

# The molecular mechanism of the spontaneous substitution mutations caused by tautomerism of bases: Post Hartree-Fock study of the DNA rare base pairs

V. I. Danilov, D. M. Hovorun, Noriyuki Kurita<sup>1</sup>

Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine  
150 Acad. Zabolotnoho st., Kyiv, 03143, Ukraine

<sup>1</sup> Toyohashi University of Technology, Department of Knowledge-Based Information Engineering  
Tempaku-cho, Toyohashi, 441-8580, Japan

---

*Gas-phase gradient optimization of the DNA rare base pairs containing lactam-lactim and amino-imino tautomers was carried out using the Hartree-Fock (HF), Density Functional Theory (DFT) and the second-order Moller-Plesset perturbation (MP2) methods at the 6-31G(d, p) basis set. It is shown that full geometry optimization at the MP2 level leads to an intrinsically nonplanar propeller-twisted and buckled geometry of G\*-T and G-T\* base pairs. The nonplanarity of the pairs is caused by pyramidalization of the amino nitrogen atoms, which is underestimated by the HF and DFT methods. This justifies the importance of geometry optimization at the MP2 level for obtaining reliable prediction of the charge distribution, molecular dipole moments and geometrical structure of the base pairs. The comparison of the formation energies for the rare base pairs shows the energetical preference of the G\*-T and A-C\* base pairs as compared with the G-T\* and A\*-C ones, respectively. It is detected that the stabilization energies of the G-T\* and A\*-C base pairs describing the interaction between monomers are essentially larger than those of the G\*-T and A-C\* base pairs, respectively. An analysis of the decomposition members for molecular HF interaction energies by Morokuma-Kitaura (MK) and the Reduced Variational Space (RVS) methods showed that the nature of a larger stability of the G-T\* and A\*-C base pairs as compared to the G\*-T and A-C\* ones is due to the electrostatic interactions by 60–65 % and the polarization and charge transfer interactions by 35–40 %.*

---

*Key words: spontaneous mutations, tautomerism, rare base pairs, Hartree-Fock, DFT, Moller-Plesset, geometry optimization, nonplanarity of amino group, propeller-twisted and buckled geometry, Morokuma-Kitaura analysis, interaction energy, stability energy.*

---

**Introduction.** One of the possible molecular mechanisms for the formation of spontaneous mutations is conditioned by the tautomerism of DNA bases [1]. The tautomerism of bases can play an important role in the formation of the Watson-Crick-like mismatched base pair. There are two ways for the formation of the rare base pairs from the tautomeric forms of DNA bases: on the template level (replication errors) and on the substrate level (insertion errors). If the

template residue or the incoming substrate (nucleoside triphosphate) is in the wrong tautomeric form when DNA undergoes semi-conservative replication, then incorrect base may be inserted. In case of the absence of correction for this error before the next replication cycle, the two daughter duplexes will have different base pairs at the position of the original mispair.

When forming a DNA double helix, guanine (G) forms a H-bonded pair with cytosine (C). On the

other hand, the rare lactim form of guanine (G\*) forms a pair with thymine (T) instead of C. Similarly, the rare imino form of cytosine (C\*) pairs with adenine (A) instead of G. After the strand separation, the counterbases will form pairs with A and T instead of G and C, respectively.

Thus, the scheme postulated in ref. [1], leads to a spontaneous G-C → A-T transition in subsequent rounds of replication, unless it is detected by the methyl-directed mismatch DNA repair system. In its turn A forms a hydrogen-bonded pair with T. Rare imino tautomer of adenine (A\*) forms pair with C instead of T. The rare lactim form of thymine (T\*) pairs with G instead of A. After the strand separation, the counterbases pair with G and C instead of A and T, respectively. As a result it leads to a spontaneous A-T → G-C transition. In these cases the frequency of mutations is governed by the concentration of template bases on DNA or free substrate nucleoside triphosphates in their minor tautomeric forms in solution.

Therefore spontaneous point mutations can arise from errors during semi-conservative replication. This, however, is very rare due to the exonuclease functions as a proofreading mechanism recognizing mismatched base pairs and excising them, but they can escape at some frequency.

The tautomeric mechanism that does not require the presence of the free rare tautomers in solution may also occur. In this case the rare tautomers are formed in the template via the simultaneous transfer of two protons in the H-bonds of the DNA base pairs. Based upon the Watson-Crick model of DNA, Lowdin [2] pointed out that there was a certain intrinsic probability of proton movement in DNA. Namely the protons in two H-bonds between paired bases change their position in time from the most favourable position to the next most favourable position. This spontaneous shift of the positions, which is characteristic and inherent to a quantum mechanical particle, transforms both bases to their tautomeric forms. The tautomeric form can make a pair only with a base different from the normal partner.

This would cause an error in the genetic information, and the accumulated errors of this kind could be responsible for mutation, aging and spontaneous tumors.

It should be noted that this modified tautomeric mechanism equally with usual tautomerism assumes that these tautomers will remain stable during DNA

unwinding and strand separation, which are the prerequisite steps for the synthesis of new DNA strand by polymerase. However, there has never been any convincing evidence demonstrating that rare tautomeric forms of the bases are responsible for the spontaneous mutagenesis. Recently Goodman suggested [3] that perhaps the rare base pairs exist in the polymerase active site and later shift to the ionized, protonated and wobble base pair structures observed by NMR and X-ray crystallography.

Therefore, the ability to identify the rare base pairs is not only interesting in itself but also may open the way to investigate their properties in relation to ionized and wobble base pairs. Moreover, irrespective of that the rare base pairs is direct or indirect reason of spontaneous mutations.

Very little information on the quantum chemical study of the rare base pairs containing lactam-lactim and amino-imino tautomers has been obtained up to now. Some geometrical and electronic properties as well as interaction energies disregarding the basis set superposition error (BSSE), obtained at the HF/STO-3G and HF/4-31G levels, were presented in the work [4]. During the geometry optimization of the base pairs, only the distances between the bases and their mutual orientations were optimized whereas the coplanarity of base pairs was maintained and the base geometries obtained by the HF/STO-3G method were kept rigid.

The interaction energies of the bases in the G\*-T and A-C\* base pairs calculated by B3LYP/6-31++G(d, p) method were presented in other works [4, 5]. At the same time for the other two rare base pairs (G-T\* and A\*-C) such data in [5] are absent. Besides, the authors of these works [4, 5] proposed that the interaction energy of the bases in the base pairs is the index of the energetic preference of their formation.

As a result the geometrical structure of DNA rare base pairs, energetic aspects of formation and physical picture of their H-bonded pairing with the standard basis set are not yet determined. Moreover, the influence of electron correlation on different properties of the rare base pairs remains obscure. In particular, the detailed analysis of nucleic bases interactions at the HF and post HF levels of the theory is extremely important.

This work presents the results of HF, DFT and MP2 *ab initio* studies at 6-31G(d, p) level of a number of properties of the rare base pairs (dipole

moments, optimal geometries and interaction energies) that have not been previously studied at these levels of the quantum-mechanical theory.

**Methods.** In order to elucidate the above-mentioned questions, we carried out a study of the G\*-T, G-T\*, A\*-C and A-C\* rare base pairs at the HF level, DFT level with functionals B3LYP and B3PW91 and MP2 *ab initio* level to find stationary points on the potential energy surfaces. In addition, the single point calculations have been performed for the studied systems at the MP2/6-311++G(d, p)//MP2/6-31G(d, p) level of the theory.

The correlated calculations are performed within the frozen core approximation [6]. The standard split-valence 6-31G basis sets augmented by a set of Cartesian d-polarization functions placed on heavy atoms and p-polarization functions assigned to hydrogen atoms [6] were used for geometry optimization of the rare base pairs at the HF, DFT and MP2 levels of the theory.

The MP2 geometry optimization was started from the HF/6-31G(d, p) optimized geometry of rare base pairs [7]. Geometry optimization has been continued until the largest component of gradient is smaller than 0.00003 Hartree/Bohr and the root means square gradient is less than 1/3 of the maximal gradient component. Since the intermolecular interaction energy calculated in a finite basis set is a subject to a BSSE, the calculated energy term should be appropriately corrected. The counterpoise corrected MP2 formation energy for the H-bonded AB complex,  $E^F$ , from the usual form of the free base and the rare tautomeric form of the free base is given by [7, 8]

$$E^F = E^{\text{TAU}} + E^{\text{DEF}} + E^{\text{INT}}, \quad (1)$$

where

$$E^{\text{TAU}} = E(\text{rare tautomer}) - E(\text{usual tautomer}) \quad (1a)$$

is tautomerization energy;

$$E^{\text{DEF}} = E^{\text{MP2}}_{\text{A}_A}(\text{AB\_MP2}) - E^{\text{MP2}}_{\text{A}_A}(\text{A\_MP2}) + E^{\text{MP2}}_{\text{B}_B}(\text{AB\_MP2}) - E^{\text{MP2}}_{\text{B}_B}(\text{B\_MP2}) \quad (1b)$$

is the deformation energy describing the effects of the geometry relaxation of subsystems A and B in the dimer;

$$E^{\text{INT}}(\text{AB\_MP2}) = E^{\text{MP2}}_{\text{AB\_AB}}(\text{AB\_MP2}) - E^{\text{MP2}}_{\text{A\_AB}}(\text{AB\_MP2}) - E^{\text{MP2}}_{\text{B\_AB}}(\text{AB\_MP2}) \quad (1c)$$

is the MP2 stabilization energy of the base pair describing the interaction between monomers.

According to the Moller-Plesset perturbation theory [9]

$$E^{\text{INT}} = E^{\text{HF}} + E^{\text{COR}}, \quad (2)$$

where

$$E^{\text{HF}}(\text{AB\_MP2}) = E^{\text{HF}}_{\text{AB\_AB}}(\text{AB\_MP2}) - E^{\text{HF}}_{\text{A\_AB}}(\text{AB\_MP2}) - E^{\text{HF}}_{\text{B\_AB}}(\text{AB\_MP2}) \quad (2a)$$

is the HF interaction energy between bases;

$$E^{\text{COR}} = E^{\text{INT}}(\text{AB\_MP2}) - E^{\text{HF}}(\text{AB\_MP2}) \quad (2b)$$

is the correlation interaction energy within the framework of the second-order Moller-Plesset perturbation theory.

In the above-mentioned expressions the following designations are used:  $E^Y_{X,Z}$  is energy of a system X computed by the Y method with basis set Z. The (AB\_MP2) symbols indicate the geometry of complex AB optimized by the MP2 method.

The  $E^F$  term was corrected by the conventional counterpoise correction method, which eliminates the BSSE. The counterpoise corrected DFT formation energy for the H-bonded AB complex is determined by analogy with the MP2 formation energy. The  $E^F$  term in our analysis includes deformation energy  $E^{\text{DEF}}$  because monomers change their geometry upon formation of the complex. The  $E^{\text{DEF}}$  was calculated as the difference between the energies of the bases in the optimized dimer geometry and the optimized isolated bases.

It should be noted that in the expressions (1), (2) and (2a) the energy terms of the Hartree-Fock interaction energy for the base pairs were calculated on the MP2-optimized geometry since the Hartree-Fock solution appears as the zero-order approximation in the MP2 method (see [9]).

For the elucidation of the nature of the hydrogen bonding and stability of the rare base pairs the Morokuma-Kitaura (MK) [10] and the Reduced Variational Space (RVS) [11] methods of the decomposition for the molecular HF interaction energies ( $E^{\text{HF}}$ ) were used. Different energy contributions determining hydrogen bonding in the rare base pairs, namely the electrostatic energy,  $E^{\text{ES}}$ , exchange repulsion energy,  $E^{\text{XR}}$ , polarization energy,  $E^{\text{PL}}$ , charge transfer energy,  $E^{\text{CT}}$ , and the higher order coupling term,  $E^{\text{MIX}}$ , were evaluated by these methods.

Since the intermolecular interaction energy calculated in a finite basis set is a subject to a basis set BSSE, the calculated energy terms should be appro-

Table 1\*

Dipole moments ( $D$ ), bond lengths ( $A$ ) and bond angles (deg.) of hydrogen bonds of the Guanine (lactam) — Thymine ( $G^*-T$ ), Guanine — Thymine (lactam) ( $G-T^*$ ), Adenine (imino) — Cytosine ( $A^*-C$ ) and Adenine — Cytosine (imino) ( $A-C^*$ ) rare base pairs optimized at HF/6-31G(d, p) and MP2/6-31G(d, p) levels of theory

Property	$G^*-T$	Value	$G-T^*$	Value
Dipole moment	—	5.17 (5.70)	—	5.12 (5.77)
Bond length	O-H( $G^*$ )...O(T)	1.76 (1.89)	O(G)...H( $T^*$ )-O	1.57 (1.73)
	N( $G^*$ )...H(T)-N	1.83 (2.00)	N-H(G)...N( $T^*$ )	1.82 (1.96)
	N-H( $G^*$ )...O(T)	2.02 (2.15)	N-H(G)...O( $T^*$ )	1.97 (2.09)
	O( $G^*$ )-H( $G^*$ )	0.99 (0.96)	N(G)-H(G)	1.03 (1.01)
	N( $G^*$ )-H( $G^*$ )	1.01 (0.99)	N(G)-H(G)	1.02 (1.00)
	N(T)-H(T)	1.04 (1.01)	O( $T^*$ )-H( $T^*$ )	1.02 (0.97)
Bond angle	O( $G^*$ )-H( $G^*$ )...O(T)	170.2 (172.0)	O(G)...H( $T^*$ )-O( $T^*$ )	171.6 (175.8)
	N( $G^*$ )...H(T)-N(T)	175.7 (174.5)	N(G)-H(G)...N( $T^*$ )	170.9 (169.5)
	N( $G^*$ )-H( $G^*$ )...O(T)	175.9 (179.3)	N(G)-H(G)...O( $T^*$ )	174.2 (172.8)

Property	$A^*-C$	Value	$A-C^*$	Value
Dipole moment	—	3.86 (5.30)	—	2.68 (3.39)
Bond length	N( $A^*$ )...H(C)-N	1.87 (1.94)	N-H(A)...N( $C^*$ )	1.98 (2.14)
	N-H( $A^*$ )...N(C)	1.79 (1.84)	N(A)...H( $C^*$ )-N	1.81 (2.00)
	N( $A^*$ )-H( $A^*$ )	1.04 (1.01)	N(A)-H(A)	1.02 (1.00)
	N(C)-H(C)	1.04 (1.02)	N( $C^*$ )-H( $C^*$ )	1.04 (1.01)
Bond angle	N( $A^*$ )...H(C)-N(C)	175.5 (176.1)	N(A)-H(A)...N( $C^*$ )	173.0 (173.9)
	N( $A^*$ )-H( $A^*$ )...N(C)	178.7 (179.3)	N(A)...H( $C^*$ )-N( $C^*$ )	177.4 (178.2)

\*The HF values are given in parentheses.

priately corrected. The  $E^{\text{HF}}$ ,  $E^{\text{XR}}$ ,  $E^{\text{CT}}$ ,  $E^{\text{MIX}}$  energies, the correlation interaction energy,  $E^{\text{COR}}$  and total complex formation energy,  $E^{\text{T}}$  were corrected by the conventional counterpoise correction method, which eliminates the BSSE. It should be noted that the BSSE correction is partially generated by the RVS energy decomposition scheme for the  $E^{\text{XR}}$ ,  $E^{\text{CT}}$  and  $E^{\text{MIX}}$  terms (see [12]).

All *ab initio* calculations were performed by the GAMESS US set of programs ([13], Granovsky Alex A. [www http://classic.chem.msu.su/gran/gamess/index.html](http://classic.chem.msu.su/gran/gamess/index.html)).

**Results and Discussions.** Table 1 summarizes bond lengths and bond angles involved in hydrogen bonds between the bases of the  $G^*-T$ ,  $G-T^*$ ,  $A^*-C$  and  $A-C^*$  rare base pairs as well as dipole moments based on the HF/6-31G(d, p) and MP2/6-31G(d, p) geometry optimization calculations.

As seen from Table 1 the dipole moments predicted by the MP2 methods are noticeably different from those predicted at the HF level. It agrees with

the known fact that the HF approximation overestimates dipole moments. According to the MP2/6-31G(d, p) results, the inclusion of electron correlation during the geometry optimization reduces the dipole moment of the  $G^*-T$  and  $G-T^*$  base pairs by 10 % and 11 %, respectively, in comparison with the dipole moment values calculated by the HF method. A larger reduction of the dipole moment is predicted by the MP2 method for the  $A^*-C$  and  $A-C^*$  base pairs. This reduction constitutes 27 % and 21 %, respectively.

The data obtained show that the neglect of electron correlation certainly distorts the molecular structures of the rare base pairs. Electron correlation brings both subsystems closer to each other. This can be seen from the decrease in the distances between the atoms H and Y in X-H...Y intermolecular H-bonds of the base pairs. At the same time according to the MP2/6-31G(d, p) calculations the distance between the atoms X and Y in the intermolecular H-bond of the  $G^*-T$  and  $G-T^*$  rare base pairs for the studied basis set reduces by 0.10–0.14 Å and 0.10–

Table 2  
Nonplanarity of amino group in the DNA rare base pairs determined by the DFT and MP2 methods in the 6-31G (d, p) basis set\*

Base pair (Method)	Dihedral angles		Sum of NH <sub>2</sub> valency angles
	N3C2N2H21	N1C2N2H22	
<i>Guanine</i>			
G*-T (DFT/B3LYP)	-7.1°	-9.4°	358.1°
G*-T (DFT/B3PW91)	+5.0°	-6.8°	359.0°
G*-T (MP2)	+15.0°	-25.3°	348.2°
G-T* (DFT/B3LYP)	+3.4°	-4.3°	359.6°
G-T* (DFT/B3PW91)	+2.6°	-3.6°	359.7°
G-T* (MP2)	+15.3°	-19.2°	352.2°
Base pair (Method)	Dihedral angles		Sum of NH <sub>2</sub> valency angles
	C5C4N4H41	N3C4N4H42	
<i>Cytosine</i>			
A*-C (DFT/B3LYP)	-0.4°	+0.9°	360.0°
A*-C (DFT/B3PW91)	-0.3°	+0.7°	360.0°
A*-C (MP2)	-1.3°	+1.3°	360.0°
Base pair (Method)	Dihedral angles		Sum of NH <sub>2</sub> valency angles
	C5C6N6H61	N1C6N6H62	
<i>Adenine</i>			
A-C* (DFT/B3LYP)	+0.4°	-0.7°	360.0°
A-C* (DFT/B3PW91)	+0.4°	-0.6°	360.0°
A-C* (MP2)	+2.7°	-2.5°	359.8°

\*A value equal to 360° corresponds to a planar amino group. Decline from the planar state serves as a pyramidalization level.

0.11 Å, respectively, in comparison with the HF optimized geometry. Analogous decreases in the A\*-C and A-C\* base pairs constitute 0.02–0.05 Å and 0.14–0.15 Å. This is rather a significant shortening caused by the dispersion attraction, which is taken into account by the MP2 method.

It is interesting to note that the bond lengths of the same H-bonds of the stereoisomer mismatches containing rare tautomers of bases are noticeably different. Especially it concerns O-H...O hydrogen bond of the G-T\* base pair and N-H...N hydrogen bond of the A\*-C base pair that are shorter by 0.2 Å and 0.1 Å, respectively.

The geometric data characterizing the amino

groups of bases for the rare base pairs optimized by the post HF methods are given in Table 2 (the numbering system corresponds to the IUPAC recommendations on Nomenclature of Organic Chemistry).

It can be seen that according to the MP2 calculations the amino group of guanine in the G\*-T and G-T\* base pairs is essentially nonplanar. At the same time the amino group of cytosine in the A\*-C and adenine in the A-C\* base pairs is planar.

The analysis of the optimized molecular structure for the G\*-T and G-T\* base pairs shows that the amino group hydrogens of guanine deviate from the base plane in the direction, opposite to the shift of the amino group nitrogen atom (the sp<sup>3</sup>-hybridized struc-

Table 3

The geometrical and energetical (kcal/mol) characteristics of hydrogen-bonded rare base pairs, obtained at the DFT(B3LYP/6-31G(d, p)), DFT(BPW91/6-31G(d, p)), MP2/6-31G(d, p) and MP2/6-311++G(d, p)//MP2/6-31G(d, p)-optimized geometries

Base Pair (Method/Basis)	PT	Buckle	$E^{\text{TAU}}$	$E^{\text{DEF}}$	$E^{\text{INT}}$	$E^{\text{F}}$
G*-T (B3LYP/6-31G(d, p))	3°	3°	0.82	3.41	-19.33	-15.10
G*-T (B3PW91/6-31G(d, p))	3°	3°	0.69	3.99	-19.51	-14.83
G*-T (MP2/6-31G(d, p))	10°	7°	1.07	2.63	-17.74	-14.04
G*-T (MP2/6-311++G(d, p)//MP2/6-31G(d, p))	10°	7°	0.06	2.62	-16.69	-14.31
G-T* (B3LYP/6-31G(d, p))	1°	3°	13.07	8.09	-35.15	-13.99
G-T* (B3PW91/6-31G(d, p))	1°	3°	13.07	8.09	-35.15	-13.99
G-T* (MP2/6-31G(d, p))	5°	2°	13.35	6.13	-30.71	-11.23
G-T* (MP2/6-311++G(d, p)//MP2/6-31G(d, p))	5°	2°	11.92	6.19	-29.64	-11.53
A*-C (B3LYP/6-31G(d, p))	0°	1°	12.06	2.88	-23.07	-8.13
A*-C (B3PW91/6-31G(d, p))	0°	1°	13.72	3.42	-23.68	-6.54
A*-C (MP2/6-31G(d, p))	0°	1°	12.36	2.68	-22.29	-7.25
A*-C (MP2/6-311++G(d, p)//MP2/6-31G(d, p))	0°	1°	12.71	2.80	-21.80	-6.29
A-C* (B3LYP/6-31G(d, p))	0°	1.5°	1.25	1.74	-14.66	-11.67
A-C* (B3PW91/6-31G(d, p))	0°	1.5°	1.52	2.11	-14.90	-11.27
A-C* (MP2/6-31G(d, p))	0°	1°	0.76	1.70	-14.33	-11.87
A-C* (MP2/6-311++G(d, p)//MP2/6-31G(d, p))	0°	1°	1.57	1.91	-14.16	-10.68

ture of the nitrogen atom valence shell of an NH<sub>2</sub>-group or the partial pyramidalization of the amino group of the DNA base). Examination of the base pairs structure suggests that the source of the nonplanarity is a geometrical peculiarity of the NH<sub>2</sub> group of guanine.

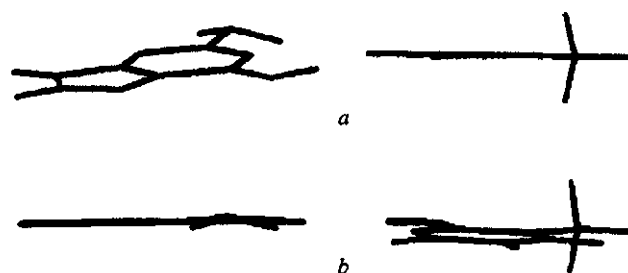
At the same time it should be noted that the DFT approach suggests a very weak nonplanarity of the amino groups of bases (see Table 2). As a result of underestimation of the pyramidalization effect, the DFT method leads to almost planar or perfect planar structure for the rare (see Table 2) and canonical base pairs (see, for example, [7]). It should be noted that our results on the geometry optimization of the rare base pairs by the HF/4-31G method without any geometrical constraint (compare to ref. [4]) as well as HF/6-31G(d, p) led to the planar structures.

Meanwhile the question about the nonplanarity of the nucleic acid bases and base pairs may have important consequences for the realization of the particular structures of these compounds in various molecular complexes. The potential biological importance of interactions involving nonplanar amino groups of bases was repeatedly stressed by Sponer et al. in the corresponding reviews [14, 15]. The data characterizing the molecular structure of the rare base pairs that have been obtained in our calculations is of

interest. The geometry optimization conducted at the DFT and MP2 levels leads to an intrinsically nonplanar canonical G\*-T and G-T\* base pairs (see Table 3). In other words, the bases in a base pair are not coplanar, instead they are twisted about the hydrogen bonds that connect them.

As it is well known the orientation of individual bases within each base pair can be described in terms of propeller twist (PT) and buckle parameters [16] which characterize the rotation of bases around the long axis of a base pair and the inclination of mean planes of bases with respect to each other. In fact, the PT angle is the dihedral angle that defines the noncoplanarity, and buckle angle is the dihedral angle between bases along their short axis. PT and buckle angles are secondary parameters, which simply describe the imperfections, i. e. nonplanarity, of a given base pair.

Especially noticeable are the angular characteristics of two base pairs (magnitudes of propeller and buckle angles) obtained by the MP2 method (see Figure, *a* and *b*). So the propeller angles between the base planes consist of 10° and 5° for the G\*-T and G-T\* base pairs, respectively, whereas the buckle angles are 7° and 2°. Supposedly, the main reason of nonplanarity of the G\*-T and G-T\* base pairs is pyramidalization of the amino group of guanine. The



Structures of the rare base pairs determined by MP2/6-31G(d, p) method: *a* — G\*-T; *b* — G-T\*

conventional planar structure of bases, as would be expected, gives only the coplanar type of H-pairing. Among other possible distortion factors of nonplanarity of the pairs are a wide variety of secondary long-range electrostatic interactions, involving the hydrogen atoms bonded to ring carbon atoms, and steric reasons (see more information [14, 15]). Therefore the nonplanarity of isolated DNA rare base pairs as well as that of the isolated Watson-Crick ones (see [7]) is their intrinsic property. The propeller base twisting and buckling in isolated rare base pairs has been obtained without any attraction of the stacking perturbation hypothesis that was marked earlier [7, 17].

According to the geometric selection mechanism of bases as a principal determinant of DNA replication fidelity [17–20] the geometrical and electrostatic properties of the polymerase active site are likely to have a profound influence on nucleotide-insertion specificities. This influence would strongly favor the insertion of bases having an optimal geometry, such that the C1'(N9)-C1'(N1) distances and bond angles most closely approximate those of the Watson-Crick base pairs.

The detailed study of the obtained geometric characteristics for the optimized rare and Watson-Crick base pairs showed the following. The distance between the bonds joining the bases to the deoxyribose groups in the G\*-T and G-T\* rare base pairs is close to the corresponding canonical distance in the G-C base pair, while this distance in the A\*-C and A-C\* base pairs is close to that in the A-T base pair. Moreover, in each pair of stereoisomers the C1'-N9 and C1'-N1 bonds make an angle with C1'(9)-C1'(1) line that is close to the corresponding

values in one of the Watson-Crick canonical base pairs. The analogous conclusion made earlier Topal and Fresco [21] who studied each of the above-mentioned rare base pairs by the model building and showed that these pairs were sterically compatible with a Watson-Crick helix. Therefore the formation of the DNA rare base pairs with such geometry is compatible with the geometric constraints of the standard double helical DNA. If these mispairs were to be incorporated in a standard Watson-Crick double helix, the helix would not likely be highly distorted and its stability significantly did not reduce.

At the same time it should be noted that the experimentally detected the G-T, G-A, and C-A mispairs have markedly different C1'(9)-C1'(1) distances and glycosyl bond angles than those of the A-T and G-C pairs. The striking geometric identity of the Watson-Crick A-T and G-C base pairs is not matched by the A-C protonated wobble and G-T wobble base mispairs or by the G(anti)-A(syn) base mispair. Therefore the geometric constraints imposed on the substrate and template bases at the polymerase active site are not the only ones for incorporation of the non-Watson-Crick base pairs in DNA.

In the Table 3 the values of the tautomerization energy  $E^{\text{TAU}}$  (see (1a)), deformation energy  $E^{\text{DEF}}$  (see (1b)), stabilization energy  $E^{\text{INT}}$  (see (1c)), formation energy  $E^{\text{F}}$  (see (1)) are also given. The comparison of the formation energies of the canonical [7] and rare base pairs (see Table 3) shows that the formation of the Watson-Crick G-C base pair is most preferable among all studied base pairs. At the same time the formation of the G\*-T base pair is more preferable than the A-T one. The A-C\* and G-T\* base pairs are only slightly energetically less preferable than the A-T base pair. The direct comparison of the computed energies for two pairs of stereoisomers also shows the energetic preference of the G\*-T and A-C\* base pairs towards the G-T\* and A-C ones, respectively. Our results led to the interesting fact that the stability of the G-T\* base pair is much larger than that of the canonical Watson-Crick G-C base pair (see [7]) which is considered at present as the largest among all studied base pairs.

It is also seen from Table 3 that the stabilization energy of the G-T\* base pair is essentially larger than that of the G\*-T, irrespective of computation methods. However, for the formation of the G-T\* rare base pair the usual form of T must be replaced by rare tautomeric form of T\* that requires the energy

Table 4  
Components of molecular interaction energy for the DNA rare base pairs, kcal/mol

Base Pair (ED)	$E^{ES}$	$E^{XR}$	$E^{PL}$	$E^{CT}$	$E^{MIX}$	$E^{HF}$	$E^{COR}$	$E^{DEF}$	$E^T$
G*-T (MK)	-34.48	32.92	-5.60	-9.04	1.50	-14.70	-3.04	2.63	-15.11
G*-T (RVS)	-34.48	32.34	-5.66	-8.00	—	-15.80	-3.04	2.63	-16.21
G-T* (MK)	-50.48	44.81	-12.00	-13.03	0.66	-30.04	-0.67	6.13	-24.58
G-T* (RVS)	-50.48	44.20	-11.75	-10.90	—	-28.93	-0.67	6.13	-23.47
A*-C (MK)	-37.51	32.48	-8.71	-8.52	1.65	-20.61	-1.68	2.68	-19.61
A*-C (RVS)	-37.51	31.96	-7.77	-7.04	—	-20.36	-1.68	2.68	-19.36
A-C* (MK)	-27.82	27.12	-5.11	-6.97	1.83	-10.95	-3.38	1.70	-12.63
A-C* (RVS)	-27.82	26.63	-4.67	-5.88	—	-11.84	-3.38	1.70	-13.52

Note. ED — energy decomposition scheme; MK — Morokuma-Kitaura energy decomposition scheme; RVS — reduced variational space energy decomposition scheme, complex formation energy disregarding tautomerization energy  $E^T = E^F - E^{TAU}$ .

consumption equal to 11.92—13.35 kcal/mol. Furthermore, the formation of such base pair is accompanied by expense of 6.13—8.09 kcal/mol of the deformation energy describing the effects of the geometry relaxation of its G and T\* bases. As a result, the formation of the G-T\* base pair becomes energetically less preferable than that of the G\*-T one. Table 3 also shows that the stabilization energy of the A\*-C base pair is essentially larger than that of the A-C\*. On the other hand, for the formation of the A\*-C base pair the usual form of A must be replaced by rare tautomeric form of A\* that requires the energy consumption equal to 12.06—13.72 kcal/mol. Moreover, for the geometry relaxation of A\* and C bases in this base pair the expenses of 2.68—3.62 kcal/mol are required. So the formation of the A\*-C base pair is energetically less preferable than that the A-C\* one.

Therefore, the obtained data directly show that the calculated interaction energies of bases in the rare base pairs are insufficient in order to characterize the relative ease or difficulty of incorporating base pairs into a double helix. In particular, it was done in the work [4, 5]. An appreciably larger stability of the G-T\* and A\*-C base pairs is of particular interest in view of large similarity of their molecular structure to the corresponding stereoisomers. For understanding this result, the MK and RVS analysis of molecular interaction energy components for the DNA rare base pairs was carried out. The calculated values of these components are given in Table 4. The analysis of the decompositions terms showed that the nature of a larger stability of the G\*-T and A\*-C base pairs, as

compared to those of the G-T\* and A-C\* ones, respectively, by 60—65 % is due to the electrostatic interactions. At the same time polarization and charge transfer interactions also make considerable contribution (by 35—40 %) to a larger stability of the above-mentioned base pairs. It should be emphasized that according to the Table 4 the correlation interaction makes a noticeably larger contribution to the stability of the G\*-T and A-C\* base pairs than to that of the G-T\* and A\*-C ones. Therefore, a larger stability of the G-T\* and A\*-C base pairs is not related to the correlation interaction.

As a result of our calculations, the following conclusion can be done. According to the energetical point of view, the formation of the G\*-T and A-C\* base pairs must lead to the spontaneous mutations more often in comparison with the G-T\* and A\*-C ones, respectively. The analysis of possible transitions during the replication and insertion errors and the obtained formation energies of the base pairs show that energetically the most probable transitions are those for which tautomeric transformations occur in the G. The transitions for which tautomeric transformations occur in the C and T are less probable. In other words, in the case of lactam-lactim and amino-imino tautomerism the replication errors mostly lead to the transitions G-C → A-T, whereas the insertion errors mostly lead to the transitions A-T → G-C. The transitions for which tautomeric transformations occur in the T are even less probable. The least probable transitions are those conditioned by tautomeric transformations in the A.

The obtained data might testify in favour of the



possibility of origin of the spontaneous mutations due to the tautomerism of bases. Unfortunately, the origin of the spontaneous mutations is conditioned not only by energy factors, but also by other reasons. Among them are such important factors as the entropy and the possible geometric selection mechanism in the exonuclease active site, enhancing the excision of non-Watson-Crick base pairs. So, the discussed problem is in need of the further investigation.

*В. І. Данілов, Д. М. Говорун, Норіюкі Куріта*

Молекулярний механізм спонтанних мутацій заміщення, обумовлених таутомерією основ: ПостХартри-Фоківське вивчення рідкісних пар основ ДНК

Резюме

Здійснено газофазну градієнтну оптимізацію рідкісних пар основ ДНК, які включають у себе лактам-лактимні і аміноімінні таутимери, за допомогою методу Хартри-Фока (ХФ), теорії функціонала густини (ТФГ) та другого порядку теорії збурень Моллера-Плессета (МП2) у базисі 6-31G(d, p). Показано, що повна оптимізація геометрії на рівні МП2 веде до внутрішньої неплоскої пропелер-обертальної і вигнутої геометрії пар основ G\*-T і G-T\*. Неплощинність пар обумовлена пірамідалізацією атомів азоту аміногрупи, яка недооцінюється методами ХФ і ТФГ. Це виправдовує важливість оптимізації геометрії на рівні МП2 для отримання поміркованого передбачення розподілу зарядів, молекулярних дипольних моментів і геометричної структури пар основ. Порівняння енергій формування рідкісних пар основ демонструє енергетичну перевагу пар основ G\*-T і A-C\* стосовно пар G-T\* і A\*-C. Виявлено, що величини енергій стабілізації пар основ G-T\* і A\*-C, які описують взаємодію між мономерами, значно більші за аналогічні значення пар основ G\*-T і A-C\* відповідно. Використовуючи аналіз членів розкладання ХФ енергій молекулярної взаємодії методами Морокуми-Китаури і зменшеного варіаційного простору, знайдено, що природа більшої стабільності пар основ G-T\* і A\*-C порівняно з парами G\*-T і A-C\* на 60—65 % обумовлена електростатичними взаємодіями і на 35—40 % — поляризаційними взаємодіями і взаємодіями з перенесенням заряду відповідно.

Ключові слова: спонтанні мутації, таутомерія, рідкісні пари основ, Хартри-Фок, ТФГ, Моллер-Плессет, оптимізація геометрії, неплоска аміногрупа, пропелер-обертальна і вигнута геометрія, метод Морокуми-Китаури, енергія взаємодії, енергія стабілізації.

*В. И. Данилов, Д. Н. Говорун, Норіюкі Куріта*

Молекулярний механізм спонтанних мутацій заміщення, обумовлених таутомерією оснований: ПостХартри-Фоківське вивчення рідких пар оснований ДНК

Резюме

Осуществлена газофазная градиентная оптимизация редких пар оснований ДНК, содержащих лактам-лактимные и аминиминные таутимеры, с помощью метода Хартри-Фока (ХФ), теории функционала плотности (ТФП) и теории возмущений Моллера-Плессета второго порядка (МП2) в базисе 6-31G(d, p). Показано, что полная оптимизация геометрии на уровне МП2 приводит к внутренне неплоской пропеллер-вращатель-

ной и согнутой геометрии пар оснований G\*-T и G-T\*. Неплощинность пар обусловлена пираимидализацией атомов азота аминогруппы, недооцениваемой методами ХФ и ТФП. Это оправдывает важность оптимизации геометрии на уровне МП2 для получения разумного предсказания распределения зарядов, молекулярных дипольных моментов и геометрической структуры пар оснований. Сравнение энергий образования редких пар оснований демонстрирует энергетическую предпочтительность пар оснований G\*-T и A-C\* относительно пар G-T\* и A\*-C. Обнаружено, что величины энергий стабилизации пар оснований G-T\* и A\*-C, описывающие взаимодействие между мономерами, существенно больше аналогичных значений пар оснований G\*-T и A\*-C соответственно. Используя анализ членов разложения для ХФ энергий молекулярного взаимодействия методами Морокуми-Китаури и сокращенного вариационного пространства, найдено, что природа большей стабильности пар оснований G-T\* и A\*-C по сравнению с парами G\*-T и A-C\* на 60—65 % обусловлена электростатическим взаимодействием и на 35—40 % — поляризационными взаимодействиями и взаимодействиями с переносом заряда соответственно.

Ключевые слова: спонтанные мутации, таутомерия, редкие пары оснований, Хартри-Фок, ТФП, Моллер-Плессет, оптимизация геометрии, неплоская аминогруппа, пропеллер-вращательная и изогнутая геометрия, метод Морокуми-Китаури, энергия взаимодействия, энергия стабилизации.

## REFERENCES

1. Watson J. D., Crick F. H. C. The structure of DNA // Cold Spring Harbor Symp. Quant. Biol.—1953.—18.—P. 123—131.
2. Lowdin P. O. Quantum genetics and the aperiodic solid. Some aspects of the biological problems of heredity, mutations, aging and tumors in view of the quantum theory of the DNA molecule // Adv. Quant. Chem.—1965.—2.—P. 213—360.
3. Goodman M. F. Mutations caught in the act // Nature.—1995.—378.—P. 237—238.
4. Jiang S.-P., Raghunathan G., Ting K.-L., Xuan J. C. Geometries, charges, dipole moments, and interaction energies of normal, tautomeric and novel bases // J. Biomol. Struct. and Dyn.—1994.—12.—P. 367—382.
5. Krechkivska O. M., Kosach D. A., Sudakov O. O., Hovorun D. M. Is the complementarity of canonical nucleotide bases in DNA their inherent physical-chemical characteristics? Results of the non-empirical quantum-chemical investigation // Biopolimery i kletka.—2003.—19, N 4.—P. 382—385.
6. Hehre W. J., Radom L., Schleyer P. v. R., Pople J. A. *Ab initio* molecular orbital theory.—New York: J. Wiley, 1986.
7. Danilov V. I., Anisimov V. M. Post Hartree-Fock studies of the canonical Watson-Crick DNA base pairs: Molecular structure and the nature of stability // J. Biomol. Struct. and Dyn.—2004.—N 4.—P. 471—482.
8. Chalasiński G., Szczesniak M. M. Origins of structure and energetics of van der Waals clusters from *ab initio* calculations // Chem. Rev.—1994.—94.—P. 1723—1765.
9. Moller C., Plesset M. S. Note on an approximation treatment for many-electron systems // Phys. Rev.—1934.—46.—P. 618—622.
10. Morokuma K., Kitaura K. Energy decomposition analysis // Chemical applications of atomic and molecular electrostatic potentials / Eds P. Politzer, D. G. Truhlar.—New York: Plenum press, 1981.—P. 215—242.
11. Stevens W. J., Fink W. H. Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to

- the water dimer // *Chem. Phys. Let.*—1987.—139.—P. 15—22.
12. *Gordon M. S., Jensen J. H.* Wavefunctions and chemical bonding: interpretation // *The Encyclopedia of Computational Chemistry.*—New York: J. Wiley, 1998.—P. 3198—3214.
  13. *Schmidt M. W., Baldridge K. K., Boatz J. A., Elbert S. T., Gordon M. S., Jensen J. H., Koseki S., Matsunaga N., Nguyen K. A., Su S. J., Windus T. L., Dupuis M., Montgomery J. A.* General atomic and molecular electronic structure system // *J. Comput. Chem.*—1993.—14.—P. 1347—1363.
  14. *Hobza P., Sponer J.* Toward true DNA base-stacking energies: MP2, CCSD(T), and complete basis set calculations // *J. Amer. Chem. Soc.*—2002.—124.—P. 11802—11808.
  15. *Sponer J., Leszczynski J., Hobza P.* Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases // *Biopolymers (Nucl. Acid Sci.)*—2002.—61.—P. 3—31.
  16. *Dickerson R. E., Bansal M., Calladine C. R., Diekmann S., Hunter W. N., Kennard O., von Kitzing E., Lavery R., Nelson H. C. M., Olson W. K., Saenger W., Shakked Z., Sklenar H., Soumpasis D. M., Tung C.-S., Wang A. H.-J., Zhurkin V. B.* Definitions and nomenclature of nucleic acid structure parameters // *J. Mol. Biol.*—1989.—208.—P. 787—791.
  17. *Komarov V. M., Polozov R. V., Konoplev G. G.* Non-planar structure of nitrous bases and non-coplanarity of Watson-Crick pairs // *J. Theor. Biol.*—1992.—155.—P. 281—294.
  18. *Sloane D. L., Goodman M. F., Echols H.* The fidelity of base selection by the polymerase subunit of DNA polymerase III holoenzyme // *Nucl. Acids Res.*—1988.—16.—P. 6465—6475.
  19. *Echols H., Goodman M. F.* Fidelity mechanisms in DNA replication // *Ann. Rev. Biochem.*—1991.—60.—P. 477—511.
  20. *Goodman M. F.* Hydrogen bonding revisited: Geometric selection as a principal determinant of DNA replication fidelity // *Proc. Nat. Acad. Sci. USA.*—1997.—94.—P. 10493—10495.
  21. *Topal M. D., Fresco J. R.* Complementary base pairing and the origin of substitution mutations // *Nature.*—1976.—263.—P. 285—289.

УДК 577.112.7

Надійшла до редакції 03.08.03