Theoret. chim. Acta (Berl.) 23, 138—146 (1971)

Semi-Empirical Molecular Orbital Calculations on Some Tautomers of Cytosine

DONNA L. BREEN and R. L. FLURRY, Jr.
Department of Chemistry, Louisiana State University in New Orleans, New Orleans, Louisiana 70122, USA

Received February 22, 1971

The stability of a series of cytosine tautomers has been investigated using both all valence electron and π-electron SCF methods. The effect of general and selective scaling of the bond lengths was studied in the CNDO/2 calculations. The enol tautomer of cytosine exceeded the keto form in stability in these calculations.

π-electron SCF CI calculations were performed on the cytosine tautomers using two different parameterization schemes. The calculated transitions obtained for the enol form showed a close proximity to that of the keto and imino structures. Comparison was made with the experimental electronic spectrum in water and acetonitrile. No clear distiction between the keto and enol forms could be made on the basis of these calculations.

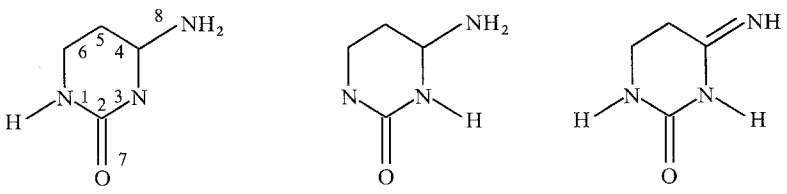
Die Stabilität einer Reihe von Cytosin-Tautomeren wurde mit einer Allvalenzelektronen- und mit einer π-Elektronen-SCF-Methode untersucht. Die Auswirkung einer allgemeinen als auch selektiven Skalierung der Bindungslängen wurde bei den CNDO/2-Berechnungen untersucht. Das Enoltautomere des Cytosins übertrifft nach diesen Rechnungen die Ketoform an Stabilität. Bei den π-Elektronen-SCF-CI-Rechnungen wurden zwei Methoden der Parametrisierung angewandt. Die für die Enolform berechneten Übergänge zeigten eine große Ähnlichkeit zu denjenigen der Keto- und Iminostrukturen. Die berechneten Werte wurden mit dem experimentellen Spektrum von Cytosin in Wasser sowie Acetonitril verglichen. Auf Grund der Berechnungen konnte keine klare Unterscheidung zwischen der Keto- und der Enolform getroffen werden.

Etude de la stabilité d'une série de tautomères de la cytosine par les méthodes SCF à électrons de valence et à électrons π. L'effet d'ajustement global ou selectif des longueurs de liaison a été étudié dans le cadre CNDO/2. Le tautomère énolique de la cytosine s'avère dans ces calculs plus stable que la forme cétonique. Des calculs d'électrons π SCF CI ont été effectués sur les tautomères de la cytosine en utilisant deux schémas différents pour la paramétrisation. Les transitions calculées obtenues pour la forme énolique montrent une grande analogie avec celles des structures cétoniques et iminiques. La comparaison a été faite avec le spectre électronique expérimental dans l'eau et l'acétonitrile. Aucune distinction claire entre les formes cétoniques et énoliques n'a pu être faite à partir de ces calculs.

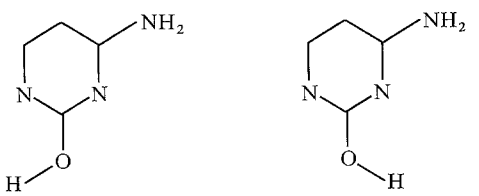
Introduction

It has recently been suggested that carcinogenesis may be produced by the RNA polymerase virus forming mutant DNA genes. Mutation in nucleic acid has been the subject of much experimental and theoretical research. The majority of the theoretical investigations have centered on the stabilities of various tautomers of the bases and base pairs known to form the inner structure of the helical nucleic acids.

This paper reports a molecular orbital study of cytosine and several of its tautomers. Cytosine is a pyrimidine-type base capable of existing in a number



Form A Form B Form C



Form D

Fig. 1. Some tautomeric structures for cytosine

Table 1. *Experimental bond distances and angles for anhydrous cytosine*[a]

Bond	Distance	Bond Angle	
N_1–C_2	1.374 Å	N_1–C_2–N_3	118.1°
C_2–N_3	1.364	C_2–N_3–C_4	119.9
N_3–C_4	1.337	N_3–C_4–C_5	122.0
C_4–C_5	1.424	C_4–C_5–C_6	117.3
C_5–C_6	1.352	C_5–C_6–N_1	120.1
C_6–N_1	1.357	C_6–N_1–C_2	122.7
C_2–O_7	1.234	N_1–C_2–O_7	119.8
C_4–N_8	1.330	N_3–C_2–O_7	122.2
		N_3–C_4–N_8	118.2
		C_5–C_4–N_8	119.9

[a] Ref. [1].

of different tautomeric forms. The structures studied here are given in Fig. 1. The experimental geometry [1] is as in Table 1. Form *A* is the tautomer assumed to be present in molecules of DNA and RNA and that experimentally found in anhydrous crystalline cytosine [1]. Other possible structures in nuclei acids would include forms *B* and *C*. Much evidence for the existence of tautomerism in cytosine and its nucleosides comes from NMR measurements [2, 3]. Other evidence arises from the x-ray structure of cytosine recrystallized from various solvents. The molecules in the crystals take on different tautomeric forms depending upon the solvent; Form *A* from methanol [1], Form *B* from water [4], and a zwiterion for cytosine-5-acetic acid [5].

Morita and Nagakura [6] investigated the tautomerization of cytosine, first by means of electronic spectra in different solvents, and then by the P-P-P π-electron SCF method with CI. Three tautomers studied by them are included in this work; forms *A*, *B*, and *C*, along with one form they did not consider.

The object of this work is to determine, by use of the CNDO/2 all valence electron [7] and the Pariser, Parr, Pople (PPP) π-electron [8, 9] calculations, and by comparison with the spectral study of Morita and Nagakura, the most probable tautomeric form for cytosine.

Methods

The usual CNDO/2 approximations and parameters were chosen for the present calculations [10]. These included the core Coulomb integrals and core resonance parameters suggested by Pople and co-workers. Effective orbital exponents were taken from Slater's rules. The exponent for hydrogen was taken as 1.2.

Calculations on simple polyatomic molecules have shown that the CNDO method reproduces equilibrium bond angles with generally good agreement. The equilibrium bond distances are usually somewhat too large. Its treatment of total energies, dissociation energies, and ionization energies can only be considered as indicating general trends and are by no means quantitative. These factors should be taken into consideration whenever the stability of tautomeric structures is the sought-after information.

For the P-P-P π-electron calculations, the method and program used were described by Flurry and Bell [11]. Two sets of parameterization were chosen. The first set was identical to that given by Morita and Nagakura [6]. Their valence state ionization potentials for carbon, nitrogen, and oxygen $2p$ orbitals were adjusted to make the calculated transitions duplicate the observed values. The second parameterization scheme was that of Flurry and Bell [11].

Results and Discussion

CNDO/2 *Calculations*

The most stable form of cytosine, according to most investigators, is Form *A*. This is in agreement with the x-ray diffraction data of Barker and Marsh [1] on the crystal structure of anhydrous cytosine. Although other tautomers are known to exist in other crystals [4, 5] and may exist in solution in various solvents, the predominant form in most cases is assumed to be *A*. This form is also that proposed for DNA and RNA.

Using Barker and Marsh's geometry for the heavy atoms on a series of cytosine tautomers, the CNDO/2 method gives their total energies as in Table 2.

Table 2. CNDO/2 *total energies of cytosine tautomers without general coordinate scaling (energies in* eV*)*

Form	Unscaled	C–N scaled	C–O scaled
A	−2327.388	−2327.521 (1.01)	−2327.634 (1.04)
B	−2326.748	−2326.879 (1.01)	−2326.998 (1.04)
C	−2326.486	−2326.499 (0.99)	−2326.701 (1.04)
D	−2327.412	−2327.563 (1.02)	−2328.582 (1.09)

The lowest total energy form is that tautomer with a protonated oxygen, Form *D* or the enol of Form *A*. We calculated both enol isomers, that is, with the hydroxyl proton on either side of the oxygen, and found their total energies equivalent to 10^{-3} eV.

The acknowledged stable form, *A*, is next, with the imino form, *C*, the highest in energy. Calculations were also performed on nonplanar tautomers, having out of plane hydrogens on the amino group. These structures showed an order of stability identical to the planar forms and were therefore omitted from further calculations for simplicity.

It is probably naive to think that each tautomer would retain the same geometries as that found for the keto form in the anhydrous crystal. Since the necessary computer time to treat each bond independently was not available, it was decided to select the two bonds which should most effect the energy and scale them, retaining the angular geometries. The position vectors of these atoms were multiplied by a scaling factor which was varied from 0.95 to 1.15. The two bonds selected were the extracyclic C–O and C–N bonds.

The results from this series of scaled bond vectors are given in the second column of Table 2. They show the same order of stability as the unscaled molecules. The scaling parameters for the energy minimum are given in parentheses. These factors reflect changes in the structure upon tautomerization. For example, the C–N bond has shortened slightly in the imino form, Form *C*, reflecting double bond character. The C–O bond has lengthened considerably to 1.09 times its original length to lower the energy of the enol tautomer, Form *D*.

Taking into account what seems, in these compounds, to be a general lengthening of calculated bond lengths in the CNDO/2 method presented a problem. It was decided to uniformly scale all of the coordinates for each tautomer and find the single scaling parameter minimizing the energy. Taking this overall scaling factor into account should compensate for some of the deficiency of the method in this respect and make succeeding calculations on the equilibrium C–O and C–N bonds more reliable. The overall scaling factor for the coordinates was the same for each tautomer considered. The energy minimum appeared at a scale factor of 1.02 for all structures. The values are given in Table 3 in the first column. The lowest energy form is now changed to Form *A*, the keto

Table 3. CNDO/2 *total energies of cytosine tautomers with general coordinate scaling (energies in* eV*)*

Form	Scaled	C–O scaled	C–N scaled	C–O and C–N min. then rescaled
A	−2327.933 (1.02)	−2327.979 (1.02)	−2327.933 (1.00)	−2327.979 (1.00)
B	−2327.298 (1.02)	—	—	—
C	−2327.098 (1.02)	−2327.126 (1.01)	−2327.269 (0.99)	−2327.296 (1.00)
D	−2327.869 (1.02)	−2328.516 (1.07)	−2327.876 (1.01)	−2328.735 (0.99)

isomer. The enol form, *D*, is very close to it but higher in energy by 0.06 eV. The other two tautomers, Forms *B* and *C*, retained the same order as in the unscaled isomers.

The identical scaling factor of 1.02 obtained for each tautomer suggested scaling the overall geometries by this parameter and then minimizing the C–O and C–N bond lengths. In Table 3, the second and third columns, are given the results of the scaling of the C–O and C–N position vectors after the general scaling. In the enol form *D*, the C–O distance lengthened to 1.07 times its distance in the crystal structure whereas, for the keto and imino forms, *A* and *C*, the energy was minimized at C–O scale factors of 1.02 and 1.01. The order of stability was the same as in the tautomers with no adjusted bond lengths. For the C–N scaling, there were only slight changes in the bonds to produce stability for Forms *C* and *D* and no change whatsoever for Form *A*. However, the order of the total energies has changed from the C–O scaling. This order is now the same as in column 1 for the totally scaled tautomers. This means that the experimentally predicted stable form of cytosine, Form *A*, becomes the lowest energy form only when only the C–N bond is scaled.

To test the effect on the total energies of minimizing both extracyclic bonds, another calculation was made on the tautomers. The fourth column of Table 3 shows the results of scaling both C–O and C–N bonds by the factors of columns two and three. A general overall scaling was again applied; however, the energies tended to be satisfied with the new geometries and to be effectively minimized with respect to both C–O and C–N bonds. This is evidenced by the fact that the energy minima hovered at a scale factor of 1.00, suggesting that this is probably the lowest obtainable from this kind of approximation.

SCF MO CI *π-Electron Calculations*

Morita and Nagakura [6] concluded from the absorption spectra in water and in acetonitrile that there were actually two tautomers present. Form *A* was suggested as the predominant form in H_2O and Form *C* as that in CH_3CN.

To support this conclusion, the P-P-P SCF method with CI was used to calculate transition energies and oscillator strengths. Good agreement with the observed values was obtained; however, the parameters for the VSIP's were varied to reproduce the observed transition energies. This tends to reduce the validity of this theoretical support, particularly in view of the fact that the predicted and experimental transitions for all the tautomers considered were rather close.

Calculations were performed here on two of the tautomers of Morita and Nagakura, Forms *A* and *C* of Fig. 1, with a third, Form *D*; the structure found to be most stable in the majority of the present CNDO/2 calculations.

The three tautomers were subjected to two sets of calculations. In first set, the adjusted parameterization scheme of Morita and Nagakura was used. Nine singly excited configurations were employed, as compared to their eight, for calculating the excited states.

Table 4 gives the transition energies and oscillator strengths from this set of calculations. Tautomers with both minimized and unminimized geometries were calculated. The transition energies are very similar, however, the oscillator

Table 4. *π-electron calculations of transition energies and oscillator strengths for some cytosine tautomers using parameterization scheme of Ref.* [6]

Form	ΔE (eV)			f		
	unscaled[a]	scaled[b]	M & N[c]	unscaled[a]	scaled[b]	M&N[c]
A	4.7243	4.7078	4.708	0.0390	0.0367	0.075
	5.4570	5.4393	5.479	0.0205	0.0161	0.011
	6.2934	6.2126	6.432	0.6645	0.6533	1.370
	6.6947	6.6400	6.681	0.1005	0.0899	0.163
	7.7953	7.6804	7.700	0.2004	0.2183	0.458
C	4.4904	4.5058	4.554	0.1366	0.1340	0.224
	5.3022	5.3408	5.445	0.2934	0.2949	0.328
	5.5651	5.5630	5.684	0.1566	0.1679	0.497
	6.0524	6.0118	6.222	0.2646	0.2477	0.698
	7.7628	7.7858	7.917	0.0227	0.0212	0.043
D	4.7724	4.7662	–	0.0508	0.0426	–
	5.3992	5.3457	–	0.0608	0.0625	–
	6.6660	6.6622	–	0.4532	0.4671	–
	6.9341	6.8993	–	0.4075	0.4140	–
	7.6671	7.6846	–	0.0471	0.0444	–

[a] Atomic coordinates are as given in Ref. [1].
[b] Distances for the C–O and C–N extracyclic bonds have been scaled according to factors from columns 2 and 3 of Table 2.
[c] Values taken from Ref. [6].

Table 5. *π-electron* SCF *transition energies and oscillator strengths using parameterization scheme of Flurry and Bell* [11]

Form	ΔE		f	
	unscaled[a]	scaled[b]	unscaled[a]	scaled[b]
A	4.3606	4.3093	0.0708	0.0667
	5.6991	5.6561	0.0508	0.0696
	6.3428	6.2917	0.5698	0.5428
	7.1469	7.0620	0.0762	0.0708
	7.1872	7.1119	0.0116	0.0211
C	5.0221	5.0366	0.2206	0.2014
	5.5276	5.5445	0.2156	0.2291
	5.9936	6.0122	0.0434	0.0477
	6.6235	6.5945	0.2399	0.2637
	6.9604	6.9657	0.0958	0.0718
D	4.6504	4.6787	0.0590	0.0525
	5.6785	5.6742	0.1245	0.1303
	6.5015	6.5257	0.4016	0.4011
	6.9695	6.9838	0.2807	0.2946
	7.4444	7.4749	0.1311	0.1276

[a] Atomic coordinates for each tautomer are as given in Barker and Marsh [1].
[b] Distances for the C–O and C–N extracyclic bonds have been scaled according to factors from columns 2 and 3 of in Table 2.

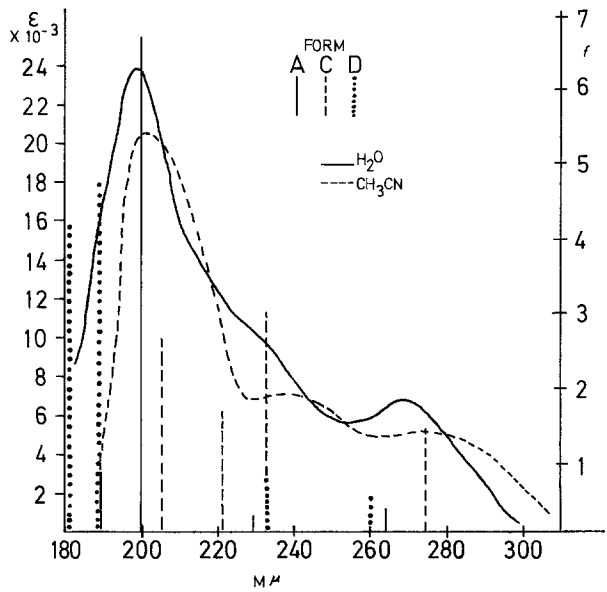


Fig. 2. Absorption spectrum of cytosine in H_2O and acetonitrile (from [6]). Vertical lines represent calculated (π-electron SCF CI) transitions. Parameters of Morita and Nagakura [6]

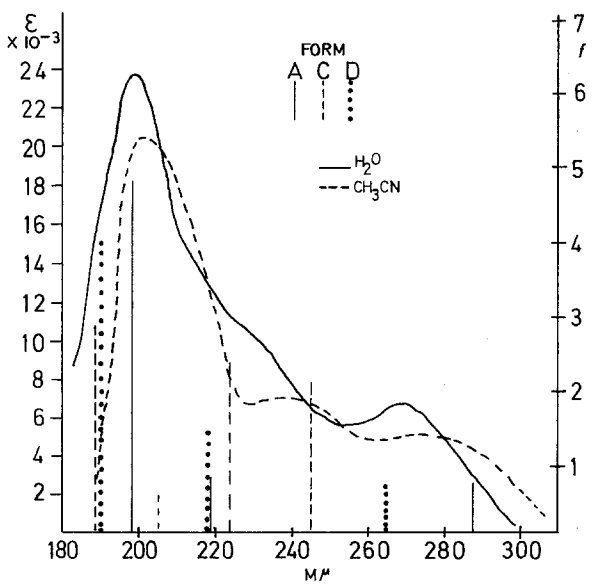


Fig. 3. Absorption spectrum of cytosine in H_2O and acetonitrile (from [6]). Vertical lines represent calculated (π-electron SCF CI) transitions. Parameters of Flurry and Bell [11]

strengths are different due to the extra CI singlet state. What is important here is that the enol tautomer, Form *D*, shows transition energies near those of the other two forms.

The results from the second set of calculations are in Table 4. For this, the Flurry and Bell approximations [11] were used. The transition energies and oscillator strengths were again calculated within the π-electron SCF framework.

Fig. 2 compares the experimental spectra with the calculated transitions using the Morita and Nagakura parameters. The values are taken from Table 4, using scaled tautometer geometries. Line heights are an indication of calculated oscillator strengths. This clearly suggests that the enol tautomer, Form *D*, could be present in H_2O solution.

The same possibility is present if the Flurry and Bell parameterization scheme is used. Fig. 3 shows the results. The calculated transitions for the scaled tautomers are again given by the vertical lines along with the experimental absorption curves of cytosine in H_2O and CH_3CN.

Conclusions

These calculations show the danger of drawing conclusions as to the stability of the tautomers of cytosine from calculations based only on the experimental crystal geometry of the keto tautomer. Using this geometry and the CNDO/2 method, one cannot rule out the possibility of finding a tautomer of higher stability than Form *A*.

It is particularly prevalent from the CNDO/2 results that allowing the C–O bond distance to increase, causes the keto tautomer, Form *A*, to become exceeded in stability by the enol form. In turn, the stability of Form *A* is preserved over the other tautomers only if the C–N bond distance is scaled without scaling the C–O bond.

From the π-electron calculations, it can be seen that the calculated transitions of the enol form of cytosine agree with the experimental spectrum of cytosine in H_2O solution to roughly the same extent as the keto isomer suggested previously. This is evident from both adjusted and non-adjusted parameterization calculations.

Due to the limitations of the semi-empirical methods employed, we cannot say with certainty that our Form *D* is the most stable tautomer of free cytosine. We do, however, feel that there is sufficient evidence to warrent further experimental investigations into the matter. On the other hand, the energies of the other tautomers which can exist in the nucleic acids (where position 1 is bound to the sugar of the backbone) are sufficiently high that form *A* can probably be reliably assigned as the structure of cytosine in these acids.

References

1. Barker, D.L., Marsh, R.E.: Acta Crystallogr. **17**, 1581 (1964).
2. Kokko, J.P., Goldstein, J.H., Mandell, L.: J. Amer. chem. Soc. **83**, 2909 (1961).
3. Miles, H.T.: J. Amer. chem. Soc. **85**, 1007 (1963).
4. Jeffrey, G.A., Kinoshita, Y.: Acta Crystallogr. **16**, 20 (1963).
5. Marsh, R.E., Bierstedt, R., Eichhorn, E.L.: Acta Crystallogr. **15**, 310 (1962).

6. Morita, H., Nagakura, S.: Theoret. chim. Acta (Berl.) **11**, 279 (1968).
7. Pople, J. A., Segal, G. A.: J. chem. Physics **44**, 3289 (1966).
8. Pariser, R., Parr, R. G.: J. chem. Physics **21**, 466 (1953).
9. Pople, J. A.: Trans, Faraday Soc. **49**, 1375 (1953).
10. — Beveridge, D. L.: Approximate molecular orbital theory. New York: McGraw-Hill, 1970. 1970.
11. Flurry, Jr., R. L., Bell, J. J.: J. Amer. chem. Soc. **89**, 525 (1967).

Prof. R. L. Flurry
Department of Chemistry
Louisiana State University in New Orleans
Lake Front
New Orleans, Louisiana 70122, USA