Tautomeric Equilibria in 3- and 5-Hydroxyisoxazole in the Gas Phase and in Aqueous Solution: a Test of Molecular Dynamics and Continuum Models of Solvation

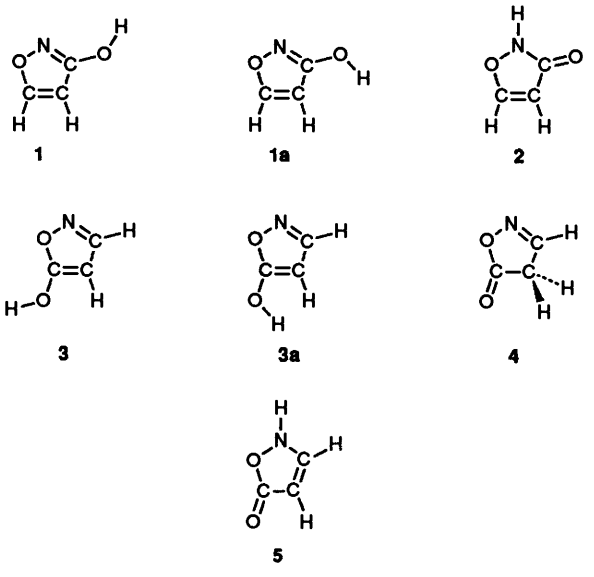
Stephen Woodcock,[a] Darren V. S. Green,†[a] Mark A. Vincent,[a] Ian H. Hillier,[a] Martyn F. Guest[b] and Paul Sherwood[b]
[a] *Chemistry Department, University of Manchester, Manchester, M13 9PL, UK*
[b] *SERC, Daresbury Laboratory, Warrington, WA4 4AD, UK*

The keto–enol tautomerism of 3- and 5-hydroxyisoxazole in the gas phase and in aqueous solution has been investigated by a range of theoretical methods. In the gas phase, *ab initio* studies using basis sets with polarisation functions, and the inclusion of electron correlation at the MP4 level, predict that the hydroxy form of 3-hydroxyisoxazole, and the 4*H*-oxo form of 5-hydroxyisoxazole are the only forms that will be observed. Free energies of hydration, predicted by free energy perturbation studies using molecular dynamics, and by the polarisable continuum model due to Tomasi *et al.* yield consistent results which differ from those predicted by the self-consistent reaction field (SCRF) model. All methods studied here predict that the enol form of 3-hydroxyisoxazole is dominant in aqueous solution, in agreement with experiment, whilst it is only the SCRF model that correctly predicts that the 2*H*-oxo form of 5-hydroxyisoxazole is observed in aqueous solution.

The tautomeric equilibria of heterocycles[1,2] continue to be investigated by a variety of theoretical methods ranging from semiempirical molecular orbital (MO) methods[3] to more sophisticated calculations which include electron correlation.[4] A major problem in judging the accuracy of these calculations is the sparsity of gas phase data which are available, so that estimates of solvent effects are needed to compare theoretical predictions with experimental data which are often obtained in aqueous solution. Here too there are a number of theoretical methods available to estimate solvent–solute interactions, ranging from quite simple reaction field models to computationally demanding simulation studies.

The combination of high-level *ab initio* calculations of the isolated tautomers together with estimates of differential solvation effects by the free energy perturbation (FEP) method has led to excellent correlation with experiment for a number of systems.[5,6] However, a self-consistent reaction field (SCRF) method[7] implemented within the AM1 semiempirical MO scheme[8] has also been found to be successful in predicting tautomer preferences in aqueous solution for a wide range of heterocycles.[3,9]

Among the remaining problems in heteroaromatic tautomerism, those that involve compounds having contiguous heteroatoms are probably the most challenging. The case of 1,2,3- and 1,2,4-triazole has been studied both as the isolated molecule and in aqueous solution[6] with essentially quantitative agreement between theory and experiment. It is of interest to examine the situation in the oxo-heterocycles of this type, since they are almost the only class for which the oxo-tautomer cannot be presumed to be dominant. They thus provide a sensitive test of predictive methods. In this paper we study theoretically two such heterocycles. The first, 3-hydroxyisoxazole, can be written as the hydroxy (**1**, **1a**) and oxo forms (**2**). The two hydroxy forms written differ in the orientation of the hydroxy group. This species is known to exist in solution as the hydroxy form.[10] However, for 5-hydroxyisoxazole, it is the 2*H*-oxo form (**5**) that exists in aqueous solution rather than the



4*H*-oxo form (**4**) or the hydroxy form (**3**, **3a**).[2] It is the purpose of the present work to study the effects that are responsible for the observed tautomer preferences, and to construct a model for the prediction of tautomer preferences. We note that for the species investigated herein no quantitative data on tautomer ratios are available to provide a more stringent test of the predictive value of the theory. We now describe the computational methods employed which involve both free molecule calculations and the modelling of solvent effects, and the results obtained from these methods.

Computational Methods

Free Molecule Calculations.—The case of the 2-hydroxypyridine–2-pyridone equilibrium is well-studied, since gas phase data are available. It has been found that geometry optimisation using at least a double-zeta basis, followed by energy calculations using a large basis with polarisation functions and

† *Present address:* Glaxo Group Research, Greenford Road, Greenford, Middlesex UB6 0HE.

including electron correlation, are needed to obtain tautomer energy differences accurate to *ca.* 1 kcal mol^{-1}.[11] Such a computational procedure has also been found to be effective in the study of tautomerism in the pyrimidine bases. Thus, the species studied here were optimised at the 3-21G[12] level and characterised as minima by calculations of their harmonic frequencies. Subsequently, energy evaluations were carried out using a 6-31G** basis[13] with electron correlation included using Møller–Plesset theory[14] both at second order (MP2) and fourth order (MP4). These calculations were carried out using the programs GAMESS[15] and GAUSSIAN88.[16]

Modelling of Solvent–Solute Interactions.—Both continuum models, and those that include explicit solvent–solute interactions were studied. The SCRF method allows the incorporation of Onsager's reaction field model into molecular orbital calculations, to allow for the consideration of non-specific solvent–solution interactions in terms of the solute dipole moment and the relative permittivity (ε) of the solvent. The formalism described by Tapia and Goscinski[7] has been implemented[8] in the semiempirical MO package, MOPAC. We have implemented this method in the *ab initio* package GAMESS. These SCRF calculations were carried out at the SCF level, using a 6-31G** basis, at the optimised gas phase geometry obtained by an SCF calculation using a 3-21G basis. In view of the approximate nature of this calculation, we assume that the corrections due to electron correlation are the same as those calculated for the isolated molecules. The total solvent–solute interaction obtained from the SCRF model arises from the interaction in the absence of solute polarisation together with the increased interaction resulting from solute polarisation.

A more realistic shape for the solvent cavity and a more accurate representation of the solute charge distribution are incorporated in a polarisable continuum model (PCM) developed by Tomasi and coworkers.[17] This method involves the generation of a cavity from spheres centred at each atom in the molecule and the calculation of virtual charges on the cavity surface to represent the polarisation of the solvent. The magnitude of these charges is proportional to the derivative of the solute electrostatic potential at each point directly calculated from the electronic wavefunction. The PCM treatment has also been implemented by us in the program GAMESS.

The only parameters in the SCRF and PCM models, in addition to those present in the MO treatments, are those related to the cavity geometry. The radius of the spherical cavity in the SCRF model can be chosen either from simple geometric considerations or from the molar volume. We have chosen a constant value (2.5 Å) for the cavity radius for the species studied here, using the former approach, following Karelson *et al.*[8] For the PCM treatment, the individual sphere radii must naturally depend upon atom type, but should also vary with formal atomic charge, and are also expected to be basis-set dependent. Appropriate parameters have been developed[18] to allow the atomic radii to be calculated in terms of Mulliken charges and basis sets, and are used herein.

The second strategy for estimating solute–solvent interaction energies considers such interactions explicitly using a free energy perturbation method[19] in molecular dynamics simulations to obtain directly the contribution of solvation ($\Delta\Delta G_{solv}$) to the total free energy difference ($\Delta\Delta G_{tot}$) between pairs of tautomers. The calculations were carried out using the program AMBER.[20] The atomic partial charges of the species studied were obtained using the strategy proposed by Singh and Kollman.[21] Following the geometry optimization at the 3-21G level, single-point calculations were carried out using a 6-31G* basis. This wavefunction was then used to calculate the electrostatic potential around the molecule which leads to the determination of the partial atomic charges. As far as other molecular mechanics parameters are concerned, the van der Waals parameters were taken from Weiner *et al.*[22] and the intramolecular parameters for the tautomers were chosen to obtain a good fit to the optimised 3-21G geometries. The molecular dynamics simulations were carried out at T = 300 K and 1 atm of pressure in a water bath containing 580 TIP3P water molecules. In the free energy perturbation method, the transformation between the two species involved in the equilibrium is carried out in a number of small steps involving the coupling parameter λ. The molecular mechanics energy of the system is expressed in terms of this parameter λ.[23] The perturbations were carried out in a series of 21 'windows' where the values of λ differed by 0.05. For each value of λ, 1000 steps of equilibration followed by 1000 steps of data collection with a time step of 0.002 ps at constant pressure and temperature were performed using periodic boundary conditions.

Computational Results

3-Hydroxyisoxazole.—The results of the *ab initio* calculations are presented in Tables 1 and 2. At the 3-21G level (3-21G//3-21G), the keto form **2** is predicted to be the most stable. However, all calculations using the basis containing polarisation functions on all atoms (6-31G**) predict the '*anti*' hydroxy form **1** to be the most stable, the ordering of the '*syn*' hydroxy form **1a** and the keto form **2** being dependent upon the inclusion of correlation effects. At both the MP2 and MP4 level, electron correlation preferentially stabilises the hydroxy form, so that at the highest level of theory used here [6-31G**(MP4)//3-21G] the relative energies are **1** < **1a** < **2**, with the keto form being 7 kcal mol^{-1} * higher than the most stable hydroxy form. The AM1 results (Table 2) are more in line with the 3-21G//3-21G results, the keto form **2** being predicted to be most stable. The *ab initio* values presented in Table 2 do not include zero-point terms. For the three species (**1**, **1a**, **2**) the zero-point energies, calculated at the 3-21G//3-21G level, using the harmonic approximation, are to within 0.2 kcal mol^{-1}, and thus will not affect the conclusions of this study.

We turn now to the prediction of the hydration energies obtained from the FEP method, and the SCRF and PCM treatments, summarised in Tables 3, 4 and 5 respectively. The MD simulations predict that the solvation energies of the keto and '*syn*' hydroxy forms are essentially the same, being 2 kcal mol^{-1} greater than that for the most stable gas phase tautomer ('*anti*' hydroxy **1**). When combined with the gas phase energies (Table 2), the predicted ordering in water is unchanged, with **1** and **1a** being predicted to be significantly more stable than the keto form. This prediction is in agreement with experiment, where the observed tautomer in aqueous solution is the hydroxy form.

We next consider the predictions of the two continuum models. In the SCRF model, the solvation energy is strongly dependent upon the dipole moment of the solute, being proportional to the square of this quantity in the absence of solute polarisation. This is the reason for the greater stabilisation of the '*syn*' hydroxy form **1a** compared with that of tautomers **1** and **2** seen in Table 4, so that **1a** is predicted to be the most stable form. The SCRF model also predicts a greater solvation energy for **2** than for **1**, owing to its larger dipole moment (Table 4), so that this treatment predicts the relative energies in water to be in the order **1a** < **2** < **1**. Thus, the SCRF model, like the MD simulations treatment, predicts the hydroxy form to be observed in water, although the two methods predict a different conformer to be predominant.

In Table 5, we show the solvation free energies predicted by

* 1 cal = 4.184 J.

Table 1 Total energies (a.u.) of tautomers calculated by *ab initio* methods

Tautomer	3-21G//3-21G	6-31G**//3-21G	6-31G**(MP2)//3-21G	6-31G**(MP4)//3-21G
1	−317.6574	−319.4554	−320.3951	−320.4333
1a	−317.6481	−319.4476	−320.3880	−320.4264
2	−317.6620	−319.4509	−320.3825	−320.4219
3	−317.6594	−319.4558	−320.3947	−320.4330
3a	−317.6538	−319.4522	−320.3916	−320.4296
4	−317.6660	−319.4674	−320.4011	−320.4456
5	−317.6619	−319.4511	−320.3864	−320.4249

Table 2 Relative energies (kcal mol^{-1}) of tautomers calculated by *ab initio* and semiempirical methods

Tautomer	3-21G//3-21G	6-31G**//3-21G	6-31G**(MP2)//3-21G	6-31G**(MP4)//3-21G	AM1 [3]
1	0	0	0	0	0
1a	5.8	4.9	4.6	4.3	—
2	−2.9	2.8	7.9	7.1	−0.1
3	4.2	7.3	4.1	8.0	15.0
3a	7.7	9.5	6.0	10.0	—
4	0	0	0	0	0
5	2.6	10.2	9.3	13.0	9.8

Table 3 Free energy differences (kcal mol^{-1}) of tautomers in water obtained from MD simulations

Tautomer pair	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{tot}$ [a]
1 ⇌ **2**	1 ⟶ 2	
	−2.2 ± 0.4	+4.9 ± 0.4
1 ⇌ **1a**	1 ⟶ 1a	
	−1.7 ± 0.5	+2.6 ± 0.5
3 ⇌ **4**	3 ⟶ 4	
	−1.8 ± 0.3	−9.8 ± 0.3
3 ⇌ **3a**	3 ⟶ 3a	
	−1.5 ± 0.2	+0.5 ± 0.2
4 ⇌ **5**	4 ⟶ 5	
	−2.1 ± 0.1	+10.9 ± 0.1

[a] Evaluated using 6-31G**(MP4)//3-21G energies (Table 2).

Table 4 Tautomer energy differences (kcal mol^{-1}) in water obtained from the self consistent reaction field (SCRF) method

Tautomer	AM1 [3]	*ab initio* [a]	μ(6-31G**//3-21G)(D) [b]
1	0	0	2.5 (3.5)
1a	—	−12.8	6.2 (8.5)
2	2.6	−1.8	3.8 (7.3)
3	19.7	27.4	2.0 (3.3)
4	0	0	5.4 (9.1)
5	3.4	−2.3	6.6 (12.9)

[a] Evaluated at 6-31G**//3-21G level with gas phase values at 6-31G**(MP4)//3-21G level. [b] The values in parentheses are the dipole moments calculated for the hydrated species.

the second continuum model employed, the PCM treatment. For tautomers **1** and **2** the relative free energies of solvation are in excellent agreement with the results of the MD simulations. We have not considered tautomer **1a**, which on the basis of MD simulations appear to be less stable than form **1** in water. On the basis of these results for 3-hydroxyisoxazole, it is suggested that the SCRF model overestimates the solvation of the tautomers with large dipole moments.

5-Hydroxyisoxazole.—For this molecule we have studied the two conformers of the hydroxy tautomer **3** and **3a** and the two oxo forms **4** and **5**. It is the 2*H*-oxo form **5**, which predominates in aqueous solution. For the gas phase species (Table 1), the

Table 5 Solvation free energy (ΔG_{solv}) [a] and total relative free energies (ΔG_{tot}) [b] (kcal mol^{-1}) in water obtained by PCM treatment

Tautomer	ΔG_{solv}	ΔG_{tot}	μ/D
1	−12.6	0	3.3
2	−15.0	4.7	5.1
3	−12.3	12.0	2.8
4	−16.3	0	7.4
5	−17.2	12.1	8.6

[a] Evaluated at 6-31G**//3-21G level. [b] Evaluated using gas phase values at 6-31G**(MP4)//3-21G level.

changes in relative stabilities predicted at the different levels of theory parallel those found for 3-hydroxyisoxazole. Thus, a 3-21G basis overestimates the relative stability of the keto form **5**, the correlation energy being greater for the hydroxy than for the keto forms, although the inclusion of correlation effects does not affect the energy ordering of the tautomers which is **4** < **3** < **3a** < **5**. It is to be noted that there are considerable changes in the relative energies of the tautomers on increasing the level of electron correlation from MP2 to MP4, unlike the situation for 3-hydroxyisoxazole. Thus, we cannot define the relative energies of the tautomers in the gas phase as precisely as we would wish. However, it is clear from the results of Table 2 that for tautomer **5** to be predicted to be the dominant form in aqueous solution requires at least 10 kcal mol^{-1} greater solvation energy for **5** compared with the 4*H*-oxo form **4** which is predicted to be the predominant tautomer in the gas phase. The latter form is also predicted by the AM1 method to be the predominant form in the gas phase.

The modelling of solvation using MD simulations (Table 3), predicts that the solvation energy for the hydroxy form **3** is 2–4 kcal mol^{-1} less than that for the other three structures. Such a small differential solvation effect does not change the ordering of the tautomers from that predicted for the gas phase, with the 4*H*-oxo form (**4**) being of considerably lower energy. The predictions of the PCM treatment (Table 5) are again very similar to those of the MD simulations, although the differential solvation energies between **3** and **4**, **5** are somewhat larger (4–5 kcal mol^{-1}). However, the 4*H*-oxo form is again predicted to be the only form observed in aqueous solution, in disagreement with experiment. The SCRF model yields considerably different predictions, with substantially increased differential

solvation for the oxo forms (**4**, **5**) compared with **3** and a much larger solvation energy for **5** compared with **4**. These effects result in the oxo forms being considerably more stable in aqueous solution compared with **3**, with the 2*H*-oxo form **5** being slightly more stable than the 4*H*-oxo form **4**, in agreement with experiment. This result arises both from the different gas-phase dipole moments of the tautomers (Table 4), with that of **5** being the largest, and from the greater polarisation energies of **4** and **5**, compared with **3**, and of **5** compared with **4**, as witnessed by the calculated dipole moments of the hydrated species. It is to be noted that the increases in the dipole moments upon solvation are considerably greater for the SCRF compared with the PCM treatment (Tables 4 and 5). The greater polarisation of **5** compared with **4** may be understood in terms of the greater conjugation of **5**.

Conclusions

The *ab initio* calculations described here have led to confident predictions of the tautomeric species present in the gas phase, although experimental data are lacking at present. Such data could probably be provided by infrared studies of matrix-isolated species which have been used with great success in identifying the tautomers present in other systems, particularly the pyrimidine bases.[24] Large basis set *ab initio* calculations have been shown[24] to be valuable in the assignment of such infrared spectra. The results we have obtained show the need for polarisation functions to predict accurate energetics, and the relative stabilisation of the hydroxy tautomers produced by the inclusion of electron correlation. The results of modelling hydration are less definitive. Of particular note is the success of the PCM treatment in giving results close to those obtained by MD simulations with considerably less expenditure of computer time. However, although both treatments are successful in their predictions for 3-hydroxyisoxazole, they both fail to predict the most stable tautomer for hydrated 5-hydroxyisoxazole. This latter deficiency is not shown by the SCRF model, where the greater stabilisation of the 2*H*-oxo form is attributed both to the dipole approximation, and to the large degree of electron polarisation occurring upon hydration, predicted by this model. There is a clear need for more accurate treatments of tautomerism in 5-hydroxyisoxazole, including both terms beyond the dipole approximation and the accurate treatment of polarisation effects.

Acknowledgements

We thank P. J. Taylor for helpful discussions, and the SERC (UK) for support of this research.

References

1 A. R. Katritzky and J. M. Lagowski, *Adv. Heterocycl. Chem.*, 1963, **1**, 339.
2 J. Elguero, C. Marzin, A. R. Katritzky and P. Linda, *The Tautomerism of Heterocycles*, Academic Press, New York, 1976, p. 283.
3 M. M. Karelson, A. R. Katritzky, M. Szafran and M. C. Zerner, *J. Chem. Soc., Perkin Trans. 2*, 1990, 195; H. S. Rzepa, M. Yi, M. M. Karelson and M. C. Zerner, *J. Chem. Soc., Perkin Trans 2*, 1991, 635.
4 H. B. Schlegel, P. Gund and E. M. Fluder, *J. Am. Chem. Soc.*, 1982, **104**, 5347; M. J. Scanlan, I. H. Hillier and A. A. MacDowell, *J. Am. Chem. Soc.*, 1983, **105**, 3568.
5 P. Cieplak, P. Bash, U. C. Singh and P. A. Kollman, *J. Am. Chem. Soc.*, 1987, **109**, 6283.
6 J. R. Cox, S. Woodcock, I. H. Hillier and M. A. Vincent, *J. Phys. Chem.*, 1990, **94**, 5499.
7 O. Tapia and O. Goscinski, *Mol. Phys.*, 1975, **29**, 1653.
8 M. Karelson, T. Tamm, A. R. Katritzky, S. J. Cato and M. C. Zerner, *Tetrahedron Comput. Methodol.*, 1989, **2**, 295.
9 A. R. Katritzky and M. M. Karelson, *J. Am. Chem. Soc.*, 1991, **113**, 1561.
10 A. R. Katritzky, *Handbook of Heterocyclic Chemistry*, Pergamon Press, Oxford, 1985, pp. 121–123.
11 M. Moreno and W. H. Miller, *Chem. Phys. Lett.*, 1990, **171**, 475; O. G. Parchment, I. H. Hillier and D. V. S. Green, *J. Chem. Soc., Perkin Trans. 2*, 1991, 799.
12 J. S. Binkley, J. A. Pople and W. J. Hehre, *J. Am. Chem. Soc.*, 1980, **102**, 939.
13 P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta.*, 1973, **28**, 213.
14 C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618.
15 M. F. Guest and J. Kendrick, GAMESS User Manual CCP1/86/1, Daresbury Laboratory, 1986.
16 M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Ragahavachari, J. S. Binkley, C. Gonzalez, D. J. De Frees, D. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol and J. A. Pople, GAUSSIAN88, Gaussian Inc., Pittsburg PA.
17 S. Miertus, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117.
18 M. A. Aguilar and F. J. Olivares del Valle, *Chem. Phys.*, 1989, **129**, 439.
19 W. L. Jorgenson and C. Ravimohan, *J. Chem. Phys.*, 1985, **83**, 3050.
20 U. C. Singh, P. K. Weiner, J. W. Caldwell and P. A. Kollman, AMBER (UCSF) Version 3.0, Department of Pharmaceutical Chemistry, University of California, San Francisco, 1986.
21 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129.
22 S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comput. Chem.*, 1986, **7**, 230.
23 U. C. Singh, F. K. Brown, P. Bash and P. A. Kollman, *J. Am. Chem. Soc.*, 1987, **109**, 1607.
24 H. Rostkowska, K. Szczepeniak, M. J. Nowak, J. Leszczynski, K. KuBulat and W. B. Person, *J. Am. Chem. Soc.*, 1990, **112**, 2147.

Paper 2/04129D
Received 31*st July* 1992
Accepted 29*th September* 1992