A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules

Wendy D. Cornell,[†] Piotr Cieplak,[‡] Christopher I. Bayly,[§] Ian R. Gould,[⊥] Kenneth M. Merz, Jr.,[∥] David M. Ferguson,[&] David C. Spellmeyer,[#] Thomas Fox, James W. Caldwell, and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received November 10, 1994[⊗]

Abstract: We present the derivation of a new molecular mechanical force field for simulating the structures, conformational energies, and interaction energies of proteins, nucleic acids, and many related organic molecules in condensed phases. This effective two-body force field is the successor to the Weiner *et al.* force field and was developed with some of the same philosophies, such as the use of a simple diagonal potential function and electrostatic potential fit atom centered charges. The need for a 10−12 function for representing hydrogen bonds is no longer necessary due to the improved performance of the new charge model and new van der Waals parameters. These new charges are determined using a 6-31G* basis set and restrained electrostatic potential (RESP) fitting and have been shown to reproduce interaction energies, free energies of solvation, and conformational energies of simple small molecules to a good degree of accuracy. Furthermore, the new RESP charges exhibit less variability as a function of the molecular conformation used in the charge determination. The new van der Waals parameters have been derived from liquid simulations and include hydrogen parameters which take into account the effects of any geminal electronegative atoms. The bonded parameters developed by Weiner *et al.* were modified as necessary to reproduce experimental vibrational frequencies and structures. Most of the simple dihedral parameters have been retained from Weiner *et al.*, but a complex set of ϕ and ψ parameters which do a good job of reproducing the energies of the low-energy conformations of glycyl and alanyl dipeptides has been developed for the peptide backbone.

Introduction

The application of computer-based models using analytical potential energy functions within the framework of classical mechanics has proven to be an increasingly powerful tool for studying molecules of biochemical and organic chemical interest. These applications of molecular mechanics have employed energy minimization, molecular dynamics, and Monte Carlo methods to move on the analytical potential energy surfaces. Such methods have been used to study a wide variety of phenomena, including intrinsic strain of organic molecules, structure and dynamics of simple and complex liquids, thermodynamics of ligand binding to proteins, and conformational transitions in nucleic acids. In principle, they are capable of giving insight into the entire spectrum of non-covalent interactions between molecules, and, when combined with quantum mechanical electronic structure calculations, modeling covalent bonding changes, essentially all molecular reactions and interactions. Given their importance, much effort has gone into consideration of both the functional form and the parameters that must be established in order to apply such analytical potential energy functions (or "force fields").

In the area of organic molecules, the book by Allinger and Burkert[1] provides a thorough review pre-1982 and the subsequent further development of the MM2[2] and MM3[3] force fields by Allinger and co-workers has dominated the landscape in this area. The number of force fields developed for application to biologically interesting molecules is considerably greater, probably because of the greater complexity of the interactions which involve ionic and polar groups in aqueous solution and the difficulty of finding an unequivocal test set to evaluate such force fields. Many of these force fields developed prior to 1987 are described briefly by McCammon and Harvey.[4]

Given the complexities and subjective decisions inherent in such biological force fields, we have attempted to put the development of the force field parameters on a more *explicitly stated* algorithmic basis than done previously, so that the force field could be extended by ourselves and others to molecules and functional groups not considered in the initial development. This is important, because, if the assumptions, approximations, and inevitable imperfections in a force field are at least known, one can strive for some cancellation of errors.

Approximately a decade ago, Weiner *et al.*[5,6] developed a force field for proteins and nucleic acids which has been widely

[†] Graduate Group in Biophysics, University of California, San Francisco.
[‡] Permanent address: Department of Chemistry, University of Warsaw, Pasteura 1, 02-093, Warsaw, Poland.
[§] Current address: Merck Frosst Canada, Inc., C.P. 1005 Pointe Claire-Dorval, Quebec H9R 4P8, Canada.
[⊥] Current address: Department of Chemistry, University of Manchester, Lancs M13 9PL, U.K.
[∥] Current address: Department of Chemistry, The Pennsylvania State University, State College, PA 16802.
[&] Current address: Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455.
[#] Current address: Chiron Corporation, Emeryville, CA 94608.
* Author to whom correspondence and reprint requests should be addressed.
[⊗] Abstract published in *Advance ACS Abstracts*, April 15, 1995.

(1) Burkert, U.; Allinger, N. J. *Molecular Mechanics;* American Chemical Society: Washington, DC, 1982.
(2) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134 and subsequent versions, e.g. MM2–87, MM2–89, MM2–91.
(3) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566, 8566–8576, 8576–8582.
(4) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
(5) 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.
(6) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230–252.

0002-7863/95/1517-5179$09.00/0

used. Important independent tests of this force field were performed by Pavitt and Hall for peptides[7] and Nilsson and Karplus[8] for nucleic acids and it was found to be quite effective. Nonetheless, it was developed in the era before one could routinely study complex molecules in explicit solvent. Weiner *et al.* attempted to deal with this issue by showing that the same force field parameters could be effectively used both without explicit solvent (using a distance-dependent dielectric constant ($\epsilon = R_{ij}$)) and with explicit solvent ($\epsilon = 1$) on model systems. Further support for this approach was provided by molecular dynamics simulations of proteins[9-11] and DNA[12,13] which compared the implicit and explicit solvent representations.

As computer power has grown, it has become possible to carry out more realistic simulations which employ explicit solvent representations. It is therefore appropriate that any new force field for biomolecules focus on systems modeled in the presence of an explicit solvent representation. This approach has been pioneered by Jorgensen and co-workers in their OPLS (Optimized Potentials for Liquid Simulations) model.[14] In particular, the development of parameters which reproduce the enthalpy and density of neat organic liquids as an essential element ensures the appropriate condensed phase behavior. The OPLS non-bonded parameters have been combined with the Weiner *et al.* bond, angle, and dihedral parameters to create the OPLS/Amber force field for peptides and proteins,[15] which has also been effectively used in many systems.[16]

We have been influenced by the OPLS philosophy of balanced solvent−solvent and solute−solvent interactions in our thoughts about a second-generation force field to follow that of Weiner *et al.*[5,6] The Weiner *et al.* force field used quantum mechanical calculations to derive electrostatic potential (ESP) fit atomic centered charges, whereas the OPLS charges were derived empirically, using mainly the liquid properties as a guide. For computational expediency, Weiner *et al.* relied principally on the STO-3G basis set for their charge derivation. This basis set leads to dipole moments that are approximately equal to or smaller than the gas-phase moment but tends to underestimate quadrupole moments. Thus, it is not well *balanced* with the commonly used water models (SPC/E,[17] TIP3P,[18] TIP4P[18]) which have dipole moments that are about 20% higher than the gas-phase value for water. These water models, which have empirically derived charges, include condensed-phase electronic polarization implicitly. Kuyper *et al.*[19] suggested that the logical choice of a basis set for deriving ESP-fit partial charges for use in condensed phases is the 6-31G* basis set, which uniformly overestimates molecular polarity. Standard ESP charges derived with that basis set were shown to lead to excellent relative free energies of solvation for benzene, anisole, and trimethoxyanisole.[19]

A 6-31G* based ESP-fit charge model, like the OPLS model, is capable of giving an excellent reproduction of condensed-phase *inter* molecular properties such as liquid enthalpies and densities and free energies of solvation.[20] A major difference between such a model and most others is the magnitude of the charges on hydrocarbons. For example, 6-31G* standard ESP charges derived from the trans conformation of butane have values of −0.344 for the methyl carbon and 0.078 for the methyl hydrogen. In both cases, however, the carbon and hydrogen charges offset each other, resulting in small net charges on the methyl groups of −0.110 and −0.059 for the trans and gauche charges, respectively. Furthermore, free energy perturbation calculations involving the perturbation of methane with standard ESP charges ($q_C = -0.464$ and $q_H = 0.116$) to methane with charges of 0.0 in solution yield essentially no change in free energy.[21] The standard ESP charges also result in conformational energies for butane which are in reasonable agreement with experiment, when used with a 1−4 electrostatic scale factor of 1/1.2.[20]

Nevertheless, the 6-31G* standard ESP charges are less than ideal for two reasons. First, when charges generated using different conformations of a molecule are compared, there is often considerable variation seen. This was demonstrated by Williams, who studied the conformational variation of ESP-fit charges in alanyl dipeptide for 12 different conformations.[22] Butane is another example, where charges from the gauche conformation have values of −0.197 and 0.046 for the methyl carbon and hydrogen, respectively. Another example is propylamine, which was studied at length by Cornell *et al.*[20] Five low-energy conformations can be identified for propylamine, and the 6-31G* standard ESP charges calculated for each conformation show significant variation. The average and standard deviation for the charge on a given atom over the five conformations are as follows: α-carbon $q_{av} = 0.339$ and $\sigma = 0.059$, β-carbon $q_{av} = 0.033$ and $\sigma = 0.060$, and γ-carbon $q_{av} = -0.205$ and $\sigma = 0.146$. This inconsistency is potentially problematic in terms of deriving other force field parameters which may be sensitive to the variation. Furthermore, it reduces the reproducibility of a particular calculation, which is not a problem in other force fields where the charges are assigned empirically.

The second reason that the 6-31G* standard ESP charges are less than ideal is that the charges on "buried" atoms (such as the sp^3 carbons described above for butane and propylamine) are statistically underdetermined and often assume unexpectedly large values for nonpolar atoms. Bayly *et al.*[23] found that the electrostatic potential of methanol could be fit almost equally well using either the standard ESP charges determined by the linear least-squares fit or an alternative set of charges derived with the methyl carbon constrained to have a much smaller value.

Considering the problems associated with the standard ESP charge model, it might seem tempting to adopt the OPLS approach of empirically derived charges. However, any empirically derived charge model cannot easily describe transition states and excited states, as can an electrostatic potential fit

(7) Pavitt, N.; Hall, D. *J. Comput. Chem.* **1984**, *5*, 441−450
(8) Nilsson, L.; Karplus, M. *J. Comput. Chem.* **1986**, *7*, 591−616.
(9) Tilton, R. F.; Singh, U. C.; Weiner, S. J.; Connolly, M. L.; Kuntz, I. D., Jr.; Kollman, P. A.; Max, N.; Case, D. *J. Mol. Biol.* **1986**, *192*, 443−456.
(10) Guenot, J. M.; Kollman, P. A. *Protein Sci.* **1992**, *1*, 1185−1205.
(11) York, D. M.; Wlodawer, A.; Redersen, L.; Darden, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8715−8718.
(12) Singh, U. C.; Weiner, S. J.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 755−759.
(13) Seibel, G. L.; Singh, U. C.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6537−6540.
(14) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008−2010.
(15) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(16) (a) Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1990**, *112*, 2773−2781. (b) Orozco, M.; Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1993**, *32*, 12864−12874.
(17) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(18) Jorgensen, W. L.; Chandreskhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1982**, *79*, 926−935.
(19) Kuyper, L.; Ashton, D.; Merz, K. M., Jr.; Kollman, P. A. *J. Phys. Chem.* **1991**, *95*, 6661−6666.
(20) Cornell, W.; Cieplak, P.; Bayly, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.
(21) (a) Sun, Y. X.; Spellmeyer, D.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1992**, *114*, 6798−6801. (b) Sun, Y. X.; Kollman, P. A. Hydrophobic Solvation of Methane and Nonbond Parameters of the TIP3P Water Model. *J. Comput. Chem.*, in press. Pang, Y. P.; Kollman, P. A., unpublished.
(22) Williams, D. E. *Biopolymers* **1990**, *29*, 1367−1386.
(23) Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

model. Furthermore, the conformational dependence of *N*-methylacetamide (NMA) is better represented with an ESP-fit model.[24] Finally, the requirement of Monte Carlo calculations on requisite liquids including appropriate fragments makes it more problematic to make an empirical charge model that will cover most or all of chemical/biochemical functionality.

Given the above-mentioned deficiencies in the standard ESP model, along with the desire to retain the general strategy of fitting charges to the electrostatic potential, Bayly *et al.*[23] were motivated to develop the RESP (restrained ESP-fit) charge model. The RESP model still involves a least-squares fit of the charges to the electrostatic potential, but with the addition of hyperbolic restraints on charges on non-hydrogen atoms. These restraints serve to reduce the charges on atoms which *can* be reduced without impacting the fit, such as buried carbons. The final RESP model requires a two-stage fit, with the second stage needed to fit methyl groups which require equivalent charges on hydrogen atoms which are not equivalent by molecular symmetry. The new charge model has been shown to perform well at reproducing interaction energies and free energies of solvation. When used with a 1−4 electrostatic scale factor of 1/1.2 (as opposed to the scale factor of 1/2 employed by Weiner *et al.*), both the RESP (and standard ESP) charges also result in good conformational energies for many of the small molecules studied to date without the necessity for an elaborate dihedral potential.[20]

In addition to the new charges which have been tailored for condensed phase simulations, new van der Waals (VDW) parameters have also been adopted and developed which are optimized for reproducing liquid properties. The VDW parameters in the Weiner *et al.*[5,6] force field are primarily a modification of a set originally proposed by Hagler−Euler−Lifson,[25] which were fit to lattice energies and crystal structures of amides. The new VDW parameters for aliphatic and aromatic hydrogens take into account the effects of any vicinal electronegative atoms.[26,27]

High-level quantum mechanical data are now available on the conformational energies of the glycyl and alanyl dipeptides[28] and these data are critical for developing ϕ and ψ dihedral parameters for the peptide backbone. Because such high-level data were unavailable at the time the Weiner *et al.* force field was developed, torsional parameters for the ϕ and ψ angles were left as 0.0 kcal/mol since the resulting molecular mechanical energies seemed to be in reasonable agreement with the best theoretical data available at that time. That force field led to conformational energies for glycyl dipeptide where the C5 extended conformation was about 1 kcal/mol too high in energy and for alanyl dipeptide where the C5 conformation was nearly 2 kcal/mol too high in energy but the $C7_{ax}$ conformation was about 1 kcal/mol too low in energy. The error in the alanyl dipeptide $C7_{ax}$ energy is not critical since it is rarely found in proteins[29] (only in γ-turns), but the errors in the energies of the C5 conformations are more important since that is the conformation found in β-sheets. Any errors in the energies of the C5 conformations are multiplied by the length of the secondary structure. The new force field includes V_1, V_2, V_3, and V_4 dihedral parameters for ϕ and ψ which result in good agreement between the molecular mechanical and quantum mechanical energies of the dipeptides.

Finally, the benzene molecule as modeled by the Weiner *et al.* all-atom force field has been shown to possess excessive flexibility for out-of-plane distortions.[30] This was caused by the use of the V_2 potential derived for the united atom model. This underestimate of the benzene V_2 parameter is noteworthy, because it affects not only the flexibility of benzene and benzene-like moieties but also the interpolation scheme used for determining the V_2 barriers for X−C−N−X and X−C−C−X dihedrals in conjugated rings. These V_2 parameters are determined by interpolating according to the bond length either between a pure single bond and a partial double bond (benzene) or between a partial double bond and a pure double bond. The excessive out-of-plane motion of benzene has been easily fixed by adjusting the V_2 parameter from 5.5 to 14.5 kcal/mol to match the experimental normal mode frequencies.

General Description of the Model

The model presented here (eq 1) can be described as "minimalist" in its functional form, with the bond and angles represented by a simple diagonal harmonic expression, the VDW interaction represented by a 6−12 potential, electrostatic interactions modeled by a Coulombic interaction of atom-centered point charges, and dihedral energies represented (in most cases) with a simple set of parameters, often only specified by the two central atoms. Electrostatic and van der Waals interactions are only calculated between atoms in different molecules or for atoms in the same molecule separated by at least three bonds. Those non-bonded interactions separated by exactly three bonds ("1−4 interactions") are reduced by the application of a scale factor.

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] \quad (1)$$

Our assumption is that such a simple representation of bond and angle energies is adequate for modeling most unstrained systems. The goal of this force field is to accurately model conformational energies and intermolecular interactions involving proteins, nucleic acids, and other molecules with related functional groups which are of interest in organic and biological chemistry.

A. Atom Types. The atom types employed are similar to those defined previously and are given in Table 1. The one significant departure is the definition of new atom types for hydrogens bonded to carbons which are themselves bonded to one or more electronegative atoms. This is similar in spirit to the electronegativity based bond length correction used in MM2 and MM3.

B. Bond and Angle Parameters. The r_{eq}, θ_{eq}, K_r, and K_θ values[5,6] were used as starting values and adjusted as necessary to reproduce experimental normal mode frequencies. These values were initially derived by fitting to structural and vibrational frequency data on small molecular fragments that make up proteins and nucleic acids. For example, in complex fragments such as the nucleic acid bases, the r_{eq} and θ_{eq} values have been taken from X-ray structural data, the K_r values

(24) Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **1991**, *12*, 1232−1236.
(25) Hagler, A.; Euler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319−5327.
(26) Gough, C.; DeBolt, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 963−970.
(27) Veenstra, D.; Ferguson, D.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 971−978.
(28) (a) Gould, I. R.; Kollman, P. A. *J. Phys. Chem.* **1992**, *96*, 9255−9258. (b) Gould, I. R.; Cornell, W. D.; Hillier, I. H. *J. Am. Chem. Soc.* **1994**, *116*, 9250−9256.
(29) Creighton, T. E. *Proteins*, 2nd. ed.; W. H. Freeman: New York, 1984.
(30) Lipkowitz, K. B.; Peterson, M. A. *J. Comput. Chem.* **1993**, *14*, 121−125.

Table 1. List of Atom Types[a]

atom	type	description
carbon	CT	any sp^3 carbon
	C	any carbonyl sp^2 carbon
	CA	any aromatic sp^2 carbon and (Cϵ of Arg)
	CM	any sp^2 carbon, double bonded
	CC	sp^2 aromatic in 5-membered ring with one substituent + next to nitrogen (Cγ in His)
	CV	sp^2 aromatic in 5-membered ring next to carbon and lone pair nitrogen (e.g. Cδ in His (δ))
	CW	sp^2 aromatic in 5-membered ring next to carbon and NH (e.g. Cδ in His (ϵ) and in Trp)
	CR	sp^2 aromatic in 5-membered ring next to two nitrogens (Cγ and Cϵ in His)
	CB	sp^2 aromatic at junction of 5- and 6-membered rings (Cδ in Trp) and both junction atoms in Ade and Gua
	C*	sp^2 aromatic in 5-membered ring next to two carbons (e.g. Cγ in Trp)
	CN	sp^2 junction between 5- and 6-membered rings and bonded to CH and NH (Cϵ in Trp)
	CK	sp^2 carbon in 5-membered aromatic between N and N-R (C8 in purines)
	CQ	sp^2 carbon in 6-membered ring between lone pair nitrogens (e.g. C2 in purines)
nitrogen	N	sp^2 nitrogen in amides
	NA	sp^2 nitrogen in aromatic rings with hydrogen attached (e.g. protonated His, Gua, Trp)
	NB	sp^2 nitrogen in 5-membered ring with lone pair (e.g. N7 in purines)
	NC	sp^2 nitrogen in 6-membered ring with lone pair (e.g. N3 in purines)
	N*	sp^2 nitrogen in 5-membered ring with carbon substituent (in purine nucleosides)
	N2	sp^2 nitrogen of aromatic amines and guanidinium ions
	N3	sp^3 nitrogen
oxygen	OW	sp^3 oxygen in TIP3P water
	OH	sp^3 oxygen in alcohols, tyrosine, and protonated carboxylic acids
	OS	sp^3 oxygen in ethers
	O	sp^2 oxygen in amides
	O2	sp^2 oxygen in anionic acids
sulfur	S	sulfur in methionine and cysteine
	SH	sulfur in cysteine
phosphorus	P	phosphorus in phosphates
hydrogen	H	H attached to N
	HW	H in TIP3P water
	HO	H in alcohols and acids
	HS	H attached to sulfur
	HA	H attached to aromatic carbon
	HC	H attached to aliphatic carbon with no electron-withdrawing substituents
	H1	H attached to aliphatic carbon with one electron-withdrawing substituent
	H2	H attached to aliphatic carbon with two electron-withdrawing substituents
	H3	H attached to aliphatic carbon with three electron-withdrawing substituents
	HP	H attached to carbon directly bonded to formally positive atoms (e.g. C next to NH_3^+ of lysine)
	H4	H attached to aromatic carbon with one electronegative neighbor (e.g. hydrogen on C5 of Trp, C6 of Thy)
	H5	H attached to aromatic carbon with two electronegative neighbors (e.g. H8 of Ade and Gua and H2 of Ade)

[a] See refs 5 and 6.

determined by linear interpolation between pure single and double bond values using the observed bond distances and the K_θ value taken from vibrational analysis of a simple sp^2 atom containing fragments such as benzene and NMA. That this approach was reasonably successful is supported by the reasonable agreement found in nucleic acid base vibrational analysis and suggested by the critical analysis of Halgren of the diagonal force constants used in different force fields.[31]

One "difficulty" arose in the development of this new force field compared to that of Weiner *et al.* which was related to the switch to the 6-31G* basis set for charge derivation. With 6-31G* standard ESP charges and a 1−4 electrostatic scale factor of 1/1.2 rather than 1/2.0 (see below), we found that the exocyclic $-NH_2$ groups of the bases moved considerably away from their r_{eq} and θ_{eq} values upon energy minimization. This problem was considerably reduced with RESP charges and a 1−4 electrostatic scale factor of 1/1.2, so we chose not to selectively increase the K_θ values around the $-NH_2$ group to force it to more "canonical" geometries.

In general, however, one might have resorted to a more complex optimization of r_{eq}, θ_{eq}, K_r, and K_θ to ensure that the geometries of simple fragments were as close as possible to experiment *after energy minimization*, rather than taking r_{eq} and θ_{eq} from experiment and assuming little distortion would occur (which is generally the case, with the slight exception of the case of the $-NH_2$ groups noted above). We chose not to undertake a more time-consuming iterative self-consistent derivation of geometrical parameters, because of our assumption that any such errors which we were making were of much smaller consequence for accurately representing conformations and intermolecular interactions than the inaccuracies remaining in the dihedral and non-bonded (charge and VDW) parameters.

C. Dihedral Parameters. Weiner *et al.*[5,6] developed a limited set of general and specific dihedral parameters which were appropriate for the functionalities found in proteins and DNA and calibrated to adjust the energies of small model compounds. In this strategy, a dihedral parameter is optimized on the simplest molecule possible and then applied to larger and more complex molecules. This approach is in contrast to one employed by many other force field developers where the parameters are optimized to best reproduce the conformational energies of a large number of molecules. An advantage of our approach is the lack of dependence of the resulting parameters on the particular molecules chosen for the test set.

For the most part, a minimalist approach has been retained with regards to dihedral parameters. For example, we have only a 3-fold Fourier component (V_3) for dihedrals around −C−C− bonds, with the exception of cases such as E−C−C−E′ where E and E′ are electronegative atoms like O or F. In these cases, there is a "gauche" effect which stabilizes the gauche conformation over the trans and this can be modeled with a 2-fold Fourier component (V_2). The rotation around phosphorus−ester bonds (CT−OS−P−OS) also requires a 2-fold component. In these cases, we have been able to go beyond the Weiner *et al.* force field by making use of reasonably high level *ab initio* models (MP2/6-31G*) to fit the values of such V_2 Fourier components.

Two exceptions were made to the principle of adding extra Fourier terms to the dihedral energies only in the presence of a compelling physical basis. These exceptions are the dipeptide ψ and ϕ and the nucleoside χ dihedrals. Here we used additional Fourier components to try to reproduce as well as possible the relative energies of the alanyl and glycyl dipeptides and a model nucleoside fragment calculated at a high level of theory without the requirement of "a physical picture". An alternative approach would be to empirically adjust the atomic partial charges to achieve the same aim. Given the power of the RESP methodology for deriving atomic partial charges which lead to good representations of intermolecular interactions and the importance of maintaining an accurate balance between intra- and intermolecular interactions, we chose to empirically adjust the terms in the Fourier series for ψ and ϕ as well as χ.

(31) Halgren, T. A. *J. Am. Chem. Soc.* **1990**, *112*, 4710−4723.

Table 2. Standardized Parameters for Scaling Algorithms

bond	r_{eq}[a]	K_r[b]
pure C−C	1.507[c]	317[d]
pure C=C	1.336[e]	570[f]
pure C−N	1.449[g]	337[h]
pure C=N	1.273[i]	570[j]
torsion	r_{eq}[a]	V_2[k]
pure X−C−C−X	1.507[c]	0.0[l]
partial X−C=C−X	1.397[m]	14.5[n]
pure X−C=C−X	1.336[e]	30.0[o]
pure X−C−N−X	1.449[g]	0.0[p]
partial X−C=N−X	1.335[q]	10.0[r]
pure X−C=N−X	1.273[i]	30.0[s]

[a] In Å. [b] In kcal/(mol Å^2). [c] Microwave data from acetone (ref 32). [d] Value taken from MM2, ref 2. [e] Microwave data from propene (ref 32). [f] Default from NMA normal mode analysis for carbonyl force constant. [g] Benedetti structural data (ref 33). [h] Value derived from normal mode analysis on NMA. [i] Microwave data from methylenimine (ref 32). [j] Default value, see footnote f. [k] In kcal/mol. [l] Assumed free rotation about pure C−C single bond. [m] Structural data from benzene (ref 32). [n] From normal modes analysis of benzene. [o] Approximate rotational barrier of ethylene is ~60 kcal/mol (see ref 34). [p] Assumed free rotation about a pure single C−N bond. [q] Benedetti structural data (ref 33). [r] Reference 35. [s] Calculated rotational barrier in methylenimine is 57.5 kcal/mol (see ref 36).

In our previous force field, the bond length and V_2 parameters for X−C−N−X and X−C−C−X fragments involving sp^2 hybridized atoms were determined by a linear interpolation approach (according to the experimental bond length) between the known barriers of pure single, pure double, and partial double bonded systems (benzene for X−C−C−X and NMA for X−C−N−X). We have used the same approach here, but have adjusted the V_2 term of benzene to more accurately describe its out-of-plane frequencies (Weiner *et al.*[5,6] had used the V_2 derived for a united atom model of benzene, which was significantly different). Table 2 presents the parameters used. For example, given a C(sp^2)−C(sp^2) bond length, its bond stretching force constant is linearly interpolated between the values for pure single bond and double bond given in Table 2. Its V_2 torsional potential is interpolated between the values for pure double and partial double or between partial double and single, depending on whether the bond length is greater or less than the 1.397 Å of benzene. This is exactly the procedure used by Weiner *et al.*[5,6]

D. VDW Parameters. Given the success of the OPLS approach in modeling liquids, we have developed all-atom sp^3 carbon and aliphatic hydrogen VDW parameters by carrying out Monte Carlo simulations on CH_4, C_2H_6, C_3H_8, and C_4H_{10} liquids and empirically adjusting R^* and ϵ for the C and H to reproduce the densities and enthalpies of vaporization of these liquids.[37] Such parameters have also been employed in calculations of relative free energies of solvation of CH_4, C_2H_6, and C_3H_8.[21,38] We also derived VDW parameters for sp^2 C and aromatic H employing Monte Carlo simulations on benzene liquid and adjusting the R^* and ϵ of these atoms to reproduce the density and enthalpy of liquid benzene.[37] At the time these parameters were developed, such all-atom parameters were unavailable for the OPLS force field. These Monte Carlo simulations were the first calculations carried out as part of the development of this new force field, and as such employed 6-31G* standard ESP charges. The electrostatic contribution for the *n*-alkanes was very small regardless of the charge model—at most a few tenths of a kcal/mol. We note that the standard ESP charges for benzene (q_C= −0.145 and q_H= 0.145) accurately reproduce the quadrupole moment of that molecule.

We have taken most of the remaining VDW parameters from the OPLS model[15]—sp^2 and sp^3 N; sp^2 O, ether ester (OS), hydroxyl (OH) and TIP3P water (OW) sp^3 oxygens; and sulfur (SH and S)—since it has been optimized for reproducing liquid properties. The Weiner *et al.*[5,6] phosphorus (P) parameters were not re-optimized since that atom is most frequently found buried inside of four other heavy atoms.

The VDW model is minimalist as well, with some exceptions. A standard VDW parameter is used for a given atom and hybridization, e.g. all sp^2 carbons have the same VDW parameters. The only heavy atom exceptions are sp^3 O, where oxygens in water (OW), alcohol (OH), and ether (OS) have slightly different parameters, as found in OPLS. We suspect that this is due to the use of a zero VDW radius on hydrogens bound to oxygen, so that an effectively larger R^* is required for a water oxygen than alcohol than ether.

A significant departure has been made from the previous model in the treatment of hydrogens. The current model does not employ 10−12 hydrogen bonding H···X parameters, although these are still supported within the AMBER software. The original Hagler *et al.*[25] and OPLS approach[14,15] suggested a zero R^* and ϵ for hydrogen binding hydrogens. Thus the TIP3P water model has R^* and ϵ equal to 0.0 for its hydrogen (HW). We opted not to develop a new water model, but to use the TIP3P one.

Hydrogen and helium are unique in the periodic table in not having an inner shell of electrons. Consequently, it makes physical sense for the hydrogen VDW radius, unlike other atoms, to be very sensitive to its bonding environment. This has been extensively analyzed for the hydrogen R^* in X−C−*H* systems by Gough *et al.* and Veenstra *et al.*,[26,27] who demonstrated the sensitivity of R^* to the electron-withdrawing properties of X. For example, a "normal" C−*H* has VDW R^* = 1.487 Å; whereas in CF_3−*H* it is ~0.3 Å shorter and in $CH_3NH_3^+$ it is ~0.4 Å shorter still.

We have employed the following approach here. A C−*H* has R^* = 1.487 Å and, based on nucleic and base pairing energy minimization, an N−*H* has R^* = 0.6 Å. This qualitative dependence on electronegativity makes physical sense. Based on the Veenstra *et al.*[27] studies we have chosen to reduce the R^* on sp^3 C−*H* atoms by 0.1 Å for each electronegative (O, N, F, S) substituent. The hydrogen atom types are then defined as H1, H2, and H3 for 1, 2, and 3 electronegative groups, respectively. The hydrogen R^* is reduced by 0.4 Å for each neighboring positively charged group (atom type HP). For sp^2 C−*H*, R^* has been reduced by 0.05 Å for each electronegative neighbor (atom types H4 and H5).

Given our retention of the simplicity of a 6−12 rather than a 6-exponential VDW representation, we have continued to reduce 1−4 VDW interactions since the 6−12 approximation and the lack of polarization in the model both will lead to exaggerated short-range repulsion. It is difficult to determine the scale factor unambiguously so we have retained the value of 1/2.0 used by Weiner *et al.*[5,6]

E. Electrostatic Energies. In Cornell *et al.*[20] and Cieplak *et al.*,[39] we have extensively analyzed the development of our

(32) Harmony, M.; Laurie, V.; Kuezkowski, R.; Schwendeman, R.; Ramsay, D.; Lovas, F.; Lafferty, W.; Maki, A. *J. Phys. Chem. Ref. Data* **1979**, *8*, 619−721.
(33) Benedetti, E. In *Peptides-Proceedings of the 5th American Peptide Symposium*; Goodman, M., Meienhofer, J., Eds.; J. Wiley and Co.: New York, 1977; pp 257−273.
(34) Douglas, J.; Rabinovich, B. S.; Looney, F. *J. Chem. Phys.* **1955**, *23*, 315−323.
(35) Momany, F.; McGuire, R.; Burgess, A.; Scheraga, H. *J. Phys. Chem.* **1975**, *79*, 2361−2381.
(36) Lehn, J.; Munsch, B.; Millie, P. H. *Theor. Chim. Acta* **1970**, *16*, 351−372.
(37) Spellmeyer, D., unpublished.
(38) Sun, Y.; Kollman, P. A. Hydrophobic Solvation of Methane and Nonbond Parameters of the TIP3P Water Model. *J. Comput. Chem.*, accepted for publication.

electrostatic model, which relies on the use of 6-31G* derived electrostatic potentials, multiple molecules, multiple conformations, and the RESP fitting approach. The multiple molecules/conformations and RESP fitting all serve to reduce the problem of statistically under-determined charges on buried atoms. We have further validated these models in their ability to calculate liquid enthalpies and densities[40] and free energies of solvation[20] of the prototypal polar molecules methanol and NMA in good agreement with experiment. We have not used lone pairs on sulfur in the new force field, despite their importance in hydrogen bond directionality[5,6] because of the PDB analysis of Gregoret *et al.*, which showed that neutral sulfur functions only extremely rarely as a proton acceptor in proteins.[41]

The new RESP charge model employs a scale factor of 1/1.2 for 1−4 electrostatics, which was calibrated on 1,2-ethanediol and also performed well on tests on simple alcohols, amines, and butane.[20] The RESP and standard ESP charge models were shown by Howard *et al.* to perform better than MM2 and MM3 in the conformational analysis of substituted 1,3 dioxanes,[42] requiring only the addition of a single dihedral parameter optimized on 2,4-dioxapentane.

Methods

ESP and RESP charges were calculated from electrostatic potentials derived using the Gaussian 90 and Gaussian 92 programs.[43] These programs were also employed for *ab initio* calculations of conformational energies. All minimization and normal mode calculations reported for this work were carried out using the AMBER package.[44] Scale factors of 1/1.2 and 1/2 were applied to 1−4 electrostatic and VDW interactions, respectively.

Free energy perturbation calculations for perturbing methanethiol to methanol and dimethyl thioether to dimethyl ether were carried out using the AMBER program and the slow growth method.[45] Simulations were run for 200 ps with a time step of 2 fs. SHAKE[46] was applied to constrain all bonds and perturbed bonds were shrunk. Only the solution perturbation was carried out (with TIP3P water[18] and periodic boundary conditions) and the intramolecular components were not included. Calculations were carried out in both the forward and reverse directions. The PMF correction was included to account for the free energy change associated with perturbed bonds.[47]

Free energy perturbation calculations for the perturbation of 9-methyladenine to methane were carried out using the SPASMS[48] module of the AMBER program and the windows method using the acceptance ratio[49] approach and decoupling the electrostatic and VDW perturbations. All intramolecular components were included. The gas-phase electrostatic runs were carried out with 11 windows with 5K (5000) steps of equilibration and 10K steps of data collection. The gas-phase VDW runs were carried out with 51 windows with 1K steps of equilibration and 5K steps of data collection. The solution perturbation was carried out with TIP3P water and periodic boundary conditions. The electrostatic part of the solution calculation was carried out analogously to the gas-phase electrostatic calculation. The VDW part of the solution calculation was carried out with 51 windows, 1K steps of equilibration, and 4K steps of data collection. A 9.0 Å cut-off with no switch functions was employed for non-bonded interactions and the time step was 1 fs. The coupling constants were 0.2 (temperature) and 0.4 ps (pressure).

Molecular dynamics simulations of ubiquitin were carried out using the AMBER program.[44] The simulations were carried out at 300 K with a time step of 1.5 fs and a non-bonded cut-off of 8.0 Å. SHAKE[46] was applied to bonds containing hydrogens.

Table 3. Results for Hydrocarbons (Energies in kcal/mol, Angles in deg)

parameter	this work	MM3[a]	experiment or high-level theory
	Ethane		
ΔE(eclipsed staggered)	2.89	2.41	2.88[b]
	Butane		
ΔE(gauche-trans)	0.67	0.81	0.75 ± 0.25[c]
ΔE(cis-trans)	5.16	4.83	4.56,[g] 4.89[h]
structural parameters			
ϕ(gauche)	68.0	64.5	71 ± 5[c]
θ(C−C−C)(cis)	117.2		
θ(C−C−C)(trans)	111.3	112.4	113 ± 4[c]
θ(C−C−C)(gauche)	113.5	113.7	
	Propane		
ΔE(V1)[d]	3.30		3.3[f]
ΔE(V2)[d]	3.74		3.9[f]

[a] Reference 3. [b] Reference 50. [c] Reference 51. [d] Energy for methylene group to eclipse first methyl group, relative to all staggered conformation (V1) and energy for methylene to eclipse second methyl group, relative to first eclipsed conformation (V2) (ref 52). [e] Reference 53. [f] Reference 54. [g] Reference 52. [h] Reference 53.

Results

We begin the development of the force field with ethane, the fundamental unit for hydrocarbons. The general V_3(X−CT−CT−X) dihedral was changed from 1.3 to 1.4 kcal/mol in order to reproduce the experimental barrier to rotation (Table 3). Ethane charges have been shown to be particularly sensitive to the conditions of the esp fit.[55] Nonetheless, changing the charges on hydrogen from 0.0 to 0.1 changes the barrier only from 2.89 to 2.92 kcal/mol. In contrast to MM2/MM3,[2,3] only this general V_3 dihedral potential is used for hydrocarbons. As one can see in Table 3, the conformational energies and structures are well represented for the simple model hydrocarbons with such an approach. At this point, we should note the difference of our approach from that of MM3,[3] where the rotational barrier in ethane is ~0.5 kcal/mol smaller than

(39) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformation RESP Methodology to Biopolymers: Derivation for DNA, RNA and Proteins. *J. Comput. Chem.*, in press.

(40) Caldwell, J.; Kollman, P. The Structures and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-Methyl Acetamide. *J. Phys. Chem.*, in press.

(41) Gregoret, L. M.; Rader, S. D.; Fletterick, R. J.; Cohen, F. E. *Proteins-Struct. Funct. Genet.* **1991**, *9*, 99−107.

(42) Howard, A.; Cieplak, P.; Kollman, P. A. A Molecular Mechanical Model that Reproduces the Relative Energies for Chair and Twist-Boat Conformations of 1,3- Dioxanes. *J. Comput. Chem.*, in press.

(43) (a) Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M. A.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seger, R.; Melius, C. F.; Baker, J.; Martin, L. R.; Kahn, L. R.; Stewart, J. J. P.; Topiol, S.; Pople, J. A. *Gaussian 90*; Gaussian, Inc.: Pittsburgh, PA, 1990. (b) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian 92*, Revision A; Gaussian, Inc.: Pittsburgh, PA, 1992.

(44) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Seibel, G. L.; Singh, U. C.; Weiner, P. A.; Kollman, P. A. *AMBER 4.0 (UCSF)*; Department of Pharmaceutical Chemistry, University of California: San Francisco, CA, 1991.

(45) van Gunsteren, W. F.; Berendsen, H. J. C. *J. Comput. Aided Mol. Des.* **1987**, *1*, 171−176.

(46) (a) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311−1327. (b) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(47) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532−4545.

(48) Spellmeyer, D. C.; Swope, W. C.; Evensen, E.-R.; Ferguson, D. M. *SPASMS*; University of California: San Francisco, CA, 1994.

(49) Ferguson, D. M. *J. Chem. Phys.* **1993**, *99*, 10086−10087.

(50) Hirota, E.; Emdo, Y.; Saito, S.; Duncan, J. *J. Mol. Spectrosc.* **1981**, *89*, 285−295.

(51) Heenan, R. K.; Bartell, L. S. *J. Chem. Phys.* **1983**, *78*, 1270−1274.

(52) Compton, D. A. C.; Montero, S.; Murphy, W. F. *J. Phys. Chem.* **1980**, *84*, 3587−3592.

(53) Allinger, N. L.; Grev, R. S.; Yates, B. F.; Schaefer, H. F., III *J. Am. Chem. Soc.* **1990**, *112*, 114−118.

(54) See review by: Payne, P.; Allen, L. C. In *Modern Theoretical Chemistry, Applications of Electronic Structure Theory*; Schaefer, H. F., Ed.; Plenum: New York, 1987; Chapter 2.

(55) Miller, M., Personal communication.

Table 4. Results for Alcohols and Ethers (Energies in kcal/mol, Angles in deg)

parameter	this work	MM3[a]	experiment
	Dimethyl Ether		
ΔE(eclipsed−staggered)	2.74	2.45	2.72[b]
θ(C−O−C)(staggered)	112.3	111.9	111.8[b]
θ(C−O−C)(eclipsed)	113.3		
	Tetrahydrofuran		
$\Delta E(C_s-C_2)$	0.12	0.094	0 ± 0.3[c]
$\Delta E(C_{2v}-C_2)$	3.98	4.41	3.5[d]
structural parameters			
C_2 conformation			
q[e]	0.40		0.39[c]
θ(C−O−C)	108.8	108.7	110.5[c]
θ(C−O−O)	106.8	106.7	106.5[c]
θ(C−C−C)	100.4	101.1	101.8[c]
C_s conformation			
q[e]	0.38		0.364,[c] 0.38[f]
θ(C−O−C)	105.4	104.0	106.2[c]
θ(C−O−O)	105.1	105.0	105.0[c]
θ(C−C−C)	103.6	103.6	104.1[c]
	Methyl Ethyl Ether		
ΔE(gauche-trans)	1.46	1.49	1.5 ± 0.2[g]
ΔE(cis-trans)	6.46	6.02	7.01[h]
structural parameters			
ϕ(gauche)	76.0	74.5	84 ± 6[i]
θ(C−O−C)(trans)	112.3	112.1	111.7[j]
θ(C−C−O)(trans)	108.3	108.7	108.9[k]
	Methanol		
ΔE(eclipsed−staggered)	1.03	0.78	1.06[k]

[a] Reference 56. [b] Reference 57. [c] Reference 58. [d] Reference 59. [e] Reference 60. [f] Reference 61. [g] Reference 62. [h] *Ab initio* MP2/6-31G*//HF/6-31G* calculations. [i] Reference 63. [j] Reference 64. [k] Reference 65.

Table 5. Dimethyl Phosphate Energies, Structures, and Low-Frequency Vibrational Modes

Relative Energies[a] (kcal/mol)

conformation[b]	E(MM)	E(QM)
g,g	0.00	0.00
g,t	1.42	1.41
t,t	2.83	3.45

Geometrical Parameters (Angles in deg)[b,c]

	MM	QM	X-ray[d]
ϕ_1,ϕ_2(g,g)	67.7, 67.7	75.2, 75.2	73, 73
ϕ_1,ϕ_2(g,t)	74.2, 179.2	73.7, 189.4	74, 169
θ(COP)(g,g)	122.1	118.5	121.7
θ(OPO)(g,g)	103.8	99.3	104.8
θ(OPO′)(g,g)	108.2	107.5	110.6
θ(O′PO′)(g,g)	119.3	124.9	119.7
θ(COP)(g,t)	120.5	118.0	
θ(OPO)(g,t)	102.5	96.7	
θ(O′PO)(g,t)	108.2	108.5	
θ(O′PO′)(g,t)	120.1	122.8	
θ(COP)(t,t)	120.2	116.5	
θ(OPO′)(t,t)	103.0	94.3	
θ(O′PO)(t,t)	108.2	109.6	
θ(O′PO′)(t,t)	119.9	120.9	

Vibrational Frequencies < 500 cm^{-1} (cm^{-1})

MM	exp[e]	MM	exp[e]	MM	exp[e]
78		262	210	359	357
109		295	321	383	393
196	195	302	345	421	503

[a] Absolute energies for g,g conformations are −40.77 kcal/mol (MM) and −720.606019 au (QM). The quantum mechanical calculations used the model MP2//6-31G*//HF/6-31G*. [b] Dihedral angles around C−O−P−O. [c] Bond angles, O is ester oxygen and O′ is anionic oxygen. [d] See Table 4 in ref 5. [e] Reference 66.

experiment. The parameters in MM3 were derived by fitting to a wide variety of data for hydrocarbons, whereas our approach is to start with ethane as the simplest model and add additional dihedral parameters in a conservative way. As one can see, the barriers and geometry of *n*-butane are well described with such a model, as is the energy to eclipse the first and second methyl group of propane with the methylene.

We next turn to the alcohols and ethers (Table 4). Here we begin with only two general V_3 dihedrals, as in Weiner *et al.*,[5,6] for X−CT−OH−X and X−CT−OS−X. This leads to essentially exact reproductions of the dihedral barriers in methanol and dimethyl ether. The cis-trans energy difference is about 0.5 kcal/mol greater than that calculated by the Weiner *et al.* force field; however, the Weiner *et al.* value matched the experimental data originally used. When these dihedral parameters are applied to methyl ethyl ether (MEE) and tetrahydrofuran (THF), one finds that a small V_2(CT−CT−OS−CT) dihedral of 0.1 kcal/mol (Weiner *et al.* had such a parameter with magnitude 0.2 kcal/mol) leads to an excellent reproduction of the g/t energy difference in MEE and a slight preference for C_2 THF over C_s, as inferred from experiments. The calculations overestimate the barrier to planarity of THF, but not by as much as MM3.

We next turn to dimethyl phosphate, the model for the backbone of nucleic acids. We have carried out *ab initio* calculations (MP2/6-31G*//HF/6-31G*) on dimethyl phosphate in its g,g; g,t; and t,t conformations and adjusted the V_2(OS−P−OS−CT) parameter to reproduce the (g,g)/(g,t) energy difference of 1.41 kcal/mol. These results are reported in Table 5. The reoptimized V_2 parameter has a value of 1.20 as opposed to the value of 0.75 determined by Weiner *et al.* with the V_3 parameter of 0.25 left unchanged. Reasonable agreement with *ab initio* calculations and consensus structural values from X-ray data has been achieved. The normal mode frequencies calculated with such a model are also compared with those developed based on experimental frequencies of diethyl phosphate.[66] Given the difference in molecules, the agreement between calculation and experiment for the low-frequency modes reported in Table 5 is acceptable.

The low-frequency modes for the simple hydrocarbons, alcohols, ethers, and thio compounds are presented in Table 6. The average error between the calculated and experimental frequencies is 31 cm^{-1} for the 36 low-frequency examples where experimental data are available, compared to an error of 21 cm^{-1} with MM3. Again, it should be noted that our parameters have been optimized using this limited set of simple molecules whereas the test set of molecules used to derive the MM3 parameters is much larger.

Next to consider in the development of a force field for nucleic acids are the bases. Elsewhere, we have reported the

(56) Allinger, N. L.; Rahman, M.; Lii, J.-H. *J. Am. Chem. Soc.* **1990**, *112*, 8293−8307.
(57) Blukis, U.; Kasei, P. H.; Myers, R. J. *J. Chem. Phys.* **1963**, *38*, 2753−2760.
(58) Almenningen, A.; Seip, H. M.; Willadsen, T. *Acta Chem. Scand.* **1969**, *23*, 2748−2745.
(59) Engerholm, G. G.; Luntz, A. C.; Gwinn, W. D.; Harris, D. O. *J. Chem. Phys.* **1969**, *50*, 2446−2457.
(60) Cremer, D.; Pople, J. *J. Am. Chem. Soc.* **1975**, *97*, 1354−1358.
(61) Geise, H.; Adams, W.; Bartell, L. *Tetrahedron* **1969**, *25*, 3045−3052.
(62) Kitagawa, T.; Miyazawa, T. *Bull. Chem. Soc. Jpn.* **1968**, *41(8)*, 1976−1976; MP2/6-31G*//HF/6-31G* *ab initio* calculations lead to ΔE = 1.4 kcal/mol.
(63) Oyanagi, K.; Kutchitsu, K. *Bull. Chem. Soc. Jpn.* **1978**, *51*, 2237−2242.
(64) Hayashi, M.; Adachi, M. *J. Mol. Struct.* **1982**, *78*, 53−62.
(65) Lees, R. M.; Baker, J. G. *J. Chem. Phys.* **1968**, *48*, 5299−5318.
(66) Brown, E.; Peticolas, W. *Biopolymers* **1975**, *14*, 1259−1271.
(67) Allinger, N. L.; Quinn, M.; Rahman, M.; Chen, K. *J. Phys. Org. Chem.* **1991**, *4*, 647−658.
(68) Allinger, N. L.; Quinn, M.; Rahman, M.; Chen, K. *J. Phys. Org. Chem.* **1991**, *4*, 659−666.

Table 6. Low-Frequency (<1000 cm^{-1}) Vibrational Modes for Small Hydrocarbons, Ethers, Alcohols, and Sulfur Compounds

symm	ϑ(this work)	ϑ(MM3)[a]	ϑ(exp)[b,c]	mode[d]
		Ethane		
A_{2u}	312	279	283	CH_3-CH_3 torsion
E_{2u}	811	908	822	CH_3 asym rocking
E_{2u}	811	908	822	CH_3 asym rocking
A_{1g}	898	962	995	C−C stretch
		Propane		
A_2	231	208	217	CH_3-CH_2 torsion
B_2	275	255	265	CH_3-CH_2 torsion
A_1	356	375	379	C−C−C bend
B_2	733	803	748	CH_2 rock + CH_3 def
A_1	809	850	868	CH_3 rock + sym C−C str/str
B_1	866	938	921	CH_3 rock + asym C−C str/str
A_2	877	961	899	CH_2 twist + CH_3 def
		Butane		
A_u	127	122	121	CH_2-CH_2 torsion
A_u	236	216		CH_3-CH_2 torsion
B_g	271	245	266	CH_3-CH_2 torsion
B_u	272	287		asym C−C−C bend + C−C−C bend
A_g	364	394	427	sym C−C−C bend + C−C−C bend
		Methanol		
A″	297	263	270	CH_3−O torsion
A′	867	1052	1034	C−O stretch
		Dimethyl Ether		
A_2	212	188	198	CH_3−O sym torsion
B_2	279	273	242	CH_3−O asym torsion
A_1	416	400	424	C−O−C bend
A_1	798	924	918	C−O sym stretch
		Methyl Ethyl Ether		
A″	123	114		C_2H_5−O torsion
A″	225	216		CH_3−C torsion + CH_3−O torsion
A″	271	257	238	CH_3−O torsion + CH_3−C torsion
A′	283	296	308	C−O−C bend + C−C−O bend
A′	404	420	472	C−C−O bend + C−O−C bend
	755	870	820	CH_3 rock + CH_2 rock + CH_2 twist
	806	897	855	C−O str + CH_3 wag + C−C str
		Methanethiol		
	707	695	704	C−S
	801	823	803	C−S−H
		Dimethyl Sulfide		
	279	285	282 (285)	C−S−C
	691	683	691 (683)	S−C sym
	720	702	741 (704)	S−C asym
		Dimethyl Disulfide		
	105	116	102 (106)	C−S−S−C torsion
	236	241	239 (242)	S−S−C bend
	275	279	272	S−S−C bend
	509	514	509 (514)	S−S stretch
	710	701	689 (694)	S−C stretch
	713	703	(694)	S−C stretch

[a] References 2, 56, 67, and 68. [b] See references 2, 5, 56, 67, and 68 for experimental frequencies. [c] Experimental frequencies given in parentheses refer to those used as reference for MM3 values. [d] See references 2, 5, 56, 67, and 68 for the mode assignments.

hydrogen bond energies and structures of A:T and G:C pairs and these appear to be in good agreement with the highest level of *ab initio* data currently available.[69] However, a critical element in the development of planar functionalities such as the bases is the dihedral potential for out-of-plane motion, as discussed by Weiner *et al.* As in the development of our previous force field, normal mode analyses of benzene and NMA are important. The results for the normal mode analyses applied to these molecules are presented in Table 7. We have readjusted the X−CA−CA−X V_2 value and the improper out-of-plane dihedral X−X−CA−HA to ensure correct representation of the lowest frequency modes of benzene, with the four lowest modes (<700 cm^{-1}) in good agreement with experiment.[71]

We next turn to NMA, the model for the peptide backbone. With a few adjustments to the Weiner *et al.*[5,6] bonded parameters, the agreement with experiment[70] for the six lowest frequency modes is again excellent. In NMA, a key adjustment

Table 7. Normal Modes of *trans*-NMA and Benzene (cm^{-1})

nmode no.	symm	this work	experiment[a]	mode
		trans-NMA		
1	A″	44		
2	A″	97		
3	A″	184	192	
4	A′	286	289	
5	A′	440	439	
6	A″	587	600	
7	A′	591	628	
8	A″	696	725	
9	A′	801	883	
10	A′	963	991	
11	A″	1037	1044	
12	A″	1046		
13	A′	1075	1114	
14	A′	1082	1161	
15	A′	1209	1300	
16	A′	1395	1374	
17	A′	1398	1414	
18	A″	1402	1441	
19	A″	1407	1451	
20	A′	1428	1458	
21	A′	1516	1471	
22	A′	1614	1569	
23	A′	1693	1660	
24	A′	2868	2935	
25	A′	2869	2935	
26	A″	2980	2981	
27	A′	2982	2981	
28	A′	2982	2994	
29	A′	2983	2994	
30	A′	3304	3307	
		Benzene		
1	e2u	410	410	ring def
2	e2g	609	606	ring def
3	a2u	661	673	CH bend
4	b2g	704	703	ring def
5	e1g	900	849	CH bend
6	e2u	979	975	CH bend
7	a1g	941	992	ring stretch (breathing)
8	b2g	947	995	CH bend
9	b1u	1167	1010	ring def
10	e1u	1124	1038	CH bend
11	b2u	1194	1150	CH bend
12	e2g	1129	1178	CH bend
13	b2u	1331	1310	ring stretch (kekule)
14	a2g	1729	1326	CH bend
15	e1u	1493	1486	ring stretch + def
16	e2g	1706	1596	ring stretch
17	e2g	3064	3047	CH stretch
18	a1g	3062	3062	CH stretch
19	e1u	3064	3063	CH stretch
20	b1u	3068	3068	CH stretch

[a] Reference 70 for *trans*-NMA. Reference 71 for benzene.

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

(70) Rey-Lafon, M.; Ford, M. T.; Garrigen-Lagrange, C. *Spectrochim. Acta, Part A* **1973**, *29A*, 471−486.

(71) Shimananchi, T. *Tables of Molecular Vibrational Frequencies*; National Stand. Ref. Data Ser.; National Bureau of Standards: Washington, DC, 1967; Parts 1−3.

was the V_1(H−N−C−O) dihedral potential, which, given the change in electrostatic and non-bonded parameters from Weiner *et al.*, had to be modified from 0.65 to 2.00 kcal/mol to ensure that the in vacuo cis/trans NMA energy difference was ~2.3 kcal/mol.

The re-optimized X−CA−CA−X parameter was used to interpolate new V_2 dihedral potentials for X−C−N−X and X−C−C−X dihedrals in conjugated rings. The normal mode frequencies for the four nucleic acid bases—guanine, adenine, cytosine, and thymine—were then calculated. The calculated and experimental[72-75] frequencies for modes ~<600 cm^{-1} are reported in Table 8. The agreement is qualitatively reasonable; in particular, the cost of out-of-plane distortion is approximately correct in these lowest frequency modes.

We then proceeded to the study of a larger fundamental unit of nucleic acids, deoxy adenosine nucleoside (dA). Table 9 presents the results of calculations of the energy of dA as a function of sugar pucker and the dihedral angles γ(C5′−O5′−C4−C3′) and χ(O1′−C1′−N9−C4), using both a pure gas phase ($\epsilon = 1$) and an implicit solvent ($\epsilon = 4$) model. Although this force field is primarily intended for use with explicit solvent, calculations by Sun *et al.* on conformational free energies of 18-crown-6 suggest that a model with $\epsilon = 4$ provides an approximate and qualitatively reasonable representation of aqueous free energies.[78]

Encouragingly, the C2′ endo/C3′ endo energy difference is 0.6−1.0 kcal/mol, in good agreement with experiment.[79] The barrier between these conformations through the O1′ endo conformation is ~1.9−2.9 kcal/mol, somewhat larger (and perhaps more realistic) than that found by Weiner *et al.* The barrier through O1′ exo is not ϵ dependent and is ~5.9 kcal/mol, which is in reasonable agreement with what is known. Experimentally, it is known that a γ in the g^+ range is preferred for nucleosides in solution, followed by trans, with little g^- observed.[79] The relative conformational energies with $\epsilon = 4$ are quite consistent with this trend, whereas the gas-phase values ($\epsilon = 1$) are not.

Finally, adenosine and deoxyadenosine are known to prefer the anti conformation[79] over the syn conformation, but the syn conformation is low enough in free energy to be observable. The gas-phase ($\epsilon = 1$) energy difference between anti and syn is very large, but the $\epsilon = 4$ value is much more reasonable. However, we wished to assess the reasonability of our calculated energies as a function of χ with an *ab initio* model. We thus constructed a simple test case where adenine is attached to CH(OH)−CH_3, with the dihedrals constrained to mimic the C2′ endo conformation of a sugar ring (Figure 1) and carried out MP2/6-31G*//HF/6-31G* *ab initio* calculations as a function of χ with this model. As one can see from Table 10, with no additional dihedral parameters, the energy difference between the syn and anti minima is significantly overestimated with our initial model. We thus chose to add explicit dihedrals (V_1 and V_2) (see Table 14) around the glycosidic bond to bring the two minima into qualitative agreement. This has very little effect on the γ and sugar pucker energies, so only the values of the final parameter set are reported in Table 9.

We next turned to studies of peptide conformations. Table 11 presents the local minima and Figures 2a and 2b the (ϕ,ψ) maps for glycyl and alanyl dipeptides. Here, as in the case of glycosidic χ, we were forced to add explicit dihedral parameters (see Table 14) in order to reproduce the *ab initio* quantum mechanical energies for these models. As one can see, the

Table 8. Low-Frequency Normal Modes of the Bases (cm^{-1})

nmode no.	this work	experiment[a]	mode[a]
		Adenine	
in-plane vib			
1	467	337	
2	529	540	
3	556	620	
4	667	665	
out-of-plane vib			
1	193	184	ring torsion
2	248	194	ring torsion
3	290	238	ring torsion
4	292	310	ring torsion
5	456	331	C6−N6 torsion
6	553	550	C6−N6 wag
7	569	624	
8	623	655	
9	672	686	
		Cytosine	
in-plane vib			
1	320	400	C−NH_2 bend
2	518	533	C=O bend
3	539	549	ring def
4	659	600	ring def
out-of-plane vib			
1	202	197	C−NH_2 wag
2	217	232	C=O wag
3	414	421	ring def
4	493	485	NH_2 wag
5	536	548	NH_2 rock
6	602	566	ring def
		Guanine	
in-plane vib			
1	301	343	C−NH_2 bend
2	344	400	C=O bend
3	505	501	ring def (py)
4	534	557	ring def (py)
5	554	645	ring def (py)
6	644	690	ring def (Im)
out-of-plane vib			
1	140	142	C−NH_2 wag
2	185	170	C=O wag
3	233	214	ring (butterfly) def
4	300	243	ring (propeller) def
5	445	416	ring (py) def
6	448	490	ring (Im) def
7	534	601	ring (Im) def
8	593	654	NH_2 rock
9	674	690	ring (py) def
		Thymine	
in-plane vib			
1	348	321	C−CH_3 bend
2	372	392	C=O bend (out-of-phase)
3	463	475	ring def
4	549	560	ring def
5	592	617	C=O bend (in-phase)
out-of-plane vib			
1	43		CH_3 rot
2	132		C−CH_3, C=O wag
3	188	206	C=O wag
4	332	285	C−CH_3 wag
5	454	433	ring def
6	599	635	ring def

[a] Reference 72 for adenine. Reference 73 for cytosine. Reference 74 for guanine. Reference 75 for thymine.

(72) Dhasuadi, Z.; Ghomi, M.; Austin, Y. C.; Girling, R. B.; Hester, R. E.; Mojres, P.; Chinosky, L.; Tarpin, P. Y.; Coulombeau, C.; Yobic, H.; Tomhinson, Y. *J. Phys. Chem.* **1993**, *97*, 1074−1084.
(73) Susi, H.; Ard, Y. S.; Purcell, Y. M. *Spectrochim. Acta* **1973**, *29A*, 725−753.
(74) (a) Delaber, J.-M.; Majoube, M. *Spectrochim. Acta* **1978**, *37A*, 129−140. (b) Delabar, J.-M. *J. Raman Spectrosc.* **1978**, *7*, 261−267.
(75) Beetz, C. P., Jr.; Ascarelli, G. *Spectrochim. Acta* **1980**, *36A*, 299−313.
(76) Cremer, D.; Pople, J. *J. Am. Chem. Soc.* **1975**, *97*, 1354−1358, 1358−1366.
(77) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: Tokyo, 1984.
(78) Sun, Y.; Kollman, P. A. *J. Chem. Phys.* **1992**, *97*, 5108−5112.
(79) Davis, D. R. *Prog. Nucl. Magn. Reson. Spectrosc.* **1978**, *12*, 135−225.

Table 9. Conformational Energies for Deoxyadenosine (Angles in deg, Energies in kcal/mol)[a]

pucker	q[a]	W[a]	γ[b]	χ[b]	3'OH[c]	5'OH[d]	E[e]	ΔE[f]
Sugar Pucker Profile								
				$\epsilon = 1$[g]				
C2'endo	0.40	146.1	51.3	−158.4	176.5	171.4	−52.40	0
C3'endo	0.37	5.7	56.9	−162.3	−178.3	−179.3	−51.87	0.63
O4'endo	0.38	65.3	54.1	−156.7	−175.5	−175.4	−49.53	2.87
O4'exo	0.29	276.6	42.6	−178.7	175.8	178.5	−46.54	5.86
				$\epsilon = 4$[g]				
C2'endo	0.39	144.9	55.3	−153.1	176.2	−179.2	−1.65	0
C3'endo	0.38	14.1	56.3	−156.1	178.1	179.9	−0.61	1.04
O4'endo	0.39	56.6	55.6	−153.1	179.1	179.5	0.21	1.86
O4'exo	0.30	285.0	42.7	176.9	177.3	179.8	4.03	5.68
Gamma Dependence								
				$\epsilon = 1$[g]				
C2'endo	0.40	146.1	51.3	−158.4	176.5	171.4	−52.40	0
C2'endo	0.42	141.9	−168.6	−168.6	179.6	−179.9	−50.39	2.01
C2'endo	0.42	151.7	−62.3	−169.9	−179.7	−179.7	−50.92	1.48
				$\epsilon = 4$[g]				
C2'endo	0.39	144.9	55.5	−153.1	176.2	179.2	−1.65	0
C2'endo	0.40	148.0	−172.9	−162.4	180.0	−179.9	−1.31	0.34
C2'endo	0.40	150.0	−66.6	−169.9	−179.9	−179.9	−0.13	1.52
χ Dependence								
				$\epsilon = 1$[g]				
C2'endo	0.40	146.1	51.3	−158.4	176.5	171.4	−47.46	0
C2'endo	0.40	166.7	63.3	60.8	−179.1	179.8	−41.52	5.94
				$\epsilon = 4$[g]				
C2'endo	0.39	145.4	55.3	−141.6	176.5	179.2	3.24	0
C2'endo	0.40	144.0	52.6	37.6	180.0	179.6	1.84	−1.40

[a] q and W defined in ref 76. [b] γ and χ defined in ref 77. Above, the first entry corresponds to an "anti" conformation for χ, the second to "syn". [c] 3'OH refers to C4'−C3'−O3'−HO3' dihedral. [d] 5'OH refers to HO5'−O5'−C5'−C4' dihedral. [e] Absolute molecular mechanical energy. [f] Relative conformational energy. [g] ϵ is the dielectric screening factor used in eq 1.

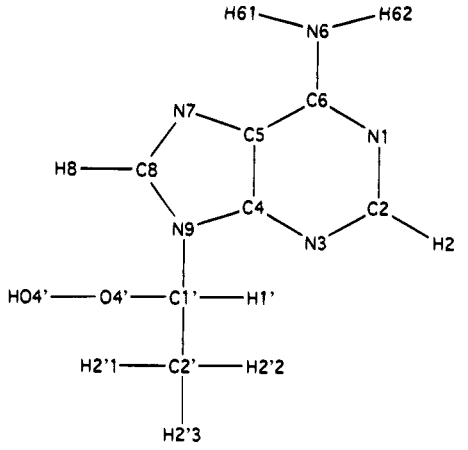


Figure 1. Model of deoxyadenosine employed in the quantum mechanical and molecular mechanical conformational studies reported in Table 10. In the quantum mechanical calculations, the HO4'−O4'−C1'−N9 and H2'3−C2'−C1'−N9 dihedrals were held fixed at values characteristic of a C2'-endo sugar, in order to mimic the conformation of the sugar ring. In the molecular mechanical calculations, the dihedrals were restrained to those values with dihedral restraints of 500 kcal/mol.

agreement with high-level *ab initio* data is very good for all but alanyl dipeptide C_{7ax} and glycyl dipeptide α_R. The ala C_{7ax} conformation is rarely found in proteins and gly occurs relatively infrequently in α-helices, due to the loss of conformational entropy, so these conformations were reasonable ones in which to tolerate any error.

Table 10. χ Angle Profile for Base with Sugar Fragment (kcal/mol)

		AMBER ($\epsilon = 1$)	
χ[a,b]	*ab initio* MP2/6-31G*//HF/6-31G*	no specific dihedral	with specific dihedral[c]
	Model of Deoxyadenosine		
60	0.94	4.63	1.53
min	0.63 (74.7°)[d]	4.62 (61.1°)[d]	1.45 (54.5°)[d]
120	3.37	5.48	5.07
180	0.06	0.38	0.40
min	0.00 (198.2°)[d]	0.00 (196.7°)[d]	0.00 (197.2°)[d]
210	0.22	0.20	0.15
240	1.45	1.58	1.20
300	5.33	6.57	3.61
360		9.68	4.72
	Model of Deoxythymidine		
60	2.27	6.00	2.94
min	2.02 (72.3°)[d]	5.99 (61.0°)[d]	2.83 (55.4°)[d]
120	7.02	8.48	8.05
180	0.74	1.37	1.43
min	0.00 (210.0°)[d]	0.00 (205.2°)[d]	0.00 (205.5°)[d]
210	0.00	0.05	0.03
240	1.29	1.77	1.40
300	8.15	8.94	5.82
360		13.11	8.18

[a] Reference 77. [b] Degrees. [c] Specific V_1 and V_2 dihedral terms were added for OS−CT−N*−CK (purines) and OS−CT−N*−CM (pyrimidines) dihedral angles. [d] Minimized value of χ.

Table 11. Conformational Energies of Glycyl and Alanyl Dipeptides (kcal/mol)

glycyl dipeptide			alanyl dipeptide		
	E(MM)	E(QM)[a]		E(MM)	E(QM)[a]
C7	0.0	0.0	$C7_{eq}$	0.0	0.0
C5	1.9	2.0	$C7_{ax}$	1.5	2.1
α_R	6.0	4.0	C5	1.5	1.5
			α_R	3.9	3.9

[a] Quantum mechanical energies calculated at the MP2/TZP//HF/6-31G* level on methyl-blocked versions of the dipeptides. See ref 28 for further details.

One of the important features in our force field is the attempt to reproduce the solvation free energies of a representative set of molecules. In Table 12, we present such a representative set. As one can see, the absolute solvation free energy of methane is somewhat (0.5 kcal/mol) too large with our model, but the relative solvation free energies of methane, ethane, and propane are within 0.3 kcal/mol of experiment. For our protypal polar molecules, methanol and NMA, the agreement with experimental solvation free energies is within ~0.5 kcal/mol. We wished also to assess the solvation free energies for sulfur compounds and the relative solvation free energies of those are in reasonable agreement with experiment (again within 0.5 kcal/mol). The calculated free energy of 9-methyladenine is a prediction, because there are no precise experiments,[85] but the relative free energies of this force field and that of Weiner *et al.*[5,6] suggest that the experimental determination of this quantity would be of great interest. Turning to the ionic molecules, our results make clear that a typical two-body additive force field will tend to overestimate ion solvation (when corrected for long-range cut-off) unless its parameters are significantly modified, but fully non-additive calculations with exactly the same parameters reproduce experiment very well.

(80) Ben Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016−2027.
(81) Wolfenden, R. *Biochemistry* **1978**, *17*, 201−204.
(82) Meng, E.; Cieplak, P.; Caldwell, J.; Kollman, P. Accurate Solvation Free Energies of Acetate and Methylammonium Calculated with a Polarized Water Model. *J. Am. Chem. Soc.*, in press.
(83) Analyzed by Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305−8311.
(84) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 292−298.
(85) Ferguson, D. M.; Pearlman, D. A.; Swope, W. C.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 362−370.

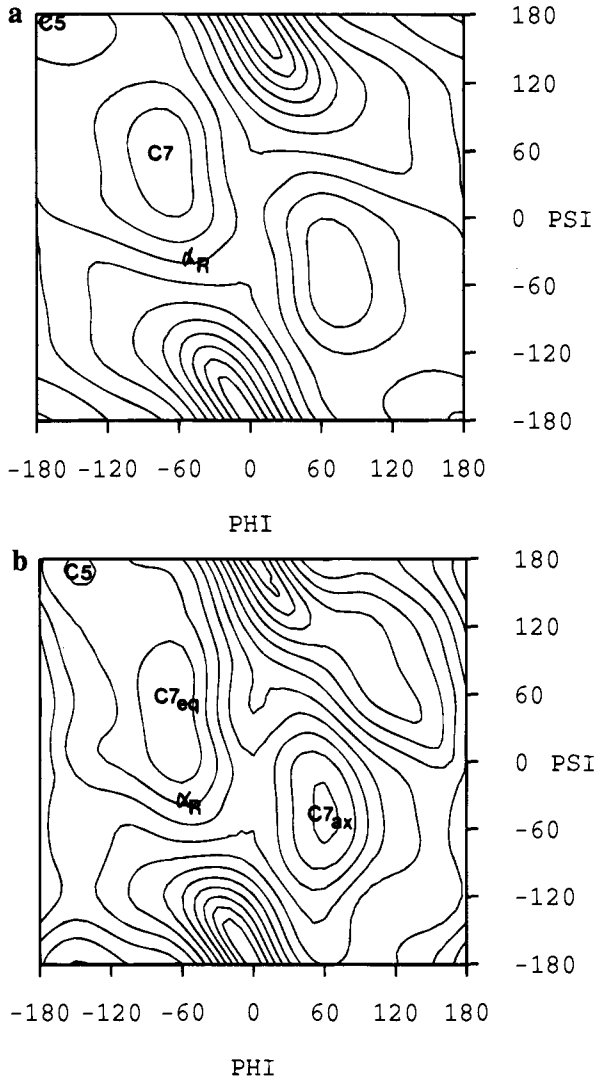


Figure 2. (a) The molecular mechanical (ϕ,ψ) map for methyl-blocked glycyl dipeptide generated using the force field presented here. Contours are drawn every 2 kcal/mol. (b) The molecular mechanical (ϕ,ψ) map for methyl-blocked alanyl dipeptide generated using the force field presented here. Contours are drawn every 2 kcal/mol.

Table 12. Solvation Free Energies for Model Compounds (kcal/mol)

molecule	$\Delta\Delta G$(calc)	$\Delta\Delta G$(exp)
$CH_4 \rightarrow$ nothing	-2.5 ± 0.1^a	-2.0^b
$C_2H_6 \rightarrow CH_4$	-0.1 ± 0.1^c	0.2^b
$C_3H_8 \rightarrow C_2H_6$	-0.2 ± 0.1^c	-0.2^b
$CH_3OH \rightarrow CH_3CH_3$	6.9 ± 0.1^d	6.9^b
NMA $\rightarrow CH_4$	11.6 ± 0.2^d	12.1^e
$CH_3NH_3^+ \rightarrow$ nothing	87.6 ± 2.0 (75.4 ± 1.7)^f	$77-79^g$
$CH_3CO_2^- \rightarrow$ nothing	87.1 ± 1.2 (71.6 ± 1.0)^f	$70-71^h$
$CH_3SCH_3 \rightarrow CH_3OCH_3$	0.9 ± 0.1^h	0.4^i
$CH_3OH \rightarrow CH_3SH$	3.5 ± 0.1^h	3.7^i
9-CH_3 adenine $\rightarrow CH_4$	18.3 ± 2.6,[j] 13.9 ± 0.4^j	15.6 ± 1.1^j

[a] Reference 21b. [b] Reference 80. [c] Reference 21a. Because of the uncertainty in the electrostatic potential derived charges for ethane, the average of the free energies for the electrostatic potential derived and Mulliken charges for these free energy calculations are presented. [d] Reference 20. [e] Reference 81. [f] Reference 82, additive potential, values in parentheses are for nonadditive potential. [g] Reference 83. [h] This paper. [i] Reference 84. [j] See Note Added in Proof.

The results described above were obtained on model systems that were relatively very simple[37] (neat liquids) and/or small (dipeptides and nucleosides). In order to test the performance of the new force field on a more complex system, we carried out an MD simulation of ubiquitin in water with periodic boundary conditions. The RMS difference was calculated for structures along the trajectory relative to the crystal structure[86] for (1) the backbone atoms and (2) all of the heavy atoms. These results were then compared to those obtained with the Weiner *et al.*[5,6] force field (Figure 3). The RMS values are reported for the first 72 residues only, since the four residues of the carboxy terminus were mobile. The behavior of the new force field presented here is better in two ways. First, the protein structure seems to have stabilized after 50 ps of simulation with the new force field, while the RMS deviation continues to increase throughout the trajectory with the Weiner *et al.*[5,6] force field. Second, the RMS deviation for all of the heavy atoms after 180 ps of simulation is about 2.0 Å with the force field presented here and about 2.5 Å with the Weiner *et al.*[5,6] force field. Alonso and Daggett have also reported the results of a long MD simulation of ubiquitin, and they found a backbone RMS deviation of 1.4 Å from the crystal structure, comparable to the deviation found here.[87] A referee has pointed out that smaller deviation from a crystal structure could simply be a consequence of an "unrealistically stiff" force field. We cannot rule this out, but stress that we did not, in our force field derivation on the fragments described above attempt to add "stiffness".

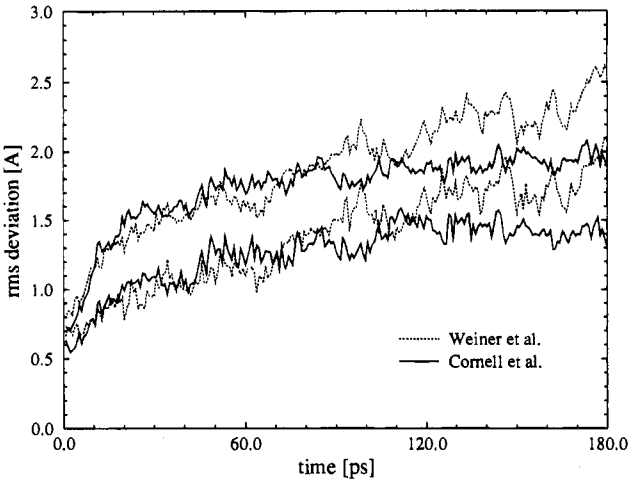
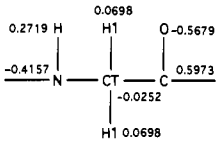


Figure 3. RMS deviation (Å) between the crystal structure of ubiquitin and structures along an MD trajectory as modeled by the Weiner *et al.*[6] and Cornell *et al.* (this work) force fields. The lower lines correspond to the RMS deviation of the heavy backbone atoms only and the upper lines to the RMS deviation for all heavy atoms.

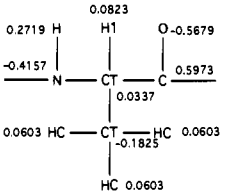
Even closer agreement with a protein crystal structure has been obtained by York *et al.*,[11] who carried out a 1000 ps MD simulation of BPTI with the long-range electrostatic forces of the crystal environment treated using the particle mesh Ewald method and the Weiner *et al.*[5,6] force field. With this model they obtained an RMS deviation from the crystal structure of 0.33 Å for backbone atoms. These results serve to illustrate the difference between errors arising from the force field itself and those arising from its implementation in a given calculation. Currently, most MD simulations employ an 8 or 9 Å cutoff for nonbonded interactions in order to reduce this rate-determining part of the calculation. In systems where long-range electrostatics play an important role, this approximation is clearly inadequate. Although the Ewald method is only fully appropriate for periodic crystal systems, other methods also exist which allow for the more accurate treatment of long-range electrostat-

(86) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531–544.

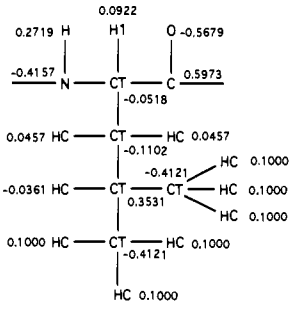
(87) Alonso, D. O. V.; Daggett, V. Molecular Dynamics Studies of Partially Unfolded Conformations of Ubiquitin in Methanol and Their Refolding in Water, submitted for publication.



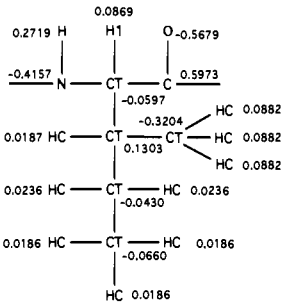
GLY



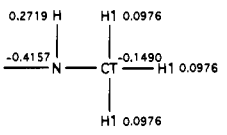
ALA



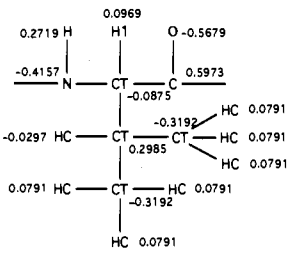
LEU



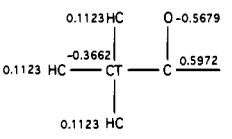
ILE



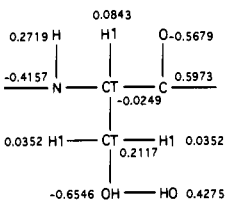
NME



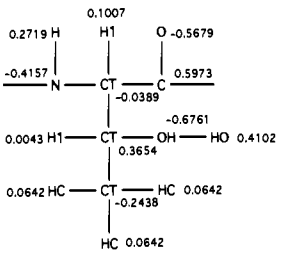
VAL



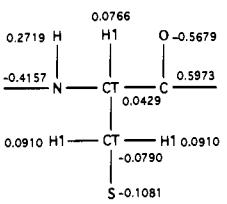
ACE



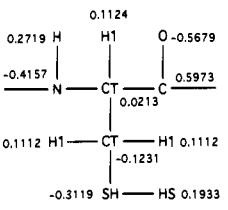
SER



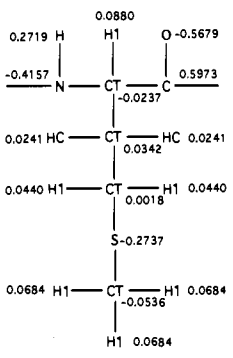
THR



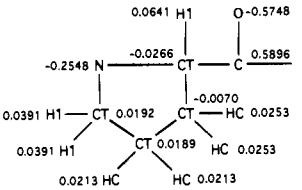
CYX



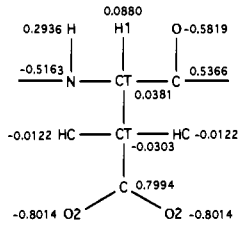
CYS



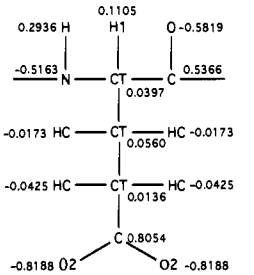
MET



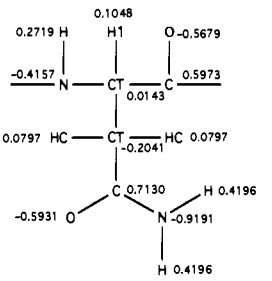
PRO



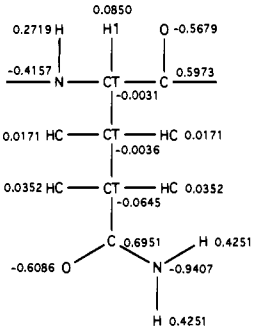
ASP



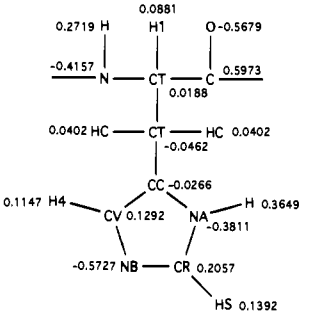
GLU



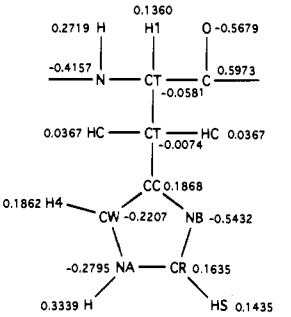
ASN



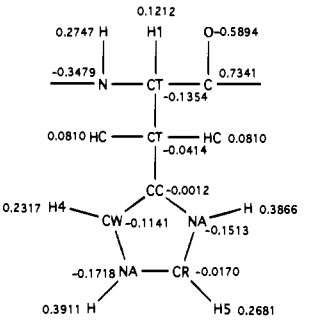
GLN



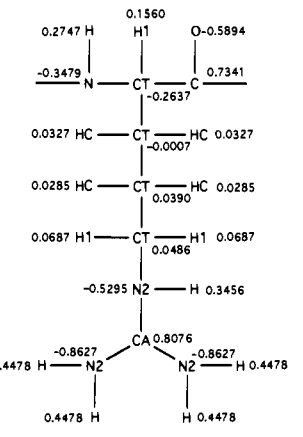
HID



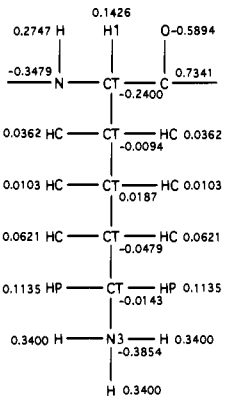
HIE



HIP



ARG



LYS

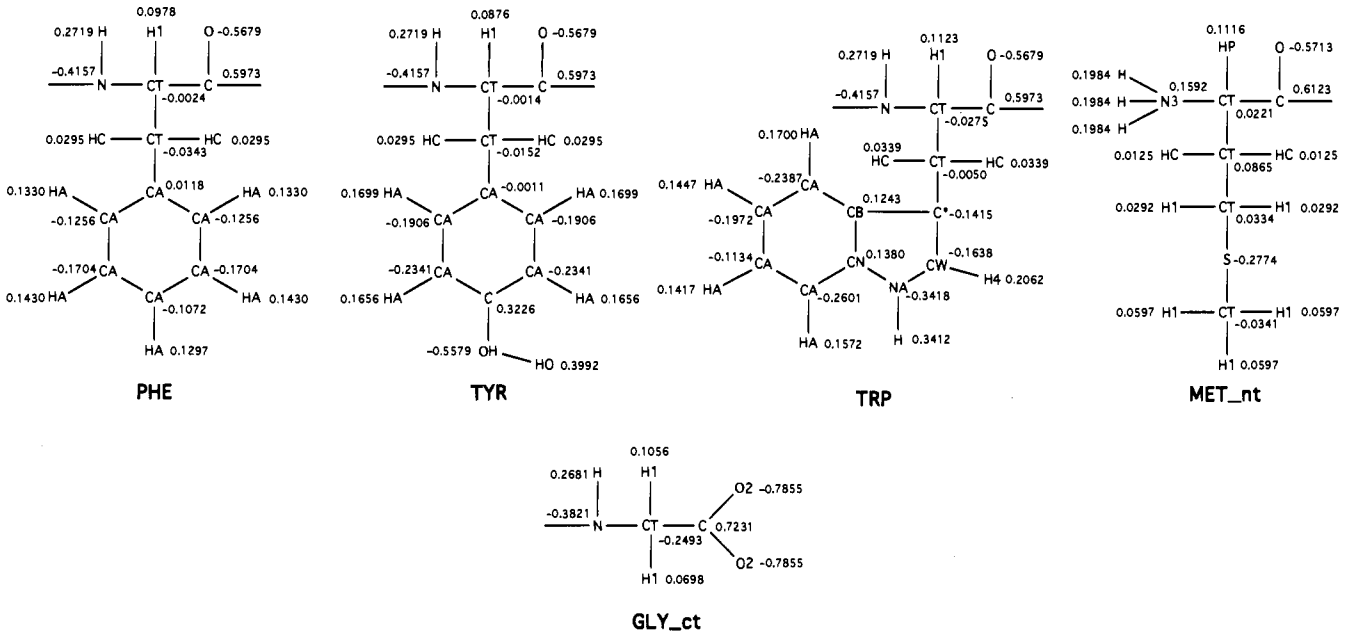


Figure 4. Charges for the peptide fragments. All the charges for the non-terminal amino acids are presented. For histidine, the three protonation states are presented (HID, HIE, HIP). For cysteine, both the disulfide bonded forms (CYX) and the reduced form (CYS) are presented. The N terminal and C terminal blocking groups are presented (ACE and NME, respectively). The N and C terminal amino acids for ubiquitin (MET_nt and GLY_ct) are shown; the remaining N and C terminal residues are available by anonymous ftp as are the protonated forms of GLU and ASP and the deprotonated form of LYS. See ref 39 for a description of how these charges were derived.

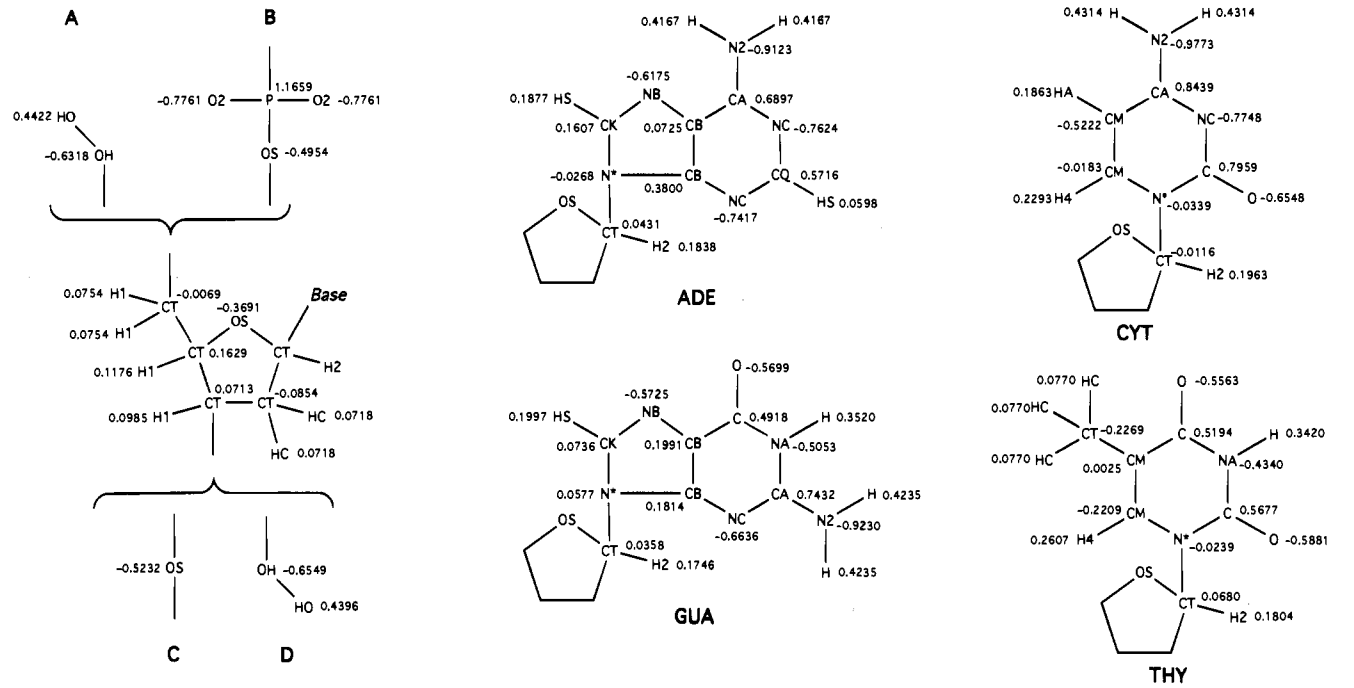


Figure 5. Charges for DNA. The four bases and the C1′ and H1′ charges are shown separately. These are combined with any of four combinations of sugar/backbone charges. A nucleoside corresponds to fragments A and D with the sugar. A 5′ terminal residue corresponds to fragments A and C and the sugar; a 3′ terminal residue corresponds to combining B and D with the sugar and a central residue corresponds to combining B and C with the sugar. See ref 39 for how these charges were derived.

ics.[88] Thus, it appears that the way electrostatic interactions are handled is significantly more important than the detailed force field parameters in ensuring that a molecular dynamics trajectory stays near an experimental (X-ray or NMR) structure. We suggest, however, that comparing two force fields with the same cutoff protocol can be illustrative and we conclude, on that basis, that the new force field performs at least as well as, or slightly better than, that of Weiner *et al.*[5,6] for full solution simulations.

(88) Saito, M. *J. Chem. Phys.* **1994**, *101*, 4055–4061.

Discussion

We have presented the development and the description of a new force field for proteins, nucleic acids, and organic molecules. Previously, we have attempted to give a coherent description of the underlying basis for the Weiner *et al.* force field,[5,6] in order that it could be extended by others as well as ourselves for studies of molecular interactions and conformations. We should emphasize again that our goal is to describe molecular conformational energies and structures as accurately as possible in condensed phases with a simple, transferable, and

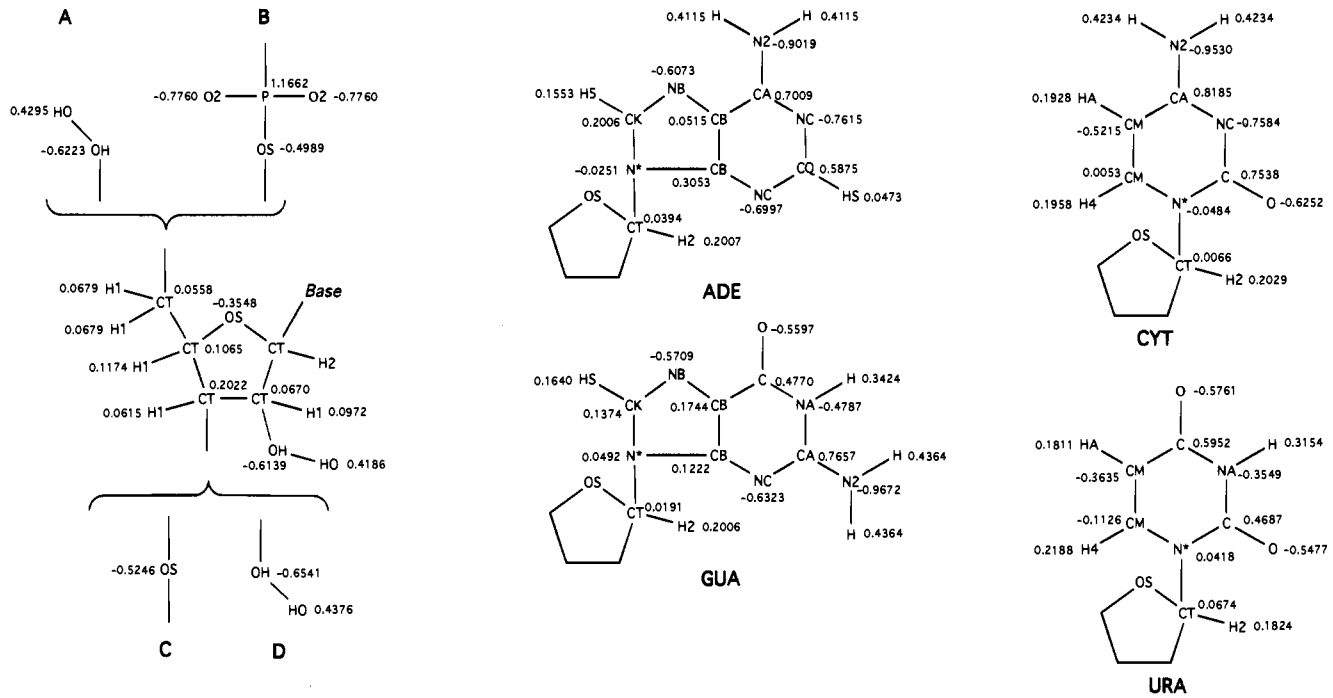


Figure 6. Charges for RNA. Notation is the same as in Figure 5.

Table 13. Comparison of Cornell *et al.*, Weiner *et al.*, CHARMm, OPLS/AMBER, and GROMOS Force Fields

force field	electrostatics	van der Waals	VDW combining rules[a]	torsions
CHARMm[93] (1983)	empirical fit to QM dimers	empirical (x-tals)	R^* arithmetic mean; ϵ geometric mean	single bond path[b]
GROMOS[92]	empirical	empirical (x-tals)	A and B "non-standard" geometric mean[c]	user specified
OPLS/AMBER[15] (1990)	empirical (MC on liqs)	empirical (liquids)	A and B geometric means	equal division among equiv bond paths
Weiner *et al.*[5,6]	ESP fit (STO-3G)	empirical (x-tals)	R^* arithmetic mean; ϵ geometric mean	equal division among equiv bond paths
this work	RESP fit (6-31G*)	empirical (liquids)	R^* arithmetic mean; ϵ geometric mean	equal division among equiv bond paths

[a] $A = \epsilon R^{*12}$ and $B = 2\epsilon R^{*6}$. [b] In CHARMm22, the torsion representation was changed to the more commonly used equal division of the energy along equivalent bond paths. [c] GROMOS employs the geometric mean method for calculating VDW interactions, but for water−methyl interactions, for example, a smaller VDW radius is assumed for the water since it is no longer in a hydrogen bonding interaction. This has been shown to result in a "too hydrophilic" methyl group.[95,96]

general model. This goal has framed our approach, which has been to focus mainly on the electrostatic, VDW, and dihedral energies and use both *ab initio* calculations, empirical liquid and solvation data, and experiment to calibrate the model. However, our approach differs significantly from that of many in building from the ground up with the simplest model and defining relatively few general principles, which are elucidated in the section General Description of the Model above.

We will attempt to summarize the salient features of some of the more commonly used force fields here, in order to compare and contrast our approach with theirs. They can be roughly grouped into four different categories, depending on the nature and complexity of the force field equation: (1) those with rigid or partially rigid geometries, (2) those without electrostatics, (3) simple diagonal force fields, and (4) more complex force fields.

The ECEPP force field of Scheraga[89] employs rigid internal geometries which allow a more efficient exploration of conformational space. This approach has the disadvantage that it can cause certain conformations and conformational barriers to be too high in energy. A second force field which uses only partially rigid geometries is JUMNA,[90] developed by Lavery and co-workers. This force field has been developed for nucleic acids and allows flexibility in the sugar ring but uses mainly internal geometries and keeps the bases rigid.

The SYBYL force field[91] has been developed for the calculation of internal geometries and conformational energies. Because it contains no electrostatic term, it is inappropriate for studying detailed condensed-phase properties. The YETI force field,[92] developed by Vedani and Huhta, is a modification of the Weiner *et al.* force field with highly damped electrostatics and an angular dependent hydrogen bond (and metal ligation) potential added. This approach could be valuable in some modeling situations, where large and difficult to handle electrostatic energies are present, but it is also unlikely to be general and extendable to condensed-phase phenomena.

The category of simple diagonal force fields includes the Weiner *et al.*,[5,6] GROMOS,[93] CHARMm,[94] and OPLS/AMBER[15] force fields. All of these force fields employ a simple harmonic diagonal representation for the bond and angle terms. Descriptions of the nonbonded and dihedral energies are given in Table 13. The Weiner *et al.* force field derived charges from

(89) (a) Roterman, I. K.; Lambert, M. H.; Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1989**, *7*, 421−453. (b) See also: Kollman, P. A.; Dill, K. A. **1991**, *8*, 1103−1107. Gibson, K. D.; Scheraga, *J. Biomol. Struct. Dyn.* **1991**, *8*, 1109−1111.

(90) Lavery, R.; Hartmann, B. *Biophys. Chem.* **1994**, *50*, 33−45.

(91) Clark, M.; Cramer, R. D.; van Oppenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982−1012.

(92) Vedani, A.; Huhta, D. A. *J. Am. Chem. Soc.* **1990**, *112*, 4759−4767.

(93) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulations (GROMOS) Library Manual*; Biomos: Groningen, 1987.

(94) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; Slater, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

fits to the electrostatic potential of a molecule whereas the other two force fields used empirical fits to interaction energies (CHARMm) or liquid and solid state data (GROMOS). The Weiner *et al.*, CHARMm, and GROMOS force fields all employ VDW parameters derived from crystal data, whereas the VDW parameters in the OPLS/AMBER and Cornell *et al.* force fields are derived from liquid simulations. (The OPLS/AMBER and GROMOS force fields specify values for "A" and "B", the repulsive and attractive coefficients, respectively, whereas Weiner *et al.*, Cornell *et al.*, and CHARMm specify values for R^* and ϵ. Some force fields use "C" instead of "B". See Table 13 for the relationship between A, B, ϵ^*, and R^*.) For heteronuclear interactions, the OPLS/AMBER and GROMOS force fields determine values for A and B using geometric mean combining rules. By comparison, Weiner *et al.*, Cornell *et al.*, and CHARMm employ arithmetic mean combining rules for R^* and geometric mean combining rules for ϵ. GROMOS makes the further distinction of using different values for A and B for a particular atom type, depending on the second atom involved in the interaction. This has been shown to result sometimes in anomalous behavior.[95,96]

Two new sets of CHARMm hydrocarbon VDW parameters have recently been published[97,98] and tested by Kaminski *et al.*[99] for their ability to reproduce condensed-phase properties. The CHARMm92[97] parameters resulted in a density for liquid butane which was 63% in error. The CHARMm94[99] parameters performed much better, reproducing the density and heat of vaporization of butane with an average error of 3.2% and 4.5%, comparable to the results obtained with the AMBER parameters reported in ref 21 and used here, where the average error for butane was 1.7% and 3.0%.[37,100] Nonetheless, the CHARMm94 model is more complex, using a different R^* and ϵ for CH_2 and CH_3 carbons. Kaminski *et al.* also reported new all-atom VDW parameters for the OPLS force field, and these were shown to result in average errors of 0.9% and 1.7% for the density and heat of vaporization of ethane, propane, and butane. The OPLS all-atom parameters also performed better at reproducing the relative free energies of solvation of methane, ethane, and propane than the Sun *et al.* parameters.[21] It should be noted that while the OPLS parameters result in the lowest overall error for the systems described/included above, this is achieved at the expense of fitting the neat liquid properties of methane (errors in density and ΔH vaporization ~10%). The main difference between the Sun *et al.*[21] and Kaminski *et al.*[99] parameters is the van der Waals well depths for hydrogen (0.0157 and 0.030 kcal/mol, respectively) and carbon (0.1094 and 0.060 kcal/mol, respectively), with compensating differences between the van der Waals radii. The Sun *et al.* values for carbon are more in line with the magnitude of the well depth for the other first row atoms in the force field presented here. However, we wish to stress that both the OPLS and the Sun *et al.* parameters are appropriate and effective models to use in condensed-phase studies of organic molecules that are not highly strained or have very short nonbonded distances involving hydrogen.

While all five force fields employ a simple Fourier expansion to represent the dihedral energy, some variation is also seen in the assignment of that energy, with Weiner *et al.*, Cornell *et al.*, OPLS/AMBER, and later versions of CHARMm distributing the energy equally among equivalent bond paths (such as the nine HC−CT−CT−HC dihedrals in ethane), and GROMOS allowing user specification of that parameter. In earlier versions of CHARMm the dihedral energy was assigned to only one specific bond path (quartet of atoms).

Finally, the category of "more complex" force fields includes not only the MM2 and MM3 force fields for small molecules[2,3] but also two other force fields. These force fields go beyond the simple diagonal potential function in their inclusion of higher order terms as well as cross-terms for representing bonds and angles. The MM3 force field is the state-of-the-art for modeling organic molecules in the gas phase and has been carefully calibrated to reproduce many properties of these molecules. The focus of MM3 is quite different from that of the force field presented here in that it is not oriented toward the representation of polar and ionic molecules in condensed phases, although, for example, some crystal minimizations were used to calibrate some of the nonbonded parameters. Its complex functional form is necessary for reproducing vibrational frequencies and subtleties of molecular geometries. The use of a 6-exponential nonbonded potential is more accurate than the 6−12 used here, particularly for close contacts such as those found in highly strained organic molecules. The MM2/MM3 model uses a point dipole approach for electrostatic interactions which has often worked well for modeling intramolecular properties but has not been rigorously established as a general model for modeling intermolecular interactions. MM2/MM3 has a large number of dihedral parameters specific to four-atom bond quartets which have been fit to a large set of data.

A second complex force field is the "Class II" one under development by Hagler and co-workers.[101] This force field has a functional form of similar complexity to that of MM2/MM3, but it differs in the extensive use of quantum mechanical energies and gradients for its calibration. The developers of this force field are pioneering new ways of deriving parameters and analyzing molecular interactions. This force field currently suffers, however, from the lack of a general charge model of the same caliber as the other parameters.

The third complex force field is the Merck Molecular Force Field (MMFF) under development by Halgren.[102] The stated purpose of this force field is to be able to handle all of the functional groups of interest in pharmaceutical design. The nonbonded function is a "buffered" 7−14 potential, which Halgren found to give the best fit to rare gas interactions, and an empirical bond dipole model is used to assign partial charges. The key calibration test set is a series of conformational energies calculated at a very high level of ab initio theory (MP4SDQ/TZP//MP2/6-31G*). Thus far, no condensed-phase simulations have been carried out, but they are planned. This approach has the advantage of generality to a large number of molecules, but at the expense of the use of a simple, empirical, generic charge model and a large number of dihedral parameters.

Conclusion

We have described the development of a second generation force field for the simulation of proteins, DNA, and organic molecules primarily in the condensed phase. The strengths of the approach presented here are: (1) the general and algorithmic strategy employed to develop the force field; (2) the emphasis

(95) Mark, A. E.; van Helden, S.; Smith, P. E.; Janssen, L. H. M.; van Gunsteren, W. *J. Am. Chem. Soc.* **1994**, *116*, 6293−6302.
(96) Aqvist, J.; Medina, C.; Samuellson, J. E. *Protein Eng.* **1994**, 385−391.
(97) Smith, J. C.; Karplus, M. *J. Am. Chem. Soc.* **1992**, *114*, 801−812.
(98) Woolf, T. B.; Roux, B. *J. Am. Chem. Soc.* **1994**, *116*, 5916−5926.
(99) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077−13082.
(100) Nagy, J.; Weaver, D. F.; Smith, V. A. A Comprehensive Study of Alkane Non-Bonded Empirical Force Fields. Suggestions for Improved Parameter Sets, submitted for publication.
(101) (a) Maple, J. R.; Dinur, U.; Hagler, A. T. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5350−5354. (b) Dinur, U.; Hagler, A. T. *J. Chem. Phys.* **1989**, *91*, 2949−2958. (c) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *J. Comput. Chem.* **1994**, *15*, 162−182.
(102) Halgren, T. *J. Am. Chem. Soc.* **1992**, *114*, 7827−7843. Halgren, T., results to be submitted for publication.

Table 14. Molecular Mechanical Parameters[a]

Bond Parameters

bond	K_r[b]	r_{eq}[c]	bond	K_r[b]	r_{eq}[c]	bond	K_r[b]	r_{eq}[c]	bond	K_r[b]	r_{eq}[c]
C−CA	469.0	1.409	CA−HA	367.0	1.080	CM−HA	367.0	1.080	CT−S	227.0	1.810
C−CB	447.0	1.419	CA−N2	481.0	1.340	CM−N*	448.0	1.365	CT−SH	237.0	1.810
C−CM	410.0	1.444	CA−NA	427.0	1.381	CN−NA	428.0	1.380	CV−H4	367.0	1.080
C−CT	317.0	1.522	CA−NC	483.0	1.339	CQ−H5	367.0	1.080	CV−NB	410.0	1.394
C−N	490.0	1.335	CB−CB	520.0	1.370	CQ−NC	502.0	1.324	CW−H4	367.0	1.080
C−N*	424.0	1.383	CB−CN	447.0	1.419	CR−H5	367.0	1.080	CW−NA	427.0	1.381
C−NA	418.0	1.388	CB−N*	436.0	1.374	CR−NA	477.0	1.343	H−N	434.0	1.010
C−NC	457.0	1.358	CB−NB	414.0	1.391	CR−NB	488.0	1.335	H−N*	434.0	1.010
C−O	570.0	1.229	CB−NC	461.0	1.354	CT−CT	310.0	1.526	H−N2	434.0	1.010
C−O2	656.0	1.250	CC−CT	317.0	1.504	CT−F	367.0	1.380	H−N3	434.0	1.010
C−OH	450.0	1.364	CC−CV	512.0	1.375	CT−H1	340.0	1.090	H−NA	434.0	1.010
C*−CB	388.0	1.459	CC−CW	518.0	1.371	CT−H2	340.0	1.090	HO−OH	553.0	0.960
C*−CT	317.0	1.495	CC−NA	422.0	1.385	CT−H3	340.0	1.090	HO−OS	553.0	0.960
C*−CW	546.0	1.352	CC−NB	410.0	1.394	CT−HC	340.0	1.090	HS−SH	274.0	1.336
C*−HC	367.0	1.080	CK−H5	367.0	1.080	CT−HP	340.0	1.090	O2−P	525.0	1.480
CA−CA	469.0	1.400	CK−N*	440.0	1.371	CT−N	337.0	1.449	OH−P	230.0	1.610
CA−CB	469.0	1.404	CK−NB	529.0	1.304	CT−N*	337.0	1.475	OS−P	230.0	1.610
CA−CM	427.0	1.433	CM−CM	549.0	1.350	CT−N2	337.0	1.463	OW−HW	553.0	0.9572
CA−CN	469.0	1.400	CM−CT	317.0	1.510	CT−N3	367.0	1.471	S−S	166.0	2.038
CA−CT	317.0	1.510	CM−H4	367.0	1.080	CT−OH	320.0	1.410			
CA−H4	367.0	1.080	CM−H5	367.0	1.080	CT−OS	320.0	1.410			

Angle Parameters

angle	K_θ[d]	θ_{eq}[e]	angle	K_θ[d]	θ_{eq}[e]	angle	K_θ[d]	θ_{eq}[e]	angle	K_θ[d]	θ_{eq}[e]
C−CA−CA	63.0	120.00	CA−CT−HC	50.0	109.50	CN−NA−H	30.0	123.10	H1−CT−N2	50.0	109.50
C−CA−HA	35.0	120.00	CA−N2−CT	50.0	123.20	CR−NA−CW	70.0	120.00	H1−CT−OH	50.0	109.50
C−CB−CB	63.0	119.20	CA−N2−H	35.0	120.00	CR−NA−H	30.0	120.00	H1−CT−OS	50.0	109.50
C−CB−NB	70.0	130.00	CA−NA−H	30.0	118.00	CR−NB−CV	70.0	117.00	H1−CT−S	50.0	109.50
C−CM−CM	63.0	120.70	CA−NC−CB	70.0	112.20	CT−C−N	70.0	116.60	H1−CT−SH	50.0	109.50
C−CM−CT	70.0	119.70	CA−NC−CQ	70.0	118.60	CT−C−O	80.0	120.40	H2−CT−H2	35.0	109.50
C−CM−H4	35.0	119.70	CB−C−NA	70.0	111.30	CT−C−O2	70.0	117.00	H2−CT−N*	50.0	109.50
C−CM−HA	35.0	119.70	CB−C−O	80.0	128.80	CT−C*−CW	70.0	125.00	H2−CT−OS	50.0	109.50
C−CT−CT	63.0	111.10	CB−C*−CT	70.0	128.60	CT−CC−CV	70.0	120.00	H4−CM−N*	35.0	119.10
C−CT−H1	50.0	109.50	CB−C*−CW	63.0	106.40	CT−CC−CW	70.0	120.00	H4−CV−NB	35.0	120.00
C−CT−HC	50.0	109.50	CB−CA−H4	35.0	120.00	CT−CC−NA	70.0	120.00	H4−CW−NA	35.0	120.00
C−CT−HP	50.0	109.50	CB−CA−HA	35.0	120.00	CT−CC−NB	70.0	120.00	H5−CK−N*	35.0	123.05
C−CT−N	63.0	110.10	CB−CA−N2	70.0	123.50	CT−CT−CT	40.0	109.50	H5−CK−NB	35.0	123.05
C−CT−N3	80.0	111.20	CB−CA−NC	70.0	117.30	CT−CT−H1	50.0	109.50	H5−CQ−NC	35.0	115.45
C−N−CT	50.0	121.90	CB−CB−N*	70.0	106.20	CT−CT−H2	50.0	109.50	H5−CR−NA	35.0	120.00
C−N−H	30.0	120.00	CB−CB−NB	70.0	110.40	CT−CT−HC	50.0	109.50	H5−CR−NB	35.0	120.00
C−N*−CM	70.0	121.60	CB−CB−NC	70.0	127.70	CT−CT−HP	50.0	109.50	HC−CT−HC	35.0	109.50
C−N*−CT	70.0	117.60	CB−CN−NA	70.0	104.40	CT−CT−N	80.0	109.70	HO−OH−P	45.0	108.50
C−N*−H	30.0	119.20	CB−N*−CK	70.0	105.40	CT−CT−N*	50.0	109.50	HP−CT−HP	35.0	109.50
C−NA−C	70.0	126.40	CB−N*−CT	70.0	125.80	CT−CT−N2	80.0	111.20	HP−CT−N3	50.0	109.50
C−NA−CA	70.0	125.20	CB−N*−H	30.0	125.80	CT−CT−N3	80.0	111.20	HS−SH−HS	35.0	92.07
C−NA−H	30.0	116.80	CB−NB−CK	70.0	103.80	CT−CT−OH	50.0	109.50	HW−OW−HW	100.0	104.52
C−NC−CA	70.0	120.50	CB−NC−CQ	70.0	111.00	CT−CT−OS	50.0	109.50	N−C−O	80.0	122.90
C−OH−HO	35.0	113.00	CC−CT−CT	63.0	113.10	CT−CT−S	50.0	114.70	N*−C−NA	70.0	115.40
C*−CB−CA	63.0	134.90	CC−CT−HC	50.0	109.50	CT−CT−SH	50.0	108.60	N*−C−NC	70.0	118.60
C*−CB−CN	63.0	108.80	CC−CV−H4	35.0	120.00	CT−N−CT	50.0	118.00	N*−C−O	80.0	120.90
C*−CT−CT	63.0	115.60	CC−CV−NB	70.0	120.00	CT−N−H	30.0	118.04	N*−CB−NC	70.0	126.20
C*−CT−HC	50.0	109.50	CC−CW−H4	35.0	120.00	CT−N2−H	35.0	118.40	N*−CK−NB	70.0	113.90
C*−CW−H4	35.0	120.00	CC−CW−NA	70.0	120.00	CT−N3−H	50.0	109.50	N*−CT−OS	50.0	109.50
C*−CW−NA	70.0	108.70	CC−NA−CR	70.0	120.00	CT−OH−HO	55.0	108.50	N2−CA−N2	70.0	120.00
CA−C−CA	63.0	120.00	CC−NA−H	30.0	120.00	CT−OS−CT	60.0	109.50	N2−CA−NA	70.0	116.00
CA−C−OH	70.0	120.00	CC−NB−CR	70.0	117.00	CT−OS−P	100.0	120.50	N2−CA−NC	70.0	119.30
CA−CA−CA	63.0	120.00	CK−N*−CT	70.0	128.80	CT−S−CT	62.0	98.90	NA−C−O	80.0	120.60
CA−CA−CB	63.0	120.00	CK−N*−H	30.0	128.80	CT−S−S	68.0	103.70	NA−CA−NC	70.0	123.30
CA−CA−CN	63.0	120.00	CM−C−NA	70.0	114.10	CT−SH−HS	43.0	96.00	NA−CR−NA	70.0	120.00
CA−CA−CT	70.0	120.00	CM−C−O	80.0	125.30	CV−CC−NA	70.0	120.00	NA−CR−NB	70.0	120.00
CA−CA−H4	35.0	120.00	CM−CA−N2	70.0	120.10	CW−CC−NA	70.0	120.00	NC−C−O	80.0	122.50
CA−CA−HA	35.0	120.00	CM−CA−NC	70.0	121.50	CW−CC−NB	70.0	120.00	NC−CQ−NC	70.0	129.10
CA−CB−CB	63.0	117.30	CM−CM−CT	70.0	119.70	CW−NA−H	30.0	120.00	O−C−O	80.0	126.00
CA−CB−CN	63.0	116.20	CM−CM−H4	35.0	119.70	F−CT−F	77.0	109.10	O2−C−O2	80.0	126.00
CA−CB−NB	70.0	132.40	CM−CM−HA	35.0	119.70	F−CT−H1	35.0	109.50	O2−P−O2	140.0	119.90
CA−CM−CM	63.0	117.00	CM−CM−N*	70.0	121.20	H−N−H	35.0	120.00	O2−P−OH	45.0	108.23
CA−CM−H4	35.0	123.30	CM−CT−HC	50.0	109.50	H−N2−H	35.0	120.00	O2−P−OS	100.0	108.23
CA−CM−HA	35.0	123.30	CM−N*−CT	70.0	121.20	H−N3−H	35.0	109.50	OH−P−OS	45.0	102.60
CA−CN−CB	63.0	122.70	CM−N*−H	30.0	121.20	H1−CT−H1	35.0	109.50	OS−P−OS	45.0	102.60
CA−CN−NA	70.0	132.80	CN−CA−HA	35.0	120.00	H1−CT−N	50.0	109.50	P− OS−P	100.0	120.50
CA−CT−CT	63.0	114.00	CN−NA−CW	70.0	111.60	H1−CT−N*	50.0	109.50			

Table 14 (Continued)

Torsional Parameters

torsion	no. of paths[f]	$V_n/2$[g]	γ[h]	n[i]	torsion	no. of paths[f]	$V_n/2$[g]	γ[h]	n[i]
X−C−CA−X	4	14.50	180.0	2.0	X−CT−OH−X	3	0.50	0.0	3.0
X−C−CB−X	4	12.00	180.0	2.0	X−CT−OS−X	3	1.15	0.0	3.0
X−C−CM−X	4	8.70	80.0	2.0	X−CT−S−X	3	1.00	0.0	3.0
X−C−CT−X	4	0.00	0.0	2.0	X−CT−SH−X	3	0.75	0.0	3.0
X−C−N−X	4	10.00	180.0	2.0	X−CV−NB−X	2	4.80	180.0	2.0
X−C−N*−X	4	5.80	180.0	2.0	X−CW−NA−X	4	6.00	180.0	2.0
X−C−NA−X	4	5.40	180.0	2.0	X−OH−P−X	3	0.75	0.0	3.0
X−C−NC−X	2	8.00	180.0	2.0	X−OS−P−X	3	0.75	0.0	3.0
X−C−OH−X	2	1.80	180.0	2.0	C−N−CT−C	1	0.0	0.0	−4.0
X−C*−CB−X	4	6.70	180.0	2.0	C−N−CT−C	1	0.0	180.0	−3.0
X−C*−CT−X	6	0.00	0.0	2.0	C−N−CT−C	1	0.20	180.0	−2.0
X−C*−CW−X	4	26.10	180.0	2.0	C−N−CT−C	1	0.00	180.0	1.0
X−CA−CA−X	4	14.50	180.0	2.0	CT−CT−C−N	1	0.100	0.0	−4.0
X−CA−CB−X	4	14.00	180.0	2.0	CT−CT−C−N	1	0.000	0.0	−3.0
X−CA−CM−X	4	10.20	180.0	2.0	CT−CT−C−N	1	0.07	0.0	−2.0
X−CA−CN−X	4	14.50	180.0	2.0	CT−CT−C−N	1	0.000	180.0	1.0
X−CA−CT−X	6	0.00	0.0	2.0	CT−CT−N−C	1	0.50	180.0	−4.0
X−CA−N2−X	4	9.60	180.0	2.0	CT−CT−N−C	1	0.15	180.0	−3.0
X−CA−NA−X	4	6.00	180.0	2.0	CT−CT−N−C	1	0.00	180.0	−2.0
X−CA−NC−X	2	9.60	180.0	2.0	CT−CT−N−C	1	0.53	0.0	1.0
X−CB−CB−X	4	21.80	180.0	2.0	CT−CT−OS−CT	1	0.383	0.0	−3.0
X−CB−CN−X	4	12.00	180.0	2.0	CT−CT−OS−CT	1	0.1	180.0	2.0
X−CB−N*−X	4	6.60	180.0	2.0	CT−S−S−CT	1	0.60	0.0	3.0
X−CB−NB−X	2	5.10	180.0	2.0	CT−S−S−CT	1	3.50	0.0	−2.0
X−CB−NC−X	2	8.30	180.0	2.0	H−N−C−O	1	2.50	180.0	−2.0
X−CC−CT−X	6	0.00	0.0	2.0	H−N−C−O	1	2.00	0.0	1.0
X−CC−CV−X	4	20.60	180.0	2.0	N−CT−C−N	1	0.40	180.0	−4.0
X−CC−CW−X	4	21.50	180.0	2.0	N−CT−C−N	1	0.0	0.0	−3.0
X−CC−NA−X	4	5.60	180.0	2.0	N−CT−C−N	1	1.35	180.0	−2.0
X−CC−NB−X	2	4.80	180.0	2.0	N−CT−C−N	1	0.75	180.0	1.0
X−CK−N*−X	4	6.80	180.0	2.0	OH−CT−CT−OH	1	0.144	0.0	−3.0
X−CK−NB−X	2	20.00	180.0	2.0	OH−CT−CT−OH	1	1.00	0.0	2.0
X−CM−CM−X	4	26.60	180.0	2.0	OH−P−OS−CT	1	0.25	0.0	−3.0
X−CM−CT−X	6	0.00	0.0	3.0	OH−P−OS−CT	1	1.20	0.0	2.0
X−CM−N*−X	4	7.40	180.0	2.0	OS−CT−CT−OH	1	0.144	0.0	−3.0
X−CN−NA−X	4	6.10	180.0	2.0	OS−CT−CT−OH	1	1.00	0.0	2.0
X−CQ−NC−X	2	13.60	180.0	2.0	OS−CT−CT−OS	1	0.144	0.0	−3.0
X−CR−NA−X	4	9.30	180.0	2.0	OS−CT−CT−OS	1	1.00	0.0	2.0
X−CR−NB−X	2	10.00	180.0	2.0	OS−CT−N*−CK	1	0.50	180.0	−2.0
X−CT−CT−X	9	1.40	0.0	3.0	OS−CT−N*−CK	1	2.50	0.0	1.0
X−CT−N−X	6	0.00	0.0	2.0	OS−CT−N*−CM	1	0.50	180.0	−2.0
X−CT−N*−X	6	0.00	0.0	2.0	OS−P−OS−CT	1	0.25	0.0	−3.0
X−CT−N2−X	6	0.00	0.0	3.0	OS−P−OS−CT	1	1.20	0.0	2.0
X−CT−N3−X	9	1.40	0.0	3.0	S−CT−N*−CM	1	2.50	0.0	1.0

Improper Torsions

torsion	$V_n/2$[g]	γ[h]	n[i]	torsion	$V_n/2$[g]	γ[h]	n[i]	torsion	$V_n/2$[g]	γ[h]	n[i]
X−CT−N−CT	1.0	180.0	2.0	X−X−CQ−H5	1.1	180.0	2.0	CK−CB−N*−CT	1.0	180.0	2.0
X−N2−CA−N2	10.5	180.0	2.0	X−X−CR−H5	1.1	180.0	2.0	CM−C−CM−CT	1.1	180.0	2.0
X−O2−C−O2	10.5	180.0	2.0	X−X−CV−H4	1.1	180.0	2.0	CM−C−N*−CT	1.0	180.0	2.0
X−X−C−O	10.5	180.0	2.0	X−X−CW−H4	1.1	180.0	2.0	CT−CM−CM−C	1.1	180.0	2.0
X−X−CA−H4	1.1	180.0	2.0	X−X−N−H	1.0	180.0	2.0	CW−CB−C*−CT	1.1	180.0	2.0
X−X−CA−H5	1.1	180.0	2.0	X−X−N2−H	1.0	180.0	2.0	NC−CM−CA−N2	1.1	180.0	2.0
X−X−CA−HA	1.1	180.0	2.0	X−X−NA−H	1.0	180.0	2.0	NA−CV−CC−CT	1.1	180.0	2.0
X−X−CK−H5	1.1	180.0	2.0	CA−CA−C−OH	1.1	180.0	2.0	NA−CW−CC−CT	1.1	180.0	2.0
X−X−CM−H4	1.1	180.0	2.0	CA−CA−CA−CT	1.1	180.0	2.0	NA−NC−CA−N2	1.1	180.0	2.0
X−X−CM−HA	1.1	180.0	2.0	CB−NC−CA−N2	1.1	180.0	2.0	NB−CW−CC−CT	1.1	180.0	2.0

Van der Waals Parameters

atom type	R^{*}[j]	ϵ[k]	atom type	R^{*}[j]	ϵ[k]	atom type	R^{*}[j]	ϵ[k]	atom type	R^{*}[j]	ϵ[k]
C[l]	1.9080	0.0860	H2	1.2870	0.0157	HS	0.6000	0.0157	O2	1.6612	0.2100
CA	1.9080	0.0860	H3	1.1870	0.0157	HW	0.0000	0.0000	OH	1.7210	0.2104
CM	1.9080	0.0860	H4	1.4090	0.0150	IP	1.8680	0.00277	OS	1.6837	0.1700
Cs	3.3950	0.0000806	H5	1.3590	0.0150	K	2.6580	0.000328	OW	1.7683	0.1520
CT	1.9080	0.1094	HA	1.4590	0.0150	Li	1.1370	0.0183	P	2.1000	0.2000
F	1.75	0.061	HC	1.4870	0.0157	N[m]	1.8240	0.1700	Rb	2.9560	0.00017
H	0.6000	0.0157	HO	0.0000	0.0000	N3[n]	1.875	0.1700	S	2.0000	0.2500
H1	1.3870	0.0157	HP	1.1000	0.0157	O	1.6612	0.2100	SH	2.0000	0.2500

Table 14 (Footnotes)

[a] See eq 1. [b] kcal/(mol Å^2). [c] Å. [d] kcal/(mol radian^2). [e] deg. [f] Number of bond paths that the total $V_n/2$ is divided into. This is equal to the product of the number of bonds to each of the middle two atoms. For example, since X−C−CA−X has 4 bond paths, each of them has a $V_n/2$ of 14.5/4 kcal/mol assigned to it. [g] Magnitude of torsion in kcal/mol. [h] Phase offset in deg. [i] The periodicity of the torsion. A negative value is not used in the calculation but signifies more than one component around a given bond. [j] van der Waals R^* for a given atom in Å. The value used in eq 1 for an interaction of atom i and atom j is $R_{ij}^* = R_i^* + R_j^*$. [k] van der Waals well depth for a given atom in kcal/mol. The value used in eq 1 for an intersection between atoms i and $j = \epsilon_{ij} = (\epsilon_i\epsilon_j)^{1/2}$. Note that $A_{ij} = \epsilon_{ij}^*(R_{ij}^*)^{12}$ and $B_{ij} = 2^*\epsilon_{ij}^*(R_{ij}^*)^6$. [l] All sp^2 carbons have these parameters. [m] All sp^2 nitrogens have these parameters. [n] sp^3 nitrogen parameters, see ref 109; these parameters were derived subsequent to the ubiquitin simulation, which used the common parameter for sp^2 and sp^3 parameters of atom type N.

on the accurate reproduction of electrostatic interactions—a demonstrated strength of the Weiner *et al.* force field;[5,6] (3) the use of a new approach for deriving electrostatic potential fit charges (multiconformer RESP) which are better behaved than the previous standard ESP model; (4) general and algorithmic approaches to describe the nonbonded interactions, particularly for hydrogens; and (5) a minimalist approach to adding dihedral potentials to the energy function. Through our approach we have minimized the coupling between the different terms in the force field equation. Although only the total energy can be compared directly with experiment, the force field has the potential of providing additional qualitative insight when the results agree with experiment.[42]

How can one extend this model to new organic molecules? First, one must carry out quantum mechanical calculations at the 6-31G* level to derive restrained electrostatic potential (RESP) charges, ideally with multiple conformations[39,103] to minimize statistical errors. Secondly, one can use the van der Waals parameters presented here, or from the OPLS model, if appropriate liquids have been simulated involving the requisite atom types. With a few exceptions, most of the van der Waals parameters are likely to be already available. The bond, angle, and dihedral parameters can come from experimental data, using initially "generic" torsional parameters such as X−C(sp^3)−C(sp^3)−X, as suggested above. Then, 6-31G*/MP2 quantum mechanical conformational analysis can be carried out on appropriate flexible fragments of the molecule(s) of interest. By comparison with the energies calculated with the molecular mechanical model of these fragments, additional specific torsional potentials can be added to ensure as accurate a representation of the intrinsic conformational energies as possible. Based on our experience, additional explicit torsional potentials are likely to be required for well-understood "anomeric effects", such as in 1,3 dioxanes,[42] or in cases of large internal electrostatic interactions/intramolecular H-bonds such as those involved with the peptide ψ,ϕ or the nucleoside χ angles.

Further applications will be required to assess how successful the new model is. In the studies described above, the major weakness was the necessity of adding dihedral potentials for the ψ and ϕ of peptides and χ of nucleic acids without obvious physical justification. This effect is at least partially due to the somewhat too large polarity of the 6-31G* RESP model, which is needed to accurately simulate solvation at the effective two-body level. The magnitude of the re-optimized ψ and ϕ dihedral parameters is considerably reduced in a non-additive force field with reduced gas-phase-like polarity,[104] and the magnitudes are slightly reduced for χ.[105] A better behaved set of charges which yielded more accurate conformational energies and still reproduced solvation free energies could possibly be derived through empirical adjustment. But then the generality and simplicity of the model would be sacrificed. These examples do emphasize the degree to which the nonbonded and dihedral terms dominate any complex intramolecular function, particularly when the charges are optimized for an effective two-body model to reproduce the energies of polar and ionic molecules in solution.

This new force field has retained some of the features of the Weiner *et al.* force field,[5,6] with its emphasis on the accurate representation of electrostatics and simple representation of bond and angle energies, while offering electrostatic and VDW parameters which are optimized for state-of-the-art condensed-phase simulations. Further work is being carried out in this laboratory to investigate the improved performance of models which incorporate either off-center charges (lone pairs)[106] or electronic polarization.[40,82,104,105,107,108] It is our belief, however, that with this new force field we have reached the limit for accurately representing biomolecular systems with an effective two-body additive potential employing quantum mechanically derived atom centered charges.[109]

Acknowledgment. When this article appears in print, we will provide an anonymous FTP address for anyone to download the complete set of topology files for amino acids and nucleic acids and the complete parameter file (see Note Added in Proof). P.A.K. is pleased to acknowledge research support from the NIH (GM-29072 and CA-25644) and NSF (CHE-91-13472). W.D.C. was the recipient of a Rosenberg graduate fellowship from the University of California at San Francisco. P.C. acknowledges the support of DARPA (MDA-91-Y-1013) and The Polish Committee for Scientific Research. We are grateful to the UCSF Computer Graphics Lab (RR-1081, T. Ferrin, P.I) for graphics support and to the San Diego and Pittsburgh Supercomputer Centers. We are grateful to Allison Howard, Steve DeBolt, Craig Gough, and David Veenstra for their contributions to this work. We thank J. L. Miller and J. Nagy for their unpublished data.

Note Added in Proof. Computer readable files for all the parameters of this force field and any others stored in the amber/dat directory can be retrieved by anonymous ftp from ftp.amber.ucsf.edu or by visiting the WWW page at http://www.amber.ucsf.edu.

We have carried out further calculations on the relative free energy of solvation of 9-methyladenine and methane (Table 12) using the methodologies in AMBER 4.0 (Gibbs) to compare with those reported using the SPASMS module (J. L. Miller and P. Kollman, studies in progress). The calculated free energies for the electrostatic part of the perturbation are in excellent agreement with each other (12.16 kcal/mol with SPASMS and 11.94 kcal/mol with GIBBS for the relative solvation free energy of 9-methyladenine and methane) and

(103) Reynolds, C. A.; Essex, J. W.; Richards, W. C. *J. Am. Chem. Soc.* **1992**, *114*, 9075−9079.
(104) Cornell, W. D.; Caldwell, J. W.; Kollman, P. A., manuscript in preparation.
(105) Cieplak, P.; Kollman, P. A., unpublished results.
(106) Dixon, R.; Kollman, P. A., work in progress.
(107) Caldwell, J. W.; Kollman, P. A. Cation-π Interactions. Non-additive Effects are Critical in Their Accurate Representation. *J. Am. Chem. Soc.*, accepted for publication.
(108) Sun, Y. C.; Caldwell, J. W.; Kollman, P. A. Molecular Dynamics and Free Energy Perturbation Study of Spherand Complexation with Metal Ions Employing Additive and Non-Additive Force Fields. *J. Phys. Chem.*, accepted for publication.
(109) Morgantini, P.-Y.; Kollman, P. A. Solvation Free Energies of Amides and Amines: Disagreement Between Free Energy Calculations and Experiment. *J. Am. Chem. Soc.*, accepted for publication.

involve simulations up to 800 ps in length. On the other hand, the second leg of the simulation, which involves the disappearance of the van der Waals interactions of the base atoms and changing the N−C bond to the H−C bond of methane, gives very different free energies for the two protocols. The calculated value is 6.16 kcal/mol with SPASMS and 1.94 kcal/mol with GIBBS, for simulations as long as 500 ps (with SPASMS) and 800 ps (with GIBBS). This leads to the values of 18.3 and 13.9 reported in Table 12. Using the calculated free energy of solvation of methane (2.5 kcal/mol), this leads to an absolute solvation free energy for 9-methyladenine of 15.8 kcal/mol with SPASMS and 11.4 kcal/mol with GIBBS. Interestingly, these two values bracket the extrapolated experimental value of 13.6 ± 1.1 kcal/mol (see Ferguson et al. (Ferguson, D. M.; Pearlman, D. A.; Swope, W. C.; Kollman P. A. *J. Comp. Chem.* **1992**, *13*, 362−372) for a discussion on how this "experimental" value was determined). Further calculations are required to sort out this issue.

JA943664F