*Mathematical Biology and Bioinformatics*
*2023. V. 18. № 1. P. 33–48. doi: 10.17537/2023.18.33*

========== **MATHEMATICAL MODELING** ==========

# Proton Reaction Path in Base Pairs of DNA Molecule According to the Complete Active Space Self-Consistent Field Method

**Simon K.V.*[1], Tulub A.V.†[2]**

[1]*Ioffe Institute, St. Petersburg, Russia*
[2]*Saint Petersburg State University, Russia*

***Abstract.*** The double proton transfer reaction paths in AT and CG base pairs of DNA molecule are calculated in the Complete Active Space Self-Consistent Field method and compared with the same paths in Density Functional Theory with B3LYP approximation approach. We found that an essential increase of an activation energy, which significantly reduces the probability of spontaneous mutations in DNA via double proton transfer. There exist two transition points on the singlet potential energy surface divided by a flat region for GC base pair. The applicability of various quantum-chemical methods for description of double proton transfer reactions was discussed.

***Key words:*** *proton transfer, DNA, base pairs, ab initio CASSCF method.*

## INTRODUCTION

There are four different DNA bases: adenine (A), guanine (G), cytosine (C), and thymine (T). The genetic stability of DNA is dependent on proton transfer along the hydrogen bonds in GC base pairs; it could lead to tautomeric structures, schematically

$$GC \leftrightarrow G^*C^* \rightarrow G^*T,$$

where the star denotes the tautomeric form. There arise hence the point mutations; guanine becomes bound with thymine instead of cytosine according to the Watson–Crick pairing rule.

The theoretical treatments of proton transfer between base pairs represent as the first step the calculation of potential energy surface (PES) and corresponding activation barriers. This was done in a number of papers, mainly in the framework of density functional theory (DFT). The review [1] and literature therein and the work [2] may serve as examples for a post-Hartree-Fock *ab initio* study including the estimation of the Gibbs free energy of the reaction. The next problem is the calculation of the quantum tunneling contribution to the proton transfer rate; it depends on the accuracy of PES finding.

A great practical interest arises in treatment the influence of different environments on the proton motion, for instance, the inclusion of silver clusters [1], gold surface [3], water molecules [4–6], and some kinds of drugs [7] into the quantum system. These surroundings play more or less the role of a perturbation on the proton transfer energies in the base pairs. The infrared spectrum change from GC to G*C* and from AT to A*T* was investigated in [7]. It was pointed out that the hydrogen atom Ha in A*T* (see figure in Table 1 below) is weakly bonded to oxygen in T* state near the equilibrium.

The reaction path in the AT and GC base pairs was done in [8, 9] in PBE, PBE (MBD) and in B3LYP approximations. Only in the last case the proton transfer in GC↔G*C* was accompanied by the relative stable product state G*C*. The quantum tunneling contribution to the proton transfer rate was calculated in [10] on the basis of previous estimation of the

*k.v.simon@mail.ioffe.ru
†a.v.tulub@gmail.com

reaction path [8] in B3LYP method and the analytic approximation of PES in the corresponding region of distances. Some advance of B3LYP approach is connected with the choice of parameters, allowing to take into account some part of the correlation energy at MP2 level, the reaction path in HF+MP2 theory is close to that in the B3LYP approximation. Cooping in mind this fact, we consider below the power of CASSCF method, which allows reproducing the main part of correlation energy with the geometry optimization on the reaction path. It is compared below with that in the B3LYP approach.

The base pairs are surrounded in double-stranded DNA by phosphate backbones. Thus, water molecules have little effect on the hydrogen bonds in base pairs, the gas phase calculations give an important information.

The genetic code information, initially stored in DNA, transfers from mRNA to a charged tRNA. The hydrogen bonding in this region obeys the Watson-Crick pairing rule. The role of a perturbation can play in this case different ions in a solution, the magnesium cations in the first place [11]. The treatment of DNA molecule can be useful also in this case.

## COMPUTATIONAL DETAILS

The calculation of double proton transfer reaction between nucleobases in DNA was performed following the technique of work [9] with IRC (Intrinsic Reaction Coordinate) algorithm. The IRC represents the steepest descent path from transition state point to products and reactants. Thus, to calculate the reaction path, one needs to locate the appropriate transition state points, which is described below. Within this work for quantum-chemical description of the system, we employed the multiconfiguration theory of self-consistent field (MCSCF) [12] in the version of complete active space self-consistent field (CASSCF) [13]. For the sake of comparison with earlier obtained results in [9] the reaction paths in density functional theory (DFT) and second-order perturbation theory (RHF+MP2) were also performed.

The CASSCF description of the system demands the choice of the active molecular orbitals (MO) space. Some recommendations for construction of active spaces can be found in the article [14]. The main idea is to construct the starting set of localized MOs for active space. All necessary for the process chemical bonds should be described by the pairs of MOs (bonding-antibonding), and also lone pairs may be added into the active space as needed. The constructed active space is denoted as CASSCF ($N$,$M$), where $N$ – is the number of active electrons, and $M$ states for number of active orbitals needed for allocation of active electrons in reference configuration plus the appropriate number of vacant MOs.

In the case of this double proton transfer processes the homolytical bond clevages do not take place, which should be described by the electron pairs. This is also indicated by the fact, that DFT calculations do not destroy the spacial symmetry of α- and β-subsets of MOs. Thus in principle, there is no need to include the disrupted bonds in the active space. Nevertheless the bonds of Ha and Hb atoms (see figures below) were included in the active space along with the lone pairs of nitrogen and oxygen atoms, for which the transfer is directed. In addition, the active space was augmented by two pairs of orbitals bonding-antibonding, which describe π-character bonds C-O in the carbonyl groups (guanine and thymine), and localized π-bonds in the rings near the nitrogen atoms, which accept Hb atom (adenine and cytosine). The constructed in this way active space includes 12 active electrons among 10 active MOs and is denoted CASSCF (12,10). The shapes of obtained active natural MOs in the optimal geometry of AT and GC base pairs are presented in Appendix 1 and 2 respectively. The CASSCF and RHF+MP2 calculations were performed with 6-31++G(d,p) basis set, but DFT calculations were performed with 6-311++G(d,p) basis following the authors of the work [9]. The quantum-chemical task was treated by GAMESS-US(2022) [15], and the visualization of obtained results was performed by the wxMacMolPlt program [16].

It should be noted that CASSCF approach takes into account static correlation effects, but it takes into account only a small part of dynamic correlation effects. Therefore, the described below results should be considered as the first approximation. One needs additional calculation step to take into account the dynamic correlation. This may change the described below reaction paths and activation barriers. Unfortunately, the most of such methods today do not have nuclear gradients available and do not allow geometry optimization. The described in this paper results are needed to be revised, if such quantum-chemical methods would appear.

## OBTAINED RESULTS

Both pairs of adenine-thymine AT and guanine-cytosine GC along with designation of atoms are shown in Figures 1a and 1b. In the course of reaction of double proton transfer, the atoms Ha and Hb are transferred. The atom Hc in guanine-cytosine pair remains inactive. The same result was obtained by the authors of article [9] in the framework of B3LYP reaction path calculation.

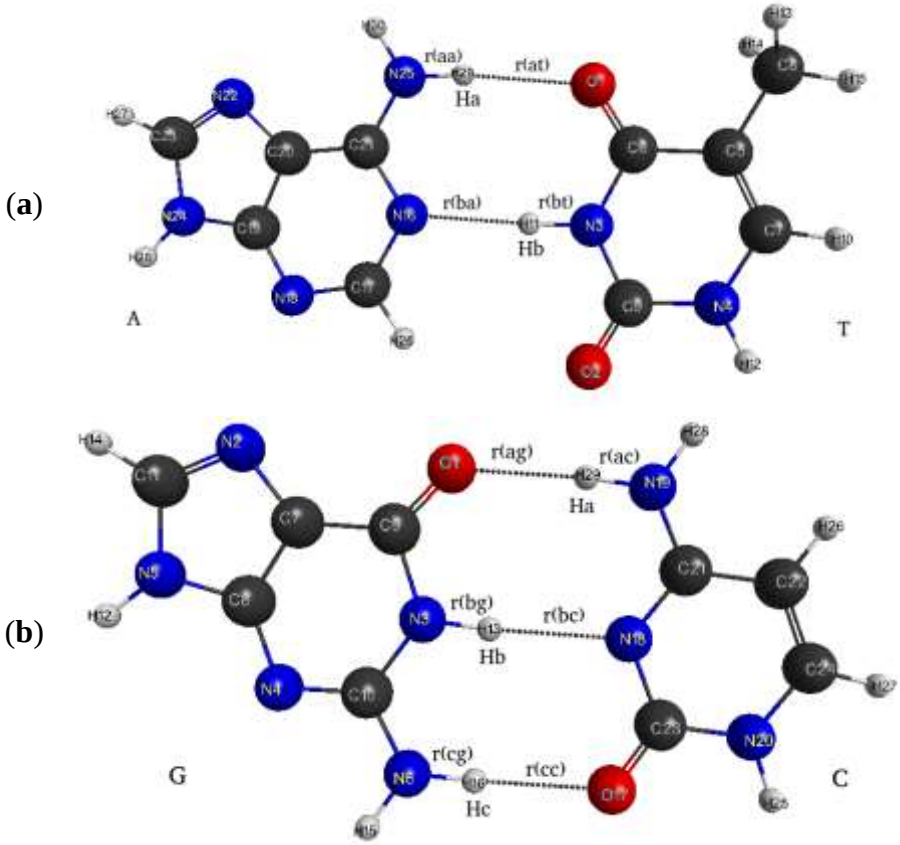
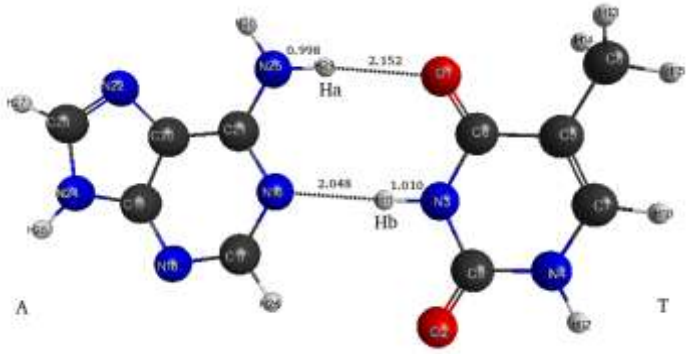


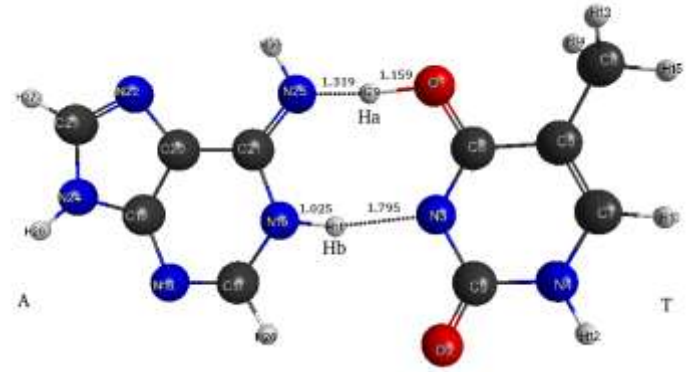

**Fig. 1.** The adenine-thymine AT and guanine-cytosine GC pairs before proton transfer with atom designation.

The calculated stationary points along the reaction path of double proton transfer AT→A*T* in adenine-thymine base pair are presented in figures collected in Table 1. The reaction begins with transfer of Hb proton, which carries Mulliken charge 0.66 in comparison with 0.44 for Ha. The process of Hb proton transfer does not yield any stationary points, and immediately after its end the Ha proton moves, giving distinctly expressed transition point with activation energy about 29.2 kcal/mol. The final structure A*T* defines the effect of this reaction about 26.8 kcal/mol, which implies this process to be endothermic. The symmetry of entire system remains approximately planar ($C_S$) along the reaction path.

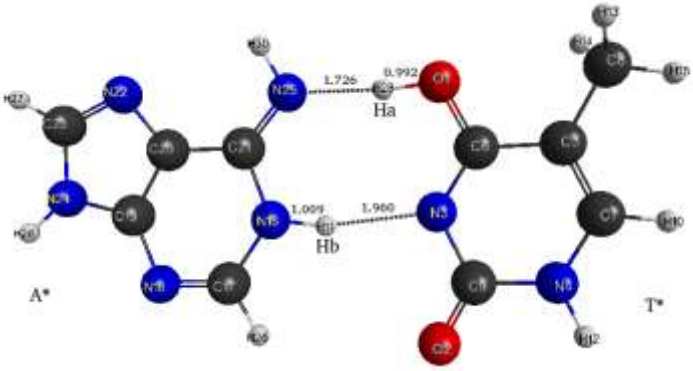
**Table 1.** The stationary points along the path of mutual proton exchange between thymine and adenine (the CASSCF(12,10) method, bond distances in angstrom)




(AT) $E$ = –916.215016 a.u. (0 kcal/mol)
q(Ha)=+0.44, q(Hb)=+0.66.




(TS) $E$ = –916.168414 a.u. (29.24 kcal/mol)




(A*T*) $E$ = –916.172360 a.u. (26.77 kcal/mol)

The dependece of relative energy along the reaction path in IRC-calculation is depicted in Figure 2 with black line. The colored lines represent obtained reaction paths in different DFT and RHF+MP2 methods. All these methods give significantly underestimated activation energies and final structure A*T* energy in comparison with CASSCF method. The red line shows reaction path in B3LYP method according to work [9]. The M06 method is sufficiently close to B3LYP, but the M06-HF with 100 % exchange of Hartree-Fock method gives activation energy less than 10 kcal/mol. One can notice the proximity of B3LYP and RHF+MP2 curves, which implies, that B3LYP was calibrated on the basis of MP2 method. It is worth to note, that various methods have different lengths of reaction path in the space of mass-weighted coordinates. The first and the last points are defined by zero gradient geometries of reactants and products.

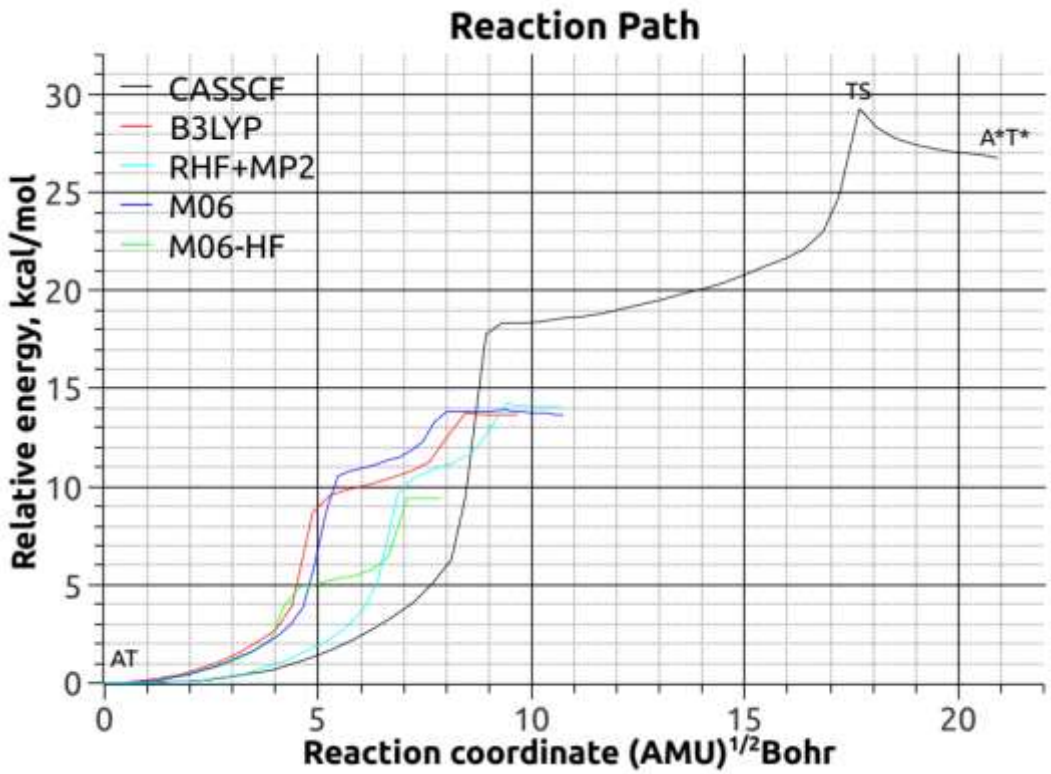



**Fig. 2.** The dependence of relative energy along the reaction path AT→A*T* in CASSCF, DFT and RHF+MP2 methods.

All considered methods do not predict the existence of stable structure of A*T*, the potential bath for which in CASSCF method does not exceed 2.5 kcal/mol. However, if according to DFT and RHF+MP2 methods the A*T* structure may appear as nonstable geometric conformation, the CASSCF method does not show of such possibility, because the activation energy of 29.2 kcal/mol is too high.

The dependence of interatomic distances along the reaction path AT→A*T* is shown in Figure 3. One can notice, that proton Hb does transfer in the first turn, the distances r(bt) and r(ba) are shown with black and red lines respectively. The change in position of Hb proton corresponds to the bend of the potential energy curve about 9$(AMU)^{1/2}$Bohr, where the transition state point could be. The Ha proton is transferred in the second turn, which is reflected in distances r (aa) and r(at) shown with green and blue lines respectively. The change of position of Ha proton corresponds transition state point on the potential curve of CASSCF method.

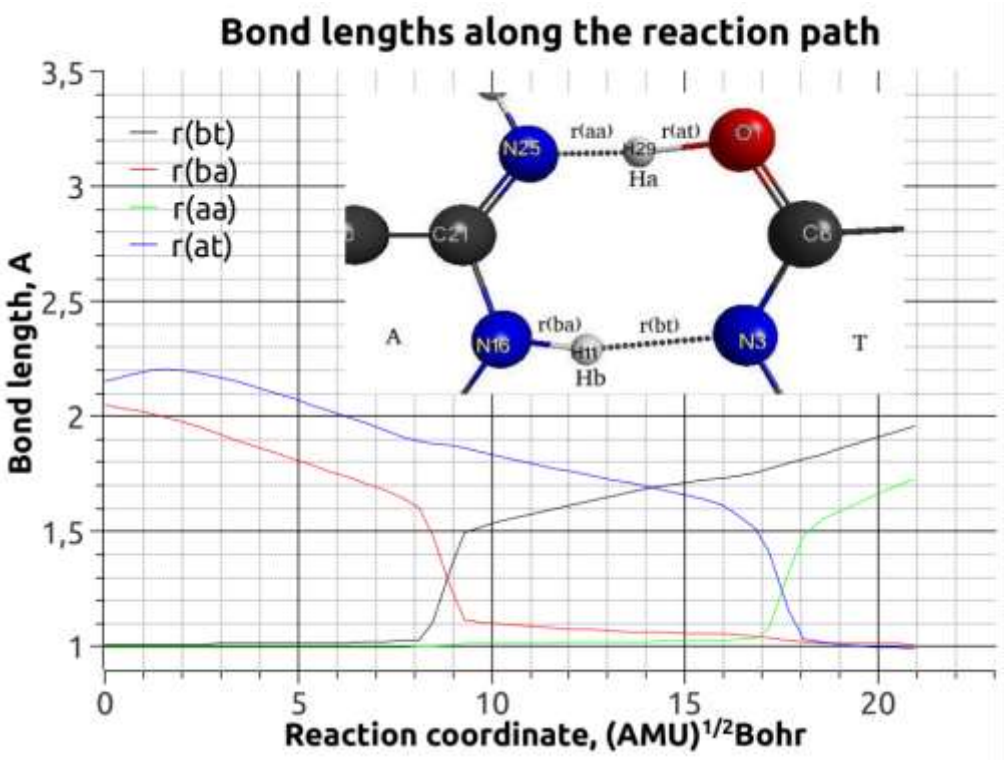
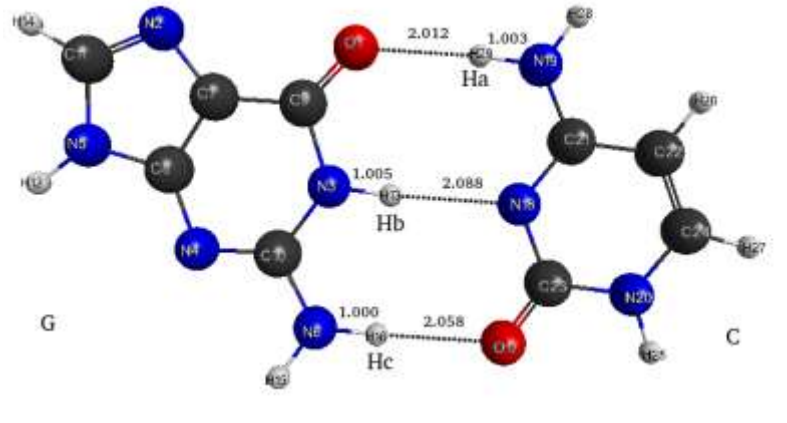


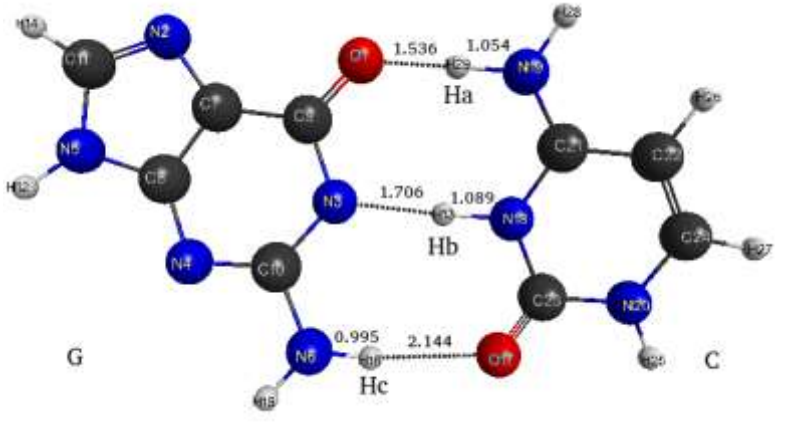

**Fig. 3.** The dependence of bond lengths (angstrom) along the reaction path of transfer of protons AT→A*T* in the CASSCF method.

In the case of guanine-cytosine base pair the reaction path, strictly speaking, is characterized in CASSCF method by two points of transition states (TS1 and TS2) and existence of intermediate minimum (IM). The energies of these geometry configurations are very close, and therefore only geometry of intermediate minimum is shown in the Table 2. As in the case of adenine-thymine base pair, the reaction GC→G*C* begins with transfer of mostly charged Hb atom from guanine to cytosine. Through the first transition state point with relative energy 30.89 kcal/mol the reaction path leads to the presented in Table 2 geometry of intermediate minimum with relative energy 30.78 kcal/mol. The second part of the reaction is the transfer of Ha proton from cytosine to guanine, it has the transition state point (TS2) with relative energy 31.66 kcal/mol, which defines the full activation energy of the overall process GC→G*C*. The total energy effect of this reaction is 21.62 kcal/mol, this means the reaction is also endothermic. The planar symmetry ($C_S$) of the initial GC structure is destroyed in TS1, IM and TS2, but approximately recovers in G*C* structure.

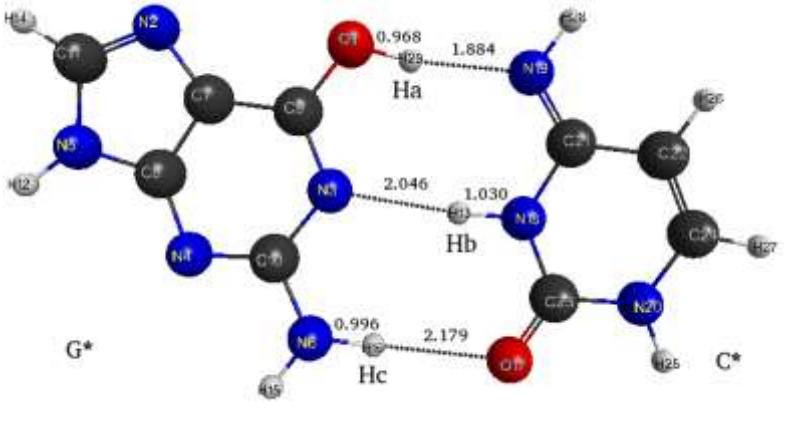
**Table 2.** The stationary points along the path of mutual exchange of protons between guanine and cytosine (CASSCF(12,10) method, bond distances in angstrom)




(GC) $E$ = –932.220034 a.u. (0 kcal/mol)
q(Ha) = +0.46, q(Hb) = +0.63, q(Hc) = +0.47.




(IM) $E$ = –932.170975 a.u. (30.78 kcal/mol)




(G*C*) $E$ = –932.185575 a.u. (21.62 kcal/mol)

The dependence of relative energy along the reaction path GC→G*C* is depicted in Figure 4. As in the case of adenine-thymine base pair, the B3LYP method two times underestimates the value of activation energy. In contrast to CASSCF, the B3LYP path has only one transition state point. The same result was also obtained in the work [9]. Nevertheless the B3LYP method gives the value of potenial bath for G*C* structure about

5.7 kcal/mol. This value also agrees with [9]. For CASSCF method the depth of potential bath for structure G*C* turns out to be 10 kcal/mol, which makes it possible the existence of complex with transferred protons, as low-stable structure relative to reverse reaction G*C*→GC. The high activation energy of the direct reaction GC→G*C* and relatively small value of the potential bath for G*C* make extremely impossible the process of DNA mutations in guanine-cytosine base pairs because of double proton transfer.

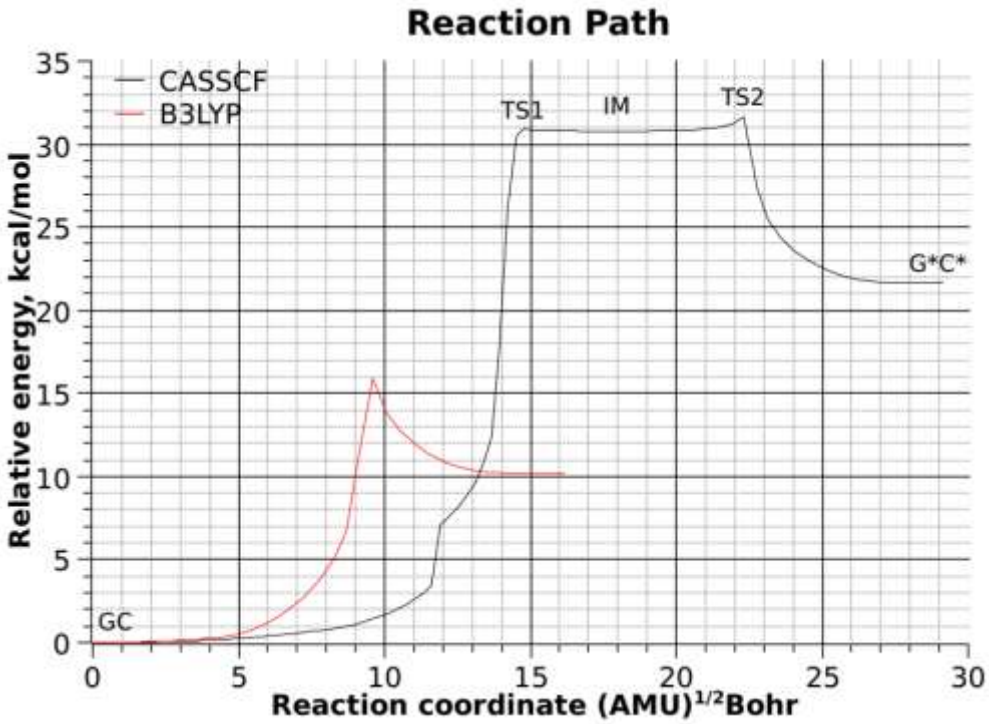



**Fig. 4.** The dependence of relative energy along the path of reaction GC→G*C* in CASSCF and B3LYP methods.

The dependence of bond lengths for trasferred protons along the reaction path GC→G*C* is presented in Figure 5. The reaction begins with transfer of Hb proton, which is characterized by the distances r(bg) and r(bc) (black and red lines). After that the proton Ha is transferred according to the curves of r(ac) and r(ag) distances (green and blue lines). The first transition state point TS1 corresponds to the change of position of Hb proton, and the second transition state point TS2 corresponds to the change of position of Ha proton.

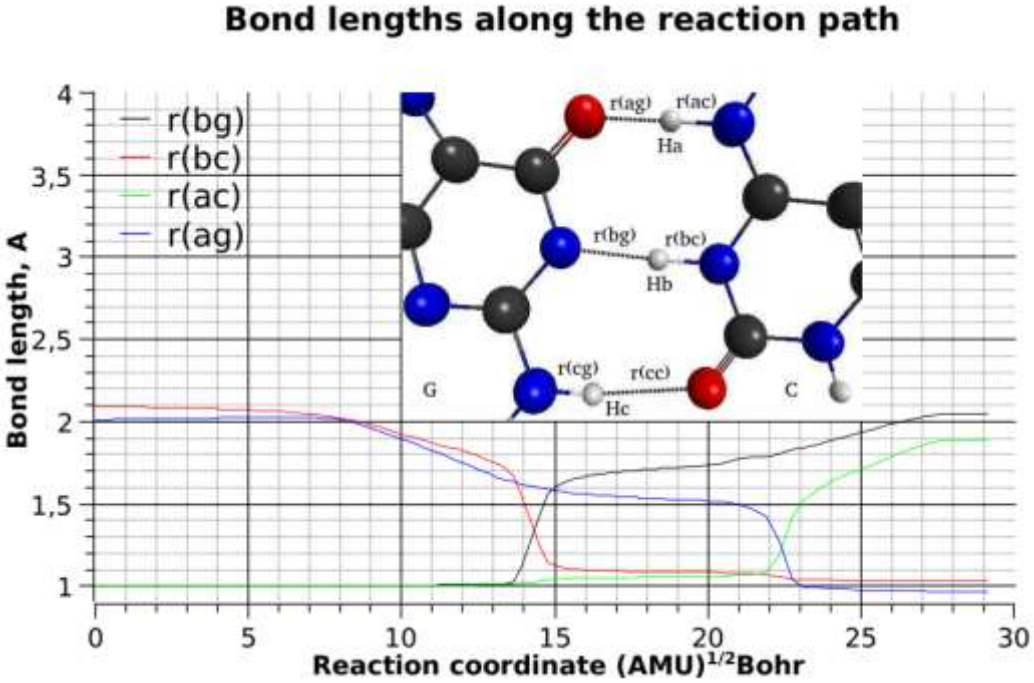



**Fig. 5.** The dependence of bond lengths (angstrom) along the path of transfer of protons GC→G*C* in the CASSCF method.

## CONCLUSIONS

The proton reaction path is characterized in the treated system by a large change of a proton coordinate. We can conclude from Figure 2 and Figure 4 that the energy derivation for the both B3LYP and CASSCF methods is small for proton coordinate near equilibrium value. The situation is drastically changes for a large derivation, B3LYP method gives about one-half of an activation energy compared to that in CASSCF. Even more, the derived two TSs for GC→G*C* reaction change drastically the form of a barrier and consequently the tunneling rate and the time scale.

The G*C* complex is able to exist as a low-stable system relative to reverse reaction G*C*→GC, the activation energy of which is approximately 10kcal/mol. The probability of formation of G*C* pairs in DNA is extremely low due to sufficiently high activation energy of direct reaction, which is 31.66kcal/mol.

Among the other observations, the A*T* complex is not able to exist. This fact was pointed out in [2] on the ground of the estimation of the Gibbs free energy inside a barrier. We can add this is true due to relatively high activation energy (29.24 kcal/mol) and extremely small depth of potential bath for the A*T* structure (less than 2.5 kcal/mol). The AT base pairs can be surely excluded from mutagenesis consideration of DNA via double proton transfer.

The performed CASSCF calculations may indicate that quantum-chemical methods, such as DFT and RHF + MP2, are poorly applicable for description of these reactions of proton transfer. The scope of applicability of such methods may be restricted by the vicinity of equilibrium geometry.

The authors are grateful to the Ioffe Institute for the provided computational resources and thank V.D. Lakhno for literature suggestions.

This work was performed in the framework of the Program of State Orders, project no. 0040-2019-0023.

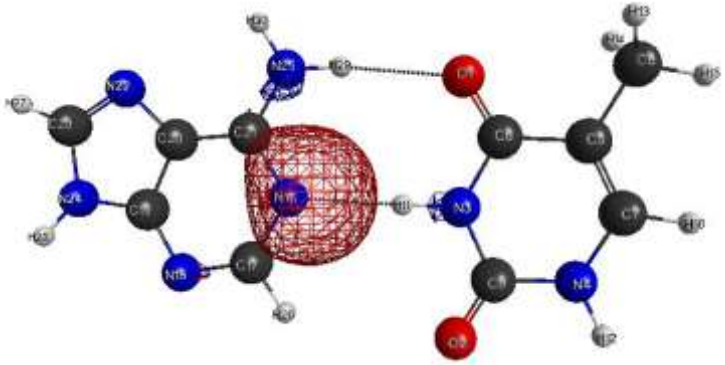
## REFERENCES


1. Srivastava R. *Front. Chem.* 2019. V. 7. P. 536–517.
2. Gorb L., Podolyan Y. Dziekonski P., Sokalski W.A., Leszczynski J. *J. Am. Chem. Soc.* 2004. V. 126. P. 10119–10129.
3. Freitas R.R.Q., Rivelino R., Mota F. de B., Gueorguiev G.K., de Castilho C.M.C. *J. Phys. Chem. C.* 2015. V. 119. № 27. P. 15735–15741.
4. Kumar A., Sevilla M.D. *J. Phys. Chem. B.* 2009. V. 113. № 33. P. 11359–11361.
5. Zengtao Lv, Shouxin Cui, Feng Guo, Guiqing Zhang. *AIP Advances.* 2019. V. 9. № 1. P. 015015.
6. Hayashi T., Mukamel S. *Israel Journal of Chemistry.* 2004. V. 44. № 1–3. P. 185–191.
7. Bezbaruah B., Medhi C. *Indian Journal of Advances in Chemical Science.* 2016. V. 4. № 3. P. 314–320.
8. Slocombe L., Al-Khalili J.S. Sacchi M. *Phys. Chem. Chem. Phys.* 2021. V. 23. P. 4141–4150.
9. Umesaki K., Odai K. *J. Phys. Chem. B.* 2020. V. 124. P. 1715−1722.
10. Slocombe L., Sacchi M., Al-Khalili J. Communications Physics. 2022. V. 5. P. 109–109. doi: 10.1038/s42005-022-00881-8
11. Tulub A.A. *RSC Adv.* 2016. V. 6. № 85. P. 81666–81671.
12. Wahl A.C., Das G. The Multiconfiguration Self-Consistent Field Method. In: *Methods of Electronic Structure Theory. Modern Theoretical Chemistry*. V. 3. Ed.: Schaefer H.F. Boston, MA: Springer. 1977. doi: 10.1007/978-1-4757-0887-5_3
13. Roos B.O. The Complete Active Space Self-Consistent Field Method and its Applications in Electronic Structure Calculations. In: *Advances in Chemical Physics: Ab Initio Methods in Quantum Chemistry*. Ed.: Lawley K.P. 1987. V. 69. P. 399–445. doi: 10.1002/9780470142943.ch7
14. Schmidt M.W., Gordon M.S. *Annu. Rev. Phys. Chem.* 1998. V. 49. P. 233–266.
15. Barca G.M.J., Bertoni C., Carrington L., Datta D., De Silva N., Deustua J.E., Fedorov D.G., Gour J.R., Gunina A.O., Guidez E. et al. *J. Chem. Phys.* 2020. V. 152. № 15. P. 154102–154126.
16. Bode B.M., Gordon M.S. *J. Mol. Graphics and Modelling.* 1998. V. 16. № 3. P. 133–138.

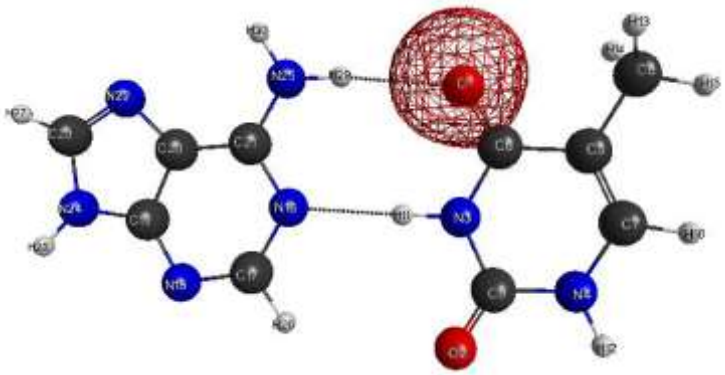

Received 08.02.2023.
Revised 20.02.2023.
Published 26.02.2023.

# APPENDIX 1

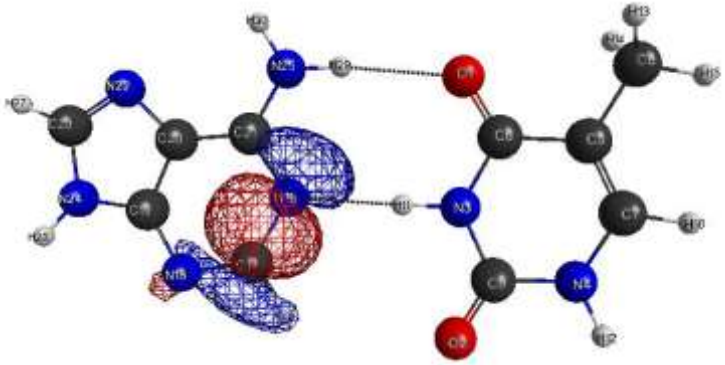
**Table A1.** The shapes of active natural MOs for adenine-thymine base pair in the CASSCF(12,10) method



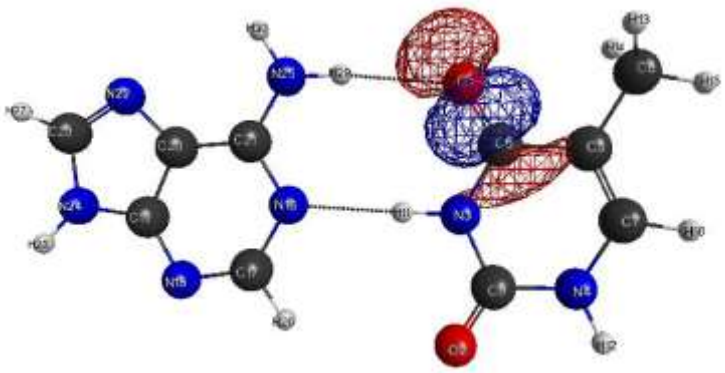
NO #63 $n$ = 1.998



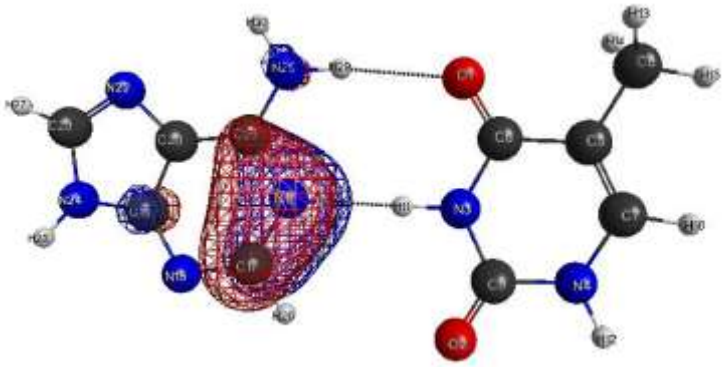
NO #64 $n$ = 1.998



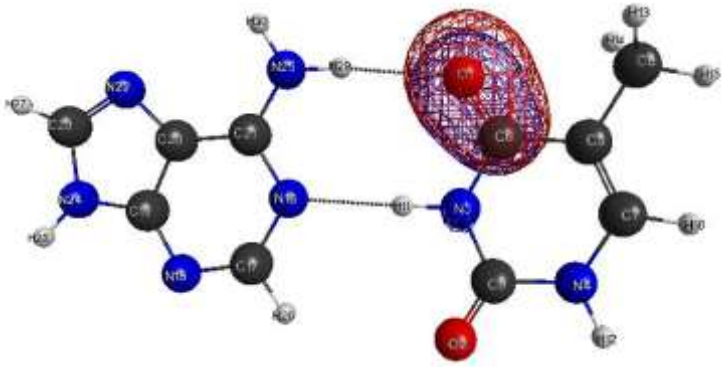
NO #65 $n$ = 1.980



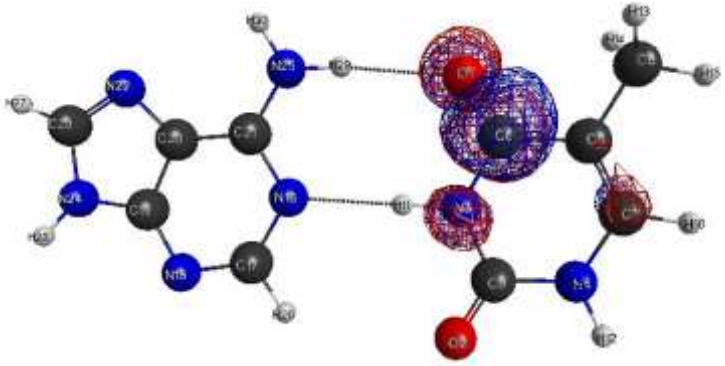
NO #66 $n$ = 1.979



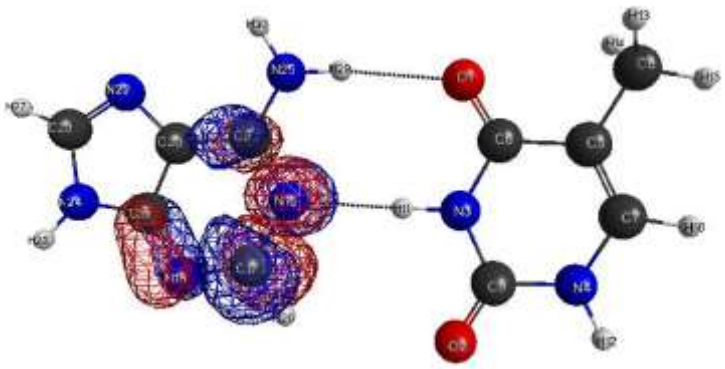
NO #67 $n$ = 1.965



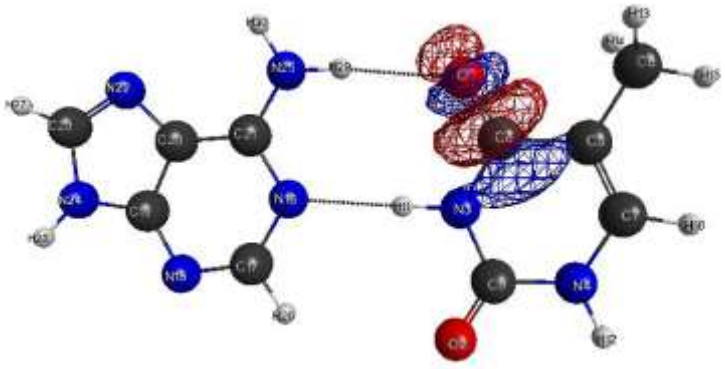
NO #68 $n$ = 1.955



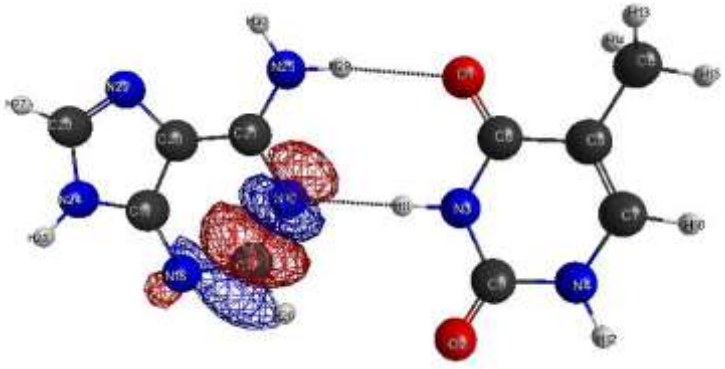
NO #69 $n$ = 0.046



NO #70 $n$ = 0.036



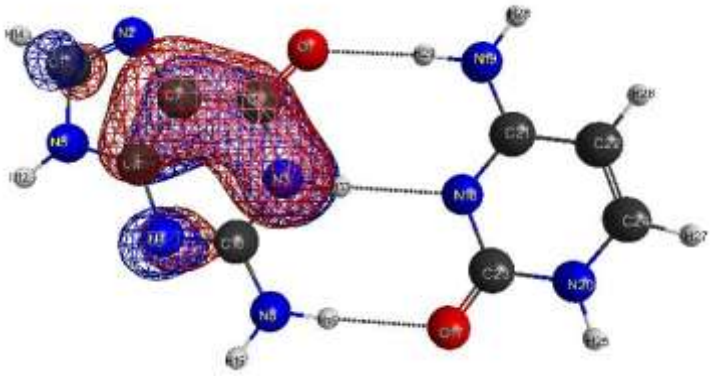
NO #71 $n$ = 0.022



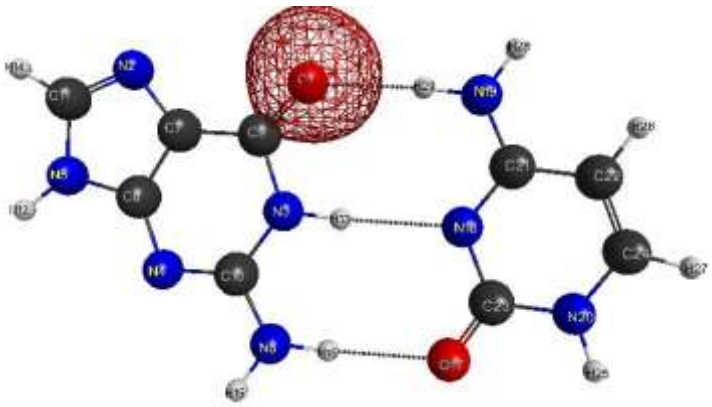
NO #72 $n$ = 0.021

# APPENDIX 2

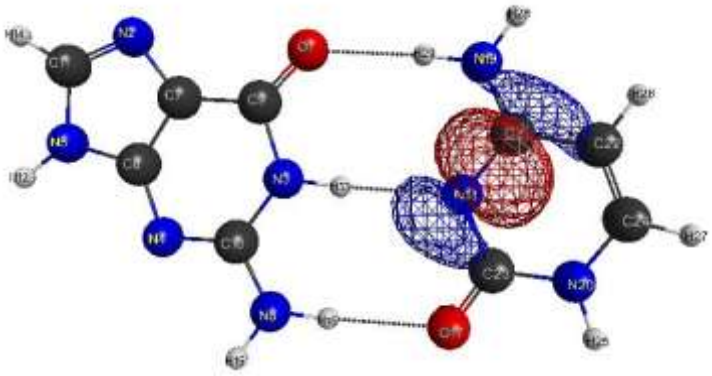
**Table A2.** The shapes of active natural MOs for guanine-cytosine base pair in the CASSCF(12,10) method



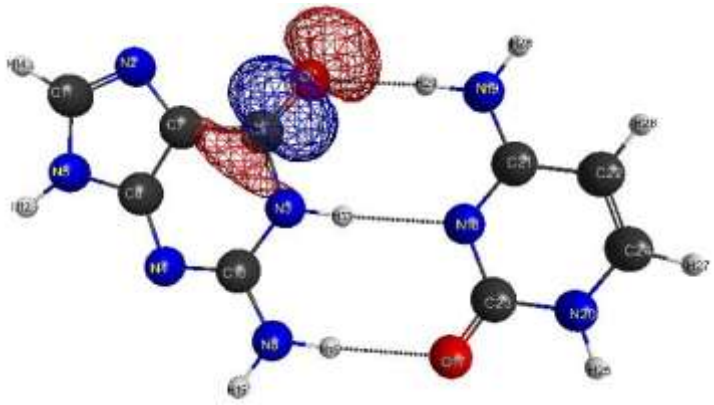
NO #63 $n$ = 1.999



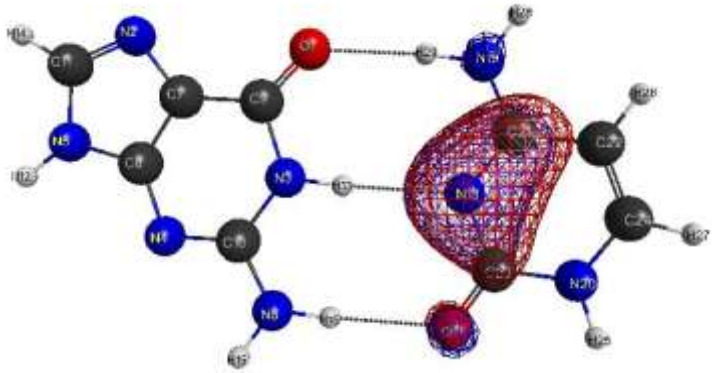
NO #64 $n$ = 1.999



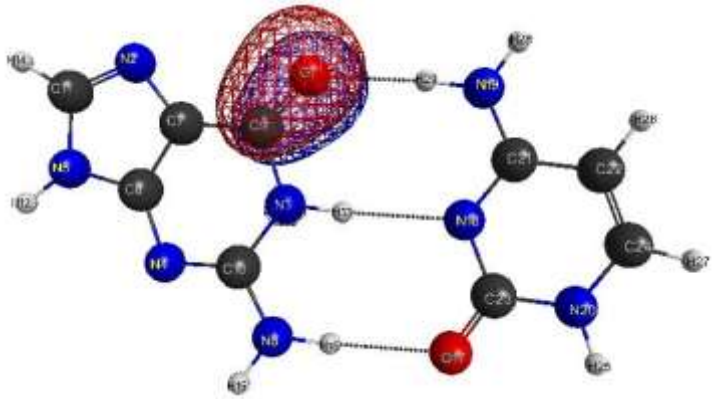
NO #65 $n$ = 1.980



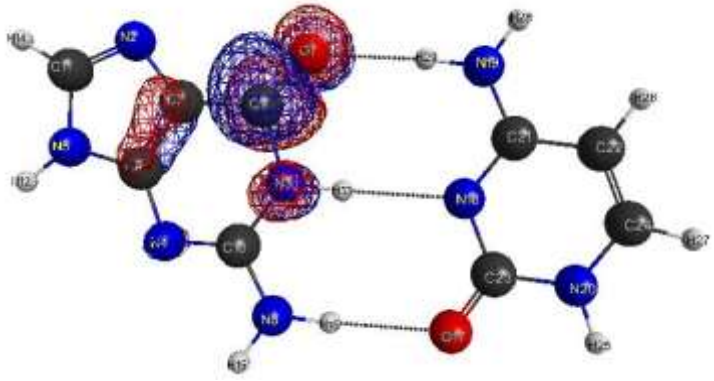
NO #66 $n$ = 1.979



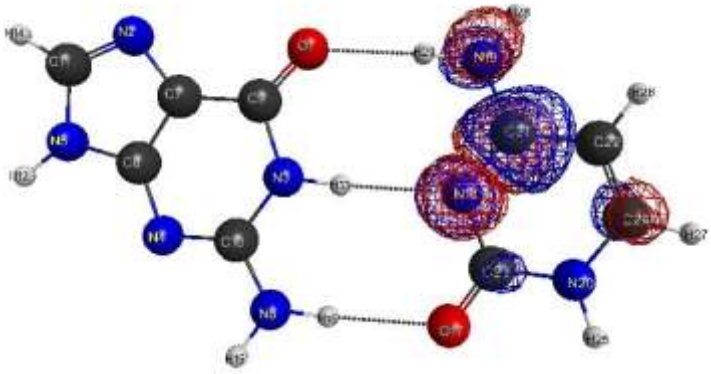
NO #67 $n$ = 1.966



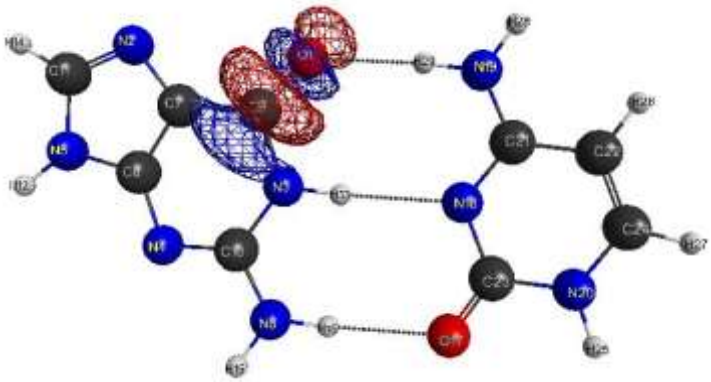
NO #68 $n$ = 1.958



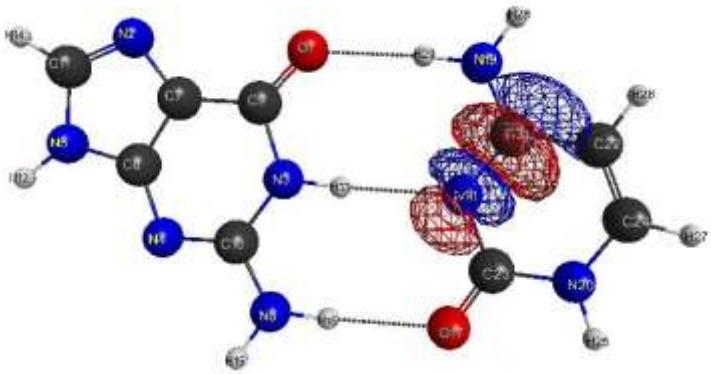
NO #69 $n$ = 0.044



NO #70 $n$ = 0.034



NO #71 $n$ = 0.022



NO #72 $n$ = 0.020