Correlation and Solvation Effects on Heterocyclic Equilibria in Aqueous Solution

Christopher J. Cramer* and Donald G. Truhlar*

Contribution from the Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455-0431

Received March 29, 1993®

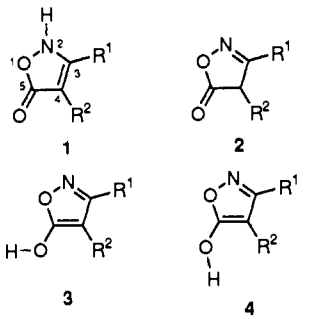
Abstract: We report extended-basis-set electronic structure calculations with high levels of electron correlation for heterocyclic tautomerizations, augmented by a detailed analysis of the aqueous solvation free energy differences which includes both electronic and geometric relaxation in aqueous solution, according to the Austin Model 1–Solvation Model 2 (AM1–SM2). The equilibria used as test cases are the competition between the hydroxy and oxo forms of 5-hydroxyisoxazoles; these involve two oxo (keto) and two hydroxy (enol) forms. For the unsubstituted parent system, it is found that the energy differences between the two oxo forms and between oxo and hydroxy forms are both very sensitive not only to extending the basis sets and including electron correlation but also to including electron correlation at levels higher than second order, indicating the difficulty of treating sp^2 and sp^3 centers on an equal footing in ring systems. We also find that treatments of electrostatic components of solvation free energies based on the popular Onsager model underestimate the solvation energy of the *syn*-hydroxy form because local bond moments have significant effects on the bulk electric polarization even when they largely cancel in the net dipole moment. Finally we note that there is a significant difference in first-hydration-shell effects for the oxo and hydroxy forms over and above that accounted for by electrostatic polarization. The effects of methyl substitution on the isoxazole ring are explored, and the calculated equilibrium shifts are consistent with available experimental data, which are thereby explained in terms of a combination of changes in both the gas-phase and solvation free energies.

1. Introduction

Solvation may have a dramatic effect on tautomeric equilibria, especially in heterocyclic systems.[1-5] For instance, the equilibrium constant for the 4-hydroxypyridine/4-pyridone tautomerization is changed more than one-million-fold on transfer from the gas phase to aqueous solution.[3] The significant changes in solvation energy upon tautomerization as well as the importance of aqueous heterocyclic equilibria in biochemistry makes such systems particularly interesting test cases for theoretical models of solvation.[4,5] For example, Woodcock and co-workers[5] recently presented a stimulating comparison of the performance of a variety of solvation models for the four tautomers (Chart I) 5(2*H*)-isoxazolone (**1a**), 5(4*H*)-isoxazolone (**2a**), *syn*-5-hydroxyisoxazole (**3a**), and *anti*-5-hydroxyisoxazole (**4a**) employing both *ab initio* techniques and semiempirical theory.

Experimental measurements of the aqueous equilibrium concentrations for the various tautomers have been accomplished for methylated homologs of this parent system.[6a,b] Thus, in the 3-methyl series (**1b–4b**), the aqueous equilibrium is about 70:30 **2b**:**1b**, with no detectable amounts of the two hydroxy isomers. For the 3,4-dimethyl series (**1d–4d**), only the NH tautomer **1d** is observed. For the 4-methyl series (**1c–4c**), aqueous data are not reported, although **1c** is the only observed tautomer in aqueous sulfuric acid. Apparently the relative instability of the parent isoxazole in aqueous solution has foiled experimental attempts to perform this measurement for **1a–4a**.[6c,d] Given the observed sensitivity to both solvent and substitution pattern, this heterocycle offers an excellent opportunity to explore in detail both the gas-phase and the solvation contributions to the overall equilibria.

Chart I



The purpose of this paper is to examine the equilibrium for **1a–4a** at a more definitive *ab initio* level and with more complete semiempirical[7,8] solvation models. This necessitates some reinterpretation of previous conclusions and yields a better appre-

* Correspondence may be addressed to either author.

® Abstract published in *Advance ACS Abstracts*, September 1, 1993.

(1) Katritzky, A. R. *Handbook of Heterocyclic Chemistry*; Pergamon Press: Oxford, England, 1985, pp 121–123.

(2) Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry*; VCH, New York, 1990; pp 98–104.

(3) Beak, P. *Acc. Chem. Res.* **1977**, *10*, 186.

(4) (a) Kwiatkowski, J. S.; Zielinski, T. J.; Rein, R. *Adv. Quantum Chem.* **1986**, *18*, 85. (b) Karelson, M. M.; Katritzky, A. R.; Szafran, M.; Zerner, M. C. *J. Org. Chem.* **1989**, *54*, 6030. (c) Karelson, M. M.; Katritzky, A. R.; Szafran, M.; Zerner, M. C. *J. Chem. Soc., Perkin Trans. 2* **1990**, 195. (d) Katritzky, A. R.; Karelson, M. M. *J. Am. Chem. Soc.* **1991**, *113*, 1561. (e) Rzepa, H. S.; Yi, M. Y.; Karelson, M. M.; Zerner, M. C. *J. Chem. Soc., Perkin Trans. 2* **1991**, 635. (f) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8552, 9901 (Additions and Corrections). (g) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. *J. Am. Chem. Soc.* **1992**, *114*, 1645. (h) Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1992**, *198*, 74; **1993**, *202*, 567 (Errata).

(5) Woodcock, S.; Green, D. V. S.; Vincent, M. A.; Hillier, I. H.; Guest, M. F.; Sherwood, P. *J. Chem. Soc., Perkin Trans. 2* **1992**, 2151.

(6) (a) Elguero, J.; Marzin, C.; Katritzky, A. R.; Linda, P. *The Tautomerism of Heterocycles*; Academic: New York, 1976; pp 300–303. (b) Katritzky, A. R.; Lagowski, J. M. *Adv. Heterocycl. Chem.* **1963**, *2*, 27. (c) Katritzky, A. R. Personal communication. (d) Two recent papers[4b,5] state that experiment establishes **1a** as the only observed tautomer in aqueous solution; both reports cite ref 1 of the present paper for this. However, as confirmed with the author,[6c] neither ref 1 nor the original source literature contains any data for unsubstituted **1a–4a** in aqueous solution. The most one can say is that the amount of NH isomer increases with increasing solvent polarity for the ring system in general.

0002-7863/93/1515-8810$04.00/0

ciation for the quantitative electron correlation effects on the relative tautomer energies and a better qualitative understanding of the nature of the solvation process, the magnitudes of the solvation energies, and the reliability of various alternative approaches[9-13] for modeling solvation energies.

In addition, we apply the semiempirical solvation models to the series of methylated homologs in order both to explore the nature of the substituent effects and to test the theoretical approach against the available experimental results.

1.1. Components of Solvation

Various effects may contribute to the free energy of solvation. One key property of a solvent is its ability to be electrically polarized, a first measure of which is provided by the bulk dielectric constant, which is 78.3 for water. Upon passing from the vacuum (or dilute gas-phase) dielectric constant of unity into solution, the structure and charge distribution within a solute will generally relax to permit greater charge separation; these effects increase with increasing dielectric constant and are referred to as solute polarization. Since it represents a distortion from the optimum gas-phase structure, solute polarization affects (raises) the internal free energy of the solute as well. Thus polarization effects on the solvent and solute should be treated self-consistently, since the latter involves work by the solvent to distort the solute and this partially cancels the gain in free energy due to more favorable interactions of the polarized subsystems. When the favorable *inter*molecular consequences of further polarization are overcome by the *intra*molecular costs of further distortion, relaxation is complete. These polarization interactions, being governed to a large extent (for neutral solutes) by partial charges and local dipole terms, operate over a long range.

The effects just described contribute a term to the free energy of solvation that we label ENP, for electronic, nuclear, and polarization. In particular, this term includes the change in the electronic and nuclear energies of the solute and the free energy of electric polarization of the bulk dielectric solvent.[7,8,14] The latter is dominated by the reorientational polarization of the solvent water molecules throughout the volume of the dielectric. In addition, there are terms which are more specifically associated with the first solvation shell. One example is the cavitation energy required to fit the solute into the solvent. There are also attractive dispersion forces between the solute and the nearby solvent molecules. Finally, there are local structural changes in the solvent as a result of the insertion of the solute— one key example is solute–solvent hydrogen bonding and another is the especially strong change in solvent–solvent hydrogen bonding in the first solvation shell. While the electrostatic component of hydrogen bonding may be included to some degree in the dielectric polarization term, it also has short-range components that cannot be accounted for in a uniform dielectric. We refer to the sum of these first-hydration-shell effects as the CDS terms, for cavitation, dispersion, and solvent-structural rearrangement.

Available models consider or ignore these types of contributions to solvation in various ways.

1.2. Solvation Modeling

Current efforts in solvation modeling in general follow one of two alternative approaches. The first involves the explicit consideration of hundreds or thousands of solvent molecules.[13,15-17] The supermolecular system consisting of these molecules plus the solute, when statistically sampled as a canonical ensemble, serves as a basis for simulations from which thermodynamic data related to solvation may be extracted. A molecular mechanics force field is generally employed, since the significant size of the system makes the analysis of forces and energies at the quantum mechanical level difficult and combinations[18] of quantal treatments of the solute with explicit water molecules representing the bulk have not yet been widely employed. The molecular mechanics approach almost always ignores the contribution of solute polarization to the ENP terms. Furthermore, although classically polarizable solvent models have been employed in specific simulations,[19] they are not yet general; therefore, electric polarization of the solvent does not include many-body effects on the solvent electronic polarization in most simulations. The explicit-water simulations do, however, include the CDS terms to the extent that they are well represented by the force field.

The main alternative to these simulation procedures replaces the explicit solvent molecules with a continuum having the appropriate bulk dielectric constant.[20] Having simplified the solvent, it is now possible to employ quantum mechanical approaches for the ENP relaxation of electronic and molecular structure in solution. Since the properties of the continuum solvent must represent an average over solvent configurations, such approaches are quantum-statistical models.

Incorporation of the electric field of the bulk dielectric into the electronic structure calculation of the solute yields the self-consistent-reaction-field (SCRF) approach; several methods have been suggested. The simplest is based on the Onsager[9] model of solute–solvent interactions. This truncates the solute charge distribution at the dipole term in the classical multipolar expansion; for neutral solutes, one thus includes only the dipole interaction with the continuum. That interaction has a simple analytic form if the solute is modeled as residing in a spherical cavity of the dielectric, as is usually done. The accuracy and applicability of this approach are somewhat restricted by the limitations of the spherical cavity assumption. Moreover, the simplification of the electronic distribution does not permit non-zero ENP contributions to the solvation free energy for neutral molecules whose dipole moments vanish as a result of symmetry. Nevertheless, the model is particularly simple to implement; it is available in standard programs,[11,12] and it is widely employed.

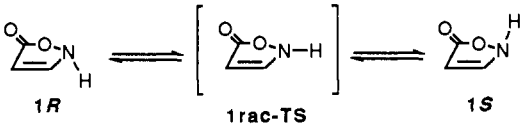
(7) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305, 9901 (Additions and Corrections).
(8) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.
(9) (a) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486. (b) Kirkwood, J. G. *J. Chem. Phys.* **1939**, *7*, 911. (c) Tapia, O.; Goscinski, O. *Mol. Phys.* **1975**, *29*, 1653.
(10) Miertius, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1990**, *55*, 117.
(11) Karelson, M.; Tamm, T.; Katritzky, A. R.; Cato, S. J.; Zerner, M. C. *Tetrahedron Comput. Methodol.* **1989**, *2*, 295.
(12) (a) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. *J. Chem. Phys.* **1991**, *95*, 8991. (b) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. *J. Am. Chem. Soc.* **1992**, *114*, 523. (c) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *GAUSSIAN92*; Gaussian, Inc.: Pittsburgh, PA, 1992.
(13) (a) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050. (b) Lybrand, T. P.; Ghosh, I.; McCammon, J. A. *J. Am. Chem. Soc.* **1985**, *107*, 7793. (c) Mezei, M.; Beveridge, D. L. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 1. (d) Bash, P.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564.
(14) (a) Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *13*, 1089. (b) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629.
(15) (a) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: London. 1987. (b) Haile, J. M. *Molecular Dynamics Simulation*; Wiley-Interscience: New York, 1992. (c) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley-Interscience: New York, 1991.
(16) (a) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184. (b) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Chem.* **1989**, *18*, 431.
(17) (a) Kollman, P. A.; Merz, K. M. *Acc. Chem. Res.* **1990**, *23*, 246. (b) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175.
(18) (a) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700. (b) Tapia, O.; Colonna, F.; Angyan, J. G. *J. Chim. Phys.* **1990**, *87*, 875. (c) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199. (d) Gao, J. *J. Phys. Chem.* **1992**, *96*, 537. (e) Gao, J.; Pavelites, J. J. *J. Am. Chem. Soc.* **1992**, *114*, 1912.
(19) (a) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647. (b) Dang, L. X.; Rice, J. E.; Caldwell, J.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 2481. (c) Xu, Y. W.; Wang, C. X.; Shi, Y. Y. *J. Comput. Chem.* **1992**, *13*, 1109. (d) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161.
(20) (a) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Viellard, A., Eds.; Reidel: Dordrecht, The Netherlands, 1980; Vol. 2, p 25. (b) Tomasi, J.; Alagona, G.; Bonaccorsi, R.; Ghio, C. In *Modelling of Structure and Properties of Molecules*; Maksic, Z. B., Ed.; Horwood: Chichester, England, 1987; p 330.

Generalizations of the Onsager model have appeared which extend the multipole series to arbitrarily high order.[10,21,22] Efficient optimization of solvated geometries motivates approaches that retain idealized cavities (e.g., ellipsoids) for the solutes,[21] but the formalism has been applied extensively with multipolar expansions fitted to completely arbitrary surfaces.[10,22]

Another way to represent the solute–solvent interaction is the generalized Born formalism,[7,8,14,23] which is based on a distributed monopole representation, i.e., distributed partial charges on the atoms, rather than a single-center multipole expansion. The atom-centered monopoles in principle generate all of the required multipoles to describe the electronic distribution. We note this only for comparative purposes, though, since calculation of the ENP terms by the generalized Born formalism does not actually involve the multipole moments explicitly.

The description of the quantum-statistical approach so far takes account of the ENP solvation terms but not of the CDS terms. To include both effects on a consistent footing, we developed the SMx aqueous solvation models,[7,8,14] of which the most successful are called AM1–SM2[8] and AM1–SM1a,[7] where SM denotes "solvation model". These models employ the Austin Model 1 (AM1)[24] semiempirical gas-phase Hamiltonian plus added terms representing solvation. The models are based on the neglect of diatomic differential overlap (NDDO)[25] level of electronic structure theory, which permits the treatment of even fairly large solutes quantum mechanically. They employ the generalized Born formalism to calculate the ENP terms, and we have added atomic parameters which account for the local CDS effects by assigning unique surface tensions to the solvent-accessible surface areas of various functional groups within the solute. The relationship between solvent-accessible surface area[26,27] and the energetics of such phenomena as cavitation, dispersion, and hydrogen bonding has been noted before in a variety of contexts.[27,28] The solvation model parameters are fit against experimental[29] free energies of solvation. The models use a realistic molecular-shaped cavity for both the ENP and CDS terms, and they specifically calculate the *absolute* free energy of solvation—a quantity not easily obtained with explicit-water simulation approaches. The development and performance of the SM$_x$ models have been recently reviewed.[14b] The mean unsigned errors in predicted free energies of solvation are about 0.7 kcal/mol for a data set of 150 neutral solutes which spans a wide variety of functionalities.

Chart II



2. Results and Discussion

All free energies in this paper correspond to 298 K. For absolute free energies, the standard state is taken to correspond to a concentration of 1 mol/L in both the gas phase and solution. One way to determine the relative free energies of two species in solution is to determine relative free energies in the gas phase and to add relative free energies of solvation. This was the approach followed by Woodcock et al.,[5] and it is also followed here for solvation models 2 and 1a. In addition, as mentioned above, these models yield absolute free energies of solvation, which are of interest in their own right and also help one to judge the reliability of the approach. Section 2.1 presents large-basis-set calculations of the gas-phase free energies, Section 2.2 discusses the absolute and relative solvation free energies, and Section 3 combines these results to make predictions of free energies in solution.

2.1. Gas-Phase Free Energies

Parent Compounds. We employ both Hartree–Fock (HF) and correlated (e.g., Møller-Plesset perturbation-theory[30] of order n, MPn[31]) methods for electronic structure calculations of the gas-phase energetics. Ideally, one would raise the level of theory until the energies and calculated geometries converge to stable values. Of course, computational practicality may prevent one from achieving that ideal. Thus, Woodcock et al.[5] calculated geometries and vibrational frequencies at the 3-21G[31] level in their study, and they calculated energies at the MP4/6-31G**[31] level. In order to determine the adequacy of such calculations, we have extended them to much larger basis sets with different levels of correlation. These *ab initio* calculations were carried out with the GAUSSIAN92[12c] computer program.

In the process of verifying the results of Woodcock et al.,[5] we discovered that, in spite of their methodological statement that all structures had been verified as local minima by calculation of analytical force constants, in fact the structure they employed for **1** was the C_s planar transition state for inversion at nitrogen (Chart II). The correct local minimum structure is considerably lower in energy (e.g. 4.5 kcal/mol at their employed level of HF/6-31G**//HF/3-21G), has a pyramidal nitrogen atom, and is of C_1 symmetry. The observation of Woodcock et al. that the relative energy of **1a** rose significantly on going to basis sets larger than 3-21G is a manifestation of the well-known tendency[32] for extended basis sets without d functions to underestimate inversion barriers at nitrogen.

We have employed the correlation-consistent polarized valence double- and triple-ζ basis sets of Dunning,[33] both with and without augmentation in the form of both diffuse functions (s and p on hydrogen, s, p, and d on heavy atoms) and functions of higher angular momentum (d on hydrogen, f on heavy atoms). The basis sets are abbreviated[33] cc-pVDZ, aug-cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ and correspond respectively to 99, 165, 222, and 345 contracted basis functions for the present system, as compared to 60 contracted basis functions for the 3-21G basis set. Rather than use the lengthy abbreviations for these basis sets, in the rest of this paper we refer to them simply as DB1,

(21) (a) Rivail, J. L. In *New Theoretical Concepts for Understanding Organic Reactions*; Bertrán, J., Czismadia, I. G., Eds.; Luwer: Dordrecht, The Netherlands, 1989; pp 219–230. (b) Pappalardo, R. R.;. Sánchez-Marcos, E.; Ruiz-López, M. F.; Rinaldi, D.; Rivail, J. L. *J. Phys. Org. Chem.* **1991**, *4*, 148. (c) Rinaldi, D.; Rivail, J. L.; Rguini, N. *J. Comput. Chem.* **1992**, *13*, 675.

(22) (a) Tomasi, J.; Bonaccorsi, R.; Cammi, R.; Olivares del Valle, F. J. *THEOCHEM* **1991**, *234*, 401. (b) Negre, M.; Orozco, M.; Luque, F. J. *Chem. Phys. Lett.* **1992**, *196*, 27. (c) Wang, B.; Ford, G. P. *J. Chem. Phys.* **1992**, *97*, 4162. (d) Chudinov, G. E.; Napolov, D. V. *Chem. Phys. Lett.* **1993**, *201*, 250.

(23) (a) Hoijtink, G. J.; de Boer, E.; van der Meij, P. H.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487. (b) Paradejordi, F. *Cah. Phys.* **1963**, *17*, 393. (c) Jano, I. *C. R. Hebd. Seances Acad. Sci.* **1965**, *261*, 103. (d) Costanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1. (e) Kozaki, T.; Morihasi, M.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547. (f) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164. (g) Still, W. C.; Tempczak, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(24) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(25) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.

(26) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(27) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.

(28) (a) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199. (b) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.

(29) (a) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 287. (b) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563. (c) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.

(30) (a) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618. (b) Goldstone, J. *Proc. R. Soc. London, A* **1957**, *239*, 267.

(31) For explanations of the MP2, MP3, and MP4 methods of treating electron correlation in electronic structure calculations and of the split-valence basis sets and standard notation, see: Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986. We note in particular that $x/y//z/w$ denotes a calculation with method x and basis y at a geometry optimized with method z and basis w.

(32) Rauk, A.; Allen, L. C.; Clementi, E. *J. Chem. Phys.* **1970**, *52*, 4133.

(33) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

Table I. Relative Gas-Phase Energies (in kcal/mol)

level of theory	1a	2a	3a	4a
Electronic Energy, Including Nuclear Repulsion				
a. At the MP2/DB1 Optimized Geometry				
HF/DB2	4.0	0.0	6.5	8.4
HF/DB3	3.3	0.0	5.7	7.7
HF/DB4	3.1	0.0	5.8	7.7
MP2/DB1	6.4	0.0	3.5	4.8
MP2/DB2	5.055	0.0	3.1	4.3
MP2/DB3	4.0	0.0	1.3	2.6
MP2/DB4	3.6	0.0	1.3	2.6
MP3/DB1	7.1	0.0	4.5	5.9
MP3/DB2	6.1	0.0	4.7	6.0
MP4/DB1	8.0	0.0	7.0	8.4
MP4/DB2	6.6	0.0	6.5	7.9
CCSD/DB1	7.9	0.0	6.8	8.4
CCSD(T)/DB1	7.9	0.0	6.7	8.2
CCSD/DB2	6.8	0.0	6.8	6.9
b. Optimized at the Same Level				
AM1[a]	9.8	0.0	12.7	15.0
HF/3-21G[b]	1.1	0.0	4.1	7.6
HF/DB1	4.9	0.0	6.4	8.3
HF/DB3	4.0	0.0		
MP2/DB2	5.060	0.0		
composite:[c]	5.3	0.0	5.0	5.2
Free Energy, $G°_{298}$(gas)[d]				
HF/DB1[e]	5.8	0.0	7.0	8.9
composite[c]	6.2	0.0	5.6	5.8

[a] The gas-phase AM1 energies are not used for our final calculations but are presented for comparison purposes. [b] Value for planar nitrogen inversion transition state of **1** is −317.661 89 au (2.6 kcal/mol relative). [c] See text. [d] Calculated using the rigid-rotor/harmonic-oscillator approximation. [e] Calculated at the HF/DB1 geometry.

DB2, DB3, and DB4, respectively (denoting "Dunning bases 1–4"); this is sufficient for our ends, since the main point of our use of these basis sets is to proceed along a well-defined series to demonstrate or learn about convergence or near convergence of the calculated quantities. We note for purposes of placing the discussion in context that the DB1 basis set is a full "polarized double-ζ" basis and is already more complete than the widely used, double-ζ-partially-polarized 6-31G*[31] basis set. The further extensions in bases DB2–DB4 produce what may be called state-of-the-art large basis sets.

Geometries were optimized for all four tautomers at the HF/DB1 and MP2/DB1 levels. Calculations with larger basis sets or higher-order treatment of correlation employed the MP2/DB1 geometries except for a few calculations on **1a** and **2a** where optimizations were carried out with larger basis sets as discussed below. Frequency calculations at the HF/DB1 level established all structures as local minima. Table I lists the relative electronic energies for **1a–4a** calculated at various levels, typically rounded to the nearest tenth of a kcal/mol but showing extra digits in a few cases where relevant to the discussion of convergence. Absolute energies of all *ab initio* calculations are given in the supplementary material.

It is evident that convergence in the relative HF energies is not reached until the DB3 basis set; augmentation to the DB4 set changes the relative energies by only one or two tenths of a kcal/mol. For our biggest basis set, the relative energies in kcal/mol at the Hartree–Fock level are as follows: **1a**, 3.1; **2a**, 0.0; **3a**, 5.8; and **4a**, 7.7.

It is well established that improvement of basis sets beyond the level required for Hartree–Fock convergence is required to accurately capture all of the correlation energy.[34,35] Consistent with this, Table I shows that for the present cases the second-order perturbation theory treatment of electron correlation does indeed converge more slowly than Hartree–Fock calculations with respect to increasing the basis set. For the largest basis set at the second order perturbation theory level, the relative energies in kcal/mol are as follows: **1a**, 3.6; **2a**, 0.0; **3a**, 1.3; and **4a**, 2.6.

To assess the importance of higher-order terms in the perturbation expansion, we have additionally performed MP3 and MP4[31,36,37] calculations with the DB1 and DB2 basis sets. It is apparent that higher-order effects are not negligible. Moreover, the MP*n* energies do not appear to have converged with respect to increasing *n*. This is not surprising, since conjugated π systems have been observed to exhibit slow convergence with many-body perturbation theory treatments.[34,38] We therefore also carried out calculations by the more accurate coupled-cluster[12c,37,39] treatment of electron correlation, in particular using the coupled-cluster single-and-double-excitation method, CCSD, and the coupled-cluster method with all single and double excitations and perturbative triple excitations,[40] CCSD(T). The calculations at the former level were carried out with the DB1 and DB2 bases, and the latter were carried out with the DB1. Comparison of the two sets of DB1 calculations indicates that the unlinked triples are of relatively minor importance, so the lower CCSD level of theory combined with the larger DB2 basis is probably the most reliable of our coupled-cluster calculations.

Although Table I shows encouraging agreement between the MP4 and CCSD relative energies, it is clear that the present equilibria appear to be particularly sensitive to basis set effects and higher levels of electron correlation, evidently because of the difficulties of treating the sp^2 and sp^3 centers on an equal footing. Some apparent trends are worth commenting on explicitly. First, increasing the size of the basis set lowers the energies of **1a**, **3a**, and **4a** relative to **2a** at both the Hartree–Fock and correlated levels. For every isomer, the drop in the second-order correlated energy on going from the polarized double-ζ to the largest basis set is more than 2 kcal/mol. Higher orders of correlation appear to be working in the opposite direction to these trends. Relative to **2a**, every isomer increases in energy with increasing level of treatment of electron correlation; for **1a** this effect is only roughly half as large as those for **3a** and **4a**.

Since aqueous equilibria for this heterocycle seem to be dominated by tautomers **1** and **2**, we have examined the geometries of these two isomers at higher levels. To investigate the effect of going from double-ζ to triple-ζ in basis set quality, we compared geometries optimized at the Hartree–Fock level with the DB1 and DB3 basis sets. The largest change observed was for the proton bound to nitrogen in **1a**, where the bond length decreased by 0.008 Å. In general, all other bond lengths changed by less than 0.005 Å and bond angles by less than 0.5°. Moreover, the energy of **1a** drops by only 0.04 kcal/mol relative to that of **2a** on going from HF/DB3//HF/DB1[31] to HF/DB3//HF/DB3. We also compared geometries optimized at the MP2/DB2 level to those optimized at the MP2/DB1 level. Changes in ring bond lengths of up to 0.016 Å were observed, but here the relative energy change of **1a** and **2a** is only 0.005 kcal/mol. Finally, we optimized the location of the NH proton in **1a** at the CCSD/DB1 level keeping all other coordinates fixed. The only notable effect was to decrease the out-of-plane angle by 1.0°, decreasing the absolute energy by only 0.01 kcal/mol. It is apparent, then, that higher-level geometry optimizations have only marginal effects

(34) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Macmillan: New York, 1982.
(35) Feller, D.; Davidson, E. R. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1990; Vol. 1, p 373.
(36) (a) Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quantum. Chem., Symp.* **1977**, *11*, 149. (b) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum. Chem., Symp.* **1979**, *13*, 325. (c) Krishnan, R.; Pople, J. A. *Int. J. Quantum. Chem.* **1978**, *14*, 91.
(37) (a) Bartlett, R. J.; Purvis, G. D., III. *Int. J. Quantum Chem.* **1978**, *14*, 561. (b) Bartlett, R. J.; Purvis, G. D., III. *Phys. Scr.* **1980**, *21*, 255.
(38) (a) Nobes, R. H.; Pople, J. A.; Radom, L.; Handy, N. C.; Knowles, P. J. *Chem. Phys. Lett.* **1987**, *138*, 481. (b) Falvey, D. E.; Cramer, C. J. *Tetrahedron Lett.* **1992**, *33*, 1705.
(39) (a) Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35. (b) Purvis, G. D., III; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910. (c) Scuseria, G. E.; Schaefer, H. F., III. *J. Chem. Phys.* **1989**, *90*, 3700. (d) Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 1697. (e) Lee, T. J.; Rendell, A. P.; Taylor, P. R. *J. Phys. Chem.* **1990**, *94*, 5463.
(40) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

Table II. Relative AM1 Gas-Phase Energies (in kcal/mol)

substitution pattern	1	2	3	4
	Relative to **2**			
a	9.8	0.0	12.7	15.0
b	8.7	0.0	12.8	15.2
c	6.3	0.0	9.8	11.8
d	5.2	0.0	10.1	12.1
	Effect of Methylation			
a	0.0	0.0	0.0	0.0
b	−1.1	0.0	0.1	0.2
c	−3.5	0.0	−2.9	−3.2
d	−4.6	0.0	−2.6	−2.9

on structure and essentially negligible effects on the relative energies of **1a** and **2a**.

Having examined all these convergence issues, it seems reasonable to use as our best estimate for higher-order corrections beyond the MP2 level the difference between the CCSD/DB2 and MP2/DB2 calculations and to add these differences to the MP2/DB4 results to get a composite estimate of the simultaneous effects of higher-order correlation and extended basis sets. Doing this, we get the values listed as "composite" in Table I. Frequency calculations at the HF/DB1 level were used to estimate thermal enthalpic and entropic vibrational contributions to the gas-phase free energies at 298 K. Adding the unscaled thermal vibrational contributions as well as thermal rotational contributions to the composite energies, we arrive at the following relative gaseous free energies, $G°_{298}$(gas), in kcal/mol: **1a**, 6.2; **2a**, 0.0; **3a**, 5.6; and **4a**, 5.8.

We note that in Table 2 of Woodcock et al.[5] the AM1 gas-phase energies for isomers **3a** and **4a** (numbered **3** and **3a** in that work) have been interchanged. The literature is disturbingly filled with errors for this heterocycle, due in part to initial reports[4c,e] for gas-phase and solvated species for which geometry optimizations were inadequate. The correct values are provided in Table I.

The geometries for **1a–4a** at the highest levels of theory at which they were calculated are provided in the supplementary material.

Methyl Derivatives. We note the rather poor comparison of AM1 results to the correlated *ab initio* predictions for the parent compounds in Table I. Most apparent is that **1a**, **3a**, and **4a** are much less stable relative to **2a** at the AM1 level. With this lesson in mind, the AM1 relative energies for the methylated series **b**, **c**, and **d** are not expected to be reliable in an absolute sense, but we will use them to learn more about the effect of methylation. These results are in Table II. Table II gives relative energies; the absolute energies are given in the supplementary material as an aid to reproducibility for later workers.

2.2. Free Energies of Solvation

We have calculated free energies of solvation for all sixteen structures **1a–4d** at the AM1–SM2[8] and AM1–SM1a[7] levels using the AMSOL program,[41] with the results presented in Table III. These calculations were performed under the constraints of permitting neither the electronic distribution nor the geometry to relax when the solute is placed in water (called NOPOL), permitting only electronic relaxation (called 1SCF), and without constraints, permitting full relaxation (full). Solute polarization increases the free energies of solvation by 9–29%, which is typical[4h,14b] of similar systems. Also typical is the fact that electronic relaxation is much more important than nuclear (geometric) relaxation.

The AM1–SM1a model adopts an approach to the calculation of the CDS terms in the solvation free energy that is considerably different from that adopted in the AM1–SM2 model. In particular, it treats protons in an environmentally sensitive fashion by applying different surface tensions depending on the atoms to which they are attached. In addition, hybridization at nitrogen and oxygen is taken account of with unique surface tensions. As a result, the model is less general than AM1–SM2 (the user must input hybridization data and assign the proton attachments, which is not always unambiguous along reaction paths for instance), but it does have a smaller overall error of 0.6 kcal/mol for the aforementioned test set of 150 neutral molecules. (We note that the test set contained no isoxazoles but did contain five other unsaturated heterocycles and six saturated heterocycles.) The ENP energies are changed by only 0.0–0.3 kcal/mol between the AM1–SM2 and AM1–SM1a models. However, the CDS terms are more negative in every case by 1.5–3.5 kcal/mol. As shown in Table VI, the net effect in every case is to stabilize every isomer relative to **2** in aqueous solution. Because of the generality of the method, the AM1–SM2 results are emphasized in this section for discussion purposes, but the AM1–SM1a results are equally as likely to be accurate for the present problem, and we will consider both sets of results in Section 2.3 when we predict the experimental result for the parent system and compare to experiment for the methyl derivatives.

We first compare our results to the simple Onsager[9,11,12] approach. In previous work,[5] only this model gave relative solvation free energies which, when added to the previous[5] gas-phase results, predicted **1a** to be the predominant tautomer in aqueous solution. However, this result deserves further examination. Since **1a** has the largest gas-phase dipole moment of all of the tautomers, it is clear that it will have the largest electrostatic solvation free energy in an Onsager calculation. Moreover, the magnitude of that quantity, and its separation from the other isomers, will increase in inverse proportion to the cube of the radius of the spherical cavity employed in the Onsager formalism. Woodcock et al. used a spherical cavity of radius 2.5 Å for **1a–4a**. While they do not report what the electrostatic solvation energies are, we have repeated their calculations (reproducing their

Table III. Calculated Free Energies of Solvation and Components Thereof (in kcal/mol)[a]

substitution pattern	treatment of relaxation	terms	1	2	3	4
		Absolute, AM1–SM2				
a	NOPOL	ENP	−3.7	−2.7	−2.3	−2.5
		CDS	−4.7	−3.8	−6.4	−6.3
		total	−8.5	−6.4	−8.7	−8.8
a	1SCF	ENP	−5.8	−4.4	−3.1	−3.5
		CDS	−4.7	−3.8	−6.4	−6.3
		total	−10.5	−8.1	−9.5	−9.8
a	full	ENP	−6.2	−4.5	−3.2	−3.6
		CDS	−4.7	−3.7	−6.4	−6.3
		total	−11.0	−8.2	−9.5	−9.9
b	full	total	−9.3	−7.6	−9.0	−9.3
c	full	total	−9.8	−7.2	−8.8	−8.9
d	full	total	−8.6	−6.8	−8.4	−8.5
		Relative, AM1–SM2				
a	full	total	−2.7	0.0	−1.3	−1.7
b	full	total	−1.7	0.0	−1.4	−1.7
c	full	total	−2.5	0.0	−1.6	−1.7
d	full	total	−1.8	0.0	−1.6	−1.7
		Absolute, AM1–SM1a				
a	full	total	−14.0	−10.0	−11.6	−12.1
b	full	total	−12.6	−9.3	−11.1	−11.7
c	full	total	−12.9	−9.0	−10.8	−11.3
d	full	total	−11.9	−8.8	−10.5	−11.1
		Relative, AM1–SM1a				
a	full	total	−4.0	0.0	−1.6	−2.2
b	full	total	−3.3	0.0	−1.8	−2.4
c	full	total	−3.9	0.0	−1.8	−2.3
d	full	total	−3.1	0.0	−1.7	−2.3

[a] See Section 2.2 for explanation of the relaxation treatments and terms.

(41) Cramer, C. J.; Lynch, G. C.; Truhlar, D. G. *AMSOL*, version 3.0.1; Quantum Chemistry Program Exchange Program No. 606. *QCPE Bull.* **1993**, *13*, 9.

Table IV. SCRF/6-31G** Onsager-Reaction-Field Calculations of Free Energies of Solvation (in kcal/mol)[a,b]

r_{cavity} (Å)	1a	2a	3a	4a
2.5	−30.6[c]	−22.4	−3.0	−18.5
3.60–3.62[d]	−6.2	−5.0	−0.7	−4.3

[a] Electrostatic component of solvation free energies in kcal/mol. [b] HF/3-21G geometries are used. [c] Value for planar TS of **1a** is −38.0. Using r_{cavity} = 3.6 Å gives −7.6. [d] Calculated using the WWF isodensity surface; see ref 4g and text.

Table V. Comparison of Calculated Free Energies of Solvation (in kcal/mol)

	1a	2a	3a	4a
Absolute				
SCRF/6-31G**, Onsager model[a]	−6.2	−5.0	−0.7	−4.3
SCRF/6-31G**, Tomasi model[b]	−17.2	−16.3	−12.3	NA[c]
explicit water, AMBER[b]	NA	NA	NA	NA
AM1–SM2, ENP only	−6.2	−4.5	−3.2	−3.6
AM1–SM2, full	−11.0	−8.2	−9.5	−9.9
Relative to **2**				
SCRF/6-31G**, Onsager model[a]	−1.2	0.0	4.3	0.7
SCRF/6-31G**, Tomasi model[b]	−0.9	0.0	4.0	NA
explicit water, AMBER[b]	−2.1	0.0	1.8	0.3
AM1–SM2, ENP only	−1.7	0.0	1.3	0.9
AM1–SM2, full	−2.7	0.0	−1.3	−1.7

[a] r_{cavity} = 3.60–3.62 Å. [b] Taken from ref 5. [c] NA = not available.

reported dipole moments in solution), and we find that the Onsager solvation free energy for planar **1** is −38.0 kcal/mol, which is *very* large. For the pyramidal geometry, this value is still very large, −30.6 kcal/mol. Both values are far larger than experimental[29] free energies of solvation for similar molecules for which such values have been measured, and the discrepancy is even more remarkable when we consider that the Onsager model does not include CDS terms, which make a negative and non-negligible contribution (about −5 kcal/mol) for this hydrophilic molecule in the more complete AM1–SM2 calculations. The source of the problem is the choice of 2.5 Å for the cavity radius, which cannot be justified in any reasonable way.

From the geometries of **1a–4a**, it is clear that if one were to replace each atom with a sphere having the appropriate van der Waals radius,[42] a cavity radius of roughly 4 Å would be required to completely contain the molecule. An approach adopted by Wong, Wiberg, and Frisch (WWF)[4g] and shown to give good agreement with experiment in other heterocyclic systems is to calculate the molecular volume as 1.44 times that enclosed by the isodensity surface corresponding to 10^{-3} esu/a_0^3 [1a_0 = 0.529 Å] and choose the appropriate cavity radius as 0.5 Å greater than the value required to reproduce that volume. This is usually close to the van der Waals result, and indeed for **1a–4a** we employed[12c] the WWF prescription and obtained cavity radii of 3.60–3.62 Å. Table IV presents the results from this latter approach and compares to the previous[5] results.

Another approach to which we can compare is Tomasi's polarized continuum model (PCM),[10,20b,22a] using a multipolar expansion in a molecular cavity with a realistic molecular shape. Woodcock et al.[5] employed this approach to calculate the ENP contribution to the free energies of solvation for **1a–4a**, using the atomic radii suggested by Aguilar and Olivares del Valle.[43] These results are in Table V. It is also interesting to compare the predicted ENP values for the two continuum techniques employed previously to those from AM1–SM2, and this is also done in Table V. We find the reported absolute free energies of the Tomasi continuum method to be surprisingly large, again by comparison to experiment[29] for analogous molecules. It is not entirely clear why this should be so, although it must be pointed out that the ENP energies calculated with the PCM approach are quite sensitive to basis set.[20b]

(42) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(43) Aguilar, M. A.; Olivares del Valle, F. J. *Chem. Phys.* **1989**, *129*, 439.

For *relative* free energies, agreement between the three continuum polarization methods (Onsager with WWF prescription for the cavity radius, Tomasi model, and the ENP part of AM1–SM2) is qualitatively reasonable. With regard to *absolute* free energies of solvation, the Onsager energies from calculations using the WWF guidelines agree within 1 kcal/mol with AM1–SM2 ENP free energies for every isomer but **3a**, where the generalized Born model is presumed to be more accurate because it takes account of important higher-order multipoles. The general agreement of the two *ab initio* continuum models with the electrostatic part of the AM1–SM2 approach for the isomers with larger dipole moments is encouraging for future applications of AM1–SM2, which is in addition more valid for systems with canceling bond dipoles.[14b] Table III shows that the full AM1–SM2 model gives quite different results from the electrostatic only calculations, and again previous experience[14b] shows that inclusion of the cavity-dispersion-structural effects that are responsible for this difference improves the accuracy in comparison to experiment. We will base our final predictions of the solvation effects on the AM1–SM2 and AM1–SM1a models primarily for the two reasons discussed in this paragraph, i.e., these models include first-hydration shell effects rather than only electric polarization effects and the treatment of electric polarization includes higher-order multipoles implicitly through the distributed monopole charge distribution.

We note here that while the AM1–SM2 model was parametrized[8] to reproduce experimental free energies of solvation, the individual ENP and CDS terms are less quantitatively reliable in that systematic errors in either term (which might derive from the AM1 charges, for instance) may be balanced by a change in the parameters for the other, so as to deliver overall agreement with experiment. Thus, a comparison of ENP terms such as that given above is useful primarily for the qualitative insight it affords.

It is informative to compare unrelaxed, i.e., NOPOL,[41] calculations to explicit-solvent-molecule approaches, since neither permits solute polarization, but they both include the physical effects responsible for the AM1–SM2 CDS terms. It is interesting to note that the difference in solvation free energies for **1a** and **2a** predicted by NOPOL calculations, 2.1 kcal/mol, is in exact agreement with the reported molecular dynamics simulation results of Woodcock et al.[5] employing the AMBER force field;[44] see Tables III and V. [However, the incorrect (i.e., planar at N) geometry was used for **1a** in ref 5.][45] The AM1–SM2 NOPOL relative solvation free energy for isomers **1a** and **3a**, on the other hand, is in much poorer agreement with the explicit-water calculations (−0.2 vs 1.8 kcal/mol, respectively). It is unclear which model is more accurate, since the effect of the planar geometry at nitrogen on the simulation is not obvious.

In addition to the points discussed above with reference to **1a**, the differences between the various solvation models for the relative free energies of solvation of **2a** and **3a** are quite dramatic and deserve further analysis. We have noted above the importance of higher-order multipoles which contribute to a more sizable ENP energy for **3a** relative to **2a** than is found with the Onsager or Tomasi models. It is worth emphasizing that this effect is two-fold: the AM1–SM2 model, by recognizing the higher-order multipoles, gives not only a larger *unrelaxed* polarization contribution but also a larger *relaxation* contribution. An additional point of interest is the surprisingly large disagreement between the AM1–SM2 continuum model and the AMBER simulation results for the relative solvation free energies of **2a** and **3a** (−1.3 and +1.8, respectively). Here the AM1–SM2 model predicts a significantly more negative CDS solvation contribution for the hydroxylic **3a** than for **2a**, which has a somewhat less hydrophilic ketone moiety.

(44) Singh, U. C.; Wiener, P. K.; Caldwell, J. W.; Kollman, P. A. AMBER (UCSF) Version 3.0. Department of Pharmaceutical Chemistry, University of California, San Francisco, 1986.
(45) Gould, I. R.; Hillier, I. H. Submitted for publication. We thank Professor Hillier for a preprint of this manuscript.

Table VI. Calculated Relative Aqueous Free Energies (in kcal/mol)

substitution pattern	level	terms	1	2	3	4
Composite Gas-Phase plus AM1–SM2 Solvation Model						
a	full	total	3.5	0.0	4.3	4.1
b	full	total	3.4	0.0	4.3	4.3
c	full	total	0.2	0.0	1.1	0.9
d	full	total	−0.2	0.0	1.4	1.2
Composite Gas-Phase plus AM1/SM1a Solvation Model						
a	full	total	2.2	0.0	4.0	3.6
b	full	total	1.8	0.0	3.9	3.6
c	full	total	−1.2	0.0	0.9	0.3
d	full	total	−1.5	0.0	1.3	0.6

2.3. Relative Free Energies in Aqueous Solution: Prediction of the Dominant Isomers of the Equilibrium Mixture

Parent Compounds. When the AM1–SM2 free energies of solvation in the full/total row of Table III are added to our best *ab initio* estimate of the relative gas phase free energies for **1a–4a** in the bottom row of Table I, we find that the relative aqueous free energies in kcal/mol are as follows: **1a**, 3.5; **2a**, 0.0; **3a**, 4.3; **4a**, 4.1. These results are given in the first row of Table VI. (Other rows in this table are discussed below). In order to assess the trustworthiness of these values, we considered three possible sources of inaccuracy. In particular, it is possible that (1) our gas-phase correlation energies may be insufficiently converged with respect to further augmentation of the basis set, even for the very-large-scale calculations reported here, (2) the rigid-rotor, harmonic approximation may be inadequate for the rotational and vibrational partition functions used in the computation of the gas-phase free energies, and (3) the errors of the AM1–SM2 model may be in opposite directions for different isomers, leading to a larger than expected error in their relative energies. We next expand upon each of these points in turn.

Concerning point 1, higher-level geometry optimizations have only marginal effects on structure and essentially negligible effects on the relative energies of **1a** and **2a**. Moreover, it is clear from Table I that the relative MP2 energies of **2a–4a** show good convergence with respect to augmentation of the triple-ζ basis set; i.e., for the final affordable basis increase from DB3 to DB4, the relative energy of **1a** drops only 0.4 kcal/mol. If augmentation functions provide such an effect, it is possible that improving the quality of the basis in regions of greater electron density (e.g. quadruple-ζ) would further enhance this effect; however, it seems unlikely that more than another 0.5 kcal/mol in relative energy would be gained.

As for the rigid-rotor, harmonic approximation, we note first that we have employed unscaled Hartree–Fock frequencies in our thermochemical calculations. It is well-established that Hartree–Fock frequencies generally require scaling by some factor on the order of 0.9 in order to bring them into better agreement with experiment.[31] Given that **1a** is calculated to have a larger zero-point energy than **2a** by 0.9 kcal/mol, scaling the zero-point energy would move the calculated aqueous equilibrium by only 0.1 kcal/mol. As far as anharmonicity is concerned, the largest effects often occur for low-frequency modes. The calculated frequencies include one vibration between 180 and 325 cm^{-1} for each isomer, with the lowest belonging to **2a**. We anticipate therefore that relative energies would not be changed in a major way by a better treatment of vibrations.

With regard to point 3, in order to assess possible errors in the AM1–SM2 model, we now consider calculated free energies of solvation at the AM1–SM1a level. Although we usually prefer the use of AM1–SM2 over AM1–SM1a because of AM1–SM2's wider applicability, it is often useful to compare the two models given their different treatments of the CDS terms. When they are in reasonable agreement, one may take more confidence in the predicted results. If the SM1a solvation energies are added to the best gas-phase estimates, we obtain for the parent system the following relative aqueous free energies: **1a**, 2.2; **2a**, 0.0; **3a**, 4.0; **4a**, 3.6. These results are given in the second **a** row of Table VI. The table shows that the AM1–SM1a model predicts a moderate stabilization for **1a** as compared to the AM1–SM2 result.

Given that most of the probable corrections to the various factors we have proposed as contributing to the overall error seem to be small, we feel confident that the aqueous equilibrium for the parent system will be dominated by the CH tautomer **2a**.

Methyl Derivatives. To provide some comparison to systems which have been observed experimentally, we correct the AM1 gas-phase relative energies for the methylated species for all substitution patterns of **1–4** by the same amounts required to bring the parent system into agreement with our composite *ab initio* calculations. (This is easily accomplished by combining the results of the **a** rows in Table VI with **b** − **a**, **c** − **a**, and **d** − **a** relative solvation energy differences from Table III and with the effects of methylation in the bottom half of Table II.)

We observe in every case that the hydroxy tautomers **3** and **4** are too high in energy to be the primary or secondary tautomeric structure at equilibrium. Using the AM1–SM1a results, the equilibrium between **1** and **2** favors the CH structure **2b** in the 3-methyl **b** series (experimentally 70:30 **2b**:**1b**),[6a,b] favors the NH structure **1c** in the 4-methyl **c** series, and favors the NH structure **1d** in the 3,4-dimethyl **d** series (experimentally 100% **1d**).[6a,b] It is apparent that combining large-scale *ab initio* calculations for gas-phase free energies with free energies of solvation computed by the AM1–SM1a method reproduces the qualitative trend in substitution pattern nicely for the cases where experimental data are available, although small quantitative discrepancies remain (about 1.2 kcal/mol for the relative solvation energies of **1** and **2** on the basis of the results for series **b** and **d**). The ability to predict trends is very encouraging for the many other heterocyclic equilibria of interest in organic and biological chemistry. With the AM1–SM2 solvation energies, the trends are similarly reproduced, but the NH structure **1** is found to be about 1.3 kcal/mol less well solvated in every case.

2.4. Key Components Influencing the Aqueous Equilibria

By performing a detailed analysis of both the relative gas-phase and solvation free energies, we are able to identify a number of general, important factors which affect the four different equilibria under discussion.

First, although the hydroxy tautomers **3** and **4** are not predicted to play a major role in any of the aqueous equilibria, it is noteworthy that they have very similar solvation free energies in every case; this is shown in Table VII. It is especially interesting that the difference in the ENP terms for these isomers favors **4** only slightly in each series (0.1–0.4 kcal/mol, a difference of 3–13%). Although similar solvation energies of **3** and **4** are reasonable given that they differ merely as hydroxyl rotamers, this is not at all what would be expected from the Onsager approach on the basis of the corresponding dipole moments for these isomers, as has already been noted in series **a**. Modeling the ENP term purely as a function of the dipole moment would predict the ratio of the ENP terms to vary as the square of the ratio of the dipole moments—a factor of 7–9 in these instances. However, our distributed monopole treatment of electric polarization effects, which implicitly sums the multipole series to infinite order, predicts ENP contributions that agree within 16%. Since the solvent accessible surface areas for the different functional groups in these isomers are also minimally affected by the hydroxyl rotation, they are found to have similar net free energies of solvation. As expected, the AM1–SM1a model is more sensitive to the rotameric CDS differences, since these differences derive almost entirely from the hydroxyl hydrogen locations and hydrogens are not treated explicitly in the CDS terms in AM1–SM2.

A more general observation is that the major effect contributing to the equilibrium predominance of the NH tautomer **1** by

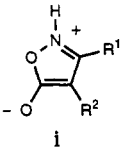
Table VII. Calculated Dipole Moments (in D) and Free Energies of Solvation and Components Thereof (in kcal/mol) for **3** and **4**[a]

substitution pattern	terms	3	4
	Absolute, AM1–SM2		
a	ENP	−3.2	−3.6
	CDS	−6.4	−6.3
	total	−9.5	−9.9
b	ENP	−3.1	−3.4
	CDS	−5.9	−5.9
	total	−9.0	−9.3
c	ENP	−3.0	−3.1
	CDS	−5.8	−5.8
	total	−8.9	−8.9
d	ENP	−3.0	−3.2
	CDS	−5.4	−5.4
	total	−8.4	−8.5
	Absolute, AM1–SM1a		
a	ENP	−3.2	−3.7
	CDS	−8.3	−8.5
	total	−11.6	−12.1
b	ENP	−3.1	−3.6
	CDS	−8.0	−8.1
	total	−11.1	−11.7
c	ENP	−3.1	−3.1
	CDS	−7.7	−8.2
	total	−10.8	−11.3
d	ENP	−3.2	−3.2
	CDS	−7.3	−7.9
	total	−10.5	−11.1
	Gas-Phase AM1 Dipole Moment		
a		1.31	3.89
b		1.31	3.97
c		1.52	4.07
d		1.53	4.13

[a] See Section 2.2 for explanation of the relaxation treatments and terms.

comparison to **2** in series **c** and **d** is a gas-phase effect. In particular, each of these two cases involves methylation at C(4). For every isomer but **2**, this methylation occurs at an sp^2 center, and the well-known hyperconjugative stabilization afforded more substituted double bonds lowers these isomers in relative energy by comparison to **2**, where the methylation occurs at an sp^3 center. By comparison of the **a** − **c** and **b** − **d** differences in Table II, it is apparent that this effect is worth about 3.5 kcal/mol in favor of isomer **1**.

There is a trend in the relative solvation free energies for **1** vs **2** which is also particularly interesting. From Table III, it is apparent that, with either solvation model 1a or 2, isomer **1** is better solvated than **2** by slightly less than 1 kcal/mol for series **a** and **c** in comparison to series **b** and **d**. The difference is almost entirely a result of variations in the ENP terms, the CDS terms being consistently in the range of 0.8–1.0 kcal/mol in favor of **1**. For series **b** and **d**, the SM2 NOPOL difference in ENP terms is 0.6 kcal/mol in favor of **1**, and this increases in each case with full relaxation to 0.9 kcal/mol. For series **a** and **c**, on the other hand, the difference starts out larger, at 1.0 kcal/mol in each case; moreover, relaxation plays a larger role, with the final differences being 1.7 and 1.5 kcal/mol, respectively. A careful analysis suggests that this is a steric effect of methylation at C(3).



i

The reason isomer **1** is particularly well-solvated in aqueous solution is that it is a vinylogous carbamate, and it thus enjoys some character of the zwitterionic canonical resonance structure i. This resonance structure obviously becomes more important in solvents of increasing polarity, like water. However, unlike the noncharged resonance contributor we have used to represent **1** up until now, i is planar at nitrogen. This results in increased steric interactions between the attached hydrogen and any substituent at C(3), suggesting that when the 3-position is methylated, **1** will be less able to take advantage of the charge separation implicit in i. Geometrical analysis of the optimized solvated structures bears out this analysis. For the non-C(3)-methylated series **a** and **c**, the out-of-plane angle of the hydrogen is less (i.e., closer to nitrogen planarity) at 49.6 and 50.5°, respectively. For the C(3)-methylated series **b** and **d**, it is 52.1° and 52.4°, respectively. Additionally, the C–N bond is shorter (i.e., more double bond character) for the former pair, 1.424 and 1.422 Å, respectively, than for the latter, 1.443 and 1.444 Å, respectively.

In summary, a detailed decomposition of trends either between series or within a single series finds important contributions from both the gas-phase and solvation free energies. The AM1–SM2 and AM1–SM1a models provide additional information with regard to the relative importance of electrostatic and more localized (i.e., first-hydration-shell) solvation effects. On the basis of these investigations, it appears that quantitative predictions of heterocyclic equilibria can be reasonably achieved neither in the absence of high-quality gas-phase calculations nor with the use of overly simplified solvation models employing multipole expansions truncated after only the dipole term.

3. Concluding Remarks

Large-scale electronic structure calculations of heterocycles show that even energy differences among tautomers are slowly convergent with respect to basis set and also require calculations beyond second-order perturbation theory for the treatment of electron correlation. The extended-basis-set electronic structure calculations show that the energy differences between the two oxo forms and between oxo and hydroxy forms are very sensitive to extending the basis sets beyond the polarized double-ζ level, presumably because of the difficulty of treating sp^2 and sp^3 nitrogen and oxygen ring sites equally accurately. Given the biological importance of such heterocycles as the nucleic acids, this has wide-ranging implications. Accurate modeling of standard and nonstandard base pairing, the relative energies of hydrogen-bonded and non-hydrogen-bonded binding situations, triple-helix formation, etc. may prove to require very-high-quality basis sets and very-high-level treatments of correlation.

In addition we have studied aqueous solvation effects on heterocyclic tautomers. Our final estimates of solvation energies are based on the AM1–SM2 and AM1–SM1a models, which have the advantages that they include both first-hydration-shell effects and local bond moments, are readily applicable to large solutes, and are not restricted by the (popular) spherical cavity assumption. The present comparisons show clearly the importance of including both bulk electric polarization and first-hydration-shell effects in solvation calculations; the latter increase the absolute solvation free energies by factors of 1.8–3.6. We also find that treatments of electrostatic components of solvation free energies based on the popular Onsager model underestimate the solvation energy of structures in which local bond moments partly cancel; e.g., the local bond moments of the *syn*-hydroxy form **3a** have significant effects on the bulk electric polarization even though they largely cancel in the net dipole moment, another finding of wide general importance.

Acknowledgment. We thank Alan Katritzky for confirming the experimental situation and Vasilios Melissas for providing the Dunning basis sets in a machine-readable form. This work was supported in part by the National Science Foundation.

Supplementary Material Available: Geometries and absolute energies for structures **1a**–**4a** and AM1 calculated heats of formation for all tautomers of **1**–**4** (4 pages). Ordering information is given on any current masthead page.