Scientific paper

Theoretical Modeling of the Ground State Intramolecular Proton Transfer in Cytosine: DFT Level Study

Vassil B. Delchev* and Maria V. Nenkova

Dept. Physical Chemistry, University of Plovdiv, Tzar Assen Str. 24, BG-4000 Plovdiv, Bulgaria

** Corresponding author: vdelchev@uni-plovdiv.bg*

Received: 19-06-2007

Abstract

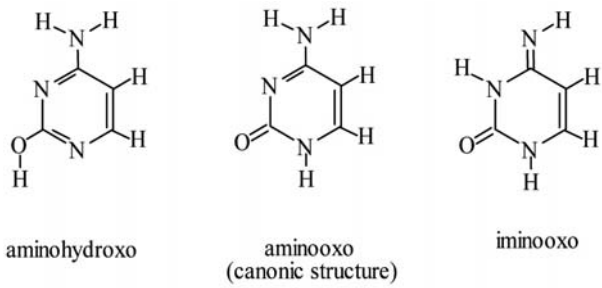
Five isomers of cytosine and their mutual interconversions were studied theoretically at the B3LYP level using basis sets 6-31G and 6-311G and a different number of polarization and diffuse functions. It was demonstrated that the canonic aminooxo tautomer of cytosine is the most stable one. However it has a non-planar geometry. It was shown that the energies and energy barriers of the studied systems are sensitive to the inclusion of polarization functions in the basis set, but they have lesser sensitivity toward inclusion of diffuse functions.

Keywords: Cytosine, density functional calculations, intramolecular proton transfer

1. Introduction

Nitrogen bases are major structural units of nucleic acids. These are of two types: purine and pyrimidine derivatives. Nucleic acid bases in the DNA structure are linked through H-bonds, as in this way they conserve and code the genetic information in living world. Each nucleic acid base can exist in several tautomeric forms but for their biological function only one tautomer is of great importance.

The experimental analysis of the cytosine (Cyt) tautomeric forms has revealed that this base exists in several tautomeric forms, as the most stable are aminooxo and aminohydroxo ones,[1–3] as shown in Scheme 1.



Scheme 1. Experimentally detected tautomers of cytosine[1–5]

These results have been reconfirmed by means of theoretical calculations on the cytosine tautomeric forms.[4,6,7] The presence of small amounts of iminooxo tautomer has been experimentally proved.[4,5] In water solution only the aminooxo cytosine tautomer exists.[8–14] In the DNA macromolecule only the aminooxo form takes part in an H-bonding with guanine,[15] however if the iminooxo cytosine is available it links through H-bonds to adenine.[16] This causes mutations in the DNA structure with all further consequences.

Topal et al.[17] have found that the equilibrium constant of the transformation aminooxo Cyt ⇄ iminooxo Cyt are within the interval 10^{-4}–10^{-5}. This equilibrium between the two tautomeric forms is entirely reachable during the DNA synthesis in cells,[16] and therefore the probability for mutations is very large.

In the scientific literature there is a gap concerning the mechanisms (transition states, energy barriers etc.) of mutual transformations of cytosine tautomers. Therefore, the purpose of the current paper is to throw light upon the mechanism of tautomeric equilibria, using the tools of computational chemistry, in particular the DFT hybrid functional B3LYP. It has been demonstrated that the B3LYP functional yielded accurate normal mode frequencies compared with experiment.[18] Moreover, the B3LYP functional has given geometries of several aromatic systems which agree well with experiment.[19,20]

2. Computational Details

The geometries of the five isomers were optimized at the B3LYP level using the standard gradient procedure with no symmetry restrictions. Subsequent frequency calculations were carried out in order to establish the structures as local or global minima (with no imaginary frequencies in their vibra-

tion spectra). The QST2 and QST3 algorithms were used to locate each transition state between two minima.

All calculations were performed with GAUSSIAN 03 program package.[21] The programs MOLDA and CHEMCRAFT were used for visualization of molecular structures and vibration spectra.[22, 23]

3. Results and Discussion

Our preliminary investigations started with X-ray powder analysis of Cyt (Fluka). The roentgenogram was recorded on a TÜR-MA-62 apparatus, working tension of 32 kV, Cu-anticathode, $\lambda_{\alpha 1}$ = 1.5405 Å, $\lambda_{\alpha 2}$ = 1.5443 Å, $\bar{\lambda}$ = 1.5424 Å. The recorded spectrum is shown in Fig. 1.

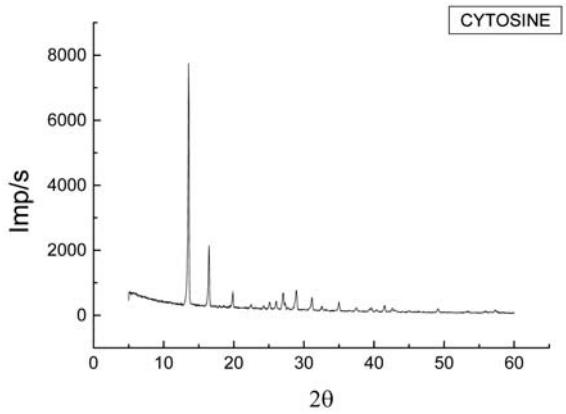


Fig. 1. X-ray powder spectrum of cytosine (Fluka)

The spectrum clearly shows the crystal structure of cytosine. The calculated interplane distance in the Cyt crystal using the Bragg's equation is 6.533 Å.[24] In order to check theoretically this value a model with two co-planar Cyt molecules was built. The attempts to calculate (AM1 and PM3) this "sandwich" model failed since the program was not able to find the minimum of such system. The energies found at the two semiempirical levels are E_{AM1} = –3043.8 eV and E_{PM3} = –2662.9 eV. The AM1 optimizations of a system with four molecules (co-planar by couples) led to the system with four Cyt molecules approximately situated in one plane (see Fig. 2) with energy E_{AM1} = –6088.1 eV.

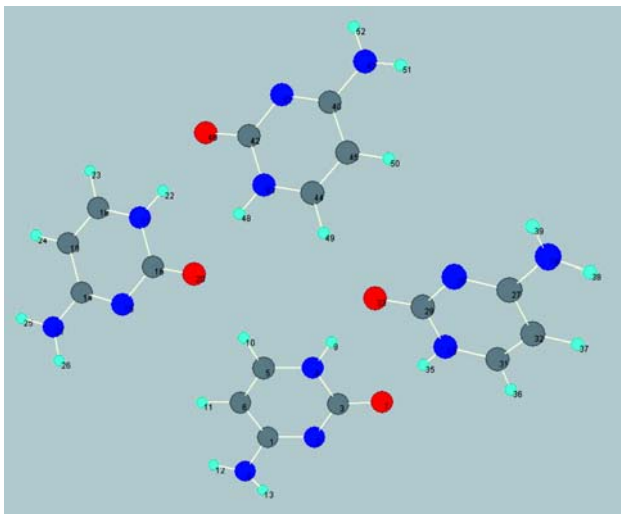
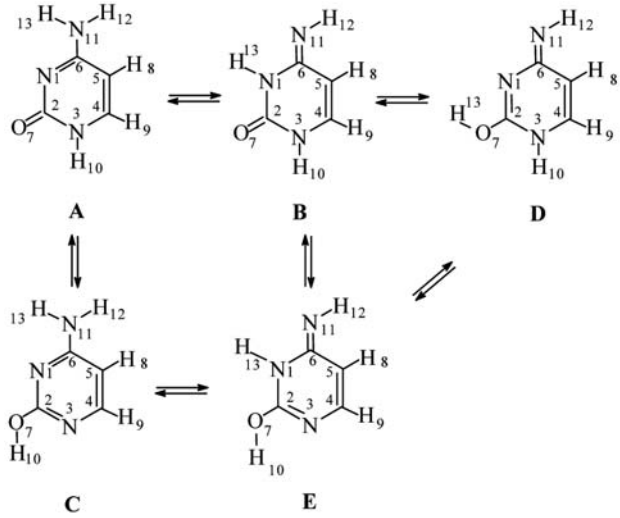


Fig. 2. Four cytosine molecules (AM1 optimizations) visualized by MOLDA.22

Five isomers and their mutual interconversions were studied in the current paper. They are illustrated in Scheme 2.



Scheme 2. Tautomers and tautomeric interconversions of cytosine (the numeration does not follow the IUPAC nomenclature)

Concerning the planarity of the isomers, several major dihedral angles are listed in Table 1.

Only the calculations with the basis sets 6-31G, 6-31G(d,p), and 6-311G predicted a full conjugation of the amino group nitrogen with the aromatic pyrimidine ring in the tautomer **A**. Obviously, the inclusion of diffuse functions considerably enhances the accuracy of the computations and implies a considerable pyramidalization of the amino group in the tautomer **A**. The pyramidal character of the amino group in the nucleic acid basis has been a subject of many papers.[25,26]

The isomers **B**, **D**, and **E** have completely planar structures. It should be expected that the intramolecular proton transfers in these forms occur in the molecular plane, and thus to provoke lower energy barriers as compared to the proton transfers in tautomers **A** and **C**. Our computations for tautomer **B** are in accord with the data reported by Hobza and Sponer.[27] The valent bond C_6=N_{11} in isomer **D** an **E** changes in the interval 1.257–1.294 Å and 1.259–1.293 Å, respectively. It is interesting to mention that the inclusion of d- and p-polarization functions in the basis set leads to a drastic shortening of this bond. However the additional inclusion of diffuse functions provokes an insignificant elongation.

The calculated energies and relative energies with respect to the most stable isomer are given in Table 2.

Table 1. Dihedral angles (in deg.) of the cytosine tautomers

Method / basis set	A			B			C			D			E		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
B3LYP/6–31G	0.0	0.0	180.0	0.0	0.0	180.0	0.0	0.0	180.0	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–31G(d)	19.7	0.4	180.0	0.0	0.0	180.0	20.7	0.3	179.9	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–31G(d,p)	0.0	0.0	180.0	0.0	0.0	180.0	18.3	0.2	179.9	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–31+G(d,p)	9.8	0.1	180.0	0.0	0.0	180.0	14.3	0.1	179.9	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–31++G(d,p)	10.1	0.1	180.0	0.0	0.0	180.0	14.5	0.1	179.9	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–311G	0.0	0.0	180.0	0.0	0.0	180.0	0.0	0.0	180.0	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–311G(d)	15.1	0.3	180.0	0.0	0.0	180.0	17.5	0.2	179.9	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–311G(d,p)	14.2	0.2	180.0	0.0	0.0	180.0	16.7	0.1	179.9	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–311+G(d,p)	11.1	0.2	180.0	0.0	0.0	180.0	14.8	0.1	179.9	0.0	0.0	180.0	0.0	0.0	180.0
B3LYP/6–311++G(d,p)	11.6	0.2	180.0	0.0	0.0	180.0	15.0	0.1	179.9	0.0	0.0	180.0	0.0	0.0	180.0

1: d(12,11,6,5); 2: d(1,2,3,4); 3: d(7,2,3,4).

Table 2. Energies (E, a.u.) and relative energies (E_{rel}, kJ mol^{-1}) of the isomers

Method / basis set	A	B	C	D	E
			E		
B3LYP/6–31G	–394.806159	–394.803135	–394.795112	–394.759103	–394.773421
B3LYP/6–31G(d)	–394.927857	–394.925759	–394.925911	–394.890230	–394.905010
B3LYP/6–31G(d,p)	–394.941352	–394.939345	–394.941363	–394.905937	–394.920537
B3LYP/6–31+G(d,p)	–394.963177	–394.960224	–394.962482	–394.927554	–394.941508
B3LYP/6–31++G(d,p)	–394.963363	–394.960409	–394.962654	–394.927774	–394.941680
B3LYP/6–311G	–394.907495	–394.905034	–394.896465	–394.861733	–394.875971
B3LYP/6–311G(d)	–395.028597	–395.026343	–395.023983	–394.988887	–395.003555
B3LYP/6–311G(d,p)	–395.041030	–395.039323	–395.039693	–395.005233	–395.019548
B3LYP/6–311+G(d,p)	–395.053064	–395.050585	–395.051545	–395.017539	–395.031098
B3LYP/6–311++G(d,p)	–395.053158	–395.050681	–395.051648	–395.017648	–395.031195
			E_{rel}		
B3LYP/6–31G	0.0	7.9	29.0	123.6	86.0
B3LYP/6–31G(d)	0.0	5.5	5.1	98.8	60.0
B3LYP/6–31G(d,p)	0.0	5.3	0.0	93.0	54.7
B3LYP/6–31+G(d,p)	0.0	7.8	1.8	93.5	56.9
B3LYP/6–31++G(d,p)	0.0	7.8	1.9	93.5	56.9
B3LYP/6–311G	0.0	6.5	29.0	120.2	82.8
B3LYP/6–311G(d)	0.0	5.9	12.1	104.3	65.8
B3LYP/6–311G(d,p)	0.0	4.5	3.5	94.0	56.4
B3LYP/6–311+G(d,p)	0.0	6.5	4.0	93.3	57.7
B3LYP/6–311++G(d,p)	0.0	6.5	4.0	93.2	57.7

In accordance with the experiment the isomer **A** was found to be the most stable one (B3LYP/6-311++G(d,p)).[1–3] The isomer **D** was found to be the most unstable one. The preliminarily calculated relative energies of the isomers at the HF level showed that the inclusion of diffuse and polarization functions in the basis set change the stability of the isomers **A** and **C** (isomer **C** gets more stable than **A**). Obviously, these combinations of method and basis sets do not give accurate results neither for the structural parameters nor for the energies.

The values of the relative energies listed in Table 2 can be interpreted as heat effects of transformations of isomer **A** to all remaining. If these transformations occur without considerable entropy changes (as known for the intramolecular proton transfers)[28] the relative energies could be assigned as Gibbs energy changes (ΔG) of the transformations.

As seen from Scheme 2 the direct transformation of isomer **A** to **D** is impossible, because it includes simultaneous proton exchange between atoms N_{11} and N_1 as well as between N_1 and O_7. As known the probability of simultaneous realization of two events is very small. The same reason makes the one-step transformation of **A** to **E** an impossible process.

The transition states of the isomer transformations shown in Scheme 2 were found. As clearly seen the transformations **A** ⇄ **B**, **A** ⇄ **C**, **B** ⇄ **D**, **C** ⇄ **E**, and **B** ⇄ **E** are tautomeric conversions, whereas the transformation **E** ⇄ **D** is a conformation conversion.

All transition states (Fig. 3) were found as a first order saddle points. In the vibration spectrum of each was found one imaginary frequency corresponding to a parallel mode along the reaction coordinate.

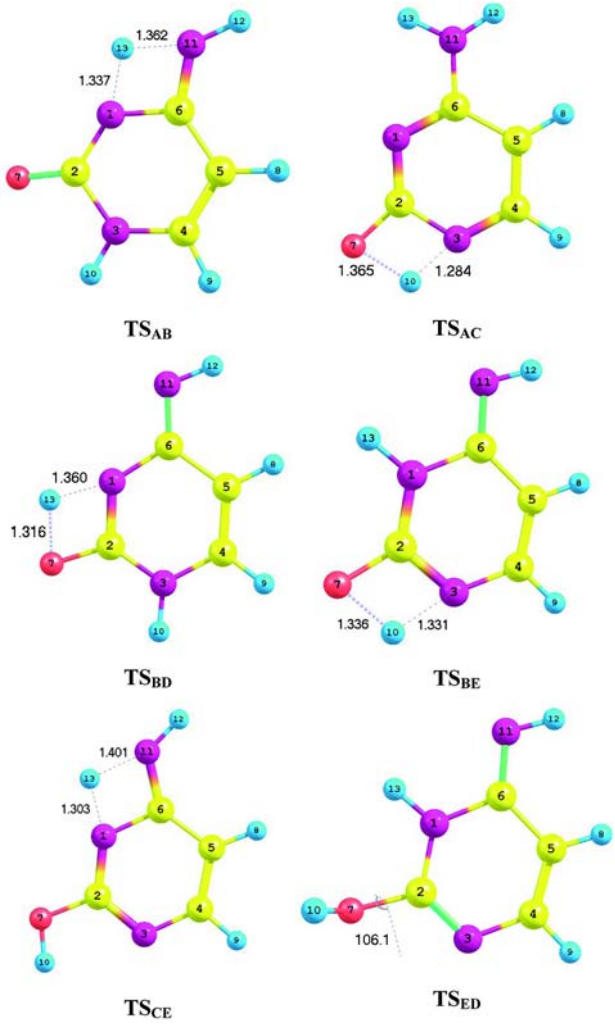


Fig. 3. Transition states of the isomer transformations (B3LYP/6-311++G(d,p))

All tautomeric conversions pass through planar transition states – the proton transfer occurs in the molecular plane. As concerns the tautomers **A** and **C** and the transition states of the tautomer conversions they take part, one can say that the transition states have conjugated N atom (from the amino group) with the pyrimidine ring. This would reflect on the energy barriers of the conversions in which these tautomers take part. They should be lower as compared to the energy barriers of the remaining conversions. The energy barriers and thermodynamic parameters of the conversions are listed in Table 3.

As seen from Table 3 the isomer conversions are enthalpically disfavored ($\Delta H > 0$). They also have positive variation of the Gibbs free energies and small steric changes along the reaction pathways. All that show that according to the Leffler-Hammond postulate the transition states are product-like or so-called "late" transition states.[29,30]

As mentioned above the direct transformation of isomer **E** into **A** is impossible. This isomerization can occur in two steps, passing through isomer **C** (Fig. 4).

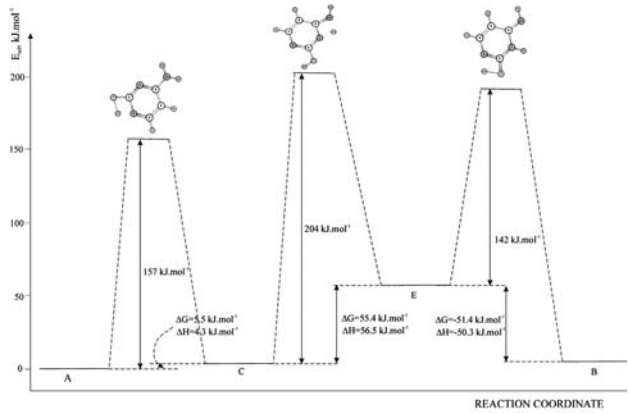


Fig. 4. Energy curves of the isomer conversions (B3LYP/6-311++G(d,p))

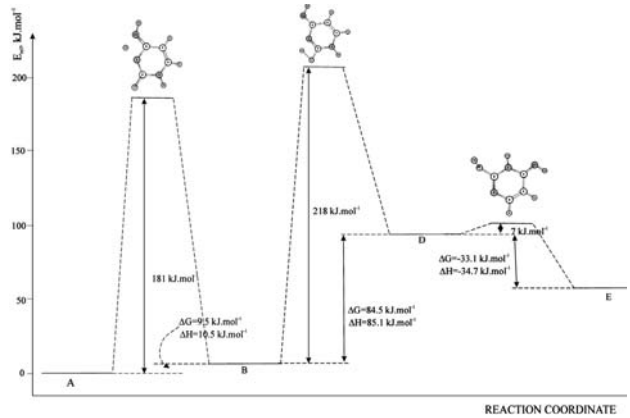


Fig. 5. Energy curves of the isomer conversions (B3LYP/6-311++G(d,p))

As seen the equilibrium is shifted towards isomer **A**. The transformation of isomer **D** into **A** passes through two steps as well. Initially **D** isomerizes to isomer **B** through an energy barrier of 131 kJ mol^{-1}. After that isomer **B** transforms into **A** through an energy barrier of 174 kJ mol^{-1} (Fig. 5).

4. Conclusions

On the basis of the performed calculations at the B3LYP level the following major conclusions can be deduced: (i) in accordance with the experiment,[1–3] all basis

Table 3. Energy barriers and thermodynamic parameters of the conversions (kJ mol^{-1})

Method /	**B ⇄ D**					**C ⇄ E**				
basis set	E_{forw}	E_{rev}	**ΔH**	**ΔG**	**TΔS**	E_{forw}	E_{rev}	**ΔH**	**ΔG**	**TΔS**
B3LYP/6–31G	180	172	8.8	9.5	–0.7	173	144	27.4	28.0	–0.6
B3LYP/6–31G(d)	183	177	9.6	8.7	0.9	154	149	5.2	6.3	–1.1
B3LYP/6–31G(d,p)	175	169	28.2	8.5	19.7	146	146	19.2	1.4	17.7
B3LYP/6–31+G(d,p)	180	173	11.6	10.7	0.9	153	151	2.1	3.2	–1.1
B3LYP/6–31++G(d,p)	180	173	11.6	10.7	0.9	153	151	2.1	3.2	–1.1
B3LYP/6–311G	182	176	7.1	7.7	–0.6	175	146	27.6	28.2	–0.6
B3LYP/6–311G(d)	186	181	10.0	9.1	0.9	165	153	12.3	13.4	–1.1
B3LYP/6–311G(d,p)	149	145	8.6	7.7	0.9	154	151	3.8	4.9	–1.1
B3LYP/6–311+G(d,p)	181	174	10.6	9.6	1.0	157	153	5.7	5.5	0.1
B3LYP/6–311++G(d,p)	181	174	10.5	9.5	1.0	157	153	4.3	5.5	–1.2
	B ⇄ D					**C ⇄ E**				
B3LYP/6–31G	242	126	111.7	111.0	0.7	204	147	56.4	56.6	–0.2
B3LYP/6–31G(d)	219	126	91.1	90.6	0.5	206	151	57.7	56.7	1.0
B3LYP/6–31G(d,p)	211	123	85.9	85.4	0.5	198	144	57.5	56.5	1.0
B3LYP/6–31+G(d,p)	214	128	84.1	83.5	0.6	203	148	57.8	56.7	1.0
B3LYP/6–31++G(d,p)	214	128	84.0	83.3	0.7	203	148	57.8	56.8	1.0
B3LYP/6–311G	243	130	110.4	109.7	0.7	205	152	53.2	53.4	–0.1
B3LYP/6–311G(d)	230	131	96.4	96.0	0.4	210	156	56.6	55.6	1.0
B3LYP/6–311G(d,p)	218	128	87.7	87.3	0.5	201	148	55.8	54.8	1.0
B3LYP/6–311+G(d,p)	218	131	85.1	84.6	0.5	204	150	55.2	55.4	–0.2
B3LYP/6–311++G(d,p)	218	131	85.1	84.5	0.6	204	150	56.5	55.4	1.1
	B ⇄ D					**C ⇄ E**				
B3LYP/6–31G	48	11	36.7	35.9	0.8	211	133	75.0	75.0	–0.1
B3LYP/6–31G(d)	45	6	37.8	36.2	1.5	191	137	53.3	54.4	–1.0
B3LYP/6–31G(d,p)	44	5	37.4	35.9	1.4	183	134	48.5	49.5	–1.0
B3LYP/6–31+G(d,p)	43	6	35.8	34.3	1.5	189	140	48.3	49.2	–0.9
B3LYP/6–31++G(d,p)	43	6	35.7	34.1	1.6	189	140	48.3	49.3	–1.0
B3LYP/6–311G	48	11	36.7	35.9	0.8	213	137	73.7	73.8	–0.1
B3LYP/6–311G(d)	47	9	37.5	36.1	1.5	202	142	58.9	59.9	–1.0
B3LYP/6–311G(d,p)	44	7	36.7	35.3	1.4	191	139	51.0	52.0	–0.9
B3LYP/6–311+G(d,p)	43	7	34.8	33.2	1.6	193	142	50.3	51.4	–1.0
B3LYP/6–311++G(d,p)	43	7	34.7	33.1	1.7	193	142	50.3	51.4	–1.1

sets showed that the aminooxo Cyt tautomer has the lowest energy. According to the B3LYP/6-311++G(d,p) calculations iminohydroxo isomer (**D**) has the lowest stability with a relative energy of 93.2 kJ mol^{-1}. All other isomers follow the next stability pattern: A > C > B > E > D. (ii) Isomers **A** and **C** have non-planar structure with respect to the amino group. The isomers **D** and **E** have planar structures found with all basis sets. (iii) All tautomerizations pass through planar transition states, therefore the intramolecular proton transfer occurs in the molecular plane. (iv) For all geometries studied in the current paper it was established that the inclusion of polarization functions in the basis set leads to a drastic reduction of the energy. The energy barriers are not so sensitive toward inclusion of diffuse functions in the basis set.

5. Acknowledgements

We thank the National Science Fund (Ministry of Education and Science of Bulgaria) for the financial support (Project "Young Scientists – Chemistry – 1504").

6. References

1. M. J. Nowak, L. Lapinski, J. Fulara, *J. Spectrochim. Acta* **1989**, *45A*, 229–242.
2. I. R. Gould, M. A. Vincent, I. H. Hiller, L. Lapinski, M. J. Nowak, *J. Spectrochim. Acta* **1992**, *48A*, 811–817.
3. A. Jaworski, M. Szczepaniak, K. Szczepaniak, K. Kubulat, W. B. Person, *J. Mol. Stuct.* **1990**, *223*, 63–92.

4. R. Kobayashi, *J. Phys. Chem. A* **1998**, *102*, 10813–10818.
5. R. D. Brown, P. D. Godfrey, D. McNaughton, A. P. Pierlot, *J. Am. Chem. Soc.* **1987**, *111*, 2308–2310.
6. A. Les, L. Adamowicz, R. J. Bartlett, *J. Phys. Chem.* **1989**, *93*, 4001–4005.
7. L. Gorb, Y. Podolyan, J. Leszczynski, *J. Mol. Struct. (THEOCHEM)* **1999**, *487*, 47–55.
8. J. S. Kwiatkowski, W. B. Person, K. Szczepaniak, M. Szczesniak, *Studia Biophysica* **1974**, *46*, 79–97
9. B. D. Sharma, J. F. McConnell, *Acta Crystallogr.* **1965**, *19*, 797–806.
10. P. R. Lowe, C. H. Schwalbe, G. J. B. Williams, *Acta Crystallogr.* **1987**, *C43*, 330–333.
11. C. Helene, P. C. R. Dozou, *Acad. Sci. (Paris)* **1964**, *259*, 4387–4392.
12. D. J. Brown, T. Teitei, *Aust. J. Chem.* **1965**, *18*, 559–568.
13. H. Morita, S. Nagakura, *Teor. Chim. Acta* **1968**, *11*, 279–295.
14. A. Kaito, M. Hatano, T. Ueda, S. Shibuya, *Bull. Chem. Soc. Jpn.* **1980**, *53*, 3073–3078.
15. A. Lehninger: Principles of Biochemistry, 4th ed., Freeman, **2004**, pp. 608–610.
16. R. Knippers: Molekulare Genetik, Thieme, Stuttgart, **1997**.
17. M. D. Topal, J. R. Fresco, *Nature* **1976**, *263*, 285–289.
18. A. Müller, M. Losada, S. Leutwyler, *J. Phys. Chem. A* **2004**, *108*, 157–165.
19. J. Gu, J. Leszczynski, *J. Phys. Chem. A* **1999**, *103*, 2744–2750.
20. A. M. Mebel, K. Morokuma, C. M. Lin, *J. Chem. Phys.* 1995, *103*, 7414–7421.
21. Gaussian 03, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, **2004**.
22. H. Yoshida, MOLDA 6.5 Rev. 2.0, **2004**.
23. G. A. Zhurko, D. A. Zhurko, CHEMCRAFT 1.5 (build 275).
24. W. J. Moor: Physical Chemistry, 5th ed., Longman, London, **1974**, pp. 838–838.
25. L. Gorb, J. Leszczynski, *J. Am. Chem. Soc.* **1998**, *120*, 5024–5032.
26. V. B. Delchev, H. Mikosch, *Monatsh. Chem.* **2004**, *135*, 1373–1387.
27. P. Hobza, J. Sponer, *Chem. Rev.* **1999**, *99*, 3247–3276.
28. V. B. Delchev, H. Mikosch, Scientific works of the Union of Scientists – Plovdiv **2005**, *V*, 7–12.
29. G. S. Hammond, *J. Am. Chem. Soc.* **1955**, *77*, 334–338.
30. J. E. Leffler, *Science* **1953**, *117*, 340–341.

Povzetek

Z uporabo B3LYP metode (z baznimi seti 6-31G in 6-311G ter z vrsto različnih polarizacijskih in difuznih funkcij) smo teoretično raziskali pet izomerov citozina in njihove medsebojne pretvorbe. Pokazali smo, da je od vseh tavtomerov citozina najbolj stabilen kanonski aminookso tavtomer, vendar pa ima neplanarno geometrijo. Pokazali smo tudi, da so izračunane energije in energijske bariere v preučevanih sistem precej odvisne od tega, če v bazni set vključimo polarizacijske funkcije, medtem ko je občutljivost na vključitev difuznih funkcij manjša.