JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 8, 395–399 (1995)

AQUEOUS SOLVATION EFFECT ON THE PROTOTROPIC TAUTOMERISM OF 2-THIOCYTOSINE

J. GUILLERMO CONTRERAS* AND JOEL B. ALDERETE
Facultad de Ciencias Quimicas, Universidad de Concepcion, Casilla 3-C, Concepcion, Chile

Free energies of solvation were calculated for the four most stable gas-phase tautomeric forms of 2-thiocytosine (TC) using a local field SCF procedure with the solvation model SM2. The calculated changes in the free energies for each pair of tautomers reveal that all six possible equilibria tend to produce the thione–amino TC(1, 8, 8) and TC(3, 8, 8) species. This is in agreement with UV and ionization constant data.

INTRODUCTION

2-Thiocytosine is a minor component of tRNA and possesses important biological implications.[1–6] In fact, the replacement of the major nucleic acid base cytosine by thiocytosine may produce significant changes in the DNA structure as a result of perturbation in the base-pairing process. Hence it is of interest to study the prototropic tautomerism since thiocytosine can exist in any of several tautomeric forms that may play an important role in the replication process and in spontaneous mutations.

2-Thiocytosine can give rise to six tautomeric forms as a result of the thiol thione and amino imino equilibria. In fact, the above equilibria can produce two thione–amino species, one thione–imino, one thiol–amino and two thiol–imino forms. These tautomeric forms are named TC(i, j, k), where i, j, k represent the numbers of atoms to which tropic hydrogens are attached.

In a previous investigation,[7] we studied the prototropic tautomerism of 2-thiocytosine in the gas phase using the MNDO method. In an attempt to learn more about the predominant species, we calculated the thermodynamic properties in the range 100–500 K and found that the tautomers TC(7, 8, 8) and TC(1, 8, 8) predominate at low and room temperature, whereas TC(1, 7, 8) could become important at higher temperatures. However, in the light of the present results, these conclusions are doubtful. In general, the semi-empirical methods do not provide a good description of the atomic charge distributions and energies. This is particularly true for the AM1 method when applied to sulphur-containing molecules, since the sulphur parameters were developed from few experimental data.[8] On the other hand, *ab initio* calculations at high HF levels have shown that AM1 produces poor relative gas-phase energies for heterocycles.[9–11]

Based on the above conclusions and in an attempt to place the prototropic tautomerism on a more reliable basis, we have carried out calculations on this thiobase in both the gas phase and aqueous solution. The calculation of the gas-phase free energies of the six tautomeric forms of 2-thiocytosine were performed at the *ab initio* level using the 6–31G* basis set. To account for the solvent effect, the SM2 solvation model was included.[12,13] Although for some tautomeric forms the relative energies are greater than 10 kcal mol^{-1} (1 kcal = 4·184 kJ) above the most stable species, we included them in the solvation calculations as their high dipole moments can lead to great stabilization in polar solvents. The standard free energy of a species in solution (G°_{soln}) is obtained by adding the free energy of solvation to the relative gas-phase free energy obtained at the HF/6–31G* level.[14]

CALCULATIONS

Gas-phase calculations

We used the program package MOPAC 5.0[15] updated with sulphur parameters.[8] The *ab initio* MO calculations were performed using the Gaussian 92 series of programs.[16] The geometry optimization of all tautomeric forms was carried out with a 6–31G* basis set at the HF level. Frequency calculations at the equilibrium geometries yielded all real frequencies and hence all structures correspond to local minima. The energy calculations were performed using the 6–31G* basis set. To obtain

* Author for correspondence.

CCC 0894–3230/95/060395–05

Received 23 September 1994
Revised 18 January 1995

the true energy, a zero point energy (unscaled) correction was added to the total energy for each tautomer. In fact, a scaling of ZPE by 0·9, as usual, produces no effect on the conclusions of this study.

Solvation model

To calculate the aqueous solvation free energies, the program AMSOL 3.0,[17] which contains the SM1, SM2 and SM3 solvation models, was used. The local field SM2 model calculates the polarization effects (G_P) and cavitation–dispersion–structure changes of the solvent (G°_{CDS}) contributions to the free energy in aqueous solution. The polarization effects are calculated using a form of the generalized Born equation[18,19] included in the Fock operator during the SCF calculation. The contributions to the free energy due to G°_{CDS} are based on the superficial accessible area of the solvent[20,21] and the semi-empirical parameters of atomic solvation.[13,19] The free energy of the solute in aqueous solution is $G^\circ_{soln} = G^\circ_{(g)} + \Delta G^\circ_s$, where G°_{soln} is the free energy of the solute in solution, $G^\circ_{(g)}$ is the free energy of the solute in the gas phase and ΔG°_s is the aqueous free energy of solvation. ΔG°_s can be partitioned as follows:

$$\Delta G^\circ_s = E_{EN(aq)} + G_{P(aq)} + G^\circ_{CDS(aq)} - E_{EN(g)}$$

where E_{EN} is the nuclear and electronic energy and $\Delta G^\circ_s = \Delta G_{ENP} + G^\circ_{CDS(aq)}$. For the tautomeric equilibria A B, the difference $\Delta G^\circ_{soln} = G^\circ_{soln(B)} - G^\circ_{soln(A)}$ gives an indication of the stabilization of B with respect to A. For each tautomeric form, both the geometry and wavefunction were optimized in solution to produce the 'relaxed' solvation free energy.[12,13]

RESULTS AND DISCUSSION

Figure 1 gives the general labelling of the atoms of 2-thiocytosine and the optimized geometries of all six tautomers. Table 1 gives the *ab initio* gas-phase calculations and the AM1 heats of formations and dipole moments for all tautomeric forms of 2-thiocytosine. The enthalpy changes were obtained by adding $\Delta(ZPE)$ and thermal corrections $\Delta(H - H_0)$ to ΔE at the HF/6–31G* level. The free energy changes were calculated from $\Delta G = \Delta H - T\,\Delta S$. The calculated gas-phase free energies are referred to the most stable tautomer, i.e. TC(7,8,8). From comparison of the *ab initio* and AM1 relative energies, it can be seen that the stability orders given by these two quantum mechanical methods are almost the same, with the exception that the tautomeric form TC(1,3,8) (a thione–imino species) is stabilized by *ca* 6 kcal mol^{-1} over the TC(3,8,8) at the HF/6–31G* level, whereas in the AM1 calculation the TC(3,8,8) tautomer is stabilized by *ca* 2 kcal mol^{-1} over TC(1,3,8). When comparing the calculated relative energy values, the situation is different. In fact, the AM1 energy values are unreliable for the reasons

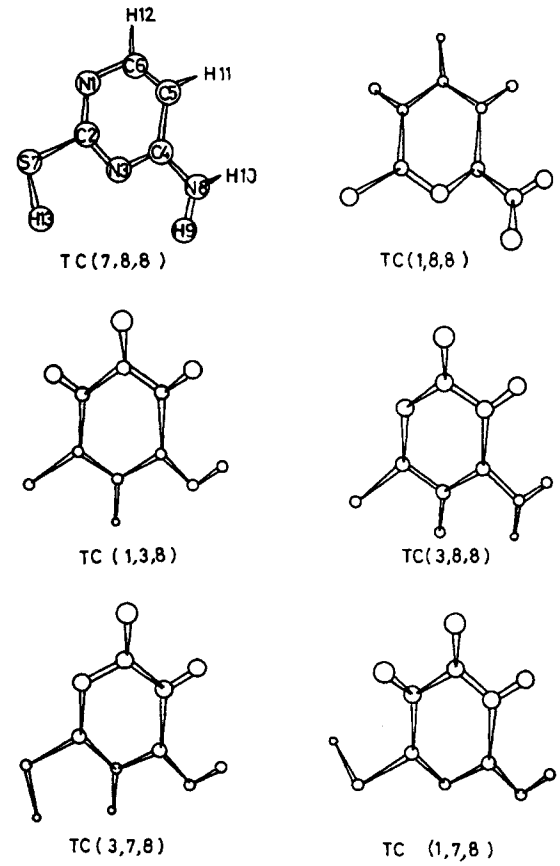

Figure 1. Optimized geometries and general labelling of atoms in 2-thiocytosine (TC)

given before. Figure 2, drawn with the data given in Table 1 and those in Ref. 7, shows the relative energies of the various tautomers calculated by different quantum mechanical methods. From Table 1 and Figure 2, it can be inferred that the stability order calculated by the MNDO method is different from that obtained by the AM1 and *ab initio* methods. In conclusion, the *ab initio* method predicts that the automeric forms TC(7,8,8), TC(1,8,8), TC(1,3,8) and TC(3,8,8) are the most stable. The thiol–amino TC(7,8,8) form has a clear predominance over the other three.

In a solution of high dielectric constant, such as water, one can expect a greater stabilization of the TC(3,8,8) tautomer owing to its larger dipole moment, but the cost of the reorganization of electronic distribution and nuclear relaxation (see below) make the TC(1,8,8) species the most stable in solution.

In this work we studied the polarization effect of thiocytosine in aqueous solution in order to calculate thefree energy of solvation of the four most stable gas-

Table 1. Calculated energies[a,b] and dipole moments (μ) of the six tautomers of 2-thiocytosine (TC) in the gas phase

Parameter	TC(7,8,8)	TC(1,8,8)	TC(3,8,8)	TC(1,3,8)	TC(3,7,8)	TC1,7,8)
E^a	−715·25131	−715·24812	−715·23605	−715·24634	−715·22045	−715·20270
ZPE	63·10	65·27	65·08	66·15	63·46	62·96
$H-H_0$	4·21	4·20	4·13	3·91	4·23	4·50
S	81·73	82·26	81·55	80·60	82·61	85·12
μ^c	4·79	9·30	10·10	6·43	0·44	7·74
μ^d	3·77	7·64	8·08	2·90	2·34	4·05
$\Delta H_f{}^d$	46·49	55·37	56·77	58·99	60·74	64·52
Relative values (kcal mol^{-1})						
ΔE^d	0·00	8·88	10·28	12·50	14·25	18·03
ΔE^a	0·00	2·00	9·58	3·12	19·36	30·50
$\Delta(ZPE)^e$	0·00	2·17	1·98	3·05	0·36	−0·14
$\Delta(H-H_0)$	0·00	−0·01	−0·08	−0·30	0·02	0·29
ΔH	0·00	4·16	11·48	5·87	19·74	30·65
$T\Delta S$	0·00	0·15	−0·05	−0·34	0·26	1·01
ΔG^f	0·00	4·01	11·53	6·21	19·48	29·64

[a] Calculations based on HF/6–31G*//6–31G*.
[b] E in hartree; *ZPE*, $H-H_0$, ΔE, ΔH, $T\Delta S$ and ΔG in kcal mol^{-1}; *S* in cal mol^{-1} K^{-1}.
[c] HF/6–31G* values, in debye.
[d] AM1 values.
[e] Unscaled values.
[f] Based on ΔE calculated at the HF/6–31G* level.

phase tautomers, using the SM2 solvation model. Table 2 gives the solvation energies for all six tautomeric forms of 2-thiocytosine in aqueous solution.

From Table 2, it can be inferred that the tautomer TC(1,8,8) is the most important species in solution, whereas the TC(7,8,8), TC(3,8,8) and TC(1,3,8) tautomers are *ca* 8 kcal mol^{-1} above TC(1,8,8) in aqueous solution. In fact, the latter three species possess roughly the same importance. This implies that in aqueous solution the thione–amino TC(1,8,8) is the predominant species, whereas the TC(3,8,8) and the TC(7,8,8) and the thione–imino TC(1,3,8) tautomers are present to a lesser extent.

For the equilibria given in Table 3, we also calculated the changes in the standard free energies in solution, ΔG°_{soln}. It is found that in solution all the less important species produce the most stable (TC(1,8,8). It is likely

Table 2. Solvation energies (kcal mol^{-1}) for the six tautomeric forms of 2-thiocytosine

Tautomer	$\Delta E_{EN}{}^a$	G_p	G°_{CDS}	ΔG°_s	$G^\circ_{soln}{}^b$
TC(7,8,8)	5·08	−14·38	−7·04	−16·34	−16·34
TC(1,8,8)	19·16	−37·49	−10·38	−28·72	−24·71
TC(3,8,8)	22·71	−40·23	−10·28	−27·81	−16·28
TC(1,3,8)	10·90	−22·64	−10·71	−22·45	−16·24
TC(3,7,8)	5·61	−13·26	−7·34	−14·99	4·49
TC(1,7,8)	10·93	−22·79	−7·45	−19·31	10·33

[a] $\Delta E_{EN} = E_{EN(aq)} - E_{EN(g)}$.
[b] Calculated from $G^\circ_{soln} = G^\circ_{(gas)} + \Delta G^\circ_s$.

Table 3. Prototropic tautomerization free energy changes for the various equilibria in 2-thiocytosine (kcal mol^{-1})

Equilibrium	ΔG°_{soln}
TC(7,8,8) ⇄ TC(1,8,8)	−8·37
TC(7,8,8) ⇄ TC(3,8,8)	0·06
TC(7,8,8) ⇄ TC(1,3,8)	0·10
TC(1,8,8) ⇄ TC(1,3,8)	8·47
TC(3,8,8) ⇄ TC(1,3,8)	0·04
TC(3,8,8) ⇄ TC(1,8,8)	−8·43

Table 4. Aqueous net atomic charges and dipole moments (*D*) for the more stable tautomers of 2-thiocytosine

Atom	TC(7,8,8)	TC(1,8,8)	TC(3,8,8)	TC(1,3,8)
N-1	−0·345	−0·264	−0·319	−0·245
C-2	−0·140	0·172	0·159	0·242
N-3	−0·313	−0·330	−0·252	−0·306
C-4	0·172	0·227	0·282	0·143
C-5	−0·364	−0·298	−0·325	−0·208
C-6	−0·011	0·071	0·049	0·021
S-7	0·456	−0·647	0·638	−0·606
N-8	−0·335	−0·278	−0·297	−0·285
H-9	0·260	0·272	0·287	0·166
H-10	0·253	0·293	0·306	0·271
H-11	0·175	0·216	0·214	0·210
H-12	0·203	0·246	0·228	0·246
H-13	−0·010	0·320	0·307	0·351
μ^a	2·578	14·685	16·437	7·034

Table 3. Prototropic tautomerization free energy changes for the various equilibria in 2-thiocytosine (kcal mol^{-1})

that both TC(7,8,8) and TC(3,8,8) are also formed, but according to Table 3 the equilibria relating these species are largely displaced to TC(1,8,8) formation. The above results agree well with the UV data and ionization constants derived from 2-thiocytosine in aqueous solution. In fact, Brown and Teitei[22] found that the tautomeric forms TC(1,8,8) and TC(3,8,8) exist as a mixture in aqueous solution with a clear predominance of the former. In addition, although no direct comparison can be made, it is worth nothing that 2-thiocytosine in the solid state shows simultaneous intermolecular hydrogen bonding at both N-1 and N-3[23] and that the calculated gas-phase proton affinities (*PA*) for these two atoms show that they are the preferred sites of protonation. The difference in the calculated PA is just 1·4 kcal mol^{-1}.

For all four important species in solution, the atomic charges and dipole moments for the 'relaxed' gas-phase molecular structure and the reorganized solute nuclear and electronic structure for TC(1,8,8) are given in Table 4. The large changes in the dipole moment on electronic relaxation come from the migration of the negative charge from the entire molecule to the C=S bond region. For the TC(7,8,8), TC(3,8,8) and TC(1,3,8), the SM2 solvation model predicts $\Delta G_{ENP(aq)}$ values of −9·3, −17·52 and −11·74 kcal mol^{-1}, respectively. The reorganizations of the atomic charges cost 19·16 kcal mol^{-1} for TC(1,8,8), 22·71 kcal mol^{-1}

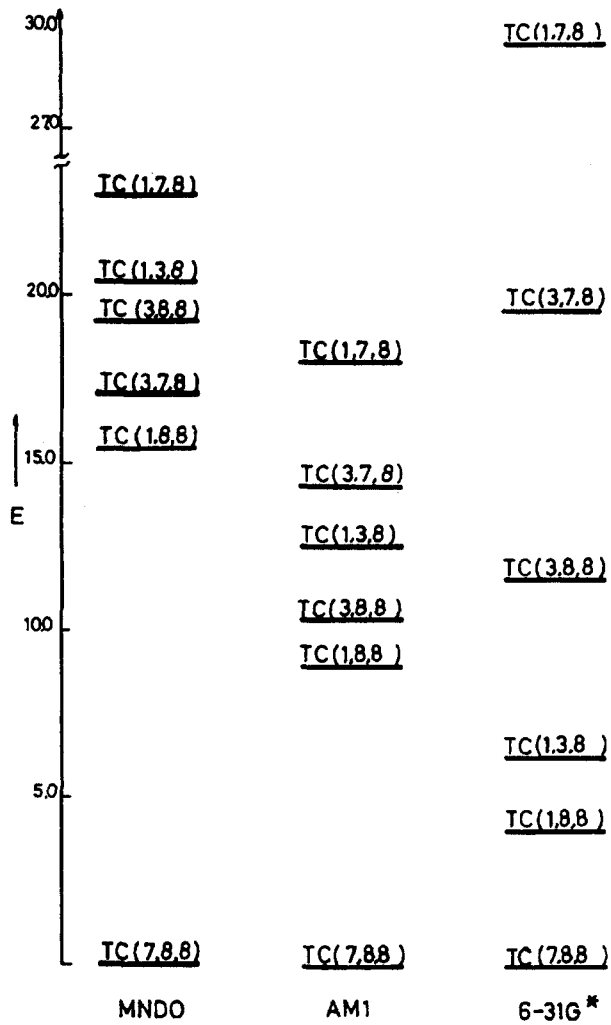

Figure 2. Relative energies of 2-thiocytosine tautomers in the gas phase calculated by different methods

for TC(3,8,8) and 10·90 kcal mol^{-1} for TC(1,3,8), yielding G_P values of −37·49, −40·23 and −22·64 kcal mol^{-1}, respectively. The polarizable nature of these species allows the reorganization costs to be overcome by the large G_P values obtained. This can also explain the large variation of the dipole moment on going from the gas phase to the related structure case.

Finally, it is worth noting that the results on the tautomerism of 2-thiocytosine in aqueous solution just obtained do not necessarily apply to 2-thiocytosine in tRNA, where a number of other effects can operate.

ACKNOWLEDGEMENTS

This work was supported by operating grants from FONDECYT (Nos 1930541 and 2930038) and the Universidad de Concepcion (Nos 94.23.03-1 and 92.013.010-6. J.B.A. is grateful to Fundacion Andres for a Scholarship.

REFERENCES

1. W. Saenger, *Principles of Nucleic Acids Structure*. Springer, Berlin (1984).
2. U. Thewald and C. E. Bugg, *J. Am. Chem. Soc.* **94**, 8892 (1972).
3. W. Saenger and D. Suck, *Eur. J. Biochem.* **82**, 473 (1973).
4. A. G. Lezius and K. H. Scheit, *Eur. J. Biochem.* **3**, 85 (1969).
5. K. H. Scheit and E. Gartner, *Biochim. Biophys. Acta* **182**, 10 (1969).
6. E. Gottschalh, E. Kopp and A. G. Lezius, *Eur. J. Biochem.* **24**, 168 (1971).
7. J. G. Contreras and J. B. Alderete, *J. Mol. Struct. (THEOCHEM)* **251**, 195 (1991).
8. M. J. S. Dewar and Y. Ch. Yuan, *Inorg. Chem.* **29**, 3881 (1990).
9. S. Woodcock, D. V. S. Green, M. A. Vincent, I. H. Hilliers, M. F. Guest and P. J. Sherwood, *J. Chem. Soc., Perkin Trans. 2* 2151 (1992).
10. J. G. Contreras and J. B. Alderete, *J. Mol. Struct. (THEOCHEM)* **283**, 283 (1993).
11. J. G. Contreras, S. T. Madariaga and J. B. Alderete, submitted for publication.
12. C. J. Cramer and D. G. Truhlar, *Chem. Phys. Lett.* **198**, 74 (1992).
13. C. J. Cramer and D. G. Truhlar, *J. Comput. Aided Mol. Des.* **6**, 692 (1992).
14. C. J. Cramer and D. G. Truhlar, *J. Am. Chem. Soc.* **115**, 8810 (1993).
15. J. J. P. Stewart, *Quantum Chemistry Program Exchange*. University of Indiana, Bull. 7(4) (1987).
16. M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Reploge, R. Gomperts, J. L. Andrews K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart and J. A. Pople, *Gaussian 92, Version D2*. Gaussian, Pittsburgh (1992).
17. C. J. Cramer, G. C. Lynch and D. G. Truhlar, *Quantum Chemistry Program Exchange*. University of Indiana, Program No. 606 (1992).
18. R. Constanciel and R. Contreras, *Theort. Chim. Acta* **65**, 1 (1984).
19. W. C. Still, A. Tempczac, R. C. Hawley and T. Hendrichson, *J. Am. Chem. Soc.* **112**, 6127(1990).
20. R. B. Hermann, *J. Phys. Chem.* **76**, 2754 (1972).
21. T. J. Richmond, *J. Mol. Biol.* **178**, 63 (1984).
22. D. J. Brown and T. Teitei, *Aust. J. Chem.* **18**, 559 (1965).
23. S. Furburg and L. H. Jensen, *Acta Crystallogr., Sect. B* **26**, 1260 (1970).