.85[d]
	vdw	0.3	2.7					
	internal	−0.8	−0.3					

Methylated Hoogsteen Base Pairing

pair	interaction energy, kcal/mol cdiel	rdiel		geometry cdiel	rdiel	exp
A-T	−13.3	−14.6	N6–O2	2.89	2.74	2.86[e]
			N7–N3	2.87	2.81	2.93[e]
A-U	−12.5	−13.4	N6–O2	2.91	2.92	
			N7–N3	2.87	2.81	

Base Dimers, Minimum Energy Interactions

pair	interaction energy, kcal/mol cdiel	rdiel	exp
A-A	−10.4	−12.6	
T-T	−11.7	−14.1	−9.0[b]
U-U	−11.8	−13.9	−9.5[b]
C-C	−18.9	−15.4	−16.0[b]

[a] Values were calculated by determining the interaction energy of rigid minimized monomers. [b] See ref 77. [c] See ref 43. [d] See ref 74. [e] See ref 78.

The use of the Urey–Bradley term in the optimization of the C_2N geometry (see above) yielded good agreement for the mode C_2N deformation as compared to experiment. The introduction of a Urey–Bradley term avoided the need for the addition of a new atom type, while allowing for a satisfactory reproduction of the geometry and vibration. The NH_2 rock at 984 cm^{-1} is again somewhat underestimated to allow for better agreement for the torsional and wagging modes in the range of 402 to 557 cm^{-1}. The wagging modes of the C_6O (189 cm^{-1}), C_2N (141 cm^{-1}), N_1H (599 cm^{-1}), and N_9H (900 cm^{-1}) moieties are all well represented by the empirical energy function. The butterfly motion of the 6-membered and 5-membered rings makes a significant contribution to the mode at 285 cm^{-1}, similar to the adenine results.

(b) Interaction Parameters. As stated in the Introduction, the nonbonded parameters were determined so as to reproduce a variety interactions involving the bases. One aspect of the parametrization was to use van der Waals parameters for the carbonyl oxygen and the nitrogen atoms to fit the interaction geometries of the Watson–Crick base pairs observed in crystal structures,[43,74] as shown in Table 1. Results obtained with a constant dielectric of unity (cdiel) and a distance dependent dielectric (rdiel)[75] are included. It can be seen that the good agreement is obtained for the cdiel results and that the rdiel distances are shorter in almost all cases. This suggests that caution is needed in the use of rdiel for these systems, because it appears to lead to an imbalance of the base-pair hydrogen bonding interactions.

The van der Waals parameters were not changed throughout the remainder of the parametrization, i.e., all carbonyl oxygens are identical and all nitrogen atoms are identical. The carbon atom van der Waals were determined based on base stacking geometries and base crystal calculations. The hydrogen atom van der Waals parameters were transferred unchanged from the protein polar and aliphatic hydrogen values. The partial atomic charges were evaluated making use of *ab initio* calculations at the 6-31G* level of the interactions between water molecules and various hydrogen bonding sites on the bases, as shown in Figure 1. Starting with charges obtained from the Pranata *et al.* parametrization of the nucleic acid bases,[76] the point charges were determined to reproduce the scaled (1.16) *ab initio* interaction energies and the minimum energy geometries with distances approximately 0.2 Å shorter than the *ab initio* values. The present methodology ensures that the solute–solvent interactions are balanced with the solvent–solvent interactions of the modified (TIP3P) water model for which these parameters are designed.

In the *ab initio* 6-31G* optimizations of the adenine and guanine interactions with water the base 3-21G optimized structures[58] and the water experimental gas-phase geometry[56] were fixed and only the distances and in some cases a single angle, as shown in Figure 1, were optimized. Interaction energies and geometries determined at the *ab initio* 6-31G* level by the same procedure for uracil and cytosine were previously published.[76] The charges were adjusted to yield water–base interaction energies equal to the *ab initio* results scaled by a factor of 1.16. Simultaneous consideration was given to fitting the experimental base pairing energies,[77] the dipole moments of the bases calculated at the HF/6-31G* level, and the heats of sublimation of uracil, 1-methylthymine, and 9-methyladenine.[77] For the empirical calculations the 3-21G optimized base geometries were employed initially to determine the water to base interactions. The resulting charges were then used in the optimization of the internal force field, as stated above, followed by readjustment of the charges using the new empirical geometry, and so on until convergence was achieved in an iterative fashion. Electrostatic groups of unit charge used in the bases ranged from 2 to 7 atoms. These groups were selected to be as small as possible without compromising the quality of the charge distribution.

The *ab initio* and empirical results for the base–water interactions are listed in Table 2. Scatter plots of the *ab initio* (scaled by 1.16) versus empirical interaction energies and of the *ab initio* offset distances (reduced by 0.2 Å) versus the empirical minimum distances are shown in Figure 3, parts A and B, respectively. The correlation coefficient from a least-squares fit of the interaction energies is 0.96 while that for the distances is 0.86. The agreement between the *ab initio* and scaled empirical interaction energies is satisfactory.

The largest discrepancies in the present results between the scaled *ab initio* and empirical interaction energies occur in the purines adjacent to the ring fusion sites (interaction 5 for guanine and interation 4 for adenine; see Figure 1 and Table 2). These interactions occur in the region of base atoms 6 and 7 (Figure 1B). The simultaneous interactions of the water molecule with the position 6 constituent and the nitrogen at position 7 limit the ability of the empirical potential to reproduce the *ab initio*

(74) Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Kim, J. J. P.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 109–144.
(75) Gelin, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2002–2006.
(76) Pranata, J.; Wierschke, S. G.; Jorgensen W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
(77) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149–1170.

Table 2. Comparison of *ab Initio* and Empirical Base to Water Interaction Energies and Geometries[a]

	1.16 × 6-31G*			empirical		
interaction	IE	R_{min}	angle	IE (L-J)	R_{min}	angle
(1) UraO4−HW[b]	−5.30	2.07		−5.35 (0.78)	1.79	
(2) UraH3−OW[b]	−6.81	1.96		−6.97 (1.20)	1.82	
(3) UraO2−HW[b]	−5.27	2.08		−5.31 (0.77)	1.79	
(4) UraH1−OW[b]	−8.18	1.98		−8.32 (1.47)	1.80	
(1) CytH42−OW[b]	−6.09	2.13		−6.03 (0.38)	1.94	
(2) CytH41−OW[b]	−5.72	2.03		−5.45 (0.80)	1.88	
(3) CytN3−HW[b]	−9.71	2.06		−9.71 (0.88)	1.88	
(4) CytO2−HW[b]	−8.69	2.03	113	−8.34 (0.61)	1.79	109
(5) CytO2−HW[b]	−10.06	1.91	114	−10.35 (1.96)	1.73	105
(6) CytH1−OW[b]	−6.52	2.01		−7.00 (1.23)	1.82	
(1) AdeN1−HW	−6.99	2.10		−6.86 (0.89)	1.89	
(2) AdeH61−OW	−4.66	2.08		−4.69 (0.82)	1.88	
(3) AdeH62−OW	−5.30	2.00		−5.39 (0.99)	1.85	
(4) AdeN7−HW	−7.09	2.08		−6.31 (0.67)	1.92	
(5) AdeH9−OW	−7.25	2.01		−7.37 (1.16)	1.83	
(6) AdeN3−HW	−7.09	2.12		−6.65 (0.75)	1.91	
(7) AdeH2−OW	−1.51	2.49		−1.88 (0.28)	2.39	
(1) GuaH22−OW	−6.11	2.03		−5.97 (0.88)	1.87	
(2) GuaH21−OW	−8.12	2.08		−8.22 (0.65)	1.90	
(3) GuaH1−OW	−7.17	2.04		−7.47 (0.53)	1.88	
(4) GuaO6−HW	−9.98	1.92	113	−9.75 (1.40)	1.76	104
(5) GuaO6−HW	−6.26	2.08	137	−4.83 (−0.12)	1.89	131
(6) GuaN7−HW	−5.14	2.22		−5.41 (0.48)	1.94	
(7) GuaH9−OW	−6.83	2.03		−7.00 (1.17)	1.83	
(8) GuaN3−HW	−4.54	2.15		−4.54 (0.32)	1.94	

[a] See Figure 1 for geometries; energies in kcal/mol, distances in Å, and angles in deg. For the empirical results the term in parentheses next to the interaction energy (IE) is the Lennard-Jones contribution. [b] *Ab initio* Results from Pranata *et al.*[76]

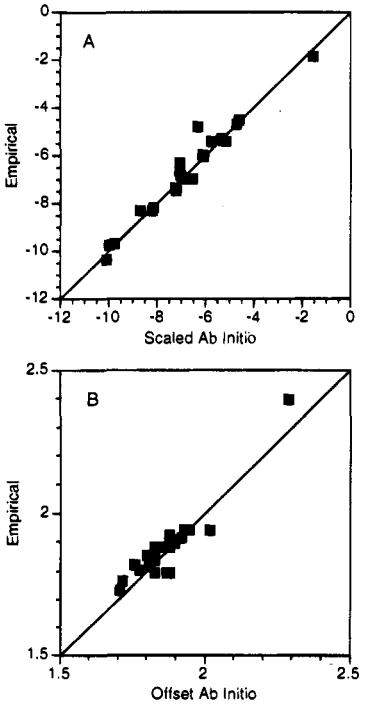

Figure 3. Scatter plots comparing the (A) scaled 6-31G* ab initio and empirical interaction energies and (B) the 6-31G* ab initio and empirical minimum interaction distances. The *ab initio* interaction energies have been scaled by 1.16 and the minimum distances offset by 0.2 Å.

results. Increasing the charge on N7 did improve the interaction energy of the adenine 4 site. However, such a change increased the dipole moment of the molecule to 3.22 D versus the experimental estimate of 3.0 D and led to an overestimation of the heat of sublimation of 9-methyladenine (see below). The poorest agreement concerning the distances (uracil interactions 1 and 3 and guanine interaction 7) occurred for instances where the empirical model gave good agreement with the interaction energies. The angles, where they were optimized, are in good agreement with the *ab initio* values.

Calculated interaction energies and experimental enthalpies of the Watson−Crick base pairs, the Hoogsten base pairs, and several dimers for methylated bases are shown in Table 1. The available experimental interaction enthalpies from gas-phase field ionization mass spectrometry[77] are presented. In the experimental study of the C−C dimer it was observed that the measured interaction enthalpy became more favorable as the ionization voltage decreased; in other cases only a single voltage was used. This suggests that the measured values may be smaller than the actual values. As the ionization potential decreases, the interaction enthalpy becomes more favorable; if extrapolation to zero ionization voltage were feasible, a better estimate of the true interaction energies could be obtained. A 4-31G *ab initio* study of the G-C and A-T base pairing energies gave approximately −30 and −15 kcal/mol, respectively.[58] Comparison of these values with the experimental interaction energies shows the overestimation of the interaction energy by the *ab initio* calculations for G-C to be significantly larger than for A-T and supports the values obtained in the present study.

Interaction energies and experimental enthalpies for the A-T and A-U Hoogsteen base pairing and several base dimers are also given in Table 1. The Hoogsteen base pairing interaction energies are slightly less favorable than the Watson−Crick values. Comparison of the A-T interaction distances shows satisfactory agreement with experimental crystal values;[78] the values are 2.93 Å (N7−N3) and 2.86 Å (N6−O2). Analysis of the base dimer interaction energies shows the calculated interaction energies to be too favorable as compared to experiment; again, this may be associated with effects of ionization field strength on the measured interaction energy, as discussed above. The ordering of the interaction energies (C-C > U-U > T-T) is in agreement with that obtained from experiment.

Interaction enthalpies are reported in Table 3 for the methylated bases in the gas phase. They include the change in internal energy at the minimum, the vibrational energies, and the rotational plus translational ideal gas corrections. The interaction energies in Table 3 are more favorable than those reported in Table 1 due to the inclusion of internal strain upon base pairing in the latter values. Inspection shows both the zero-point energies and rotational/translational/ideal gas corrections to be significant, so that the interaction enthalpies are smaller in magnitude than the interaction energies. The agreement is excellent for the G-C base pair, while the A-T and A-U calculated values are smaller than the experimental values.

Table 11 of the supporting information presents the Watson−Crick base pair interaction energies from the present parametrization and compares them with the values reported for AMBER,[5] for OPLS,[76] and from restrained electrostatic potential fitting procedures (RESP[79]) for a constant dielectric, as well as with the experimental interaction enthalpies.[77] The CHARMM 22 values are larger than those from both AMBER and OPLS, but smaller than the RESP value for G-C. If the vibrational, rotational, translational, and ideal gas corrections are not taken into account, CHARMM 22 gives the best agreement with experiment for the A-T base pair while AMBER has the best agreement of the G-C base pair. However, when those

(78) Frey, M. N.; Koetzle, T. F.; Lehmann, M. S.; Hamilton, W. C. *J. Chem. Phys.* **1973**, *59*, 915−924.
(79) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

Table 3. Base Pairing Interaction Energies, Zero Point Energies, and Heats of Interaction for the Watson–Crick, Hoogsteen, and Homo Dimer Interactions Using a Constant Dielectric for the Methylated Bases[a]

interaction energy		vibrational energy				ΔH interaction	
rigid	relaxed				ΔE_{vib}	calcd	exp[b,c]
Watson–Crick	AT	A	T	A-T			
−14.0	−14.38	84.89	85.71	172.03	1.43	−10.58	13.0
	A-U	A	U	A-U			
−14.0	−14.38	84.89	68.98	155.31	1.44	−10.57	14.5
	G-C	G	C	G-C			
−24.8	−25.33	88.70	76.03	166.47	1.75	−21.22	21.0
Hoogsteen	A-T	A	T	A-T			
−13.3	−13.66	84.89	85.71	171.84	1.24	−10.06	
	A-U	A	U	A-U			
−12.5	−13.41	84.89	68.98	155.58	1.71	−9.34	
homodimers	A-A	A	T	A-A			
	−10.84	84.89	84.89	171.23	1.46	−7.02	
	−9.49	84.89	84.89	171.12	1.35	−5.78	
	−10.86	84.89	84.89	171.24	1.47	−7.03	
	T-T	T	T	T-T			
	−11.88	85.71	85.71	172.58	1.16	−8.36	9.0
	−11.77	85.71	85.71	172.56	1.14	−8.28	
	−11.76	85.71	85.71	172.76	1.34	−8.06	
	U-U	U	U	U-U			
	−11.63	68.98	68.98	139.13	1.17	−8.10	
	−11.84	68.98	68.98	139.56	1.60	−7.88	
	−11.91	68.98	68.98	139.11	1.15	−8.40	9.5
	C-C	C	C	C-C			
	−19.30	76.03	76.03	153.35	1.29	−15.65	16.0

[a] Energies in kcal/mol. Zero point vibrational energies were calculated at 300 K. The $4RT$ correction includes the rotational ($3/2RT$), translational ($3/2RT$), and ideal gas (PV) contributions. $\Delta H_{interaction}$ calculated equals the sum of the interaction energy, the zero point energy of the dimer minus the sum of the monomer zero point energies, and the $4RT$ correction for the rotational, translational, and ideal gas terms. [b] The experimental data are from ref 77. [c] It has been suggested that the A-T experimental interaction enthalpy is associated with the Hoogsteen base pairing (see ref 80).

corrections are taken into account (see above) and assuming the vibrational correction to be similar for AMBER, OPLS, and RESP, the increase in the interaction energy of approximately 4 kcal/mol leads to AMBER and OPLS being less favorable than experiment for both the A-T and G-C base pairs; the RESP G-C base pairing interaction enthalphy is in good agreement with experiment. The lowering of the interaction energy is most significant in the case of the A-T base pair. As discussed above, the corrections in the case of CHARMM22 lead to good agreement for the G-C interaction while the A-T interactions are smaller in magnitude than experiment. It should be noted that the experimental data for the A-T base pair may correspond to the Hoogsteen, rather than the Watson–Crick, geometry[80] (see below).

Following completion of the present parameter optimization, new *ab initio* results on the interaction strengths of the methylated guanine–cytosine and adenine–thymine base pairs were reported.[80] Full geometry optimization was performed on the bases as well as the base pairs at the HF/3-21G and HF/6-31G* levels of theory, electron correlation was treated via MP2 calculations on the Hartree–Fock optimized geometries, basis set superposition error was taken into account, and ΔH_{298} values were estimated; it is not clear that the ΔV term was included. Although these calculations are at a considerably higher level than previously reported *ab initio* calculations, it is not clear that they are accurate enough to be used for molecular mechanics energy function optimizations, particularly when one considers the water dimer results at the same *ab initio* level. It is of interest that the best G-C and A-T values are considerably larger than experiment. As to the A-T value, the authors suggest from their calculations that the experimental measurements are in fact for the Hoogsteen rather than the Watson–Crick dimer. This would be a very important result, and if true, it would yield *ab initio* values in very good agreement with experiment. However, this would raise the question why the G-C value disagrees significantly with experiment, while the A-T value does not. Our parametrizations predict that the Hoogsteen A-T interaction is slightly less favorable than the Watson–Crick interaction.

Figure 2 shows the partial atomic charges for all the bases and their methylated analogs. Charges for the methyl groups are equivalent to those for the all-hydrogen aliphatic groups[34] except in the case of thymine. They are 0.09 for H and −0.27 for C. For thymine, the use of 0.09 on the hydrogens led to a nonplanar minimized structure. There appears to be repulsion between the H5 of the methyl group and the H6 atom. This leads to a non-symmetric methyl group and causes the C5 methyl atom to be 0.3° out-of-plane; the ring deviates 0.1° or less from planarity. The introduction of a smaller (0.07) charge on the methyl hydrogens alleviated this problem. Table 12 of the supporting information lists the calculated base dipoles along with the experimental values[81] and the theoretical results at the 6-31G* level from the published 3-21G structures.[58] The dipole moments of adenine and uracil are in satisfactory agreement with the experimental estimates. The dipole moments of guanine, uracil, and thymine are somewhat underestimated and those of cytosine and adenine are somewhat overestimated as compared with the *ab initio* values.

Table 13 of the supporting information includes the CHARMM 22, AMBER,[35] OPLS,[76] X-ray,[82] and HF/6-31G* partial atomic charges and dipole moments for thymine, cytosine, adenine, and guanine. Generally, the signs and magnitudes of the charges are similar. The magnitudes of the charges used in AMBER and from the HF/6-31G* Mulliken population analysis tend to be larger than the others. Examples include the C2 atom of thymine and the C4 atom of cytosine. The AMBER charges were derived from a fit to an *ab initio* electrostatic potential from minimal basis set (STO-3G) calculations. The CHARMM 22 charges are more similar to the OPLS and X-ray charges. Both the CHARMM 22 and OPLS values are derived so as to reproduce HF/6-31G* interactions of the bases or base analogs with water molecules. Comparison of the dipole moments show the CHARMM 22 values to exceed the HF/6-31G* values for cytosine and adenine and give better agreement than AMBER and OPLS for thymine and guanine. Overestimation of the dipoles is required to compensate for the lack of polarizability in the potential energy function.

Although the approaches to determining charges in OPLS and CHARMM are similar, there are some differences. To simplify the *ab initio* base–water interaction calculations, Pranata *et al.*[76] used the model compounds 4-aminopyrimidine and 2-aminopyrimidin-4-one to represent the 6-membered rings of adenine and guanine, while imidazole was used to model the 5-membered ring. This led to some significant differences in the interaction energies and the resulting charges. Examples include the use of imidazole for the N7 site in the purines; the in-plane *ab initio* water interaction energy is −6.26 kcal/mol in imidazole as compared to −4.43 and −6.11 kcal/mol in adenine and guanine, respectively. Other examples include the

(80) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493–2499.

(81) DeVoe, H.; Tinoco, I., Jr. *J. Mol. Biol.* **1962**, *4*, 500. Brown, R. B.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1988**, *110*, 2329–2330.

(82) Pearlman, D. A.; Kim, S.-H. *J. Mol. Biol.* **1990**, *211*, 171–187.

Table 4. Base Stacking Interactions in the B-DNA Dickerson Dodecamer[a,b]

		interaction energy				
		CHARMM22			X-ray charges[c]	
	bases	total	VDW	elec	total	elec
(1)	(5′)C-G(3′)					
	(3′)G-C(5′)	−18.6	−13.1	−5.5	−28.3	−15.2
(2)	G-C					
	C-G	−27.8	−15.8	−12.0	−41.2	−25.4
(3)	C-G					
	G-C	−22.2	−13.5	−8.7	−29.5	−16.0
(4)	G-C					
	A-T	−16.0	−17.2	1.2	−8.3	8.9
(5)	A-T					
	A-T	−1.6	−16.2	14.6	17.3	33.0
(6)	A-T					
	T-A	−11.8	−18.2	6.5	−20.2	−2.6
(7)	T-A					
	T-A	−2.3	−16.7	14.4	18.5	34.9
(8)	T-A					
	C-G	−16.6	−16.9	0.3	−9.0	7.8
(9)	C-G					
	G-C	−20.8	−13.4	−7.5	−27.4	−14.0
(10)	G-C					
	C-G	−24.4	−13.8	−10.6	−34.1	−20.4
(11)	C-G					
	G-C	−14.1	−11.1	−3.0	−20.7	−9.6

[a] Reference 83 for the geometry. [b] Interaction energies in kcal/mol are the sum of the intrastrand stacking energies calculated for the base atoms for the two strands. [c] Charges from ref 82 and VDW energies are identical to the CHARMM22 values.

interactions trans to the N1 atom with the position 6 carbonyl oxygen in 2-aminopyrimidin-4-one and the amino group in 4-aminopyrimidine which yield *ab initio* interaction energies of −6.74 and −4.98 kcal/mol, respectively. These may be compared to values of −5.40 and −4.56 kcal/mol for the comparible sites in guanine and adenine, respectively. The use of the model compounds also eliminates the sites adjacent to the fusion of the 6-membered and 5-membered rings in the purines, which introduced some difficulties in fitting the water interaction energies in the present study. Pranata *et al.*[76] adjusted their charges based on the TIP4P water model rather than TIP3P. Additionally, the OPLS charges are chosen to give direct agreement with the 6-31G* *ab initio* results, rather than the 1.16 scaled interaction energies used in CHARMM. Use of the 6-31G* *ab initio* interaction energies themselves may lead to an underestimation of the solute−solvent nonbonded interactions as compared to the solvent−solvent interactions (see above). Based on the TIP4P[32] to 6-31G* [33] water dimer interaction energy ratio of 1.10, a scale factor of 1.10 would be appropriate. In spite of these differences, the parameters of Pranata *et al.*[76] have been used successfully for studying the nucleic acid bases, particularly in nonaqueous solvents. This may indiciate that some simulation results are relatively insensitive to the details of the charge distribution.

The stacking interaction energies and geometries of the various base combinations were examined. As a first case, the interaction energies of the bases in the CGCGAATTCGCG B-DNA dodecamer[83] were calculated. The stacking energy was obtained as the sum of the intrastrand 5′→3′ and 3′→5′ interaction energies in the crystal geometry (Table 4); i.e., the sum was used because the two sets of values are very similar. The stacking energies vary between −27.8 and −1.6 kcal/mol. Separation into the van der Waals and electrostatic components shows the van der Waals contribution to be attractive and similar for all the stacked pairs, while the electrostatic component is responsible for the large differences in the total interaction energies. *Ab initio* studies of individual base pairs with ideal A or B DNA geometries show that the stacking energy is strongly dependent on the base pair.[84,85] To check the present stacking energy variation the experimentally determined base charges of Pearlman and Kim[82] were used with the CHARMM van der Waals parameters to calculate the dodecamer stacking energies. The results in Table 4 show that the variation in interaction energy with the X-ray charges is even larger than that of the CHARMM values.

Further analysis of the stacking interactions was performed by optimizing the distance between bases stacked in the "ideal" B-DNA configuration[86] with the 3-21G optimized internal geometries of the bases; the bases were used by themselves without the sugar or phosphate moieties. In these calculations only the distances between the bases were varied while the planes of the bases were kept parallel and the angle between the stacked bases was maintained in the ideal B-DNA configuration [Θ = 36°]. Table 14 of the supporting information shows the minimum distances for all possible pairs to be clustered around 3.40 Å, in agreement with the standard B-DNA geometry.[86] However, several stacked pairs show no minima out to 8 Å; this is responsible for the low interaction energies of sets 5 and 7 in Table 4. The variation in the energies is due to both the van der Waals and electrostatic term, with the latter repulsive in almost all cases (see Table 4). Verification of the calculated interaction energies is difficult. A number of experimental[77,87−89] and theoretical[84,85,90,91] studies indicate stacking enthalpies in the range of 5 to 15 kcal/mol and support the ordering of the stacking interaction energies in the present study. The calculated minimum distances and rms fluctuations between bases in crystals from the B-DNA dynamics simulations are shown in Table 15 of the supporting information (see also Section Vb); the results are generally in good agreement with experiment. In the molecular orbital calculations,[84,85] the primary component responsible for the difference in stacking energies is electrostatic. This is in accord with the present results. The variation in the stacking energies calculated for nucleic acid bases suggests that these interactions, in addition to the base pairing hydrogen bonds, may have a significant effect on the dependence of nucleic acid structure and dynamics on the primary sequence. Also, the unfavorable electrostatic contribution to base stacking shows that B-DNA does not have a geometry that optimizes the base stacking interactions.

(c) Crystal Minimizations and Simulations of Nucleic Acid Bases. The parameters for the nucleic acid bases were tested by using them for crystal minimizations and simulations. The systems studied were 1-methylthymine,[92] 9-methyladenine,[93] uracil,[94] the 9-methyladenine, 1-methylthymine base pair,[78] and the 1-methylcytosine, 9-ethylguanine base pair.[95] All crystal calculations were performed with the asymmetric unit as the primary atoms. Thus, the primary atoms included a single

(83) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, S.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179.
(84) Aida, M.; Nagata, C. *Int. J. Quantum Chem.* **1986**, *29*, 1253−1261.
(85) Aida, M. *J. Theor. Biol.* **1988**, *130*, 327−335.
(86) Arnott, S.; Hukins, D. W. L. *J. Mol. Biol.* **1973**, *81*, 93−105.
(87) Suurkuusk, J.; Alvarez, J.; Freire, E.; Biltonen, R. *Biopolymers* **1977**, *16*, 2641−2652.
(88) Marky, L. A.; Breslauer, K. J. *Biopolymers* **1982**, *21*, 2185−2194.
(89) Petersheim, M.; Turner, D. H. *Biochemistry* **1983**, *22*, 256−263.
(90) Ornstein, R. L.; Rein, R.; Breen, D. L.; Macelroy, R. D. *Biopolymers* **1978**, *17*, 2341.
(91) Haram, T. E.; Berkovich-Yellin, Z.; Shakked, Z. *J. Biomol. Struct. Dyn.* **1984**, *2*, 397.
(92) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 28−38.
(93) Stewart, R. F.; Jensen, L. H. *J. Chem. Phys.* **1964**, *40*, 2071−2075.
(94) Stewart, R. F.; Jensen, L. H. *Acta Crystallogr.* **1967**, *23*, 1102−1105.
(95) O'Brien, E. J. *Acta Crystallogr.* **1967**, *23*, 92−106.

monomer or a base pair. No water molecules are present in any of the crystals and all of the molecules are neutral. The crystal simulations were used to determine minimized and average dynamics structures and to determine heats of sublimation; experimental values are available for the monomers.[77] Details on the crystal calculations provide a very stringent test of the parameters because much more experimental data are available for comparison than from solution studies. There is information about specific interactions, as well as more global properties, such as unit cell parameters and pressures.

Figure 4 presents stereodiagrams of the five crystals. The figures show the asymmetric units (the primary atoms) in bold for uracil, 1-methylthymine, 9-methylguanine, 1-methylthymine−9-methyladenine, and 1-methylcytosine−9-ethylguanine crystals. In addition the figures include all unique nearest neighbor atoms in the crystal within 4 Å of the primary atoms, nonbonded interactions used in the analysis of the crystals (dotted lines connecting atoms), and the edges of the unit cells (dotted lines marked with crosses) including the origin, O, and the a, b, and c edges. In the minimizations and simulations all unique image atoms in the crystal which were within the specified cutoff distance of any of the primary atoms were generated using the CRYSTAL module in CHARMM.

Prior to examining the various crystals, it was necessary to determine cutoff distances that adequately represented the polar interactions (see Section II). Table 16 of the supporting information shows the lattice parameters, total energies, and lattice energies obtained after minimization as a function of cutoff distance. The convergence of the lattice parameters is not monotonic as the cutoff distance increased. However, the fluctuations are small at the longer cutoffs and the 22−21−19 cutoff scheme appears adequate. The total energies tend to converge toward more favorable values at the longer cutoff; this convergence was monotonic except for 9-methyladenine. The lattice energies do not vary monotonically, but the difference for the two longest cutoffs is less than 0.5 kcal/mol. Final rms gradients were less than 10^{-3} (kcal mol^{-1})/Å following the 1000-step minimization with the lattice parameters varied, except with the 1-methylthymine minimization, where the gradient was 0.05 (kcal mol^{-1})/Å.

The number of neighbors included in the simulations depends on the unit cell, as well as the particular axis. In uracil the length of the C axis is 3.6 Å, which corresponds to one molecule in a stacked configuration. For a 22-Å cutoff, 6−7 unit cells and 6−7 molecules are present in the calculation along the C direction. The A axis is 11.9 Å, which corresponds to two molecules lying side by side, such that 2 unit cells and 4 molecules are present with a 22-Å cutoff.

(i) Crystal Minimization Results. Table 17 of the supporting information compares the lattice parameters calculated with the 22−21−19 cutoff and the experimental values. In all cases the unit cell volumes were observed to contract. This is expected since the minimizations corresponds to 0 K structure while the experimental crystal structures were measured at room temperature. Crystal minimizations of imidazole, L-histidine and L-*N*-acetylhistidine have yielded contraction of the unit cell of 11.6, 5.1, and 4.4%, respectively,[96] in accord with the range of contraction calculated in the present study. The largest contraction occurred with 1-methylthymine. In addition to the problems observed with the final rms gradient (see above), significant deviations were found for the lattice parameters; i.e., the differences are on the order of 10 to 30%, instead of 1 to 8% calculated for the other crystals. Analysis of the structure of 1-methylthymine shows the presence of an approximate C2 symmetry axis of the pyrimidine ring corresponding to the line connecting N3 to C6; the asymmetry is due to the nitrogen at N1 and the carbon at C5.[87] In the determination of the crystal structure the conformation shown in Figure 4B was adopted arbitrarily. Crystal minimizations of this structure in the present study led to the problems mentioned above and shown in Table 17 of the supporting information. The minimizations were redone following a 180° rotation about the N3 to C6 axis. This was performed by exchanging the coordinates of the C5M atom with the C1 atom, N1 with C5, C2 with C4, and O2 with O4, followed by generation of the images using the standard procedures. Improved final rms gradients were obtained (see above) and the calculated and experimental lattice parameters were in much better agreement (see Table 17 of the supporting information). In molecular dynamics simulations, the rotated structure yielded significantly worse results than those of the reported crystal orientation. It is possible that 1-methylthymine is polymorphic and/or that the crystals are disordered. Further analysis of this possibility would be of interest; the remainder of the present study includes only the results based on the reported crystal structure.

Contraction of the unit cell in all cases is not symmetric with the percent differences being 7.7% or less, excluding the 1-methylthymine calculation. In uracil the A axis expands while the B and C axes contract. Examination of the crystals in Figure 4 shows certain nonbonded interactions to be aligned approximately parallel to a specific axis of the unit cell. Such alignment makes possible changes in the individual unit cell axis to be discussed in terms of the specific nonbonded interactions. The A axis is approximately parallel to the N3 to O4 interaction direction, the B axis is approximately parallel to the C5 to O2 and O4 to N1 interaction directions, and the C axis involves the stacking interactions (see Figure 4A). The changes in the nonbonded interaction distances for uracil (Table 18 of the supporting information) that take place on minimization are not correlated with the changes observed in the lattice parameters; i.e., the N3 to O4 distance decreases the most, while the A axis expands and the contraction of the B axis is associated with only a slight shortening of the C5 to O2 distance and an increase in the O4 to N1 distance. These interactions correspond to the only hydrogen bonds present in the crystal. The largest change, which involves the C axis, is associated with a change in the stacking such that the angle between the rings in the same plane is altered. Changes in stacking in Table 19 of the supporting information include an increase in the C6 to N3 distance from 3.52 to 3.59 Å and decreases in the N1 to C2 and C5 to C4 distances from 3.33 and 3.33 Å to 3.19 and 3.29 Å, respectively. As the stacking changes, the angle of the ring planes with respect to the A−B plane changes. This change leads to the alterations in the A and B axes. For 1-methylthymine (see Figure 4b), the C6 to O2 and N3 to O4 interactions, listed in Table 18 of the supporting information, are aligned with the B axis. These interaction distances are well maintained during the minimization and lead to the good agreement with experiment along the B axis when the molecule has the crystal orientation. The large increase in the C5M to C5M distance is due to a tilting of the molecule about the N3 to C6 axis. This leads to the large contraction of the A axis and the expansion of the C axis. The tilting occurs such that the spatial relation of the C6 to O2 and N3 to O4 interactions is maintained. In the 9-methyladenine crystal (see Figure 4C), minimization led to a contraction of the A and C axes, while the B axis expanded. Although none of the short-range interactions are oriented directly along a given axis, the N6 to N1, N7 to N6, C8 to C9M, and C9M to N3 interactions all have a large component

(96) MacKerell, A. D., Jr.; Karplus, M. Manuscript in preparation.

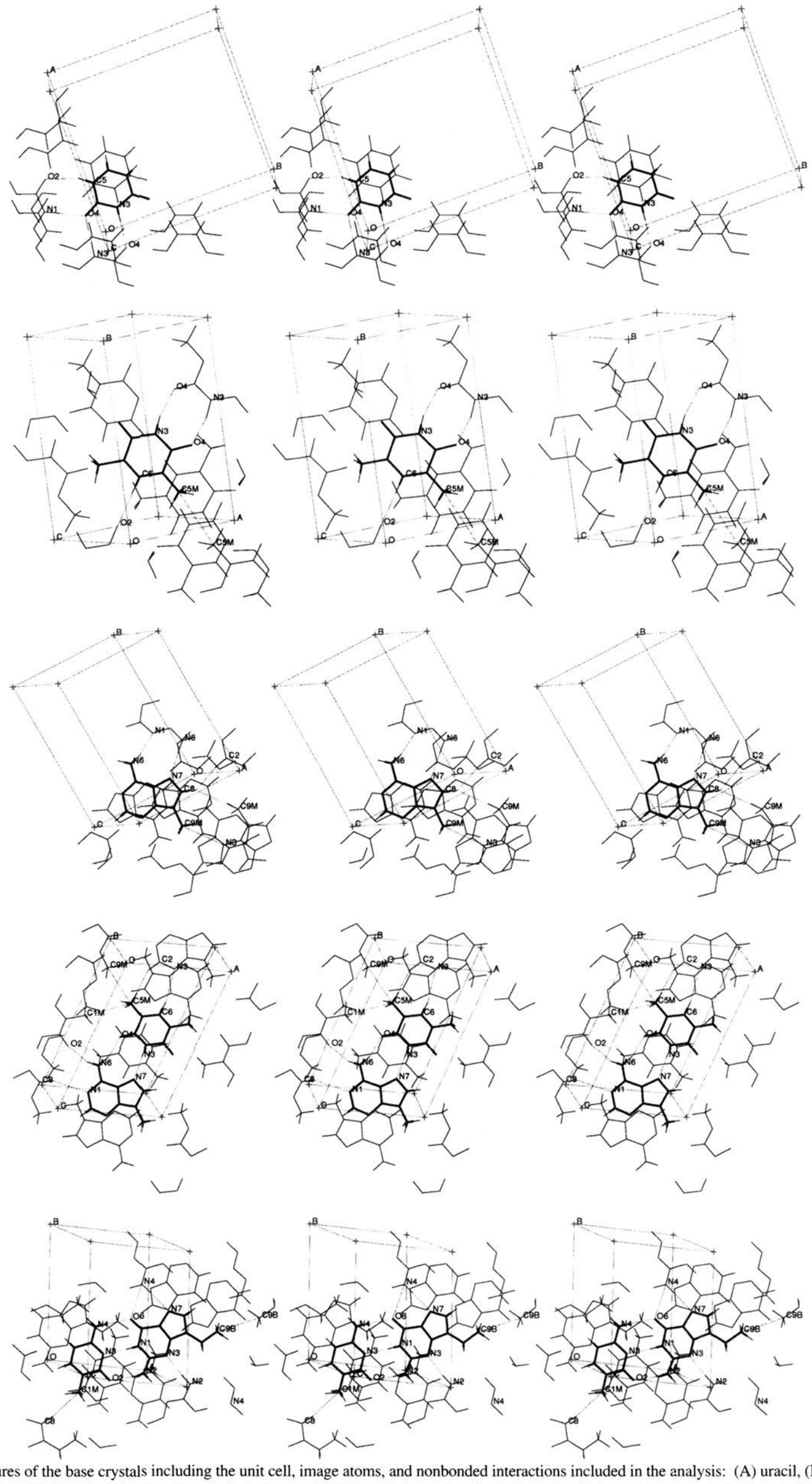

Figure 4. Structures of the base crystals including the unit cell, image atoms, and nonbonded interactions included in the analysis: (A) uracil, (B) 1-methylthymine, (C) 9-methyladenine, (D) 1-methylthymine, 9-methyladenine, and (E) 1-methylcytosine, 9-ethylguanine. Primary atoms are in bold and all unique image atoms within 4 Å of the primary atoms are included. The unit cell is indicated by the dashed line, with the origin (O) and A, B, and C axes indicated.

along the B axis. The small changes in these interaction distances lead to the small change in the B axis upon minimization. The shortening of the C8 to C2 distance is responsible for the contraction of the C axis. The A axis is primarily associated with stacking interactions; the calculated contraction is reasonable. The contraction of 2.1% of the 9-methyladenine, 1-methylthymine crystal was fairly uniform; i.e., the B axis changes little, while the A and C axes contract slightly. The shortening of the Hoogsteen interaction distances, N3 to N7 and O4 to N6, contributes to the contracton of the C axis. Further contributions to the C axis contraction come from the C5M to C2 and C6 to N3 interactions. The contraction of the O2 to N6, N1 to C8, and C5M to C1M distances leads to the A axis contraction. The increased C5M to C9 distance is related to the larger contraction of the C5M to C2 distance relative to that of C6 to N3. The consequence is a slight rotation of the primary thymine base relative to the adjacent adenine base. The B axis contraction was very small; this is in accord with the correct representation of the stacking interactions (Table 19 of the supporting information). The 0.14 Å contraction of the N1 to C8 distance indicates the van der Waals parameters at the C8 site to be appropriate. In the CHARMM22 histidine parameters,[96] the analogous site in imidazole has a van der Waals radius of 0.9 Å which is smaller than the aliphatic radius of 1.32 Å used here. The use of a smaller hydrogen radius in imidazole is associated with the carbon being situated between the two nitrogens. This introduces more polar character and allows the site to act as a hydrogen bond donor, as evident in the *ab initio* calculations. The present results suggest that the aliphatic parameter is sufficient. Also, the C5M to C1M distance is well maintained, supporting the transferability of the aliphatic parameters developed for the proteins based on small aliphatic model compounds[34] to the nucleic acids. Use of the aliphatic parameters for the methyl and ethyl groups in the nucleic acids is appropriate. The 1-methylcytosine, 9-ethylguanine crystal (Figure 4E) showed a significant contraction of the A axis while the B and C axes expanded only slightly. As the unit cell is triclinic the three angles defining the unit cell were included in the optimization. As may be seen the α angle increased, the β angle decreased, and the γ angle was essentially unchanged. The three Watson–Crick interaction distances listed in Table 18 of the supporting information, along with the C9B to C9B distance, are closely parallel to the A axis while the remaining four interactions are directed more along the B axis. The contraction of the A axis is associated with little change in the Watson–Crick interaction distances and the C9B to C9B distance. Correspondingly, the three interactions aligned with the B axis contract a small amount while the B axis expands. The minimized structure shows that the relative orientation of cytosine and guanine in the asymmetric unit has shifted slightly, such that guanine translated toward the O2 side of cytosine. This decreased the O2 to N2 distance and increased the N4 to O6 distance. Analysis of the relationship of the guanine shift to the unit cell parameters suggests that the γ angle between the A and B axes should be the most affected; however, only a small change is calculated. The contraction of the A axis and expansion of the B axis appear to be associated with this shift. The expansion of the C axis could have been associated with stacking effects (Table 19 of the supporting information); however, the calculated expansion of 0.08 Å is minimal.

(ii) Crystal Dynamics Results. Molecular dynamics simulations in the NVT ensemble with the lattice parameters fixed at the experimental values were performed to analyze the influence of temperature on the calculated interactions. The average distances and rms fluctuations from the 20-ps simulations are given in Table 18 of the supporting information; for comparison the experimental values and results from minimization are also listed. In all cases, the dynamics results for the interaction distances are in good agreement with the experimental values. The distances are generally longer than those from minimizations. The increased distances are due to the presence of kinetic energy in the system and to the fixed lattice parameters. In the minimizations the unit cell volume decreased in all cases and the largest differences in the interaction distances, such as C8 to C2 in 9-methyladenine and C5M to C9 in 1-methylthymine–9-methyladenine, are due to the latter. In the 1-methylthymine crystal, a significant improvement between the calculated and experimental distances was obtained, in contrast to the minimization results. This discrepancy points toward the importance of using dynamics as well as crystal minimizations in parameter testing. The use of NPT dynamics simulations, where the lattice parameters are allowed to vary, may yield additional insight.

Table 20 of the supporting information gives the rms positional deviations following a least-squares superposition with the crystal. The individual molecular geometries are accurate in both types of simulations. Pressures were calculated from the NVT simulations as shown in Table 21 of the supporting information. This provides an important check on the results. The pressure calculation in constant volume simulations is analogous to the calculations of densities in constant pressure simulations. The calculated values ranged from −2800 to 1000 atm. These values may be compared to results obtained for L-histidine.[96] In that study both NVT and NPT simulations were performed. The NVT calculations yielded pressures of approximately 8000 atm and the total unit cell volume expanded by approximately 4% in the NPT simulation. If the relationship of calculated pressure to volume change is similar for the present crystal simulations, the calculated pressures are satisfactory; i.e., an expansion of the unit cell volume of −1.4 to 0.5% would be expected. The rms fluctuations in the calculated pressures appear to be typical,[96] reflecting the sensitivity of the calculated pressures to the change in forces during the molecular dynamics simulation.

Heats of sublimation, as reported in Table 22 of the supporting information, were calculated from the 22−21−19 cutoff minimizations and NVT simulations and compared with experimental values.[77] Results from the minimizations show the calculated heats of sublimation to be overestimated by 5.5 to 9.2%. This is associated with the contraction of the unit cell upon minimization, which leads to an overestimate of the lattice energies that dominate the calculated heats of sublimation; other contributions come from the rotation–vibration energy and internal energy. The presence of kinetic energy in the system in the NVT simulations and the use of the experimental lattice parameters lowers the heats of sublimation, so that the calculated values were 0.9 to 14.4% less than the experimental values. Both uracil and 9-methyladenine are in good agreement with experiment, while the heat of sublimation of 1-methylthymine is significantly underestimated. The 14.4% discrepancy for 1-methylthymine may be related to the problems with the crystal minimization mentioned above. As with the interaction energies the change upon going from minimizations to simulations is associated with the presence of kinetic energy in the system and the lattice parameters being held at their experimental values. To gauge the influence of lattice relaxation on the heats of sublimation, they were calculated for uracil, 1-methylthymine, and 9-methyladenine based on 200-step ABNR minimization with the lattice fixed. The results are 30.3, 27.4, and 34.0 kcal/mol, respectively. These values are 1 to 3

kcal/mol smaller than the fully relaxed minimized values in Table 22 of the supporting information. Thus, it is likely that the calculated heats of sublimation would increase somewhat if the lattice parameters were allowed to relax, but the present agreement is satisfactory. A related study using minimization of the full unit cells[97] compared a number of parameter sets including the AMBER[5] and a previous CHARMM extended atom sets.[6] AMBER yielded values of 22.7, 22.9, and 28.6 and the CHARMM extended atom set yielded values of 26.2, 25.4, and 34.5 kcal/mol for uracil, 1-methylthymine, and 9-methyladenine, respectively. These values tend to be too low as compared to experiment with the exception of the extended CHARMM 9-methyladenine value. The present parameters yield better agreement with experiment when compared with the older parameter sets. This improved agreement is based on both the optimization of the partial atomic charges and the van der Waals parameters of the carbons in the purine and pyrimidine rings.

A decomposition of the interaction energy between the primary and image atoms in Table 22 of the supporting information shows that both the Lennard-Jones and Coulombic terms make significant contributions to the heat of sublimation. During the optimization of the van der Waals parmeters the well depths of the ring carbons were adjusted to obtain optimal agreement for the heats of sublimation. This yielded values of 0.14 kcal/mol for the ring carbons attached to the carbonyl oxygens and the amine groups, and 0.18 kcal/mol for the C5 and C6 pyrimidine carbons and for the remaining carbons in the purines. This emphasizes the importance of obtaining a proper balance between two terms contributing to the interaction portion of the potential energy.

The distances involved in nonbonded interactions are analyzed in Table 18 of the supporting information. The average difference between experiment and minimized values for hydrogen bonding heavy atoms was −0.05 Å. The average difference involving carbon to N or O atoms was 0.02 Å and that involving carbon to carbon interactions was 0.10 Å. The C to C interactions, however, included some extreme differences, ranging from 1.60 Å for the C5M to C5M interaction in 1-methylthymine to −0.44 Å for the C8 to C2 interaction in 9-methyladenine. Overall, the agreement with experiment is satisfactory, although for the nonbonded C to C interactions, some large discrepancies did occur. These structural changes emphasize the delicate balance necessary to maintain the crystal structure. Since polar interactions involving heteroatoms are the strongest, they tend to dominate the overall structure at the expense of the nonpolar interactions. The lattice parameters dominated by the stacking interactions (axis A for 9-methyladenine, axis B for 1-methylthymine−9-methyladenine, and axis C for uracil, 1-methylthymine, and 1-methylcytosine−9-ethylguanine) show an average deviation from experiment for the five crystals of 2.9%. This demonstrates that the stacking interactions are well represented by the potential energy function.

The presented CHARMM22 parameters for the nucleic acid bases satisfactorily reproduce a variety of experimental gas and condensed phase and *ab initio* data. For the nucleic acid bases in the gas phase the present parameters adequately reproduce survey data for the geometries and both the frequencies and normal mode assignments of both experimental and *ab initio* vibrational data. Limitations in the geometric data on the positions of the hydrogen atoms and in the vibrational data on the assignments of the normal modes, especially in the frequencies below 500 cm^{-1}, make additional optimization difficult. The availability of high-level *ab initio* data will allow some of these limitations to be overcome. Partial atomic charges and Lennard-Jones parameters were optimized to reproduce experimental and *ab initio* data. The obtained interaction parameters satisfactorily reproduce experimental heats of interaction for Watson−Crick base pairing and for several base homodimers, interaction energies, and geometries based on *ab initio* calculations for a variety of water−base interactions and experimental heats of sublimation of uracil, 1-methylthymine, and 9-methyladenine. Crystal minimizations and simulations show the parameters to reproduce both intermolecular interactions and the unit cell parameters. Base stacking, which is difficult to compare directly to experiment, appears to be modeled satisfactorily based on the changes in the stacking interaction energies in a B-DNA dodecamer, the position of minima of bases stacked in an ideal B-DNA geometry, comparison with low-level *ab initio* data, and the reproduction of stacking interactions in various crystal structures. Problems were encountered in the thymine crystal minimizations; however, results from the crystal simulations gave adequate agreement with experiment. Overall, the ability of the parameters to accurately treat the intramolecular portion of the force field along with the intermolecular interactions ensures their usefulness in condensed phase simulations.

(97) Norberg, J.; Nilsson, L. Personal comunication.

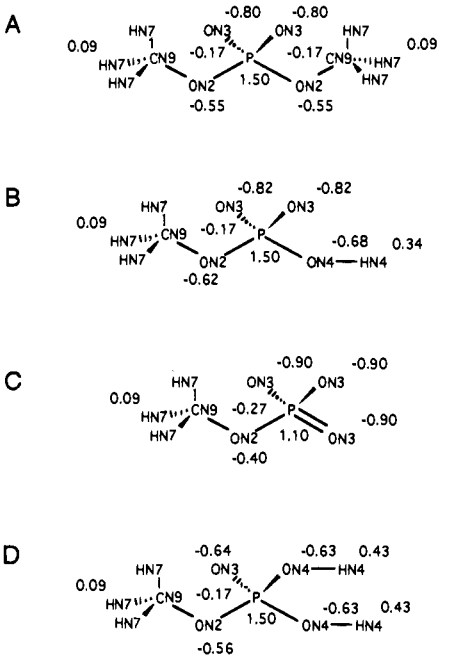

Figure 5. Diagrams of (A) dimethyl phosphate, (B) anionic methyl phosphate, (C) dianionic methyl phosphate including the atom types and charges, and (D) neutral methyl phosphate.

IV. Phosphate Parametrization

Parametrization of the phosphate group used the model compounds dimethyl phosphate (DMP), anionic methyl phosphate, and dianionic methyl phosphate (see Figure 5). Parameters were initially developed for DMP and then transferred to the methyl phosphates. In the DMP parametrization the energies and geometries of the three minima associated with the O1−P1−O2−C2 and O2−P1−O1−C1 dihedral angles (Figure 6) (gauche,gauche (g,g), gauche,trans (g,t), and trans,trans (t,t)) were considered in the optimization. Parameters not present in DMP were then optimized to fit the methyl phosphate data. Figure 5 shows the atom types and charges for the three model phosphate compounds.

(a) Internal Parametrization. Internal geometries for DMP were based on data from a survey of the CCDB[31] and are presented in Table 23 of the supporting information with the

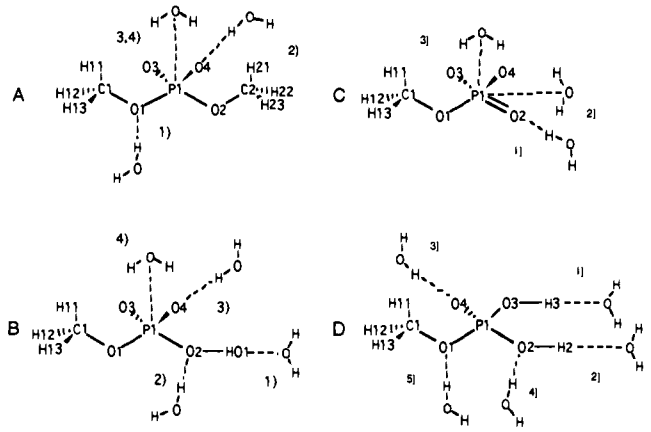

Figure 6. Water–phosphate interaction geometries for (A) dimethyl phosphate, (B) anionic methyl phosphate, (C) dianionic methyl phosphate, and (D) neutral methyl phosphate.

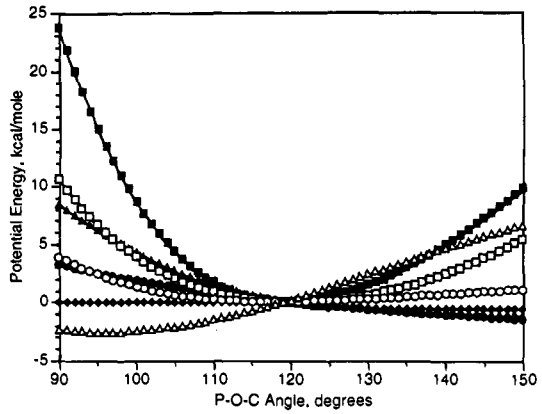

Figure 7. Adiabatic potential energy surface associated with the P–O–C angle of dimethyl phosphate. Included are the total (■), electrostatic (●), van der Waals (▲), dihedral (◆), angle (□), bond (○), and Urey–Bradley (△) energy contributions.

results from the present parameters. The empirical results include the structures for the (g,t) and (t,t) conformers, as well as the experimental (g,g) conformer (see below). Agreement between the survey and empirical results is good, with only the O3–P–O4 angle being outside the uncertainity of the survey data. To obtain the correct P–O1–C1 angle a Urey–Bradley term was introduced. The equilibrium value for that angle had to be set smaller (114°) than the desired angle (120°). This was necessitated by the electrostatic and van der Waals repulsion between the methyl and phosphate moieties of DMP. Vacuum simulations of DMP (Langevin Dynamics, 5-ps equilibration, 50-ps production) showed the average values of the P–O–C angles to be slightly larger than the minimized values in Table 23 of the supporting information. These simulations yielded average values of 121.1° (4.5) and 121.2° (4.1) for the P–O–C angles, where the value in parentheses represents the rms fluctuation about the average. The change upon going from the minimized structure to the dynamics average is due to the anharmonic effective potential acting on the P–O–C angles. The anharmonicity is such that the increase in energy is smaller as the angle opens than when the angle decreases.

Such a change in geometry upon going from the minimized to the dynamics averaged structure does not normally occur for bond length or bond angle degrees of freedom with calculations based on empirical potential energy functions. The dynamics and minimized values are generally equivalent, as expected from the stiff harmonic nature of the potential function. Examples include the DMP dynamics averages for the O–P–O angles. They are 108.8° (2.8) for O3–P–O1, 101.6° (3.1) for O1–P–O2, and 117.6° (2.5) for O3–P–O4 where the fluctuations are in parentheses; all of these are within 0.2° of the minimized values. However, if the effective potential is very anharmonic, as it is for the P–O–C angle and the P–O–H angle in anionic methyl phosphate (see below), nonbonded terms can significantly influence the energy surface of a particular internal coordinate so that the dynamics average structure deviates from the minimized structure.

Figure 7 shows the anharmonic nature of the DMP P–O–C angle adiabatic surface, including the contributions from the individual components in the potential energy function. Decreasing the P–O–C angle leads to a significantly larger increase in energy than to the opening of the angle. The increased energy upon closure is associated with the angle, van der Waals, bond, and electrostatic terms; the Urey–Bradley term partially compensates for the other terms because its equilibrium value was set shorter than the distance associated with the minimized angle to counteract the repulsive contributions of the nonbonded terms that lead to an opening of the angle. As the angle decreases the repulsive terms, due to the movement of the methyl group toward the phosphorus, increase rapidly, leading to anharmonicity. It is not that the opening is "softer". Such effects are more commonly observed in dihedral angle energy surfaces since the internal energy terms are much weaker. They emphasize the requirement that an empirical potential function must maintain consistency between the internal and external interactions and that dynamics as well as minimization studies are needed to evaluate the parameters.

For the anionic and dianionic methyl phosphates, survey results from the CCDB[31] were used to optimize the equilibrium internal parameters that do not occur in DMP; they are terms associated with the addition of a proton to one of the phosphate anionic oxygens and the change of the oxygen atom type from ON3 to ON4. Table 24 of the supporting information contains the results from the CCDB survey and the empirical minimized anionic and dianionic methyl phosphate structures. Comparison of the survey results with the empirical optimized geometries shows the agreement to be satisfactory. The agreement is generally somewhat poorer than that for DMP due to the transfer of parameters already mentioned. For the bonds the agreement is 0.022 Å or better. The O(C)–P=O, O(H)–P=O, and O=P=O angles (3.8, −6.0, and 1.6%, respectively) are outside the statistical errors from the survey results; see Table 24 of the supporting information. The O=P=O and O(C)–P=O parameters are taken directly from DMP, precluding additional optimization, while the O(H)–P=O angle is unique to anionic methyl phosphate. Attempts were made to further optimize this angle; however, alterations of the parameters led to larger deviations in some of the other angles. The P–O–H angle, as in DMP, is influenced by the strong effect of the external portion of the force field on the optimized geometry. This is due to electrostatic interactions between the OH proton and the oxygen atoms. A Urey–Bradley term for the P–O–H angle was used to obtain the value listed in Table 24 of the supporting information. This value is in agreement with the 6-31+G* optimized value of 106.6°. The average P–O–H angle from an anionic methyl phosphate vacuum simulation, with the same methodology as for DMP, was 106.2° (5.0).

As with anionic methyl phosphate, the optimization of the dianionic species was limited by the transfer of parameters directly from DMP and the anionic species. For the bond lengths the values are within 0.03 Å of the CCDB values and the largest discrepancy for the angles is 0.4°. Thus, the transferability of the phosphate parameters from DMP to anionic

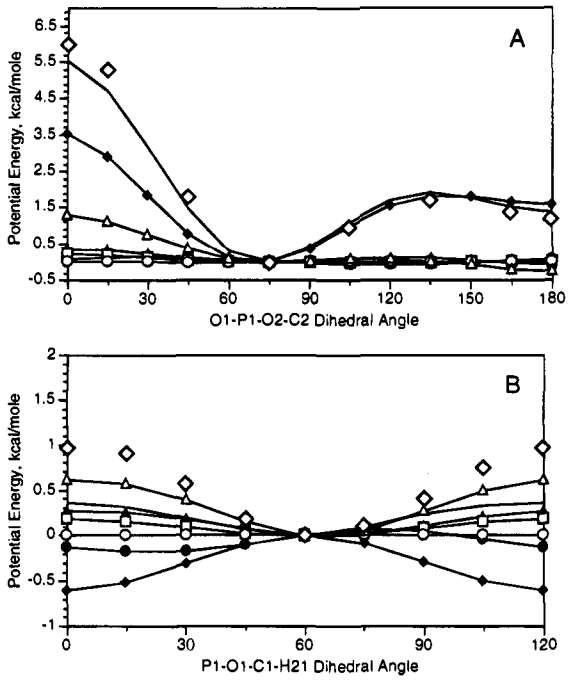

Figure 8. Adiabatic potential energy torsional surfaces of dimethyl phosphate for the (A) O1−P1−O2−C2 surface and the (B) P1−O1−C1−H21 surface. Included are the 6-31G* *ab initio* values (◇) and the empirical total (−), electrostatic (●), van der Waals (▲), dihedral (◆), angle (□), bond (○), and Urey−Bradley (△) energy contributions.

and dianionic methyl phosphate yields reasonable results. Improvement could be introduced by the addition of new atom types.

For the parametrization of the DMP internal force constants, 3-21G* *ab initio* frequencies for the three DMP minima were calculated. The resulting frequencies are included in Table 25 of the supporting information along with experimental values for the lowest energy g,g rotamer.[98] Comparison of the 3-21G* results with recently published vibrational data calculated at the HF/6-31G* level shows satisfactory agreement.[99] The assignments of Jayaram *et al.*[100] were used to compare the experimental and *ab initio* data allowing for the determination of the ratios between the 3-21G* and experimental frequencies for the g,g rotamer to be used for scaling of the *ab initio* frequencies prior to fitting. Negative frequencies associated with modes 1 and 2 of the t,t conformer are related to the torsional surface in that region being exceptionally "flat" (see Figure 8). The average ratio between the assigned experimental frequencies and the HF/3-21G* values for the g,g conformer was 0.895. It is close to the 0.90 scale factor generally employed and was used to scale the remaining *ab initio* frequencies. The results are given in Tables 26−28 of the supporting information and include the scaled frequencies and assignments. Those tables also contain the CHARMM22 frequencies for the gg, g,t, and t,t conformers of DMP, respectively. The empirical and *ab initio* frequencies and assignments for three DMP isomers are in satisfactory agreement. The average percent difference between the scaled *ab initio* results and the empirical values is 3.8%, when the two lowest (negative) frequency t,t conformer modes are ignored. In the high-frequency region above 1300 cm^{-1} the agreement is satisfactory. These modes are associated with parameters for the methyl groups that were taken directly from the aliphatic parameter set;[34] the agreement between the *ab initio* and empirical values for the methyl stretches and deformations supports the approach based on a transfer of parameters. For the modes including and below 1350 cm^{-1}, which are associated with force constants adjusted in the present parametrization, the average percent difference is 5%. Agreement for the methyl rocks, the P=O stretch, the C−O stretch, and the P−O stretch is good for both the frequencies and assignments for all three DMP conformers. In the case of the P=O stretch the lower frequency, at approximately 1100 cm^{-1}, is better fit than the higher frequency at approximately 1300 cm^{-1}. Throughout the remainder of the low-frequency region the *ab initio* spectra for all three conformers are well reproduced by the empirical force field. In the 300−500-cm^{-1} region of the spectra there are some incorrect assigments. For example, in the g,g conformer, the *ab initio* results predict the O−P wagging and twisting modes to occur at 454 and 375 cm^{-1}, respectively, while the empirical model yields a reversal of the ordering with values of 412 and 419 cm^{-1}, respectively. In the g,t and t,t conformers, however, the ordering of the modes is the same in both the *ab initio* and empirical calculations. Modes in this region are associated with bending, wagging, twisting, and rocking modes involving the phosphorus atom. Considering the large degree of mixing of these modes the present empirical spectra are adequate. As mentioned above, Urey−Bradley terms were introduced to allow both the geometry and frequency associated with the P−O−C angles to be optimized. The empirical P−O−C frequency is in good agreement with the scaled 3-21G* values for all three conformers. Optimization of the frequencies associated with rotation around the C−O torsion requires the use of negative force constants for the corresponding X−C−O−X torsion parameter (see Appendix, supporting information). This was necessary because the nonbonded interactions between the methyl and phosphate moieties led to frequencies that were too high, even when the force constants associated with that torsion were set to zero. Agreement between the HF/3-21G* and empirical frequencies for the P−O torsions is good for the g,g and g,t conformers. In contrast to the negative P−O torsion frequencies obtained in the *ab initio* calculation for t,t, the empirical model gave positive results that are similar to the g,g and g,t conformer values.

To check the transferability of the DMP parameters to the methyl phosphates, *ab initio* frequencies at the 6-31G* level were calculated for anionic methyl phosphate. Table 29 of the supporting information lists the 6-31G* frequencies, scaled by 0.90, and the assignments along with the empirical results. Additional terms associated with the alcoholic proton and a new atom type, ON4, allowed for some flexibility in the fitting. For the anionic methyl phosphate the average deviation for all frequencies is 4.7%. The average deviation is 1.7% for frequencies above 1350 cm^{-1} and 5.9% for those below. The overall agreement is satisfactory. The frequency associated with the P−OC torsion is high (by 48 cm^{-1}) while the C−O torsional mode is in satisfactory agreement with the *ab initio* values. The P−OH torsion is slightly higher than the *ab initio* value; however, the agreement is reasonable considering the strong electrostatic contribution and resulting anharmonic behavior (see above). Urey−Bradley terms are used for the P−O−C angle bend, as discussed for DMP, and the P−O−H angle bending mode.

The proper energy ordering of the three gas phase conformers (g,g, g,t, and t,t) is essential for DMP. The three conformers were initially optimized at the 3-21G* and 6-31G* levels.[101] MP2/6-31G* energies were obtained subsequently with the

(98) Garrigou, C.; Bouloussa, O.; Clement, C. *Can. J. Spectrosc.* **1976**, *21*, 75−82.
(99) Liang, C.; Ewig, C. S.; Stouch, T. R.; Hagler, A. T. *J. Am. Chem. Soc.* **1993**, *115*, 1537.
(100) Jayaram, B.; Ravishanker, G.; Beveridge, D. L. *J. Phys. Chem.* **1988**, *92*, 1032−1034.

6-31G* geometries. The relative energies of the three conformers are reported in Table 30 of the supporting information, which also contains the present empirical results. The g,g conformer has the lowest energy in the gas phase, followed by the g,t and t,t conformers, in that order (see also ref 100). Agreement between the empirical and the *ab initio* results concerning both the order and relative energy differences of the conformers is good. The empirical energy difference of the conformers is larger than that used in previous work,[5,6,102] but it is in accord with the higher level *ab initio* calculations. The energy differences in the empirical energy function calculation are due primarily to the presence of a 2-fold dihedral, in addition to the 3-fold term, associated with the O−P−O−C torsion (see Appendix, supporting information). This term was adjusted to obtain approximate agreement between both the tt/gg and gt/gg *ab initio* energy differences; the gt/gg energy difference was emphasized because it plays a more important role in thermodynamic properties. Counterbalancing the dihedral angle terms for the different conformers are the van der Waals and Urey−Bradley terms. The van der Waals energy is more favorable in the more extended g,t and t,t structures, as is the Urey−Bradley term, which was included for the P−O−C angle to obtain the correct geometries and frequencies (see above).

The adiabatic O1−P1−O2−C2 and P1−O1−C1−H11 dihedral angle rotation surfaces and the two-dimensional O1−P1−O2−C2 versus O2−P1−O1−C1 adiabatic surface were checked to verify the parametrization of these important dihedral angles. The O1−P1−O2−C2 and P1−O1−C1−H11 adiabatic surfaces are compared with the 6-31G* *ab initio* adiabatic surface in Figure 8, parts A and B, respectively. The empirical maps include the individual component contributions to the potential energy function. For the O1−P1−O2−C2 surface, the agreement with the *ab initio* surface is good for the barrier heights and the energetic ordering of the minima. The dihedral angle terms dominate the surface and it can be seen how the use of a combination of 2-fold and 3-fold terms controls the relative ordering of the minima. The other significant contribution comes from the Urey−Bradley term associated with the P−O−C angle, which raises the barrier at 0° as the P−O−C angle increases to 123.7° from 119.2° at 75° and 118.4° at 180°. The agreement between the empirical and calculated values of P1−O1−C1−H11 (Figure 8B) is less good. Although the overall shape is correct, the barrier height is too low. This is due to the dihedral contribution associated with the negative force constant required to optimize the vibrational spectra. Since the frequencies associated with the O−C torsion yielded good agreement with the scaled *ab initio* frequencies for the three DMP confomers and for anionic methyl phosphate, further optimization was not performed. The adiabatic DMP torsional potential energy contour map is shown in Figure 9. The g,g and g,t minima are rather localized while the minimum in the region of the t,t conformer is rather broad. A recent study calculated the adiabatic energy surface of DMP for dihedral angles of 0 to 180° in 30° increments for the O1−P1−O2−C2 angle with O2−P1−O1−C1 fixed at 60°.[23] The results showed an energy of 10.75 kcal/mol above the global minimum at (60°, 0°) for the two dihedral angles. This decreased to 0.47 kcal/mol at (60°, 77°), increased to 1.94 kcal/mol at (60°, 150°), and again decreased to 1.57 kcal/mol at (60°, 180°). Comparison of these values with the 2-dimensional adiabatic map in Figure 9 shows the agreement to be satisfactory. The DMP contour map from the CHARMM extended atom parameter set[6]

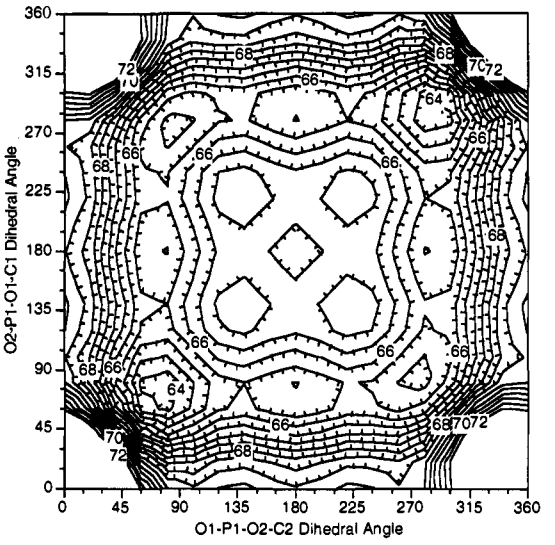

Figure 9. Dimethyl phoshate empirical adiabatic torsion map, O1−P1−O2−C2 versus O2−P1−O1−C1. Energy contours are plotted in increments of 0.5 kcal/mol.

had an overall topology similar to the present map. In the present map the barriers between the minima are higher than previously and the minimum in the t,t region is broader. These differences are due to the use of higher level *ab initio* results as the data for the optimization of the phosphate parameters.

Analysis of the contributions of the individual terms in the potential energy function to the surfaces presented in Figure 8 reveals that although the interaction terms are significant, their overall effect is minimal. In the O1−P1−O2−C2 surface, this is due to a balance of the electrostatic interactions between the carbon and hydrogen atoms on the methyl group and the anionic oxygen atoms and phosphorus atoms in the phosphate. In the P1−O1−C1−H11 surfaces the electrostatic contribution to the barrier at 0° is favorable while the Lennard-Jones contribution is unfavorable. This balance is due to the favorable electrostatic interaction between the methyl hydrogens and the anionic oxygens at the minimum and the simultaneous van der Waals repulsion between those atoms. These contributions again emphasize the importance of balancing the intra- and intermolecular portions of the force field and the electrostatic and van der Waals contributions to the intermolecular portion.

(b) Interaction Parameters. Interaction parameters for the phosphate groups were derived from *ab initio* calculations of the interaction energy and geometry for water complexed with dimethyl phosphate, anionic methyl phosphate, and dianionic methyl phosphate as shown in Figure 6, parts A, B, and C, respectively. Also included are results for neutral methyl phosphate (Figure 6D), which have been used to obtain parameters for continuum electrostatic calculations.[103] *Ab initio* calculations were performed at the 6-31G* level with only the distances shown in Figure 6 optimized; i.e., the orientations of the water molecules were fixed as shown in the figures. For DMP (g,g), anionic methyl phosphate, and neutral methyl phosphate, the 6-31G* optimized monomer geometries were used while the 6-31+G* optimized geometry was used for dianionic methyl phosphate. The diffuse function in the dianionic methyl phosphate was included due to the highly charged nature of the compound.[104] The 6-31G* and empirical interaction energies and geometries are listed in Table

(101) Schlenkrich, M. Ph.D. Thesis, Department of Chemistry, Darmstadt Technical High School, Germany, 1992.

(102) Jayaram, B.; Mezei, M.; Beveridge, D. L. *J. Comp. Chem.* **1987**, *8*, 917−942.

(103) MacKerell, A. D., Jr.; Sommer, M. S.; Karplus, *J. Mol. Biol.* **1995**, *247*, 774−807.

(104) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, T. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986.

Table 5. Empirical and *ab Initio* 6-31G* Water Interaction Energies and Geometries for Dimethyl Phosphate, Anionic Methyl Phosphate, Dianionic Methyl Phosphate, and Neutral Methyl Phosphate[a]

	6-31G*		empirical[b]	
interaction	E_{min}	R_{min}	E_{min} (LJ)	R_{min}
Dimethyl Phosphate				
(1) O2−HW	−7.37	2.02	−6.94 (−0.05)	1.89
(2) O3−HW	−13.32	1.85	−13.60 (1.99)	1.66
(3) P−OW[c]	−10.48	3.73	−11.37 (−0.53)	3.59
(4) P−OW[c]	−16.64	3.38	−17.21 (1.59)	3.28
Anionic Methyl Phosphate				
(1) H2−OW	2.31	2.14	1.58 (0.11)	1.92
(2) O2−HW	−11.34	1.95	−10.88 (0.54)	1.83
(3) O3−HW	−13.99	1.85	−13.74 (2.24)	1.65
(4) P−OW	−17.18	3.39	−17.17 (1.64)	3.27
Dianionic Methyl Phosphate				
(1) O3−HW	−24.28	1.71	−21.42 (3.27)	1.60
(2) P−OW	−22.71	3.22	−24.44 (1.01)	3.18
(3) P−OW	−30.91	3.32	−31.04 (3.80)	3.21
Neutral Methyl Phosphate (6-31G* Energies Scaled by 1.16)				
(1) H2−OW	−9.69	1.86	−9.26 (1.44)	1.77
(2) H3−HW	−9.91	1.84	−10.05 (1.71)	1.76
(3) O4−HW	−6.78	2.03	−6.75 (0.99)	1.74
(4) O2−HW	−3.39	2.11	−3.32 (0.19)	1.89
(5) O3−HW	−3.64	2.13	−3.65 (−0.12)	1.91

[a] Energies in kcal/mol and distances in Å. [b] Empirical values were obtained using the CHARMM22 optimized gas phase geometry, except in the case of the neutral methyl phosphate where the 6-31G* geometry was used. [c] Interaction 3 has the plane of the water molecule in the same plane as the O=P=O atoms while in interaction 4 the planes are perpendicular. The total interaction energy is given with the van der Waals contribution in parentheses.

5. In contrast to the base−water interactions the *ab initio* interaction energies are not scaled because of the phosphate being charged. This is in accord with the analysis of Jorgensen and co-workers[7] and with studies in our laboratory on a variety of both cationic and anionic model compounds.[105] In the optimizations the only variables were the partial atomic charges, whose sum was constrained to be an integer, and the phosphorus van der Waals parameters. The remaining van der Waals parameters had been previously determined and were transferred directly to the present systems (see above). The anionic oxygens van der Waals parameters were obtained from the carboxylic acid parameters (proteins), the ester or alcohol oxygens and the polar hydrogens were equivalent to the CHARMM TIP3P oxygen, and the methyl group van der Waals parameters were from the aliphatic parameter set. As with the internal parameters optimization was first performed on DMP. The empirical parameters overestimate the *ab initio* values by 2.1 to 8.5% for the three strongest interactions while the weakest interaction is underestimated by 5.8%. Such a balance of the interactions is required because the simplicity of the potential energy function does not allow the individual interactions to be reproduced exactly. Another DMP force field[102] has interaction energies greater than those in the present work, apparently due to the use of lower level *ab initio* calculations to obtain the data which were used for determining the parameters. The minimum interaction distances are between 0.1 and 0.2 Å shorter than the *ab initio* values; the greatest decrease occurs for the neutral phosphate O4 to HW interaction which is 0.29 Å shorter. The combination of the interaction energies being equivalent to the HF/6-31G* values and the distances being shorter has been shown to yield agreement with experiment concerning heats of solvation in condensed phase simulations.[8,106] The optimization of the phosphorus van der Waals parameter aided in the balancing of the four interactions. The Lennard-Jones contributions to the empirical interaction energies vary from −0.53 to 1.99 kcal/mol. Their important role is consistent with earlier studies of polar species.[38] The value of ϵ for phosphorus, 0.585 kcal/mol, corresponds to that determined via the Slater and Kirkwood equation[107,108] as reported by Nilsson and Karplus.[6] The radius, R_{min}, however, was increased to 2.15 from 1.9 Å to better balance the interaction energies. The R_{min} and ϵ of phosphorus are both large in magnitude as compared to the remaining van der Waals parameters (see Appendix); similar values have been obtained in the parametrization of sulfur[109] and may be necessary for second-row atoms, in general, for the type of potential energy function used here. The partial atomic charges of anionic and dianonic methyl phosphate were optimized using the same phosphorus van der Waals parameters as for DMP. The interaction energy and distance results shown in Table 5 are again in good agreement with the 6-31G* values for both compounds. For the anionic compound, the phosphorus van der Waals term aided in balancing the interaction energies of the various water positions, while for the dianionic compound it tended to increase the deviation. Since the interaction energies with water are all very large for the dianionic methyl phosphate, their relative values are probably less important. The neutral methyl phosphate water interactions, also included in Table 5, show good agreement with the scaled 6-31G* values; i.e., a scaling factor of 1.16 was used for the *ab initio* interaction energies, consistent with the parametrization used with other polar neutral compounds (see above[38]). Of significance in the neutral methyl phosphate results are the proper ordering of the first and second interactions. At these sites the atoms are identical, except for their relative position with respect to the remainder of the molecule. Achieving the correct order provides an indication of the accuracy of the overall charge distribution. Figure 5 shows the charges of the various phosphate compounds. The magnitudes of the partial atomic charges are relatively large and similiar to the charges from a Mulliken population analysis of the 6-31G* calculations (see Table 31 of the supporting information).

(c) Sugar Parametrization. The details of the parametrizaton for the ribose and deoxyribose sugars for use with the present all-hydrogen nucleic acid parameters will be presented elsewhere.[42] Included with the Appendix are the sugar parameters and Figure 10 contains the partial atomic charges and atom types for the sugar moieties. The interaction parameters were obtained from the protein aliphatic[34,47] and alcohol groups.[8] Optimization of the internal terms was based on survey results from the CCDB.[110−112] To obtain agreement with the survey results concerning sugar puckering as defined by the pseudorotation model a single 2-fold term was used in addition to the standard 3-fold terms. The 2-fold term aids in reproducing the relative energies of the sugar conformations.[113] In the present set the single 2-fold term is used for the Cl′−C2′−C3′−O3 dihedral angle; adjustment of the force constant associated with that term allows for a proper fit of the ribose and deoxyribose

(105) Kuczera, K.; MacKerell, A. D., Jr.; Gao, J.; Karplus, M. Manuscript in preparation.

(106) Gao, J. Ph.D. Thesis, Purdue University, Department of Chemistry, 1987.

(107) Slater, J. C.; Kirkwood, J. G. *Phys. Rev.* **1931**, *37*, 682.

(108) Pitzer, K. S. In *Advances in Chemical Physics*; Prigogine, I., Ed.; Interscience: New York, 1959; Vol. 2.

(109) Nguyen, D.; Lau, F. T. K.; MacKerell, A. D., Jr.; Karplus, M. Manuscript in preparation.

(110) Arnott, S.; Hukins, D. W. L. *Biochem. J.* **1972**, *130*, 453−465.

(111) De Leeuw, H. P. M.; Haasnoot, C. A. G.; Altoona, C. *Isr. J. Chem.* **1980**, *20*, 108−126.

(112) Olsen, W. K.; Sussman, J. L. *J. Am. Chem. Soc.* **1982**, *104*, 270−278.

(113) Olsen, W. K. *J. Am. Chem. Soc.* **1982**, *104*, 278−286.

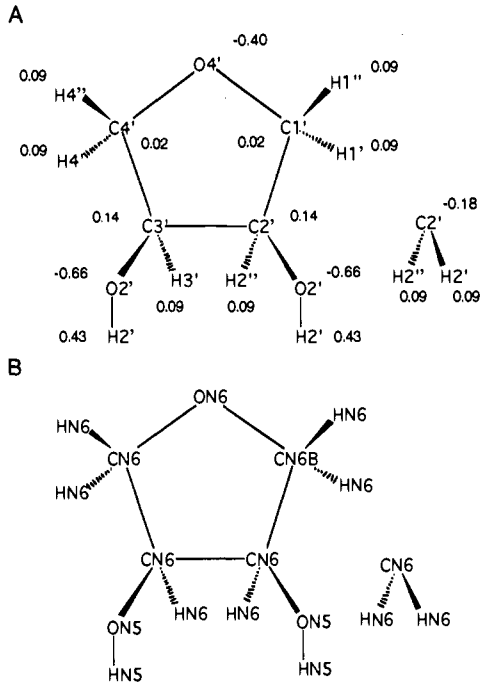

Figure 10. Atom names and partial atomic charges (A) and atom types (B) for ribose. Insets contain the partial atomic charges (A) and atom types (B) required for the treatment of deoxyribose.

pseudorotation surfaces. This represents a decrease in the number of "gauche" terms used to reproduce the pseudorotation surface, as first described by Olsen[113] and subsequently used in other parameter sets.[4-6] Optimization of other internal force constants was based on *ab initio* frequency calculations at the 3-21G level. Overall, the resultant geometries, frequencies, and pseudorotation maps were in good agreement with the survey and *ab initio* values, and the parameters yield agreement with the crystal results for nucleic acid structures, as described below.

V. Parameter Testing via Crystal Simulations

A global test of the parameters was performed by simulating a GpC dimer,[43] a B-DNA decamer,[44] and a Z-DNA hexamer[45] in full crystal environments, including solvent and counterions. Of the three test systems the GpC dimer is the best resolved with an *R* factor of 5.4% based on data collected out to 0.89 Å.[43] It has a C2 unit cell and contains four GpC molecules, 36 waters, and 4 sodium ions. The B-DNA decamer, CCAACGTTGG, has been refined to a resolution of 1.4 Å with a final *R* factor of 17.9%. The crystal is of space group *C*2 with 1 strand of DNA, 4 Mg^{2+} water clusters, and 49 additional waters per asymmetric unit. The Z-DNA hexamer, CGCGCG, has a resolution of 1.25 Å to an *R* factor of 17.6%. The crystallographic asymmetric unit comprises 1 Z-DNA hexamer duplex, 4 ions, and 102 water molecules of space group $P2_12_12_1$. The high resolution of these structures makes possible comparison of the experimental and dynamics averaged properties of the molecules. Additionally, the presence of ions and solvent in the X-ray structures allows for the analysis of hydration effects. GpC was simulated in the same way as the base crystal simulation (see above) using an NVT ensemble (20 ps), while the B- and Z-DNA simulations were performed (100 and 110 ps, respectively) in the NVE ensemble following an initial heating and equilibration stage. Although none of these simulations are very long, they are sufficient for a comparison of the local properties of the simulated crystal with experiment. Further, the fact that many of the ions and water molecules are observed in the crystal suggests that they are localized sufficiently so that averaging over widely different configurations is not required. Details of the Z-DNA simulation will be presented elsewhere,[114] but a brief description is included in the present manuscript since it is important for validating the parameter set.

(a) GpC Crystal Simulation. The full unit cell was simulated to avoid problems from the fact that two water molecules in the asymmetric unit (which contains 1 GpC, 9 waters, and 1 Na^+)[43] are on crystallographic 2-fold axes. All sodium ions and water molecules are well resolved in the crystal. As with the base calculations described in Section III, the GpC crystal was minimized with several truncation distances. These calculations are shown in Table 32 of the supporting information. There is a convergence in the behavior as the nonbond truncation distance is increased. The 22−21−19 cutoff scheme used for the bases was employed here for analysis. Table 6 compares the experimental and calculated lattice parameters from the minimization, the average pressures from the 20-ps NVT dynamics simulation, and the Watson−Crick base pairing hydrogen bonding distances. The minimizations show that the unit cell contracts slightly along the A and C axes and expands along the B axis, leading to an overall contraction of the unit cell by 1%. This contraction was less than that observed in the nucleic acid base calculations (Table 17 of the supporting information) and the average pressure is approximately 4500 atm from the NVT simulation. This pressure, as compared with L-histidine[91] (see above), corresponds to a 2−4% expansion of the unit cell.

The Watson−Crick base pairing interaction distances obtained from minimization and dynamics are also shown in Table 6. The contraction of all the distances is the same (0.1 Å) in the minimizations. In the dynamics simulations the distances increase, with the largest deviation from experiment being 0.05 Å. The agreement with experiment is excellent and there is no systematic deviation. This supports the derivation of the charge scheme as outlined above. In particular, the optimization of the short interaction distances relative to the *ab initio* calculation is seen to work well in a crystal environment.

Data on the internal structures are presented in Tables 33 and 34 of the supporting information[43] for the nucleic acid backbone and the sugar pseudorotation parameters, respectively. The backbone dihedrals are defined as follows:[115]

P--O5'--C5'--C4'--C3'--O3'--P
α---β---γ---δ---ε---ζ

χ, purine: O4'--C1'--N9--C4 and χ, pyrimidine: O4'--C1'--N1--C2

and the sugar puckering is defined in terms of the pseudorotation model.[116] Minimization of GpC led to only minor changes in the nucleic acid backbone; the NVT molecular dynamics simulations led to some larger changes. The largest differences involve the β and ϵ backbone dihedral angles, though even those are relatively small (less than 14°). These angles are important for the conformation of the phosphodiester linkage between the nucleotides. The differences occur in an anticorrelated fashion (i.e., β shifts are negative and ϵ shifts are positive) such that the overall change in position of the phophates and the sugars is minimal. In all cases the dihedral changes leave the calculated values in the same well as that observed experimentally. Values of the pseudorotation angle and amplitude are presented in Table

(114) Jung, S.-H.; MacKerell, A. D., Jr.; Karplus, M. Manuscript in preparation.
(115) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1983.
(116) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205−8212.

Table 6. Experimental and Calculated Values for the GpC Crystal[a]

lattice parameters	exp	calcd	% diff
A	21.460	21.384	−0.4
B	16.297	16.383	0.5
C	9.332	9.223	−1.2
β	90.56	90.40	−0.2
vol	3264	3231	−1.0
dynamics pressure	1 atm	4556 ± 4118 (ext), 4558 ± 2850 (int)	

H bond distances	exp	calcd	diff
minimization			
guaN1−CytN3	2.95	2.84, 2.84, 2.84, 2.84	−0.11
guaN2−CytO2	2.86	2.76, 2.76, 2.76, 2.76	−0.10
guaO6−CytN4	2.91	2.80, 2.80, 2.80, 2.80	−0.11
dynamics			
guaN2−CytO2	2.86	2.85 ± 0.12, 2.81 ± 0.10, 2.86 ± 0.12, 2.85 ± 0.12	−0.01, −0.04, 0.01, 0.00
guaN1−CytN3	2.95	2.98 ± 0.10, 2.97 ± 0.10, 2.95 ± 0.08, 2.99 ± 0.09	0.03, 0.02, 0.00, 0.04
guaO6−CytN4	2.91	2.95 ± 0.15, 2.96 ± 0.16, 2.92 ± 0.13, 2.96 ± 0.14	0.04, 0.05, 0.01, 0.05

[a] The 22−21−19 cutoff scheme was used (see Table 32, supporting information). All distances in Å and angles in deg.

Table 7. Nucleic Acid−Water and Sodium−Water Interactions from the GpC Crystal[a,b]

	average position, Å			hydration number		
site[c]	crystal	MD	diff	crystal	MD	diff
phosphate						
anionic O	3.02 ± 0.29	3.05 ± 0.28 (2.65)	0.03	3.0	3.80	0.80
ester O	3.22 ± 0.24	3.10 ± 0.23 (2.95)	−0.12	1.0	1.35	0.35
bases						
Cyt N4		3.35 ± 0.12 (3.45)		0	0.33	0.33
Cyt O2	2.80 ± −	2.95 ± 0.28 (2.75)	0.19	1.0	1.24	0.24
Gua N2	2.90 ± −	3.13 ± 0.21 (2.95)	0.23	1.0	1.80	0.80
Gua N3	2.84 ± −	2.99 ± 0.19 (2.85)	0.15	1.0	0.86	−0.14
Gua N7	2.91 ± −	2.94 ± 0.15 (2.85)	0.38	2.0	2.02	0.02
Gua O6	3.03 ± −	3.05 ± 0.27 (2.75)	0.02	2.0	1.84	−0.16
sugar						
O2′	2.97 ± 0.33	3.03 ± 0.23 (2.85)	0.06	2.0	2.07	0.07
O3′	3.01 ± 0.20	3.01 ± 0.26 (2.75)	0.00	3.0	2.71	−0.29
O4′†	4.24 ± 0.37	4.49 ± 0.54 (3.35)	0.25	2.0	2.43	0.43
O5′	3.46 ± −	3.21 ± 0.19 (3.25)	−0.25	1.0	1.08	0.08
Na^+ to water O	2.39 ± 0.07	2.34 ± 0.17 (2.25)	−0.05	4.0	4.05	0.05
Na^+ to phosphate*						
anionic O	2.33 ± −	2.52 ± 0.17 (2.45)	0.19	1.0	1.0	0.0
ester O	3.29 ± −	3.23 ± 0.15 (3.25)	−0.06	0.5	0.43	−0.07

[a] All distances in Å. The 22−21−19 truncation was used. [b] Peak positions were determined from the average distance to all the water O atoms within 3.5 Å and the hydration number is the sum over all water oxygens within 3.5 Å. [c] The dagger (†) and asterisk (*) indicate the use of 5 and 8 Å for the averaging and hydration number determination. Values in parentheses adjacent to the MD average positions are the peak positions from the calculated radial distribution functions. If no rms fluctuation is given only 2 or fewer data points were available from the crystal structure.

34 of the supporting information. In the minimization the guanine pseudorotation angles decreased, and an increase occurred for the cytosine sugar moieties. In the dynamics simulations the guanine pseudorotation angles remained low, while the cytosine values shifted in a manner to yield better agreement with experiment as compared with the minimized results. The pseudorotation amplitudes of both the G and C values decreased upon minimization, while during the dynamics simulation the values increased to be larger than the experimental values. It should be noted that the fluctuations in the dynamics are very large because of the small energy differences involved.[42] The trend of the cytosine pseudorotation amplitude being larger than the guanine value is maintained in the calculations. As with the backbone dihedral angles, no significant changes were observed in the sugar puckering. In all cases the calculated pseudorotation angles are in the same minima as the experimental values. The values are systematically closer to zero, but the energy differences are very small (less than *kT*).

Table 7 compares a number of the nucleic acid−water interaction distances, along with the sodium to water and sodium to phosphate distances. It should be noted that the cutoff for the distance averaging and hydration number determination is somewhat arbitrary (see Table 7 legend) and influences the reported values, as discussed in Section II. All interactions within a selected distance were used to obtain the average distances and hydration numbers. A cutoff of 3.5 Å was chosen to ensure that the first hydration layer observed experimentally was included; in certain instances the cutoff was increased to include at least one of the identified interactions in the experimental structure. First peak positions from the simulation radial distribution functions are also included in Table 7. The presence of symmetry in the crystal leads to each site being represented only once for the bases and sodium ion and twice for the phosphate oxygens. Thus, the number of each type of interaction is limited. However, the use of the full unit cell in the simulations increases the number of solute−solvent interactions available for analysis.

The simulation satisfactorily reproduced both the interaction distances and the hydration numbers. The relationship between the anionic and ester phosphate oxygens is maintained for both the interaction distance and hydration number, supporting the charge distribution developed for dimethyl phosphate. For the bases the calculated interaction average distances tend to be longer than the X-ray values. However, as expected (see Methods), the agreement between the peak values from the

radial distribution functions is generally better. For the sugar to water interactions the agreement for both the distances and the hydration numbers is satisfactory. The use of 3.5-Å cutoff for the average interaction distances from the simulations tends to increase the values due to excursions by water molecules not in the first hydration layer within the 3.5-Å cutoff. For example, if the cutoff distance is decreased to 3.0 Å, the average distance will be shorter. The agreement of the base hydration numbers is also satisfactory. The largest discrepancy occurs at the guanine N2 site; it is associated with the decrease in the N3 hydration number. Since those sites are adjacent, the surrounding water molecules are shared by both and contibute to the individual hydration numbers.

The interactions of sodium with water and with the phosphate oxygens also were analyzed in Table 7. The sodium to water and sodium to phosphate distances and hydration numbers are reproduced well in the simulation. The octahedral coordination of the sodium, which includes four water molecules and an oxygen from two different phosphate groups, is maintained. The overall good agreement emphasizes the balance between solute–solvent and solute–solute interactions in the present parameter set.

The present hydration results may be compared with a Monte Carlo simulation of the hydration of GpC in solution using AMBER 3.0 with the TIP4P water model.[117] That study used a single rigid GpC duplex and somewhat different criteria for defining the hydration numbers. Hydration of the phosphate favored the anionic oxygens versus the ester oxygens, with values of 4.9 and 1.4, respectively; this differs from the present study. More generally, base hydration was lower in ref 117 than in the present study, with the exception of CytN4 where a hydration number of 1.0 (sum of N4 and amino hydrogen water coordination numbers) was reported in contrast to the value of 0.33 in the present study. Hydration of the hydroxyls was slightly lower for the O2′ and O3′ sites and higher for the O5′ site. Although the comparison is limited by differences in the simulation and analysis protocols, the two sets of results indicate that there may be a significant effect of the parametrization on the calculated hydration properties; the rigid/flexible difference for GpC may also play a role.

(b) B-DNA Decamer Crystal Simulation. The B-DNA decamer was simulated for 100 ps in an NVE ensemble following an initial heating and equilibration stage of 5 and 5 ps, respectively. Parts A, B, and C of Figure 11 show the total energy, the potential energy, and the rms difference of the DNA heavy atom positions versus the crystal structure, respectively, as a function of time. Both the total and potential energies of the system increased during the heating stage of the simulation. The potential energy then relaxed over the next 20 ps prior to fluctuating around a constant value. Discrete jumps in the total energy of the system were observed during the equilibration stage of the simulation followed by a steady value during the remainder of the simulation. Energy was well conserved throughout the simulation; the total energy increased from −8755.8 to −8752.5 kcal/mol. The average temperature of the simulations was 302.5 ± 3.0 K. Relaxation of the rms difference with respect to the crystal also occurred during the initial 20 ps of the simulation; the value then fluctuated about 0.8 Å. The magnitude of the difference is within the error of the X-ray structure at 1.4-Å resolution.

The rms differences of the 10- to 110-ps time-averaged structure versus the X-ray structure are shown in Figure 12. The values have been broken down into the contributions of the bases, the sugars, and the phosphate moieties. Residue numbering is from the 5′ to 3′ end of strand 1 followed by strand 2, such that residue 1 is base paired with 20, 2 with 19, and so on. The rms difference for all heavy atoms in the time-averaged structure is 0.94 Å. This value differs from that in Figure 9 due to distortions associated with the averaging procedure; i.e., motions of the planar bases upon averaging yield atom positions that are shifted toward the center of the rings. There is no

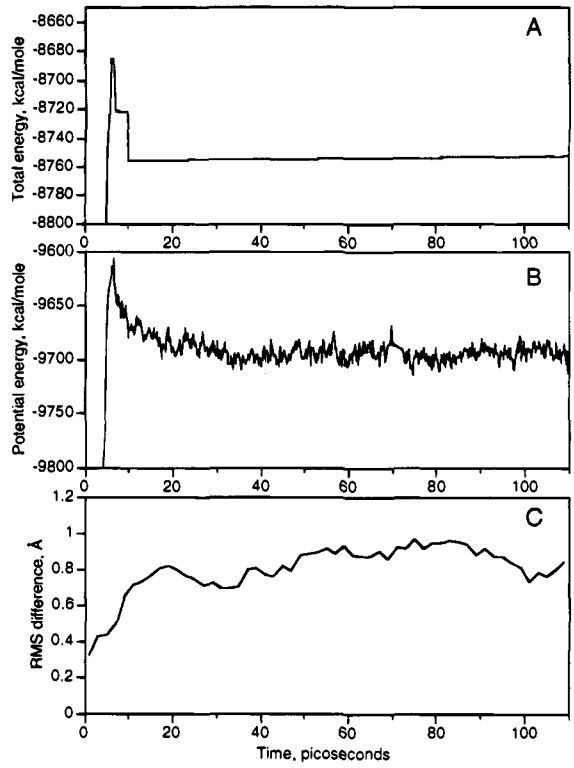

Figure 11. Total energy (A), potential energy (B), and rms difference of all DNA heavy atoms versus the crystal structure (C) as a function of time for the B-DNA decamer crystal simulation. Presented energies were averaged over 0.2 ps, the rms differences is for time-averaged coordinates averaged over 2 ps windows, and the heavy atoms were least-squares fit to the X-ray atoms prior to determination of the rms difference.

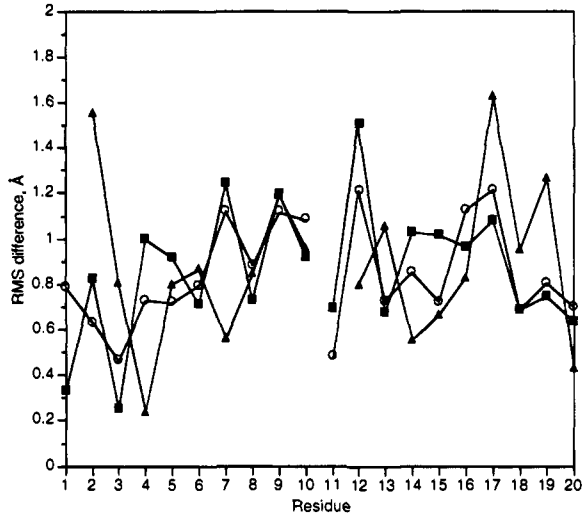

Figure 12. The rms differences of the B-DNA time-averaged simulation structure versus the X-ray structure for the base (■), phosphate (▲), and sugar (○) heavy atoms. Phosphate atoms included the phosphorus and 2 anionic oxygen atoms and sugar atoms include the 5 ring atoms and the 2 ester oxygens. Heavy atoms were least-squares fit to the X-ray atoms prior to determination of the rms difference.

(117) Subramanian, P. S.; Pitchumani, S.; Beveridge, D. L.; Berman, H. M. *Biopolymers* **1990**, *29*, 771.

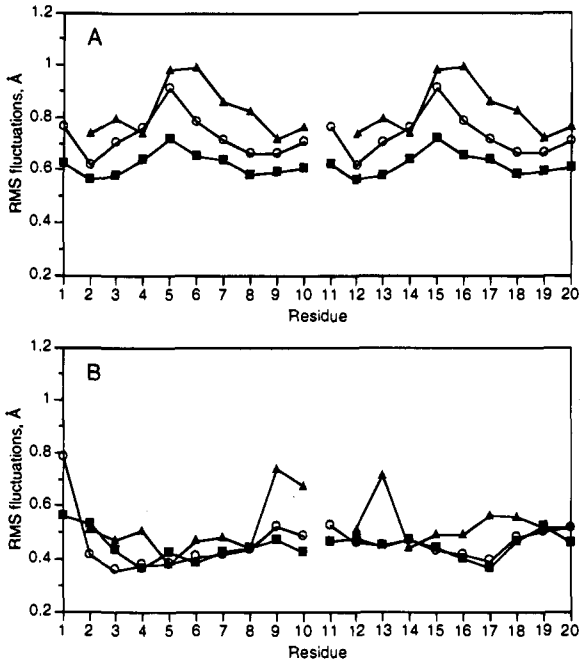

Figure 13. The rms fluctuations from the B-DNA X-ray temperature factors (A) and from the MD simulation (B) for the base (■), phosphate (▲), and sugar (○) heavy atoms. Phosphate atoms included the phosphorus and 2 anionic oxygen atoms and sugar atoms include the 5 ring atoms and the 2 ester oxygens.

distinct pattern concerning the observed differences, with large residue to residue changes occurring along the DNA strand. For the bases a larger change in strand 1 is accompanied by a larger change in the adjacent base in strand 2, as expected due to the maintainence of the Watson–Crick base pairing. Such correlation does not appear to occur for the sugar or phosphate moieties. This suggests that the backbone atoms of the individual DNA strands move independently. Comparison of the rms differences of the individual residues shows that the base and sugar differences are correlated. In contrast, little correlation between the phosphate and the bases or sugar is observed; i.e., the phosphates move independently of the surrounding sugars and bases. Such motions are allowed due to anticorrelated variations about the β and ϵ dihedral angles.

Root-mean-square fluctuations estimated from the X-ray temperature factors and calculated from the molecular dynamics simulation are shown in Figure 13, parts A and B, respectively. In the simulation the fluctuations of the bases and the sugars are similar and the phosphate fluctuations are generally larger. This is in reasonable agreement with the X-ray results, though the differences are larger in the experimental results. The simulation indicates more flexibility at the termini of the helix with lower amplitude motion toward the center; the X-ray results also show more flexibility at the termini along with more flexibilty toward the center of the helix. The difference in magnitude between the calculated and experimental results may be due to the short duration of the simulation, as well as the presence of lattice disorder in the crystal which is not present in the simulation. This is supported by the overall greater mobility in the X-ray study as compared with the simulation.

Comparison of the X-ray and dynamics averaged Watson–Crick base-pairing interaction distances is presented in Table 8. Since a single strand of DNA comprises the asymmetric unit the first 5 and second 5 base pairs in the generated DNA duplex are identical; this symmetry is lost in the simulation. Thus, there are two calculated results for each experimental point. This indicates that the simulations are not fully converged. Generally, the interaction distances from the simulations are slightly longer than the X-ray results, with the majority of the differences being less than 0.1 Å. Specifically, the C-G N4 to O6 interaction tends to be too long. The fluctuations associated with this interaction are the largest in Table 8 and may be contributing to the difference. Also, the magnitude of the difference between the X-ray and calculated results is smaller than the rms fluctuations of the distances. The corresponding interaction was well reproduced in the GpC crystal simulation (Table 6). In contrast, the A-T N6 to O4 interaction is consistently shorter than the X-ray value. This again may be associated with the fluctuations, which are larger than those for the adjacent N1 to N3 distance. It should be noted that the B-DNA simulation was performed with different angle parameters associated with the adenine N6 atom than those reported in the Appendix. This difference, since corrected, led to the N1–C6–N6 angle being 116° in the minimized adenine base, rather than the values of 120.1° reported in Table 4 of the supporting information. The previously smaller N1–C6–N6 angle tended to move the amine group closer to the thymine O4 atom in the Watson–Crick base pair and is likely to be responsible for the short calculated distance. This problem emphasizes the importance of self-consistency between the internal and external empirical terms. The corrected parameters, as shown in the Appendix, were used for all results in this paper, excluding the B-DNA simulation, which was not redone because of the large amount of additional computer time that would have been required.

Table 8. Watson–Crick Base Pair Interactions from the B-DNA Decamer Crystal and the MD Simulation[a]

	distances		
residue	crystal	dynamics[b]	diff
Cyt1 to Gua10			
N4 to O6	2.90	3.10 ± 0.24, 3.08 ± 0.19	0.20, 0.18
N3 to N1	2.93	2.97 ± 0.11, 3.02 ± 0.12	0.04, 0.09
O2 to N2	2.82	2.88 ± 0.17, 2.94 ± 0.16	0.06, 0.12
Cyt 2 to Gua 9			
N4 to O6	2.88	3.02 ± 0.19, 2.93 ± 0.15	0.14, 0.05
N3 to N1	2.96	2.97 ± 0.10, 2.94 ± 0.09	0.01, −0.02
O2 to N2	2.89	2.85 ± 0.14, 2.90 ± 0.14	−0.04, 0.01
Ade 3 to Thy 8			
N1 to N3	2.80	2.89 ± 0.10, 2.95 ± 0.11	0.09, 0.15
N6 to O4	3.00	2.89 ± 0.16, 2.90 ± 0.16	−0.11, −0.10
Ade 4 to Thy 7			
N1 to N3	2.84	2.90 ± 0.10, 2.87 ± 0.08	0.06, 0.03
N6 to O4	3.15	3.01 ± 0.19, 2.92 ± 0.15	−0.14, −0.23
Cyt 5 to Gua 6			
N4 to O6	2.86	3.05 ± 0.20, 2.98 ± 0.17	0.19, 0.12
N3 to N1	2.94	2.99 ± 0.11, 2.98 ± 0.10	0.05, 0.04
O2 to N2	2.83	2.86 ± 0.13, 2.89 ± 0.13	0.03, 0.06

[a] All distances in Å. Experimental values from ref 44. [b] In the crystal the first 5 and the second 5 base pairs are equivalent by symmetry. Due to the use of 2 asymmetric units comprising the full DNA duplex, the symmetry was lost in the simulation and the values from the first and second 5 base pairs are given. Errors are rms fluctuations from the simulations.

The overall configuration of DNA can be described by the backbone dihedral angles and the sugar puckering (see above for the definitions). The X-ray and dynamics averaged values for the dihedral angles and differences between the measured and calculated values are presented in Tables 9 and 10, respectively. The overall rms difference between the experimental and calculated backbone dihedrals is 24°. Analysis of Table 9 shows this number arises from several particularly large values, with the agreement being significantly better for the majority of the dihedral angles. The largest difference occurs for the Cyt1 γ dihedral which flips from 55° to −57 or −40°. This dihedral angle concerns the ester linkage from sugar 1 to

Table 9. Backbone Dihedral Angles from the B-DNA Dodecamer Crystal and Dynamics Simulation[a]

		χ	α	β	γ	δ	ϵ	ζ
					Residue Crystal			
Cyt1		−121			55	127	−158	−92
Cyt2		−83	−74	177	52	140	−110	167
Ade3		−84	−67	147	51	145	−180	−88
Ade4		−95	−71	173	51	126	−161	−124
Cyt5		−117	−45	147	51	114	−167	−126
Gua6		−112	−44	167	42	141	−177	−94
Thy7		−123	−72	177	62	107	−170	−94
Thy8		−92	−55	179	43	140	−97	166
Gua9		−80	−71	150	45	144	−175	−94
Gua10		−102	−67	168	51	112		
					Dynamics			
Cyt1	1	−130 ± 16			−57 ± 41	150 ± 10	−128 ± 16	−80 ± 11
	2	−126 ± 18			−40 ± 44	139 ± 13	−106 ± 13	−126 ± 19
Cyt2	1	−74 ± 12	−88 ± 13	151 ± 11	57 ± 10	134 ± 7	−118 ± 15	166 ± 12
	2	−76 ± 8	−75 ± 11	129 ± 10	69 ± 7	144 ± 7	−126 ± 18	−174 ± 17
Ade3	1	−105 ± 10	−69 ± 10	127 ± 8	65 ± 8	142 ± 11	−165 ± 10	−100 ± 10
	2	−110 ± 11	−69 ± 14	135 ± 11	63 ± 8	139 ± 8	−160 ± 8	−97 ± 9
Ade4	1	−114 ± 10	−69 ± 10	163 ± 11	55 ± 8	117 ± 11	−168 ± 8	−97 ± 10
	2	−130 ± 13	−70 ± 9	154 ± 9	59 ± 8	101 ± 13	−168 ± 8	−90 ± 11
Cyt5	1	−107 ± 9	−60 ± 8	156 ± 9	56 ± 8	139 ± 6	−105 ± 17	175 ± 14
	2	−111 ± 9	−65 ± 9	163 ± 8	57 ± 6	141 ± 9	−153 ± 9	−159 ± 11
Gua6	1	−137 ± 11	−69 ± 11	134 ± 11	53 ± 8	130 ± 11	−162 ± 10	−99 ± 10
	2	−162 ± 7	−60 ± 10	137 ± 10	59 ± 7	88 ± 6	−150 ± 8	−77 ± 9
Thy7	1	−119 ± 12	−70 ± 9	175 ± 14	54 ± 9	123 ± 13	−168 ± 11	−102 ± 14
	2	−125 ± 11	−60 ± 10	170 ± 10	64 ± 7	137 ± 8	−160 ± 13	−103 ± 14
Thy8	1	−103 ± 10	−63 ± 10	169 ± 11	51 ± 8	138 ± 7	−117 ± 24	−175 ± 14
	2	−97 ± 11	−65 ± 11	169 ± 15	52 ± 8	141 ± 7	−123 ± 19	−174 ± 16
Gua9	1	−123 ± 19	−62 ± 14	127 ± 12	58 ± 8	112 ± 16	−129 ± 18	−110 ± 23
	2	−103 ± 17	−62 ± 11	136 ± 14	60 ± 8	137 ± 10	−150 ± 15	−94 ± 10
Gua10	1	−111 ± 8	−74 ± 11	147 ± 16	56 ± 8	88 ± 8		
	2	−111 ± 10	−78 ± 11	157 ± 10	53 ± 9	88 ± 10		
					Difference: Strand 1, Strand 2			
Cyt1		−9, 5			−111, −95	23, 13	30, 52	12, −34
Cyt2		9, 7	−14, 0	−26, −48	6, 18	−6, 4	−8, −17	−1, 19
Ade3		−20, −25	−1, −2	−20, −12	14, 12	−2, −6	15, 20	−12, −9
Ade4		−19, −36	2, 1	−10, −19	4, 8	−9, −25	−7, −7	27, 34
Cyt5		10, 5	−15, −20	10, 16	5, 6	25, 27	61, 14	−60, −33
Gua6		−25, −50	−25, −16	−33, −30	11, 16	−11, −53	15, 27	−4, 18
Thy7		4, −2	3, 13	−2, −7	−8, 2	16, 30	2, 10	−8, −9
Thy8		−11, −5	−8, −10	−9, −10	8, 9	−2, 0	−20, 26	19, 20
Gua9		−42, −23	9, 9	−23, −15	13, 15	−32, −7	46, 25	−15, 0
Gua10		−10, −10	−8, −11	−21, −11	5, 2	−24, −24		

[a] All angles in deg. The rms fluctuations from the simulations are given. [b] Experimental values from ref 44.

the phosphate of the second residue, where large deviations of the β dihedral are also observed. The γ and β differences tend to act in an anticorrelated fashion so that the change in the position of the second and twelfth phosphate groups is less than 1.5 Å (Figure 12). Additionally, the largest rms fluctuations occur in the terminal γ dihedral angle. Analysis of the calculated values of χ, the glycosidic linkage angle, shows good agreement for the pyrimidines. However, there is a trend in the purines for χ to decrease; calculated values differ from the experimental values by −10 to −42°. In all cases, the dihedral angles maintain the *anti* configuration. These changes are not associated with a rotation of the base itself, as indicated by the maintenance of the Watson−Crick base pairing, but rather are due to changes in the sugar conformation (Table 10). In the purines the sugar pseudorotation angles shift toward the C3′-endo conformation from the C2′-endo region of the pseudorotation surface typically observed in B-DNA. Such a shift is not unexpected as the C3′-endo sugar pucker is often observed in purines.[115] Additionally, the shift in χ toward more negative values in the purines is consistent with previous observations, where χ values between −138° and −180° are associated with a C3′-endo sugar pucker.[115]

The largest differences for the backbone dihedral angles occur at the termini and the middle of the DNA strand. Plotted in Figure 14 are the rms differences of the dihedrals as a function of the base; the rms difference includes all values reported in the difference section of Table 9. The only exception to the trend is at the terminal guanine. The trend is similar to that observed in the rms fluctuation versus residue plot for the X-ray data shown in Figure 13A. This similarity suggests the possibility that the experimental rms fluctuations are related to conformational substates occurring in the duplex that are not sampled in the present simulation. Support for this is provided by the lower amplitude rms fluctuations observed in the calculated results as compared with experiment (Figure 13). The calculated rms fluctuations may be a consequence of the region of conformational space that is sampled rather than being determined by the empirical potential energy function. This could explain the better agreement between the X-ray and calculated backbone dihedral angle and sugar pucker parameters in the GpC than the B-DNA simulations; the conformation space for GpC is likely to be much more restricted than that for B-DNA. Further study of this phenomenon requires either drastically increasing the duration of the simulation or running a number of simulations differing in the initial conditions that direct the simulations to different regions of conformational space (Caves *et al.*, to be published).

Sugar puckering is known to be related to DNA helix type.[115]

Table 10. Sugar Pseudorotation Angle and Amplitude in the B-DNA Crystal and the Dynamics Simulation[a,b]

	exp	molecular dynamics	
		strand 1	strand 2
		Pseudorotation Angle	
Cyt1	151.7	122.2 ± 106.2 (−29.5)	124.8 ± 84.8 (−26.9)
Cyt2	153.1	138.2 ± 8.5 (−14.9)	151.2 ± 20.3 (−1.9)
Ade3	185.4	140.1 ± 80.8 (−45.3)	142.9 ± 64.4 (−42.5)
Ade4	138.8	122.0 ± 16.3 (−16.8)	98.2 ± 21.3 (−40.6)
Cyt5	124.6	142.3 ± 7.4 (17.7)	147.1 ± 8.3 (22.5)
Gua6	171.1	137.6 ± 23.0 (−33.5)	5.1 ± 9.5 (−166.0)
Thy7	117.7	131.0 ± 22.7 (13.3)	149.5 ± 50.6 (31.8)
Thy8	151.9	144.1 ± 9.7 (−7.8)	146.0 ± 27.0 (−5.9)
Gua9	180.4	98.6 ± 37.4 (−81.8)	95.4 ± 121.9 (−85.0)
Gua10	117.0	28.9 ± 22.8 (−88.1)	61.3 ± 21.6 (−55.7)
		Pseudorotation Amplitude	
Cyt1	30.9	40.6 ± 5.0 (9.7)	36.3 ± 5.9 (5.4)
Cyt2	43.2	45.7 ± 4.3 (2.5)	44.1 ± 4.0 (0.9)
Ade3	36.0	37.4 ± 4.9 (1.4)	36.8 ± 4.5 (0.8)
Ade4	41.4	41.9 ± 5.2 (0.5)	43.9 ± 4.9 (2.5)
Cyt5	46.7	47.7 ± 3.7 (1.0)	45.0 ± 4.1 (−1.7)
Gua6	39.6	38.9 ± 5.0 (−0.7)	38.5 ± 5.1 (−1.1)
Thy7	42.1	38.9 ± 5.7 (−3.2)	32.6 ± 6.2 (−9.5)
Thy8	44.1	43.8 ± 4.2 (−0.3)	42.7 ± 4.9 (−1.4)
Gua9	35.2	31.9 ± 8.6 (−3.3)	33.2 ± 5.7 (−2.0)
Gua10	34.8	34.9 ± 6.9 (0.1)	40.0 ± 6.0 (5.2)

[a] See footnote of Table 28 (supporting information). [b] The values in parentheses are the difference between the calculated and experimental values.

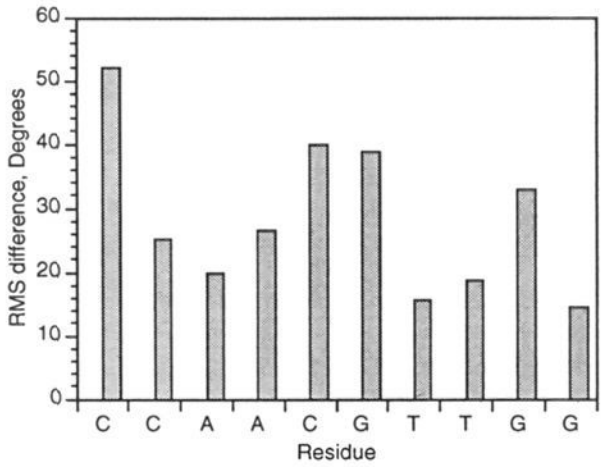

Figure 14. The rms difference between the X-ray and B-DNA dynamics averaged backbone dihedrals versus base number. The rms differences were obtained from the difference values reported in Table 9.

In B-DNA the distribution of sugar pseudorotation angles is broad, varying from C2′-endo (P = 144°) to C3′-endo (P = 15°).[44] In the present calculations the averge pseudorotation values span this range with values from 5 to 150° for Gua6, strand 2 and Thy7, strand 2 (Table 10). For the pseudorotation amplitude, the agreement between experiment and the dynamics averages is good. In all cases, excluding Cyt1, strand 1 and Thy7, strand 2, the differences are smaller than the rms fluctuation obtained from the dynamics simulation.

The relationship between the sugar pucker and the glycosidic linkage angle may be examined by analyzing the correlation between the two terms. Parts A and B of Figure 15 present plots of δ versus χ for the X-ray and the dynamics average values, respectively. The angle δ is substituted for the pseudorotation angle in accord with previous studies.[109,111] In the X-ray structure the points are scattered about δ, χ values of 150°, −110°. Shown in the figure is the least-squares fitted line from which a slope of 0.39, a y-intercept of −158°, and a correlation coefficient of 0.33 were determined. Similar relationships between δ and χ have been observed in a number of other B-DNA structures.[115,118] Results from the dynamics run in Figure 15B show a general decrease in both the δ and the χ

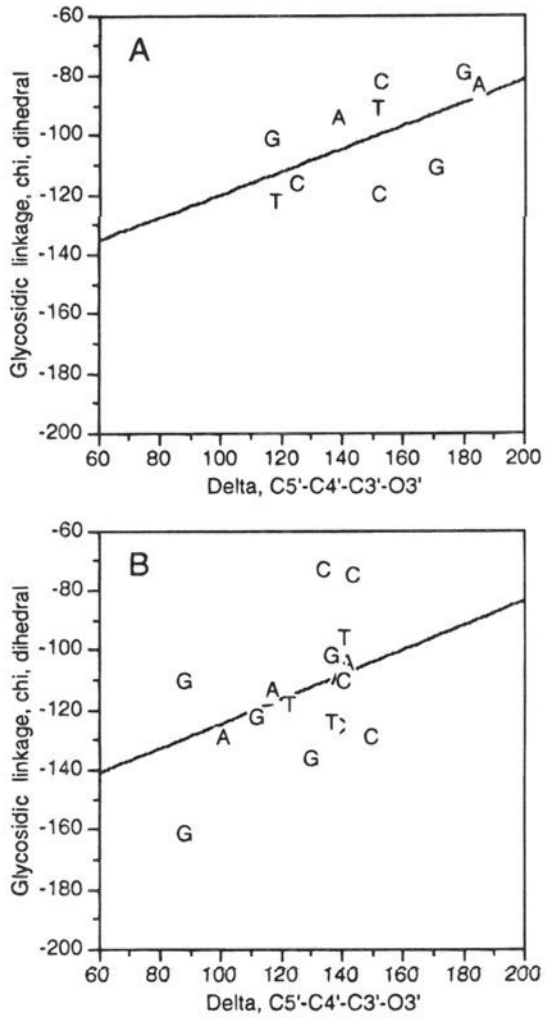

Figure 15. Scatter plot between the backbone dihedrals δ and χ from the X-ray structure (A) and from the dynamics averages (B). The line represents the least-squares fit to the presented points.

values. Nevertheless, both the slope and y-intercept of 0.41 and −166°, respectively, are comparable to the experimental result (dynamics correlation coefficient of 0.17). The shift in δ is greatest in the purines and is associated with the changes in the pseudorotation angle discussed above. Thus, the known correlation between the glycosidic linkage angle and sugar pucker angles[115,118] is reproduced by the empirical force field, even though the actual values of the angles deviate from experiment.

Correlations have also been observed between the backbone dihedral angles δ and ϵ and ζ and δ.[118] Figures 16 and 17 present δ versus ϵ and ζ versus δ plots, respectively, for the X-ray structure and the dynamics averages structure; neither correlation is very precise. The slope and y-intercept of a least-squares fit to the base are comparable (X-ray: slope = 0.49, y-intercept = −219°, r^2 = 0.05; dynamics: slope = 0.51, y-intercept = −209°, r^2 = 0.14). The larger differences are observed at the termini and the middle of the helix, as discussed above. For the ζ versus δ dihedral angles the X-ray structure shows two overlapping points in the region of −190°, 140°, with the remaining points scattered about −100°, 130°. A similar trend is observed in the dynamics run. Thus, the empirical energy function reproduces the dihedral correlations observed between base type, sugar pucker, and backbone dihedrals in a variety of B-DNA structures.

Hydration properties of B-DNA and the Mg^{2+} ions are presented in Table 11 for the X-ray structure and from averages over the dynamics simulation. Solvation of the phosphate showed the anionic oxygens to be more favorably solvated than the ester oxygens. This trend is maintained in the simulation. The dynamics average distances are slightly shorter than the X-ray values, with the difference being significantly smaller

(118) Dickerson, R. E.; Kopka, M. L.; Drew, H. R. Structural Correlations in B-DNA. In *Structre and Dynamics: Nucleic Acids and Proteins*; Clement, E., Sarma, R. H., Eds.; Adenine Press: New York, 1982.

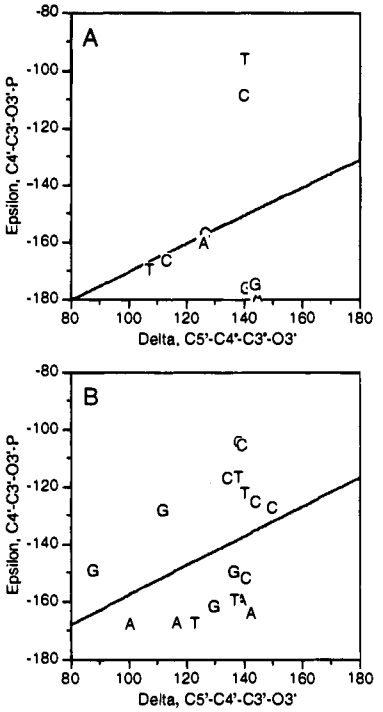

Figure 16. Scatter plot between the backbone dihedrals δ and ϵ from the X-ray structure (A) and from the dynamics averages (B). The line represents the least-squares fit to the presented points.

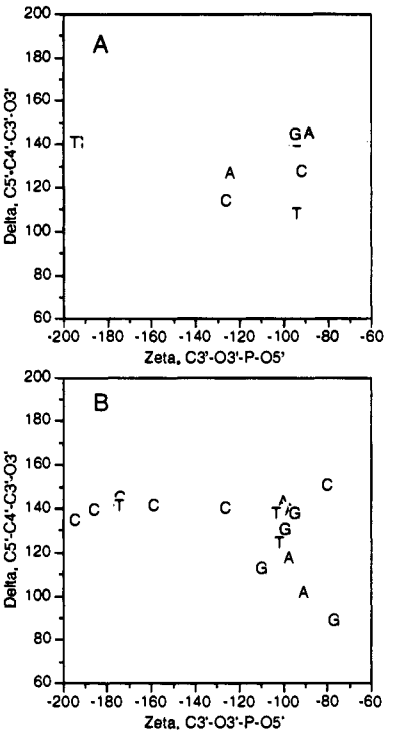

Figure 17. Scatter plot between the backbone dihedrals ζ and δ from the X-ray structure (A) and from the dynamics averages (B).

than the rms fluctuations. Both the anionic and ester oxygen hydration numbers increase slightly in the simulation. However, the use of the 3.5-Å cutoff in the determination of the distances and hydration numbers influences the calculated number (see Methods). Parts A and B of Figure 18 present the anionic O to water O and ester O to water O radial distribution functions, respectively. Both the X-ray and experimental results are shown in the figures. The curves are in very good agreement, although the experimental values correspond to relatively few discrete points. For the anionic oxygen the first peaks coincide, although the calculated peak is shifted slightly inward as compared to the experimental curve. A similar effect is observed when the TIP3P O−O radial distribution function is compared with experiment.[32] This difference appears to be required in order to obtain the correct solvent density and is associated with the lack of many-body polarization effects in the empirical force field. Comparison of the X-ray and dynamics ester oxygen to water oxygen radial distribution function (Figure 18B) shows the first peak to be situated 0.4 Å in from the experimental peak. This difference is larger than that observed for the anionic oxygen and still occurs even though the ester O to water monomer interaction energy is somewhat less favorable than the anionic interaction (Table 5). Considering the data related to the ester O solvation in the GpC crystal (Table 7), no further optimization of those parameters was performed. In both the anionic and ester oxygen radial distribution functions the agreement with the X-ray data for the second peak is also quite good for both position and magnitude. Again, this comparison is limited by the number of experimental points available, but supports the accuracy of charge distribution used in the phosphodiester linkage.

For the base heteroatom hydrogen bonding sites the dynamics averaged values tend to be longer than the experimentally observed values, while the dynamics hydration numbers overestimate the X-ray values. For the experimental results the average distance represents the first hydration layer, as supported by the hydration numbers being 1.0 or 0.5; this corresponds to one or two waters for the two sites. On the other hand, the calculated values include contributions from waters moving into the 3.5-Å cutoff, as represented by the hydration numbers being larger than 1. Problems associated with the 3.5-Å cutoff are not present in the X-ray data due to a water molecule being either inside or outside the cutoff distance. Since both the average distances and hydration numbers are dependent on the cutoff distance used in the averaging procedure the peak positions of the calculated radial distribution functions are reported in Table 11. These peak values, calculated to a resolution of 0.1 Å, are generally in good agreement with the X-ray average value. A detailed comparison is not possible because of the small number of experimental data points. It appears that the present empirical force field adequately represents the solvation of both the phophate moiety and the bases in B-DNA. The proper base hydration is of special interest for in-depth studies of the spines of hydration important for the structure and function of DNA.[115] There is a spine of waters present in both the major and the minor groove; those waters, in some instances, interact with the phosphates as well as the bases.

Three sites on the sugars are available for hydrogen bonding to solvent; the O4′ ether oxygen in the ring and the O3′ and O5′ hydroxyl groups. The latter sites are only available at the termini of the DNA strands where they are not involved in a phosphodiester linkage. For the O3′ and the O4′ sites the agreement with experiment is good: both the calculated interaction distances and hydration numbers reflect experiment. For the O5′ site, however, there is a significant difference, in that no water molecules in the region of the O5′ are identified in the crystal structure. However, only about 50% of the waters are observed in the crystal (G. Privé, private communication). In this case, it may be associated with the mobility of the atom; i.e., the rms fluctuations of O5′ are 0.93 and 1.09 Å on strands 1 and 2, respectively. This can lead to disordered water neighbors hindering strong interactions with the O5′ atom. In the simulation, water molecules move into the vicinity of the O5′ atom, forming hydrogens bonds similar to those at the O3′

Table 11. B-DNA Water Interactions from the Crystal and the MD Simulation[a,b]

site	average position, Å			hydration number		
	crystal	MD[d]	diff	crystal	MD	diff
phosphate						
anionic O	2.92 ± 0.26	2.85 ± 0.26 (2.65)	−0.07	2.11	2.97	0.86
ester O	3.26 ± 0.11	3.14 ± 0.23 (2.95)	−0.12	0.39	0.78	0.39
bases						
Cyt N4	3.03 ± 0.17	3.18 ± 0.20 (3.05)	0.15	1.0	1.42	0.42
Cyt O2	2.82 ± 0.18	3.01 ± 0.26 (2.75)	0.19	1.0	1.16	0.16
Gua N2	3.22 ± 0.28	3.13 ± 0.20 (2.95)	−0.09	1.0	1.43	0.43
Gua N3	2.96 ± 0.19	3.13 ± 0.22 (2.95)	0.17	1.0	1.04	0.04
Gua N7	2.64 ± 0.10	3.02 ± 0.23 (2.85)	0.38	1.0	1.66	0.66
Gua O6	3.01 ± 0.24	3.04 ± 0.26 (2.75)	0.03	1.0	1.32	0.32
Ade N3	2.79 ± 0.06	3.10 ± 0.21 (2.85)	0.31	1.0	1.27	0.27
Ade N6	2.78 ± −	3.08 ± 0.21 (2.85)	0.30	0.5	1.23	0.73
Ade N7	2.97 ± 0.35	3.12 ± 0.22 (2.95)	0.15	1.0	1.34	0.34
Thy O2	2.61 ± 0.10	3.05 ± 0.27 (2.75)	0.44	1.0	1.43	0.43
Thy O4	2.71 ± 0.13	3.02 ± 0.25 (2.85)	0.31	1.0	1.00	0.0
sugar						
O3′	2.88 ± 0.19	2.98 ± 0.24 (2.85)	0.10	2.0	2.52	0.52
O4′	3.05 ± 0.27	3.08 ± 0.23 (2.85)	0.03	0.8	1.09	0.29
O5′		2.99 ± 0.25 (2.75)			1.82	
Mg^{2+} to water O	2.10 ± 0.02	2.04 ± 0.17 (2.05)	0.06	5.43	6.08	0.65
Mg^{2+} to phosphate[c]						
anionic O	6.38 ± 1.06	6.15 ± 1.28	−0.23			
ester O	6.67 ± 0.77	6.53 ± 0.81	−0.14			

[a] See footnote of Table 26 (supporting information). [b] Peak positions were determined from the average distance to all the water O atoms within 3.5 Å and the hydration number is the sum over all water oxygens within 3.5 Å. [c] Average distance determined over all distances 8 Å or less. [d] Values in parentheses adjacent to the molecular dynamics average positions are the peak positions from the calculated radial distribution functions.

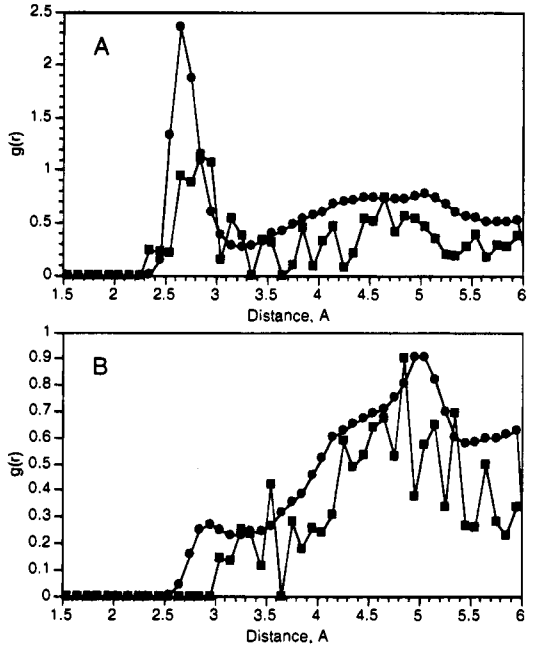

Figure 18. Radial distribution functions of the water oxygens to the phosphate (A) anionic and (B) ester oxygens from the BDNA X-ray structure (■) and the dynamics simulation (●).

atom. The lower O5′ versus O3′ hydration number indicates a greater mobility of the adjacent solvent.

Table 11 also includes information on the hydration of Mg^{2+} and the Mg^{2+} to phosphate interactions. The interaction distance between Mg^{2+} and water is well reproduced by the potential energy function, with the calculated values being 0.06 Å shorter than the X-ray value. Further, the hydration number, associated with the octahedral coordination of the Mg^{2+} ion, is well reproduced. To check for significant changes in the Mg^{2+} to phosphate spatial relationships the average distances between the anionic and ester phosphate oxygens were determined for all interactions less distant than 8.0 Å. The calculations reproduce the X-ray data, with a slight shortening of the average interaction distances. Importantly, the difference between the anionic O to Mg^{2+} and the ester O to Mg^{2+} distances is maintained. Thus, it appears that the balance between Mg^{2+} to water and Mg^{2+} to phosphate interactions is represented by the present parameter set.[119]

Results from the present simulation of the B-DNA decamer can be compared with a number of published B-DNA simulations. However, the comparisons are limited because previous studies were not performed in a crystal, but rather in vacuum or in the presence of water and counterions. Levitt performed a simulation in vacuum on the d(CGCGAATTCGCG) duplex in both the presence and absence of electrostatic interactions.[16] Results showed the simulation without electrostatic contributions to be stable with large excursions from the average structure. In the presence of electrostatic interactions the helix was observed to unwind. A vacuum simulation of a B-DNA pentamer using the extended atom AMBER parameters[4] and in the presence of an extended cation representing a hexacoordinated sodium[120] reported rms differences of 0.8 to greater than 2 Å with respect to the initial structure in the dynamics. Other reported results by Singh *et al.*[120] include rms fluctuations of the Watson−Crick hydrogen bonds of 0.16 to 0.70 Å, of the backbone dihedral angles of 9 to 21°, and of the sugar pucker amplitude of 0.05 to 0.08 Å. Another study of a B-DNA d(CGCGAATTCGCG) duplex was performed in vacuum also using the AMBER extended atom parameters with an extended cation.[24] That study used a 100-ps simulation initiated from the canonical B-DNA structure; a 3.2 Å rms difference with respect to the starting structure was observed. With the same AMBER parameter set, Siebel *et al.* repeated the B-DNA pentamer simulation (see above) in the presence of a sphere of 830 water molecules and 8 sodium ions.[19] In the solvated simulation, a decrease in the fluctuations of the phosphodiester dihedral ω' was observed (from ±20 to ±12°) while the fluctuations in the sugar pucker amplitude were similar (0.05−0.07 Å). No data on rms differences from the starting canonical

(119) Prod'hom, B.; Karplus, M. Manuscript in preparation.
(120) Singh, U. C.; Weiner, S. J.; Kollman. P. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 755−759.

B-DNA structure were reported. Solvation properties from the simulation yield a sodium to oxygen distance of 2.5−2.6 Å, an average hydration number of 3.0 for the anionic phosphates with an average distance of 2.7−2.8 Å, an average hydration number of 1.7 and 0.9 for the phosphate ester oxygen O3′ and O5′ atoms, respectively, with an average distance of 3.1−3.2 Å, and an average hydration number of 1.1 for the sugar O1′ atom. Solvent to nucleic acid interactions were also reported for the bases; a hydration number and distance of 1.0 and 2.8−2.9 Å were found for the pyrimidine O2 atom, values of 0.9 and 3.1 Å for the purine N3 and 0.7, and 2.9−3.0 Å for the guanine N2 amino group. Larger hydration numbers are reported for the guanine O6 and thymine O4 atoms, ranging from 1.4 to 2.7, with an average interaction distance of 2.9 Å. These hydration properties are qualitatively similar to those obtained from the present parameter set, although there are some significant differences. The hydration number of 0.8 in the present work for the phosphate ester oxygens is lower, the hydration number of the base nitrogen and O2 atoms are larger, and the O6 and O4 hydration numbers are lower. Such differences, although not possible to evaluate statistically, may be associated with differences in the partial charge distributions used in the different parameter sets.

Hydration properties of BDNA were also studied by Swamy and Clementi.[20] In that work the hydration numbers were 2.4 to 2.5 and 0.0 to 0.6 for the anionic and ester phosphate oxygens, respectively. Values of 0.1 to 0.8 were obtained for the base heteroatoms and a value of 0.7 for the O4′ oxygen in the sugar ring. That simulation was only 4 ps, so that there was no exchange with the bulk solvent. Two recent simulations of B-DNA have been published using the GROMOS parameter set.[121] One study simulated a $(dG)_6$-$(dC)_6$ minihelix in the presence of a droplet of 292 water molecules and 10 sodium ions.[122] Results from that study indicate that the B conformation was maintained. However, a detailed analysis of the results was not given. A second study using the GROMOS force field was performed on a d(CGCGAATTCGCG) dodecamer, including 1927 water molecules and 22 sodium ions, using periodic boundary conditions.[123] In that study harmonic constraints of 5 kcal/mol were placed on the Watson−Crick hydrogen bonds. A rms difference of 2.3 Å from canonical B-DNA was obtained, with an overall rms deviation of approximately 45° of the backbone dihedrals. AMBER parameters have been used for 150-ps simulations of the d(CGCGAATTCGCG) dodecamer in solution, including 22 sodium ions.[124] In that study the structure was stable for the first 60 ps of the simulation. However, it deviated significantly from the B-DNA structure in the second half of the simulation; base pair opening and severe distortions of the helical structure were observed. In addition, a wide variety of sugar conformations, including C3′-endo, O4′-endo, O4′-exo, and C1′-endo, were sampled in the simulations. Simulations on the d(CGCGAATTCGCG) dodecamer in solution have also been performed using the GROMOS force field modified to include explicit hydrogen bonding interactions for the Watson−Crick interactions and the phosphate charges decreased to −0.24 eu to account for counterion effects.[27] These simulations have been extended up to 1 ns in duration. Results show rms differences with respect to canonical B-DNA in the range of 4.5 Å; further comparison with the present results is limited due the difference in environment (solution versus crystal). Studies in progress using the present parameter set to simulate the d(CGCGAATTCGCG) dodecamer in solution have yielded stable structures in both the absence and presence of counterions (MacKerell, A. D., Jr., work in progress).

A test simulation of a Z-DNA hexamer crystal[45] has also been performed and compared with earlier work using the CHARMM19 DNA energy function.[6,22] Details of the comparison will be presented elsewhere. However, a brief summary of the results follows as it provides another validation of the present parameter set. The simulation used a protocol similar to that outlined for B-DNA (see Methods); the crystal, solvent, and both sodium and magnesium ions were included. A 110-ps simulation was performed. The rms difference of the 10−110-ps time-averaged structure was 0.59 Å with respect to the X-ray structure. This difference is well within the 1.25-Å resolution of the crystal and represents an improvement over the CHARMM19 rms difference of 0.9 Å. As with the B-DNA simulation, rms fluctuations were smaller than those observed experimentally; however, the decrease relative to experiment was not as large. Watson−Crick base pairing was well maintained with the majority of interaction distances within 0.1 Å of experiment. Larger differences did occur, which were associated with the interaction of a sodium ion with the base pair and with increased rms fluctuations of a second base pair. Comparison of the calculated and experimental backbone dihedral angles showed a total rms difference of 12°. The calculated sugar pseudorotation values stayed in the C2′-endo region for the cytosine residues and in the C3′-endo region for the guanine residues. Correlations between backbone dihedrals observed in the X-ray structure were maintained in the calculations. Hydration properties were also well reproduced by the present parameter set including solvation of the bases, phosphates, and sodium ions. As with the GpC and B-DNA simulations, the present parameters satisfactorily reproduced the internal geometries, hydrogen bond interactions, and solvation properties observed in the X-ray structure.

VI. Conclusion

Parameters for nucleic acids have been developed for an empirical energy function used with the CHARMM program. The model treats all atoms, including hydrogens, explicitly. This is referred to as an all-atom model, in contrast to a model that treats only polar hydrogens explicitly and represents nonpolar hydrogens as parts of extended atoms. The parameters were determined by fitting an extended set of experimental and *ab initio* results. A self-consistent approach was employed to obtain a proper balance between the intramolecular and intermolecular portions of the potential energy function. Emphasis was placed on a balance of the solvent−solvent, solvent−solute, and solute−solute portions of the intermolecular portion of the potential energy. This is essential for accurate condensed phase simulations from which both structural and thermodynamic information can be obtained. A list of the parameter values is included in an Appendix in the supporting information.

Use was made of results for model compounds, including the nucleic acid bases, dimethyl phosphate and anionic and dianionic methyl phosphate, ribose, and deoxyribose. Internal parametrizations (bond length, bond angle, Urey−Bradley, dihedral, and improper dihedral terms) were chosen to reproduce geometries and vibrational spectra from experimental crystal structures, infrared and Raman spectroscopic data, and *ab initio* calculations. Interaction parameters (electrostatic and van der Waals terms) were derived from 6-31G* *ab initio* interaction energies and geometries for water molecules bonded to polar

(121) van Gunsteren, W. F.; Berendsen, H. J. C. Groningen Molecular Simulations System, BIOMOS B.V., Laboratory of Physical Chemistry, University of Groningen, Groningen, The Netherlands, 1986.
(122) Zielinski, T. J.; Shibata, M. *Biopolymers* **1990**, *29*, 1027−1044.
(123) Swaminathan, S.; Ravishanker, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 5027−5040.
(124) Miaskiewicz, K.; Osman, R.; Weinstein, H. *J. Am. Chem. Soc.* **1993**, *115*, 1526.

sites of the model compounds and from the experimentally measured gas-phase Watson−Crick base pair energies and geometries, base heats of sublimation, and experimental and 6-31G* *ab initio* dipole moments.

The CHARMM22 parameters for the nucleic acid bases give satisfactory results for a variety of experimental gas and condensed phase data and *ab initio* calculations. For the nucleic acid bases the parameters reproduce the geometries and the vibrational frequencies and normal mode assignments. Limitations in the geometric data on the positions of the hydrogen atoms and in the vibrational data on the assignments of the normal modes, especially for frequencies less than 500 cm^{-1}, make additional optimization difficult. The availability of high-level *ab initio* data will allow some of these limitations to be overcome. Partial atomic charges and Lennard-Jones parameters were optimized to reproduce experimental and *ab initio* results. The interaction parameters in the potential energy function satisfactorily reproduce experimental heats of interaction for Watson−Crick base pairs and for several base homodimers; interaction energies, and geometries from *ab initio* calculations of a variety of water−base interactions; and the experimental heats of sublimation of uracil, 1-methylthymine, and 9-methyladenine. Crystal minimizations and simulations show the parameters to reproduce both intermolecular interactions and the unit cell parameters. Base stacking, which is difficult to compare directly to experiment, appears to be modeled satisfactorily. Analyses are made of the changes in the stacking interaction energies in a B-DNA dodecamer, the positions of minima of bases stacked in an ideal B-DNA geometry, low-level *ab initio* data, and stacking interactions in various crystal structures. Problems were encountered in the thymine crystal minimizations; however, results from the crystal molecular dynamics simulations gave adequate agreement with experiment. Overall, the ability of the parameters to accurately treat the intramolecular portion of the force field along with the intermolecular interactions ensures their usefulness in condensed phase simulations.

In the various parameter tests that were made, a number of results of general interest were found. In contrast to most other heavy atom bond angles, the adiabatic potential surface for the P−O−C angle in dimethyl phosphate was found to be highly anharmonic. Decreasing the P−O−C angle led to a significantly larger increase in energy than the opening of the angle. The increased energy upon closure is associated with the angle, van der Waals, bond, and electrostatic terms; the Urey−Bradley term partially compensates for the other terms. Such effects are more commonly observed in dihedral angle surfaces since the internal energy terms are much weaker. They emphasize the requirement that an empirical potential function must maintain consistency between the internal and external interactions and that dynamics as well as minimization studies are needed to evaluate the parameters. Also of interest is the fact that, in certain cases, the interaction terms have only a small effect on the dihedral angle potential. In the O1−P1−O2−C2 dihedral angle surface, for example, this is due to a balance of the electrostatic interactions between the carbon and hydrogen atoms on the methyl group and the anionic oxygen atoms and phosphorus atoms in the phosphate. The conformations of dimethyl phosphate and anionic methyl phosphate show significant differences between the minimized and dynamics averaged structures due to anharmonicity.

The crystal calculations for the various nucleic acid bases showed that accurate structural properties are obtained and that there is good agreement for the experimental heats of sublimation. This requires the maintenance of the two or three hydrogen bonds between two adjacent fluctuating bases. To maintain these intermolecular interactions, significant distortions of the internal geometries of the bases must occur. For the hydrogen bonds to be maintained in duplex DNA, a similar balance is required for the Watson−Crick base pairs. This balance between interactions and flexibility is not evident from experimental studies.

The stacking energies in the B-DNA dodecamer were calculated and it was found that they vary over a wide range; e.g., for G−C stacked pairs a stacking energy of −13.8 kcal/mol was obtained while the stacking energies are unfavorable for the T-G, G-G, and T-T stacked pairs (Table 14 of the supporting information). The variation in the energies is due primarily to the electrostatic term, which is strongly repulsive in almost all cases. Although such a result may seem surprising, it is in accord with molecular orbital calculations. The variation in the stacking energies calculated for nucleic acid bases suggests that these interactions, in addition to the base pairing hydrogen bonds, may have a significant effect on the dependence of nucleic acid structure and dynamics on the primary sequence. The present results show that B-DNA does not have a geometry that optimizes the base stacking interactions.

For the B-DNA decamer, the agreement between the simulation results and the crystal structure is very good. There is no distinct pattern in the observed differences between the calculated and the X-ray structure. For the bases a larger change in strand 1 is accompanied by a larger change in the adjacent base in strand 2, as expected due to the maintenance of the Watson−Crick base pairing. Such correlations are less evident for the sugars and do not occur for the phosphate moieties. This suggests that the backbone atoms of the individual DNA strands move independently. Comparison of the rms differences for the individual residues shows that the base and sugar differences are correlated. In contrast, little correlation between the phosphate and the bases or sugar is observed; i.e., the phosphates move independently of the surrounding sugars and bases. Such motions are allowed due to anticorrelated variations of the β and γ dihedral angles. The γ and β differences in the simulation relative to the X-ray structure (see Table 9) tend to occur in an anticorrelated fashion so that the change in the position of the second and twelfth phosphate groups is relatively small (see Figure 12).

Overall, the calculated backbone dihedral angle deviations from the X-ray structure are smaller than expected from the experimental temperature factors. This indicates that the experimental "fluctuations" are related to conformational substates of the duplex that are not sampled in the simulation. Support for this is provided by the lower amplitude rms fluctuations observed in the calculated results as compared with experiment. This is also in accord with the better agreement between the X-ray and calculated backbone dihedral angle and sugar pucker parameters in the GpC than the B-DNA simulations; i.e., the conformation space for GpC is likely to be much more restricted than that for B-DNA. Further study of this phenomenon requires either drastically increasing the duration of the simulation or running a number of simulations differing in the initial conditions that direct the simulations to different regions of conformational space.

Hydration properties are an essential part of DNA structure. The simulation of the high-resolution GpC crystal structure shows that it satisfactorily reproduces the water interaction distances and the hydration numbers for all the groups. The relationship between the anionic and ester phosphate oxygens is maintained for both the interaction distance and hydration number, supporting the accuracy of the charge distribution

developed for dimethyl phosphate. It is shown that the use of radial distribution functions provides a reliable approach for comparison with experiment. The interactions of sodium with water and with the phosphate oxygens are reproduced well in the simulation, indicating the quality of the sodium potential.[125] The octahedral coordination of the sodium, which includes four water molecules and an oxygen from two different phosphate groups, is maintained. In the B-DNA decamer simulation the hydration numbers and interaction distances were also in good agreement with experiment. Importantly, the difference between the anionic O to Mg^{2+} and the ester O to Mg^{2+} distances is maintained in the simulation; i.e., it appears that the balance between Mg^{2+} to water and Mg^{2+} to phosphate is represented by the parameter set.

Comparisons with nucleic acid simulations made with other energy functions show the importance of the parameters in determining the internal geometry and the interactions with the surroundings. It is our hope that the present parameter set, which was developed with considerable care so as to reproduce a wide range of experimental and theoretical properties with a relatively simple functional form for the potential energy, will be used by many people in CHARMM and other simulation programs. Although the parameters are the essential result of this study, they are reported as supporting information to save space in this journal.

Acknowledgment. This work was supported in part by a grant from the National Science Foundation. A.D.M. thanks the NIH for support in the form of an NIH postdoctoral fellowship. We thank William Jorgensen for sharing his nucleic acid parametrization results prior to publication, Udo Heinemann for supplying the B-DNA X-ray coordinates, Guy Privé for information on the B-DNA crystal structure, and Sun-Hee Jung for the Z-DNA coordinates.

(125) Roux, B. Ph.D. Thesis, Chemistry Department, Harvard University, 1989.

Supporting Information Available: Tables of data used for the optimization of the equilibrium bond lengths and angles along with the CHARMM optimized structures, detailed comparison of the CHARMM calculated frequencies and potential energy distributions for the five bases, Watson−Crick base pair interaction energies, calculated base dipole moments, nucleic acid base charges, base stacking minimum energies and distances, minimized lattice parameters, nonbonded interaction distances, stacking distances, heats of sublimation, internal geometries for DMP, frequencies and potential energy distributions, HF/6-31G* Mulliken population analysis charges, and pseudorotation angles and amplitudes and an Appendix giving a complete list of the bond, bond angle, dihedral angle, and Lennard-Jones parameters (60 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA941855C