Development of an All-Atom Force Field for Heterocycles. Properties of Liquid Pyrrole, Furan, Diazoles, and Oxazoles

Nora A. McDonald[†] and William L. Jorgensen*

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Received: February 18, 1998

The approach to the extension of the optimized potentials for liquid simulations all-atom (OPLS-AA) force field, which was tested for pyridine and the diazenes in the first paper of this series, has been extended to pyrrole, pyrazole, imidazole, furan, isoxazole, and oxazole. Standard OPLS Lennard-Jones parameters are used for the nonbonded interactions in conjunction with partial charges obtained from fitting to RHF/6-31G* electrostatic potential surfaces. The harmonic bond stretching and angle bending terms are adopted mostly from the AMBER force field, although the addition of two new atom types was required. The resultant force field was tested by computing the gas-phase structures of the heterocycles, heterocycle−water interaction energies, properties of the pure liquids, and the relative free energies of hydration for pyrrole and imidazole.

Introduction

Monte Carlo (MC) and molecular dynamics simulations can be successfully applied to a wide range of problems.[1,2] One area of great interest is the study of protein−ligand interactions, which are relevant to drug design.[3] Given the size of the molecular systems, these calculations are normally performed with molecular mechanics and, hence, classical force fields rather than quantum mechanics. The optimized potentials for liquid simulations-all-atom (OPLS-AA) force field includes parameters for proteins and most common organic functional groups.[4,5] Because so many drug candidates contain heterocyclic rings, one focus of recent force field development has been expansion to a wider range of heterocyclic compounds.

In the first paper of this series, development of OPLS-AA parameters for pyridine and the diazenes was reported.[6] The approach taken in that work arose from several considerations. (a) The force field is modular such that large molecules may be assembled from component fragments. (b) The utility of the force field should be tested by checking the accuracy of conformational energetics and liquid-phase properties; the average errors for computed densities and heats of vaporization using the OPLS-AA force field is only ca. 2%.[4] (c) The number of parameters should be kept to a minimum. To this end, the Lennard-Jones parameters have been found to be highly transferable; for example, all aromatic carbons have the same Lennard-Jones parameters. Some partial atomic charges have been found to be transferable as well, such as the O and H charges for alcohols, but in general charges may not easily be transferred.

The large number of heterocyclic compounds makes the assignment of partial charges a substantial challenge. Optimization for individual systems, as for common functional groups,[4,5] is not viable. The alternative approach outlined in the diazene study is to use partial charges derived from fitting to the RHF/6-31G* electrostatic potential surface (EPS) with the CHELPG procedure.[7] Standard OPLS-AA Lennard-Jones parameters are also used, along with bonded terms mostly from the AMBER94 force field.[8] EPS charges are a common choice for describing intermolecular electrostatic interactions;[8,9] it has been demonstrated that EPS charges, coupled with the OPLS-AA Lennard-Jones parameters, yield free energies of hydration in TIP4P water with an average error of 1.1 kcal mol^{-1}.[9] They have also proven useful for computing free energies of solvation in OPLS chloroform[10] and for numerous binding studies.[1,11,12]

The results from the diazene work further confirmed the utility of the EPS charges in OPLS-AA parameter development; the average errors in the computed densities and heats of vaporization for the four molecules were 0.8 and 2.7%. Free energies of hydration were also calculated for pyridine and pyrazine and produced errors of less than 1 kcal mol^{-1} in comparison to experimental data. The current work further tests the approach by extending the treatment to several five-membered heterocycles containing nitrogen and/or oxygen: pyrrole, pyrazole, imidazole, furan, isoxazole, and oxazole.

Computational Methods

Ab Initio Calculations. A variety of calculations were performed to both develop and test the force field. The structures of the heterocycles were calculated at the RHF/6-31G* level, as were the vibrational frequencies and partial atomic charges obtained by fitting to the electrostatic potential surface with the CHELPG procedure. Optimizations of the six intermolecular degrees of freedom in the heterocycle−water complexes were also performed; the structures of the heterocycles were held fixed in their optimized forms, and the gas-phase experimental geometry was used for the water monomer (r_{OH} = 0.9572 Å, θ_{HOH} = 104.52°). Single point calculations were carried out at the MP2/6-31G* level on the resultant RHF/6-31G* structures to assess the contribution due to electron correlation, and subsequently MP2/6-31G*//MP2/6-31G* calculations were also performed to reoptimize the six intermolecular degrees of freedom. Gas-phase heat capacities were calculated utilizing the RHF/6-31G* vibrational frequencies scaled by a factor of 0.91. All ab initio calculations were performed with Gaussian94.[13]

* To whom correspondence should be addressed.
[†] Present address: SmithKline Beecham Pharmaceuticals, King of Prussia, PA 19406-0939.

S1089-5647(98)01200-0 CCC: $15.00
Published on Web 09/19/1998

Force Field Calculations. The OPLS-AA force field consists of harmonic terms for the bond stretching and angle bending energies. In addition, torsional energetics are described by a Fourier series for each dihedral angle and the nonbonded energies between atoms separated by three or more bonds are represented by Coulomb and Lennard-Jones interactions (eqs 1−4). The nonbonded terms are scaled by a factor of 0.5 for

$$E_{\text{bond}} = \sum_i k_{\text{b},i} (r_i - r_{0,i})^2 \quad (1)$$

$$E_{\text{bend}} = \sum_i k_{\vartheta,i} (\vartheta_i - \vartheta_{0,i})^2 \quad (2)$$

$$E_{\text{torsion}} = \sum_i \{V_{0,i} + V_{1,i}(1 + \cos \varphi_i)/2 + V_{2,i}(1 - \cos 2\varphi_i)/2 + V_{3,i}(1 + \cos 3\varphi_i)/2\} \quad (3)$$

$$E_{\text{nb}} = \sum_i \sum_j \{q_i q_j e^2/r_{ij} + 4\epsilon_{ij}[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6]\} \quad (4)$$

intramolecular 1,4-interactions and are unscaled otherwise. As in the previous study,[6] the heterocycles were constrained to be planar, consistent with experimental and ab initio structures, and the intramolecular nonbonded energies were not included. Again, if distortion from planarity is desired, the appropriate torsional terms from the AMBER force field can be added. The only variable internal degrees of freedom, which were parametrized, were for the bond lengths and bond angles. The selection of the full set of parameters, listed in Tables 1 and 2, is discussed below. All geometry optimizations of the isolated heterocycles, of their dimers, and of the heterocycle−water complexes were performed with gradient-based procedures implemented in BOSS.[14]

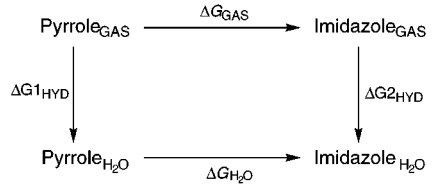
Monte Carlo Simulations. The MC methodology was previously described in detail.[6] Briefly, a periodic box of any liquid may be constructed in BOSS by providing a **Z**-matrix specifying the internal coordinates of one molecule of the liquid. Any or all of the internal degrees of freedom may be sampled. The initial MC configuration is created by overlaying the central atom of one molecule of the new liquid on each atom in a box of liquid argon, which has been scaled to the estimated volume of the solvent to be simulated. Each cubic box contained 267 solvent molecules, and for the present selection of liquids, the boxes were 30−32 Å on an edge. The cutoffs for the intermolecular interactions were at 12 Å, based roughly on the distance between ring centers, and the interaction energy was quadratically feathered to zero over the last 0.5 Å. A standard correction was applied to account for the Lennard-Jones interactions which were neglected beyond the cutoff distance.[15] All MC simulations of the pure liquids were run in the *NPT* ensemble at 25 °C and 1 atm pressure; volume changes were attempted every 1625 configurations, and the ranges for the intra- and intermolecular motions were adjusted to give ca. 25−40% acceptance ratios.

The liquid simulations consisted of a minimum of 2×10^6 configurations of equilibration and 4×10^6 configurations of averaging, while the gas-phase simulations required only half of the above values. These simulation lengths have been demonstrated to produce well-converged results for quantities normally compared to experiment, such as the heats of vaporization, liquid densities, and free energies of solvation, as well as for the radial distribution functions. The reported uncertainties (σ) were obtained from the fluctuations of the separate averages over batches of 4×10^5 configurations during averaging. The properties were computed as described previously.[4,10]

TABLE 1: Bond Stretching and Angle Bending Parameters for Heterocycles

bond stretching params			angle bending params		
bond	r_0 (Å)	k_b (kcal mol^{-1} Å^{-2})	angle	θ_0 (deg)	k_ϑ (kcal mol^{-1} rad^{-2})
NA−CW	1.381	427.0	NA−CW−CS	107.7	70.0
CS−CW	1.367	546.0	CW−NA−CW	109.8	70.0
CS−CS	1.424	469.0	CW−CS−CS	107.3	70.0
H −NA	1.010	434.0	H−NA−CW	120.0	35.0
HA−CW	1.080	367.0	HA−CW−NA	121.6	35.0
HA−CS	1.080	367.0	HA−CS−CW	125.7	35.0
NA−NB	1.349	400.0	HA−CW−CS	132.1	35.0
NB−CU	1.320	410.0	HA−CS−CS	127.5	35.0
NA−CW	1.381	427.0	NA−NB−CU	104.1	70.0
CA−CW	1.367	546.0	NB−NA−CW	113.1	56.0
HA−CU	1.080	367.0	NA−CW−CA	109.0	70.0
HA−CA	1.080	367.0	H −NA−NB	118.4	56.0
CU−CA	1.421	469.0	HA−CU−NB	118.9	35.0
NA−CR	1.343	477.0	HA−CA−CW	126.9	35.0
NB−CR	1.335	488.0	CU−CA−CW	103.8	70.0
CV−CW	1.370	520.0	NB−CU−CA	111.9	70.0
HA−CR	1.080	367.0	HA−CU−CA	128.6	35.0
HA−CV	1.080	367.0	HA−CW−CA	130.7	35.0
NB−CV	1.394	410.0	HA−CA−CU	128.2	35.0
CW−OS	1.360	340.0	NA−CR−NB	120.0	70.0
OS−NB	1.399	462.0	CR−NA−CW	109.8	70.0
OS−CR	1.357	462.0	NA−CW−CV	106.3	70.0
			H −NA−CR	120.0	35.0
			HA−CR−NA	120.0	35.0
			HA−CV−CW	128.2	35.0
			NB−CV−CW	111.0	70.0
			CR−NB−CV	110.0	70.0
			HA−CR−NB	120.0	35.0
			HA−CW−CV	130.7	35.0
			HA−CV−NB	120.0	35.0
			OS−CW−CS	110.6	70.0
			CW−OS−CW	106.5	70.0
			HA−CW−OS	113.4	35.0
			CA−CW−OS	110.6	70.0
			OS−NB−CU	105.3	70.0
			CW−OS−NB	108.9	70.0
			NB−CR−OS	115.0	70.0
			CR−OS−CW	104.0	70.0
			CV−CW−OS	108.0	70.0
			HA−CR−OS	117.0	35.0

Experimental free energies of hydration at 20 °C are also available for pyrrole and imidazole,[16] so free energy perturbation (FEP) calculations were performed and the relative free energies of hydration were determined via the thermodynamic cycle shown as follows:



$$\Delta\Delta G_{HYD} = \Delta G1_{HYD} - \Delta G2_{HYD} = \Delta G_{GAS} - \Delta G_{H_2O} \quad (5)$$

The difference in the experimental quantities, $\Delta\Delta G_{\text{HYD}}$, can then be compared to the computed quantity from ΔG_{GAS} and $\Delta G_{\text{H}_2\text{O}}$. This methodology and its utility have been well-documented.[1]

For the present case, the perturbation was performed between the new AA models of pyrrole and imidazole. The simulations were performed over 10 windows with double-wide sampling,

TABLE 2: Nonbonded Parameters for Heterocycles

molecule	atom	AMBER type	q (e)	σ (Å)	ϵ (kcal mol^{-1})
pyrrole	N	NA	−0.239	3.250	0.170
	C2	CW	−0.163	3.550	0.070
	C3	CS	−0.149	3.550	0.070
	H1	H	0.317	0.000	0.000
	H2	HA	0.155	2.420	0.030
	H3	HA	0.118	2.420	0.030
pyrazole	N1	NA	−0.059	3.250	0.170
	N2	NB	−0.491	3.250	0.170
	C3	CU	0.246	3.550	0.070
	C4	CA	−0.320	3.550	0.070
	C5	CW	−0.034	3.550	0.070
	H1	H	0.301	0.000	0.000
	H3	HA	0.072	2.420	0.030
	H4	HA	0.150	2.420	0.030
	H5	HA	0.135	2.420	0.030
imidazole	N1	NA	−0.257	3.250	0.170
	C2	CR	0.275	3.550	0.070
	N3	NB	−0.563	3.250	0.170
	C4	CV	0.185	3.550	0.070
	C5	CW	−0.286	3.550	0.070
	H1	H	0.306	0.000	0.000
	H2	HA	0.078	2.420	0.030
	H4	HA	0.075	2.420	0.030
	H5	HA	0.187	2.420	0.030
furan	O	OS	−0.190	2.900	0.140
	C2	CW	−0.019	3.550	0.070
	C3	CS	−0.154	3.550	0.070
	H2	HA	0.142	2.420	0.030
	H3	HA	0.126	2.420	0.030
isoxazole	O	OS	−0.122	2.900	0.140
	N	NB	−0.413	3.250	0.170
	C3	CU	0.405	3.550	0.070
	C4	CA	−0.455	3.550	0.070
	C5	CW	0.250	3.550	0.070
	H3	HA	0.053	2.420	0.030
	H4	HA	0.184	2.420	0.030
	H5	HA	0.098	2.420	0.030
oxazole	O	OS	−0.257	2.900	0.140
	C2	CR	0.511	3.550	0.070
	N	NB	−0.590	3.250	0.170
	C4	CV	0.169	3.550	0.070
	C5	CW	−0.148	3.550	0.070
	H2	HA	0.043	2.420	0.030
	H4	HA	0.091	2.420	0.030
	H5	HA	0.181	2.420	0.030

to give 20 free energy increments. The aqueous simulations consisted of one solute molecule in a box of 260 TIP4P water molecules, which underwent ca. 4×10^6 configurations of equilibration and 6×10^6 configurations of averaging with Metropolis sampling at 20 °C and 1 atm. Both the solvent−solvent and the solute−solvent cutoffs were 9.5 Å, based roughly on the ring centers and the water oxygen.

Results and Discussion

Force Field Parameters. The force field parameters were developed in a manner similar to that used for the six-membered heterocycles.[6] In that study, the CA−CA−CA and CA−CA−HA angle bending and all the bond stretching parameters, with the exception of the NC−NC bond in pyrazine, were taken from the AMBER94 force field. The remainder of the bonded terms was chosen to reproduce the experimental geometries, and the force constants were taken from similar types of interactions in the AMBER force field. For the present study, this procedure has again been followed; the results are presented in Table 1. It should be noted that although only previously defined AMBER atom types were used in the diazene study, two new atom types have been defined for the purposes of the present work: type CS in pyrrole and furan and type CU in pyrazole and isoxazole (Figure 1). Atom type CS is needed as an aromatic carbon, which does not have bond angles of 120° to all of its neighbors, and CU was defined for an aromatic carbon, which abuts type NB in a heteroatom−heteroatom bond. The CA−heteroatom-(−heteroatom) r_0 value in AMBER is 1.40 Å, as opposed to the experimental values of ∼1.30 Å for this type of bond in the present molecules.[17,18] Once these additions were made to the atom types list, it was possible to reproduce the experimental geometries accurately, as illustrated in Figure 2 (experimental values are given in parentheses).[17−22] As in the diazene study, the average errors are less than 0.01 Å for the bond lengths and 1° for the bond angles.

TABLE 3: Computed and Experimental Dipole Moments (*D*) for the Heterocycles

	pyrrole	pyrazole	imidazole	furan	isoxazole	oxazole
exptl[a]	1.74	2.21	3.80	0.72	2.90	1.50
OPLS-AA	1.85	2.57	3.89	0.77	3.59	1.48

[a] Reference 23.

For the nonbonded parameters, the Lennard-Jones parameters from the OPLS force field and the 6-31G* EPS charges are presented in Table 2. The dipole moments calculated with the parameters given in Tables 1 and 2, and their experimental counterparts, are presented in Table 3.[23] RHF/6-31G* EPS charges consistently lead to dipole moments which are 10−20% greater than gas-phase experimental values. These enhancements are necessary in order to reproduce condensed-phase properties, presumably because they compensate for the lack of explicit polarization in additive models. The dipole enhancement for the present molecules is not uniform and varies from ca. 24% for isoxazole to a slight decrease compared to experiment for oxazole.

It should again be noted that if accurate vibrational frequencies are desired from the force field, the force constants should be refined.[6] The focus of the OPLS-AA parameter development remains on intermolecular interactions and on accurately reproducing such thermodynamic quantities as heats of vaporization and free energies of solution.

Heterocycle−Water Complexes. The lowest energy structures of each heterocycle−water complex from the OPLS-AA calculations are presented in Figure 3. These are the global minima from a conformational search, which included all degrees of freedom with the exception of the planarity constraint on the heterocycles and the use of a rigid TIP4P water molecule.[24] The interaction energies obtained from the OPLS-AA and ab initio calculations are compared in Table 4. As noted previously,[6] ab initio interaction energies at the RHF/6-31G* level are preferable to those at the MP2/6-31G* level, as the latter overestimate the complexation energy. It should also be noted that the hydrogen bond lengths from the force field calculations are consistently 0.2−0.3 Å shorter than those obtained from ab initio calculations. Without this contraction, densities of the pure liquids are computed to be too low. For the structures in Figure 3, the 6-31G* values for the X···HO hydrogen bond lengths are 2.076, 2.228, 2.261, and 2.102 Å for pyrrole, pyrazole (to N and to N−H), and imidazole, respectively. The 6-31G* values for furan, isoxazole, and oxazole are 2.231, 2.220, and 2.158 Å. The 6-31G* structures are similar to the OPLS-AA ones with the exception that the O···HO bond for furan is more nearly in the plane of the ring, and the bifurcated hydrogen bond obtained with the force field for isoxazole is not found; instead a more nearly linear hydrogen bond to the nitrogen is preferred.

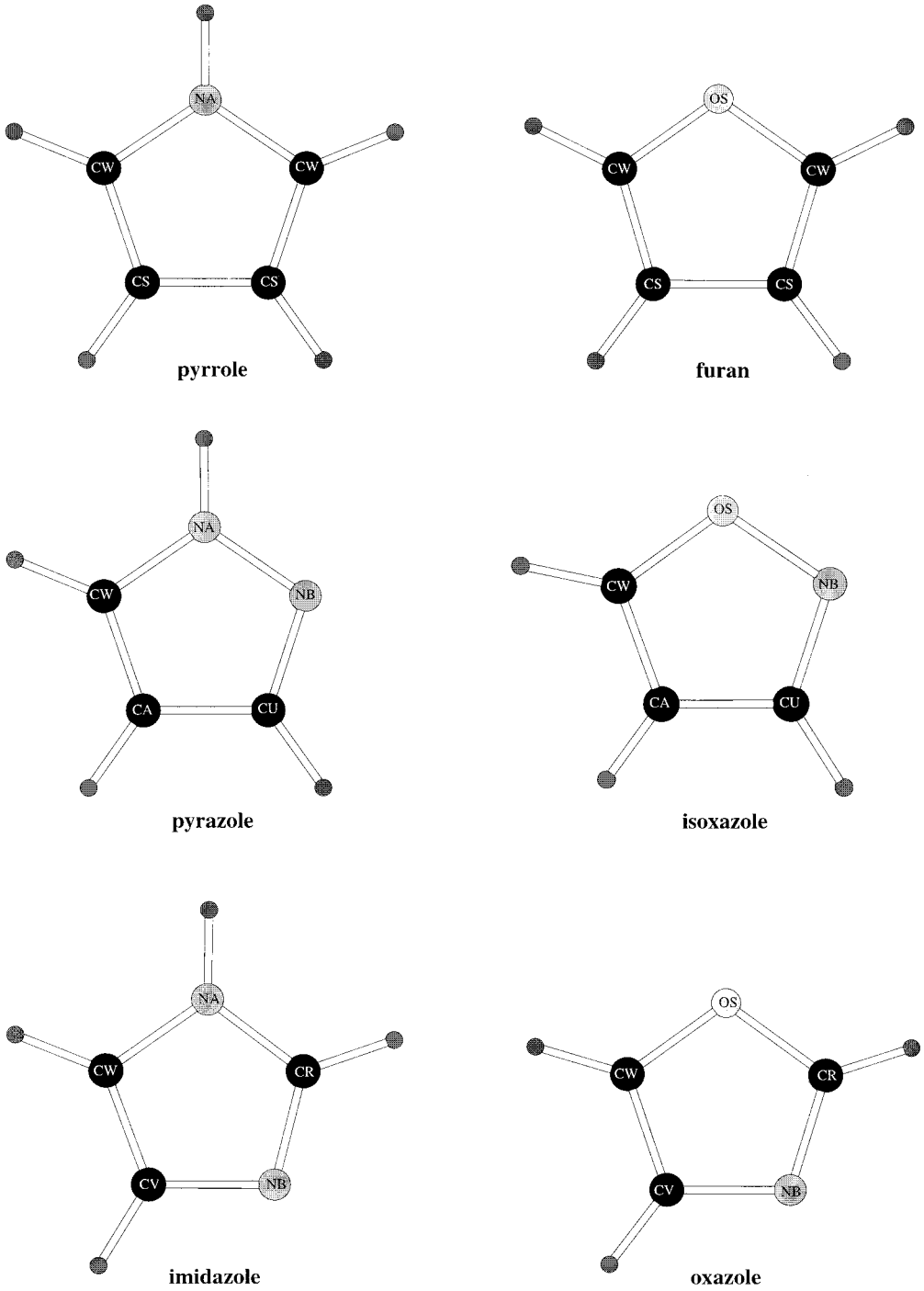


Figure 1. AMBER atom types for the heterocycles.

TABLE 4: Computed Interaction Energies (kcal mol^{-1}) for Heterocycle−Water Complexes[a]

method	pyrrole	pyrazole	imidazole	furan	isoxazole	oxazole
OPLS-AA	−5.81	−6.90	−6.72	−3.37	−5.63	−5.87
RHF/6-31G*//RHF/6-31G*	−5.29	−7.95	−7.07	−3.62	−5.25	−5.43
MP2/6-31G*//RHF/6-31G*	−7.01	−10.85	−9.12	−5.04	−6.85	−7.38
MP2/6-31G*//MP2/6-31G*	−7.20	−11.09	−9.32	−5.20	−7.04	−7.61

[a] The optimizations with the ab initio calculations only varied the six intermolecular degrees of freedom.

The OPLS-AA interaction energies for the heterocycle−water complexes differ from the RHF/6-31G* values by ca. 0.5 kcal mol^{-1} or less, with the exception of pyrazole, which is too weakly bound by 1 kcal mol^{-1}. The accord is remarkable in view of the simple origin of the charges and the point-charge model. It may also be noted that although the interaction energy is overestimated for pyrrole with the force field compared to the ab initio values, it is underestimated for both of the diazoles. The pattern is reversed for furan and the two oxazoles, as discussed further below. Nevertheless, the ordering of the interaction energies for the six molecules is the same from the force field and from all three sets of ab initio calculations. The

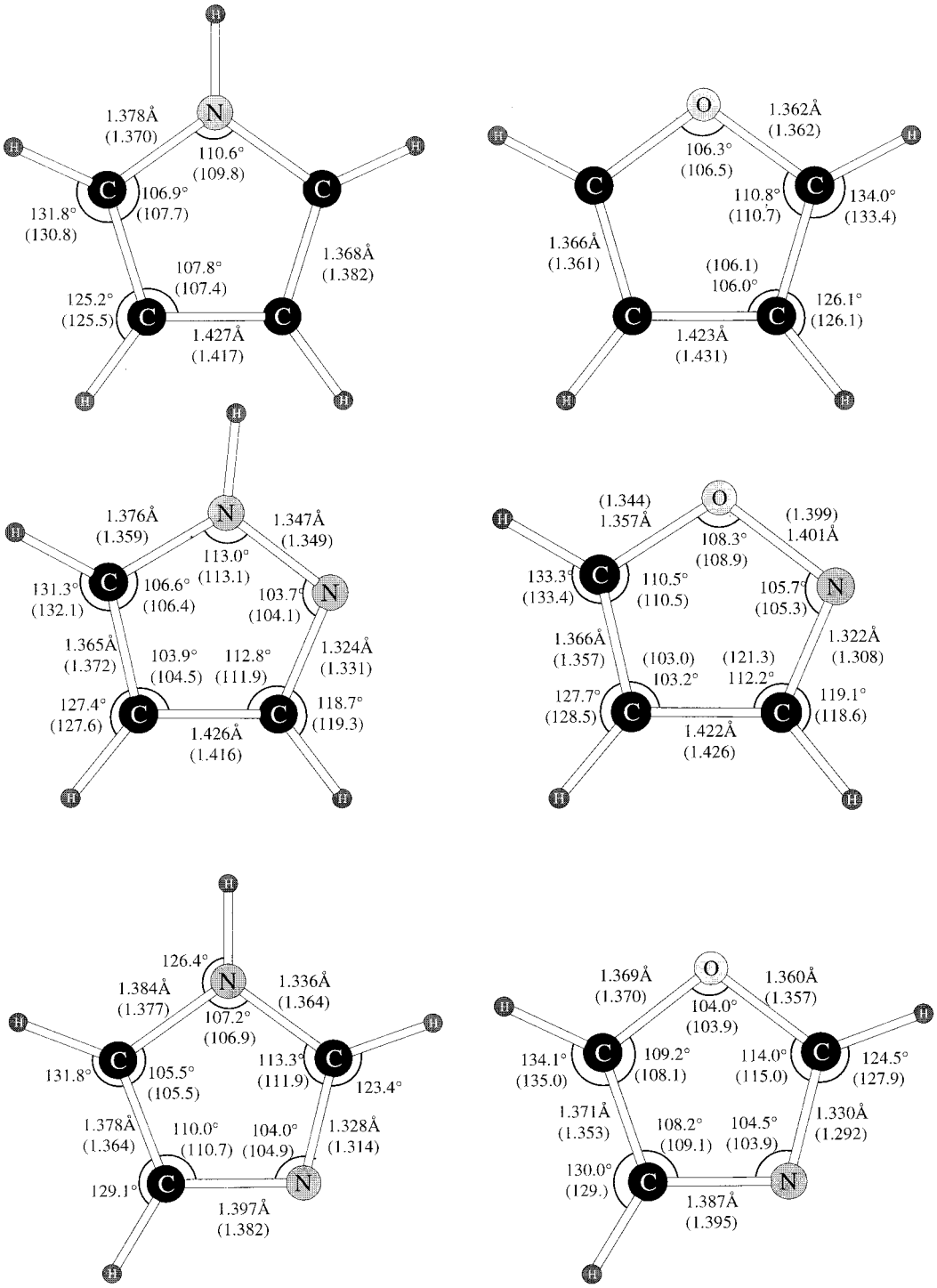


Figure 2. Computed structures for the heterocycles with the OPLS-AA force field. Experimental gas-phase results in parentheses.[17−23]

hydrogen bond strengths are in the following order: pyrazole > imidazole > oxazole > pyrrole > isoxazole > furan. This ordering matches the gas-phase proton affinities with the exception of imidazole and pyrazole, which are reversed.[25,26] The relatively stronger hydrogen bonding for pyrazole likely comes from the additional interaction that it has with the water oxygen. It is also interesting to note that the most attractive interaction energy for both the diazole and oxazole complexes is not to the heterocycle with the largest dipole moment. In the case of pyrazole vs imidazole, the cyclic hydrogen bonding with pyrazole compensates for its lower dipole moment. For isoxazole vs oxazole, the bifurcated hydrogen bond to the nitrogen and oxygen of isoxazole is slightly weaker than the one linear hydrogen bond to the oxazole nitrogen.

Examples of additional energy minima found with the force field are shown in Figure 4. These are the next-to-lowest energy structures found from the conformational searches. The interaction energies for these complexes, other than with the oxazoles, are within 0.6 kcal mol^{-1} of the global minima, so they should be detectable in gas-phase experiments. The structures found for pyrrole and furan, the mono-heteroatom species, have a hydrogen bond to the π system in a manner similar to that found for benzene.[27,28] The other structures either utilize the alternate heteroatom (oxazole) or have the water oxygen as the hydrogen-

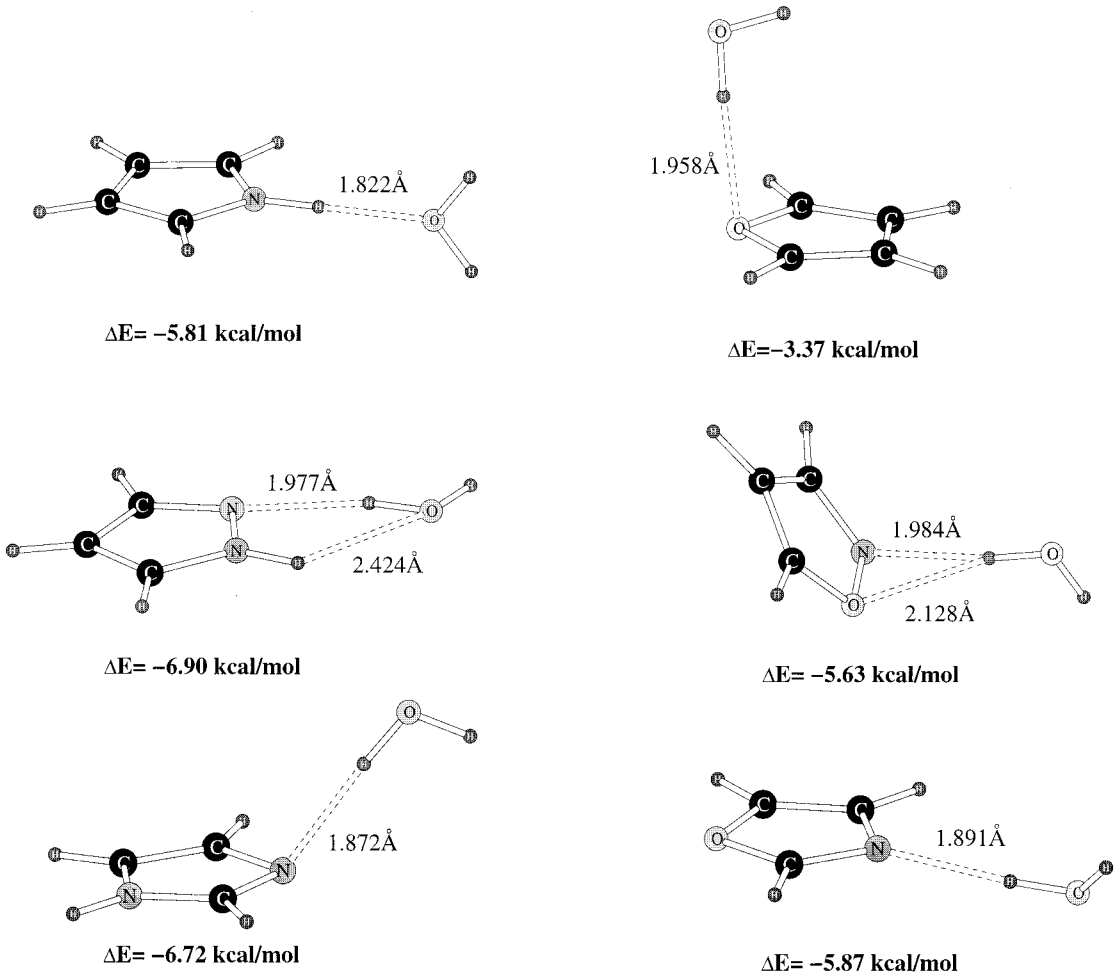


Figure 3. Lowest energy structures for the heterocycle−water complexes from the optimizations with the OPLS-AA force field.

TABLE 5: Computed and Experimental Heats of Vaporization (kcal mol^{-1})[a]

				ΔH_{vap}	
liquid	T (°C)	E(gas)	E(liquid)	calc	expt
pyrrole	25.0	5.60	−4.41	10.60	10.80[b]
pyrazole	25.0	5.12	−7.65	13.36	
imidazole	25.0	7.60	−6.75	14.94	
furan	25.0	4.69	−1.49	6.77	6.56[c]
isoxazole	25.0	4.03	−5.59	10.21	8.71[d]
oxazole	25.0	5.51	−3.01	9.11	7.77[d]

[a] Statistical uncertainties, $\pm 1\sigma$, for the computed E(gas), E(liquid), and ΔH_{vap} are 0.02, 0.02, and 0.03 kcal mol^{-1}. [b] Reference 29. [c] Reference 30. [d] Reference 31.

TABLE 6: Computed and Experimental Heat Capacities (cal mol $^{-1}$ K^{-1}) and Liquid Densities (g cm^{-3})[a]

			C_p(liquid)		density	
liquid	T (°C)	C_p°(gas)[b] calc	calc	expt	calc	expt
pyrrole	25.0	16.85	31.44	30.52[c]	0.987	0.966[c]
pyrazole	25.0	15.29	28.15		1.103	
imidazole	25.0	15.43	30.84		1.120	
furan	25.0	15.25	26.68	27.44[d]	0.943	0.931[e]
isoxazole	25.0	14.14	26.99	22.17[f]	1.102	1.07[f]
oxazole	25.0	13.95	26.92	23.07[f]	1.097	1.08[f]

[a] Statistical uncertainties, $\pm 1\sigma$, for the computed C_p(liquid) and density are 1.5 cal mol $^{-1}$ K^{-1} and 0.001 g cm^{-3}. [b] Computed with RHF/6-31G* vibrational frequencies scaled by 0.91. [c] Reference 29. [d] Reference 30. [e] Reference 32. [f] Reference 31.

bond acceptor, either from N−H (the diazoles) or from two C−H···O interactions (isoxazole).

Thermodynamics and Structure for the Pure Liquids. The heats of vaporization, heat capacities, and densities from the simulations of the pure liquids are presented in Tables 5 and 6 along with the corresponding experimental values. The computed and experimental heats of vaporization for pyrrole and furan are in good accord with experiment, with an average error of 2.5%. However, while the computed difference in heats of vaporization for oxazole and isoxazole is about right at 1 kcal mol^{-1}, the absolute values are both too high by ca. 1.5 kcal mol^{-1}, raising the overall error to 9.9%. Pyrazole and imidazole are solids at room temperature. Thus, the computed properties at 25 °C are for the supercooled liquids and are included for completeness. For the four liquids with experimental data, the computed values for the liquid densities are good, with an average error of 2.0%. The computed values for the heat capacities are also in reasonable accord with the experimental data in view of the statistical uncertainty of ca. 1.5 cal mol^{-1} deg^{-1}. The only significant errors are for isoxazole and oxazole, for which the heat capacities, like the heats of vaporization, are overestimated. These results are surprising for oxazole in view of its low dipole moment (Table 3).

The computed N1−N1 (where N1 is the hydrogen-bearing nitrogen), N−H1 (where N is the nonhydrogen-bearing nitrogen in pyrazole and imidazole), and O−O radial distribution functions (rdfs) are shown in Figures 5−7, respectively. These curves, $g_{AB}(r)$, give the probability (normalized for the bulk density) of finding an atom of type B at a distance r from an

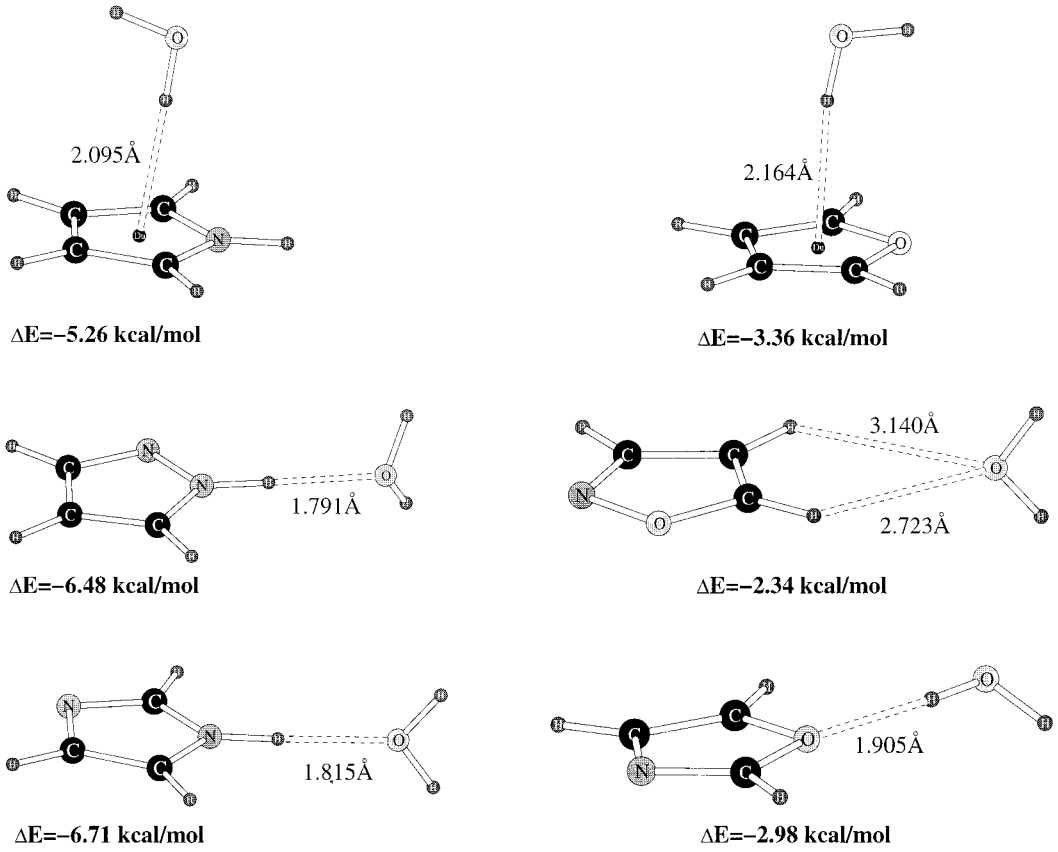


Figure 4. Additional energy minima for the heterocycle−water complexes from the optimizations with the OPLS-AA force field.

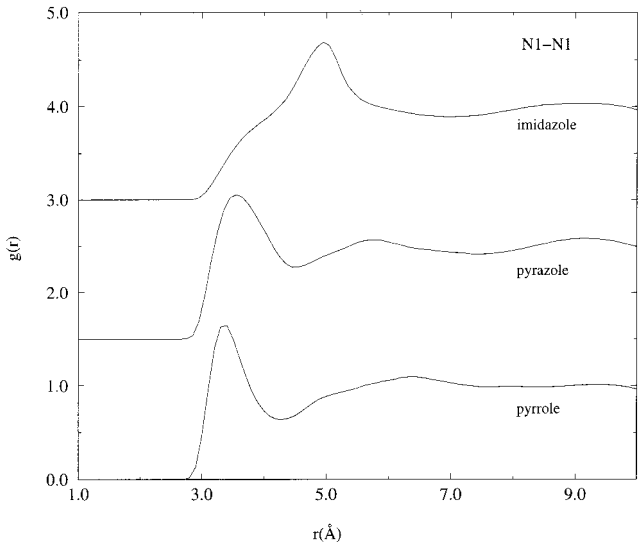


Figure 5. Computed intermolecular N1−N1 radial distribution functions for pyrrole, pyrazole, and imidazole at 25 °C. N1 is the hydrogen-bearing nitrogen.

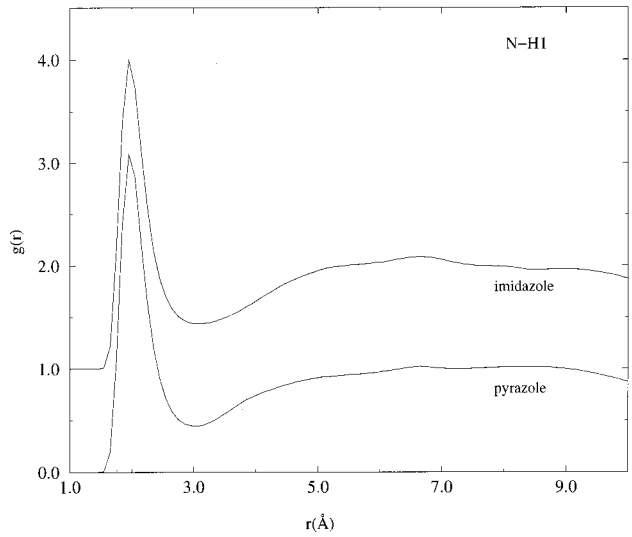


Figure 6. Computed intermolecular N−H1 radial distribution functions for pyrazole and imidazole at 25 °C. N is the nonhydrogen-bearing nitrogen, and H1 is attached to N1.

atom of type A in the liquid. For reference, the global minima and the next-to-lowest energy structures for dimers of each of the heterocycles located with the force field are presented in Figures 8 and 9, respectively. These structures were obtained from conformational searches of the six intermolecular degrees of freedom with the geometries of the monomers fixed in their optimal structures (Figure 2). Although the homodimers of the six-membered heterocycles, as found with the force field, were exclusively π-stacked,[6] the five-membered species exhibit a wider variety of structures. The pyrrole dimer is T-shaped, pyrazole is the most strongly bound with a cyclic arrangement of hydrogen bonds, and imidazole is also hydrogen-bonded with the ring planes almost perpendicular to each other. The furan dimer, like pyrrole, is T-shaped, and both isoxazole and oxazole are π-stacked, although the ring planes are not perfectly parallel.

Examination of Figure 5 reveals that the radii at the maximal peak heights correspond quite closely with the N1−N1 distances in the optimized homodimers. Pyrrole has the shortest N1−N1 distance at ca. 3.3 Å in the optimized dimer, and the maximum peak height in the rdf occurs at 3.40 Å. A shift of the first peak to longer N1−N1 distances is expected from the thermal averaging in the liquids at 25 °C. For pyrazole, there is also a shift from the optimized N1−N1 distance of 3.05 Å in the dimer to 3.50 Å in the liquid. The maximum peak height

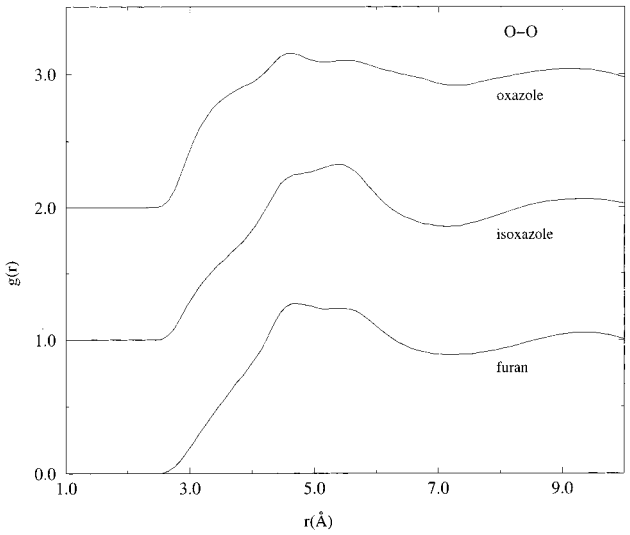


Figure 7. Computed intermolecular O−O radial distribution functions for furan, isoxazole, and oxazole at 25 °C.

for imidazole is at 5.0 Å, just under the optimized N1−N1 distance in the homodimer, and there is a substantial shoulder at distances of ca. 3−4 Å in the rdf. Although attempts were made to locate alternate gas-phase minima for imidazole, none were found. However, graphical examination of the full simulation cell reveals that, in the pure liquid, there is a significant population of both T-shaped and π-stacked dimers. Both of these feature N1−N1 distances well under 5 Å and appear to be contributing to the shoulder in the rdf at distances under 5 Å.

The N−H1 rdfs for pyrazole and imidazole in Figure 6 both have their first peak at 1.90 Å. Although this is to be expected for imidazole, it is at a shorter distance than the N−H1 distance in the homodimer for pyrazole (Figure 8). However, examination of the pyrazole cell reveals that the preponderance of the intermolecular interactions are more "head-to-tail", hence leading to chains in the liquid rather than cyclic dimers. The average length for the single hydrogen bonds is similar for imidazole and pyrazole, as illustrated in Figures 8 and 9.

Portions of the simulation cells for one representative configuration of liquid pyrazole and imidazole are shown in the stereoplots in Figures 10 and 11. One molecule near the center of the periodic box is shown with its neighbors that have at least one interatomic contact of 5 Å or less. Although both pure liquids exhibit extensive hydrogen-bonding, pyrazole forms

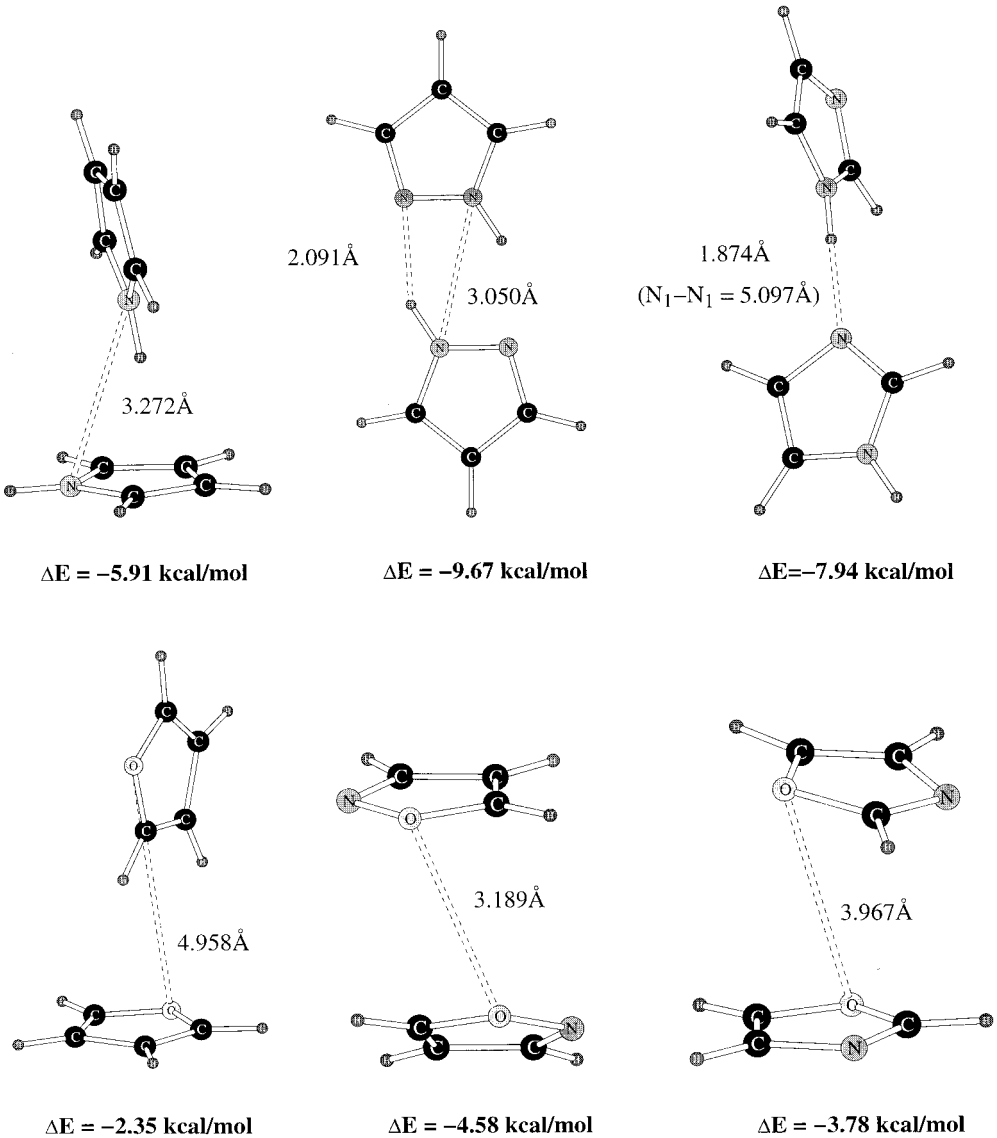


Figure 8. Lowest energy structures for homodimers of the heterocycles in the gas phase. Results from optimizations with the OPLS-AA force field.

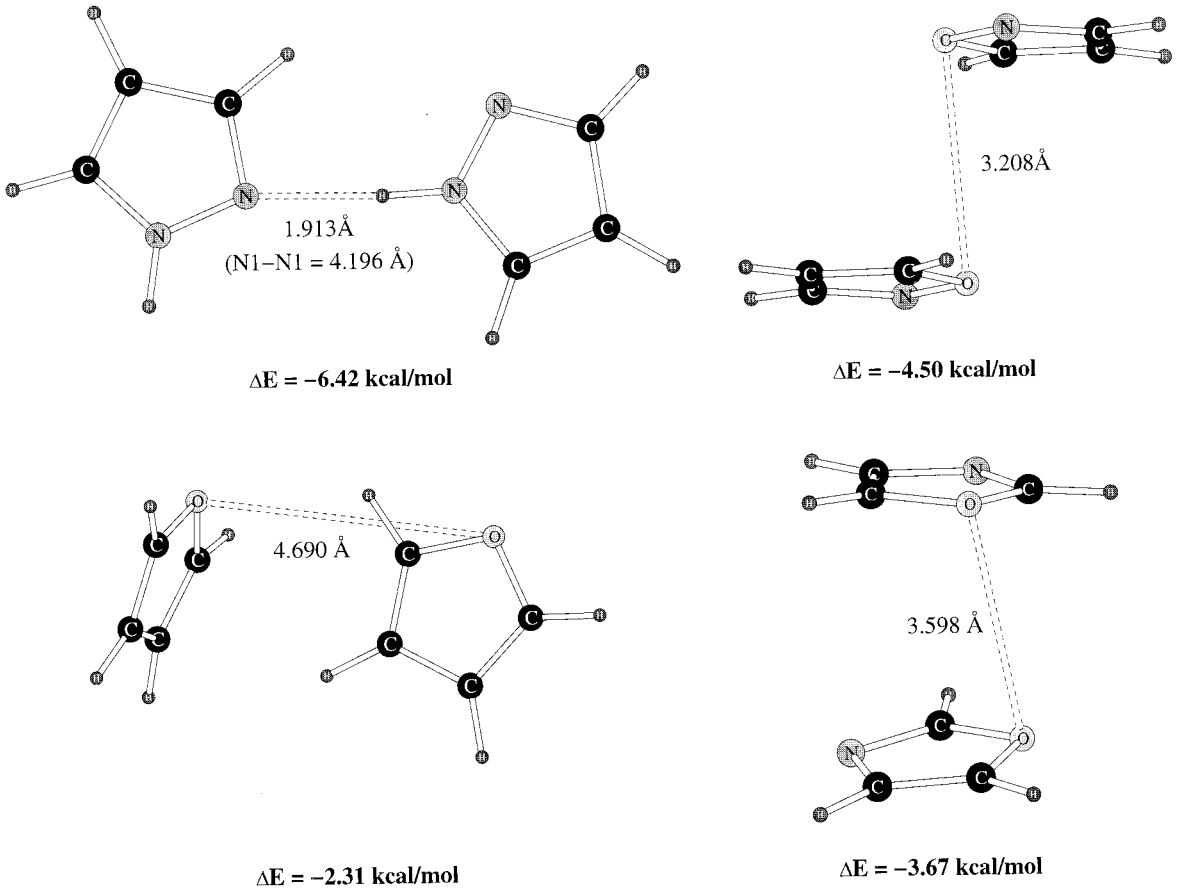


Figure 9. Next-to-lowest energy structures for homodimers of the heterocycles in the gas phase. Results from optimizations with the OPLS-AA force field.

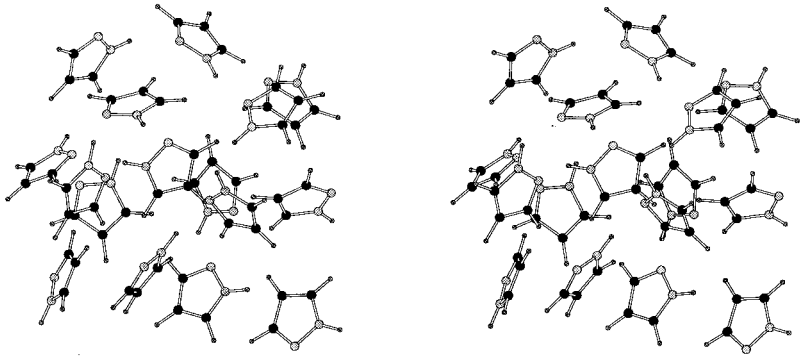


Figure 10. Stereoplot of a molecule near the center of the simulation cell and its nearest neighbors in liquid pyrazole.

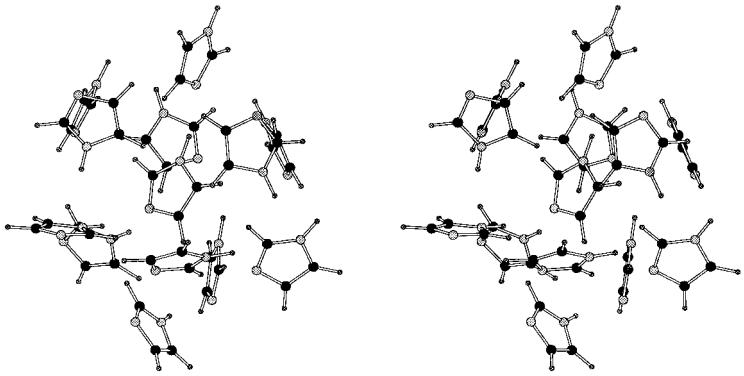


Figure 11. Stereoplot of a molecule near the center of the simulation cell and its nearest neighbors in liquid imidazole.

some hydrogen bonded chains and an appreciable number of cyclic trimers, whereas liquid imidazole mostly features hydrogen-bonded chains. Branching of the chains takes place at both N1 and N for both liquids. Integration of the N−H1 rdfs for pyrazole and imidazole out to the first minima at 3.0 Å yields a coordination number of 1.1. Thus, each monomer has an

average of 2.2 hydrogen-bonded neighbors, which is consistent with the occurrence of some chain branching.

The present oxygen-containing heterocycles cannot form hydrogen bonds and show relatively featureless rdfs typical of dipolar liquids. The shortest O−O distances in the liquids occur at 3−4 Å (Figure 7) and arise from offset stacked structures (Figure 8). The bands near 5 Å in the O−O rdfs can be assigned to T-shaped structures, as for the furan dimer in Figure 8.

Free Energies of Hydration. The FEP calculations for the mutation of pyrrole to imidazole yielded a free energy difference of 2.31 ± 0.00 kcal mol^{-1} in the gas phase and −2.11 ± 0.07 kcal mol^{-1} in TIP4P water. The experimental free energies of hydration for pyrrole and imidazole at 20 °C are −4.79 and −9.63 kcal mol^{-1}, respectively.[16] Using these values in eq 5 gives an experimental $\Delta\Delta G_{HYD}$ of 4.84 kcal mol^{-1}, and the corresponding value from the FEP calculations is 4.42 kcal mol^{-1}. This is in accord with the earlier observation that the force field relatively underestimates the imidazole−water attraction (Table 4), though the error of 0.4 kcal mol^{-1} for $\Delta\Delta G_{HYD}$ is small. The present results are also similar to those ($\Delta\Delta G_{HYD}$ = 4.1 kcal mol^{-1}) from an earlier calculation by Nagy et al. with the OPLS-UA (united atom) force field in TIP4P water.[33]

Details on the hydration of the solutes may be garnered from inspection of the heterocycle−water rdfs. Integration of the first peaks in the H1−OW (hydrogen attached to nitrogen in the heterocycles to water oxygen) rdfs to the minima at 2.5 Å indicates that pyrrole and imidazole donate an average of 0.94 and 1.04 hydrogen bonds to water. The other nitrogen in imidazole is also a hydrogen-bond acceptor site, and integration of the N−HW rdf to its minimum at 2.6 Å gives a coordination number of 0.99. Thus, the present simulations provide reasonable totals of 1 and 2 hydrogen bonds for pyrrole and imidazole, respectively. These results are similar to those reported by Nagy et al. (0.8 and 1.8).[33]

Conclusions

The general approach to developing force field parameters for heterocycles, which was initially explored in a study of pyridine and the diazenes, has been extended to include five-membered nitrogen- and oxygen-containing heterocycles. The approach features the use of standard OPLS-AA Lennard-Jones parameters and partial atomic charges obtained by the CHELPG procedure at the RHF/6-31G* level. Results presented here covered gas-phase structures, heterocycle−water complexes, heterocycle dimers, pure liquid properties, and free energies of hydration.

Although the results are generally good, the errors in the heats of vaporization for the oxazoles are larger than normal with the OPLS-AA force field. This does not reflect significant errors in the heterocycle−water interaction energies; the results from the force field are uniformly similar to those calculated at the ab initio RHF/6-31G* level (Table 4). Though the oxazole−water interaction energies are 0.4 kcal mol^{-1} more attractive with the force field, the pyrrole−water interaction is 0.5 kcal mol^{-1} too strong, but the computed properties for liquid pyrrole are fine. The computed difference in the free energies of hydration for pyrrole and imidazole is also in notably good agreement with experiment. Other potential sources of error lie in the details of the charge distributions and the simplicity of the point-charge model. Force field parametrization for amines and other nitrogenous compounds has generally proven to be challenging.[34,35] With this knowledge, and the expectation that the present difficulties might be due to the presence of two heteroatoms with lone pairs, it was decided to test EPS charges obtained with the more flexible 6-31+G* basis set. The pure liquid simulations for isoxazole and oxazole were repeated with 6-31+G* CHELPG charges. However, the results showed no improvement over those from the original calculations with the 6-31G* charges.

Consequently, the simplicity of the atom-centered charge model is suspect. As recently noted by Dixon and Kollman,[36] it can be advantageous to move the partial charges off of the Lennard-Jones sites for the heteroatoms in modeling heteroatomic compounds. They report increased accuracy in reproducing conformational energetics and the complexation energies and geometries with water. TIP4P water is, of course, another example of the utility of this approach.[24] In at least one case, they also found it more useful to simply scale the EPS charges than to add an off-atom charge site. In the present case, scaling the charges for isoxazole and oxazole by a factor of 0.8 brings the heats of vaporization and the densities to within 1.5 and 1.0%, respectively, of the experimental values.

In summary, the use of OPLS-AA Lennard-Jones parameters and 6-31G* EPS charges for the nonbonded parameters of heterocycles has now been explored for 10 molecules. The results for gas-phase complexes with water, the pure liquids, and free energies of hydration are generally impressive. Considering the simplicity of the point-charge model, some deterioration in the quality of the results can be expected as the percentage of heteroatoms in the molecules increases. This tendency appears to be reflected in the ca. 15% overestimates of the heats of vaporization for oxazole and isoxazole.

Acknowledgment. Gratitude is expressed to the National Institute of General Medical Sciences and the Office of Naval Research for support of this work.

References and Notes

(1) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.
(2) Gunsteren, W. F. v.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992.
(3) Balbes, L. M.; Mascarella, S. W.; Boyd, D. B. *A Perspective of Modern Methods In Computer-Aided Drug Design*; VCH Publishers: New York, 1994; Vol. V.
(4) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(5) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810.
(6) Jorgensen, W. L.; McDonald, N. A. *J. Mol. Struct. (THEOCHEM)* **1998**, *424*, 145.
(7) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
(8) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(9) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J. Comput. Chem.* **1993**, *14*, 1240.
(10) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. *J. Phys. Org. Chem.* **1997**, *10*, 563.
(11) McDonald, N. A.; Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1998**, *120*, 5104.
(12) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1994**, *116*, 6337.
(13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, Revision B.2; Gaussian Inc.: Pittsburgh, PA, 1995.
(14) Jorgensen, W. L. *BOSS*, Version 3.6; Yale University: New Haven, CT, 1995.
(15) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638.
(16) Wolfenden, R.; Liang, Y.-L.; Matthews, M.; Williams, R. *J. Am. Chem. Soc.* **1987**, *109*, 463.

(17) Stiefvater, O. L. *J. Chem. Phys.* **1975**, *63*, 2560.
(18) Nygaard, L.; Christen, D.; Nielsen, J. T.; Pedersen, E. J.; Snerling, O.; Vestergaard, E.; Sorensen, G. O. *J. Mol. Struct.* **1974**, *22*, 401.
(19) Bak, B.; Christensen, D.; Dixon, W. B.; Hansen-Nygaard, L.; Andersen, J. R.; Schottlander, M. *J. Mol. Spectrosc.* **1962**, *9*, 124.
(20) Christen, D.; Griffeths, J. H.; Sheridan, J. *Z. Naturforsch.* **1981**, *36a*, 1378.
(21) Kumar, A.; Sheridan, J.; Stiefvater, O. L. *Z. Naturforsch.* **1978**, *33a*, 145.
(22) Nygaard, L.; Nielsen, J. T.; Kirchheiner, J.; Maltesen, G.; Rastrup-andersen, J.; Sorensen, G. O. *J. Mol. Struct.* **1969**, *3*, 491.
(23) McClellan, A. L. *Tables of Experimental Dipole Moments;* Rahara Enterprises: El Cerrito, 1974, Vol. 2.
(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(25) Houriet, R.; Schwartz, H.; Zummack, W. *Angew. Chem., Int. Ed. Engl.* **1980**, *19*, 905.
(26) Meot-Ner, M.; Liebman, J. F.; Bene, J. E. D. *J. Org. Chem.* **1986**, *51*, 1105.
(27) Suzuki, S.; Green, P. G.; Bumgarner, R. E.; Dasgupta, S.; Goddard, W. A. G., III; Blake, G. A. *Science* **1992**, *257*, 942.
(28) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768.
(29) Scott, D. W.; Berg, W. T.; Hossenlopp, I. A.; Hubbard, W. N.; Messerly, J. F.; Todd, S. S.; Douslin, D. R.; McCullough, J. P.; Waddington, G. *J. Phys. Chem.* **1967**, *71*, 2263.
(30) Majer, V.; Svoboda, V. *Enthalpies of Vaporization of Organic Compounds*; IUPAC Chemical Data Series No. 32; Blackwell: Oxford, U.K., 1985.
(31) McCormick, D. G.; Hamilton, W. S. *J. Chem. Thermodyn.* **1978**, *10*, 275.
(32) Timmermans, J. *Physico-Chemical Constants of Pure Organic Compounds*; Elsevier: Amsterdam, 1965; Vol. 2.
(33) Nagy, P. I.; Durant, G. J.; Smith, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 2912.
(34) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775. Ding, Y.; Bernardo, D. N.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1995**, *99*, 11575. Meng, E. C.; Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 2367.
(35) Morgantini, P.-Y.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 6057.
(36) Dixon, R. W.; Kollman, P. A. *J. Comput. Chem.* **1997**, *18*, 1632.