Molecular Mechanical Models for Organic and Biological Systems Going Beyond the Atom Centered Two Body Additive Approximation: Aqueous Solution Free Energies of Methanol and N-Methyl Acetamide, Nucleic Acid Base, and Amide Hydrogen Bonding and Chloroform/Water Partition Coefficients of the Nucleic Acid Bases

PIOTR CIEPLAK,[1] JAMES CALDWELL,[2] PETER KOLLMAN[2]

[1]*Chemistry Department, Warsaw University, Warsaw, Poland*
[2]*Department of Pharmaceutical Chemistry, University of California, San Francisco*

Received 30 August 2000; accepted 18 January 2001

ABSTRACT: We have developed a methodology to derive RESP charges for molecular mechanical models that include "lone pairs" on lone-pair donor sites and atom-centered polarizabilities. This approach uses a very high level *ab initio* cc-pVTZ basis set,[1] where the multipole moments of the molecules are as accurate as possible. The partial charges are derived self-consistently so that the model, whose electrostatic potential comes from both partial charges and induced dipoles, reproduces the quantum mechanical electrostatic potential. We then study the ability of such models to reproduce the aqueous solvation free energy

Correspondence to: P. Kollman; e-mail: pak@cgl.ucsf.edu
Contract/grant sponsor: NSF
Contract/grant sponsor: NIH; contract/grant numbers: CA-25644 and GM-29072
Contract/grant sponsor: KBN; contract/grant number: 3 T09A 096 19
This article contains supplementary material available from the author upon request or via the Internet at ftp.wiley.com/public/journals/jcc/suppmat/22/10 or http://www.interscience.wiley.com/jpages/0192-8651/suppmat/v22.1048.html

of methanol and N-methyl acetamide (NMA), the base pair hydrogen bonding of the 26 base pairs analyzed by Hobza et al. and the chloroform/water partition coefficients of the five N-methyl substituted nucleic acid bases. The base pair H-bond energies are described as well as the atom centered additive model, after modifying the van der Waals parameter on the N—H to give reasonable base pair H-bond distances. The experimental solvation free energies (gas → water) of methanol and NMA are well described, and the water/$CHCl_3$ partition coefficients are improved over the additive model, without any parameter changes. J Comput Chem 22: 1048–1057, 2001

Keywords: molecular mechanical models; lone pairs; cc-pVTZ basis set

Introduction

A cornerstone of all molecular simulations is the molecular mechanical model. The accuracy of this model and the extent of conformational sampling determine the ability of the simulations to reproduce the experimental behavior of complex molecular systems. The sampling problem has been the larger difficulty in simulating biomolecular systems, but nonetheless, to increase the realism of simulations, one must proceed to improve both the molecular mechanical model and the sampling.

In this article, we present the first steps in the development of a molecular mechanical model for nucleic acids and proteins that includes both lone pair sites on electron donor atoms and inducible polarizabilities on all atoms. We build this model just as we have the effective two-body additive models, which used the minimalist model for the force field [eq. (1)] and *ab initio* 6-31G* Hartree–Fock calculations to derive RESP charges.[2–4] Given the considerable successes of such effective two-body additive models in molecular dynamics simulations of proteins, nucleic acids, and membrane systems, it makes sense to use them as a basis for deriving the nonadditive models.

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

One uses 6-31G* RESP charges for effective two-body models because these charges systematically overestimate monomer dipole moments, like the standard water models, such as TIP3P,[5] whose charges were derived empirically to fit the density and enthalpy of liquid water. One is assuming that such charges should implicitly include the effect of polarization on the molecular charge distribution, which is significant in aqueous solution, and likely also in the interior of proteins.

Nonetheless, such models systematically overestimate the chloroform/water partition coefficients for nucleic acid bases[6] because they use "too polarized" charge distributions in chloroform, and thus make the models too soluble in this liquid compared to water. As we have shown,[6] scaling the base charges by 0.9 for the chloroform simulation and maintaining the standard charges in water leads to chloroform/water partition coefficients in much better agreement with experiment. However, such empirical scaling is an unsatisfactory solution to the problem. What one needs is a model that begins with an accurate gas phase charge distribution and lets the distribution in condensed phases reflect the polarity of the environment.

The first article descibing the use of induced dipoles to represent the polarization energy was that by Warshel and Levitt.[7] Since that time, many articles have used such a model for polarization.[8–12] We have been building such models for some time. Ten years ago, we showed that a model that included atomic polarizabilities could more accurately describe water–ion interactions than could additive models in molecular dynamics[13, 14] and Monte Carlo simulations.[15] Subsequently, we have illustrated the role of nonadditive effects in cation–pi interactions,[16] amine solvation,[17] ion binding in anisole spherands,[18] and liquid properties of water, methanol, and N-methyl acetamide.[19] We have also shown the importance of off-center charges for representing lone-pair directionality in hydrogen bonding.[20]

In this article we take the next step toward a general and accurate nonadditive model for biological simulations. We first show how to derive accurate

charges with inclusion of "lone pairs" on electron donor sites of methanol, N-methyl acetamide, and the five N-Me nucleic acid bases, G, C, T, U, and A. This requires a self-consistent determination of the RESP charges at a DFT/cc-pVTZ basis set level, the self-consistency requirement resulting from the fact that when one includes intramolecular polarization with such charges, the induced dipoles have a non-negligible effect on the electrostatic potential. It is the electrostatic potential resulting from both the charges and the induced dipoles that must reproduce that found quantum mechanically, and this is accomplished with an iterative approach.

We then use this model in calculations of the aqueous solvation free energies of methanol and N-methyl acetamide, the hydrogen bonding of 26 base pairs, and the chloroform/water partition coefficients of the five N-methylated bases. We are able to get reasonable agreement with experiment or *ab initio* calculations in these cases, with small changes in the van der Waals parameters of the NH hydrogen, HO hydrogen, and alcohol (OH) oxygen.

Methods

CHARGE DERIVATION

The "lone pairs" are defined as additional points attached to electron-donating atoms, for example, sp^2, sp^3 oxygen, nitrogen, and sulphur. We found out that the best bond length between nitrogen and oxygen atoms and their lone pair is 0.35 Å, whereas it is 0.7 Å for sulphur. Orientation of the "lone pairs" is described in ref. 20. The "lone pairs" serve as additional location of point charges. Their role is to improve the total molecular charge distribution represented by point charges and to gain directionality of the intermolecular interactions.

The nonadditive interactions have been described using the simple isotropic polarization model first introduced by Applequist[21] (and subsequently developed by Warshel[7]) according to which atomic point charges polarize other atoms and interact with their induced atomic dipole moments. The electrostatic field that induces dipole moment on ith atom is calculated self-consistently in an iterative way, and the polarization energy is then calculated according to the following formula:

$$E_{\mathrm{pol}} = -\frac{1}{2}\sum^{i} \alpha_i \mathbf{E}_i^{(0)} \cdot \mathbf{E}_i \qquad (2)$$

where α_i is the isotropic polarizability of ith atom, $\mathbf{E}_i$ is the electrostatic field on atom i due to all other charges and induced dipoles, and $\mathbf{E}_i^0$ is the electrostatic field on atom ith due to permanent atomic charges only. There are no polarizabilities assigned to the "lone pairs" (atom type EP signifying "Extra Point").

The new nonadditive parametrization is based on the same force-field equation as in Cornell et al.[4] [eq. (1)] with the polarization energy [eq. (2)] added. Derivation of our new force field is based on the same algorithmic approach, which was applied to derivation of the Cornell et al. force field.[4] We used the same bond and bond angle parameters as in Cornell et al. and Wang et al.[22] force fields.

The following approach was used to derive atomic charges. As previously, we decided to use our RESP methodology[2–4] to fit charges to the quantum mechanically derived electrostatic potential around a molecule, and, for comparison, we also describe the fits using no restraints on the electrostatic potential in the charges (ESP model). Initially, the geometry of each molecule was optimized using the Hartree–Fock method and 6-31G* basis set.[23] Because we want to get closest to the gas phase experimental molecular dipole moments, and potentially accurate higher multipole moments as well, we performed high-level quantum mechanical density functional (DFT) calculations for each tested molecule. The B3LYP type exchange and correlation functionals[24] and the cc-pVTZ basis set[1] was used to obtain electrostatic potential around each molecule.

At this point one could hope that such atomic charges, which reproduce gas phase dipole moments, when combined together with properly chosen atomic polarizabilities, would constitute an appropriate force field model for a condensed phase properties modeling in molecular mechanics and dynamics calculations. Unfortunately, such a scheme does not take into account the self-polarization of a molecule, and as such, is not adequate for intermolecular interaction calculations. This is because of the fact that induced dipoles significantly alter the monomer charge distribution from that which was fit to quantum mechanical electrostatic potential. This problem has been also pointed out by Reynolds et al.[25] This effect can be seen in Figure 1, which shows the orientation of the QM dipole moment, induced and effective resultant dipole moments for the N-methyl adenine molecule.

This effect can be corrected by taking into account self-polarization in the RESP charge-fitting procedure. To do this, we calculated electrostatic potential around a molecule due to induced dipoles [ESP(ind)]. Then, a new set of ESP point

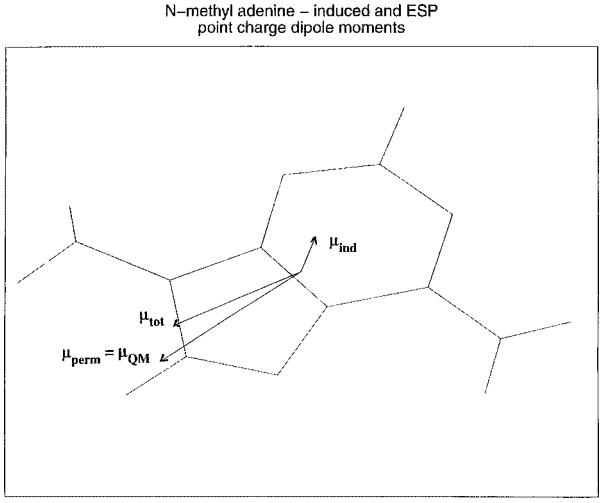

FIGURE 1. Dipole moment components (Debye) of N-methyl adenine of initial RESP fit, from induced dipoles and the vector sum of the two.

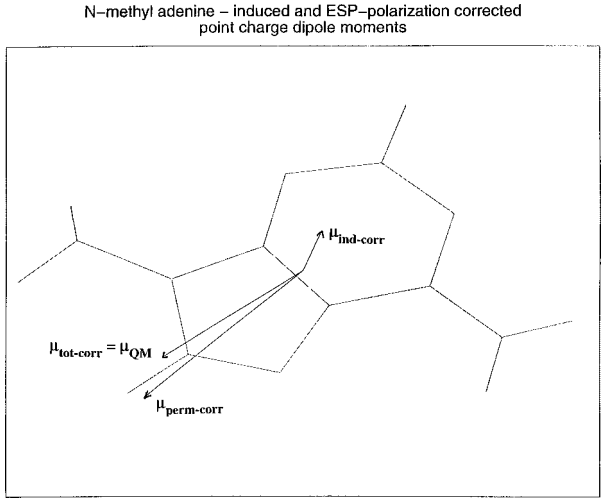

FIGURE 2. Dipole moment components (Debye) of N-methyl adenine after iterative calculation of both permanent and induced electrostatic moments as well as the total, which in this case is close to equal to the QM-value.

charges were fitted to the difference between QM and induced dipoles electrostatic potentials (e.g., to ESP[(QM)-ESP(ind)]. Such new charges were used in the next iteration to self-polarize the molecule and to calculate the new electrostatic potential due to induced dipoles.

Such procedure was repeated iteratively until the new charges, or alternatively ESP(ind), did not change in two consecutive iterations. Usually from 4 up to 12 iterations was necessary in each case to achieve convergence. Finally, the fully iterated ESP(QM)-ESP(ind) was used to perform RESP-type charge fitting. The effect of such an approach on correcting the total charge distribution is represented in Figure 2 for the N-methyl adenine molecule.

As one can see, the effective molecular dipole moment, calculated as a sum of dipole moment due to new, polarization corrected set of charges and total induced dipole moment matches quantum mechanical dipole moment very well. The results of the iterations during the charge fitting are shown in Table I.

TABLE I.
The N-Me-ADE (No "Lone Pairs") Dipole Moments (in Debye) Calculated from Point Charges at Various Stages of the Charge-Fitting Procedure.

	Dipole Moments Calculated from Charges after:[a]		
ITERATION	ESP-fit	ESP-fit + POL	RESP-fit + POL
0th ITERATION			
charge fit to ESP(QM)	2.46	1.95	2.05
1st ITERATION			
charge fit to ESP = ESP(QM) − ESP[ind(0-iter)]		2.37	
2nd ITERATION			
charge fit to ESP = ESP(QM) − ESP[ind(1st-iter)]		2.43	
3rd ITERATION			
charge fit to ESP = ESP(QM) − ESP[ind(2nd-iter)]		2.44	
4th ITERATION			
charge fit to ESP = ESP(QM) − ESP[ind(3rd-iter)]		2.45	
5th ITERATION			
charge fit to ESP = ESP(QM) − ESP[ind(4th iter)]		2.45	2.46

[a] Quantum mechanical dipole moment = 2.45 D.
ESP(QM) was calculated at the DFT/B3LYP-cc-pVTZ level.

TABLE II.
Dipole Moments for N-Methyl-adenine, Methanol, and N-methyl-acetamide and Glycine Using Various Schemes for Charge Fitting and "Lone Pairs."

Molecule	ESP Fit			RESP Fit			Number of Iterations
	Perm[a]	Ind[b]	Total[c]	Perm[a]	Ind[b]	Total[c]	
N-Me-ADE							
No "lone pairs"	3.12	0.76	2.45	3.06	0.68	2.46	5
"Lone pairs" at 0.2 Å	3.15	0.91	2.46	3.04	0.73	2.40	8
"Lone pairs" at 0.35 Å	3.08	0.77	2.45	3.00	0.75	2.37	5
Dipole QM(DFT): 2.45 D							
Methanol							
No "lone pairs"	1.54	0.12	1.58	1.83	0.11	1.85	6
"Lone pairs" at 0.2 Å	1.65	0.15	1.54	1.94	0.09	1.87	12
"Lone pairs" at 0.35 Å	1.63	0.11	1.54	1.93	0.08	1.87	12
Dipole QM(DFT): 1.57 D							
N-Methyl-acetamide							
No "lone pairs"	4.47	0.83	3.66	4.47	0.76	3.74	6
"Lone pairs" at 0.2 Å	5.45	1.80	3.68	5.25	1.05	4.20	10
"Lone pairs" at 0.35 Å	4.77	1.09	3.68	4.72	1.02	3.70	4
Dipole QM(DFT): 3.68 D							
Glycine, C5 conformation							
No "lone pairs"	3.95	0.65	3.31	3.8	0.66	3.22	6
"Lone pairs" at 0.2 Å	3.83	1.03	3.33	3.68	0.56	3.16	14
"Lone pairs" at 0.35 Å	3.85	0.57	3.32	3.74	0.55	3.20	5
Dipole QM(DFT): 3.30 D							

[a] Dipole moment due to partial charges.
[b] Dipole moment due to induced charges.
[c] Total dipole moment.

The charge-fitting procedure is independent of the choice or quality of the atomic polarizabilities taken into consideration. This is because the fitted point charges effectively compensate for inaccuracies in an assumed set of polarizability values to reproduce the quantum mechanical dipole moment. On the other hand, changing atomic polarizabilities requires that the charges have to be refitted.

The effect of adding "lone pairs" on the molecular dipole moments is summarized in Table II for a few molecules.

As one can see, polarization corrected charges, obtained here using iterated electrostatic potential, together with a self-induced molecular dipole moment, reproduce the total quantum mechanical dipole moment very well for any tested molecule. This effect is almost independent on whether "lone pairs" were considered in the fit or not, although convergence for the case of "lone pairs" located at the distance 0.2 Å is less accurate than that at 0.35 Å, and that is one of the reasons we use a distance of 0.35 Å in our standard model. The charges are also similar for ESP and RESP approaches, which supports the idea that these will lead to similar intermolecular interaction energies.

FREE ENERGY CALCULATIONS

The calculations were carried out using modified versions of the Sander and Gibbs modules of Amber5. The modifications for the extra points (EP) were: to place the EPs analytically (and symmetrically) at appropriate atoms after the SHAKE subroutine[26] had adjusted the system bond lengths and then as the derivatives of the energy were evaluated to assign the $dE(\mathrm{EP})/d(x, y, z)$ derivative back to its "parent" atom. This is done algebraically; quaternions were not invoked, as the atom–EP bond lengths are so short that the error introduced is presumed to be so small that it falls into the other "MD noise." In the case where the EP is not a formal pseudolone pair, the derivatives were farmed out to the symmetrically related atoms in direct proportion to their X-EP distance (no systems of that type are

reported here but the capability is in the molecular dynamics code).

The direct free energy runs were always done with the new charge/EP state as lambda = 1; lambda = 0 corresponded to the Parm94 charge with the EPs mutated to dummy particles.

Equilibrations were done with Sander for 200–1000 ps. The latter time was necessary when using the chloroform model, which was a slight modification of that of Fox et al.[27] It was necessary to further modify the Sander program to allow both Berendsen[28] and Anderson[29] temperature coupling simultaneously (Gibbs already had that feature). The temperature was set at 300 K and with separate solute/solvent Berendsen coupling and an (Anderson) reassignment of velocities with random Boltzmann weighted values every 500 MD steps (which were 1 fs) for chloroform solvent simulations, water simulations only used Berendsen coupling. Following equilibrations, we did some perturbation length experiments with runs from 55 to 505 ps and settled on 205 ps (41 5-ps windows) runs, because these were well converged.

To make the calculations more computationally efficient, Sander/Gibbs were also modified such that separate pair lists are kept over the regular nonbond interactions and the ones involving one or two polarizable centers. We then used TIP3P additive water and the modified Fox et al.[27] additive chloroform, decreasing the polarization overhead from ~4× the additive model to ~2×.

Results

NUCLEIC ACID HYDROGEN BONDING

The polarization corrected atomic charges have been developed for all the protein and nucleic acid residues for use in molecular mechanics and molecular dynamics calculations[30] and for all molecules considered in this article. The appropriate charges and atomic polarizabilities are presented in Figure 3 for methanol and NMA (the methylated bases appear in the supplementary material).

The new set of charges have been used for calculating nucleic acid base pairs interactions. The results are summarized in Tables III and IV. All geometries were adopted from Sponer et al.[31]

For the polarizable molecular mechanical results in Table III, a modified van der Waals R^* equal to 0.85 Å, for the Lennard-Jones parameter for a hydrogen atom connected to an sp^2 nitrogen (type HN). In the original AMBER force field[4] this parameter has a value of 0.6 Å.

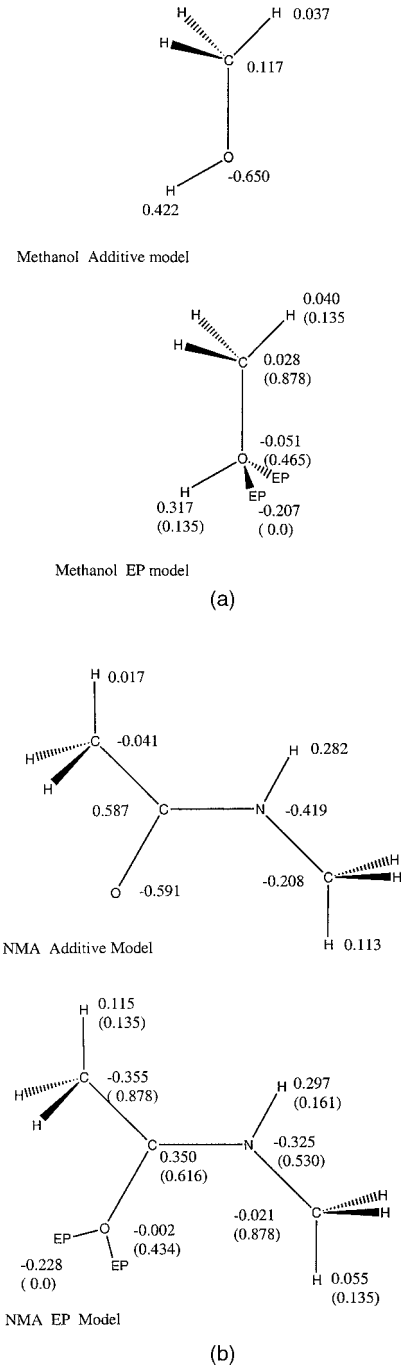

FIGURE 3. Charge models (atomic charge units) for methanol (a), N-methyl acetamide (b) with both the additive model from ref. 4 and the polarizable (EP) model presented here. In the latter case, the polarizabilities (Å^3) are in parentheses.

As one can see from Table IV, the average absolute error and correlation coefficients for linear fits are excellent for the additive force field (parm94) and for the LP = 0.35 Å nonadditive model. We

TABLE III.

H-Bond Interaction Energies for Nucleic Acid Base Pairs in kcal/mol, with Various Schemes of Using "Lone Pairs."[a]

	Molecular Mechanics				Quantum Mechanics	
Basepair	Parm94FF[b]	NO-LP[c]	LP at 0.2 Å[d]	LP at 0.35 Å[e]	MP2[a]	MP2 + def[a]
gcwc	−26.04	−22.35	−27.59	−25.11	−25.80	−23.80
gg1	−23.76	−21.54	−23.76	−22.43	−24.70	−22.20
cc	−17.67	−15.51	−21.42	−19.51	−18.80	−17.50
gg3	−17.78	−14.86	−18.94	−16.99	−17.80	−17.00
ga1	−12.98	−13.03	−14.42	−15.06	−15.20	−14.10
gt1	−14.63	−13.55	−15.62	−14.84	−15.10	−13.90
gt2	−14.36	−12.90	−13.67	−13.62	−14.70	−13.50
ac1	−12.82	−11.24	−14.66	−14.80	−14.30	−13.50
gc1corr	−14.39	−11.62	−16.56	−15.42	−14.30	−13.40
gc2corr	−12.97	−12.47	−15.65	−14.66	−14.10	−13.20
ga3	−13.12	−13.62	−14.75	−14.14	−13.80	−13.10
athoog	−13.03	−11.85	−13.18	−13.19	−13.30	−12.70
atrhoog	−13.09	−11.67	−12.65	−12.8	−13.20	−12.60
atwc	−12.17	−10.29	−11.29	−12.36	−12.40	−11.80
atrwc	−12.18	−10.06	−10.47	−11.85	−12.40	−11.70
aa1	−10.16	−8.65	−9.83	−11.67	−11.50	−11.00
ga4	−10.48	−8.74	−11.05	−12.07	−11.40	−10.70
tc2	−10.96	−10.35	−13.23	−12.40	−11.60	−10.70
tc1	−10.83	−9.85	−11.62	−11.32	−11.40	−10.70
aa2	−9.89	−9.11	−10.31	−11.14	−11.00	−10.30
tt2	−11.16	−10.19	−10.99	−11.24	−10.60	−10.00
tt1	−11.23	−10.05	−10.38	−10.87	−10.60	−10.00
tt3	−11.33	−9.97	−9.96	−10.62	−10.60	−9.90
ga2	−10.05	−9.82	−11.09	−11.15	−10.30	−9.60
gg4	−9.31	−7.50	−11.10	−11.13	−10.00	−9.30
aa3	−9.60	−9.77	−10.54	−10.50	−9.80	−9.20

[a] H-bond geometries and QM energies from ref. 31.
[b] MM energies using the additive charges.[4]
[c] This work, no "lone pairs."
[d] This work, "lone pairs" at 0.2 Å.
[e] This work, "lone pairs" at 0.35 Å.

TABLE IV.

Average Errors (kcal/mol) of the MM Interaction Energies with Respect to Quantum Mechanical Values.

	parm94	NO-LP	LP at 0.2 Å	LP at 0.35 Å
Avg. error vs. MP2[a]	0.66	1.85	1.00	0.58
Avg. error vs. MP2 + def[b]	0.68	1.08	1.35	0.98

	Parm94			No-LP			LP at 0.2 Å			LP at 0.35 Å		
	A	B	R^2	A	B	R^2	A	B	R^2	A	B	R^2
Fit to MP2[a]	0.222	0.981	0.972	−0.313	0.843	0.949	0.699	1.067	0.933	−1.717	0.882	0.97
Fit to MP2 + def[b]	0.727	1.088	0.972	0.081	0.932	0.949	0.081	0.932	0.970	−1.255	0.979	0.97

Linear Fit (Y = A + B, where Y = MM model and X = QM Model and Correlation Coefficients = R^2.
[a] See ref. 31, MP2 energy.
[b] See ref. 31, MP2 energy including intramolecular deformation upon complexation.

TABLE V.
NMA–Water Interaction Energies (kcal/mol) and Distance (Å) HOHOC(CH_3)—NH(CH_3).

Structure	*Ab Initio*[a]		MM, Additive[a]		MM pol, no lp[a]		MM pol, with lp[d]	
	R	ΔE	*R*	ΔE	*R*	ΔE	*R*	ΔE
I (CO...H) = 120°	1.99	−5.98	1.72	−7.64	1.99	−5.62	2.01	−5.75
II (CO...H) = 180°	2.04	−4.75	1.73	−7.22	2.03	−5.03	2.03	−4.95
III (CO...H) = −90°	2.19	−4.21	1.83	−6.79	2.17	−4.67	2.21	−4.09

[a] cc-pVTZ (f) (BSSE) + MP2.
[b] 6-31G* RESP charges and TIP3P water.[5]
[c] DFT-pVTZ RESP charges, no "lone pairs" for NMA and polarizable water model.[35]
[d] DFT-pVTZ RESP charges, with "lone pairs" for NMA and polarizable water model.[35]

do not present the H-bond geometries in detail, but they are within 0.1 Å rms from those described in ref. 31. As noted earlier,[32] the Cornell et al. force field gave the best fit to *ab initio* base interaction energies, and the new nonadditive model also fits those data comparably well.

WATER–AMIDE HYDROGEN BONDING

In Table V, we present the results of *ab initio* calculations of water hydrogen bonding to the trans N-methyl acetamide (NMA) carbonyl oxygen, which *ab initio* calculations suggest is the strongest H-bond acceptor site on NMA. As one can see, the standard additive model[4] underestimates the hydrogen bond distance, overestimates the interaction energy, and underestimates the H-bond directionality (preference of $\Theta = 120°$ over $\Theta = 180°$ or $\Theta = 90°$). The nonadditive model without "lone pairs" shows impressive agreement with the *ab initio* calculations for the interaction energy and distance, but also underestimates the directionality. The model with lone pair does an excellent job on all three criteria.

SOLVATION FREE ENERGIES OF METHANOL AND N-METHYL ACETAMIDE

We carried out free-energy perturbation calculations mutating the methanol of Cornell et al.[4] into the new model including "lone pairs" and polarizabilities. In the process, we also decided to modify the R^*(HO) from zero to a small, finite value ($R^* = 0.2$ Å) and also to change the R^*(OH) from its OPLS methanol value to the OPLS ether value. The reasoning was simply that, ideally, there should be a single R^* for sp^3 oxygen, and a possible reason that alcohols and water have a larger R^* than ethers is simply to compensate for the zero R^* for the HO. The $\Delta\Delta G$ between methanol from ref. 4 and this methanol was -0.25 ± 0.1. Because the Cornell et al. has an electrostatic solvation free energy of methanol that is slightly less favorable by ~0.5 kcal/mol than experiment, this new model is a slight improvement. For N-methyl acetamide (NMA), mutating the Cornell et al. model to the LP model (with $R^*(HN) = 0.85$ Å) leads to a relative solvation free energy of 0.0 ± 0.1 kcal/mol. Given the controversy about the actual ΔG solvation of NMA,[33] one can only conclude that the model is as accurate or inaccurate as in ref. 4.

PARTITION COEFFICIENTS OF THE NUCLEIC ACID BASES

In Table VI, we present the free energies for mutating the charges of the Cornell et al. model (p94) into the new model (EP) in water and the 0.9 scaled charges of the Cornell et al. model (0.9p94) into the EP model in $CHCl_3$. In each case, these ΔG are the difference between the gas phase and solution mutations. The effect on the relative solvation free energies of the bases (water vs. $CHCl_3$) in the new model vs. the CS2 model of Eksterowitz et al.[6] (which used p94 charges in water and 0.9p94 charges in $CHCl_3$) is reported in the third column of Table VI.

In Table VII, these free energies are used to calculate the $\log P$ values for this new model (third column). As one can see, the calculated partition coefficients are improved for A, C, and U, and significantly worse for G and T. The calculated dipole moments are also reported in Table VII. Whereas fixed-charge models cannot respond differentially to environment, the EP model leads to significantly larger aqueous dipole moments than in $CHCl_3$, where the dipole moment is very similar to that found in the gas. The intrinsic dipole (gas phase) of EP is significantly smaller than p94 for the pyrim-

TABLE VI.
Calculation of Solvation Free Energies of Nucleic Acid Bases.[a,b]

Base	ΔG(p94 → EP)wat[a]	$\Delta\,\Delta G$(0.9p94 → EP), $CHCl_3$[b]	$\Delta\,\Delta G$(wat-$CHCl_3$)
A	−0.9 ± 0.05	−0.6 ± 0.05	−0.3
G	−1.7 ± 0.20	−0.8 ± 0.05	−0.9
C	1.0 ± 0.05	0.0 ± 0.20	1.0
U	0.8 ± 0.02	−0.3 ± 0.02	1.1
T	2.5 ± 0.05	0.0 ± 0.05	2.5

[a] $\Delta\,\Delta G$ for aqueous determined by the average of the direct perturbation of p94 → EP in water and in the gas, and in the difference between p94 → zero charges and EP → zero charges in water vs. gas.
[b] $\Delta\,\Delta G$ for $CHCl_3$ determined by the direct perturbation of p94 → EP in chloroform and in the gas phase.

idines C, U, and T, but comparable to that of p94 for the purines A and G. This leads, upon polarization, to negative free energies in the aqueous and chloroform solvation free energies (first two columns of Table VII) for the purines and positive for the pyrimidines. For the pyrimidines, the relative C and U free energies in Table VII make sense, given the smaller dipole moments of EP than p94 (in H_2O) and EP than 0.9p94 (in $CHCl_3$), but the reason for the much greater magnitude for the aqueous ΔG for T than U is not obvious.

Discussion and Conclusions

We have begun the development and application of a new nonadditive molecular mechanical force field that includes "lone pairs" at electron donor sites. We have shown that this model gives an excellent representation of nucleic acid base–base and amide–water H-bond energies and solvation free energies of methanol and N-methyl acetamide.

In a more challenging application, we have applied the model to the chloroform/water partition coefficients of the N-methylated bases G, A, U, T, and C. The model does a reasonable job, better than fixed charge models (average absolute error (AAE) in partition coefficients 1.0–1.2), with an AAE = 0.8. The model where different fixed charges are used in $CHCl_3$ and H_2O has an AAE of 0.5 and the SM semiempirical solvation models achieve an AAE of only 0.2. But given that the SM model has been empirically calibrated on related molecules for both water and $CHCl_3$, such better representation of the partition coefficients is not surprising.

One should also not forget that the partition coefficients are calculated from the difference between two large solvation free energies, the largest being the ΔG_{solv} of G in water (~ -23 kcal/mol)

TABLE VII.
Partition Coefficients and Dipole Moments of Bases.

Base	log P(expt)[a]	log P(p94, 0.9p94)[b]	log P(EP)[c]	μ(g)[d]	$\mu(CHCl_3)$[e]	$\mu(H_2O)$[f]	μ(p94)[g]	μ(0.9p94)[h]
A	−0.8	0.3	0.1	2.5	2.6	3.0	2.9	2.6
G	−3.5	−3.7	−4.3	6.8	6.8	7.8	6.9	6.3
C	−3.0	−3.6	−3.0	5.9	6.0	7.5	8.2	7.3
U	−1.2	−1.6	−0.8	4.3	4.2	5.1	5.7	5.2
T	−0.4	−0.7	+1.4	4.2	4.1	5.1	5.8	5.2

[a] $P = [\text{base}]_{CHCl3}/[\text{base}]_{water}$. This is the experimental value.[36]
[b] Value calculated by Eksterowitz et al.[6]
[c] Value calculated by correcting those log P from Eksterowitz et al.,[6] with the $\Delta\,\Delta G$ values for Table VI. For example, the $\Delta\,\Delta G$ of −0.6 kcal/mol for A corresponds to a −0.4 change in log P because $\Delta G = -1.4 \log P$.
[d] Dipole moment of EP model in the gas phase, averaged over 100 ps.
[e] Dipole moment of EP in water, averaged over 100 ps.
[f] Dipole moment of EP in chloroform, averaged over 100 ps.
[g] Dipole moment from model of N-Methyl base in ref. 4.
[h] Dipole moment from model of N-Methyl base in ref. 4, scaling the charges by 0.9.

and in $CHCl_3$ (~ -18 kcal/mol). Thus, an average $\Delta \log P$ of 0.8, which corresponds to a $\Delta\Delta G_{solv}$ of $\sim$1 kcal/mol, suggests that our proposed model performs reasonably. There are many avenues for parameter adjustment (without changing the permanent charges). By increasing the polarizabilities, one can have a significant effect on water solvation free energies without much affecting $CHCl_3$ solvation free energies. Subtle differences in van der Waals parameters could significantly affect the $CHCl_3$ solvation free energies, without much affecting the water solvation free energies.

We should also note that our solvation free energies were calculated with a model in which the solvents were described at an additive level, and only the solute was polarizable. This was done with a software modification that has led to a more efficient nonadditive model (only $\sim$2 times slower than additive) that should be applicable to MD simulations of proteins and nucleic acids. The charge parameters for such simulations, using the new charge derivation method presented here, have been developed,[30] and the software for PME simulations of nonadditive models with the inducible dipoles and off-center charges has been developed by Darden and coworkers.[34] Thus, we are in a position to extend our model to more complex molecules.

References

1. Dunning, T., Jr. J Chem Phys 1989, 90, 1007.
2. Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. J Phys Chem 1993, 97, 10269.
3. Cieplak, P.; Cornell, W.; Bayly, C.; Kollman, P. J Comput Chem 1995, 16, 1357.
4. Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K., Jr.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. J Am Chem Soc 1995, 117, 5179.
5. Jorgensen, W.; Chandresekhar, J.; Madura, J.; Impey, R.; Klein, M. J Chem Phys 1983, 79, 926.
6. Eksterowitz, J.; Miller, J.; Kollman, P. J Phys Chem B 1997, 101, 10971.
7. Warshel, A.; Levitt, M. J Mol Biol 1976, 103, 227.
8. Stern, H.; Kaminski, G.; Banks, J.; Zhou, R.; Berne, B.; Freisner, R. J Chem Phys B 1999, 103, 4730.
9. Brooks, C., unpublished results.
10. Sprik, M. J Phys Chem 1991, 95, 2283.
11. Kaminski, G.; Jorgensen, W. J Chem Soc Perkin Trans 1999, 2, 2365.
12. Ding, Y.; Bernardo, D.; Krogh-Jespersen, K.; Levy, R. J Phys Chem 1995, 29, 11575.
13. Caldwell, J.; Dang, L.; Kollman, P. J Am Chem Soc 1990, 112, 9144.
14. Dang, L.; Rice, J.; Caldwell, J.; Kollman, P. J Am Chem Soc 1991, 113, 2481.
15. Cieplak, P.; Kollman, P. J Chem Phys 1990, 92, 6761.
16. Caldwell, J.; Kollman, P. J Am Chem Soc 1995, 117, 4177.
17. Meng, E.; Caldwell, J.; Kollman, P. J Phys Chem 1996, 100, 2367.
18. Sun, Y.; Caldwell, J.; Kollman, P. J Phys Chem 1995, 99, 1008.
19. Caldwell, J.; Kollman, P. J Phys Chem 1995, 99, 6208.
20. Dixon, R.; Kollman, P. J Comput Chem 1997, 18, 1632.
21. Applequist, J.; Carl, J.; Fung, K.-K. J Am Chem Soc 1972, 94, 2952.
22. Wang, J.; Cieplak, P.; Kollman, P. J Comput Chem 2000, 21, 1049.
23. Hariharan, P.; Pople, J. Theor Chim Acta 1973, 28, 213.
24. Becke, A. J Chem Phys 1993, 98, 5648.
25. Winn, P.; Ferenczy, G.; Reynolds, C. J Comput Chem 1999, 20, 704.
26. Rychaert, J.; Ciccotti, G.; Berendsen, H. J Comput Phys 1977, 23, 327.
27. Fox, T.; Kollman, P. J Phys Chem B 1998, 102, 8070. This ref. used ε(Cl) of 0.2550 kcal/mol; we found better liquid properties with $\varepsilon = 0.325$ kcal/mol, with all other parameters the same.
28. Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. J Chem Phys 1984, 81, 3684.
29. Anderson, H. J Chem Phys 1980, 72, 1284.
30. Cieplak, P.; Wang, J.; Caldwell, J.; Kollman, P., manuscript in preparation.
31. Sponer, J.; Leszczynski, J.; Hobza, P. J Phys Chem 1996, 100, 1965.
32. Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. J Comput Chem 1997, 18, 11336.
33. Spector, T.; Kollman, P. J Phys Chem B 1998, 102, 4004.
34. Darden, T., unpublished.
35. Dang, L.; Chang, T. J Chem Phys 1997, 106, 8149.
36. Cullis, P.; Wolfenden, R. Biochemistry 1981, 20, 3024.