The Tautomerism of Cytosine and Hydroxycytosine. A Quantum-Mechanical Study

Józef S. Kwiatkowski

Institute of Physics, Nicholas Copernicus University, 87-100 Torun, Poland,

and

Bogdan Lesyng

Department of Biophysics, Institute of Experimental Physics, University of Warsaw, 02-089 Warszawa, Poland,

and

Michael H. Palmer

Department of Chemistry, University of Edinburgh, Edinburgh EH9 3JJ, Scotland

and

Wolfram Saenger

Abteilung Chemie, Max-Planck-Institut für Experimentelle Medizin, D-3400 Göttingen, Bundesrepublik Deutschland

Z. Naturforsch. **37 c,** 937–941 (1982); received March 9, 1982

Tautomerism, Cytosine, Hydroxycytosine

Semi-empirical studies by the MNDO and MINDO-3 methods together with *ab initio* studies by the Hartree Fock method, have been performed on the tautomers of cytosine (C) and hydroxycytosine (HC), in each case with full optimisation of the molecular skeleton. The results are in agreement with recent solution studies of these systems, and show that the amino form of C and the oxime form of HC are the lowest energy tautomers. Thus the property is one of the ring system rather than environment.

Various conformers of the imino tautomer of 5-methyl-N(4)-hydroxycytosine were studied by MINDO/3 and PCILO calculations; although only partial geometry optimisation was performed, it seems unlikely that the molecule can be a complementary base to adenine in the Watson-Crick sense.

I. Introduction

The action of hydroxylamine on cytosine (C) residues in DNA both *in vitro* and *in vivo*, leads to several modified species, and in particular to N(4)-hydroxycytosine (HC). Mutagenic effects have been ascribed to the latter when it participates in base pairing with guanine and adenine, as a result of mispairing [1–5].

Both these molecules (C, HC) are capable of tautomeric equilibria. In aqueous solution cytosine exists predominantly in the amino form (AC, **1**), with a tautomeric constant $K_{AC/IC}$ of about $10^4 \sim 10^5$, for [AC]/[IC] where IC is the corresponding imino-form (**2**) [1]. In contrast, N(4)-hydroxycytosine (HC) has the ratio much closer to unity, and with the imino form (IHC, **3**) more abundant than the amino-form (AHC, **4**) [1]; the tautomeric constant [IHC]/[AHC] is about $10 \sim 10^2$ [1], and thus the change in ratio between C and HC is approximately 10^5 or greater; these equilibria are of interest to the theory of point mutations. Additional interest attached to N(4)-hydroxycytosine since its ability to base pair is dependent on the orientation of =N–OH group of the imino form relative to N(3) [6].

We have carried out full geometry optimisation of cytosine and N(4)-hydroxycytosine, by various methods, with a view to (a) obtaining information on the relative stabilities of the free molecules, and (b) obtaining electronic and molecular properties at the computed equilibrium geometry. The present paper is concerned with (a), while a following paper discusses charge distributions and also ionisation potentials as determined by photoelectron spectroscopy [7].

II. Methods

In the first phase of work, the equilibrium geometry of the tautomers **1** and **2**, **3** and **4** were

Reprint requests to M. H. Palmer.
0341-0382/82/1000-0937 $ 01.30/0

determined by semiempirical means using the MINDO/3 and MNDO methods [8, 9], and the PCILO method [10]. The most promising structures for **1–4** from the first phase then acted as starting positions for an *ab initio* study. The latter used a 7s3p atomic orbital Gaussian basis for C,N,O and a 3s basis for H, as in our previous work [11, 12]. These basis functions where then contracted to 5,2 and 3 GTO contracted basis for each of 1s, 2s and 2p for C,N,O and 3 GTO's for the $1s_H$ orbital. The energy optimised GTO's yield > 99.7% of the Hartree-Fock free atom energies, and have been shown to give good agreement with a number of experimental geometries [13, 14]. The gradient procedures were based upon the Komorniki and McIver procedure [15–17] and in the *ab initio* studies were carried out using the HONDO-6 [18] program on an ICL 2972 and a CDC 7600 computer.

It is often assumed that amino-substituted heterocycles are completely planar, although it is known that the parent hydrocarbon analogue, aniline is distinctly non-planar overall; the latter is basically two intersecting planes, one containing the C_6H_5N portion and the other the NH_2 group [19]. Thus in some of the MNDO and *ab initio* studies, this extra flexibility was allowed by lowering the symmetry constraints for **1–4** from C_s to C_1.

III. Results and Discussion

Following normal practice, the results of the semiempirical studies are given on a relative basis, with positive energies denoting lower stability relative to the reference form. The same practice is followed with the *ab initio* studies, but in order to compare with absolute energies of other molecules, the computed total energies in atomic units are also given (1 a. u. = 627.5 kcal mol^{-1} = 2626 kJ mol^{-1}). The principal results are shown in Table I for all complete 2N or 3N coordinate searches of the N atom species (C_s and C_1 respectively).

(a) The tautomerism of cytosine and hydroxycytosine

The MINDO-3 calculations suggest the order of stability **1** (AC) > **5** (IC) > **2** (IC) for cytosine, with the reverse order **6** (IHC) > **3** (IHC) > **4** (AHC) (both conformers) for hydroxycytosine. The more refined MNDO calculations lead to the same overall pattern for the amino-tautomer of cytosine, (**1**) being more stable than the imino-form (**2, 5**) and the reverse being true for the hydroxy-compounds (**3, 6** > **4**), but the energy differences are generally markedly altered. Thus the hydroxyamino forms **4** (and the related tautomer with HN and NOH

Table I. Absolute total energies (a.u.) and relative energies (in kJ/mol) of various forms of cytosine and N(4)-hydroxycytosine[a, b]

	Symmetry	MINDO/3:	MNDO:	Relative *ab initio:*	Final total energies (a.u.)
Cytosine					
amino form					
(C, a)	C_s	− 25.73	− 15.23	+ 0.74	− 391.46302
	C_1	−	− 20.08[b]	+ 0.55	− 391.46309
imino forms					
(C, i−r)	C_s	ref	ref	ref	− 391.46330
(C, i−l)[c]	C_s	− 10.00	+ 0.42	−	−
N(4)-hydroxycytosine					
amino forms					
(HC, a−r, u)	C_s	+ 22.68	+ 87.95	−	−
(HC, a−l, u)	C_1	+ 24.10	+ 100.75	55.38	− 466.05437
imino forms					
(HC, i−r, u)	C_s	ref	ref	ref	− 466.07546
(HC, i−l, u)	C_s	− 11.09	− 3.35	−	−

[a] ref, − the energy of the form taken as zero. Plus or minus indicate that the form is less or more stable, respectively, relative to the reference form.
[b] Unpublished results by A. Sygula (Department of Chemistry, Jagiellonian University, Cracow, Poland – private information to J.S.K.).
[c] Results taken from [24].

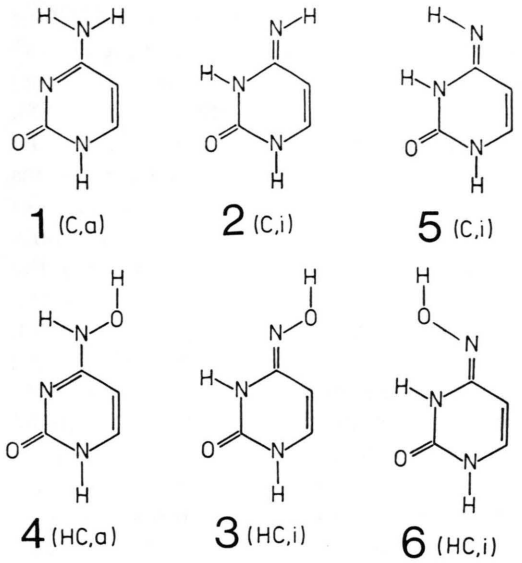

Fig. 1. Molecular structures of amino (a) and imino (i) tautomers of cytosine (C) and N(4)-hydroxycytosine (HC). The conformers of the tautomers depending on the structure and orientation of the external group (= N – H, = N – OH, H – N – OH) are indicated as the conformers *right* (r: 9, 10) or left (l: 7, 8) if the hydrogen of the substituent is located in positions *anti* or *syn* relative to the ring nitrogen N (3), respectively, and as conformers *up* (7, 9) or *down* (8, 10) if the N – H and O – H bond of the substituent is directed up or down, respectively, relative to the bonds C = N or N – O.

reversed in position) are singularly destabilised. The *ab initio* results again lead to the same overall conclusion with respect to the hydroxycytosine case, but with the energy difference between the two forms **3** and **4** being rather smaller. In contrast the *ab initio* studies suggest that the two forms of cytosine itself are effectively identical in energy. In conclusion, all three methods of investigation for the isolated molecules (as in the gas phase) yield results in agreement with the aqueous solution data cited above, although the absolute values are somewhat variable with the method. If we define the difference in energy between the amino (A) and imine (I) forms to be $\Delta E_{A,I} = E_A - E_I$, and take the difference (δE) between two pairs of molecules (C and HC) $-\delta E = \Delta E^{C}_{A,I} - \Delta E^{HC}_{A,I}$ we obtain 48, 103 and 56 kJ mol^{-1} for MINDO-3, MNDO and *ab initio* methods; it is interesting to note that an earlier CNDO-2 study, but without geometry optimisation, yielded the result $-\delta E = 58$ kJ mol^{-1} [20]. Thus it is clear that in this instance the use of standardised geometric parameters rather than fully optimised geometries is adequate.

For cytosine itself, it seems to be apparent that the energy difference between amino and imino forms declines with sophistication in the calculations as follows: $-\Delta E_{A,I} = 43$ (MINDO-2 [21]), 26 (Mindo-3), $15 \sim 20$ (MNDO) and 0.6 kJ mol^{-1} (*ab initio*). Usually the comparison of these theoretical isolated molecule differences is with thermodynamic data in solution, although this is of low validity; the experimental results for the present system **1** → **2** are $\Delta E \sim 23.0$ [22], $\Delta F \sim 28.2$ [23] and $\Delta H \sim 23.9$ kJ mol^{-1} [23]. These results suggest a rather higher proportion of amino to imino form (about 400:1 respectively) than the present studies. It has recently been shown that the amount of imino tautomer may have been somewhat underestimated [25]; thus a comparison with 3-methylated cytosine in non-polar solvents yields a ratio for **1:2** of about 30:1. The significantly lower dipole moment of the imino form (**2,** 5.55 D) relative to the amino form (**1,** 7.58 D) also suggests that the imino form will be relatively stabilised in non-polar solvents, or in the vapour phase. We will return to this topic in a future paper on the photoelectron spectra of cytosine derivatives.

(b) Conformational studies in the imino tautomer of 5-methyl-N(4)-hydroxycytosine

The previous section showed that the quantum mechanical studies for the relative stabilities of the amino *versus* imino forms are in general agreement with experiment for both cytosine and hydroxycytosine. The imino form of hydroxycytosine (**3**) in fact is alternatively named as the 4-oxime of uracil, formally arising from condensation of uracil with hydroxylamine. Thus, it seems probable that the imino compound (**3**) and its 5-methyl derivative could imitate the behaviour of uracil or thymine respectively in forming stable base pairs with adenine. This can only occur if the appropriate conformer is present, and especially if the hydrogen of the OH group is located in a position *anti* with respect to the ring nitrogen N(3). In order to predict the relative stabilities of various conformers of the imino tautomer mentioned above, we performed PCILO (and some MINDO-3) calculations for the rotational barrier of the OH group of the imino form of 5-methyl-N(4)-hydroxycytosine (Fig. 2). In these calculations we used the experimental bond lengths

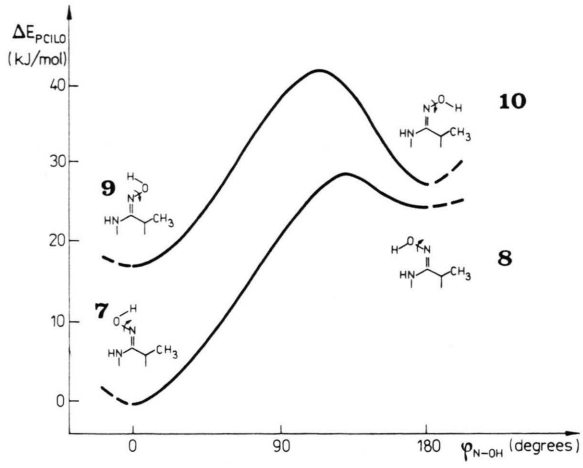

Fig. 2. Barrier for rotation of OH group about N – O bond calculated by PCILO, with partial geometry optimisation, for l and r conformers of the imino tautomer of 5-methyl-N(4)-hydroxycytosine.

and bond angles [6] of the molecule, except for five external bond angles, including those of the = N – OH group and its neighbour (H – N(4) – C(4), N(3) – C(4) – N(4), C(4) – C(5) – CH_3, C(4) – N(4) – O, N(4) – O – H), which were optimised.

The four conformers studied and their relative energies are shown in Fig. 2; it is convenient to use a fixed ring skeleton as in the oxime **3**, and to refer to the four conformers by their positions relative to the N(3) – C(4) bond; we thus have two *syn* forms (**7** and **8**) with transoid (**7**) and cisoid (**8**) C = N – OH fragments, and two *anti*-forms (**9** and **10**) with transoid (**9**) and cisoid (**10**) fragments. All four (**7** – **10**) are found to be local minima on the PCILO energy surface, with the *syn*, transoid form (**7**) being much lower in energy than the others by 17.15 (**8**), 24.68 (**9**) and 26.73 kJ mol^{-1} (**10**) respectively. The MINDO-3 calculations led to the same conclusions for **7** and **9**, with an energy difference of 12.38 kJ mol^{-1}, but led to the conclusion that the sterically overcrowded molecules (**10** and **8**) were nearly the same in energy as **9** and **7** (differences of 0.55 and 3.05 kJ mol^{-1} respectively). This appears to be a known weakness of the MINDO-3 method [26 – 29] for some conformational studies. However, both types of calculation do show that the *left* or cisoid conformers (**7** and **8**) are more stable than the *right* oriented pair (**9** and **10**). Since the former pair cannot have proton donor-acceptor properties similar to those of thymine (or uracil), it seems that N(4)-hydroxycytosine cannot be a complementary base with adenine. The H-bonded interactions of this molecule with a complementary base should be similar to those recently observed for O-methoxycytosine [30].

Acknowledgements

This work was supported in part by the Polish Ministry of Science and Higher Education under the project MR.I.5; we thank the British Council for support for J.S.K. B.L. thanks the Alexander von Humbolt foundation for a grant enabling him to perform some of the calculations in Göttingen.

[1] J. S. Kwiatkowski and B. Pullman, Adv. Heterocyclic Chem. **18,** 199 (1975). Particularly see pp. 220 – 221 of the paper.
[2] C. Janion and D. Shugar, Biochim. Biophys. Res. Commun., **18,** 617 (1965).
[3] C. Janion and D. Shugar, Acta Biochim. Polon. **15,** 108 (1968).
[4] E. I. Budovsky, E. D. Sverdlov, and T. N. Spasokukotskaya, FEBS Letters **17,** 336 (1971).
[5] R. A. Flavell, D. L. Sabo, E. F. Bandle, and C. Weismann, J. Mol. Biol. **89,** 255 (1974).
[6] D. Shugar, C. P. Huber, and G. U. Birnbaum, Biochim. Biophys. Acta **447,** 274 (1976).
[7] J. S. Kwiatkowski, B. Lesyng, M. H. Palmer, and J. W. Wheeler, in preparation.
[8] R. C. Bingham, M. J. S. Dewar, and D. H. Lo, J. Amer. Chem. Soc. **97,** 1285, 1294, 1302, 1307 (1975).
[9] M. J. S. Dewar and W. Thiel, J. Amer. Chem. Soc. **99,** 4899, 4907 (1977).
[10] S. Diner, J. P. Malrieu, F. Jordan, and M. Gilbert, Theoret. Chim. Acta (Berlin) **15,** 100 (1969).
[11] M. H. Palmer, R. H. Findlay, S. M. F. Kennedy, W. Moyes, and J. D. Nisbet, [in:] Quantum Chemistry – The State of the Art (V. R. Saunders and J. D. Brown, eds.), p. 229, Science Research Council, London 1975.
[12] M. H. Palmer, A. J. Gaskell, and M. S. Barber, Theoret. Chim. Acta (Berlin) **26,** 357 (1972).
[13] M. H. Palmer, J. R. Wheeler, R. H. Findlay, N. P. C. Westwood, and W. M. Lau, J. Mol. Structure, THEOCHEM, **86,** 193 (1981).
[14] M. H. Palmer, I. Simpson, and J. R. Wheeler, Z. Naturforsch. **36 A,** 1246 (1981).
[15] J. W. McIver, Jr. and A. Komornicki, J. Amer. Chem. Soc. **94,** 2625 (1972).
[16] J. W. McIver, Jr. and A. Komornicki, Chem. Phys. Letters **10,** 303 (1971).
[17] M. C. Flanigan, A. Komornicki, and J. W. McIver, Jr., in: Semiempirical methods of electronic structure calculation. Part B: Applications (G. A. Segal, ed.) p. 1, Plenum Press, New York 1977.
[18] Kindly supplied to us by Prof. M. Dupuis, via Dr. M. F. Guest (Daresbury Laboratory).

[19] D. G. Lister and J. K. Tyler, J. Chem. Soc., Chem. Commun., **1966,** 152; J. Mol. Structure, **23,** 253 (1974).
[20] J. S. Kwiatkowski and B. Pullman, Acta Phys. Polon. **A 45,** 693 (1974).
[21] T. J. Zielinski and R. Rein, Int. J. Quantum Chem. **14,** 851 (1978).
[22] H. Morita and S. Nagakura, Theoret. Chim. Acta (Berlin) **11,** 279 (1968).
[23] B. I. Sukhorukov, A. S. Gukovskays, L. N. Sukzoruchkina, and G. I. Lavrenova, Biofizika **17,** 5 (1972).
[24] R. Czerminski, B. Lesyng, and A. Pohorille, Int. J. Quantum Chem. **16,** 605 (1979).
[25] M. Dreyfus, O. Bensaude, G. Dodin, and J. E. Dubois, J. Amer. Chem, Soc. **98,** 6338 (1976).
[26] R. C. Bingham, M. J. S. Dewar, and D. H. Lo, J. Amer. Chem. Soc., **97,** 1294, 1302, 1307 (1975), and references cited therein.
[27] M. J. S. Dewar, D. H. Lo, and R. Ramsden, J. Amer. Chem. Soc. **97,** 1311 (1975), and references cited therein.
[28] J. A. Pople, J. Amer. Chem. Soc. **97,** 5306 (1975).
[29] W. J. Hehre, J. Amer. Chem. Soc., **97,** 5308 (1975).
[30] A. Psoda, B. Kierdaszuk, A. Pohorille, M. Geller, J. T. Kusmierek, and D. Shugar, Int. J. Quantum Chem., **20,** 543 (1981).