Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions*

THOMAS A. HALGREN

Department of Molecular Design and Diversity, Merck Research Laboratories, P.O. Box 2000, Rahway, New Jersey 07065

Received 20 March 1995; accepted 31 August 1995

ABSTRACT

This article defines the parameterization and performance of MMFF94 for intermolecular interactions. It specifies the novel "buffered" functional forms used for treating van der Waals (vdW) and electrostatic interactions, and describes the use of: (1) high quality *ab initio* data to parameterize vdW interactions involving aliphatic hydrogens; and (2) HF/6-31 G* calculations on hydrogen-bonded complexes to parameterize nonbonded interactions in polar systems. Comparisons show that appropriate trends in the HF/6-31G* data are well reproduced by MMFF94 and that intermolecular interaction energies and geometries closely parallel those given by the highly regarded OPLS force field. A proper balance between solvent–solvent, solvent–solute, and solute–solute interactions, critically important for prospective success in aqueous simulations, thus appears to be attained. Comparison of MMFF94, OPLS, CHELPG electrostatic potential fit, QEq, Gasteiger, and Abraham charges for 20 small molecules and ions also shows the close correspondence between MMFF94 and OPLS. As do OPLS and all current, widely used force fields, MMFF94 employs "effective pair potentials" which incorporate in an averaged way the increases in polarity which occur in high dielectric media. Some limitations of this approach are discussed and suggestions for possible enhancements to MMFF94's functional form are noted.

* This article includes Supplementary Material available from the author upon request or via the Internet at ftp.wiley.com/public/journals/jcc/suppmat/17/520 or http://journals.wiley.com/jcc

Journal of Computational Chemistry, Vol. 17, Nos. 5 & 6, 520–552 (1996)

CCC 0192-8651 / 96 / 050520-33

Introduction

Part I in this series[1-4] summarized the form, parameterization, and performance of the newly developed Merck Molecular Force Field (MMFF94). This second part provides a full description of the parameterization and performance of the core, computationally derived, portion of MMFF94 for nonbonded van der Waals (vdW) and electrostatic interactions. As in Part I, comparisons will also be presented to characterize the performance of our previous-generation MM2X force field. The nonbonded interactions with which this study is concerned are critical ones for a molecular force field to describe accurately, as they largely serve to determine how successful the force field will be at describing solvation and the energetics of host–guest binding.

The next section of this article fully defines the expressions used for treating vdW and electrostatic interactions and discusses the basis for the choices made in MMFF94. The third section describes MMFF94's parameterization for nonbonded interactions in aliphatic systems, using high-quality quantum-mechanical data as benchmarks, and shows the degree to which MMFF94, MM2X, MM2, MM3, and CHARMm (QUANTA 3.2 parameterization) reproduce the quantum-mechanical data. The fourth section then specifies the data and procedures used to determine initial electrostatic parameters and to adjust selected vdW and electrostatic parameters to reproduce quantum-mechanically calculated values and trends in intermolecular interaction energies and geometries. The fifth section presents representative values for the derived parameters and specifies how the complete set can be obtained; the related vdW parameters and partial atomic charges used for an enzyme-site "context molecule" in OPTIMOL/MMFF calculations[1] are specified in Appendix A (Supplementary Material). The sixth section compares the approach taken in MMFF with OPLS, CHELPG, QEq, Gasteiger, and other procedures for determining partial atomic charges, and also shows that MMFF94 charges usually agree well with OPLS charges but differ in characteristic ways from charges given by the other methods. The seventh section then shows MMFF94's ability to reproduce values and trends in the HF/6-31G* data on intermolecular interaction energies and geometries used in its parameterization. It also compares MMFF94 and OPLS results for hydrogen-bonded dimers and shows that the two approaches usually agree closely. Finally, the concluding discussion summarizes this work and notes enhancements needed to further improve the accuracy with which an empirical force field such as MMFF94 can describe nonbonded interactions.

The MMFF94 van der Waals and Electrostatic Potentials

VAN DER WAALS INTERACTIONS

As previously indicated,[1] MMFF uses a special "Buffered-14-7" (Buf-14-7) form [eq. (1)] for the van der Waals (vdW) potential in place of the Lennard–Jones or Exp-6 potentials employed in force fields such as MM2,[5] MM3,[6] AMBER,[7] VFF,[8] and CHARMm.[9] The Buf-14-7 form is defined as[10]:

$$E_{\mathrm{vdW}_{ij}} = \varepsilon_{IJ}\left(\frac{1.07R^*_{IJ}}{R_{ij} + 0.07R^*_{ij}}\right)^7\left(\frac{1.12R^{*7}_{IJ}}{R^7_{ij} + 0.12R^{*7}_{IJ}} - 2\right) \quad (1)$$

where R_{ij} is the interatomic distance. The parameters ε_{IJ} and R^*_{IJ} are the well depth in kilocalories per mole and the minimum-energy separation in angstroms. These quantities are written using capitalized subscripts to indicate that they depend on the MMFF atom types I and J to which atoms i and j correspond. As we have shown elsewhere,[10] the Buf-14-7 potential describes the experimentally well-characterized interactions of rare-gas atoms far more accurately than do the Lennard–Jones or Exp-6 potentials. As is commonly done, vdW and electrostatic interactions are computed whenever atoms i and j belong to separate domains or are separated by three or more chemical bonds. 1,4-vdW interactions are not scaled in MMFF94.

For like-pair interactions, the minimum-energy separations are obtained from assigned atomic polarizabilities, α_I, via the relationship

$$R^*_{II} = A_I\alpha_I^{0.25} \quad (2)$$

where A_I is a scale factor that is usually taken to be invariant across a row of the periodic table.[10] To obtain the minimum-energy separations R^*_{IJ} for

unlike pairs, MMFF94 uses an "augmented arithmetic mean":

$$R^*_{IJ} = 0.5(R^*_{II} + R^*_{JJ})\left(1 + B\left(1 - \exp(-\beta^* \gamma^2_{IJ})\right)\right) \quad (3)$$

$$\gamma_{IJ} = (R^*_{II} - R^*_{JJ})/(R^*_{II} + R^*_{JJ}) \quad (4)$$

This mean behaves similarly to the previously described "cubic mean" when R^*_{II} and R^*_{JJ} differ by less than about a third, as is normally the case. Both increase R^*_{IJ} relative to the commonly used[5-7,9] arithmetic mean, which gives $R^*_{IJ} = 0.5(R^*_{II} + R^*_{JJ})$. When applied to the data in Table V of a previously published report,[10] the root mean square (rms) deviation from the rare-gas minimum-energy distances is 0.042 Å for the augmented arithmetic mean when the values used in MMFF94 (B = 0.2 and $\beta = 12$) are employed. This rms deviation is comparable to that of 0.040 Å rms for the cubic-mean rule, and is substantially better than that of 0.133 Å found for the simple arithmetic mean.[10] The performance of MMFF94 therefore remains comparable to that reported for the novel combination rules recently introduced by the Waldman and Hagler,[11] who included the cubic-mean rule in their comparisons. The augmented arithmetic and cubic means differ significantly, however, as γ_{IJ} approaches unity. Depending on the methodology employed, this limit can be reached in thermodynamic integration and perturbation studies[12] when atoms are made to disappear. When the arithmetic mean is used, R^*_{IJ} approaches $0.5R^*_{II}$ as R^*_{JJ} approaches zero. In contrast, R^*_{IJ} approaches $1.0R^*_{II}$ when the cubic mean is employed, but approaches $0.6R^*_{II}$, a seemingly more reasonable limit, when the augmented arithmetic mean is used with $B = 0.2$.

For vdW well depths, a Slater–Kirkwood-based formula is used for both like- and unlike-pair interactions:

$$\varepsilon_{IJ} = \frac{181.16 G_I G_J \alpha_I \alpha_J}{(\alpha_I/N_I)^{1/2} + (\alpha_J/N_J)^{1/2}} \frac{1}{R^{*6}_{IJ}} \quad (5)$$

where N_I and N_J are the Slater–Kirkwood effective numbers of valence electrons, and G_I and G_J are scale factors. As previously described,[10] the G and N parameters are taken to depend on the atomic species (rather than on the full MMFF94 atom types) and are obtained from systematic relationships based on high-quality experimental data on vdW interactions of rare-gas atoms and of small molecules with one another. Some of the atomic polarizabilities used in computing vdW radii were obtained from additive decompositions of experimentally determined molecular polarizabilities.[10] Others, as described below, have been adjusted to best reproduce values and trends in *ab initio* intermolecular interaction energies and geometries (cf. the third and fourth sections of this article).

Modifications are made to eqs. (3) and (5) for interactions involving polar hydrogens. Specifically, when either i or j corresponds to such an atom, we use the simple arithmetic mean [or, equivalently, use $B = 0.0$ in eq. (3)] to obtain R^*_{IJ}. In addition, if I–J is classified as a donor–acceptor interaction, we further scale R^*_{IJ} by the factor DARAD = 0.8, and scale ε_{IJ} as given by eq. (5) (and calculated using the unscaled R^*_{IJ}) by the factor DAEPS = 0.5. This procedure in effect reduces polar hydrogens in size when they interact with hydrogen-bond acceptors, as we have found to be needed to properly describe hydrogen-bonding interactions, but retains their "normal" (larger) size in other cases. We found in part IV[3] that the larger vdW size is needed to reproduce conformational energies in key cases, for example, for cyclohexanol.

In addition to the article on combination rules by Waldman and Hagler,[11] two other recent studies on vdW interactions are of interest. In the first, Hart and Rappé[13] showed that the Morse function is superior to the Lennard–Jones 12-6 (LJ-12-6) and Exp-6 forms for describing the shape of the vdW potential in reference systems. The Morse form also avoids the excessive repulsion of the LJ-12-6 form at short contact distances and, like the Buf-14-7 and the experimental rare-gas potentials,[10] rises more quickly than R^{-6} for distances beyond the vdW minimum. This form, however, requires an additional "shape" parameter whose value depends on the types of the interacting atoms. In our view, the more complex exponential form and the need for a third interaction-dependent parameter make the Morse potential less attractive than the Buf-14-7 form for use in molecular-mechanics calculations.[14]

In another study, Cambi et al.[15] relate atomic polarizability to minimum-energy separations, R^*, and well depths, ε, in manners qualitatively similar to ours.[10] Furthermore, their calculation of well depths involves systematic relations for the Slater–Kirkwood effective number of electrons N_{eff} that are analogous to, but broader than, those for the equivalently defined N_I given in eqs. (20)–(24)

of our previous article.[10] Formal comparison of the respective formulas for ε shows that $1.44^2 N_{\text{eff}} = G_I^4 N_I$. Thus, their approximation for N_{eff} incorporates some of the enhancement from C_8, C_{10}, and higher order dispersion terms that our approach accomodates through the scale factors G_I. The systematic relationships they define for N_{eff} may provide some guidance for how N_I should behave outside the ranges for which we have defined it. Unfortunately, however, their approach appears to be too inflexible. For example, the well depth for He $\cdots$ He calculated using their approach is 0.068 kcal/mol, or three times the experimental value of 0.022 kcal/mol. Moreover, their relationship between atomic polarizability and R_m (R^* in our notation) leads to an infinite value for R_{IJ}^* as R_{JJ}^* goes to zero. This behavior does not seem physically appropriate.

ELECTROSTATIC INTERACTIONS

MMFF uses a simple modification of the standard coulombic form:

$$EQ_{ij} = 332.0716 q_i q_j / \left(D(R_{ij} + \delta)^n \right) \qquad (6)$$

where q_i and q_j are partial charges on atoms i and j, R_{ij} is the internuclear separation in angstroms, $\delta = 0.05$ Å is the *electrostatic buffering* constant, and D is the dielectric constant. As in the case of vdW interactions, only terms involving atoms separated by three or more bonds are retained; for reasons addressed in part IV,[3] electrostatic 1,4-interactions are scaled by a factor of 0.75. The "distance buffering" provided by the positive value for δ is needed to prevent infinite attractive electrostatic energies from overwhelming the bounded (but large) repulsive vdW term in eq. (1) as oppositely charged atomic centers coalesce. As in the case of eq. (1), the motivation for this form derives from the use of distance buffering in related contexts.[16,17] Both constant ($n = 1$) and distance-dependent ($n = 2$) dielectric models are supported; the default values are $D = 1.0$ and $n = 1$. In contrast to MM2 and MM3, which use a default dielectric constant of 1.5, MMFF94's value has been chosen to facilitate application to condensed phase processes, particularly ones with water as explicit solvent.

The partial atomic charges q in eq. (6) are assembled from initial (usually zero) integral or fractional formal atomic charges by adding *bond-charge-increment* contributions, ω_{KI}, that reflect the polarity of the bonds to i from attached atoms k. Specifically, we take:

$$q_i = q_i^0 + \sum \omega_{KI} \qquad (7)$$

where q_I^0 is the integral or fractional formal atomic charge (e.g., $+1/3$ on guanidinium nitrogens) and ω_{KI} represents the partial charge contributed to atom i from the i–k bond (note that $\omega_{IK} = -\omega_{KI}$). A similar approach has long been used in MM2X,[18] our previous-generation force field, but the assignment of the q_I^0 in some cases is more elaborate for MMFF. For certain atom types—currently, those assigned a negative integral or fractional formal atomic charge in the primary atom-typing process—a portion (e.g., half) of the initially assigned formal charge is transferred to (shared with) the bonded atom(s) k. Charge sharing is discussed in part V.[4]

Parameterization of Nonbonded Interactions in Aliphatic Systems

In this section, we describe how MMFF94 has been parameterized and validated for intermolecular interactions involving aliphatic hydrogens. We also show how well MM2X, MM2, MM3, and CHARMm (QUANTA 3.2 parameterization) reproduce the *ab initio* data on intermolecular interactions used in MMFF94's parameterization.

PARAMETERIZATION OF MMFF94

We determined the vdW parameters for aliphatic hydrogen by fitting to quantum-mechanical interaction energies calculated for the methane and hydrogen dimers as a function of separation and orientation. As in MM2 and MM3, we take the partial charge on hydrogen atoms attached to an aliphatic carbon as zero. For the methane dimer, we utilized the interaction energies calculated by Scheiner and coworkers.[19] Their calculations were performed for the six orientations A–F shown in Figure 1 of their article, and used Sadlej's "medium polarized" basis sets.[20] Though of moderate size, $(10s,6p,4d/5s,4p)$ contracted to $[5s,3p,2d/3s,2p]$, these basis sets contain sufficiently diffuse polarization functions to allow high-quality interaction energies to be obtained.[19] The interaction energies cited in Table I as $DE^{(2)}$ are the results they obtained at second order after correction for basis-set superposition errors using the Boys counterpoise method.[21]

TABLE I.
Interaction Energies (kcal / mol) for the Methane Dimer.

R_{CC} (a_o)	DE$^{(2)}$ [a]	MMFF94	MM2X	MM2	MM3	CHARMm[b]
Orientation A ($HCH_3 \cdots H_3CH$)						
6.0	0.73	0.66	−0.29	−0.46	0.41	−0.69
6.5	−0.15	−0.24	−0.59	−0.91	−0.27	−0.74
7.0	−0.33	−0.40	−0.57	−0.84	−0.41	−0.59
7.5	−0.35	−0.36	−0.46	−0.66	−0.38	−0.43
8.0	−0.29	−0.27	−0.36	−0.49	−0.31	−0.31
8.5	−0.22	−0.20	−0.28	−0.35	−0.24	−0.22
9.0	−0.16	−0.14	−0.21	−0.25	−0.18	−0.16
10.0	−0.09	−0.07	−0.12	−0.14	−0.10	−0.09
Orientation B ($HCH_3 \cdots H_2CH_2$)						
6.0	1.84	1.91	0.39	0.70	1.31	−0.13
6.5	0.29	0.12	−0.42	−0.59	0.05	−0.67
7.0	−0.20	−0.32	−0.54	−0.78	−0.31	−0.61
7.5	−0.31	−0.35	−0.48	−0.67	−0.36	−0.46
8.0	−0.28	−0.28	−0.38	−0.51	−0.31	−0.34
9.0	−0.17	−0.15	−0.22	−0.27	−0.18	−0.17
10.0	−0.09	−0.08	−0.12	−0.14	−0.10	−0.09
Orientation C ($H_2CH_2 \cdots H_2CH_2$)						
8.0	−0.28	−0.29	−0.39	−0.53	−0.31	−0.36
Orientation D ($HCH_3 \cdots HCH_3$)						
6.0	4.59	9.11	6.24	6.34	5.02	8.56
6.5	1.37	2.76	1.18	1.32	1.52	1.04
7.0	0.16	0.55	−0.11	−0.20	0.23	−0.31
7.5	−0.22	−0.11	−0.38	−0.54	−0.18	−0.45
8.0	−0.29	−0.24	−0.37	−0.51	−0.26	−0.37
8.5	−0.26	−0.21	−0.31	−0.40	−0.24	−0.27
9.0	−0.21	−0.16	−0.24	−0.30	−0.19	−0.20
10.0	−0.11	−0.09	−0.14	−0.16	−0.20	−0.10
Orientation E ($H_2CH_2 \cdots HCH_3$)						
8.0	−0.20	−0.19	−0.36	−0.47	−0.20	−0.38
Orientation F ($H_3CH \cdots HCH_3$)						
6.0	27.88	153.44	843.18	76.06	32.42	5560.65
7.0	4.08	11.10	14.73	7.70	4.89	31.50
7.5	1.37	3.18	2.52	1.88	1.51	3.80
8.0	0.33	0.72	0.24	0.17	0.32	0.24
9.0	−0.13	−0.13	−0.24	−0.29	−0.15	−0.22
10.0	−0.11	−0.10	−0.16	−0.18	−0.12	−0.12

[a] Total interaction energy through second order from ref. 19.
[b] CHARMm parameters as accessed by QUANTA (Version 3.2) from Molecular Simulations, Inc. (Burlington, MA).

For the hydrogen dimer, we calculated intermolecular interaction energies for the "crossed" (D_{2d}), "linear" ($D_{\infty h}$), "parallel" (D_{2h}), and "tee" (C_{2v}) orientations discussed by Wiberg and Murcko.[22] We used the Sadlej basis set for hydrogen[20] supplemented by one set of Gaussian d functions having an exponent of 0.075, a value chosen to be about midway between those of the two most diffuse p functions (0.1027 and 0.0324). The calculations were performed at the MP4(SDTQ) level[23] with Gaussian 88[24] using the 5D representation for the d functions. Each dimer is characterized by its orientation and by the separation parameter, R_{HH}, which in each case measures to the shortest contact distance between hydrogen atoms in different H_2 monomers.[25] As in Scheiner's work,[19] the Boys counterpoise method was used to correct for basis-set superposition error. The computed interaction energies are listed in Table II.

In parameterizing MMFF94, we used the published vdW parameters for aliphatic carbon[10] (MMFF numerical atom type 1) and varied the atomic polarizability α_I and the scale parameter A_I [eq. (2)] for hydrogen (numerical atom type 5)

TABLE II.
Interaction Energies (kcal / mol) for the Hydrogen Dimer.

R_{HH} (Å)	MP4	MMFF94	MM2X	MM2	MM3	CHARMm[a]
Crossed						
2.381	2.080	0.493	0.106	1.114	1.122	−0.019
2.646	0.728	−0.004	−0.142	0.017	0.234	−0.168
3.704	−0.037	−0.036	−0.064	−0.110	−0.065	−0.043
Linear						
1.700	2.354[b]	4.734	4.714	4.947	3.345	7.743
2.010	0.621	1.042	0.603	0.937	0.858	0.665
2.170	0.344	0.415	0.145	0.288	0.376	0.117
3.000	−0.028	−0.041	−0.073	−0.115	−0.059	−0.055
Parallel						
2.010	6.443	3.258	1.992	6.444	4.551	2.034
2.381	1.762	0.342	0.026	0.764	0.857	−0.074
2.646	0.625	−0.031	−0.150	−0.058	0.158	−0.165
3.704	−0.029	−0.034	−0.061	−0.105	−0.063	−0.041
Tee						
2.117	1.305	1.208	0.617	1.677	1.438	0.565
2.275	0.634	0.462	0.131	0.595	0.664	0.058
3.334	−0.075	−0.041	−0.071	−0.117	−0.065	−0.050

[a] CHARMm parameters as accessed by QUANTA (Version 3.2) from Molecular Simulations, Inc. (Burlington, MA).
[b] MP3 result; the MP4 calculation was unstable. See text for basis set and procedure used.

until a good fit to the calculated interaction energies was obtained. The final values were $\alpha_i = 0.25$ Å^3 and $A_i = 4.20$ Å$^{1/4}$. The value for A_i differs slightly from the published[10] H, He value of 4.40 Å$^{1/4}$, which it supersedes. These values give $\varepsilon_{II} =$ 0.0216 kcal/mol and $R^*_{II} = 2.970$ Å. These parameters are used for all hydrogens attached to carbon (i.e., all hydrogens of numerical atom type 5).[1] MM3 uses a similar well depth (0.020 kcal/mol), but employs a somewhat larger minimum-energy separation (defined as twice the vdW radius) of 3.24 Å.[6]

COMPARISON OF RESULTS FOR MMFF94, MM2X, MM2, MM3, AND CHARMm

Tables I and II show that the MMFF94 parameters fit the *ab initio* results for the methane dimer fairly well but, while clearly of the right scale, describe the hydrogen dimer less satisfactorily. For the methane dimer, only the MM3 interaction energies agree as well as do those for MMFF94. Moreover, MM3 is superior for the repulsive head-to-head orientation *F*. It appears to benefit from the hydrogen offset that foreshortens the C—H bond and increases the effective H ··· H contact distance. This foreshortening is amply supported by X-ray crystallographic studies that find comparable shortenings of ~ 0.1 Å in bond lengths to hydrogen to be required to best fit the molecular electron density with spherical atoms. A similar hydrogen offset undoubtedly would help in MMFF. For now, however, we omit it in the interest of computational speed.

Table I also shows that none of the other tested force fields performs as well as MMFF94 or MM3. The QUANTA 3.2 parameters for CHARMm give far too repulsive an interaction at short H ··· H contacts for head-to-head orientation *F*, as would be expected from the improper short-range behavior of the LJ-12-6 potential.[10] Interestingly, the same parameters also give insufficient repulsion at close contact distances for the interdigitated orientations *A* and *B* and lead to too-negative interaction energies in the region of the potential-energy minimum. The MM2 potential and the very similar MM2X potential also fare poorly. Both are much too stabilizing in the region of the vdW minima, as had previously been noted for MM2.[26]

For the hydrogen dimer, MM3 again fares best (Table II), although it appears to be too stabilizing in the attractive well region (an alternative interpretation would be that the quantum-mechanical calculations significantly underestimate the well depths). Again, MMFF94 also performs reasonably well. Except for the linear case, the orientations examined in Table II probe the vdW properties near the center of the H—H bond and thus partic-

ularly strongly highlight the utility of the bond foreshortening used in MM3.[27] MM2X again is much too stabilizing in certain instances, as to a lesser degree are MM2 and CHARMm. In each case, the well depth ϵ_H of ~ 0.040 kcal/mol appears to be too large. For CHARMm, in addition, the minimum-energy separation R^*_{HH} of 2.66 Å seems to be too small.[28]

Parameterization of vdW and Electrostatic Interactions in Hydrogen-Bonding Systems

The computational data employed in the parameterization consisted of two parts: (1) HF/6-31G* calculations[29] of molecular dipole moments for 423 HF/6-31G*-optimized geometries,[30] for most of which we also obtained electrostatic-potential fit charges by fitting to the HF/6-31G* molecular electrostatic potential using Williams' PDM88 method[31]; and (2) HF/6-31G* calculations for the 66 HF/6-31G*-optimized dimers shown in Figure 1. HF/6-31G* calculations are reported to give good descriptions of hydrogen-bonded complexes.[32] While these calculations use neither large basis sets nor correlated wave functions, it should be noted that interaction energies and dipole moments are notoriously difficult to calculate accurately even when large basis sets and correlated wave functions are used. The extensive study by Frisch et al. of the H_2O dimer and of the monomer dipole moment is both illuminating and sobering in this regard,[33] as is the length to which Bartlett and coworkers had go to define the gas-phase dimerization energy with reasonable precision.[34] Furthermore, even accurate data on gas-phase systems would be of limited utility, for MMFF94 uses effective-pair potentials that reflect in an averaged way the enhancement of the molecular dipole moment in an aqueous medium due to molecular polarizability. To examine the sensitivity of the calculated interaction energies to the choice of theoretical model, we also corrected the HF/6-31G* interaction energies for basis-set superposition error, and we carried out single-point MP2/6-31 + G** calculations at the HF/6-31G* geometries.[29] For a few systems, we also optimized the geometries at the MP2/6-31G* level. We found, in some cases, distinct differences in interaction energies and geometries (discussed later), but no indication that a parameterization based on any of these more complex computational models would prove superior.

As described in more detail in the following subsections, the parameterization of MMFF94 was carried out in three stages. In the first, we determined bond charge increments, ω_{KI} [eq. (7)], from fits to HF/6-31G* dipole moments. The HF/6-31G* model usually exaggerates observed gas-phase dipole moments by 10–20%, approximately the amount needed in a method that employs effective-pair potentials to account in an averaged way for the increase in polarity that occurs in a high-dielectric medium as a consequence of molecular polarizability.[35] In the second stage, we adjusted particular vdW and bond-charge-increment parameters to best reproduce scaled HF/6-31G* interaction energies and hydrogen-bond geometries. In these fits, the targeted interaction energies were 10–15% larger in magnitude than the HF/6-31G* values and the targeted heteroatom distances in hydrogen bonds were 0.2–0.3 Å shorter. MacKerell and Karplus[36] have recently employed a very similar approach, on which the present approach is patterned. These adjustments are also compatible with those Jorgensen and coworkers have found to be needed in parameterizing OPLS.[32,37] In the third and final stage, we achieved self-consistency by rederiving an appropriate subset of the bond charge increments from fits to scaled HF/6-31G* dipole moments, and by then refitting the scaled HF/6-31G* data for the dimers.

INITIAL FIT TO HF/6-31G* DIPOLE MOMENTS

This fit used the Biosym Consortium program PROBE[38] to determine bond-charge-increment parameters ω_{KI}. It employed a simultaneous least-squares fit to HF/6-31G* dipole moment components μ_x, μ_y, and μ_z for a set of 423 molecular conformers that covered the full range of the core MMFF94 parameterization.[1] An inherent limitation of this approach is that the dipole moment is not a "local" quantity. As a result, certain linear combinations of the ω_{KI}, and hence, of the partial atomic charges, can be varied to a large degree without substantially increasing the error in the least-squares-fit criterion. A better approach would have been to determine the ω_{KI} by carrying out a simultaneous fit to an equivalent set of HF/6-31G* electrostatic potentials, but we lacked this capability. To make the fit to the dipole moments determinant, we constrained certain of the charges (usually those of hydrogens and other terminal

atoms) by fixing the associated bond-charge-increment parameter in the following ways: (a) For hydrogen attached to saturated carbon (numerical atom types 5 and 1, respectively), we set the charge on hydrogen to zero, as do MM2 and MM3. Other procedures, for example, those discussed later (see "Comparison of Methods for Determining Partial Atomic Charges" section) typically give nonzero charges that range up to about +0.15 for aliphatic hydrogens, although the Bader procedure usually yields a negative charge of −0.05 to −0.10.[39] We show elsewhere[3] that the choice, $q_H = 0$, is a reasonable one. In particular, MMFF torsion parameters derived when a positive charge of +0.05 to +0.10 is assigned to such hydrogens are unable to reproduce certain significant patterns in conformational energies. (b) For hydrogen attached to olefinic and aromatic carbon (MMFF type 5 for hydrogen, types 2 and 37 for carbon), we fixed the charge on hydrogen at +0.15. This value has been shown to account for the trace of the molecular quadrupole moment of benzene and to be appropriate for reproducing the preference for the edge-face orientation for the benzene dimer.[40] In OPLS, Jorgensen and coworkers employ a similar, though somewhat smaller, charge of +0.115 for aromatic hydrogens.[41] (c) Because too little computational data were available for substituted cyclobutanes and cyclobutenes (MMFF atom types 20 and 30 for carbon), we set the bond charge increments to the values found in comparable acyclic systems. (d) Finally, for polar hydrogen atoms, we fixed the associated bond charge increments at values that reflected trends found in the HF/6-31G* electrostatic potential fit and/or Mulliken charges.

With these constraints in place, we carried out the least-squares fit to the HF/6-31G* dipole moments. Typically, the calculated variances in the derived bond-charge-increment parameters showed that fit was still indeterminant in a few instances. In such cases, we fixed additional bond charge increments at values chosen by analogy with trends observed in other, better determined parameters or in the dielectrostatic-potential-fit charges, and then repeated the least-squares fit. Ultimately, a determinant fit was achieved without undue difficulty.

INITIAL FIT TO HF/6-31G* INTERMOLECULAR INTERACTION ENERGIES AND GEOMETRIES

Except for polar hydrogens, the procedures described earlier provided initial values for all the vdW and electrostatic parameters involved in the HF/6-31G* dimer systems. To choose the vdW representation for polar hydrogens, we sought parameters that would reproduce the properties of the isolated linear water dimer given by the widely used SPC,[42] TIP3P,[43] and TIP4P[43] models as well as possible. These effective-pair-potential models have been shown to reproduce the properties of liquid water fairly well. Each gives an O···O distance of ~ 2.75 Å and an interaction energy of about −6.4 kcal/mol.[43] The comparison to the HF/6-31G* gas-phase values of 2.97 Å and −5.62 kcal/mol, respectively, illustrates the manner in which the HF/6-31G* results need to be adjusted[36]—that is, down for heteroatom distances, and up (in magnitude) for intermolecular-interaction energies. We then transferred the derived hydroxyl-hydrogen representation to all polar hydrogens.

In seeking to reproduce the behavior of the SCP, TIP3P, and TIP4P water models in MMFF, we found it necessary to assign a relatively small value for the atomic polarizability α for the hydroxyl hydrogen and to modify the associated A and N parameters [cf. eqs. (2) and (5)]. For reasons noted previously, we also used the simple arithmetic-mean combination rule for vdW interactions involving polar hydrogens. The derived value for α yields a minimum-energy separation R^*_{II} of 1.6 Å (vdW radius of 0.8 Å) for the hydroxyl hydrogen in water. This value is roughly half that found for aliphatic hydrogen. The comparison is not unreasonable, we think, based on the reduction in electron density around a polar hydrogen and on the likelihood that an even greater bond foreshortening than that used in MM2 and MM3 for aliphatic hydrogens should apply in this case.[44] The well depth, ε_{II}, on the other hand, has about the same value, roughly 0.02 kcal/mol, as that found for aliphatic hydrogen.

In contrast to MMFF94, OPLS,[45] AMBER,[7] and CVFF[8] all use a zero (or essentially zero) vdW radius and well depth for polar hydrogens, as indeed do the SPC, TIP3P, and TIP4P models for water.[43] Each of these other approaches compensates to some degree by adopting a somewhat larger vdW radius for the attached heteroatom. However, this approach makes polar hydrogens "invisible" to electrostatically uncharged atoms. Computational evidence suggests otherwise. Consider, for example, the linear H—F···Ne and F—H···Ne dimers of hydrogen fluoride with neon. The fact that the large basis set CEPA calculations of ONeil et al.[46] show that the F···Ne

distance at the potential-energy minimum is larger in F—H ··· Ne (3.31 Å) than in H—F ··· Ne (3.15 Å) implies that the strongly polar hydrogen in H—F has a finite vdW size. Moreover, these investigators also show that F—H ··· Ne is more stable than H—F ··· Ne. They argue, persuasively we think, that this greater stability arises because the electric field in H—F is stronger near hydrogen than near fluorine; this stronger electrostatic field in turn produces a much more strongly attractive induced-dipole (polarization) interaction with neon in F—H ··· Ne and, therefore, reduces the equilibrium F ··· Ne distance to a greater extent in F—H ··· Ne. Consequently, in the absence of the induced-dipole attraction, the F ··· Ne distance in F—H ··· Ne would exceed that in H—F ··· Ne by even more than the 0.16 Å obtained in the CEPA calculations.

We see no way to account for this qualitative difference other than by attributing a nonzero vdW radius to hydrogen in H—F. Note, however, that the theoretical calculations do not imply that the polar hydrogen radius is ~ 0.16 Å. Given that the vdW radius for fluorine is roughly 1.5 Å and that the H—F bond length is about 1 Å, a vdW radius for hydrogen of less than about 0.5 Å in effect would be hidden within that for fluorine. Thus, the real vdW radius for the hydrogen atom in HF must be somewhat greater than 0.5 Å. Additional computational support for the conclusion that polar hydrogens have significant vdW radii may be found in the anisotropy observed for exchange-repulsion interactions in the dimers of ammonia and water with argon[47] and in calculations we have performed for the linear H—F ··· He and F—H ··· He dimers[48] that find the F ··· He distances at the zero-crossing point and at the vdW minimum to be ~ 0.25 Å greater in the latter.

The adjustment of key vdW and electrostatic parameters was carried out iteratively. At each step, the HF/6-31G* dimers were optimized using the current estimates for the nonbonded parameters. Then, the optimized HF/6-31G* and MMFF binding energies were tabulated, as were the $X \cdots Z$ and $H \cdots Z$ distances and the X—H ··· Z angles, where X—H represents the hydrogen-bond donor and Z represents the heteroatom acceptor. As noted above, we sought to adjust the $X \cdots Z$ distances to make them 0.2–0.3 Å smaller than their HF/6-31G* counterparts wherever possible. These distances are affected mainly by the minimum-energy separations R_{II}^*, and hence by the atomic polarizabilities α_I [cf. eq. (2)]. Thus, where the MMFF distance was too large, we reduced α_Z and/or α_X. In contrast, the calculated binding energies depend primarily on the partial atomic charges of the interacting atoms, and thus on the associated bond charge increments. Where the MMFF binding energy was too small in magnitude relative to the scaled HF/6-31G* value, we adjusted the bond charge increment for the H—X bond or for one or more of the other bonds to X or Z to strengthen the electrostatic interaction. For the interaction between a donor water molecule and the acceptor nitrogen of pydidine, for example, we adjusted the polarity of pyridine's C—N bonds. This procedure required roughly a dozen iterations. The final parameters may partially reflect the starting values, but probably not to an inordinate degree. The quality of the fit to the *ab initio* data is discussed later.

ITERATION TO SELF-CONSISTENCY

In addition to modifying some of the vdW parameters, the just-described fit to the HF/6-31G* interaction energies and geometries increased many of the bond charge increments for the hydrogen-bonding moieties. To "systematize" this increase in polarity, we examined the fit of the resultant MMFF dipole moments to scaled HF/6-31G* dipole moments for the set of 423 structures employed previously. We found that the best fit resulted when the latter were increased by about 10%. Accordingly, we rederived the bond charge increments by fitting to HF/6-31G* dipole moments multiplied by this factor of 1.1. To preserve the adjustments made in fitting to the dimer data, we added to the constraint set the bond charge increments optimized in the dimer fit. Then, using the rederived bond charge increments as new initial values, we repeated the fit to the dimer data. These adjustment required little further change to either the vdW parameters or the bond charge increments, indicating that a reasonably self-consistent set of nonbonded parameters had been obtained.

PARAMETERS FOR METAL AND OTHER IONS

As described elsewhere, preliminary parameters have also been obtained for such protein metals as Zn^{2+}, Ca^{2+}, Mg^{2+}, Cu^{2+}, Cu^{+}, Fe^{2+}, and Fe^{3+}, as well as for the hydronium and hydroxide ions and for various halide and alkalai metal ions.[4]

MMFF94 van der Waals and Electrostatic Parameters

In this section we describe a subset of the MMFF94 nonbonded parameters. The full set of parameter, including those defined for the extended MMFF parameterization described in Part V,[4] contains approximately 500 bond charge increments and is too large to present here. The full set, however, is available in computer-readable form as supplementary material.[49] The corresponding MM2X parameters have been published as supplementary material to a previous article.[50] Listed in Tables III and IV are the parameters that arise from the following MMFF numerical atom types: 1, 2, 3, 5, 7, 10, 28, 31, and 70. The first seven describe saturated carbon, olefinic carbon, carbonyl carbon, hydrogen attached to carbon, carbonyl oxygen, amide nitrogen, and amide hydrogen, respectively. The last two are used for hydrogen and oxygen in water. Thus, these atom types cover simple alkanes, alkenes, aldehydes, ketones, and amides, in addition to water.

Table III lists the vdW parameters. Column 1 contains the MMFF numerical atom type, and column 2 contains the associated atomic polarizability (α_I) in units of cubic angstroms. Note that the effective-electron numbers N and the scale factors A and G in columns 3–5 usually depend only on the atomic species for the interacting atoms i and j, rather than on the full MMFF numeric atom type. Column 6, labeled "D_A_I" specifies whether the Ith atom type is considered to be a hydrogen-bond donor ("D"), acceptor ("A"), or neither ("—"). As noted previously, vdW interactions involving donors and acceptors are treated specially. For ease of readability, column 7 lists a representative MMFF symbolic atom type; as previously discussed, two or more symbolic atom types often map to the same numeric atom type. In the examples shown, as is generally the case, the first character of the symbolic atom type corresponds to the atomic species. Finally, the entry in column 8 indicates the origin of the parameter. The entries labeled "C94" specify parameters obtained in the fits to computational data described in this article, while entries labeled "E94" employ parameters taken from the previous study.[10] In the full parameter file, entries labeled "X94" represent parameters obtained in the extension of the parameterization described in part V.[4]

The bond-charge-increment parameters used in eq. (7) are shown in Table IV. The BT_{IK} are the "bond type indices" described in part III.[2] Note that the "0 1 2" line indicates that an olefinic carbon atom of type "2" in a "1—2" C—C bond gains a charge of −0.1382 from the polarity of the bond, while the aliphatic carbon atom of type "1" gains an equal and opposite charge of +0.1382. Thus, each bond charge increment gives the charge contributed to the second-listed atom, always the one of higher MMFF atom type. The parameter file is ordered by the integer canonical index:

$$CXQ = MC*(I*MA + K) + BT_{IK} \qquad (8)$$

TABLE III.
Partial List of MMFF94 vdW Parameters Used in Eqs. (2) – (5).[a]

MMF type	α_i	N_i	A_i	G_i	D_A_I	Symb.[b]	Origin[c]
1	1.050	2.490	3.890	1.282	—	CR	E94
2	1.350	2.490	3.890	1.282	—	C = C	E94
3	1.100	2.490	3.890	1.282	—	C = O	E94
5	0.250	0.800	4.200	1.209	—	HC	C94
7	0.650	3.150	3.890	1.282	A	O = C	C94
10	1.000	2.820	3.890	1.282	A	NC = O	E94
28	0.150	0.800	4.200	1.209	D	HNCO	C94
31	0.150	0.800	4.200	1.209	D	HOH	C94
70	0.870	3.150	3.890	1.282	A	OH2	C94

[a] The vdW parameters use the following values: B = 0.2, β = 12.0, DARAD = 0.8, and DAEPS = 0.5.
[b] See ref. 1 for a complete listing of MMFF94 numeric (column 1) and symbolic atom types. The symbolic atom types in this column are representative examples of the corresponding numeric atom type. In each case shown the first character identifies the atomic species.
[c] Indicates the source for or procedure used in developing the parameters (see text).

TABLE IV.
Partial List of MMFF94 Bond-Charge-Increment Parameters Used in Eq. (7).

Class	MMFF Types		ω_{KI}[a]	Origin / comment[b]
0	1	1	0.0000	#C94
0	1	2	−0.1382	C94
0	1	3	−0.0613	#C94
0	1	5	0.0000	#C94
0	1	10	−0.3001	C94
0	2	2	0.0000	#C94
1	2	2	0.0000	#C94
1	2	3	−0.0144	C94
0	2	5	0.1500	#C94
0	2	10	−0.1348	X94
1	3	3	0.0000	#C94
0	3	5	0.0600	#C94
0	3	7	−0.5700	#C94
0	3	10	−0.0600	C94
0	10	10	0.0000	#C94
0	10	28	0.3700	#C94
0	31	70	−0.4300	#C94

[a] See eq. (7). [b] See "MMFF94 van der Waals and Electrostatic Parameters" section.

where MC is an integer equal to at least the maximum permissible bond-type index +1, MA is an integer equal to the maximum numerical atom-type index +1, and atom type I is less than or equal to K. Parameter lines identified by the label "#C94" in the "Origin/comment" column list core MMFF94 bond charge increments that were included in the final constraint set cited previously; the "C94" label denotes parameters obtained in that fit. The full file also contains parameters labeled "X94," #X94," and "E94."[4]

The related nonbonded parameters used in OPTIMOL/MMFF calculations that utilize a rigid enzyme "active site" as a "context molecule"[18] are specified in Appendix A (Supplementary Material).[51]

Comparison of Methods for Determining Partial Atomic Charges

In this section we compare the bond-charge-increment approach with alternative approaches for determining partial atomic charges.

BOND-CHARGE-INCREMENT APPROACH

A principal virtue of the bond-charge-increment approach is that it allows MMFF94 to be applied straightforwardly to a wide range of chemical structures. This attribute is an especially important one in a pharmaceutical setting, in which it is difficult enough for a computational chemist to keep pace with the panoply of ideas and structures generated in a rapidly evolving medicinal chemistry project even when all needed parameters are at hand. Fortunately, this approach also enables conformational[3] and intermolecular-interaction energies (cf. next section) to be reproduced reasonably well. To be sure, it is too simplistic. To cite just one aspect, the polarity of a chemical bond of a given type is improperly taken to be constant, but in fact can depend significantly on the chemical environment. For example, an individual C—F bond in CF_4 is significantly less polar than the analogous bond in CH_3F because of "saturation" effects related to the finite ability of carbon to donate electrons. Nevertheless, the bond-charge-increment approach meets the tests of generality and accuracy adequately for the present purposes. Any other approach suitable for use in a force field like MMFF would also need to be able to do so.

OPLS APPROACH

Jorgensen and coworkers have derived OPLS nonbonded potentials for many common organic functional groups.[32,45] In Table V, OPLS charges for 20 small molecules and ions are compared to MMFF94 and other charge sets. The first 13 molecules were used by Carlson et al. in a study characterizing the ability of Mulliken and ESP charges to account for absolute and differential free energies of hydration.[35] To increase the structural diversity and to incorporate ionic species, we have included seven additional systems. Note that OPLS, which usually uses a united atom model, includes explicit hydrogens on carbon atoms only for benzene and for aniline (among cases considered). Except for pyridine and imidazole, however, the correspondence with MMFF is unaffected, because MMFF94 assigns zero charge to aliphatic hydrogen atoms.

While no account of the OPLS charges was taken in deriving MMFF94, comparison shows that the two charge sets usually behave very similarly. In particular, charges for polar hydrogens in uncharged molecules and for electronegative elements or groups attached to sp^3-hybridized car-

bons usually differ by 0.05 or less. A larger difference occurs for CH_3SH, where OPLS assigns a larger charge to the polar hydrogen (+0.27) than does MMFF94 (+0.18). This difference highlights one central distinction between MMFF and OPLS. Thus, while some OPLS parameters were obtained from simulations on neat liquids, others were derived by fitting to experimental enthalpies of hydration or by emulating gas-phase energies and geometries for HF/6-31G* dimers. In contrast, all core MMFF94 nonbonded parameters were derived in a common manner.

The two charge sets differ most dramatically for imidazole and for the methylammonium cation. For imidazole, the charge on N1 is −0.57 for OPLS but +0.033 for MMFF94, and that on the attached hydrogen is +0.42 for OPLS but only +0.27 for MMFF94. We note that N1 is a "pyrrole-type" nitrogen that contributes a pi-electron pair to the aromatic system and bears a formal positive charge in the "Kekule" structures that express the aromatic delocalization. In contrast, N3 and the three carbons each bear a negative formal charge in one of the resonance structures. As a result, while N1 need not bear a positive charge, it should at least be less negatively charged than N3. The MMFF94 and ESP charges (Table V) reflect this ordering, but the OPLS charges, whose origin has not been described in detail, do not.

For methylammonium, the MMFF94 charge set shows a much stronger polarization of the bonds to nitrogen. Partially as a result, the comparisons in the next section (third subsection) find the $CH_3NH_3^+ \cdots OH_2$ dimerization energy to be more negative for MMFF94 than for OPLS. MMFF's charge distribution was derived by assuming a N—H bond charge increment of 0.45, as we did not think that hydrogen attached to cationic nitrogen should bear a smaller positive charge than hydrogen attached to neutral nitrogen. We note that MMFF94 uses the same charge (+0.45) for the acidic hydrogens in the imidazolium cation and that OPLS assigns a comparable charge (+0.46) in this case. As noted earlier in this study, the MMFF94 charges for polar hydrogens have been chosen to mirror trends in ESP charges or, where these charges seemed not to be helpful (as in the case of methylammonium), in Mulliken charges. For reference, the HF/6-31G* Mulliken charges for polar hydrogens are 0.437 for methanol, 0.332 for methylamine, 0.090 for methanethiol, ~0.39 for acetamide, 0.468 for acetic acid, 0.359 for aniline, 0.399 for imidazole, 0.434 for water, 0.472 for the methylammonium cation, and 0.471 for the imadazolium cation.

ELECTROSTATIC POTENTIAL (ESP) CHARGES

This approach uses charges obtained in fits to the quantum-mechanically calculated molecular electrostatic potential performed using one of a variety of algorithms that differ in how the electrostatic potential is sampled and, in some cases, in how the least-squares criterion is defined.[31,52-55] Because they produce different sets of charges, even when the same quantum-mechanical wave function (often HF/6-31G*) is used, the reader should bear in mind that comparisons and conclusions based on one set of ESP charges may not apply when another set is used.

The ESP charges listed in Table V were obtained from HF/6-31G* wavefunctions using the CHELPG procedure[54] implemented in Gaussian 92.[56] They compare closely to the HF/6-31G*-based CHELPG charges tabulated by Carlson et al., but differ slightly because we have used HF/6-31G*-otpimized rather than standard geometries. To simplify the presentation, we have averaged the charges for hydrogens made equivalent through rotations about single bonds, as Carlson et al. did and as symmetry considerations require be done for use in molecular mechanics or dynamics calculations. Bayly et al. have shown, however, that any such *a posteriori* averaging merits careful attention, as it can change the molecular dipole moment and the strength of computed intermolecular interactions substantially.[53] For *trans-N*-methylacetamide and methanol, the cases they most closely studied, such averaging raised the calculated dipole moment by ~10%. As a further word of caution, we note, however, that we have found that the effect on the dipole moment is not uniform. Thus, while the dipole moment calculated for *trans-N*-methylacetamide also increases when either the CHELPG[54] and Besler–Merz–Kollman[55] charges for the methyl group are averaged, the dipole moment for *cis-N*-methylacetamide is unchanged or even decreases slightly. Although the effects of averaging amount to only a few tenths of a Debye in this case, they alter the relationship between the *ab initio* dipole moments of the cis and trans conformers and thus may affect the relative free energies of hydration.

For the most part, the CHELPG charges are similar to the MMFF94 and OPLS charges. One qualitative difference is that hydrogen atoms at-

tached to sp^3-hybridized carbon usually bear appreciable positive charge, often of roughly +0.1, for the CHELPG set. As we pointed out in justifying MMFF's assignment of zero charge to such hydrogens, positive charges of this magnitude leave MMFF94 unable to reproduce certain key conformational energies.[3]

Carlson et al. found that their CHELPG charges, when used in conjunction with OPLS Lennard–Jones parameters, gave reasonable values for absolute and relative free energies of hydration in most cases.[35] They noted, however, that the CHELPG charges for acetamide, when used in simulations with TIP4P water, gave a free energy of hydration that was about 4 kcal/mol too negative. They also found that the CHELPG charges gave substantially more negative interaction energies than did OPLS for dimers in which a molecule of water donates a hydrogen bond to the carbonyl oxygen either syn (−10.17 kcal/mol) or anti to the nitrogen (−8.75 kcal/mol); the OPLS interaction energies were −9.30 and −6.67 kcal/mol, respectively. Now, the fact that both MMFF and CHELPG assign a larger negative charge to carbonyl oxygen (−0.57 and −0.637, respectively) than does OPLS (−0.50) might seem a source for concern, though the MMFF94 charge set clearly is less strongly polarized. We have found, however, that MMFF94 gives interaction energies of −9.14 and −7.18 kcal/mol for the syn and anti dimers,[57] respectively, in rather close agreement with OPLS, suggesting that the MMFF94 charges may well yield good results for acetamide in molecular simulations. The difficulty encountered with the CHELPG charges, on the other hand, serves to underscore the fact that the ESP charges result from a fitting procedure that takes neither intermolecular-interaction energies nor conformational energies into account. Such charges need not give the former accurately nor be compatible with previously determined torsion parameters.

A further concern is that the ESP approach is not sufficiently general for our purposes, because a preliminary quantum-mechanical calculation is needed to obtain the requisite partial atomic charges for each system involving new chemistry; extensive libraries of ESP charges for diverse chemical fragments, coupled to robust algorithms for combining such fragments, do not yet exist. A related concern is that ESP charges are sometimes ill-determined for "interior" (or "buried") charge centers.[53a,58] Examples include the large and seemingly nonphysical variations in ESP charges with conformation observed by Williams[59] for the alanyl dipeptide (up to 0.76 electron units) and by Stouch and Williams[58] for glycerylphosphorylcholine (up to 1.32 electron units). This "flighty" aspect may may well inhibit transferability and limit the accuracy with which ESP charges for complex molecules, even in principle, could be assembled from fragment libraries.

CHARGE / ELECTRONEGATIVITY EQUILIBRATION

Several approaches have been described that make use of charge or electronegativity equilibration, among them the methods of Gasteiger,[60] Mortier and coworkers,[61] Mullay,[62] and Rappé and Goddard.[63] Each bases the charge calculation on assigned atomic (rather than atom-pair) properties, and hence should easily be able to meet the test of generality. Moreover, each produces charges that depend on the chemical topology in a more complete manner than do bond increment charges, another plus; some also yield charges that depend on the molecular conformation. Unfortunately, however, the charges given by these methods, as presently parameterized, differ significantly from those found to be required for MMFF94. For example, the Gasteiger partial equivalence of orbital electronegativity (PEOE) charges for polar hydrogens, listed in Table V,[64] are typically only half as large as those given by MMFF94, OPLS, and CHELPG, suggesting that they would give unsatisfactory small intermolecular-interaction energies.

The QEq (Rappé-Goddard) charges[63,65] for polar hydrogens also tend to be small by comparison to MMFF94 and OPLS. In contrast, the QEq charge on chlorine in CH_3Cl (−0.43) seems too large. A perhaps more serious problem is that aliphatic hydrogens typically carry far larger positive charges than would be acceptable for MMFF94.[3] Another is that hydrogen atoms that become equivalent through single-bond rotations can have quite different QEq charges. For the CH_3 group in methanol, for example, the gauche (to the hydroxyl group) and anti hydrogens carry charges of 0.138 and 0.189, respectively. As we previously noted, one either would need to average such nonequivalent charges (e.g., use the values listed in Table V), or to recompute them frequently during the simulation. But the latter approach would require readily obtainable analytical derivatives of the charges with respect to the nuclear positions, quantities that may not be straightforward to formulate.[66]

TABLE V.
Comparison of Atomic Charges for MMFF94 and Other Methods.

Molecule	Atom	MMFF type	MMFF	OPLS[a]	ESP[f]	QEq[g]	Gast.[i]	Abraham[j]
CH_3OH	C	1	0.280	0.265	0.321	−0.157	0.033	−0.036
	HC	5	0.000	—	−0.017	0.155	0.052	0.055
	O	6	−0.680	−0.700	−0.693	−0.663	−0.398	−0.458
	H	21	0.400	0.435	0.424	0.355	0.209	0.329
CH_3NH_2	C	1	0.270	0.200	0.482	−0.351	−0.019	−0.062
	HC	5	0.000	—	−0.064	0.161	0.039	0.047
	N	8	−0.990	−0.900	−1.033	−0.643	−0.333	−0.505
	H	23	0.360	0.350	0.372	0.255	0.118	0.212
CH_3CN	C	1	0.200	0.150	−0.202	−0.374	0.023	−0.045
	HC	5	0.000	—	0.101	0.134	0.038	0.049
	C	4	0.357	0.280	0.408	0.213	0.059	0.362
	N	42	−0.557	−0.430	−0.509	−0.242	−0.197	−0.464
CH_3OCH_3	C	1	0.280	0.250	0.140	−0.112	0.036	−0.033
	HC	5	0.000	—	0.024	0.128	0.052	0.055
	O	6	−0.560	−0.500	−0.427	−0.541	−0.386	−0.266
CH_3SH	C	1	0.230	0.180	0.025	−0.293	−0.021	−0.021
	HC	5	0.000	—	0.047	0.162	0.034	0.043
	S	15	−0.410	−0.450	−0.369	−0.400	−0.182	−0.237
	H	71	0.180	0.270	0.203	0.208	0.101	0.128
CH_3Cl	C	1	0.290	0.250	−0.094	−0.055	0.011	0.039
	HC	5	0.000	—	0.101	0.162	0.039	0.048
	Cl	12	−0.290	−0.250	−0.210	−0.430	−0.129	−0.184
C_2H_6	C	1	0.000	0.000	0.00	−0.464	−0.068	−0.110
	H	5	0.000	—	0.00	0.155	0.023	0.037
CH_3CONH_2	C	1	0.061	0.000	−0.399	−0.302	0.012	−0.037[k]
	HC	5	0.000	—	0.105	0.130	0.032	0.057
	C	3	0.569	0.500	0.970	0.470	0.207	0.252
	O	7	−0.570	−0.500	−0.662	−0.439	−0.276	−0.408
	N	10	−0.800	−0.850	−1.098	−0.616	−0.330	−0.378
	H	28	0.370	0.425	0.436	0.249	0.145	0.222
CH_3COOH	C	1	0.061	0.080	−0.343	−0.232	0.021	−0.044
	HC	5	0.000	—	0.106	0.134	0.033	0.049
	C	3	0.659	0.550	0.898	0.586	0.256	0.285
	O =	7	−0.570	−0.500	−0.637	−0.458	−0.266	−0.373
	O—	6	−0.650	−0.580	−0.692	−0.663	−0.332	−0.346
	H	24	0.500	0.450	0.455	0.366	0.220	0.330
$(CH_3)_2CO$	C	1	0.061	0.062	−0.367	−0.327	−0.006	−0.061
	HC	5	0.000	—	0.093	0.129	0.031	0.046
	C	3	0.448	0.300	0.764	0.335	0.127	0.190
	O	7	−0.570	−0.424	−0.589	−0.455	−0.298	−0.342
CH_3COOCH_3	C	1	0.061	0.050	−0.505	−0.265	0.021	—
	HC	5	0.000	—	0.135	0.119	0.033	—
	C	3	0.659	0.550	0.972	0.613	0.259	—
	O =	7	−0.570	−0.450	−0.640	−0.503	−0.265	—
	O—	6	−0.430	−0.400	−0.504	−0.571	−0.319	—
	C	1	0.280	0.250	0.120	−0.080	0.046	—
	HC	5	0.000	—	0.051	0.150	0.053	—
C_6H_6	C	37	−0.150	−0.115[b]	−0.100	−0.097	−0.062	−0.150[k]
	H	5	0.150	0.115	0.100	0.097	0.062	0.150
C_5H_5N	N	38	−0.620	−0.490	−0.683	−0.334	−0.263	−0.349[l]
	C1	37	0.160	0.230	0.482	0.070	0.028	0.044
	H1	5	0.150	—	0.008	0.110	0.083	0.088
	C2	37	−0.150	−0.030	−0.489	−0.105	−0.044	−0.065
	H2	5	0.150	—	0.154	0.090	0.063	0.081
	C3	37	−0.150	0.090	0.231	−0.103	−0.059	−0.024
	H3	5	0.150	—	0.065	0.106	0.062	0.079
$C_6H_5NH_2$	H	28	0.400	0.400[b]	0.369	0.228	0.124	0.215[m]
	N	40	−0.900	−0.900	−0.945	−0.654	−0.302	−0.372

(*Continues on next page*)

TABLE V.
(continued)

Molecule	Atom	MMFF type	MMFF	OPLS[a]	ESP[f]	QEq[g]	Gast.[i]	Abraham[j]
	C1	37	0.100	0.100	0.526	0.147	0.013	0.027
	C2	37	−0.150	−0.115	−0.349	−0.062	−0.044	−0.121
	H2	5	0.150	0.115	0.158	0.103	0.063	0.094
	C3	37	−0.150	−0.115	−0.024	−0.121	−0.060	−0.088
	H3	5	0.150	0.115	0.101	0.104	0.062	0.091
	C4	37	−0.150	−0.115	−0.215	−0.109	−0.062	−0.128
	H4	5	0.150	0.115	0.116	0.109	0.063	0.091
Imidazole	N1	39	0.033	−0.570[c]	−0.256	−0.518	−0.310	−0.137[k]
	H1	23	0.270	0.420	0.307	0.231	0.152	0.224
	C2	63	0.037	0.410	0.275	0.195	0.085	0.126
	H2	15	0.150	—	0.077	0.112	0.100	0.109
	N3	66	−0.565	−0.490	−0.565	−0.315	−0.247	−0.398
	C4	64	0.077	0.100	0.194	0.084	0.043	−0.079
	H4	15	0.150	—	0.072	0.096	0.084	0.102
	C5	63	−0.302	0.130	−0.294	0.011	0.013	—
	H5	15	0.150	—	0.189	0.105	0.080	—
H_2O	O	70	−0.860	−0.834[d]	−0.816	−0.702	−0.410	−0.62[n]
	H	31	0.430	0.417	0.408	0.351	0.205	0.31
$CH_3CO_2(-)$	C	1	−0.106	−0.100[e]	−0.262	−0.338	−0.025	0.043[k]
	HC	5	0.000	—	0.011	0.017	0.029	0.056
	C	41	0.906	0.700	0.969	0.582	0.039	0.009
	O	32	−0.900	−0.800	−0.872	−0.647	−0.550	−0.545
$CH_3NH_3(+)$	C	1	0.503	0.310[e]	−0.014	−0.119	−0.073	−0.039
	HC	5	0.000	—	0.113	0.265	0.006	0.059
	N	34	−0.853	−0.300	−0.288	−0.702	0.427	0.140
	H	36	0.450	0.330	0.321	0.342	0.209	0.253
Imidazolium(+)	N1	81	−0.700	−0.540[c]	−0.190	−0.443	0.250	−0.058[k]
	H1	36	0.450	0.460	0.369	0.314	0.310	0.306
	C2	80	0.650	0.500	0.142	0.299	0.242	0.114
	H2	5	0.150	—	0.211	0.218	0.158	0.158
	C5	78	0.200	0.330	−0.076	0.166	0.128	−0.037
	H5	5	0.150	—	0.221	0.205	0.112	0.153
$(-)O_2C(CH_2)_6NH_3(+)$	O	32	−0.900	−0.800[c]	−0.846	−0.475[h]	−0.550[h]	−0.545[o]
	C	41	0.900	0.700	0.916	0.671	0.042	0.009
	N	34	−0.853	−0.300	−0.552	−0.833	0.218	0.140
	H	36	0.450	0.330	0.363	0.247	0.199	0.253

[a] Unless otherwise noted, from W. L. Jorgensen, J. M. Briggs, and M. L. Contreras, *J. Phys. Chem.*, **94**, 1683–1686 (1990).
[b] From ref. 41.
[c] Based on charges for amino-acid sidechains given in ref. 45.
[d] TIP3P charges from ref. 43.
[e] From W.L. Jorgensen and J. Gao, *J. Phys. Chem.*, **90**, 2174–2182 (1986).
[f] Electrostatic potential fit charges obtained from HF/6-31G* wavefunctions for HF/6-31G*-optimized geometries using the CHELPG procedure (ref. 54) implemented in GAUSSIAN 92 (ref. 56).
[g] Obtained for MP2/6-31G*-optimized geometries using the QEq procedure (ref. 63) as implemented in Cerius Version 3.2 (Molecular Simulations, Inc., Burlington, MA).
[h] Obtained for the MMFF94-optimized geometry for the extended conformation.
[i] Gasteiger partial equilization of orbital electronegativity (PEOE) charges (ref. 60) obtained using the implementation in SYBYL Version 5.5 (Tripos Associates, Inc., St. Louis, MO).
[j] Unless otherwise indicated, CHARGE2 model charges from ref. 69a.
[k] Based on CHARGE2 charges for protein sidechains given in ref. 69a.
[l] "DM" method charges from R. J. Abraham and P. E. Smith, *J. Comput. Chem.*, **9**, 288–297 (1987).
[m] CHARGE2 charges from ref. 69b.
[n] Estimated charges based on the calculated dipole moment for water of 1.68 debyes cited in R. J. Abraham and B. Hudson, *J. Comput. Chem.*, **5**, 562–570 (1984).
[o] Inferred from the results listed for $CH_3CO_2(-)$ and $CH_3NH_3(+)$.

To be sure, the dependence of the QEq charges on conformation is potentially an important attribute. In a classic study, Jorgensen and Gao provided strong evidence that the OPLS amide-group charges needed to change substantially on going from *cis-* to *trans-N*-methylacetamide[67] (NMA) in order to account for the otherwise surprising experimental finding that the two conformers have essentially equal free energies of aqueous hydration; one might have expected the trans form to have the larger dipole moment in a fixed-charge model, and hence to be the more strongly solvated. Any such variation of atomic charge with conformation, however, must be both of the right magnitude and in the proper direction to be helpful. No assurance can be given that the classically motivated Mortier and Rappé–Goddard approaches will behave in this manner; for NMA, it appears that the QEq charges do not. In particular, they give dipole moments of 2.80 and 3.90 D, respectively, for *trans-* and *cis*-NMA when MP2/6-31G*-optimized geometries are used.[65] In contrast, MMFF94 gives dipole moments of 4.44 and 4.55 D for the same geometries, and of 4.31 and 4.46 D when MMFF-optimized geometries are used. For reference, the *ab initio* HF/6-31G* calculations give dipole moments of 3.99 and 4.18 D for HF-6-31G*-optimized geometries, and the adjusted Jorgensen–Gao OPLS charges for NMA give dipole moments of 3.85 and 4.03 D. Thus, QEq rightly gives a larger dipole moment for *cis*-NMA, but greatly overestimates the magnitude of the difference. In contrast, MMFF94's fixed-charge model not only reproduces the *ab initio* ordering but also gives reasonable absolute and relative values for the dipole moments; the unmodified OPLS charges, on the other hand, do not satisfy these criteria.

A critical limitation is evidenced by the comparison in Table V of the QEq charges for the acetate and methylammonium ions with those for the $(CH_2)_6$-bridged ammoniumcarboxylate zwitterion. In particular, the QEq group charges on the $CO_2(-)$ and $NH_3(+)$ moieites are -0.684 and $+0.326$, respectively, in the separate ions, but are -0.279 and -0.092 in the zwitterion. The $NH_3(+)$ charge in methylammonium and the $CO_2(-)$ charge in the zwitterion already seem too small, but the *negative* charge for the $NH_3(+)$ group in the latter is clearly inappropriate. Evidently, the head group charges have inappropriately leaked into the saturated hexamethylene chain, which should instead act as a buffer. This behavior reveals a key failing of the QEq method. Specifically, QEq simply treats the zwitterion as a *neutral* system; it does not understand that the zwitterion consists of essentially isolated ionic subsystems. The same difficulty of course would apply to proteins containing charged residues, and would lead to charges that change significantly when distant residues undergo a change in ionization state.

Last, we consider the Mortier[61] and Mullay[62] methods. The former, which uses an electronegativity equalization approach, gives charges that depend both on the connectivity and on the geometry. The latter equalizes "bond electronegativities" and gives charges that depend only on the connectivity. Both appear to have been parameterized to reproduce STO-3G charges, and both yield charges for the carbonyl oxygens in the alanyl dipeptide that are only half as large in magnitude as the MMFF94 charges. Interestingly, the Mullay approach, unlike QEq, handles zwitterions appropriately. Still, the published charges suggest that neither those nor any of the other methods considered in this subsection, as currently parameterized, would yield useful intermolecular-interaction energies in the context of a force field like MMFF.

OTHER APPROACHES

Finally, we briefly consider the methods of Del Re[68] and Abraham and coworkers.[69] These methods either employ a quasi–quantum-mechanical approach (Del Re) or, for extended pi-systems, include an explicitly quantum-mechanical component (Abraham). For comparison, we list a set of Abraham charges in Table V, most of which were taken from the recent "CHARGE2" parameterization.[69] The table shows that these charges correlate fairly well with MMFF94 and OPLS charges, but often are only about half as large. Clearly, such reduced charges would severely underestimate electrostatic interactions if used in empirical force-field calculations. For Del Re, Rappé and Goddard[63] give comparisons that show, for example, that carbonyl oxygens typically are assigned charges of about -0.10. Such charges would not be useful for our purposes.

DISCUSSION

In concluding this section, we note that a distinction needs to be made between: (1) the charges determined in the parameterization of a force field on *prototype systems*; and (2) the charges produced by the charge calculation/assembly procedure in *general applications*. For some methods, the distinction is an important one. Thus, while it might not

be practical to directly parameterize a given method to reproduce *interaction energies*, it might be feasible to first determine a set of desired *target charges* in fits to data on interaction energies, and then to parameterize the method to reproduce the target charges. These target charges could be those derived in the parameterization of MMFF94 or of OPLS or in some other suitable manner. To the extent, however, that the charges needed to reproduce interaction energies are determined by quantum-mechanical considerations, a classically motivated scheme based only on *atomic properties* may have insufficient flexibility.[70] In contrast, a scheme that, like MMFF's, is based on *bond properties*, while requiring more parameters, possesses an inherently greater degree of flexibility and may be capable of greater accuracy.

Performance of MMFF94 and Other Force Fields

DIPOLE MOMENTS

The MMFF94 partial atomic charges for the 423 molecular species/conformations used in its parameterization reproduced the scaled HF/6-31G* molecular dipole moments (in Debye units) with rms deviations[71] as shown:

	MMFF94	MM2X
Dipole magnitude	0.39 D	0.64 D
Dipole direction	5.5°	10.8°

In these comparisons, the HF/6-31G* geometries were used and the *ab initio* dipole moments were scaled by the factor of 1.10 used in the fitting process (discussed earlier). Listed for comparison are the results obtained for 398 MM2X conformers, also computed using the scaled HF/6-31G* dipole moments.[72] As the rms value of the scaled HF/6-31G* dipole moments is 3.42 D, the average MMFF94 error is about 10%, while that for MM2X is about 20%. Nevertheless, while respectable, the errors in dipole magnitude for MMFF94 are larger than we would like to see. The treatment of electrostatic interactions is one particular area in which improvement needs to be made.

INTERMOLECULAR INTERACTION ENERGIES AND GEOMETRIES

Table VI summarizes MMFF94's ability to reproduce trends in the HF/6-31G* data used in MMFF94's parameterization (Model 1). The leftmost column identifies the dimers examined. Except for cyclic dimers, the donor species is listed first, the acceptor species second. The second through fourth columns compare the HF/6-31G* and MMFF94 binding energies (in kilocalories per mole); the fifth through seventh compare the heteroatom distances $X\cdots Z$ (in angstroms) in the listed hydrogen bonds; and the remaining columns compare the $H\cdots Z$ distances, compare the X—$H\cdots Z$ angles (in degrees), and specify the pertinent MMFF atom types. Together with the careful representational structures shown in Figure 1, the listed information should suffice to characterize the HF/6-31G* geometries. Except for the cyclic ammonia dimer (**32**, C_{2h} symmetry) and the stacked *trans-N*-methylacetamide dimer (**27**), each dimer and its associated monomers have been fully optimized both *ab initio* and in the MMFF94 calculations. For these two systems, MMFF94 gives a saddle-point structure that has one imaginary frequency. For the former, the cyclic C_{2h} dimer is the minimum at the HF/6-31G* level and the "linear" C_s dimer (**33**) is a saddle point.[73] At a number of higher levels,[33] however, including the rather similar HF/6-31 + G* level,[73] the linear dimer becomes the stable structure, just as we find for MMFF94. Whether the stacked *trans-N*-methylacetamide dimer is a minimum on the HF/6-31G* surface is unclear, but we should note that the *ab initio* energy minimization was not carried to completion and that the two $X\cdots Z$ distances already differ slightly.

For comparison to Model 1, we list dimerization energies in Table VII for two other theoretical models (Models 2 and 3), each of which employs the HF/6-31G*-optimized geometries used in Table VI. Model 2 utilizes single-point MP2/6 − 31 + G** wave functions, while Model 3 uses HF/6-31G* wave functions, but incorporates a standard correction for basis-set superposition error.[74] This correction gives ΔE_{A-B} as

$$\Delta E_{A-B} = E_{A-B} - E_A - E_B \qquad (9)$$

where the calculations on the full system A–B and on the molecular subsystems A and B all use the same A–B dimer basis set. In Table VIII, we also report a limited number of dimer energies and geometries obtained from geometry-optimized MP2/DZP calcualtions[75] (Model 4). For each of the four models, it is possible to generate "target" dimerization energies by uniformly scaling the interaction energies so that the energy for the linear

TABLE VI.
Comparison of MMFF94 and HF / 6-31G* Hydrogen-Bond Energies and Geometries (cf. Fig. 1).

Dimer (structure number)	Energy (kcal / mol)			*X*···*Z* distance (Å)			H···*Z* distance		*X*—H···*Z* angle		
	HF	MMFF	Δ	HF	MMFF	Δ	HF	MMFF	HF	MMFF	Atom types
HOH ··· OH_2 (1)	−5.624	−6.609	−0.984	2.971	2.750	−0.221	2.025	1.776	172.3	170.4	70-31 ··· 70
HOH ··· HOH, Cyclic C_s (2)	−2.113	−4.909	−2.796	2.776	2.777	0.001	2.609	2.193	90.0	117.0	70-31 ··· 70
				2.776	2.777	0.001	2.609	2.193	90.0	117.0	70-31 ··· 70
HOH ··· $OHCH_3$ (3)	−5.544	−6.321	−0.777	2.951	2.732	−0.219	2.011	1.763	169.0	168.7	70-31 ··· 6
CH_3OH ··· OH_2 (4)	−5.590	−6.145	−0.555	2.961	2.749	−0.212	2.013	1.764	174.9	179.9	6-21 ··· 70
CH_3OH ··· $OHCH_3$ (5)	−5.538	−6.033	−0.495	2.945	2.726	−0.219	1.998	1.744	174.6	176.9	6-21 ··· 6
HOH ··· $O(CH_3)_2$ (6)	−5.311	−5.978	−0.666	2.946	2.770	−0.175	2.020	1.808	163.9	166.7	70-31 ··· 6
C_6H_5OH ··· OH_2 (7)	−7.366	−8.448	−1.082	2.901	2.686	−0.215	1.952	1.716	174.0	167.5	6-29 ··· 70
HOH ··· OHC_6H_5 (8)	−4.711	−5.085	−0.374	3.031	2.826	−0.205	2.118	1.888	160.8	159.9	70-31 ··· 6
t-vinyl alcohol ··· OH_2 (9)	−7.434	−8.481	−1.046	2.883	2.676	−0.207	1.934	1.689	175.0	176.1	6-29 ··· 70
HOH ··· *t*-vinyl alcohol (10)	−4.691	−5.464	−0.772	3.031	2.809	−0.222	2.109	1.868	163.2	160.4	70-31 ··· 6
HOH ··· furan (11)	−3.640	−3.775	−0.135	3.026	2.881	−0.145	2.225	1.966	141.6	155.4	70-31 ··· 59
H_3CCOOH ··· $HOOCCH_3$, cyclic (12)	−15.549	−17.100	−1.550	2.793	2.617	−0.177	1.829	1.634	175.9	169.2	6-24 ··· 7
				2.793	2.616	−0.177	1.829	1.634	175.9	169.2	6-24 ··· 7
H_3CCOOH ··· OHH (13)	−9.943	−11.704	−1.761	2.802	2.644	−0.158	1.895	1.674	156.3	163.4	6-24 ··· 70
				2.834	2.766	−0.068	2.149	2.108	127.9	122.8	70-31 ··· 7
trans,*trans*-oxalic acid ··· OHH, cyclic (14)	−10.209	−12.716	−2.507	2.760	2.606	−0.154	1.799	1.610	171.5	175.4	6-24 ··· 70
				2.839	2.597	−0.242	2.217	1.916	122.2	123.9	70-31 ··· 7
OHH ··· *cis*,*cis*-oxalic acid, cyclic (15)	−4.455	−6.106	−1.651	3.231	2.935	−0.296	2.437	2.021	141.1	155.7	70-31 ··· 6
				3.213	2.785	−0.428	2.345	2.688	151.7	85.4	70-31 ··· 7
OHH ··· *trans*,*trans*-oxalic acid, cyclic (16)	−3.470	−4.669	−1.199	3.154	2.860	−0.293	2.268	2.024	155.0	142.5	70-31 ··· 7
				3.492	3.198	−0.294	2.952	2.432	117.5	135.5	70-31 ··· 6
HOH ··· *skew*-methyl vinyl ether (17)	−4.731	−4.798	−0.066	2.969	2.871	−0.097	2.077	1.938	155.6	159.0	70-31 ··· 6
HOH ··· methyl formate (=O) (18)	−5.808	−6.843	−1.035	2.995	2.728	−0.267	2.061	1.748	167.0	177.6	70-31 ··· 7
HOH ··· methyl formate (—O—) (19)	−3.473	−2.629	0.845	3.018	2.951	−0.068	2.240	2.123	138.7	142.0	70-31 ··· 6
HOH ··· formaldehyde (20)	−5.284	−6.749	−1.464	2.954	2.715	−0.240	2.107	1.742	147.6	170.5	70-31 ··· 7
HOH ··· acetone (21)	−6.249	−7.063	−0.813	2.979	2.711	−0.268	2.055	1.734	163.1	173.1	70-31 ··· 7
t-NMA ··· OH_2 (22)	−5.427	−6.864	−1.436	3.116	2.839	−0.276	2.120	1.816	178.3	176.9	10-28 ··· 70
HOH ··· *t*-NMA (23)	−7.296	−8.192	−0.896	2.938	2.706	−0.232	2.000	1.727	167.6	174.2	70-31 ··· 7
Formamide dimer, cyclic (24)	−13.440	−12.264	1.176	2.997	2.813	−0.184	1.999	1.794	171.6	173.4	10-28 ··· 7
				2.996	2.813	−0.183	1.998	1.794	171.5	173.4	10-28 ··· 7
Formamide dimer, parallel (25)	−7.382	−7.692	−0.310	3.039	2.762	−0.277	2.094	1.743	156.7	174.4	10-28 ··· 7
t-NMA dimer, antiparallel (26)	−6.999	−9.082	−2.083	3.075	2.759	−0.317	2.085	1.756	171.8	165.2	10-28 ··· 7
t-NMA antiparallel stacked (27)	−3.512	−5.621	−2.109	3.969	3.608	−0.361	3.880	3.490	87.9	88.4	10-28 ··· 7
				3.987	3.608	−0.380	3.880	3.490	88.9	88.5	10-28 ··· 7
HOH ··· *N*-methylformamide (28)	−7.181	−8.061	−0.881	2.944	2.712	−0.232	2.014	1.732	164.7	174.8	70-31 ··· 7
t-N-OH,*N*-methylacetamide ··· OH_2 (30)	−7.464	−7.994	−0.529	2.852	2.703	−0.149	1.940	1.716	159.2	175.8	6-21 ··· 70
HOH ··· *t*-N-OH, *N*-methylacetamide (31)	−3.724	−3.950	−0.225	3.061	2.847	−0.214	2.124	1.873	168.7	175.8	70-31 ··· 6
H_2NH ··· HNH_2, cyclic C_{2h} (32)	−3.191	−3.023	0.167	3.271	3.106	−0.165	2.604	2.397	123.9	125.4	8-23 ··· 8
				3.271	3.106	−0.165	2.605	2.397	123.8	125.4	8-23 ··· 8
H_2NH ··· NH_3, linear C_s (33)	−3.073	−3.667	−0.594	3.428	3.037	−0.391	2.423	2.004	179.0	179.9	8-23 ··· 8
				3.428	3.037	−0.391	3.643	3.206	69.8	71.3	8-23 ··· 8
HOH ··· NH_3 (34)	−6.559	−6.828	−0.269	3.039	2.798	−0.242	2.085	1.817	176.4	173.9	70-31 ··· 8
HCH ··· NH_2CH_3 (35)	−6.429	−6.700	−0.271	3.042	2.819	−0.223	2.088	1.841	175.9	173.1	70-31 ··· 8
Imidazole ··· OH_2 (36)	−6.365	−7.918	−1.553	3.029	2.824	−0.205	2.030	1.808	178.2	175.6	39-23 ··· 70

(*Continues on next page*)

TABLE VI.
(continued)

Dimer (structure number)	Energy (kcal / mol) HF	MMFF	Δ	X···Z distance (Å) HF	MMFF	Δ	H···Z distance HF	MMFF	X—H···Z angle HF	MMFF	Atom types
HOH ··· imidazole (**37**)	−7.063	−7.935	−0.871	2.970	2.745	−0.225	2.098	1.807	151.2	158.6	70-31 ··· 66
Indole ··· OH_2 (**38**)	−5.754	−6.507	−0.753	3.050	2.856	−0.194	2.067	1.838	168.3	179.7	39-23 ··· 70
Pyrrole ··· OH_2 (**39**)	−5.358	−6.276	−0.918	3.064	2.860	−0.203	2.066	1.841	180.0	180.0	39-23 ··· 70
HOH ··· pyridine (**40**)	−6.030	−7.103	−1.072	3.013	2.790	0.223	2.118	1.852	155.8	158.9	70-31 ··· 38
Formamidine ··· H_2O, cyclic (**41**)	−10.029	−10.645	−0.615	3.057	2.967	−0.090	2.196	2.087	143.3	142.8	40-28 ··· 70
				2.932	2.755	−0.177	2.055	1.798	151.4	162.0	70-31 ··· 9
HOH ··· formaldehydeimine (**42**)	−6.131	−7.952	−1.821	2.996	2.737	−0.260	2.111	1.767	154.0	167.6	70-31 ··· 9
Guanidine ··· OHH (**43**)	−7.233	−7.476	−0.243	3.130	2.903	−0.227	2.254	1.982	145.4	148.1	40-28 ··· 70
				3.007	2.942	−0.064	2.205	2.114	141.1	141.1	70-31 ··· 40
Vinylamine ··· OH_2 (**44**)	−3.796	−4.609	−0.813	3.166	2.872	−0.294	2.193	1.845	164.2	173.3	40-28 ··· 70
Aniline ··· OH_2 (**45**)	−4.204	−4.859	−0.655	3.194	2.874	−0.320	2.199	1.850	173.2	172.8	40-28 ··· 70
HSH ··· OH_2 (**46**)	−2.664	−2.912	−0.248	3.619	3.377	−0.242	2.295	2.035	175.3	172.8	15-71 ··· 70
HOH ··· $S(CH_3)_2$ (**47**)	−3.236	−3.480	−0.244	3.390	3.351	−0.039	3.129	2.385	97.7	170.8	70-31 ··· 15
				3.390	3.351	−0.039	3.130	3.620	97.7	66.4	70-31 ··· 15
OHH ··· CH_3SSCH_3, cyclic (**48**)	−3.439	−3.221	0.217	3.673	3.549	−0.124	2.840	2.579	147.0	174.6	70-31 ··· 15
HOH ··· Thiophene (**49**)	−2.420	−2.713	−0.292	3.904	3.690	−0.214	3.500	3.026	108.4	126.8	70-31 ··· 44
HOH ··· SCH_6H_5 (**50**)	−1.998	−3.609	−1.611	4.959	4.369	−0.590	4.271	3.546	132.4	144.2	70-31 ··· 15
HSH ··· SH_2 (**51**)	−0.877	−1.251	−0.373	4.509	4.019	−0.490	3.186	2.698	175.5	167.3	15-71 ··· 15
HOH ··· SH_2 (**52**)	−2.061	−1.998	0.064	3.775	3.491	−0.284	2.828	2.548	175.5	163.2	70-31 ··· 15
CH_3COOH ··· NH_3, bidentate (**53**)	−10.850	−11.579	−0.729	2.857	2.655	−0.201	1.900	1.659	168.3	174.3	6-24 ··· 8
				3.145	3.016	−0.129	2.534	2.428	119.0	115.2	8-28 ··· 7
HOH ··· pyridine *N*-oxide (**54**)	−9.719	−10.609	−0.890	2.859	2.637	−0.222	3.238	2.978	58.5	60.3	70-31 ··· 7
Methylethylamine *N*-oxide ··· OHH (**55**)	−14.076	−14.870	−0.795	2.705	2.593	−0.112	1.832	1.657	148.9	154.6	70-31 ··· 7
				2.867	2.745	−0.123	2.174	2.097	124.2	117.6	68-23 ··· 70
Methylethylhydroxylamine ··· OH_2 (**56**)	−6.738	−7.357	−0.619	2.889	2.792	−0.096	2.107	1.941	138.5	143.2	6-21 ··· 70
				2.891	2.879	−0.012	2.183	2.099	130.2	135.1	70-31 ··· 8
HOH ··· FCH_3 (**57**)	−4.425	−4.472	−0.047	2.939	2.659	−0.280	2.057	1.687	153.9	172.7	70-31 ··· 11
HOH ··· chloropropane (**58**)	−2.805	−2.838	−0.033	3.627	3.347	−0.280	2.821	2.446	143.5	153.9	70-31 ··· 12
H_2NH ··· $O(CH_3)_2$ (**59**)	−2.837	−3.345	−0.508	3.262	2.994	−0.268	2.320	1.972	155.9	171.6	8-23 ··· 6
Methylammonium ··· OH_2 (**60**)	−19.298	−21.173	−1.874	2.813	2.619	−0.194	1.791	1.572	173.8	176.7	34-36 ··· 70
Guanidinium ··· OHH (**61**)	−18.480	−19.024	−0.544	2.972	2.779	−0.193	2.076	1.866	147.7	146.8	56-36 ··· 70
				2.972	2.779	−0.193	2.076	1.866	147.7	146.8	56-36 ··· 70
Imidazolium ··· OH_2 (**62**)	−16.950	−17.599	−0.649	2.819	2.622	−0.197	1.807	1.591	174.5	173.4	58-36 ··· 70
Formamidinium ··· OH_2 (**63**)	−16.976	−18.747	−1.771	2.818	2.605	−0.213	1.908	1.572	148.7	177.2	55-36 ··· 70
Formamidinium ··· OH_2, cyclic C_{2v} (**64**)	−19.589	−19.702	−0.113	2.951	2.783	−0.168	2.089	1.893	142.6	143.5	55-36 ··· 70
				2.950	2.782	−0.168	2.088	1.891	142.6	143.6	55-36 ··· 70
Formaldehydeiminium ··· OH_2 (**65**)	−19.984	−21.648	−1.664	2.770	2.646	−0.125	1.775	1.610	163.4	176.6	54-36 ··· 70
OHH ··· $(-)O_2CCH_3$, bidentate (**66**)	−21.848	−22.622	−0.774	2.883	2.743	−0.141	2.063	1.867	142.9	145.7	70-31 ··· 32
				2.892	2.743	−0.149	2.080	1.868	141.8	145.7	70-31 ··· 32

water dimer (**1**) is −6.5 kcal/mol,[36] and by uniformly adjusting the *ab initio* X ··· Z distances so that the O ··· O distance for the water dimer is reduced to 2.75 Å. Explicitly, the interaction energies would have to be enhanced by ~ 15.5%, 5.6%, 38%, and 3.1%, respectively, and the heteroatom distances would have to be reduced by ~ 0.2 Å for the first three methods (which all use HF/6-31G*-optimized geometries) and by ~ 0.15 Å for the fourth. As can be seen, the BSSE-

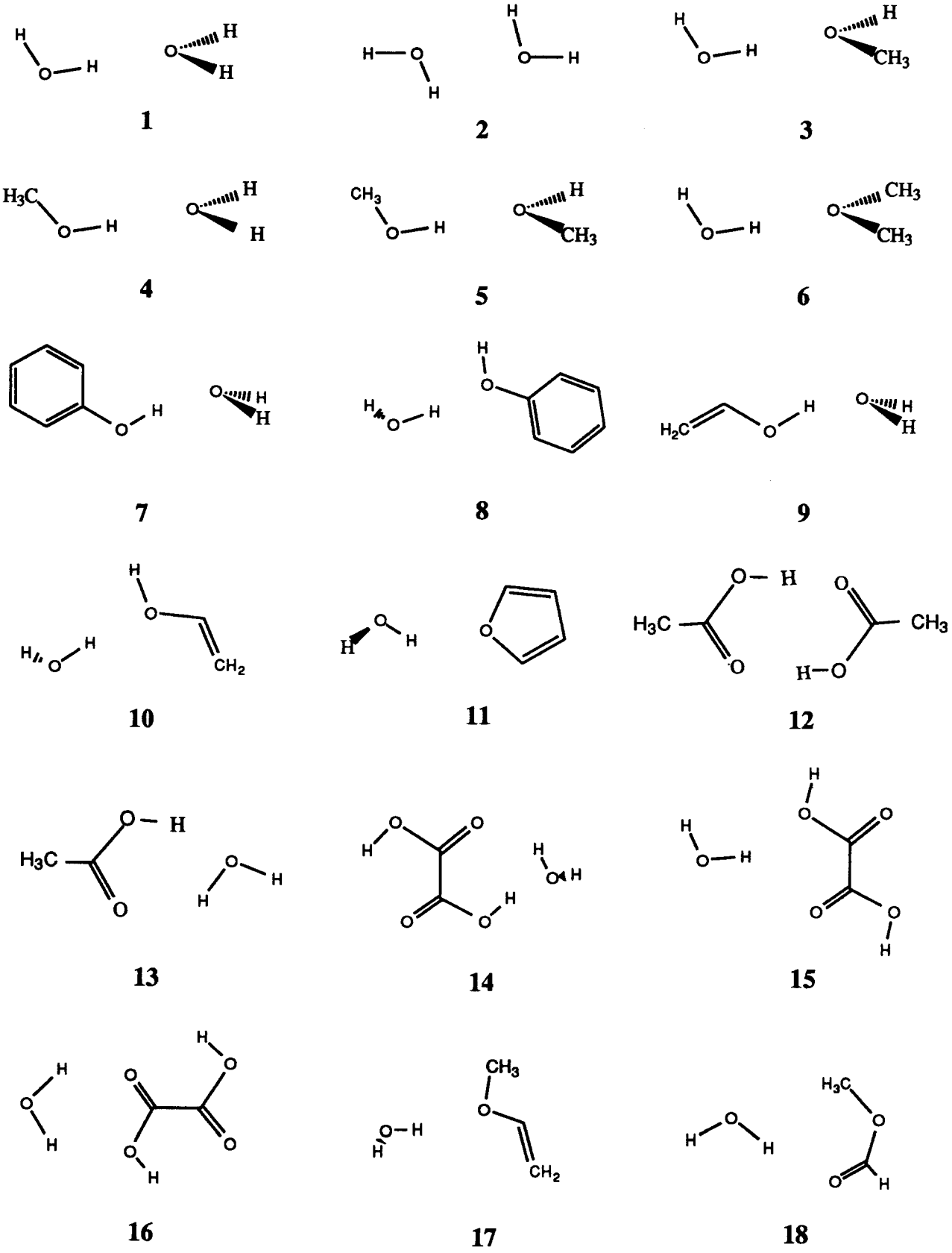
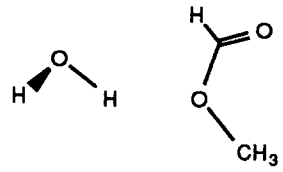
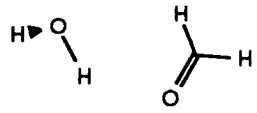


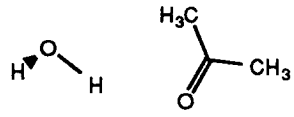
FIGURE 1. Representative structures for the HF / 6-31G* dimers used in parameterizing MMFF94. Where appropriate, the hydrogen-bond donor species is shown to the left.



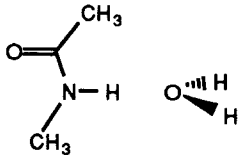
19



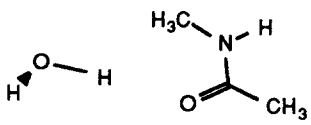
20



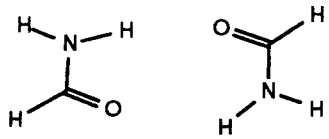
21



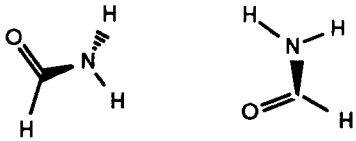
22



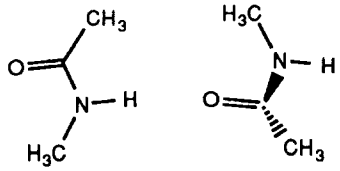
23



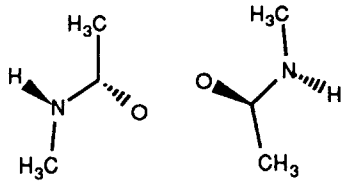
24



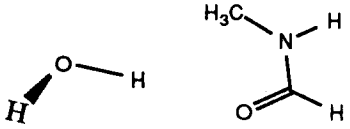
25



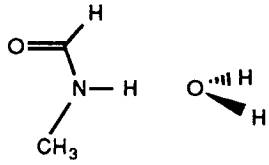
26



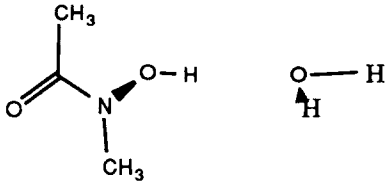
27



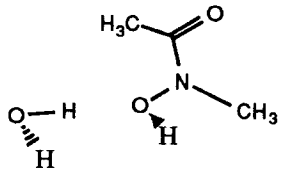
28



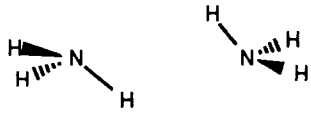
29



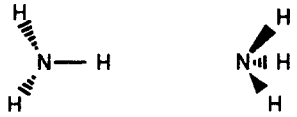
30



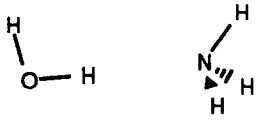
31



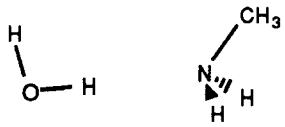
32



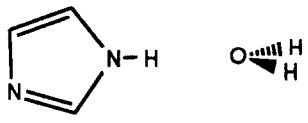
33



34



35



36

FIGURE 1. (continued)

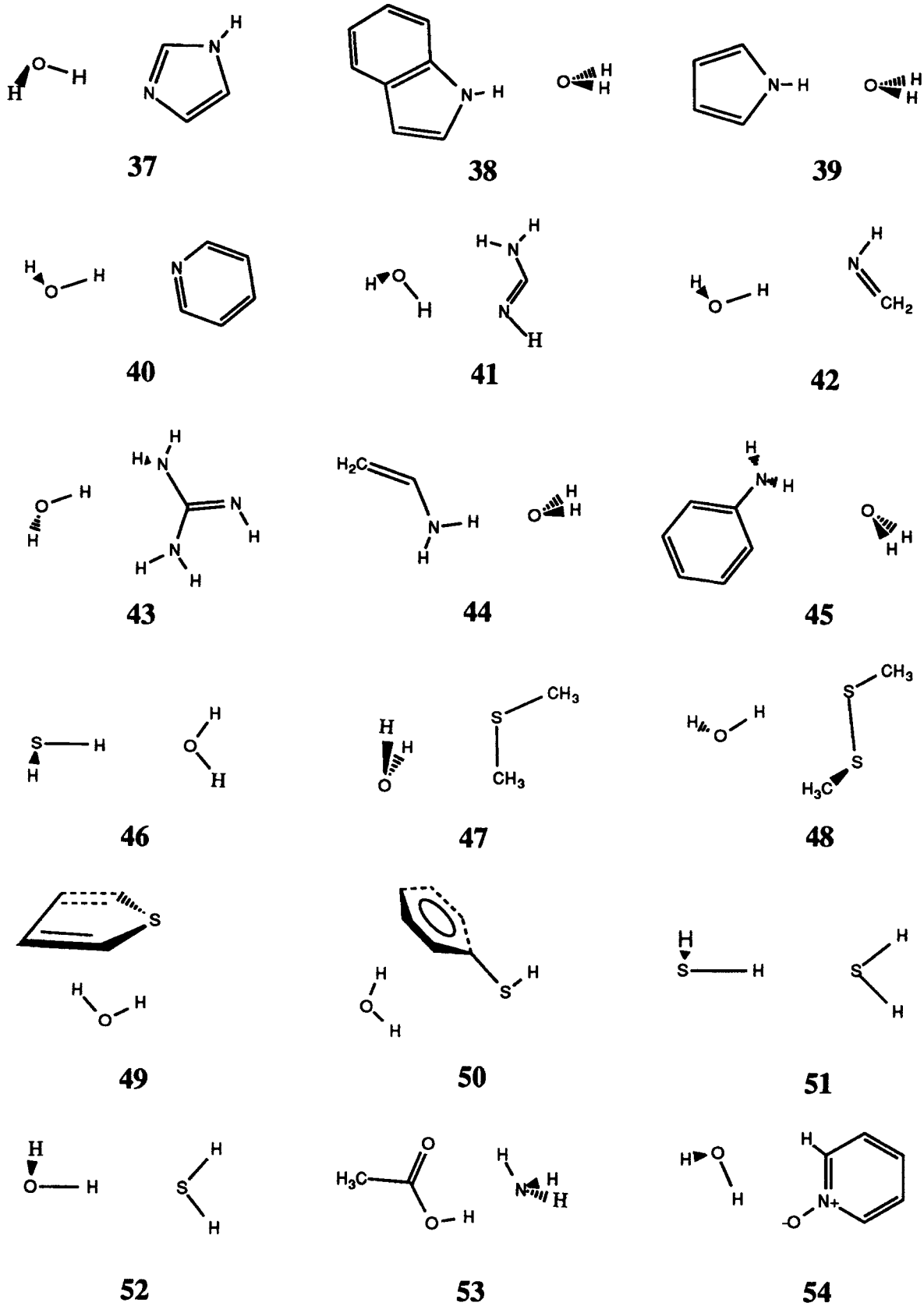


FIGURE 1. (continued)

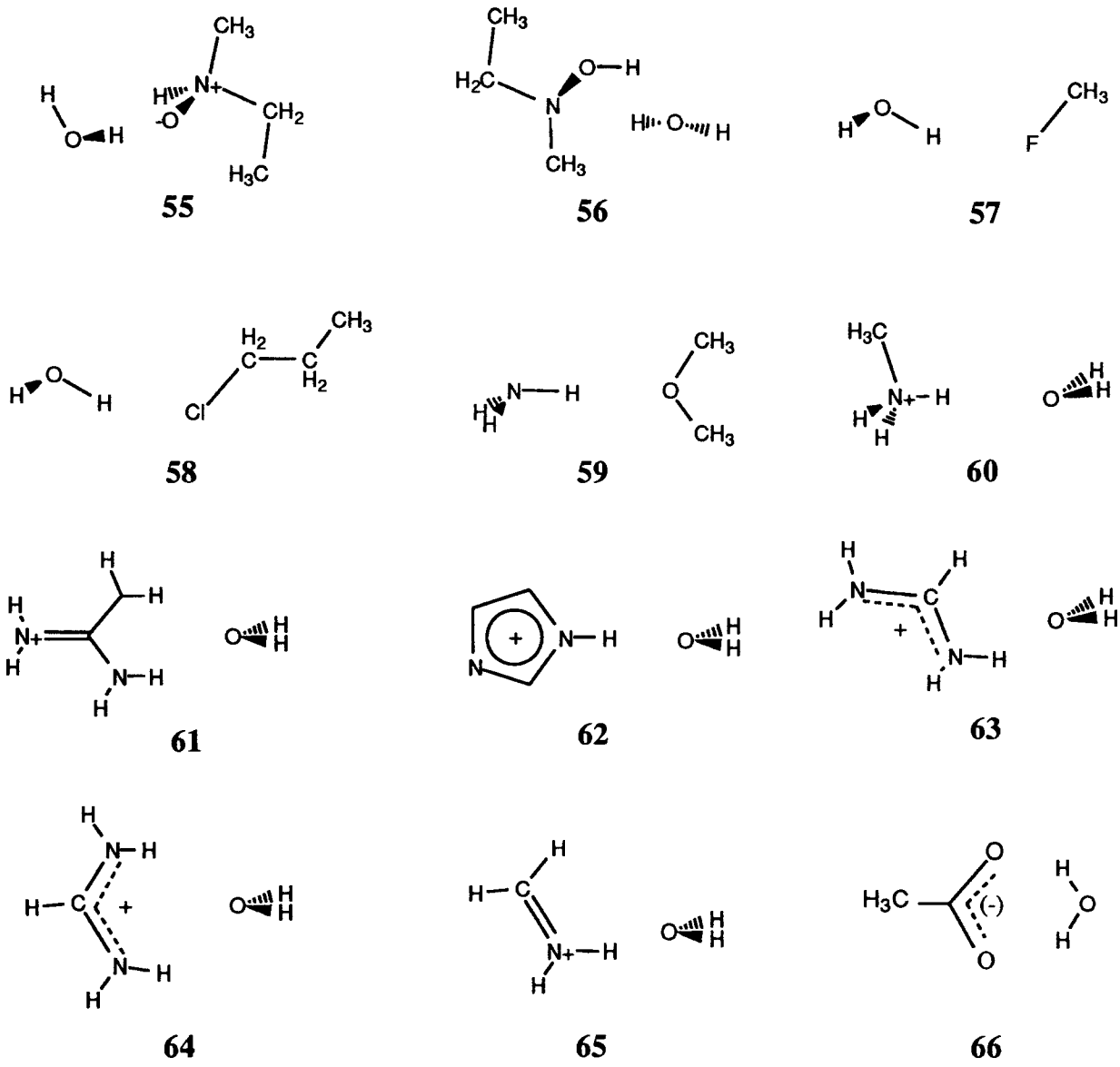


FIGURE 1. (continued)

corrected HF/6-31G* set gives the smallest dimerization energies and requires the greatest enhancements. Conversely, the MP2 models require the smallest enhancements. These MP2 models add both real correlation stabilization to the dimers and a basis-set superposition error which, at least in the case of the water dimer, is larger than that for the uncorrelated calculation when the 6-31G* basis set is used.[33] Taken together, the four models afford some measure of how well the dimerization energies we wish to reproduce may be said to be known, and provide a background against which to assess any limitations in MMFF's ability to reproduce the trends in the HF/6-31G* results shown in Table VI.

The four sets of generated target interaction energies (not shown) agree fairly well (e.g., to ± 0.5 kcal/mol) in most cases, with some notable exceptions. In particular, the BSSE-corrected set gives uniformly more negative scaled interaction energies for the strongly bound dimers, typically by 3–4 kcal/mol for the positive ion···water dimers (**60**–**65**). This set also gives the smallest scaled interaction energy for the cyclic water dimer (**2**) and gives smaller interaction energies for most dimers involving divalent sulfur. Conversely, the MP2 models give slightly less negative scaled interaction energies for the strongly bound dimers, but usually give significantly more negative values for the cyclic water dimer and the divalent

TABLE VII.
MP2 / 6-31 + G and Basis-Set-Superposition Error Corrected HF / 6-31G* Dimerization Energies (kcal / mol) for HF / 6-31G*-Optimized Geometries.**

Dimer (structure number)[a]	MP2 6-31 + G**	HF / 6-31G* BSSE[b]
HOH ··· OH_2 (**1**)	−6.157	−4.716
HOH ··· HOH, cyclic *Cs* (**2**)	−2.883	−1.331
HOH ··· $OHCH_3$ (**3**)	−6.739	−4.500
CH_3OH ··· OH_2 (**4**)	−6.202	−4.675
CH_3OH ··· $OHCH_3$ (**5**)	−6.886	−4.516
HOH ··· $O(CH_3)_2$ (**6**)	−6.724	−4.105
C_6H_5OH ··· OH_2 (**7**)	−8.604	−6.294
HOH ··· OHC_6H_5 (**8**)	−5.566	−3.166
t-vinyl alcohol ··· OH_2 (**9**)	−8.224	−6.568
HOH ··· *t*-vinyl alcohol (**10**)	−5.134	−3.362
HOH ··· furan (**11**)	−4.096	−2.283
H_3CCOOH ··· $HOOCCH_3$, cyclic (**12**)	−16.109	−14.661
H_3CCOOH ··· OHH (**13**)	−10.428	−8.697
trans,trans-oxalic acid ··· OHH, cyclic (**14**)	−11.266	−9.922
OHH ··· *cis,cis*-oxalic acid, cyclic (**15**)	−5.316	−3.710
OHH ··· *trans,trans*-oxalic acid, cyclic (**16**)	−3.990	−2.514
HOH ··· *skew*-methyl vinyl ether (**17**)	−5.788	−3.478
HOH ··· methyl formate (=O) (**18**)	−6.532	−4.761
HOH ··· methyl formate (—O—) (**18**)	−4.139	−2.209
HOH ··· formaldehyde (**20**)	−5.469	−3.767
HOH ··· acetone (**21**)	−6.574	−4.640
t-NMA ··· OH_2 (**22**)	−6.661	−4.542
HOH ··· *t*-NMA (**23**)	−7.969	−5.857
Formamide, cyclic (**24**)	−14.356	−11.918
Formamide dimer, parallel (**25**)	−7.856	−6.032
t-NMA dimer, antiparallel (**26**)	−8.954	−5.673
t-NMA dimer, antiparallel stacked (**27**)	−6.726	−2.405
HOH ··· *N*-methylformamide (**28**)	−7.654	−5.579
N-methylformamide ··· OH_2 (**29**)	−6.798	−4.853
t-N-OH,*N*-methylacetamide ··· OH_2 (**30**)	−8.909	−6.326
HOH ··· *t-N*-OH,*N*-methylacetamide (**31**)	−4.791	−2.536
H_2NH ··· HNH_2, cyclic C_{2h} (**32**)	−3.654	−2.213
H_2NH ··· NH_3, linear C_s (**33**)	−3.685	−2.512
HOH ··· NH_3 (**34**)	−7.654	−5.882
HOH ··· NH_2CH_3 (**35**)	−7.743	−5.673
Imidazole ··· OH_2 (**36**)	−7.861	−5.831
HOH ··· imidazole (**37**)	−7.843	−5.532
Indole ··· OH_2 (**38**)	−7.342	−4.890
Pyrrole ··· OH_2 (**39**)	−6.992	−4.819
HOH ··· pyridine (**40**)	−7.246	−4.604
Formamidine ··· H_2O, cyclic (**41**)	−10.773	−8.262
HOH ··· formaldehydeimine (**42**)	−6.956	−4.877
Guanidine ··· OHH (**43**)	−8.295	−5.777
Vinylamine ··· OH_2 (**44**)	−4.757	−3.148
Aniline ··· OHH (**45**)	−5.147	−3.112
SHH ··· OHH (**46**)	−3.661	−2.078
HOH ··· $S(CH_3)_2$ (**47**)	−3.047	−1.641
OHH ··· CH_3SSCH_3, cyclic (**48**)	−4.104	−2.089
HOH ··· thiophene (**49**)	−3.542	−1.584
HOH ··· SCH_6H_5 (**50**)	−3.678	−1.371
HSH ··· SH_2 (**51**)	−1.568	−0.772
HOH ··· SH_2 (**52**)	−2.878	−1.783
CH_3COOH ··· NH_3, bidentate (**53**)	−12.113	−9.684

(*Continues on next page*)

TABLE VII.
(continued)

Dimer (structure number)[a]	MP2 6-31 + G**	HF / 6-31G* BSSE[b]
HOH ··· pyridine *N*-oxide (**54**)	−9.044	−7.754
Methylethylamine N-oxide ··· OHH (**55**)	−15.457	−11.098
Methylethylhydroxylamine ··· OH_2 (**56**)	−9.037	−5.304
HOH ··· FCH_3 (**57**)	−4.300	−2.599
HOH ··· chloropropane (**58**)	−3.279	−1.741
H_2NH ··· $O(CH_3)_2$ (**59**)	−4.096	−1.938
Methylammonium ··· OH_2 (**60**)	−19.798	−18.764
Guanidinium ··· OHH (**61**)	−19.222	−17.689
Imidazolium ··· OH_2 (**62**)	−18.192	−16.429
Formamidinium ··· OH_2 (**63**)	−17.360	−16.404
Formamidinium ··· OH_2, cyclic C_{2v} (**64**)	−20.248	−18.849
Formaldehydeiminium ··· OH_2 (**65**)	−20.895	−19.492
OHH ··· $(-)O_2CCH_3$, bidentate (**66**)	−20.738	−19.220

[a] See Figure 1.
[b] Corrected for basis-set-superposition error as shown in eq. (9).

sulfur dimers. Notably, the MP2/DZP method gives a scaled value of −6.09 kcal/mol for the HOH ··· $S(CH_3)_2$ dimer (**47**), a value that seems far too large in magnitude. The correlation contribution to the basis-set-superposition error may be unusually large when second-row atoms are involved. The MP2/6-31 + G** model also gives a scaled dimerization energy for the stacked *t*-NMA dimer (**27**) of −7.10 kcal/mol, nearly twice the scaled HF/6-31G* value. This variation may partly arise from basis-set-superposition error, though one would expect the stacked configuration to be stabilized to some degree by London dispersion interactions omitted from the HF model. Taken together, these comparisons indicate serious difficulties with the MP2-based models (Models 2 and 4). While Model 3 (BSSE-corrected HF/6-31G* calculations) behaves reasonably in most respects, it does not appear superior to Model 1 (simple HF/6-31G* calculations). Accordingly, we have used Model 1, the simplest of the four, in parameterizing MMFF94.

Let us now turn to the comparison of the MMFF94 and HF/6-31G* results listed in Table VI. Clearly, the MMFF94 results correlate well with those from the *ab initio* calculations. On a

TABLE VIII.
Comparison of MP2 / DZP and HF / 6-31G* Energies and Geometries for Hydrogen-Bonded Dimers.

Dimer (structure number)[a]	Energy[b]		*X* ··· *Z* distance[c]		*X*—H ··· *Z* angle[d]	
	MP2	HF	MP2	HF	MP2	HF
HOH ··· OH_2 (**1**)	−6.303	−5.624	2.909	2.971	174.3	172.3
HOH ··· HOH, cyclic C_s (**2**)	−4.029	−2.113	2.951	2.777	99.4	90.0
H_2NH ··· HNH_2, cyclic C_{2h} (**32**)	−3.746	−3.191	3.159	3.271	115.6	123.9
			3.159	3.271	129.2	123.8
H_2NH ··· NH_3, linear C_s (**33**)	−3.925	−3.073	3.249	3.428	170.7	179.0
SHH ··· OHH (**46**)	−3.269	−2.664	3.520	3.619	177.5	175.3
HOH ··· $S(CH_3)_2$ (**47**)	−5.909	−3.236	3.268	3.390	147.3	97.7
HSH ··· SH_2 (**51**)	−1.761	−0.877	4.191	4.509	175.9	175.5
HOH ··· SH_2 (**52**)	−3.345	−2.061	3.543	3.775	170.0	175.5
Methylammonium ··· OH_2 (**60**)	−21.099	−19.298	2.731	2.813	174.1	176.7
OHH ··· $(-)O_2CCH_3$ (**66**)	−22.510	−21.848	2.842	2.883	144.8	145.7

[a] See Figure 1. [b] In kilocalories per mole. [c] In angstroms. [d] In degrees.

quantitative level, we note that 63 of the 66 listed interaction energies are indeed more negative for MMFF94 than for HF/6-31G*, often by the desired 10–15%. Overall, the average interaction energies are −7.23 kcal/mol for HF/6-31G* and −8.04 kcal/mol for MMFF94, 11% larger in magnitude. Many comparisons of donor and acceptor abilities suggested by the HF/6-31G* calculations (and supported by general chemical knowledge) are properly reflected by MMFF94. For example, MMFF94 properly understands that phenyl and vinyl alcohol are stronger hydrogen-bond donors (**7**, **9**) but weaker acceptors (**8**, **10**) than are water, aliphatic alcohols, and ethers (**1**, **3–6**), and recognizes that furan (**11**) is a weaker acceptor still. Similarly, MMFF94 properly rates ammonia as a stronger acceptor (**34**) than donor (**59**), and correctly assesses vinylamine and aniline as being stronger donors (**44**, **45**) than ammonia. In addition, MMFF94 closely parallels the HF/6-31G* assessments of the strengths of heteroaromatic amines such as imidazole, indole, pyrrole, and pyridine (**36–40**). MMFF94 also correctly assigns lesser abilities both as donors and acceptors to thiols and sulfides (**46–52**). Qualitative errors relative to HF/6-31G* occur in only a few instances—for example, for cyclic formamide (**24**) and for the less strongly bound dimer of water with methyl formate (**19**), where the calculated MMFF94 interaction energies are about 1 kcal/mol less negative than are the HF/6-31G* values. That Models 2–4 also predict more negative binding energies in these cases suggests that the MMFF94 results are in error. Two of the largest quantitative differences arise for the cyclic water dimer (2) and the stacked *t*-NMA dimer (**27**), where the MMFF94 energies are significantly more negative than those for the HF/6-31G* model. As noted above, however, the MP2 models also give more negative binding energies, suggesting that the MMFF94 results are not unreasonable.

Comparison of the X···Z heteratom distances in Table VI shows that most of the MMFF94 distances are smaller than the HF/6-31G* values by 0.2–0.3 Å, as desired. Indeed, only for the two equivalent O···O interactions in the C_s cyclic water dimer (**2**) are the MMFF94 distances longer, and then only trivially so. Moreover, the MMFF94 distances are shorter than are those for the MP2/DZP-optimized geometry (Table VIII) in this case, suggesting that the deficiency may lie in the HF/6-31G* model rather than in MMFF94. In all, the average for the listed heteroatom distances is 3.105 Å for HF/6-31G* and 2.896 Å for MMFF94, for a difference of 0.209 Å. The largest irregularities occur in weakly bound dimers involving sulfur. For example, the O···S distance is nearly 0.4 Å shorter in the HF/6-31G* dimer of water with dimethyl sulfide (**47**) than in the dimer of water with hydrogen sulfide (**52**); MMFF94 qualitatively reproduces this trend but shows a much smaller variation—as does geometry optimization at the more sophisticated MP2/DZP *ab initio* level (Table VIII).

Most X—H···Z hydrogen-bond angles are well produced. Some significant discrepancies are evident, however, a few of which appear to reflect limitations in the HF/6-31G* model rather than in MMFF. One discrepancy occurs for the second listed O—H···O interaction for the dimer of H_2O with *cis,cis*-oxalic acid (**15**), where both the long MMFF94 H···O distance and the small O—H···O angle indicate that the force-field and *ab initio* geometries differ qualitatively. Another occurs for the dimer of water with dimethyl sulfide (**47**), where HF/6-31G* gives a bifurcated structure while MMFF94 gives a single, essentially linear, O—H···S hydrogen bond. In this case, the MP2/DZP model also shows a single, though more strongly distorted, O—H···S hydrogen bond (Table VIII). A third case, though not evident from Table VI, occurs for the ammonia dimer. As noted above, the cyclic C_{2h} structure (**32**) is a minimum on the HF/6-31G* surface, but is a saddle-point for MMFF94 and at various higher levels of *ab initio* theory.[33,73] The MP2/DZP model, for example, like MMFF94, prefers the linear C_s dimer (**33**).

We also examined the abilities of the MM2X force field[76] and of an earlier version of MMFF, designated "MMFF91," to reproduce the dimer data. We found that MM2X also performs fairly well. Thus, about half of the MM2X interaction energies were more negative than the unadjusted HF/6-31G* values, and all but one of the MM2X X···Z distances was smaller. Except for the cyclic water dimer (−5.45 kcal/mol error) and the stacked *t*-NMA dimer (−7.58 kcal/mol error), the differences between the MM2X and HF/6-31G* dimer energies were never larger than about 2 kcal/mol. However, the MM2X interaction energies did not, in general, show the enhancement relative to the HF/6-31G* values we have targeted in developing MMFF94. In contrast to the methodology employed in deriving MMFF94, the MMFF91 force field used electrostatic parameters (bond charge increments) derived simply by fitting to the (unscaled) HF/6-31G* dipole moments in the manner described earlier, and for nonhydrogen

atoms used the vdW parameters previously published.[10] This force field showed much larger errors than either MMFF94 or MM2X. Thus, only 13 of the 66 dimer energies were more negative than their (unadjusted) HF/6-31G* counterparts, and only about three fourths of the $X \cdots Z$ distances were shorter. Moreover, some large errors were found; in four cases, the MMFF91 interaction energies were more positive than the *ab initio* values by 4–6 kcal/mol. The contrast with MMFF94 shows that an explicit parameterization was required to obtain the present generally good results.

COMPARISON TO OPLS

Table IX compares MMFF94 and OPLS results for 22 of the dimers included in Table VI. For the linear H_2O dimer (**1**), TIP3P results are listed, as this frequently used model is similar in construction to "MMFF" water. Most of the other listed results, however, use the TIP4P model. The tabulated comparisons show that the two sets give similar binding energies and heteroatom distances in most cases. For imidazole–water dimers (**36**, **37**), MMFF94 yields significantly more negative binding energies both when imidazole acts as a donor and when it acts as an acceptor. In this case, however, our fully optimized HF/6-31G* calculations yielded theoretical binding energies of −6.365 and −7.063 kcal/mol, respectively, whereas the partial HF/6-31G* optimizations used in deriving OPLS gave significantly smaller binding energies of −5.73 and −6.24 kcal/mol.[32] Also, MMFF94 intentionally deepens the HF/6-31G* binding energies, whereas OPLS in this case reproduces the unadjusted (and unoptimized) energies used in its derivation.

For dimers involving divalent sulfur (**46**, **47**, **51**, and **52**), OPLS yields more negative binding energies and shorter heteroatom distances. These dif-

TABLE IX.
Comparison of MMFF94 and OPLS Hydrogen-Bond Energies and Geometries.

Dimer (structure number)[a]	Energy (kcal / mol)		$X \cdots Z$ distance (Å)	
	MMFF[b]	OPLS	MMFF	OPLS
$HOH \cdots OH_2$ (**1**)	−6.61	−6.50[c]	2.75	2.74
$HOH \cdots HOH$, cyclic C_s (**2**)	−4.91	−4.77[d]	2.78	2.79
$HOH \cdots OHCH_3$ (**3**)	−6.32	−6.77[d]	2.73	2.72
$CH_3OH \cdots OH_2$ (**4**)	−6.14	−6.03[d]	2.75	2.77
$CH_3OH \cdots OHCH_3$ (**5**)	−6.03	−6.80[d]	2.73	2.73
$HOH \cdots O(CH_3)_2$ (**6**)	−5.98	−5.77[d]	2.77	2.75
t-NMA $\cdots OH_2$ (**22**)	−6.93	−6.7[e]	2.84	2.87
$HOH \cdots$ *t*-NMA (**23**)	−8.19	−7.4[e]	2.71	2.72
Formamide, cyclic (**24**)	−12.26	−14.07[f]	2.81	2.81
Formamide dimer, parallel (**25**)	−7.69	−7.96[f]	2.76	2.78
t-NMA dimer, antiparallel (**26**)	−9.08	−8.97[f]	2.76	2.79
t-NMA, antiparallel stacked (**27**)	−5.62	−5.6[f]	3.61	3.6
Imidazole $\cdots OH_2$ (**36**)	−7.92	−5.88[g]	1.84[k]	1.86[k]
$HOH \cdots$ imidazole (**37**)	−7.94	−6.24[g]	1.81[k]	1.88[k]
$HSH \cdots OH_2$ (**46**)	−2.91	−4.59[h]	3.38	3.10
$HOH \cdots S(CH_3)_2$ (**47**)	−3.48	−4.56[h]	3.35	3.08
$HSH \cdots SH_2$ (**51**)	−1.25	−2.48[h]	4.02	3.61
$HOH \cdots SH_2$ (**52**)	−2.00	−3.61[h]	3.49	3.22
Methylammonium $\cdots OH_2$ (**60**)	−21.17	−17.8[i]	2.62	2.72
Guanidinium $\cdots$ OHH (**61**)	−19.02	−16.1[j]	3.21[l]	3.33[l]
Imidazolium $\cdots OH_2$ (**62**)	−17.60	−16.0[j]	2.62	2.72
$OHH \cdots (-)O_2CCH_3$, bidentate (**66**)	−22.62	−18.9[i]	3.06[m]	3.06[m]

[a] See Figure 1. [b] From Table VI. [c] Results for TIP3P water from ref. 43. All other listed OPLS dimers involving water appear to use the TIP4P model. [d] From ref. 37. [e] W. L. Jorgensen and C. J. Swenson, *J. Am. Chem. Soc.*, **107**, 1489–1496 (1985). [f] W. L. Jorgensen and C. J. Swenson, *J. Am. Chem. Soc.*, **107**, 569–587 (1985). [g] From ref. 32. [h] W. L. Jorgensen, *J. Phys. Chem.*, **90**, 6379–6388 (1986). [i] From ref. 77. [j] From ref. 45. [k] $H \cdots Z$ distance (Å). [l] $C \cdots O$ distance (Å). [m] $O \cdots C$ distance (Å).

ferences correlate with the greater polarity of S—H bonds for OPLS noted earlier. The OPLS heteroatom distances are also considerably shorter than the optimized MP2/DZP distances (Table VIII), suggesting to us that the OPLS values may be too short. On the other hand, the MP2/6-31 + G** (Table VII) and especially the MP2/DZP binding energies for these dimers are more negative and usually agree better with OPLS than with either MMFF94 or HF/6-31G*. This better agreement, however, may trace to a correlation-enhanced basis-set-superposition error and may be artifactual. We note that the BSSE-corrected HF/6-31G* interaction energies are significantly less negative and agree less well with OPLS than with MMFF94. Thus, while the OPLS parameters have been found to describe the neat liquids well, there is room for uncertainty as to whether the OPLS or the MMFF representation is the more appropriate for simulations in aqueous solution.

Finally, we note that MMFF94 yields more strongly negative binding energies than does OPLS for the dimers of water with the methylammonium,[77] guanidinium,[45] imidazolium,[45] and acetate[77] ions (**60**–**62** and **66**). For these systems, the OPLS dimerization energies are uniformly less negative than the HF/6-31G* values listed in Table VI. The MMFF94 energies, in contrast, are uniformly more negative, by about 5% on average. Intentionally, these MMFF94 enhancements are smaller than those sought for neutral dimers. We presumed that the additional polarization induced by surrounding water molecules should be less important for the ion–water dimers because substantial polarization is already induced in the gas-phase dimers by the strong electric fields involved. Jorgensen and coworkers went further in developing OPLS: they set their ion–water dimerization energies to values *less* negative than those given by their HF/6-31G* calculations, arguing that the *ab initio* values overestimate the gas-phase dimerization energies.[77] It is not clear, however, that gas-phase energies are the right quantities to emulate; certainly, Bartlett's gas-phase dimerization energy of approximately −4.7 kcal/mol for the water–water dimer[34] would not be. We note, however, that OPLS is reported to give reasonable heats of hydration for the methylammonium and acetate ions, among others.[77] It is thus possible that MMFF94 will overestimate these quantities. Explicit simulations will be needed to determine whether a reparameterization of MMFF94 for charged systems will be needed.

Concluding Discussion

The Merck Molecular Force Field (MMFF94) uses a novel "Buf-14-7" form for vdW interactions and employs a similarly buffered coulomb term for electrostatic interactions. The latter uses partial atomic charges constructed in a straightforward manner from bond-polarity-related "bond charge increments." This approach not only confers generality but also enables MMFF94 to describe a broad range of chemical structures and interactions well. The nonbonded parameters were determined by explicit parameterization against high-quality computational data on dimerization energies for methane and hydrogen, and, using simpler HF/6-31G* calculations, by fitting to scaled diple moments and to scaled HF/6-31G* intermolecular interaction energies and geometries for hydrogen-bonded dimers. The derived parameters reproduce values and trends in the computed interaction energies and geometries quite well. MMFF94's uniform parameterization against HF/6-31G* data is meant to ensure that a proper balance is achieved between solvent–solvent, solvent–solute, and solute–solute interactions in aqueous simulations. As MacKerell, Wiórkiewicz-Kuczera, and Karplus emphasize,[36b] this balance is essential if reliable results are to be obtained in condensed-phase simulations.

Our approach to parameterizing MMFF bears some similarities to that used by Jorgensen and coworkers in developing OPLS.[32,37,45] These workers also use HF/6-31G* data, particularly when experimental condensed-phase data are lacking. Indeed, a number of our hydrogen-bonded dimers are derived from systems they described. Our approach even more strongly resembles that of MacKerell, Wiórkiewicz-Kuczera, and Karplus,[36] who have used similarly scaled HF/6-31G* data in developing a new parameterization for CHARMM. Encouragingly, MMFF94 and the frequently used TIP3P and SPC water models describe the isolated water dimer very similarly. Moreover, MMFF94 and OPLS usually yield similar charge distributions and usually give similar results for hydrogen-bonded dimers. Given the success OPLS has enjoyed in modeling applications, we believe that this close correspondence augers well for MMFF94.

Our approach, on the other hand, differs substantially from that taken by Hagler and coworkers in deriving CFF93, the Biosym Consortium

force field.[78] Like MMFF94, the Consortium force field also constructs partial atomic charges from bond charge increments. In many cases, however, the Biosym approach evidently obtains many electrostatic and vdW parameters from fits to crystallographic data. This approach is most readily applicable for hydrocarbons, carboxylic acids, and amides, functional groups for which many crystal structure exist for which heats of sublimation have been measured.[8a] Even in these cases, however, a possible concern is that the fitting of *energy minimized* force-field structures to *thermally averaged* crystal structures disregards temperature effects that may not be negligible. Thus, MacKerell, Wiórkiewicz-Kuczera, and Karplus in a theoretical study of nucleic acid base pairs have found that hydrogen bond lengths typically increase by ~ 0.1 Å when an energy-minimized (0 K) crystal is warmed to room temperature in a molecular-dynamics simulation.[36b] In qualitative agreement, Coppens and coworkers report that neutron-diffraction studies find that hydrogen bond lengths in crystals of α-oxalic acid dihydrate increase by 0.04–0.06 Å on warming from 100 K to room temperature;[79] X-ray determinations at 100 K[79] and 299 K[80] show that the volume of the unit cell concomitantly increases by ~ 4%. Further study may be needed to determine how accurately CFF93 predicts hydrogen-bond distances and crystal volumes in simulations carried out at room temperature and at 1 atmosphere pressure.[81]

Another concern with the Biosym Consortium approach is that, for most functional groups, few if any heats of sublimation are available for relevant crystal structures. It is not yet clear what approach will be taken in such cases, as it would not seem prudent to derive nonbonded *energy* parameters from fits to crystal *geometries* alone. One possible approach might be to obtain nonbonded parameters from quantum calculations on dimers.[82] Another approach might be to use the procedure recently described for polycarbonates,[83] in which fits to crystal geometries (no heats of sublimation were available) were employed to determine a uniform scaling factor to be applied to bond charge increments extracted from ESP charges. While this approach appears promising, the manner in which a proper balance between solvent–solvent, solvent–solute, and solute–solute interactions will be effected for CFF93 remains to be seen. For the polycarbonate force field, for example, the scaled bond charge increments yield charges of −0.448 for carbonyl oxygen and +0.355 for hydroxyl hydrogen in a carboxylic acid moiety. These charges are substantially smaller in magnitude than the charges of −0.57 and +0.50 we found to be needed for MMFF94 to fit the scaled HF/6-31G* dimer data. Furthermore, the CFF93 charge on the acidic hydroxyl hydrogen is smaller than the charge of +0.40 or greater used in MMFF94 and other three-point models[42,43] for the less readily ionized hydroxyl hydrogen of water.

Our approach also differs fundamentally from that used for MM3. A key difference arises from the fact that MM3 normally employs a dielectric constant of 1.5 and uses bond dipoles chosen to reproduce gas phase (i.e., not polarization-enhanced) dipole moments. To offset the first factor, MM3 fits the overall interaction potential by including relatively large Exp-6 terms that contribute up to 6 kcal/mol of stabilization to hydrogen-bonding interactions.[84] The resultant hydrogen-bonding energies, by design, then correspond to experimental gas-phase dimerization energies rather than to the larger values used by force fields that seek to account for aqueous-phase polarization in an average way. MM3(94), for example, yields an interaction energy of −4.77 kcal/mol for the linear water dimer,[85] whereas SPC, TIP3P, and MMFF94 all yield interaction energies of about −6.5 kcal/mol; the gas-phase dimerization energy has been estimated to be about −5.4 kcal/mol.[86,87] These differences suggest to us that MM3 may not be properly configured to handle molecular-dynamics simulations in explicit water.

Finally, our approach also differs from that taken by Kollman and coworkers, who obtain electrostatic potential fit (ESP) charges, most recently those derived using the newly developed RESP approach,[53] for use in AMBER calculations.[88] In assessing the suitability of ESP charges, it is important to keep in mind that MMFF94 and other contemporary force fields greatly simplify the way in which nonbonded vdW and electrostatic interactions are represented. To compensate for the approximation or outright omission of many of the nonbonded terms that would appear in the "correct" energy expression, parameters for those terms that are retained may need to be determined from explicit fits to appropriate data on nonbonded interactions. Because ESP charges are not chosen in this way, no guarantee can be given that they (or any other set of externally defined charges—even the "true" charges, could these somehow be defined) will satisfactorily reproduce intermolecular interaction energies. Thus, as discussed previously, Carlson et al. found that CHELPG ESP charges for acetamide give much too negative a

free energy of hydration.[35] While another set of ESP charges might perform better in this case, it will be difficult to establish that a given ESP method produces charges that predict interaction energies consistently well enough to achieve a proper balance between solvent–solvent, solvent–solute, and solute–solute interactions.

To focus further on the simplifications made in modeling nonbonded interactions, we note that nearly all force fields currently used in general molecular modeling not only neglect polarizability[89] but also ignore contributions from higher atomic multipoles.[90,91] In addition, most employ fixed, atom-centered charges that are independent of conformation and of the geometry of interatomic interactions (e.g., ignore charge transfer, an effect that quantum mechanical partitioning schemes suggest may play a significant, though not predominant, role in hydrogen-bonding interactions[92]). Moreover, all widely used force fields employ an overly simplified "spherical-atom" model for vdW interactions, though the nuclear offset used for C—H bonds in MM2[5] and MM3[6] is helpful. As envisioned in part I,[1] we expect future force fields to remedy these and other deficiencies by employing a superior physical representation that will allow the following objective to be met: *to define a force field that describes gas-phase molecular properties accurately and that behaves properly when embedded in the condensed phase.* However, while much remains to be done to achieve this objective, substantial progress has already been made in understanding and reducing to practice certain key aspects of the requisite physics.[90,91,93] Undoubtedly, such more elaborate force fields will become increasingly difficult to parameterize. We believe that only computational data will suffice, and expect MMFF94 to be one of a succession of force fields derived almost solely from such data.

But while more elaborate physical models need to be explored, the current MMFF94 force field can at least be said to incorporate a broad and systematic parameterization of nonbonded interactions. This parameteration encompasses not only the range of systems examined in this article but, through the extension of the core MMFF94 parameterization described in part V, also covers a variety of inorganic ions of interest in biomolecular modeling.[4] Definitive evidence of the utility of the present MMFF parameterization, to be sure, will have to await the results of liquid-phase simulations we are undertaking in collaboration with others. Some such simulations will use the recent implementation of MMFF94 in the Merck version of CHARMm[94] or the pending implementation of MMFF94 in the academic program, CHARMM[9]; others may use the forthcoming implementation of MMFF94 in the BatchMin module of the MacroModel program suite.[95] One such study is expected to examine the ability of MMFF94, or a reparameterized or reformulated variant, to describe the properties of liquid water.[96a] Others are expected to examine free energies of solvation[96b] and free energies of binding in host–guest systems.

Supplementary Material

Computer-readable ASCII files of vdW, bond charge increment, and other MMFF94 parameter files,[49] and Appendix A (nonbonded parameters used for enzyme-site "context molecules" in OPTIMOL/MMFF calculations)[51] are included in Supplementary Material.

Acknowledgments

I thank Prof. Martin Karplus for suggesting the use of scaled HF/6-31G* interaction energies and geometries,[36] Bruce Bush and Chris Bayly for helpful discussions on ESP fitting, and Mike Miller for interfacing PDM88 to Gaussian 88.

References

1. Part I: T. A. Halgren, *J. Comput. Chem.* (this issue).
2. Part III: T. A. Halgren, *J. Comput. Chem.* (this issue).
3. Part IV: T. A. Halgren and R. B. Nachbar, *J. Comput. Chem.* (this issue).
4. Part V: T. A. Halgren, *J. Comput. Chem.* (this issue).
5. (a) N. L. Allinger, *J. Am. Chem. Soc.*, **89**, 8127 (1977); (b) U. Burkert and N. L. Allinger, *Molecular Mechanics*, American Chemical Society, Washington, DC, 1982; (c) N. L. Allinger and Y. Yuh, *QCPE*, **12**, 395 (1980).
6. N. L. Allinger, Y. H. Yuh, and J.-H. Lii, *J. Am Chem. Soc.*, **111**, 8551–8566 (1989).
7. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.*, **7**, 230–252 (1986); S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr., and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765–784 (1984).
8. (a) S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5111–5121 (1979), and references therein; (b) P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler, *Proteins*, **4**, 31–47 (1988). This is the force field used in the program DISCOVER distributed by Biosym Technologies, Inc. (San Diego, CA).

9. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187–217 (1983).
10. T. A. Halgren, *J. Am. Chem. Soc.*, **114**, 7827–7843 (1992).
11. M. Waldman and A. T. Hagler, *J. Comput. Chem.*, **14**, 1077–1084 (1993).
12. See, for example: *Computer Simulation of Biomolecular Systems*, W. F. van Gunsteren and P. K. Weiner, Eds. ESCOM, Leiden, 1989.
13. J. R. Hart and A. K. Rappé, *J. Chem. Phys.*, **97**, 1109–1115 (1992).
14. We should point out that Hart and Rappé applied the Morse potential to a wider range of reference vdW potentials than were used in testing the Buf-14-7 form. We do not know whether the Buf-14-7 form would also require additional shape parameters (e.g., by taking γ and/or δ in ref. 10 as variables) if applied to the same range of reference potentials.
15. R. Cambi, D. Cappelletti, G. Liutti, and F. Pirani, *J. Chem. Phys.*, **95**, 1852–1861 (1991).
16. (a) N. Mataga and K. Nishimoto, *Z. Physik. Chem.*, **13**, 140 (1957); (b) K. Ohno, *Theor. Chem. Acta*, **2**, 219–227 (1964); (c) J. Klopman, *J. Am. Chem. Soc.*, **86**, 4550 (1964).
17. (a) N. Gresh, P. Claverie, and A. Pullman, *Int. J. Quantum Chem.*, **22**, 199–215 (1982); (b) N. Gresh, P. Claverie, and A. Pullman, *Int. J. Quantum Chem. Symp.*, **13**, 243 (1979). (c) J. Langlet, P. Claverie, and F. Caron, In *Intermolecular Forces*, B. Pullman, Ed., Reidel, Dordrecht, 1981, pp. 397–429.
18. M. K. Holloway, J. M. Wai, T. A. Halgren, P. M. D. Fitzgerald, J. P. Vacca, B. D. Dorsey, R. B. Levin, W. J. Thompson, L. J. Chen, S. J. deSolms, N. Gaffin, A. K. Ghosh, E. A. Giuliani, S. L. Graham, J. P. Guare, R. W. Hungate, T. A. Lyle, W. M. Sanders, T. J. Tucker, M. Wiggins, C. M. Wiscount, O. W. Woltersdorf, S. D. Young, P. L. Darke, and J. A. Zugay, *J. Med. Chem.*, **38**, 305–317 (1995).
19. M. M. Szczesniak, G. Chalasinski, S. M. Cybulski, and S. Scheiner, *J. Chem. Phys.*, **93**, 4243–4253 (1990).
20. A. J. Sadlej, *Coll Czech. Chem. Commun.*, **53**, 1995 (1988).
21. F. Boys and F. Bernardi, *Mol. Phys.*, **19**, 553 (1970).
22. K. B. Wiberg and M. A. Murcko, *J. Comput. Chem.*, **8**, 1124–1130 (1987).
23. R. Krishnan, M. J. Frisch, and J. A. Pople, *J. Chem. Phys.* **72**, 4244–4245 (1980); R. Krishnan and J. A. Pople, *Int. J. Quantum Chem.*, **14**, 91 (1978).
24. M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Raghavachari, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol, and J. A. Pople, Gaussian 88, Gaussian, Inc., Pittsburgh, PA, 1988, as modified at Merck for improved I/O performance by E. M. Fluder.
25. We replicated the hydrogen dimer calculations in this work because we were not always certain of the meaning of the separation parameter cited by Wiberg and Murcko (ref. 22).
26. J. L. M. Dillen, *J. Comput. Chem.*, **11**, 1125–1138 (1990).
27. For H_2, we have assumed a foreshortening of 0.057 Å in the MM3 calculations such that each vdW center lies this distance into the bond from the parent hydrogen nucleus. This foreshortening amounts to 7.7% of the H—H bond length of 0.741 Å employed. Strictly speaking, MM3 calculations are not defined for H_2, but this fractional foreshortening is the same as is used for C—H bonds: cf. J.-H. Lii and N. L. Allinger, *J. Am. Chem. Soc.*, **111**, 8576–8582 (1989). The MM3 calculations were performed using a modified version of OPTIMOL/MM2X (ref. 18).
28. Like the MM3 calculations, the CHARMm calculations were performed using OPTIMOL. While these calculations were performed with the vdW parameters distributed with QUANTA Version 3.2, we note that the QUANTA 3.3, QUANTA 3.4, and QUANTA 4.0 values for these parameters are unchanged.
29. W. J. Hehre, L. Random, P. v. R Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986, Chapter 4. The 6-31G* and 6-31 + G** basis sets are also known as 6-31G(d) and 6-31 + G(p,d), respectively.
30. For the most part, these systems correspond to those listed in Table I of ref. 1.
31. D. E. Williams, *J. Comput. Chem.*, **9**, 745–763 (1988), as implemented by M. Miller (Merck Research Laboratories) in Gaussian 88 (ref. 24).
32. J. Pranata, S. G. Wierschke, and W. L. Jorgensen, *J. Am. Chem. Soc.*, **113**, 2810–2819 (1991), and references therein.
33. M. J. Frisch, J. E. Del Bene, J. S. Binkley, and H. F. Schaefer III, *J. Chem. Phys.*, **84**, 2279–2289 (1986).
34. K. Szalewicz, S. J. Cole, W. Kolos, and R. J. Bartlett, *J. Chem. Phys.*, **89**, 3662–3673 (1988).
35. H. A. Carlson, T. B. Nguyen, M. Orozco, and W. L. Jorgensen, *J. Comput. Chem.*, **14**, 1240–1249 (1993).
36. (a) A. D. MacKerell Jr. and M. Karplus, *J. Phys. Chem.*, **95**, 10559–10560 (1991). (b) A. D. MacKerrel Jr., J. Wiórkiewicz-Kuczera, and M. Karplus, *J. Am. Chem. Soc.* (in press).
37. W. L. Jorgensen, *J. Phys. Chem.*, **90**, 1276–1284 (1986).
38. PROBE is a computer program used to derive molecular-mechanics parameters in least-squares fits to data obtained from *ab initio* calculations. PROBE was created for the Biosym Consortium on Potential Energy Functions by Biosym Technologies, Inc. (San Diego, CA) (cf. ref. 79).
39. K. B. Wiberg, R. F. W. Bader, and C. D. H. Lau, *J. Am. Chem. Soc.*, **109**, 1001–1012 (1987).
40. Cf. G. Karlström, P. Linse, A. Wallqvist, and B. Jönsson, *J. Am. Chem. Soc.*, **105**, 3777–3782 (1983). A similar value of ca. 0.143 on hydrogen in benzene can be inferred from the structure of benzene and from the data and analysis in M. R. Battaglia, A. D. Buckingham, and J. H. Williams, *Chem. Phys. Lett.*, **78**, 421–423 (1981).
41. W. L. Jorgensen, E. R. Laird, T. B. Nguyen, and J. Tirado-Rives, *J. Comput. Chem.*, **14**, 206–215 (1993).
42. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, In *Intermolecular Forces*, B. Pullman, Ed. Reidel, Dordrecht, 1981, pp. 331–342.
43. W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, *J. Chem. Phys.*, **79**, 926–935 (1983).
44. That is, a force field like MMFF which does not displace the vdW center along the bond away from the hydrogen nucleus would have to use a compensatingly smaller vdW radius to obtain a similar interaction between vdW centers in roughly head-on contacts.
45. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657–1666 (1988).
46. S. V. Oneil, D. J. Nesbit, P. Rosmus, H.-J. Werner, and D. C. Clary, *J. Chem. Phys.*, **91**, 711–721 (1989).

47. G. Chalasinski, S. M. Cybulski, M. M. Szczesniak, and S. Scheiner, *J. Chem. Phys.*, **91**, 7809–7817 (1989); G. Chalasinski, M. M. Szczesniak, and S. Scheiner, *J. Chem. Phys.*, **94**, 2807–2816 (1991).
48. These calculations used Gaussian 88 (ref. 24) and employed the MP2 model for electron correlation in conjunction with 6-31G basis sets supplemented by single diffuse Gaussian *p* functions on H and He (exponents 0.20, 0.30) and *d* functions on F (exponent 0.35).
49. The nonbonded and other MMFF94 parameter files are included in Appendix B in the Supplementary Material for paper I (ref. 1).
50. The MM2X parameters are provided as supplementary material in ref. 18.
51. The supplementary material for this article (Appendix A), can be accessed on the Internet (see footnote * on the first page of this article).
52. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, **5**, 129–145 (1984).
53. (a) C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.*, **97**, 10269–10280 (1993); (b) W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, *J. Am. Chem. Soc.*, **115**, 9620–9631 (1993).
54. C. Brenneman and K. B. Wiberg, *J. Comput. Chem.*, **11**, 361–373 (1990). For the CHELP method on which CHELPG is based, see also: L. E. Chirlian and M. M. Franci, *J. Comput. Chem.*, **8**, 894 (1987).
55. B. H. Besler, K. M. Merz, and P. A. Kollman, *J. Comput. Chem.*, **11**, 431–439 (1990).
56. M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian 92 (Revision C), Gaussian, Inc., Pittsburgh, PA, 1992.
57. These results are for dimers between "MMFF water" and acetamide constructed in the same manner as those reported in ref. 35. Specifically, we, too, had to restrict the water molecule in the anti conformer to lie in the plane of the carbonyl group to prevent collapse to the syn form, in which the oxygen of the water molecule interacts with a hydrogen on the amide nitrogen to form a second hydrogen bond (cf. Fig. 4 of ref. 35).
58. T. R. Stouch and D. E. Williams, *J. Comput. Chem.*, **13**, 622–632 (1992).
59. D. E. Williams, *Biopolymers*, **29**, 1367–1386 (1990).
60. J. Gasteiger and M. Marsili, *Tetrahedron*, **36**, 3219–3228 (1980).
61. W. J. Mortier, K. Van Genechten, and J. Gasteiger, *J. Am. Chem. Soc.*, **107**, 829–835 (1985); W. Mortier, S. K. Ghosh, and S. Shankar, *J. Am. Chem. Soc.*, **108**, 4315–4320 (1986).
62. (a) J. Mullay, *J. Comput. Chem.*, **12**, 369–375 (1991); (b) J. Mullay, *J. Comput. Chem.*, **9**, 399–405 (1988).
63. A. K. Rappé and W. A. Goddard III, *J. Phys. Chem.*, **95**, 3358–3363 (1991).
64. The Gasteiger charges, which do not depend on the geometry, were obtained using the implementation of this method in SYBYL, Version 5.5 (Tripos Associates, Inc., St. Louis, MO).
65. The QEq charges were obtained for MP2/6-31G*-optimized geometries using the QEq implementation in CERIUS (Molecular Simulations, Inc., Burlington, MA) for MP2/6-31G* optimized geometries.
66. This objection applies particularly strongly to an iterative approach like that employed by Mortier (ref. 61). Rappé and Goddard (ref. 63) suggest in their manuscript title that their approach is applicable for molecular-dynamics simulations, but present no algorithm for computing charge derivatives analytically. Moreover, their method as described employs a matrix inversion which has a higher power dependence on system size than does the calculation of the molecular-mechanics energy function itself.
67. W. L. Jorgensen and J. Gao, *J. A. Chem. Soc.*, **110**, 4212–4216 (1988).
68. G. Del Re, *J. Chem. Soc.*, 4031 (1958); G. Del Re, B. Pullman, and T. Yonezawa, *Biochim. Biophys. Acta*, **75**, 153 (1963).
69. (a) R. J. Abraham, G. H. Grant, I. S. Haworth, and P. E. Smith, *J. Computer-Aided Mol. Design*, **5**, 21–39 (1991); (b) R. J. Abraham and P. E. Smith, *J. Computer-Aided Mol. Design*, **3**, 175–187 (1989).
70. This judgment is supported by the preliminary results of attempting to apply a charge equilibration scheme to heteroaromatic compounds: M. D. Miller and T. A. Halgren (unpublished research).
71. For dipole directions, the values cited are weighted rms deviations computed as: rms = $\{\Sigma_i w_i \Delta_i^2 / \Sigma_i w_i\}^{1/2}$, where $\Delta_i = \cos^{-1}(\mu_i^{calc} \cdot \mu_i^{ref} / (|\mu_i^{calc}||\mu_i^{ref}| + 1 \times 10^{-10}))$ and $w_i = |\mu_i^{calc}||\mu_i^{ref}|$. By weighting the squared deviations in direction by the product of the force-field calculated and (scaled) reference HF/6-31G* dipole moments and by computing the angle between the dipole moments in a manner which yields a finite value even when one or both dipole moments vanish, this formula ensures that the weighted rms deviation mainly reflects differences in direction in cases in which the dipole moment is relatively large in magnitude. In particular, it keeps large differences in direction from dominating the comparison when either of the dipole moments is small in magnitude.
72. Using unscaled HF/6-31G* dipole moments, MM2X gives an rms error of 0.58 D in the dipole magnitudes. The rms deviation for dipole directions is of course unaffected.
73. M. J. Frisch, J. A. Pople, and J. E. Del Bene, *J. Phys. Chem.*, **89**, 3664–3669 (1985).
74. G. Chalasinski and M. Gutowski, *Chem. Rev.*, **88**, 943–962 (1988).
75. Double-zeta plus polarization (DZP) basis sets: T. H. Dunning Jr., *J. Chem. Phys.*, **53**, 2823–2833 (1970); T. H. Dunning Jr. and P. J. Hay, In *Methods of Electronic Structure Theory*, H. F. Schaefer III, Ed., Plenum Press, New York, 1977, Vol. 2, pp. 1–27. The polarization exponents were those defined in ref. 3, i.e., 0.80 for hydrogen, 0.60, 0.85, 1.20, and 1.65 for C, N, O, and F, and 0.50, 0.65 and 0.80 for P, S, and Cl.
76. These calculations used the Lennard-Jones 9-6 potential for intermolecular vdW interactions, and used partial charges computed for use in intermolecular interactions (cf. ref. 18).
77. W. L. Jorgensen and J. Gao, *J. Phys. Chem.*, **90**, 2174–2182 (1986).
78. J. R. Maple, M.-J. Hwang, T. P. Stockfish, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler, *J. Comput. Chem.*, **15**, 161–182 1994; M.-J. Hwang, T. P. Stockfish, and A. T. Hagler, *J. Am. Chem. Soc.*, **116**, 2515–2525 (1994).

79. E. D. Stevens and P. Coppens, *Acta Cryst.*, **B36**, 1864–1876 (1980).

80. R. G. Delaplane and J. A. Ibers, *Acta Cryst.*, **B25**, 2423–2437 (1969).

81. For a pertinent comparison of OPLS and earlier Lifson et al. (ref. 8a) force-field results for liquid-phase simulations, see: W. L. Jorgensen and C. J. Swenson, *J. Am. Chem. Soc.*, **107**, 569–578 (1985). Consistent with the present arguments, the Hagler–Lifson force field, which was derived using a similar energy-minimization methodology, gave longer energy-minimized hydrogen bond lengths for gas-phase dimers than Jorgensen and Swenson found would allow OPLS to correctly reproduce liquid-phase properties of simple amides.

82. U. Dinur and A. T. Hagler, *J. Am. Chem. Soc.*, **111**, 5149–5151 (1989).

83. H. S. Sun, S. J. Mumbry, J. R. Maple, and A. T. Hagler, *J. Am. Chem. Soc.*, **116**, 2978–2987 (1994).

84. See, for example, the "vdW and hydrogen bonding" parameters listed in Table I of: J.-H. Lii and N. L. Allinger, *J. Comput. Chem.*, **12**, 186–199 (1991).

85. J.-H. Lii and N. L. Allinger, *J. Comput. Chem.*, **12**, 186–199 (1991).

86. L. A. Curtiss, D. J. Frurip, and M. Blander, *J. Chem. Phys.*, **71**, 2703 (1979).

87. J. Reimers, R. Watts, and M. Klein, *Chem. Phys.*, **64**, 95 (1982).

88. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179–5197 (1995).

89. MM2 and MM3 allow contributions from induced dipoles to be included in the calculation (N. L. Allinger, personal communication). This feature does not, however, appear to be used in routine work.

90. (a) U. Dinur and A. T. Hagler, *J. Comput. Chem.*, **16**, 154–170 (1994); (b) U. Dinur and A. T. Hagler, *J. Chem. Phys.*, **91**, 2949–2958 (1989); (c) U. Dinur and A. T. Hagler, *J. Chem. Phys.*, **91**, 2959–2970 (1989).

91. See, for example: (a) U. Dinur, *J. Comput. Chem.*, **12**, 91–105 (1991); (b) U. Dinur, *J. Comput. Chem.*, **12**, 469–486 (1991); (c) U. Dinur, *J. Phys. Chem.*, **94**, 5669–5671 (1990).

92. H. Umeyama and K. Morokuma, *J. Am. Chem. Soc.*, **99**, 1316–1332 (1977), and references therein. This article indicates that charge transfer contributes ~ 2 kcal/mol to the interaction energy for the water dimer, a quantity smaller than the electrostatic stabilization of ~ 8 kcal;mol but by no means negligible. The authors also note that charge transfer appears to play a larger role in weaker hydrogen-bonding complexes, and we think it possible that this may particularly be true for hydrogen bonds to such higher row atoms as sulfur.

93. See, for example: (a) G. Corongiu, M. Migliore, and E. Clementi, *J. Chem. Phys.*, **90**, 4629 (1989); (b) L. X. Dang, J. E. Rice, J. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **113**, 2481–2486 (1991), and references therein; (c) M. Sprik, *J. Phys. Chem.*, **95**, 2283–2291 (1991), and references therein; (d) S.-B. Zhu, S. Yao, J.-B. Zhu, S. Singh, and G. W. Robinson, *J. Phys. Chem.*, **95**, 6211–6217 (1991); (e) S. W. Rick, S. J. Stuart, and B. J. Berne, *J. Chem. Phys.*, **101**, 6141–6156 (1994); (f) D. N. Bernardo, Y. Ding, K. Krough-Jespersen, and R. M. Levy, *J. Phys. Chem.*, **98**, 4180–4187 (1994).

94. R. Czerminski, J. L. Banks, and T. A. Halgren (unpublished research). Currently, a version of CHARMm that supports the earlier and less widely parameterized MMFF93 force field (refs. 2–4) is available from Molecular Simulations, Inc. (San Diego, CA). However, while the local Merck code employs MMFF94, arrangements for including MMFF94 in the distributed MSI version have not yet been concluded.

95. P. S. Shenkin and T. A. Halgren (work in progress). The MacroModel program suite and its BatchMin module, developed in the laboratories of Prof. Clark Still, are available from Columbia University (New York, NY).

96. (a) J. L. Banks, R. Czerminski, and T. A. Halgren (work in progress); (b) B. L. Bush, J. L. Banks, and T. A. Halgren (work in progress).