A Theoretical Study of Tautomerism in the Gas Phase and Aqueous Solution: A Combined Use of "State-of-the-Art" ab Initio Quantum Mechanics and Free Energy Perturbation Methods

Piotr Cieplak,[†] Paul Bash,[‡] U. Chandra Singh,[⊥] and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143. *Received December 22, 1986*

Abstract: The combination of ab initio quantum mechanical calculations together with the free energy perturbation method implemented in molecular dynamics simulations gives a good description of tautomeric equilibria for 2-oxopyridine, 2-oxopyrimidine, and cytosine tautomers in water solution. Such an approach appears to be general and applicable to most systems where equilibrium constants and free energy differences of solvation between structural isomers are needed. The strength of this method is further demonstrated by showing that one can "close the thermodynamic cycle" by simulating transitions from one species to another going in different directions. In the case of gas-phase tautomeric energy differences, one still has uncertainties of the order of 1–2 kcal/mol at the highest level of ab initio theory used by us (6-31G* geometry optimization with energies at the MP2/6-31G* level) applied to the 2-oxopyridine tautomerism. Uncertainties of ~1 kcal/mol are found in the solvation contribution to free energy differences. Nonetheless, this work demonstrates the power of combining these two methods to study chemical equilibrium in the gas phase and under various solvent conditions.

Studies of tautomerism phenomena have been of interest in many areas of chemistry. Reviews of the experimental and theoretical studies in this area of chemistry and biochemistry are presented in ref 1–4. Knowledge of the energetics of tautomerism in a simple model molecule such as 2-pyridone/2-hydroxypyridine, with its keto–enol equilibrium, and cytosine or other pyrimidine or purine bases, with an imino–amino equilibrium, can provide useful information on the intrinsic stability of various tautomers of molecules. Furthermore, knowing how these tautomerization energies change in different environments can give an insight into the influence of solvent effects on molecular stability.

Tautomerism also has biological consequences. The tautomerism of purine and pyrimidine bases naturally occurring in nucleic acids, nucleotides, and enzymes may play a role in mutagenesis.[5,6] One hypothesis suggests that the frequency of mispairing, and thus, mutagenesis in DNA, is correlated with equilibrium constants for the keto–enol or amino–imino[5,6] tautomerization.

For a few protomeric equilibria, precise thermodynamic data exist. We chose to study three representative cases. The first is the system 2-pyridone/2-hydroxypyridine, which is the simplest prototype for nucleic acid bases. This system has been extensively studied experimentally by means of UV,[7,8] IR,[9] and X-ray PES[10] spectroscopy and theoretically by semiempirical CNDO/2,[11] MINDO/3,[12] different levels of ab initio calculations,[4,13–15] and Monte Carlo simulations in water solution.[16,17] For a review see ref 4.

The second system is 2-pyrimidone/2-hydroxypyrimidine, which is a precursor of all pyrimidine nucleic acid bases. For this system only qualitative experimental data are available for equilibrium constants in vacuo, in water/ethanol solution, and in low-temperature matrices.[8,18,19] A number of theoretical studies have been done on this system including MINDO/3 geometry optimization,[12] semiempirical and ab initio STO-3G calculations,[13] and Monte Carlo simulations in water,[20] CCl_4,[21] and $CHCl_3$[22] solutions.

For both of the above molecules experimental results and Monte Carlo simulations suggest that in vacuo and nonpolar solutions (CCl_4) the hydroxy form prevails. However, in the crystal state[23] and polar solutions, where the interactions with the environment play a role, the tautomeric equilibrium is shifted toward the keto form.

The third system considered in our work is the molecule cytosine, for which experimental data are available for water and other solutions[24,25] and inert gas matrices[26] as well as theoretical studies involving MINDO/3[27] and ab initio 3-21G[28] geometry

* To whom reprint requests should be addressed.
[†] Permanent address: Quantum Chemistry Laboratory, Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland.
[‡] Current address: Chemistry Department, Harvard University, Cambridge, MA 02138.
[⊥] Permanent address: Research Institute of Scripps Clinic, Department of Molecular Biology, La Jolla, CA 92037.

(1) Katrizky, A. R.; Lagowski, J. M. *Adv. Heterocycl. Chem.* **1963**, *1*, 2.
(2) Pullman, B.; Pullman, A. *Adv. Heterocycl. Chem.* **1971**, *13*, 77.
(3) Kwiatkowski, J. S.; Pullman, B. *Adv. Heterocycl. Chem.* **1975**, *18*, 199.
(4) Kwiatkowski, J. S.; Zielinski, T. J.; Rein, R. *Adv. Quantum Chem.* **1986**, *18*, 85.
(5) Topal, M. D.; Fresco, J. R. *Nature (London)* **1976**, *263*, 285.
(6) Topal, M. D.; Fresco, J. R. *Nature (London)* **1976**, *263*, 289.
(7) Beak, P.; Fry, F. S., Jr. *J. Am. Chem. Soc.* **1973**, *95*, 1700.
(8) Beak, P.; Fry, F. S., Jr.; Lee, J.; Steele, F. *J. Am. Chem. Soc.* **1976**, *98*, 171.
(9) Nowak, M. J.; Szczepaniak, K.; Barski, A.; Shugar, D. *Z. Naturforsch. C, Biosci.* **1978**, *33C*, 876.
(10) Brown, R. S.; Tse, A.; Vederas, J. C. *J. Am. Chem. Soc.* **1980**, *102*, 1174.
(11) Berndt, M.; Kwiatkowski, J. S.; Budzinski, J.; Szczdrowska, B. *Chem. Phys. Lett.* **1973**, *19*, 246.
(12) Czerminski, R.; Lesyng, B.; Pohorille, A. *Int. J. Quantum Chem.* **1979**, *16*, 1141.
(13) Krebs, C.; Hofmann, H.-J.; Kohler, H.-J.; Weiss, C. *Chem. Phys. Lett.* **1980**, *69*, 537.
(14) Schlegel, H. B.; Gund, P.; Fluder, E. M. *J. Am. Chem. Soc.* **1982**, *104*, 5347.
(15) Scanlan, M. J.; Hillier, I. H.; McDowell, A. A. *J. Am. Chem. Soc.* **1983**, *105*, 3568.
(16) Corongiu, G.; Clementi, E.; Dagnino, M.; Paoloni, L. *Chem. Phys.* **1979**, *40*, 439.
(17) La Manna, G.; Venuti, E. *J. Comput. Chem.* **1982**, *3*, 593.
(18) Beak, P. *Acc. Chem. Res.* **1977**, *10*, 186.
(19) Nowak, M. J.; Szczepaniak, K.; Barski, A.; Shugar, D. *J. Mol. Struct.* **1980**, *62*, 47.
(20) Cieplak, P.; Geller, M. *THEOCHEM* **1985**, *121*, 247.
(21) Cieplak, P.; Geller, M. *THEOCHEM* **1985**, *124*, 249.
(22) Cieplak, P., unpublished results.
(23) Almlof, J.; Krick, A.; Olovsson, I. *Acta Crystallogr.* **1971**, *27*, 1201.
(24) Dreyfus, M.; Bensaude, O.; Dodin, G.; Dubois, J. E. *J. Am. Chem. Soc.* **1976**, *98*, 6338.
(25) Katrizky, A. R.; Waring, A. J. *J. Chem. Soc.* **1963**, 3046.
(26) Szczesniak, M.; Nowak, M. J.; Szczepaniak, K. *J. Mol. Struct.* **1984**, *115*, 221.

0002-7863/87/1509-6283$01.50/0

optimization in vacuo and the modeling of the hydration by various reaction field calculations.[29] The amino–imino equilibrium is observed[24] in water soluion, but the amino form predominates. The cytosine tautomeric equilibrium is a particularly important one, because the imino form can create an abnormal base pair C:A in DNA instead of the normal C:G pair observed with the amino tautomer.

To determine the equilibium constant for tautomerization in vacuo by theoretical means, the difference in total energies between the two tautomers may be an adequate approximation. In some cases, it has been assumed that differences in vibration energies and entropy differences cancel, which is not necessarily correct.[13,14] Calculations done with semiempirical methods often gave large errors in relative energies of tautomer pairs.[11,13,27] The most successful approach has been presented by Schlegel et al.[14] in which ab initio geometry optimization with smaller basis sets is followed by single-point energy calculations with a large basis set with polarization functions added. Addition of the correlation energy with the second-order Møller–Plesset[30] method together with the differences in zero-point vibration energies improved the agreement between calculated and experimental energies.

To date, the results of all the theoretical studies[16,17,20-22,29] devoted to investigate tautomerism in solutions give at best a qualitative description of this phenomena, since none of them is capable of providing the difference in free energy of solvation quantitatively. In the different Monte Carlo simulations or reaction field models, the free energy was usually approximated only by the difference in the solvation energy (enthalpy), since the differences in solvation entropy is difficult to calculate and assumed to cancel. To quantitatively estimate the changes in free energy of solvation between different species, the application of the free energy perturbations methods in Monte Carlo or molecular dynamics simulations is appropriate, and we present such an application below.

The calculation of free energy differences by the perturbation method has been done by using molecular dynamics (MD) simulations for the description of the cavity formation in water and solubility of noble gases in water solution,[31] in Monte Carlo (MC) calculations of the differences in solvation of CH_3CH_3 and CH_3OH,[32] in MD simulations for Cl^- and Br^- ions and SC24/H^+ ion association,[33,34] in MC simulations of $\Delta\Delta G$ of hydration between simple alkali and halide ions,[35] and in various MD simulations on relative solvation free energies of amino acids and nucleic acids.[36,37] The above successes of the free energy perturbation approach have encouraged us to use this method to simulate solvation effects in tautomerism phenomena.

Given the capabilities of the free energy perturbation approach, it appears to be possible to obtain "new answers" to these "old problems". In this paper we extend the work of Schlegel et al.[14] on the 2-oxopyridone/2-hydroxypyridine equilibrium in vacuo by carrying out geometry optimization at the 6-31G* level as well as single-point MP2 calculation at that level and carry out the first quantitative theoretical analysis of the solvation free energy differences between these tautomers. We also apply both ab initio and free energy perturbation methods to the 2-pyrimidone/2-hydroxypyrimidine and imino/amino cytosine tautomerism. The agreement between calculated and observed free energies is very good and suggests that the method used here can be used in a predictive way to study chemical tautomerism in various media.

Method and Computational Details

In our studies we have used a combination of ab initio quantum mechanics and molecular dynamics/free energy perturbation methods to calculate the differences in solvation free energy between two given tautomers.

To obtain a good representation of partial charge models for the molecules as well as the best possible energy differences between tautomers in vacuo, we performed ab initio geometry optimization at the 3-21G[38,39] level followed by single-point calculations in the 6-31G*[40] basis set for all molecules considered here. The calculations were done with programs Gaussian-80-UCSF[41] and Gaussian-82.[42] In all cases the molecules were assumed to be planar. The 6-31G* wave functions were used to calculate the electrostatic potential around molecules which were then used to determine partial charges on the atoms.[43] The charges used in the molecular dynamics simulations were rescaled down by the ratio of experimental and calculated within the 6-31G* basis set dipole moments. This ratio was taken to be 0.85 in all cases considered here, because this factor was appropriate for the pyridine isomers.

In one case, the 2-pyridone tautomers, we also carried out ab initio geometry optimization at the 6-31G* basis set level, followed by a calculation of the correlation effects with use of second-order Møller–Pleset (MP2) perturbation theory.[30]

After the quantum mechanical calculations were complete, one tautomer was transformed into another in solution by using molecular dynamics/free energy perturbation[36] methods to obtain the solvation free energy differences $\Delta\Delta G$ between these two isomers. This method has been implemented into the molecular simulation program AMBER-UCSF (Version 3.0).[44] The molecular dynamics simulations were carried out at T = 300 K and 1 atm of pressure in a water bath which contained from 578 to 600 TIP3P water molecules.[45] The initial configurations of water molecules were taken from a Monte Carlo simulation of TIP3P water. The walls of the simulated box were chosen to be at least 12 Å from the most peripheral solute atoms. Thus, we have a different number of H_2O molecules for different pairs of tautomers, see Table II). The van der Waals parameters for the nucleic acid bases were taken from ref 46.

The transformation of one tautomer into another was done in several stages by means of a coupling parameter λ. The dependence of the molecular mechanical energy functions on the parameter λ is given in ref 36 and, in the case of these tautomers, involves also simultaneous vanishing of a tautomeric proton in one position (e.g. oxygen) and growing of this proton in another position (e.g., nitrogen). All the bond length, bond angle, and dihedral angle equilibrium values in the MD simulations were chosen to be those corresponding to the quantum mechanically optimized values. The solvation free energy differences were then calculated from the following expression

$$\Delta\Delta G_{solv} = \sum_{i=1}^{N}\Delta\Delta G_i = -kT\sum_{i=0}^{N} \ln\,(\exp[-\beta(U(\lambda_{i+1}) - U(\lambda_i))])_i \quad (1)$$

where $\beta = 1/kT$ and $(\,)_i$ describes the mean value calculated over the state coupled to parameter λ_i. The above expression can be easily derived from the statistical mechanical perturbation theory introduced originally by Zwanzig in 1954[47] and is further discussed in ref 35 and 36.

The perturbation calculations were carried out in a series of 21 "windows" where the values of λ differed by 0.05. For each such window we calculated the mean value and the statistical fluctuation in the free energy change. For each window, the statistical fluctuation were less than 10% of the mean value. For each value of λ we performed 500 steps of equilibration followed by 500 steps of data collection with a time step

(27) Czerminski, R.; Lesyng, B.; Pohorille, A. *Int. J. Quantum Chem.* **1979**, *16*, 605.
(28) Scanlan, M. J.; Hillier, I. H. *J. Am. Chem. Soc.* **1984**, *106*, 3737.
(29) Bartzsch, C.; Weiss, C.; Hofmann, H.-J. *J. Prakt. Chem.* **1984**, *326*, 407.
(30) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(31) Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55.
(32) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
(33) Lybrand, T. P.; Ghosh, I.; McCammon, J. A. *J. Am. Chem. Soc.* **1985**, *107*, 7793.
(34) Lybrand, T. P.; McCammon, J. A.; Wipff, G. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833.
(35) Cieplak, P.; Lybrand, T. P.: Kollman, P. A. *J. Chem. Phys.*, in press.
(36) Singh, U. C.; Brown, F. K.; Bash, P. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1987**, *109*, 1607.
(37) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564.
(38) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.
(39) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.
(40) Harlharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(41) G80-UCSF: Singh, U. C.; Kollman, P. A. *QCPE* **1982**, 446.
(42) Binkley, J. S.; Whiteside, R. A.; Raghavachari, K.; Seeger, R.; DeFrees, D. J.; Schlegel, H. B.; Frisch, M. J.; Pople, J. A.; Kahn, L. R., Carnegie-Mellon University.
(43) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
(44) AMBER (UCSF), Version 3.0: Singh, U. C.; Weiner, P. K.; Caldwell, J. W.; Kollman, P. A. Department of Pharmaceutical Chemistry, University of California, San Francisco, 1986.
(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926.
(46) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(47) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

Table I. Total Energies (au) and Dipole Moments (D) Calculated at Different Levels for 2-Oxopyridine, 2-Oxopyrimidine, and Cytosine Isomers

	2-oxopyridine			2-oxopyridine		cytosine	
	keto Ia	hydroxy Ib	hydroxy Ic	keto IIa	hydroxy IIb	amino IIIa	imino IIIB
(a) HF/3-21G	−319.77080	−319.76814	−319.74340	−335.66517	−335.66336	−390.41619	−390.41557
geometry optimization	4.6	1.7	4.0	6.6	3.5	7.2	5.3
(b) HF/6-31G*	−321.56652	−321.56565	−321.55586	−337.56249	−337.56544	−392.61417	−392.61321
geometry optimized	4.8	1.5	3.7	6.6	3.0	7.2	5.3
(c) HF/6-31G*	−321.56725	−321.56709	−321.55746				
geometry optimization	4.6	1.5	3.6				
(d) MP2/6-31G*	−322.52444	−322.52700	−322.51743				
geometry optimized in HF/6-31G*							
experimental dipole moment	4.15[53]	1.15[53]				7.0[54]	

Table II. Tautomerization Energies ΔE (kcal/mol) for 2-Oxopyridine, 2-Oxopyrimidine, and Cytosine Isomers in the Gas Phase Calculated at Different Levels

	$\Delta E = E(\text{Ib}) - E(\text{Ia})$	$\Delta E = E(\text{Ic}) - E(\text{Ia})$	$\Delta E = E(\text{IIb}) - E(\text{IIa})$	$\Delta E = E(\text{IIIa}) - E(\text{IIIb})$
MINDO/3 geometry optimization	3.75	4.27	7.05	−13.82
(a) HF/3-21G geometry optimization	1.67	17.19	1.14	−0.39
(b) HF/6-31G* geometry optimized in 3-21G	0.55	6.69	−1.85	−0.61
(c) HF/6-31G* geometry optimization	0.10	6.14		
(d) MP2/6-31G* geometry optimized in HF/6-31G*	−1.61	4.40		−1.4[b]
zero-point vibration-energy differences (from MINDO/3)	−0.73	−0.85	−0.63	0.58
estimates[a] of tautomerization energy				
(b)	−0.18	5.84	−2.48	−0.03
(c)	−0.63	5.29		
(d)	−2.34	3.55		
exptl $\Delta\Delta G$	−0.4 ± 0.7 (IR)[8]		−2.4[8]	∼0[25c]
	−0.5 ± 0.8 (UV)[8]			
	−0.6 ± 0.1 (X-PES)[10]			

[a] Values given under labels b, c, d refer to the calculations done with HF/6-31G* with geometry optimized in the 3-21G basis set, HF/6-31G* geometry optimization, and MP2/6-31G* geometry optimized in the 6-31G* basis set, respectively. [b] Taken from ref 50, geometry optimized with the 3-21G basis set and energy calculated by the MBPT(2) method in the 6-31G* basis set. [c] Estimated from the solvent dependence of the tautomeric equilibria in ref 25.

0.002 ps at constant pressure and temperature (300 K) using a periodic boundary conditions.

To more accurately estimate the tautomerization energies in vacuo we also performed MINDO/3[48] geometry optimization calculations for the molecules, which provide us an estimate of zero-point vibration-energy differences.

Results for the Gas Phase

In Table V we list the optimal geometries obtained from total geometry optimization at the 3-21G basis set level for 2-oxopyrimidine and cytosine isomers. Those for 2-oxopyridine are obtained after 6-31G* optimization, since for this system results obtained with use of the 3-21G basis set are available elsewhere.[14] Table VI gives the partial atomic charges obtained by electrostatic potential fitting[43] at the 6-31G* basis set with use of 3-21G optimized geometries.

In general, the geometries obtained from isolated molecule optimization agree very well with experimental X-ray data (Table V), although some discrepancies are observed, since we cannot take into account interactions due to crystal packing and our theoretical model is not exact. To obtain tautomerization energies as accurately as possible, one must compare the energies of totally geometry optimized molecules.[12,14] The position of the tautomeric hydrogens is critical for determining sensible energy differences.[2,11,12] In our case, we find average difference of 0.02 Å in bond lengths and 2° in bond angles for heavy atoms between the 3-21G optimized geometry and experimental values.

In Table I the quantum mechanically calculated total energies and dipole moments at different levels of accuracy (basis set/correlation energy) are compared. In Table II the tautomerization energies are summarized. Zero-point energy differences $\Delta\epsilon_0$ between tautomers calculated with MINDO/3 method are in the range 0.6–0.9 kcal/mol. Thus, these cannot be neglected since a correction of this magnitude can make a significant difference in calculated equilibrium constants. Estimates made by Beak et al.[49] for methylated 2-oxopyridines suggest that the vibrational energy differences should be less than 1.5 kcal/mol, but even this upper bound seems to be high, given the MINDO/3 results. We have neglected the differences which arise from temperature dependence of vibrational and rotational energies and entropies because at the MINDO/3 level these are less than 0.1 and 0.2 kcal/mol, respectively.

Table II lists the results for tautomerization energies and indicates that the results obtained by the MINDO/3 method give the greatest errors. This was also stated earlier by Czerminski et al.,[12] Krebs et al.,[13] and Schlegel et al.[14] In the latter two papers, two numbers were cited with the wrong sign. The MINDO/3 model favors the keto form of 2-oxopyridine and 2-oxopyrimidine in the vapor phase rather than the hydroxy form. In addition, in ref 14, the experimental values for 2-oxopyridine tautomerization energies were cited with the wrong sign.

Table II also shows that going from the 3-21G basis set to 6-31G*, which contains additional d-type polarization functions, improves agreement with experiment for the 2-pyridone/2-hydroxypyridine equilibrium. For 2-oxopyridine further geometry optimization at the 6-31G* basis set moves the position of the tautomeric equilibrium in the right direction and closer to the observed value but still the keto form is more stable. Addition of the correlation energy calculated within the 6-31G*/MP2 level stabilizes the hydroxy form by 1.7 kcal/mol. This result is opposite

(48) Bingham, R. C.; Dewar, M. J. S.; Lo, D. H. *J. Am. Chem. Soc.* **1975**, *97*, 1294.

(49) Beak, P.; Bonham, J.; Lee, J. T. *J. Am. Chem. Soc.* **1968**, *90*, 1569.

Table III. Comparison of the Free Energy Differences (kcal/mol) and Equilibrium Constants for Tautomerization for 2-Oxypyridine, 2-Oxopyrimidine, and Cytosine Isomers in the H_2O Solution at 300 K Obtained from MD Simulation and Experiment[a]

n	tautomer pair	N H_2O in MD simulation	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{tot}$	K_{calcd}	$\Delta\Delta G_{expt}$	K_{exptl}
			Ia → Ib	Ib → Ia	Ia/Ib		
			5.6	(b) −5.2 ± 0.4	6,390		Ia/Ib
1	Ib ⇄ Ia	600	Ib → Ia	(c) −4.7 ± 0.4	2,980	−4.1[8]	910
			−5.1	(d) −3.0 ± 0.4	164		
			Ib → Ic	Ic → Ib	Ib/Ic		
			−0.9	(b) −5.2 ± 0.1	5,300		
2	Ic ⇄ Ib	584	Ic → Ib	(c) −5.1 ± 0.1	5,300		
			1.0	(d) −5.1 ± 0.1	5,300		
			Ia → Ic	Ic → Ia	Ia/Ic		
			4.5	(b) −10.2 ± 0.1			
3	Ic ⇄ Ia	584	Ic → Ia	(c) −9.7 ± 0.1			
			−4.3	(d) −7.9 ± 0.1	711,000		
			Ia → Ic	Ic → Ia	Ia/Ic		
		based on	4.7	(b) −10.2 ± 0.4			
3′	Ic ⇄ Ia	1 and 2	Ic → Ia	(c) −9.7 ± 0.4			
			−4.1	(d) −7.9 ± 0.4	711,000		
			IIa → IIb				
			5.6	IIb → IIa	IIa/IIb		IIa/IIb
4	IIb ⇄ IIa	578	IIb → IIa	−2.9 ± 0.2	154	<−1.6[8]	>15
			−5.3				(ethanol)
			IIIa → IIIb				
			4.3	IIIb → IIIa	IIIa/IIIb		IIIa/IIIb
5	IIIb ⇄ IIIa	587	IIIb → IIIa	−4.2 ± 0.2	1235	−5.5 to −6.9[24]	10^4–10^5
			−4.0				

[a] See Table II for explanation of the different level of calculations b, c, d of the results for the simulations number 1, 2, 3. $\Delta\Delta G_{tot}$ is defined by eq 2. $\Delta\Delta G_{solv}$ is the value calculated according to eq 1, K_{calcd} refer to our calculated equilibrium constant from $\Delta\Delta G_{tot}$, and K_{exptl} refer to the experimental values.

in direction to that obtained by Schlegel et al.[14] at the 6-31G/MP2 level, where the keto form was stabilized by 0.8 kcal/mol. Addition of both the correlation energy and zero-point vibrational-energy differences pushes the equilibrium too much toward the hydroxy form. This rather large basis set dependence of the MP2 correlation correction is surprising and prevents a more confident analysis of the theoretical energies. In fact, leaving out the MP2 correlation correction entirely leads to essential agreement with the experimental energy difference in vacuo.

In column 3 of Table II we also show the energy differences between 2-pyridone and the other 2-hydroxypyridine, "reversed" where the N–C–O–H dihedral angle is 180° rather than 0°. It can be easily seen that such a conformation is not likely in vacuo.

The overall agreement of the calculated data for the gas-phase tautomeric equilibria for 2-oxopyrimidine is acceptable, but one should keep in mind that the estimated error for the experimental results is not known and our result does not contain correlation energy correction.

In the case of the cytosine isomers, the experimental data for the gas-phase equilibrium is qualitative[26] only. On the basis of studies of methylated derivatives of cytosine in nonpolar solvents,[24] it was suggested that the imino form of cytosine should be abundant in vapor phase. Our calculations suggest that both forms should be observed in the gas phase but that solvation in water strongly favors the amino tautomer.

Solution Studies

In Table III we summarize the results for the free energies and equilibrium constants obtained here and in experiments. The $\Delta\Delta G_{tot}$ (column 5), which is the free energy difference between tautomers in solution, can be estimated according to equation

$$\Delta\Delta G_{tot} = \Delta\Delta G_{solv} + \Delta\epsilon_0 + \Delta E_{qm} + \Delta S_{qm} \qquad (2)$$

where $\Delta\Delta G_{solv}$ is the solvation free energy free difference between the tautomers, $\Delta\epsilon_0$ is the difference in zero-point vibrational energies between two tautomers, ΔE_{qm} is the difference in energies for isolated molecules obtained from the ab initio calculations, and ΔS_{qm} is the difference in the entropies which could be calculated by quantum mechanical methods, but which we assume to be negligible, given the MINDO/3 results noted above. In general, there is qualitative agreement in all cases. For the cytosine tautomerism there is an estimate of the free energy difference between amino and imino tautomers[25] and our results are reasonably close to this value.

It is encouraging that (Tabl III) the sum of the free energies for the three simulations involving the transformations Ia → Ib, Ib → Ic, Ic → Ia, which should be exactly 0 kcal/mol, is indeed 0.0 ± 0.4 kcal/mol. This accuracy supports the "robustness" of this method to calculate free energy differences, and it was achieved despite a difference in the number of water molecules in the simulations.

The effect of solvation on the isomerization of the Ib to Ic isomer also shows a competition between intrinsic energies and solvation effects. Table III, simulation no. 2, shows that the solvation free energy difference and the zero-point vibration-energy contribution favors Ic (angle N–C–O–H = 180°), but the internal electronic energy is lower by 6.0 kcal/mol for the Ib molecule, thus, in net, the Ib form is favored. This solvent influence can be explained by comparing the dipole moments for Ib and Ic. It is well-known that the molecule with the greater dipole moment will be more stabilized in a dipolar solvent. Thus stabilization by hydration is greater for Ic, which has μ = 3.6 D, than for Ib, with its μ = 1.5 D.

In the case of equilibrium for IIa ⇄ IIb, only the data measured in ethanol solution are available. However, changing solvents from ethanol to water the keto form, with its higher dipole moment, will be preferentially stabilized. This will make the difference in free energy more negative than −1.6 kcal/mol[8] and, thus, closer to our calculated value.

We also tried to estimate the influence of the substituent effect on the tautomerization process in the case of cytosine, because only the molecule substituted at the N1 position with CH_3 rather than H is relevant to the cytosine in DNA or RNA. To achieve this, we calculated the charge distributions for cytosine isomers with a methyl group at N1 rather than a hydrogen atom, since the change in electrostatic energies is the major contribution to the $\Delta\Delta G_{solv}$ when the reference and fully perturbed molecules do not differ much geometrically. Comparison of the charges for the most important atoms, e.g., C2,N3,O2,N4 (Table VI) in N^1-methylcytosine and cytosine for their amino and imino tau-

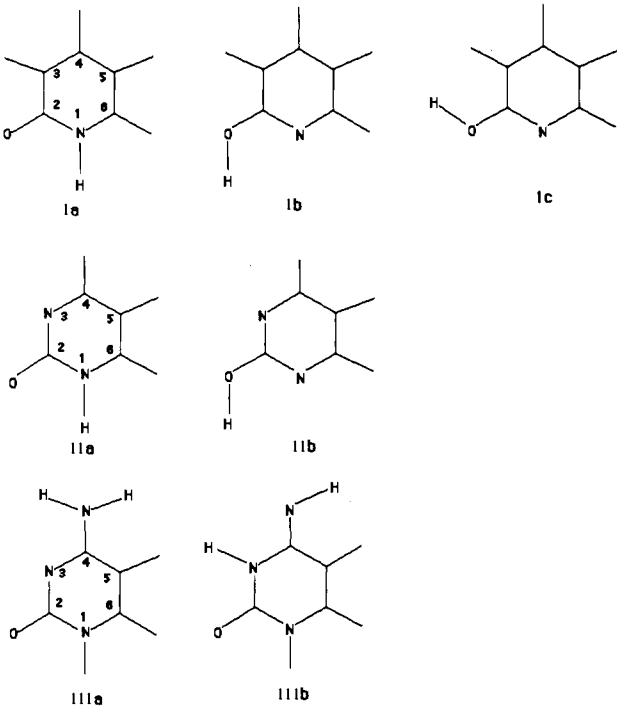

Figure 1. Tautomeric structures, atom nomenclature, and numbering scheme for 2-oxopyridine (I), 2-oxopyrimidine (II), and cytosine (III).

Table IV. The Mean Values of Water Coordination Numbers for N1, O2, N3, and N4 Atoms of Amino and Imino Tautomers of Cytosine Molecules Calculated at the Distances Corresponding to the First Minimum on the Appropriate Radial Distribution Functions[a]

	N1	O2	N3	N4
amino tautomer				
$r_{1\,min}$ (Å)	3.4	3.8	3.3	3.6
coord no.	1.8	6.0	1.5	4.1
imino tautomer				
$r_{1\,min}$ (Å)	3.6	3.5	3.3	3.8
coord no.	2.5 (1.7)	4.3 (5.6)	1.5 (1.5)	6.1 (5.3)

[a] For the imino tautomer, in parentheses, are given coordination numbers for the same distances as for the amino tautomer.

tomers, suggests that the substituent effect is insignificant. Also the total energy differences between the isolated amino and imino forms calculated by using the 6-31G* basis set, without further geometry optimization of the N^1-methylcytosine isomers, change from 0.6 kcal/mol for the unmethylated forms to 1.1 kcal/mol for N^1-methylcytosine.

We have carried out a structural analysis of the first solvation shell around cytosine tautomers. In Table IV we present the mean values of water coordination numbers for a given solute atom calculated at the distances corresponding to the first minimum on the appropriate radial distribution functions. For the imino tautomer we also consider the coordination numbers calculated using the same distances as for the amino tautomer. The changes with distance of the coordination number for the O2, N3, and N4 with the water hydrogens are shown in Figure 2. A comparison of these data show that although the number of water molecules that constitute the first solvation shell around both tautomers is almost the same, the orientational features of these waters are different. The atom O2 forms more hydrogen bonds with water in the amino cytosine tautomer (see Figure 2a), whereas the N4 atom has fewer H bonds in the amino tautomer (Figure 2c). The similar r dependance of the coordination number ($1.9 < r < 2.6$) for the water hydrogen atoms near N3 (Figure 2b) is surprising. The imino form has a hydrogen at N3 and would be expected to interact less effectively with the hydrogen of water. This suggests that some of the water hydrogens near N4 and O2 are contributing to the N3 coordination.

We also estimated the size of the interaction of cytosine tautomers with the first solvation shell of waters by calculating the

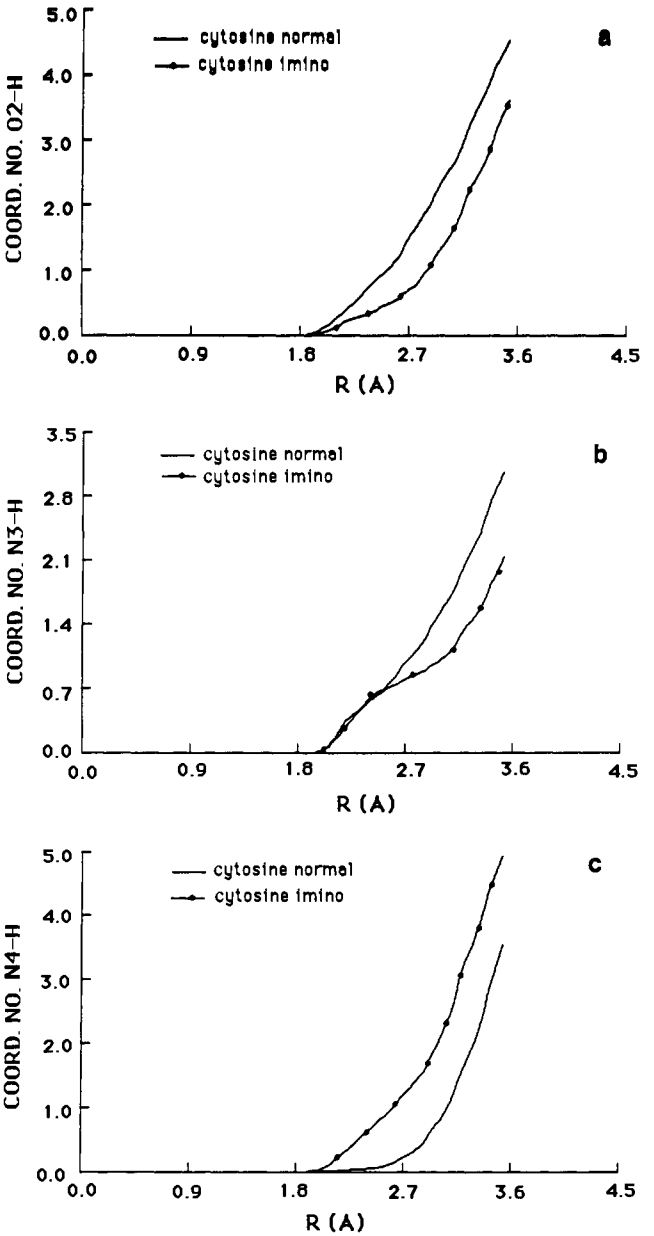

Figure 2. The R dependence of the coordination numbers of water's hydrogen atoms around O2, N3, and N4 atoms in cytosine tautomers.

mean value of the dot product between the water dipole moments and a vector parallel to the solute's dipole moment. The mean value of this product for the normal amino cytosine tautomer is −0.746 and that for the imino form is −0.204. It is thus clear that the amino tautomer, with its higher dipole moment, is able to more effectively align the waters near it than the imino tautomer. The importance of the overall dipole moment of the tautomers should not be overemphasized.[37] In Figure 3, we illustrate how, in the amino tautomer, the proton acceptor parts of the molecule (δ^-) can reinforce each other, whereas in the imino tautomer, the alternation of δ^+ and δ^- regions may preclude effective water alignment.

Discussion and Conclusions

The results of our quantum mechanical studies confirm many of the conclusions derived previously,[11,13,14] i.e., to obtain proper results for tautomerization energies in the gas phase, the proper choice for geometries of the tautomers and method for energy calculation are needed. One assumes that optimization of the geometry using a smaller 3-21G basis set followed by a single-point energy calculation with an extended 6-31G* basis set should give an energy difference that is close to that which could be obtained after full optimization with the extended basis set. In our test

Table V. Optimized Bond Lengths and Bond Angles

	2-oxopyridine (6-31G*)			2-oxopyrimidine (3-21G)		cytosine (3-21G)	
	keto	hydroxy	hydroxy reversed	keto	hydroxy	amino	imino
			A. Optimized Bond Length (in Å)[a]				
1–2	1.383 (1.37)	1.307 (1.34)	1.309	1.408 (1.37)	1.305	1.414 (1.39)	1.382
2–3	1.458 (1.41)	1.396 (1.38)	1.396	1.370 (1.36)	1.347	1.369 (1.36)	1.365
3–4	1.342 (1.35)	1.372 (1.38)	1.378	1.298 (1.31)	1.329	1.297 (1.34)	1.400
4–5	1.438 (1.39)	1.396 (1.38)	1.389	1.433 (1.39)	1.407	1.442 (1.43)	1.464
5–6	1.340 (1.35)	1.374 (1.34)	1.381	1.379 (1.35)	1.404	1.337 (1.36)	1.325
6–1	1.363 (1.36)	1.331 (1.34)	1.323	1.357 (1.35)	1.332	1.354 (1.36)	1.378
H–11	0.997			1.034		0.999	0.997
O(2)–2	1.203 (1.26)	1.336 (1.33)	1.340	1.215 (1.24)	1.320	1.211 (1.24)	1.213
H–3	1.073	1.073	1.076				1.000
H–4	1.075	1.075	1.076	1.117	1.117		
H–5	1.072	1.073	1.074	1.101	1.102	1.067	1.068
H–6	1.073	1.077	1.076	1.115	1.117	1.070	1.070
O(2)–H		0.950	0.946		0.951		
N(4)–4						1.344 (1.32)	1.258
N(4)–HA						0.998	
N(4)–HB						0.994	1.009
			B. Optimized Bond Angles				
1–2–3	113.7 (115)	124.1 (123)	123.5	111.9 (118.3)	123.8	115.1 (118.6)	113.6
2–3–4	116.3 (121)	117.3 (118)	117.8	125.0 (118.2)	117.0	122.2 (120.5)	128.0
3–4–5	121.5 (121)	119.6 (120)	119.4	122.5 (124.6)	123.1	122.9 (121.5)	113.6
4–5–6	117.6 (118)	117.6 (117)	117.4	115.5 (116.8)	115.1	116.1 (117.0)	119.6
5–6–1	121.0 (120)	123.7 (126)	124.1	118.8 (119.0)	120.4	120.7 (121.2)	122.2
6–1–2	125.1 (124)	117.7 (116)	117.9	126.3 (123.0)	120.6	123.0 (121.2)	123.0
H–1–2	114.8			119.3 (118.5)		115.7 (118.2)	116.2
O(2)–2–3	125.8 (126)	118.0 (119)	121.5	126.5 (121.9)	113.1	125.9 (122.5)	124.0
H–3–4	122.5	122.9	121.4				115.6
H–4–5	118.8	120.5	120.7	119.8 (121.6)	121.2		
H–5–6	121.0	120.9	120.9	121.4 (119.2)	122.5	122.4	121.2
H–6–1	115.7	115.7	115.5	118.3 (116.6)	118.3	116.7	115.3
H–O(2)–2		108.1	110.9		115.7		
N(4)–4–3						118.3	117.7
HA–N(4)–4						117.9	
HB–N(4)–4						122.4	115.2

[a] Experimental values (in parentheses) were taken from ref 23, 55, 56 for 2-oxopyridine, 2-oxopyrimidine, and cytosine, respectively.

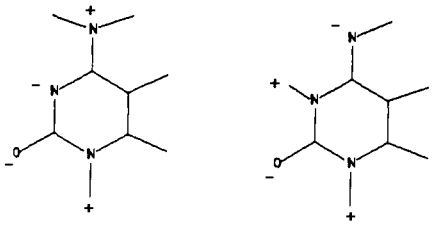

Figure 3. The schematic representation of the charge distribution for the cytosine amino and imino tautomers.

case, further optimization of the 2-pyridone and 2-hydroxypyridine at the 6-31G* level gave a 0.4-kcal/mol change in the relative tautomerization energy favoring the hydroxy form. Adding zero-point energy corrections favored the hydroxy tautomer by 0.7 kcal/mol, whereas changes in vibrational excitation energy and entropy for the tautomers at 300 K appears to be negligible.

We expect that the MINDO/3 model is adequate for calculating differences in zero-point and vibrational energies and entropies because in the case of comparison of two similar molecules, there could be significant cancellation of systematic errors in the theoretical model. This is evidently not true for relative electronic energies, where semiempirical methods are not adequate to accurately calculate these energies.

It is surprising that the MP2 correlation correction to the tautomerization energy for 2-oxopyridine at the 6-31G* level is −1.7 kcal/mol (favoring the hydroxy tautomer), but it is 0.8 kcal/mol at the 6-31G level.[14] It appears that using MP2 method with diffuse polarization functions can lead to inaccurate estimates of the correlation energies. Higher levels of correlation energy estimates are essential to determine gas-phase energy differences to better than 1 kcal/mol.

After our calculations were completed, we learned that Kwiatkowski et al.[50,51] have carried out ab initio calculations of the relative gas-phase energies for the nucleic acid bases at a similar level as our calculations for 2-oxopyridine and cytosine tautomers. They found that the MBPT(2) correlation correction calculated at the 6-31G* basis set level stabilizes the amino–keto (IIIa) isomer over keto–imino (IIIb) cytosine by 0.8 kcal/mol.

Our MD simulation results for calculation of solvation free energy differences between two isomers gave rather good agreement with experiment, and we suggest that they can be used to predict equilibrium constants in other cases where no experimental data are available. Specifically, the data are most definitive for the K_{equil} = [2-pyridone]/[2-hydroxypyridine]. Aqueous solvation changes the equilibrium constant in favor of the keto form by a factor of 2300, which corresponds to a $\Delta\Delta G_{solv}$ = 4.7 kcal/mol, in quite good agreement with the calculated $\Delta\Delta G_{solv}$ = 5.35 ± 0.25 kcal/mol. Given the simple molecular mechanical model used here to calculate solvation effects, the agreement with experiment is respectable.

In this context, we should note the reaction field calculations by Beak et al.[52] on a variety of pyridone–pyridine equilibria in a number of solvents. These authors show that the relative solvation free energies can be fit with a simple model that focuses on the dipole moments of the solutes, the dielectric constants of the solvents, and the H-bonding interactions between solute and solvent. Our results reinforce the basis of that model, in that one can in a "first principles" way calculate the solvation free energies

(50) Kwiatkowski, J. S.; Person, W. B.; Szczepaniak, K.; Szczesnaik, M. *Acta Biochim. Pol.*, in press.
(51) Kwiatkowski, J. S.; Bartlett, R. J.; Person, W. B. *J. Am. Chem. Soc.*, submitted for publication.
(52) Beak, P.; Covington, J. B.; White, J. M. *J. Org. Chem.* **1980**, *45*, 1347.
(53) Albert, A.; Phillips, J. N. *J. Chem. Soc.* **1956**, 1294.
(54) Kulakowska, I.; Geller, M. Lesyng, B.; Bolewska, K.; Wierzchowski, K. L. *Biochim. Biophys. Acta* **1975**, *407*, 420.
(55) Furberg, S.; Solbakk, J. *Acta Chem. Scand. A* **1974**, *28*, 435.
(56) Voet, D.; Rich, A. *Progr. Nucl. Acid Res.* Mol. *Biol.* **1970**, *10*, 183.

Table VI. Changes for 2-Oxopyridine, 2-Oxopyrimidine, Cytosine, and *N*-Methylcytosine

	2-oxopyridine			2-oxopyrimidine		cytosine		N^1-methylcytosine	
	keto	hydroxy	hydroxy reversed	keto	hydroxy	amino	imino	amino	imino
A. Charges Obtained from the 6-31G* Wave Function, Fitted To Reproduce Electric Field around Molecules with the Geometries Optimized in the 3-21G Basis Set[a]									
1	−0.4996	−0.7839	−0.7104	−0.7624	−0.9331	−0.7662	−0.5995	−0.2987	−0.1429
2	0.8290	0.9441	0.9960	1.2019	1.1889	1.1052	0.8776	1.0233	0.8304
3	−0.5118	−0.6341	−0.7677	−0.8987	−0.7831	−0.9741	−0.6993	−0.9629	−0.7558
4	0.1581	0.2372	0.3415	0.7126	0.5337	1.2474	0.9173	1.1566	0.9018
5	−0.4262	−0.5659	−0.6149	−0.7430	−0.7693	−0.8640	−0.6435	−0.7989	−0.6333
6	0.0770	0.4165	0.4615	0.3928	0.6087	0.3816	−0.2062	0.2052	0.0776
O(2)	−0.6629	−0.7022	−0.7218	−0.6872	−0.7020	−0.6828	−0.6422	−0.6583	−0.6284
H3	0.2019	0.2406	0.2232				0.3898		0.4060
H4	0.1201	0.1088	0.0840	0.0372	0.0748				
H5	0.1954	0.2047	0.2055	0.2354	0.2401	0.2605	0.2175	0.2575	0.2224
H6	0.1668	0.0794	0.0517	0.1296	0.0689	0.1338	0.1556	0.1687	0.1796
H(O2)		0.4546	0.4513		0.4725				
H1	0.3521			0.3816					
N4						−1.2117	−0.9613	−1.1718	−0.9522
HA						0.4893		0.4856	
HB						0.4865	0.4071	0.4720	0.4009
CH_3(N1)								0.1215	0.0939
(total)									

	keto	hydroxy	hydroxy reversed		keto	hydroxy	hydroxy reversed
B. Charges for 2-Oxopyridine after Geometry Optimization in the 6-31G* Basis Set							
1	−0.5041	−0.7843	−0.7150	H3	0.2051	0.2440	0.2266
2	0.8501	0.9547	1.0040	H4	0.1200	0.1069	0.0826
3	−0.5231	−0.6534	−0.7887	H5	0.1949	0.2085	0.2111
4	0.1567	0.2505	0.3593	H6	0.1673	0.0769	0.0473
5	−0.4280	−0.5812	−0.6410	H(O2)		0.4526	0.4519
6	0.0739	0.4182	0.4752	H1	0.3516		
O(2)	−0.6644	−0.6934	−0.7132				

[a] The charges actually used in the molecular dynamics simulations were these numbers multiplied by the factor 0.85.

with a simple molecular mechanical model in which the main difference in intermolecular interactions between the isomers and solvent is electrostatic. Thus, the quantitative calculations of solvation free energies described here and elsewhere can be complimentary to simpler solvation theories.[52]

The error bars for the calculated solvation free energy are less than 0.5 kcal/mol. The ab initio model for the gas-phase equilibrium in this example, where the most "extensive" model was used (6-31G*/MP2 + zero-point energy corrections, model d in Tabl II), favored the hydroxy form of 2-oxopyridine by 2.3 kcal/mol compared to the experimental value of 0.4–0.6 kcal/mol. We note that both less sophisticated models (b and c in Table II) are more consistent with experiment, and their use is further supported by the good agreement with experiment for the relative stabilities of the two pyrimidine isomers (II). Both quantum-mechanical gas-phase energies and the solvation free energies seem to have uncertainties in the range 1–2 kcal/mol and, in favorable cases, may be capable of providing even greater accuracy. In fact, the theoretically predicted solution equilibrium of 2-oxopyridine in which one combines the gas-phase and solution free energy esimates (Table III, model d) is 1.1 kcal/mol different from experiment, because of the fortuitous cancellation between the "gas-phase error" of −1.8 kcal/mol and the solution error of 0.65 kcal/mol.

Qualitatively, there is nothing surprising about the relative solvation free energies, since in each case the tautomer with the larger dipole moment is more effectively solvated. In the general case, as noted in our study of nucleic acid base solvation,[37] both the dipole and quadrupole moments may play a role in the relative calculation of solvation free energies of molecules.

In summary, it is clear that ab initio theory with zero-point vibrational-energy corrections can lead to excellent agreement with available experiments on gas-phase tautomerism. It is somewhat disconcerting that the most sophisticated model (model d in Table II) is not the most accurate, but it is encouraging that one high level model (b) gives rather good agreement with experiment for all three examples of gas-phase tautomerism.

The solvation free energies can be calculated rather accurately. In the case where the data are clearest, calculated and experimental results differ by only ~0.6 kcal/mol. Inaccuracies in the calculation can come from the use of a scale factor of 0.85 for the 6-31G* electrostatic potential derived charges, which may be more accurate for some molecules than others. In addition, we do not take into account possible polarization of the charges in water solution and this could lead to an additional contribution to differences in solvation energy. It is clear, however, that the calculated relative solvation free energies for tautomers are capable of being within 1 kcal/mol of experiment, which is approaching "chemical" accuracy.

It is interesting that in all three cases here, the tautomeric equilibrium is very solvent dependent, the most dramatic example being the 2-pyrimidone/2-hydroxypyrimidine equilibrium which strongly favors the hydroxy form in the gas phase and the keto form in solution. In the case of cytosine, it is clearly not an evolutionary advantage to shield the molecule from the water solvent during replication, since this would lead to a much greater percent of imino tautomers and consequent C:A mispairing than would be the case in water solution. Thus, such solvent effects on tautomeric equilibrium can have a very profound effect on biological processes, and it is exciting to be able to calculate them in a "first principles" manner.

Acknowledgment. We are glad to acknowledge research support of the National Science Foundation (CHE-85-10066) and National Institutes of Health (CA-25644). One of us (P.C.) thanks the Polish Academy of Science for partial support for this project through Grant No. CPBP 01.12. Most of these calculations were carried out at the San Diego Supercomputer Center through supercomputer support provided to P.A.K. by the National Science Foundation (CHE-85-10066 and DMB-84-19883). We also gratefully acknowledge the use of an FPS-264 array processor, purchased through grants from the National Institutes of Health (RR-02441) and National Science Foundation (DMB-84-13762) in some of this work, as well as the facilities of the UCSF Computer Graphics Laboratory (supported by RR-1081 to R. Langridge).

Registry No. 2-Oxopyridine, 142-08-5; 2-oxopyrimidine, 557-01-7; cytosine, 71-30-7.