Theoret. Chim. Acta (Berl.) 39, 1—6 (1975)

A Note on a Non-Empirical Molecular Orbital Study of Some Cytosine and Thymine Tautomers

John D. Goddard, Paul G. Mezey, and Imre G. Csizmadia
Department of Chemistry, University of Toronto

Received January 17, 1975/March 26, 1975

The relative stabilities of a series of cytosine and thymine tautomers have been investigated by means of non-empirical *(ab initio)* LCAO-MO-SCF Hartree-Fock-Roothaan calculations employing a small contracted Gaussian basis set (STO-3G) in all cases. The relative stabilities of the various tautomeric forms agree in general with the results of earlier empirical and semi-empirical calculations on these molecules. In particular, the "enol forms" of cytosine and thymine, within the isolated molecule approximation inherent in these MO calculations, are predicted to possess greater stability than the forms commonly assigned to these molecules in aqueous solution and found in the solid state by X-ray crystallographic studies.

Key words: Cytosine tautomers—Thymine tautomers

The question of the relative stabilities of the tautomers of pyrimidines and purines is clearly of some interest both chemically and biologically. Extensive semi-empirical molecular orbitals studies [1–4] have provided some insight into the problem and it was felt that an extension to the non-empirical level was warranted. Although useful, the semi-empirical calculations all suffer from the fact that in any particular case the effects of the approximations inherent in such methods are difficult to judge.

Cytosine and thymine were chosen for the first part of a non-empirical study of the tautomerism of purines and pyrimidines. The electronic absorption spectra of cytosine and thymine in various solvents have *not* been unambiguously assigned as to the tautomer present with the aid of Pariser-Parr-Pople SCF-CI pi-electron calculations [5, 6]. Simple Hückel MO calculations suggest that cytosine should have the greatest tendency to exist in rare forms [1] and a CNDO/2 study of its tautomerism has been published [4a].

LCAO-MO-SCF Hartree-Fock-Roothaan calculations were performed using the GAUSSIAN 70 series of programs [7]. The choice of basis set was limited to a minimal STO-3G basis by the large size (cytosine 58 electrons, thymine 66 electrons) of the molecules. The standard orbital exponents optimized by Pople *et al.* [8] for a large number of small polyatomic molecules were employed throughout as was the $\alpha_s = \alpha_p$ orbital exponent constraint on the radial functions. Geometry optimization for such large systems was clearly impractical and geometries designed on the basis of chemical analogies were employed throughout. In the absence of extensive geometry optimization which was not economically

Table 1. Comparison of non-empirical calculations on cytosine and thymine

		This work STO-3G	Clementi et al Minimal	Mely-Pullman Minimal	Experimental
Cytosine	Total Energy[a]	-387.51625	-390.93564	-	-
	Ionization Potential[b]	5.80	9.83	9.79	8.90[d] ± 0.02
	Dipole Moment[c]	5.74	6.40	6.76	-
Thymine	Total Energy[a]	-445.62394	-449.59107	-	-
	Ionization Potential[b]	6.44	10.54	10.16	9.43[d] ± 0.1
	Dipole Moment[c]	2.98	3.29	4.22	-

[a] In hartree, 1 hartree = 627.71 kcal mole^{-1}. [b] In eV, by Koopmans' theorem.
[c] In Debyes, 1 a.u. = 2.54154 Debye. [d] Ref. [13].

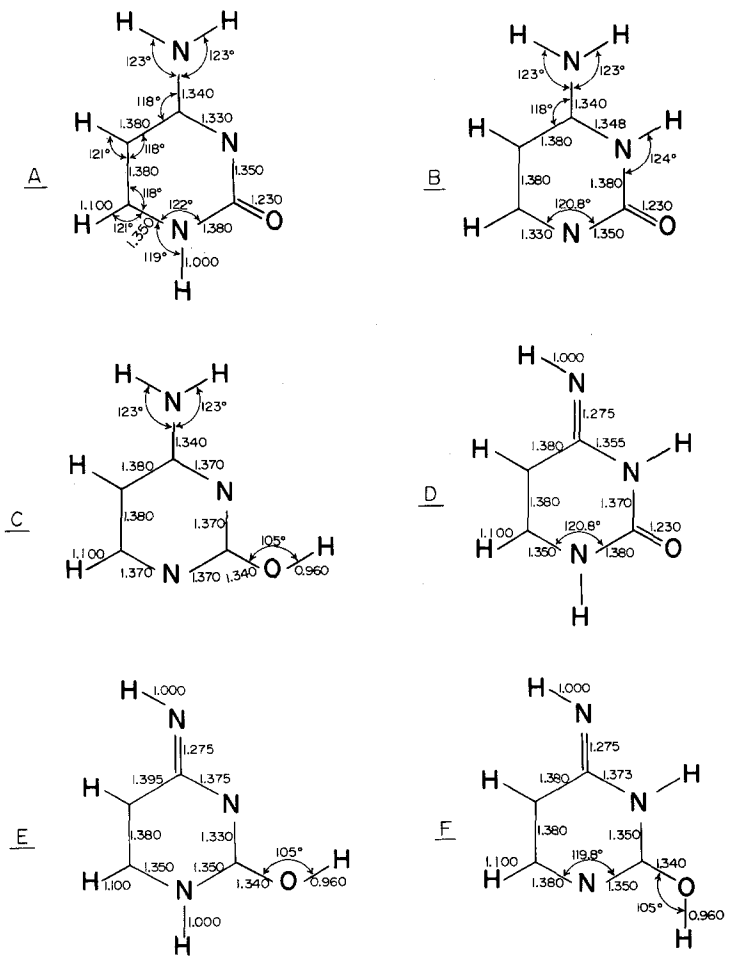

Fig. 1A–F. Geometries of the cytosine tautomers. All bond lengths in Å, 1 a.u. = 0.52917 Å. All bond angles in degrees

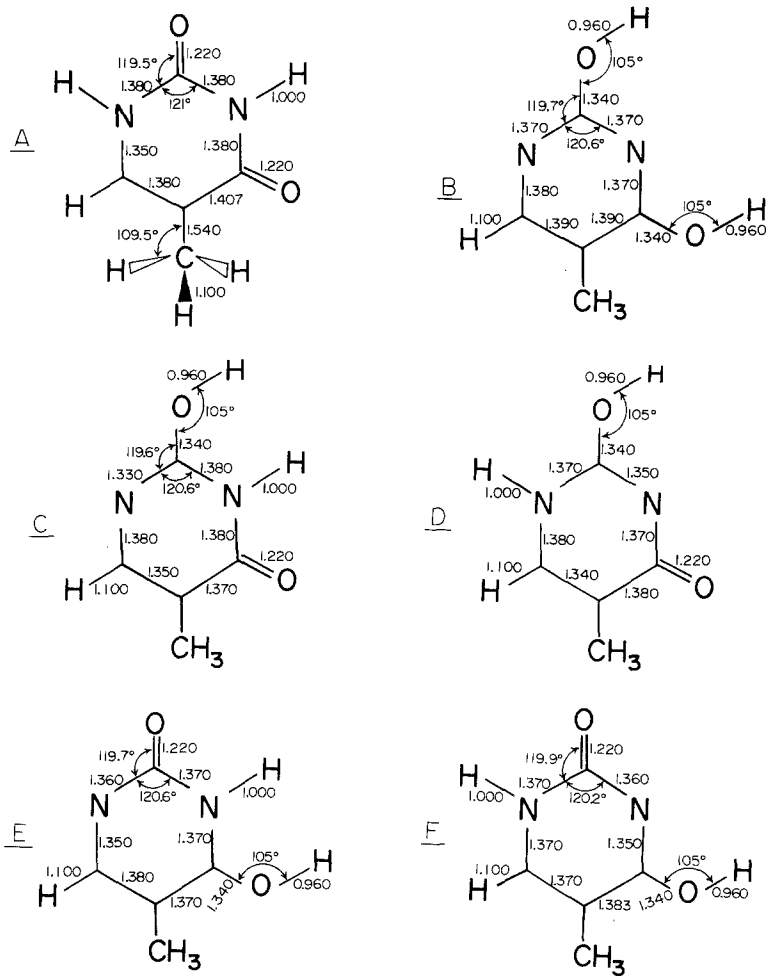

Fig. 2A–F. Geometries of the thymine tautomers. All bond lengths in Å, 1 a.u. = 0.52917 Å. All bond angles in degrees

feasible such a procedure has been adopted. Though the results must be viewed with considerable caution, a preliminary study of formamide tautomerism indicated that the chemically intuitive structures and the geometry optimized ones gave an energy difference between formamide tautomers that was constant to within approximately 1 kcal mole^{-1}.

The "keto" tautomers; form A of cytosine and form A of thymine were the object of previous non-empirical calculations by Clementi *et al.* [9] employing a 7^s- and 3^p-type Gaussian set on carbon, nitrogen, and oxygen and a 3^s-type Gaussian set on hydrogen all contracted to a minimal basis. Mely and Pullman [10] reported results using a 4^s plus 2^p heavy atom set and a 3^s hydrogen basis again contracted to a minimal basis. The results for the "keto" tautomers in this work are compared to these previous results in Table 1. The geometries for these calculations were taken from a survey of X-ray crystallographic data [11].

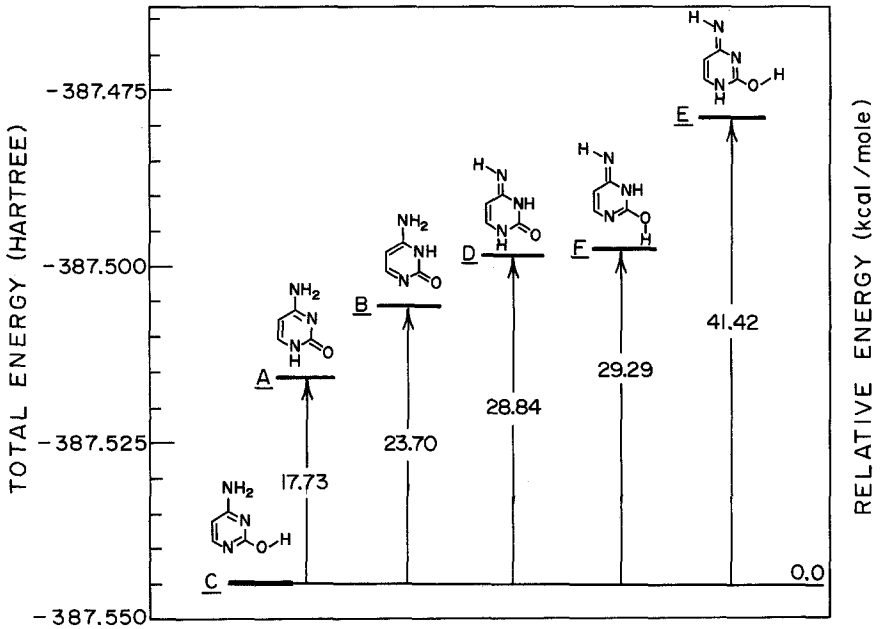

Fig. 3A–E. Total and relative energies of the cytosine tautomers

In designing geometries for cytosine and thymine by analogy to similar molecules certain structural features were assumed: all CH bonds were assigned a length of 1.100 Å, all NH bonds 1.000 Å; all C=O bonds were set equal to 1.220 for cytosine and to 1.230 Å for thymine by reference to forms A and to the earlier formamide study; all exocyclic C=N bonds were assigned the optimum length found for the formamide tautomer, i.e. 1.275 Å; OH bond lengths were taken as 0.960 Å by reference to methanol [12] all C–O bond lengths were set al 1.340 Å with reference to resorcinol [12]; in general, sp^2 hybridized atoms were assigned idealized valence angles at 120° (see Figs. 1 and 2); the COH bond angle was chosen to be 105°; and the tautomers were all constrained to be planar (with the obvious exception of the methyl hydrogens in thymine). With the above assumptions, the geometry problem was reduced to assigning bond lengths to the six-membered ring and this was achieved by reference to the known structures of the keto forms for both molecules. The geometrical information is summarized in Figs. 1 and 2. The stereochemistry and conformation, where relevant, is as illustrated.

The total electronic, nuclear repulsion, and total energies in hartree, the relative energies in kcal $mole^{-1}$ (relative to the calculated most stable tautomer) and the first ionization potentials (in electron volts) as given by Koopmans' theorem are presented in Table 3 and illustrated in Figs. 3 and 4. For an isolated cytosine molecule the "enol" tautomer (C) is predicted to be some 18 kcal $mole^{-1}$ more stable than the "keto" tautomer (A). This prediction of enhanced stability of the enol form relative to the keto is in agreement with the CNDO/2 study of Breen and Flurry even though these authors have assumed rather different

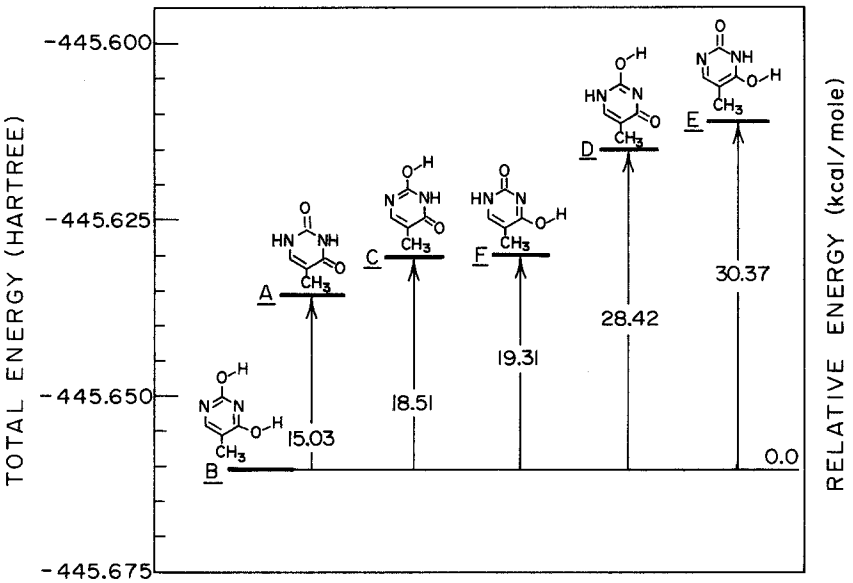

Fig. 4A–E. Total and relative energies of the thymine tautomers

geometries than those of the present work (i.e. the relative stability of enol over keto does *not* seem to be *merely* an artefact of the design of geometries). Breen and Flurry have also shown that the assignment of the absorbing species in the

Table 2. Energies and ionization potentials[a, b] of (a) the cytosine tautomers; (b) the thymine tautomers

(a) Cytosine

TAUTOMER	TOTAL ELECTRONIC ENERGY	NUCLEAR REPULSION	TOTAL ENERGY	RELATIVE ENERGY	IONIZATION POTENTIAL
E	-744.98760	357.50920	-387.47840	41.42	5.45
F	-744.73443	357.23670	-387.49773	29.29	5.47
D	-748.55944	361.06099	-387.49845	28.84	6.26
B	-746.63340	359.12677	-387.50663	23.70	5.42
A	-747.05308	359.53692	-387.51615	17.73	5.80
C	-740.15936	352.61497	-387.54439	00.00	6.80

(b) Thymine

TAUTOMER	TOTAL ELECTRONIC ENERGY	NUCLEAR REPULSION	TOTAL ENERGY	RELATIVE ENERGY	IONIZATION POTENTIAL
E	-886.88166	441.26986	-445.61180	30.37	5.78
D	-886.02080	440.40589	-445.61490	28.42	5.99
F	-885.98319	440.35378	-445.62941	19.31	6.00
C	-886.45866	440.82796	-445.63069	18.51	5.98
A	-888.13776	442.50153	-445.63623	15.03	6.53
B	-879.52952	433.86934	-445.66018	00.00	7.14

[a] Total energies in hartree. Relative energies in kcal mole^{-1}. Ionization Potentials in eV.
[b] For tautomer labels see Figs. 3 and 4.

UV spectra in water and in acetonitrile on the basis of Pariser-Parr-Pople SCF-CI pi-electron calculations is not without ambiguity and that the forms C, A, and D all yield rather similar theoretical spectra. (See also Ref. [6].) The order of stability of the other tautomers is that expected in that the other keto form B lies rather close to A while the three imine forms (the most localized structures in valence bond terms) are of higher energy. The total and relative energies of the thymine tautomers are presented as for cytosine above. To be more consistent with the structures assigned to the other tautomers a second calculation was performed on form A at a geometry designed in a similar fashion. This fact accounts for the apparent discrepancy in the energies for tautomer A in Tables 1 and 2. For an isolated molecule of thymine the "enol" form is predicted to be some 15 kcal $mole^{-1}$ more stable than the "keto" form.

This study predicts that for the isolated molecules, the fully aromatic "enol" tautomers, form C of cytosine and form B for thymine are more stable than the "keto" structures, A and B for cytosine and A for thymine. A CNDO/2 study [4] with different geometries also predicted a more stable "enol" form in the case of cytosine. Experimental structural results for cytosine and thymine in the gas phase are not available and would provide the most suitable comparison with these calculated results. In the solid state X-ray crystallographic studies show the "keto" forms to be more stable [11, 14]. The UV spectra in various solvents do not provide conclusive evidence as to the absorbing species present [5, 6].

Acknowledgements. The authors are deeply indebted to the National Research Council of Canada for continuing financial support.

References

1. For a review of π-electron methods see: Pullman, A., Pullman, B.: Advan. Quantum Chem. **4**, 267 (1968)
2. Giessner-Prettre, C., Pullman, A.: Theoret. Chim. Acta (Berl.) **9**, 242 (1968)
3. Pullman, A., Kochanski, E., Gilbert, M., Denis, A.: Theoret. Chim. Acta (Berl.) **10**, 231 (1968)
4. (a) Breen, D. L., Flurry, R. L.: Theoret. Chim. Acta (Berl.) **13**, 138 (1971)
   (b) Fujuta, H., Imamura, A., Nagata, C.: Bull. Chem. Soc. Japan **42**, 1467 (1969)
5. Tanaka, M., Nagakura, S.: Theoret. Chim. Acta (Berl.) **6**, 320 (1966)
6. Morita, H., Nagakura, S.: Theoret. Chim. Acta (Berl.) **11**, 279 (1968)
7. Hehre, W. J., Lathan, W. A., Ditchfield, R., Newton, M. D., Pople, J. A.: GAUSSIAN 70 QCPE Program No. 236, Quantum Chemistry Program Exchange. Indiana University, Bloomington, Indiana
8. Hehre, W. J., Stewart, R. F., Pople, J. A.: J. Chem. Phys. **51**, 2657 (1969)
9. Clementi, E., Andre, J. M., Andre, M. Cl., Klint, D., Hahn, D.: Hungarian Phys. Acta **27**, 493 (1969). (Volume in Honour of Professor Gómbas)
10. Mely, B., Pullman, A.: Theoret. Chim. Acta (Berl.) **13**, 278 (1969)
11. Spencer, M.: Acta Cryst. **12**, 59 (1959)
12. Bowen, H. J. M.: Tables of interatomic distances and configuration in molecules and ions. Chem. Soc. (London) Spec. Publ. (1958)
13. Lifschitz, C., Bergmann, E. D., Pullman, B.: Tetrahedron Letters **46**, 4583 (1967)
14. Barker, D. L., Marsh, R. E.: Acta Cryst. **17**, 1581 (1964)
    Gerdil, R.: Acta Cryst. **14**, 333 (1961)

Prof. Dr. I. G. Csizmadia
Department of Chemistry
University of Toronto
80 St. George St.
Toronto, Ontario, Canada, M5S 1A1