PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

How the long G·G* Watson-Crick DNA base mispair comprising keto and enol tautomers of the guanine tautomerises? The results of the QM/QTAIM investigation

Ol'ha O. Brovarets' and Dmytro M. Hovorun✉

[a]Department of Molecular and Quantum Biophysics, Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, 150 Akademika Zabolotnoho Str., 03680 Kyiv, Ukraine
[b]Department of Molecular Biotechnology and Bioinformatics, Institute of High Technologies, Taras Shevchenko National University of Kyiv, 2-h Akademika Hlushkova Ave., 03022 Kyiv, Ukraine
✉Corresponding author. Email: dhovorun@imbg.org.ua

Abstract. The double proton transfer (DPT) in the long G·G* Watson-Crick base mispair (|C6N1(G*)N1C6(G)|=36.4º; C_1 symmetry), involving keto and enol tautomers of the guanine (G) nucleobase, along two intermolecular neighboring O6H···O6 (8.39) and N1···HN1 (6.14 kcal·mol^{-1}) H-bonds, that were established to be slightly anti-cooperative, leads to its transformation into the G*·G base mispair through a single transition state (|C6N1N1C6|=37.1º; C_1), *scilicet* to the interconversion into itself. It was shown that the G·G*↔G*·G tautomerisation *via* the DPT is assisted by the third specific contact, that sequentially switch along the intrinsic reaction coordinate (IRC) in an original way: (G)N2H···N2(G*) H-bond (-25.13 ÷ -10.37) → N2···N2 van der Waals contact (-10.37 ÷ -9.23) → (G)N2···HN2(G*) H-bond (-9.23 ÷ 0.79) → (G*)N2···HN2(G) H-bond (0.79 ÷ 7.35 Bohr). The DPT tautomerisation was found to proceed through the asynchronous concerted mechanism by employing the QM/QTAIM approach and the methodology of the scans of the geometric, electron-topological, energetic, polar and NBO properties along the IRC. The 9 key points, that can be considered as the part of the tautomerisation repertoire, have been established and analyzed in detail. Furthermore, it was shown that the G·G* or G*·G base mispair is thermodynamically and dynamically stable structure with a lifetime of $8.22 \cdot 10^{-10}$ s and all 6 low-frequency intermolecular vibrations are able to develop during this time span. Lastly, our results highlight the importance of the G·G*↔G*·G DPT tautomerisation, that can have implications for the biological and chemical sensing applications.

Keywords: Spontaneous point mutations · Keto and enol tautomers · Mutagenic tautomerisation · The double proton transfer · H-bond · Van der Waals contact · Cooperativity of the H-bonds · B3LYP and MP2 · QTAIM

Introduction. One of the traditional biologically important topics of molecular and quantum biophysics, which despite its long and complicated history is far from definitive understanding at the atomic level, is the physical nature of the spontaneous point mutations [1-6]. As of today, it was found that the formation of mismatches with geometry close to the geometry of the canonical A·T and G·C Watson-Crick base pairs in the base pair recognition pocket of the DNA polymerase is their structural root cause [7]. All irregular purine·pyrimidine [8-13], pyrimidine·pyrimidine [14-17] and purine·purine [8,18-23] DNA base mispairs, that are active players on the field of the spontaneous mutagenesis, have been identified as a result of the painstaking experimental [23-25], as well as theoretical [8,9] efforts. At the same time, the first theoretical attempts have been made in order to understand how these incorrect base pairs are formed in the recognition pocket of the high-fidelity DNA-polymerase [18-20]. Thus, the atomistic molecular dynamics simulations of the G·G mismatched base pair (primer·template nucleotides) also revealed its distorted geometry in the polymerase (Pol) active site and inefficient extension, coinciding with the experimentally deduced inability of Pol β to extend this mispair [21]. In the work [22] an assumption has been expressed, that the incorrect purine·purine A*·A_{syn} [26], G·A_{syn} and A*·G*$_{syn}$ [22] base mispairs (here and below the rare, particularly mutagenic [27-33], tautomers are marked with an asterisk) with architecture similar to the Watson-Crick DNA base pairs are formed by the one and the same scenario – conformational transition through the intermediate – the long A*·A, G·A and A·G Watson-Crick base mispairs, respectively. It remains unclear whether this scenario can be applied to the fourth incorrect G·G*$_{syn}$ purine·purine base mispair [34], i.e. whether it is universal for all aforementioned purine·purine pairs.

In order to shed light on this biologically important question, in this paper we make an attempt to investigate the structural and energetic characteristics of the long G·G* and G*·G Watson-Crick base mispairs, that are one and the same (Scheme 1), and physico-chemical mechanism of their tautomerisation *via* the double proton transfer (DPT). Such statement of the problem is urgent not only from the standpoint of the theory of the spontaneous point mutagenesis, but it is topical for the establishment of the most probable spatial structure of the G·G base mispair with Watson-Crick geometry observed in RNA by X-ray analysis [35-40], as well as in biomolecular nanoelectronics, in particular at the development of the devices for the detection and identification of the DNA point defects caused by the incorrect pairing of the nucleobases [41-43].

We used a combined QTAIM analysis and quantum-mechanical (QM) approach in order to contribute to the on-going systematic investigation of the G·G*↔G*·G DPT tautomerisation, that is of the interconversion of the G·G* base mispair into itself. It was established that this DPT reaction takes place in the ground electronic state through a single transition state following the concerted mechanism, in which the proton-transfer process occurs asynchronously without forming any stable intermediates. We have provided here the theoretical survey of the thermodynamic and dynamic stability of the G·G*/G*·G base mispair. The scans of the basic physico-

chemical properties [14-17,44] of the G*·G base mispair along the intrinsic reaction coordinate (IRC) of the G·G*↔G*·G DPT tautomerisation, that are necessary to understand the nature of this intriguing reaction, have been presented in this paper.

Computational Methods. All calculations were performed using the Gaussian'09 software package [45]. Geometries and harmonic vibrational frequencies of the G·G*/G*·G DNA base mispairs and the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ of their mutagenic tautomerisation *via* the DPT were obtained using Density Functional Theory (DFT) [46] with the B3LYP hybrid functional [47], which includes Becke's three-parameter exchange functional (B3) [48] combined with Lee, Yang and Parr's (LYP) correlation functional [49] in connection with Pople's 6-311++G(d,p) basis set in vacuum. A scaling factor of 0.9668 [50-53] (see papers [54-56] about the existing approaches to the definition of this parameter) has been used in the present work at the B3LYP level of QM theory to correct the harmonic frequencies of all studied structures. The DFT method has been recommended in the literature for the investigation of the double proton transfer process [5,6,13] and vibrations of the constituents of the nucleic acids [57-59] in the H-bonded base pairs, since it represents itself a good balance between computational cost and accuracy and therefore can be suggested as the shortest way to MP2 results [60-63]. Moreover, an excellent agreement between computational and experimental NMR, UV and IR spectroscopic data [64,65] evidences that the method (B3LYP/6-311++G(d,p)) employed for the geometry optimisation and vibrational frequency calculations are reliable. We performed single point energy calculations at the correlated MP2 level of theory [66] with the 6-311++G(2df,pd)/6-311++G(3df,2pd) Pople's [67-69] and cc-pVTZ/cc-pVQZ Dunning's cc-type [70,71] basis sets for the B3LYP/6-311++G(d,p) geometries, successfully applied on similar systems to obtain barrier heights and kinetic characteristics [14-17], to take into account the electronic correlation effects.

The correspondence of the stationary points to the G·G*/G*·G DNA base mispairs or to the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$, located by means of **S**ynchronous **T**ransit-guided **Q**uasi-**N**ewton (STQN) method [72,73], on the potential energy landscape has been checked by the absence or the presence, respectively, of one and only one imaginary frequency corresponding to the normal mode that identifies the reaction coordinate. The reaction pathway was established by following the IRC in the forward and reverse directions from the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ using the Hessian-based predictor-corrector (HPC) integration algorithm [74-76] with tight convergence criteria, that ultimately guarantees that the proper reaction pathway of the G·G*↔G*·G DPT tautomerisation has been found.

The electronic interaction energies E_{int} have been computed at the MP2/6-311++G(2df,pd) level of QM theory for the geometries optimised at the DFT B3LYP/6-311++G(d,p) level of theory as the difference between the total energy of the base pair and the energies of the isolated monomers. In each case the interaction

energy was corrected for the basis set superposition error (BSSE) [77,78] through the counterpoise procedure [79,80].

The Gibbs free energy G values for all structures were obtained at room temperature (T=298.15 K) in the following way:

$$G=E_{el}+E_{corr}, \quad (1)$$

where E_{el} – the electronic energy obtained at the MP2/6-311++G(2df,pd)//B3LYP/6-311++G(d,p) level of QM theory, E_{corr} – the thermal correction obtained at the B3LYP/6-311++G(d,p) level of QM theory.

The lifetime τ of the G·G*/G*·G DNA mispairs was calculated using the formula $1/k_{f,r}$.

The time $\tau_{99.9\%}$ necessary to reach 99.9% of the equilibrium concentration of the G·G* reactant and the G*·G product of reaction in the system of reversible first-order forward (k_f) and reverse (k_r) reactions was estimated by the formula [81]:

$$\tau_{99.9\%} = \frac{ln10^3}{k_f + k_r}. \quad (2)$$

To estimate the values of the forward k_f and reverse k_r rate constants for the G·G*↔G*·G tautomerisation reaction:

$$k_{f,r} = \Gamma \cdot \frac{k_B T}{h} e^{-\frac{\Delta\Delta G_{f,r}}{RT}} \quad (3)$$

we applied the standard TS theory [81], in which quantum tunneling effects are accounted by the Wigner's tunneling correction [82], that is adequate for the DPT reactions [17,32,52,61]:

$$\Gamma = 1 + \frac{1}{24}\left(\frac{h\nu_i}{k_B T}\right)^2, \quad (4)$$

where k_B – Boltzmann's constant, T – absolute temperature, h – Planck's constant, $\Delta\Delta G_{f,r}$ – the Gibbs free energy of activation for the forward and reverse DPT reactions (T = 298.15 K), R – universal gas constant, ν_i – the magnitude of the imaginary frequency associated with the vibrational mode at the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$.

Bader's quantum theory "Atoms in molecules" (QTAIM) was applied to analyse the electron density distribution [83], using program package AIMAll [84] with all default options. Wave functions were obtained at the level of theory used for geometry optimisation. The presence of a bond critical point (BCP) [83], namely the so-called (3,-1) BCP, and a bond path between the H-bond donor and acceptor, as well as the positive value of the Laplacian at this BCP ($\Delta\rho \geq 0$), were considered as criteria for the H-bond formation [85-87].

The energies of the conventional intermolecular H-bonds in the G·G*/G*·G DNA base mispairs were evaluated by the empirical Iogansen's formula [88]:

$$E_{HB} = 0.33 \cdot \sqrt{\Delta\nu - 40}, \quad (5)$$

where Δν - the magnitude of the redshift (relative to the free molecule) of the stretching mode of the H-bonded groups involved in the H-bonding. The partial deuteration of the OH, NH and NH_2 groups was applied to eliminate the effect of vibrational resonances [52,86].

The energies of all intermolecular H-bonds under the investigation of the scans [17,32,44] of their energies were evaluated by the empirical Espinosa-Molins-Lecomte (EML) formula [89,90], which was first successfully applied for the estimation of the individual energetic contributions of the separate H-bonds in the Watson-Crick DNA base pairs [91]:

$$E_{HB}=0.5\cdot V(r), \quad (6)$$

where V(r) - the value of a local potential energy density at the (3,-1) BCPs.

The energies of the O6H···O6 and N2H···N2 H-bonds in the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ were estimated by the Nikolaienko-Bulavin-Hovorun formulas [92]:

$$E_{O6H\cdots O6} = -3.09 + 239\cdot\rho, \quad (7)$$

$$E_{N2H\cdots N2} = -2.03 + 225\cdot\rho, \quad (8)$$

where ρ is the electron density at the (3,-1) BCP of the H-bond.

Period of the intermolecular vibrations T was calculated as:

$$T = \frac{1}{\nu\cdot c}, \quad (9)$$

where ν – the frequency of vibrations, c – the speed of the light in vacuum.

The frequency *f* of the G·G*↔G*·G tautomerisation was estimated by the formula:

$$f = \frac{1}{\tau}, \quad (10)$$

where τ is the lifetime of the G·G*/G*·G DNA base mispair.

The atomic numbering scheme for the nucleobase is conventional [93].

Obtained Results and Their Discussion. This work represents itself a logical extension of our previous achievements [94,95]. Tables 1-4, Scheme 1, Figs. 1-10 and Fig. S1, ESI† present the data obtained in this paper.

1. Energetic and geometric peculiarities of the G·G*/G*·G and $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ complexes and intermolecular specific interactions stabilizing them

The G·G*/G*·G (|C6N1(G*)N1C6(G)|=36.4º; C_1 symmetry) DNA base mispairs posess *cis*-oriented N9H glycosidic bonds of the G and G* bases, that are flexible molecules [96,97], in the *anti*-orientation with

respect to the sugar moiety. Quantum-mechanical calculations at the B3LYP/6-311++G(d,p) level of theory have revealed that these mismatches are stabilized by the O6H···O6 (8.39), N1H···N1 (6.14) and (G)N2H···N2(G*) (2.73 kcal·mol^{-1}) H-bonds, whereas the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ (|C6N1N1C6|=37.1º; C_1 symmetry; ΔG_{TS}=5.51 and ΔE_{TS}=8.33 kcal·mol^{-1} obtained at the MP2/cc-pVQZ//B3LYP/6-311++G(d,p) level of QM theory in vacuum; ν_i=1110.0*i* cm^{-1}) is stabilized by the N1-H-N1 covalent bridge and the O6H···O6 (29.72) and N2H···N2 H-bonds (2.92 kcal·mol^{-1}) (Tables 1 and 2, Fig. 1).

Geometric (the $d_{H\cdots N1/N2/O6}$ distances of the intermolecular H-bonds are less than the sum of corresponding Bondi's [98] van der Waals radii; the AH H-bond donating groups elongate upon the formation of the conventional AH···B H-bonds; the angles of the H-bonding are obtuse), electron-topological (the values of the electron density ρ at the (3,-1) BCP of the H-bond are in the acceptable range and the values of the Laplacian of the electron density Δρ at the (3,-1) BCP of the H-bond are positive) and spectroscopic (the frequency of the stretching vibrational mode *ν*(AH) of the AH donor group is shifted toward the lower frequencies or toward the red end of the spectrum under the formation of the intermolecular AH···B H-bonds) characteristics, used to deepen the nature of the considered interactions [85-87,99], testify the existence of the canonical H-bonds in the investigated structures.

2. Predicted theoretical pathway of the G·G*↔G*·G DPT tautomerisation

The calculated electron-topological and geometric profiles of the G·G*↔G*·G DPT tautomerisation along the IRC exhibit a spans of the χ-like crossings of the $d_{O6H/HO6}$, $d_{N1H/HN1}$, $\rho_{O6H/HO6}$, $\rho_{N1H/HN1}$, $\Delta\rho_{O6H/HO6}$ and $\Delta\rho_{N1H/HN1}$ graphs of the intermolecular H-bonds (Figs. 4a, 4b and 5b), that enables us to suggest that the G·G*↔G*·G DPT tautomerisation is likely to follow the *asynchronous concerted mechanism*. Moreover, we have allocated the 9 key points (KPs), that are critical for the detailed understanding of this mechanism (Fig. 1).

The obtained IRC allowed us to follow in detail the atomistic mechanism of DPT tautomerisation. As depicted in Figure 1, the process of the tautomerisation starts from the transfer of the proton localized at the N1 nitrogen atom of the G base in the G·G* DNA base mispair (the KP 1; IRC=-25.13 Bohr) along the N1H···N1 H-bond towards the N1 nitrogen atom of the G* enol tautomer, reaching the KP 2 ($\Delta\rho_{H\cdots N1}$=0; IRC=-0.29 Bohr), at which the N1-H covalent bond is significantly weakened and the H···N1 H-bond becomes the H-N1 covalent bond. Notably, exactly at this point the chemical identity of the G and G* bases is lost and actually a chemical reaction starts. Further, bypassing the KP 3 ($\rho_{N1\text{-}H}=\rho_{H\text{-}N1}$; IRC=-0.08 Bohr), characterized by the equivalent N1-H/H-N1 covalent bonds involved in the N1-H-N1 covalent bridge, the complex reaches the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ (the KP 4; IRC=0.00 Bohr). Then the base mispair transforms into the KP 5 ($\Delta\rho_{N1\cdots H}$=0; IRC=0.17 Bohr), at which the N1-H covalent bond becomes the N1···H H-bond, thus arriving at the $G^-\cdot G^+$ zwitterionic non-stable intermediate. Thereafter, the second mobile proton, localized at the $O6^+$ oxygen atom of the G^+

protonated base begins to move towards the $O6^-$ oxygen atom of the G^- deprotonated base after the first proton is already transferred. At this proton migration the rupture of the old H-O6 covalent bond, corresponding to the KP 6 ($\Delta\rho_{O6\cdots H}=0$; IRC=0.42 Bohr), crossing through the KP 7 ($\rho_{O6\text{-}H}=\rho_{H\text{-}O6}$; IRC=0.62 Bohr), characterized by the equivalent O6-H/H-O6 covalent bonds involved in the O6-H-O6 covalent bridge, and the formation of the new O6-H covalent bond, corresponding to the KP 8 ($\Delta\rho_{H\cdots O6}=0$; IRC=0.79 Bohr), are observed. Finally, the G*·G base mispair (the KP 9; IRC=7.35 Bohr) is reached.

The detailed electron-topological and geometric characteristics of the intermolecular bonds revealed in the 9 key points and the polarity of the latters are given in Table 2. It is worthwhile to point out that the H-bonds detected in the 9 key points meet all the criteria of the canonical H-bonding (see ref. [86] and bibliography therein).

It should be highlighted that the obtained 9 KPs let us divide the reaction pathway for the G·G*↔G*·G DPT tautomerisation into three distinct regions: the reactant (from -25.13 to -0.29 Bohr), transition state (from -0.29 to 0.79 Bohr) and product regions (from 0.79 to 7.35 Bohr), separated by the KP 2, coinciding with the reaction force [100] minimum, and the KP 8, coinciding with the reaction force maximum, in which the extrema of the first derivative of the electronic energy with respect to the IRC are reached (Figs. 1 and 2b).

Our results indicate that the energy equal to 7.98 kcal·mol^{-1}, representing 84.8 % of the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ electronic energy relatively to the G·G*/G*·G DNA base mispair, is spent at the *reactant region* on the reciprocal structural adjustment of the G and G* bases within the G·G* base mispair, in particular on the conformational changes of the NH_2 amino groups, to bring the donor and acceptor atoms as close as possible to each other in order to acquire such mutual deformation and orientation, that eventually lead to the DPT reaction at the TS region, whereas the energy equal to 8.62 kcal·mol^{-1}, representing 91.5 % of the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ electronic energy relatively to the G·G*/G*·G DNA base mispair, is expended at the *product region* on the structural relaxation of the KP 8 to the terminal G*·G base mispair.

3. Scans of the basic physico-chemical properties of the G·G*/G*·G DNA base mispairs along the IRC of their mutual tautomerisation *via* the DPT

Our results show that the molecular properties, namely the electronic energy, the first derivative of the electronic energy with respect to the IRC, the dipole moment of the base pair, the characteristic distances and the angle of the intermolecular H-bonds, the electron density, the Laplacian of the electron density, ellipticity and the energy at the (3,-1) BCPs of the intrapair H-bonds, the NBO charges of the hydrogen atoms involved in the tautomerisation, the glycosidic angles, the distance between the glycosidic hydrogens, angle between the plane and exocyclic bond and dihedral angles, are strongly dependent on the minimum-energy path (Figs. 2-10).

The calculated electronic energy profile of the DPT tautomerisation exhibits an asymmetric structure [61] with respect to the transition state in contrast to the results obtained for the interconversion of the equivalent base pairs through the tautomerisation *via* the DPT (Fig. 2a) [15-17,32]. We are inclined to link this unaccustomed asymmetry of the electronic energy profile with the significant correlated conformational changes of both amino groups belonging to the G and G* bases nearby the IRC = -10.07 Bohr. This very interesting question is discussed in the next paragraph of this paper.

The profile of the first derivative of the electronic energy with respect to the IRC (dE/dIRC) reaches its maximum (7.49 kcal·(mol·Bohr)$^{-1}$) at the KP2 (IRC=-0.29 Bohr) and minimum (-7.86 kcal·(mol·Bohr)$^{-1}$) – at the KP 8 (IRC=0.79 Bohr) (Fig. 2b).

We have fixed quite abrupt changes of the absolute value of the dipole moment μ of the studied base pairs within the narrow range of values 8.14÷8.64 D along the IRC of the G·G*↔G*·G tautomerisation with pronounced sharp peak at the IRC=-7.42 Bohr (Fig. 3), that can be explained by the fact that the dipole moment is the result of the vector sum of two terms: an atomic dipolar polarization term and a charge transfer (or bond dipole) term [101]. Moreover, it was previously shown by the research group of Prof. Chérif F. Matta that the property of the dipole moment to rise nearby the TS can be exploited in the laser control of reactions [102,103].

It should be highlighted that the curves of the electron density ρ, the Laplacian of the electron density Δρ and the ellipticity ε for the N1-H bonds intersect with each other exactly at the KP 3 (IRC=-0.08 Bohr), whereas for the O6-H bonds – at the KP 7 (IRC=0.62 Bohr), indicating that their values are equalized at the points of intersection (Figs. 4a and 4b). The analogous crossings are observed on the graphs of the distance $d_{AH/HB}$ between the hydrogen and electronegative A or B atoms (Fig. 5b). The values of the ρ (0.032÷0.320 a.u.), Δρ (-2.225÷0.165 a.u.), ε (0.013÷0.067) and $d_{AH/HB}$ (0.998÷1.935 Å) parameters, lying within a wide range of values, are in a good agreement with the previous results [5,6,13,14] (Figs. 4a, 4b, 4c and 5b).

Our studies of the dependencies of the H-bond energies, calculated by the EML formula [89,90] at the (3,-1) BCPs of the H-bonds, on the IRC presented in Figure 4d, indicate that the O6H···O6 and N1H···N1 H-bonds are weakly anti-cooperative [104] both in the G·G* ($dE_{N1H\cdots N1}/dE_{O6H\cdots O6}$=-0.78 at the IRC=-25.13 Bohr) and G*·G ($dE_{N1H\cdots N1}/dE_{O6H\cdots O6}$=-2.13 at the IRC=7.35 Bohr) DNA base mispairs.

It is worthwhile to point out that the upper O6H···O6 H-bond from the side of the major groove (13.61 ÷ 51.96 kcal·mol^{-1}) and the middle N1H···N1 H-bond (7.08 ÷ 37.65 kcal·mol^{-1}) exist within the range from the KP 1 to KP 6 (from -25.13 to 0.42 Bohr) and from the KP 1 to KP 2 (from -25.13 to -0.29 Bohr), achieving at this their maximum values at the KPs 6 and 2, respectively, whereas the upper O6H···O6 H-bond (14.76 ÷ 54.32 kcal·mol^{-1}) from the side of the major groove and the middle N1H···N1 H-bond (8.19 ÷ 38.64 kcal·mol^{-1}) exist within the range from the KP 8 to KP 9 (from 0.79 to 7.35 Bohr) and from the KP 4 to KP 9 (from 0.00 to 7.35 Bohr), attaining their maximum values at the KPs 8 and 4, respectively (Fig. 4d).

We have also observed the changes in the distance between the O6 oxygen atoms (2.434÷2.672 Å) and the N1 nitrogen atoms (2.595÷2.960 Å) involved in the O6H···O6 and N1H···N1 H-bonds, respectively, and also the $R(H_{N9}\text{-}H_{N9})$ distance between the glycosidic hydrogens (11.913÷12.361 Å), that results in the compression or the so-called "respiration" of the G·G* DNA base mispair during the reaction path, especially at the TS region (Figs. 5a and 9a). Also the noticeable oscillations of the ∠O6H···O6 (171.6÷174.8º) and ∠N1H···N1 (168.3÷176.1º) H-bond angles and the α_1 (∠N9H(G)H(G*))) (41.4÷44.6º) and α_2 (∠N9H(G*)H(G)) (42.0÷45.0º) glycosidic angles have been established (Figs. 5c and 9b).

The tautomerisation process of the base mispair is accompanied by the significant correlated conformational changes of both amino groups belonging to the G and G* bases, that is reflected in the changes of the β_1 (∠HN2C2N3(G)) (-9.1÷13.6º) and β_2 (∠HN2C2N3(G*)) (-11.5÷15.9º) dihedral angles along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT (Fig. 10b). Herewith, particularly dramatic changes are observed for the angle of inclination of the HN2H plane of the amino group relatively to the C2N2 exocyclic bond τ (Fig. 10a). In particular, the amino group of the G base becomes planar (τ_1=0º) at the IRC=-22.86 Bohr and then it begins to dodge in the direction opposite to the C2N2 bond (Fig. 10a). The same effect (τ_2=0º) is observed for the amino group of the G* base at the IRC=-7.73 Bohr. Notably, exactly at this point the dipole moment of the base mispair, that tautomerises, reaches its maximum (8.59 D) (Fig. 3).

The scans of the NBO charges of the hydrogen atoms H_I and H_{II} involved in the $N1H_I\cdots N1$ and $O6H_{II}\cdots O6$ H-bonds do not intersect with each other along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT (Fig. 8).

4. Properties of the third specific contact assisting the G·G*↔G*·G DPT tautomerisation

This section addresses the behavior of the third N2H···N2 H-bond, which physico-chemical properties we have monitored along the IRC of the tautomerisation (Figs. 6 and 7).

A characteristic structural feature of the long G·G* Watson-Crick base pair, that distinguishes it among others long incorrect pairs of nucleotide bases [17,22,32,94], is a steric conflict between the two exocyclic C amino groups of the G and G* nucleobases. This conflict is overcome in a very interesting way - through the propeller-like geometry of the base pair (Fig. 1), as well as due to the conformational adjustment of both amino groups. As a result, the pyramidal, substantially non-equivalent amino groups are involved in the canonical intermolecular (G)N2H···N2(G*) H-bond (see Tables 1 and 2). During the G·G*↔G*·G DPT tautomerisation basic physical and chemical parameters of the (G)N2H···N2(G*) H-bond, that exists in the IRC range from -25.13 to -10.37 Bohr, change noticeably - ρ (0.0081÷0.0165 a.u.); Δρ (0.0286÷0.0489 a.u.); ε (0.074÷2.4992); $E_{(G)N2H\cdots N2(G^*)}$ (1.45÷2.79 kcal·mol^{-1}); $d_{N2\cdots N2}$ (3.198÷3.278 Å); $d_{H\cdots N2}$ (2.237÷2.671 Å) and ∠N2H···N2 (118.8÷162.9º) (Figs. 6 and 7) - and the sequential switching (G)N2H···N2(G*) H-bond → N2···N2 van der

Waals contact → (G)N2···HN2(G*) H-bond → (G*)N2···HN2(G) H-bond occurs in an original way. At the IRC=-10.37 Bohr the (G)N2H···N2(G*) H-bond is transformed without discontinuities and bifurcations into the attractive N2···N2 van der Waals contact with slightly varying properties (ρ (0.0080÷0.0081 a.u.); Δρ (0.0281÷0.0286 a.u.); ε (2.379÷3.488); $E_{N2\cdots N2}$ (1.44÷1.45 kcal·mol^{-1}) and $d_{N2\cdots N2}$ (3.275÷3.279 Å)), which exists till the IRC=-9.23 Bohr and then smoothly, i.e. without discontinuities and bifurcations is converted into the (G)N2···HN2(G*) H-bond with significantly disturbed physico-chemical characteristics (ρ (0.008÷0.022 a.u.); Δρ (0.028÷0.067 a.u.); ε (0.057÷2.229); $E_{(G)N2\cdots HN2(G^*)}$ (1.44÷4.28 kcal·mol^{-1}); $d_{N2\cdots N2}$ (3.075÷3.307 Å); $d_{H\cdots N2}$ (2.097÷2.677 Å) and ∠(G)N2···HN2(G*) (118.0÷160.5º)), which further at the IRC=0.79 Bohr is turned due to the DPT along the O6H···O6 and N1···HN1 H-bonds into the (G*)N2···HN2(G) H-bond with greatly disturbed physico-chemical properties (ρ (0.015÷0.022 a.u.); Δρ (0.044÷0.067 a.u.); ε (0.059÷0.070); $E_{(G^*)N2\cdots HN2(G)}$ (2.46÷4.28 kcal·mol^{-1}); $d_{N2\cdots N2}$ (3.075÷3.266 Å); $d_{H\cdots N2}$ (2.097÷2.286 Å) and ∠(G*)N2···HN2(G) (160.5÷162.9º)) (Figs. 6 and 7, Tables 1 and 2). Notably, the distance d_{N2H} slightly varies within 1.2 % remaining practically constant during the tautomerisation reaction.

Intriguingly, the N2···N2 van der Waals contact is dynamically unstable in the limiting points of its existence, i.e. it is modulated by the low-frequency inter- and intramolecular vibrations [59], as is indicated by a significant increasing of the value of the ellipticity ε in this region (Fig. 6c, Tables 1 and 2). It is not excluded that the described above regularities are quite universal and can be specific to any exocyclic C-amino groups that sterically clash with each other in the H-bonded pairs of nucleotide bases.

Our studies indicate that the hydrogen atom H_{III}, which is included in the donor group of the (G)N2H_{III}···N2(G*) H-bond, posesses a larger NBO charge (0.394÷0.417 e) than the hydrogen atom H_{IV} of the N2H_{IV} (G*) imino group of the G* base (0.382÷0.395 e) till the IRC=-8.35 Bohr, where these NBO charges are equalized (0.395 e) (Fig. 8). Notably, in the vicinity of this structure the singularities on the profiles of the electronic energy, the first derivative of the electronic energy with respect to the IRC of the tautomerisation and the dipole moment are observed (Figs. 2 and 3). Thereafter, the NBO charge of the hydrogen atom H_{IV} involved in the (G*)N2H_{IV}···N2(G) H-bond (0.396÷0.421 e), that has acquired its donor properties at the IRC=-9.23 Bohr, starts to exceed the NBO charge of the hydrogen atom H_{III} of the amino group N2H_{III} (G) of the G base (0.380÷0.394 e), that is reflected in the arrangement of its graph above those for the hydrogen atom H_{III}. So, it is obvious that the hydrogen atoms of the neighboring exocyclic amino groups of the G and G* bases interchange their donor properties along the IRC.

5. Thermodynamic and dynamic stability of the G·G*/G*·G DNA base mispairs

In the course of this investigation, it was found that the G·G* or G*·G DNA base mispair is thermodynamically [94,95] and dynamically [105] stable structure (Table 3). The intermolecular H-bonds in the

G·G*/G*·G DNA base mispair ($[E_{O6H\cdots O6}+E_{N1H\cdots N1}+E_{N2H\cdots N2}]/|\Delta E_{int}|$=73.8 %) represent itself a significant portion into the electronic energy of the interaction between the G and G* bases (ΔE_{int}=-23.40 kcal·mol^{-1}). These data agree remarkably well with the results in our previous papers [5,6,13-15,17,22,26,32-34].

Our study indicates that the lifetime τ of the G·G*/G*·G DNA base mispairs (τ=8.22·10^{-10} s obtained at the MP2/cc-pVQZ//B3LYP/6-311++G(d,p) level of QM theory) was established to be noticeably greater than the periods *T* (1.49·10^{-12}, 8.53·10^{-13}, 5.52·10^{-13}, 3.67·10^{-13}, 3.53·10^{-13} and 3.06·10^{-13} s, respectively) of all 6 low-frequency [59] intermolecular vibrations (22.4, 39.1, 60.4, 90.8, 94.6 and 109.0 cm^{-1}), additionally suggesting that the G·G*/G*·G base pair is dynamically stable structure [105]. Logically, the G·G*↔G*·G Löwdin's [106,107] tautomerisation *via* the DPT can be considered as the periodic dipole-active process (Table 2, Figure 3) that occurs with the frequency 1220 MHz.

The time $\tau_{99.9\%}$ necessary to reach 99.9% of the equilibrium concentration of the starting G·G* and the final G*·G base pair is equal to 3.79·10^{-9} s (Table 3).

6. Comparative analysis of the long purine·purine Watson-Crick pairs of nucleotide bases

Finally, let's sum the brief results concerning investigated by us wrong purine-purine base pairs that belong to the so-called long Watson-Crick DNA base pairs (see Table 4).

First, these base pairs are intermediates of the acquisition of the geometry close to those of the canonical Watson-Crick base pairs in the base-pairing recognition pocket of the high-fidelity DNA polymerase. It is known that namely this process provides the catalytic competence of these base pairs in a closed conformation of the DNA polymerase. The A·A*↔A*·A$_{syn}$, G·G*↔G·G*$_{syn}$, G·A↔G·A$_{syn}$ and A*·G*↔A*·G*$_{syn}$ structural transitions are possible due to the availability of the so-called "molecular joint arthroplasty" – the exocyclic NH_2- or OH- groups – at the 6 position of one of the bases in the pair. We have already started the investigation of the nature of these conformational transitions and the first obtained results are very encouraging. Secondly, some of the investigated incorrect purine·purine base pairs, namely the A·A*, G·G* and H·H* base pairs, can be the source of the mutagenic tautomers at their dissociation in the recognition pocket of the DNA polymerase.

Conclusions. In this study the methodology of the scans of the basic physico-chemical parameters, namely energetic, electron-topological, geometric, polar and NBO properties, along the IRC combined with QM/QTAIM analysis has been used for the investigation of the G·G* DNA base mispair tautomerisation into the G*·G DNA base mispair *via* the migration of the protons involved in the intermolecular O6···HO6 (8.39) and N1H···N1 (6.14 kcal·mol^{-1}) H-bonds. On the basis of the obtained results the solid conclusions can be made that the G·G*↔G*·G DPT tautomerisation occurs through the *asynchronous concerted* mechanism.

We have paid special attention to the profiles of the physico-chemical properties of the G·G* DNA base mispair and intermolecular interactions in it, that provides new clues for the step-by-step tracking of their

changes along the G·G*↔G*·G reaction path. Furthermore, the 9 key points, that are critical for the atomistic understanding of the tautomerisation reaction, were comprehensively presented and analysed.

It was established that the tautomerisation *via* the DPT is assisted, except two others O6H···O6 and N1H···N1 H-bonds, that have been established to be anti-cooperative, by the third specific contact, sequentially switching along the IRC in an original way: (G)N2H···N2(G*) H-bond (-25.13 ÷ -10.37) → N2···N2 van der Waals contact (-10.37 ÷ -9.23) → (G)N2···HN2(G*) H-bond (-9.23 ÷ 0.79) → (G*)N2···HN2(G) H-bond (0.79 ÷ 7.35 Bohr).

We have also predicted that the G·G*/G*·G DNA base mispairs are thermodynamically and dynamically stable structures with lifetime of $8.22 \cdot 10^{-10}$ s and all 6 low-frequency intermolecular vibrations are able to develop during this lifetime. These observations are key causes to consider the G·G*↔G*·G tautomerisation *via* the Löwdin's mechanism as the periodic dipole-active process occuring with the frequency of 1220 MHz.

The possible practical application of the investigated tautomerisation deserves a more detailed discussion. Thus, we hope that our results on the incorrect purine·purine nucleobase pairs that can be integrated into the DNA double helix would be useful at the creation of the nano-biomolecular sensors for the detection of the point structural defects in DNA associated with the formation of the mismatches. On the other hand, these knowledges can guide the planning of the experiments for the studying of the new types of the nucleotide pairs of the metallized DNA [108-110] and microstructural interpretation of their results (Figure S1, ESI†).

Acknowledgments. This work was partially supported by the State Fund for Fundamental Research (SFFR) of Ukraine within the Ukrainian-Japanese project № F 52.2/001 for 2013-2014 years and by the Science and Technology Center in Ukraine (STCU) within the project № 5728 for 2012-2014 years. O.O.B. was supported by the Grant of the President of Ukraine to support scientific research of young scientists for 2014 year from the State Fund for Fundamental Research of Ukraine (project № GP/F56/074). This work was performed using computational facilities of joint computer cluster of SSI "Institute for Single Crystals" of the National Academy of Sciences of Ukraine and Institute for Scintillation Materials of the National Academy of Sciences of Ukraine incorporated into Ukrainian National Grid.

References.

1. Drake J.W., *Annu. Rev. Genet.*, 1991, **25**, 125–146.
2. von Borstel R.C., *Mut. Res.*, 1994, **307**, 131-140.
3. Friedberg E.C., Walker G.C., Siede W., Wood R.D., Schultz R.A., Ellenberger T. *DNA repair and mutagenesis*. Washington D.C.: ASM Press, 2006.
4. Brovarets' O.O., Kolomiets' I.M., Hovorun D.M. Elementary molecular mechanisms of the spontaneous point mutations in DNA: a novel quantum-chemical insight into the classical understanding. In: Tada T. (ed.), Quantum chemistry – molecules for innovations. Rijeka: In Tech Open Access, 2012, pp 59–102.
5. Brovarets' O.O., Hovorun D.M., *J. Biomol. Struct. Dynam.,* 2014, **32**, 127-154.
6. Brovarets' O.O., Hovorun D.M., *J. Biomol. Struct. Dynam.,* 2013. DOI: 10.1080/07391102.2013.822829.
7. Kool E. T., *Annu. Rev. Biochem.*, 2002, **71**, 191–219.
8. Topal M.D., Fresco J.R., *Nature*, 1976, **263**, 285-289.

9. Poltev V.I., Shulyupina N.V., Bruskov V.I., *Mol. Biol., **32***, 1998, 233-240.
10. Watson J.D., Crick F.H.C., *Nature*, 1953, **171**, 964–967.
11. Poltev V.I., Bruskov V.I., *Mol. Biol.,* 1977, **11,** 661–670.
12. Danilov V.I., Anisimov V.M., Kurita N., Hovorun D., *Chem. Phys. Lett.,* 2005, **412**, 285–293.
13. Brovarets' O. O., Hovorun D. M., *J. Biomol. Struct. Dynam.,* 2013. DOI: 10.1080/07391102.2013.852133.
14. Brovarets' O.O., Hovorun D.M., *J. Comput. Chem.*, 2013, **34**, 2577-2590.
15. Brovarets' O.O., Hovorun D.M., *Phys. Chem. Chem. Phys.*, 2013, **15**, 20091-20104.
16. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *Chem. Phys. Lett.*, 2014, **592**, 247-255.
17. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *Chem. Phys. Lett.*, 2013, **578**, 126-132.
18. Brovarets' O. O., Hovorun D. M., *Ukr. Biokhim. Zh.,* 2010, **82**, 55-60.
19. Brovarets' O. O. Physico-chemical nature of the spontaneous and induced by the mutagens transitions and transversions (PhD thesis). Taras Shevchenko National University of Kyiv: Kyiv, 2010.
20. Kochina O. S., Yurenko Ye. P., Hovorun D. M. *Biopolym. Cell*, 2007, **23**, 167-171.
21. Yang L., Beard W.A., Wilson S.H., Roux B., Broyde S., Schlick T., *J. Mol. Biol.*, 2002, **321**, 459–478.
22. Brovarets' O.O., Hovorun D.M., *Phys. Chem. Chem. Phys.*, 2014, **16**, 9074-9085.
23. Brown T., Hunter W. H., Kneale G., Kennard O., *Proc. Natl. Acad. Sci. USA,* 1986, **83**, 2402-2406.
24. Bebenek K., Pedersen L.C., Kunkel T.A., *Proc. Natl. Acad. Sci. USA*., 2011, **108**, 1862-1867.
25. Wang W., Hellinga H.W., Beese L.S., *Proc. Natl. Acad. Sci. USA*., 2011, **124**, 17644-17648.
26. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *Phys. Chem. Chem. Phys.*, 2014, **16**, 3715-3725.
27. Platonov M.O., Samijlenko S.P., Sudakov O.O., Kondratyuk I.V., Hovorun D.M., *Spectrochim. Acta A Mol. Biomol. Spectrosc.,* 2005, **62,** 112–114.
28. Kosenkov D., Kholod Y., Gorb L., Shishkin O., Hovorun D.M., Mons M., Leszczynski J., *J. Phys. Chem. B*, 2009, **113**, 6140-6150.
29. Brovarets' O.O., Hovorun D.M., *Biopolym. Cell*, 2010, **26**, 72–76.
30. Brovarets' O.O., Hovorun D.M., *Biopolym. Cell*, 2010, **26**, 295–298.
31. Furmanchuk A., Isayev O., Gorb L., Shishkin O.V., Hovorun D.M., Leszczynski J., *Phys. Chem. Chem. Phys.*, 2011, **13**, 4311-4317.
32. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *J. Mol. Model.,* 2013, **19**, 4223-4237.
33. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *J. Comput. Chem.*, 2014, **35**, 451-466.
34. Brovarets' O.O., Hovorun D.M., *Mol. Phys.*, 2014, in press.
35. Ippolito J.A., Steitz T., *J. Mol. Biol.*, 2000, **295,** 711–717.
36. Hung L.-W., Holbrook E.L., Holbrook S.R., *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 5107–5112.
37. Schuwirth B.S., Borovinskaya M.A., Hau C.W., Zhang W., Vila-Sanjurjo A., Holton J.M., Cate J.H., *Science*, 2005, **310,** 827–834.
38. Timsit Y., Bombard S., *RNA*, 2007, **13,** 2098–2107.
39. Lee J.C., Gutell R.R., *J. Mol. Biol.*, 2004, **344**, 1225–1249.
40. Johnson S.J., Beese L.S., *Cell,* **116**, 803-816.
41. Kobori A., Peng T., Hayashi G., Nakatani K., *Nucl. Acids Symp. Ser.*, 2004, 129-130.
42. Jauregui L.A., Seminario J.M., *IEEE Sensors J.,* 2008, **8**, 803-814.
43. Otero-Navas I., Seminario J.M., *J. Mol. Model.*, 2012, **18**, 91-101.
44. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *J. Mol. Model.*, 2013, **19**, 4119-4137.

45. Frisch M.J., Trucks G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Pople J.A. et. al. (2010) GAUSSIAN 09 (Revision B.01), Gaussian Inc. Wallingford, C.T.
46. Parr R.G., Yang W. *Density-functional theory of atoms and molecules.* Oxford University Press: Oxford. 1989.
47. Tirado-Rives J., Jorgensen W.L., *J. Chem. Theory. Comput.*, 2008, **4**, 297–306.
48. Becke A.D., *Phys. Rev. A*, 1988, **38**, 3098-3100.
49. Lee C., Yang W., Parr R.G., *Phys. Rev. B*, 1988, **37**, 785-789.
50. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *Biopolym. Cell*, 2010, **26**, 398–405.
51. Brovarets' O.O., Hovorun D.M., *Biopolym. Cell*, 2011, **27**, 221–230.
52. Brovarets' O.O., Hovorun D.M., *J. Biomol. Struct. Dynam.*, 2013, **31**, 913-936.
53. Brovarets' O. O., Hovorun D. M., *Ukr. Biokhim. Zh.,* 2010, **82**, 55-60.
54. Palafox M.A., Iza N., Gil M., *J. Mol. Struct. (Theochem)*, 2002, **585**, 69-92.
55. Palafox M.A., Rastogi V.K., *Spectrochim. Acta A Mol. Biomol. Spectrosc.,* 2002, **58A**, 411-440.
56. Palafox M.A., *Int. J. Quantum Chem.*, 2000, **77**, 661-684.
57. Pelmenschikov A., Hovorun D.M., Shishkin O.V., Leszczynski J., *J. Chem. Phys.,* 2000, **113**, 5986–5990.
58. Hovorun D.M., Gorb L., Leszczynski J., *Int. J. Quantum. Chem.*, 1999, **75**, 245–253.
59. Shishkin O.V., Pelmenschikov. A., Hovorun D.M., Leszczynski J., *Chem. Phys.*, 2000, **260,** 317–325.
60. Matta C.F., *J. Comput. Chem.*, 2010, **31**, 1297–1311.
61. Arabi A.A., Matta C.F., *Phys. Chem. Chem. Phys.*, 2011, **13**, 13738-13748.
62. Lozynski M., Rusinska-Roszak D., Mack H.-G., *J. Phys. Chem. A*, 1998, **102,** 2899–2903.
63. Palafox A.M., *J. Biomol. Struct. Dynam.*, 2014, **32**, 831-851.
64. Brovarets' O.O., Hovorun D.M., *Opt. Spectrosc.*, 2011, **111,** 750–757.
65. Samijlenko S.P., Yurenko Y.P., Stepanyugin A.V., Hovorun D.M., *J. Phys. Chem. B,* 2011, **114**, 1454-1461.
66. Frisch M.J., Head-Gordon M., Pople J.A., *Chem. Phys. Lett.*, 1990, **166,** 281-289.
67. Frisch M.J., Pople J.A., Binkley J.S., *J. Chem. Phys.,* 1984, **80**, 3265−3269.
68. Hariharan P.C., Pople J.A., *Theor. Chem. Accounts. Theor. Comput. Model.,* 1973, **28**, 213−222.
69. Krishnan R., Binkley J.S., Seeger R., Pople J.A., *J. Chem. Phys.,* 1980, **72**, 650−654.
70. Dunning Jr. T.H., *J. Chem. Phys.,* 1989, **90**, 1007−1023.
71. Kendall R.A., Dunning Jr.T.H., Harrison R.J., *J. Chem. Phys.,* 1992, **96**, 6796−6806.
72. Peng C., Schlegel H.B., *Isr. J. Chem.*, 1993, **33**, 449–454.
73. Peng C., Ayala P.Y., Schlegel H.B., Frisch M.J., *J. Comput. Chem.*, 1996, **17**, 49–56.
74. Hratchian H.P., Schlegel H.B., *J. Chem. Phys.,* 2004, **120**, 9918-9924.
75. Hratchian H.P., Schlegel H.B. Finding minima, transition states, and following reaction pathways on *ab initio* potential energy surfaces. In: Dykstra C.E., Frenking G., Kim K.S., Scuseria G. (eds.) Theory and applications of computational chemistry: The first 40 years, Elsevier, Amsterdam, 2005, pp. 195-249.
76. Hratchian H.P., Schlegel H.B., *J. Chem. Theory Comput.*, 2005, **1,** 61-69.
77. Boys S.F., Bernardi F., *Mol. Phys.*, 1970, **19**, 553-566.
78. Gutowski M., Van Lenthe J.H., Verbeek J., Van Duijneveldt F.B., Chalasinski G., *Chem. Phys. Lett.,* 1986, **124**, 370–375.
79. Sordo J.A., Chin S., Sordo T.L., *Theor. Chim. Acta*, 1988, **74**, 101-110.
80. Sordo J.A., *J. Mol. Struct. (THEOCHEM)*, 2001, **537**, 245-251.
81. Atkins P.W. *Physical chemistry,* Oxford University Press, Oxford, 1998.

82. Wigner E., *Zeit. Physik. Chem.* B, 1932, **19**, 203−216.
83. Bader R.F.W. *Atoms in molecules: A quantum theory,* Oxford University Press, Oxford, 1990.
84. Keith T.A. (2011) AIMAll (Version 11.12.19) Retrieved from http://aim.tkgristmill.com.
85. Ponomareva A.G., Yurenko Y.P., Zhurakivsky R.O., van Mourik T., Hovorun D.M., *Phys. Chem. Chem. Phys.*, 2012, **14,** 6787-6795.
86. Brovarets' O.O., Yurenko Y.P., Hovorun D.M., *J. Biomol. Struct. Dynam.,* 2014, **32**, 993-1022.
87. Ponomareva A.G., Yurenko Y.P., Zhurakivsky R.O., van Mourik T., Hovorun D.M., *J. Biomol. Struct. Dynam.,* 2014, **32**, 730-740.
88. Iogansen A.V., *Spectrochim. Acta A Mol. Biomol. Spectrosc.,* 1999, **55**, 1585–1612.
89. Espinosa E., Molins E., Lecomte C., *Chem. Phys. Lett.*, 1998, **285**, 170–173.
90. Mata I., Alkorta I., Espinosa E., Molins E., *Chem. Phys. Lett.,* 2011, **507**, 185–189.
91. Matta C.F., Castillo N., Boyd R.J., *J. Phys. Chem. B*, 2006, **110**, 563-578.
92. Nikolaienko T.Y., Bulavin L.A., Hovorun D.M., *Phys. Chem. Chem. Phys.,* 2012*,* **14**, 7441–7447.
93. Saenger W. *Principles of nucleic acid structure*, Springer, New York, 1984.
94. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *Mol. Phys.*, 2014. DOI: 10.1080/00268976.2013.877170.
95. Brovarets' O.O., Zhurakivsky R.O., Hovorun D.M., *J. Biomol. Struct. Dynam.*, 2014, DOI: 10.1080/07391102.2014.897259.
96. Govorun D.N., Danchuk V.D., Mishchuk Ya.R., Kondratyuk I.V., Radomsky N.F., Zheltovsky N.V., *J. Mol. Struct.*, 1992, **267**, 99–103.
97. Nikolaienko T.Y., Bulavin L.A., Hovorun D.M., *J. Biomol. Struct. Dynam.*, 2011, **29**, 563-575.
98. Bondi A.J., *J. Phys. Chem.,* 1964, **68**, 441-451.
99. Kaplan I. *Intermolecular interactions: physical picture, computational methods and model potentials* (Wiley Series in Theoretical Chemistry), John Wiley & Sons Ltd, Chichester, 2006.
100. Politzer P., Murray J.S., Jaque P., *J. Mol. Model.*, 2013, **19**, 4111-4118.
101. Matta C. F., Sowlati-Hashjin S., Bandrauk A. D., *J. Phys. Chem. A*, 2013, **117**, 7468-7483.
102. Bandrauk A.D., Sedik E.S., Matta C.F., *J. Chem. Phys.*, 2004, **121**, 7764-7775.
103. Bandrauk A.D., Sedik E.S., Matta C.F., *Mol. Phys.*, 2006, **104**, 95-102.
104. Mishchuk Ya.R., Potyagaylo A.L., Hovorun D.M., *J. Mol. Struct.*, 2000, **552**, 283-289.
105. Brovarets' O.O., Yurenko Y.P., Dubey I.Ya., Hovorun D.M., *J. Biomol. Struct. Dynam.*, 2012, **29**, 1101–1109.
106. Löwdin P.-O., *Rev. Mod. Phys.*, 1963, **35**, 724-732.
107. Löwdin P.-O. Quantum genetics and the aperiodic solid: some aspects on the biological problems of heredity, mutations, aging, and tumors in view of the quantum theory of the DNA molecule. In Löwdin, P.-O. (Ed.) *Advances in Quantum Chemistry* (1966, **2**, pp. 213-360). New York, USA, London, UK: Academic Press.
108. Ono A., Torigoe H., Tanaka Y., Okamoto I., *Chem. Soc. Rev.*, 2011, **40**, 5855–5866.
109. Ono A., Cao S., Togashi H., Tashiro M., Fujimoto T., Machinami T., Oda S., Miyake Y., Okamoto I., Tanaka Y., *Chem. Commun.*, 2008, 4825–4827.
110. Torigoe H., Ono A., Kozasa T., *Chem. Eur. J.*, 2010, **16**, 13218–13225.

Table 1 Electron-topological, structural, vibrational and energetic characteristics of the intermolecular H-bonds revealed in the G·G*/G*·G DNA base mispairs and the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ obtained at the B3LYP/6-311++G(d,p) level of QM theory in vacuum.

Complex	AH···B H-bond	ρ[a]	$\Delta\rho$[b]	$100\cdot\varepsilon$[c]	$d_{A\cdots B}$[d]	$d_{H\cdots B}$[e]	Δd_{AH}[f]	∠AH···B[g]	$\Delta\nu$[h]	E_{HB}[i]
G·G*/G*·G	O6H···O6	0.050	0.138	2.48	2.646	1.649	0.036	171.2	687.1	8.39
	N1H···N1	0.035	0.091	6.64	2.920	1.893	0.021	172.4	385.6	6.14
	N2H···N2	0.016	0.048	6.71	3.231	2.244	0.006	163.9	108.6	2.73
$TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$	O6H···O6	0.100	0.124	1.81	2.464	1.398	-	174.7	-	20.72*
	N2H···N2	0.022	0.066	5.97	3.079	2.104	-	160.1	-	2.92*

[a]The electron density at the (3,-1) BCP, a.u.
[b]The Laplacian of the electron density at the (3,-1) BCP, a.u.
[c]The ellipticity at the (3,-1) BCP
[d]The distance between the A (H-bond donor) and B (H-bond acceptor) electronegative atoms involved in the AH···B H-bond, Å
[e]The distance between the H and B atoms of the AH···B H-bond, Å
[f]The elongation of the H-bond donating group AH upon the AH···B H-bonding, Å
[g]The AH···B H-bond angle, degree
[h]The redshift of the stretching vibrational mode ν(AH), cm^{-1}
[i]Energy of the H-bonds, calculated by Iogansen's [88] or Nikolaienko-Bulavin-Hovorun (marked with an asterisk) [92] formulas, kcal mol^{-1}

Table 2 Electron-topological and structural characteristics of the intermolecular bonds revealed in the 9 key points and the polarity of the latters along the IRC of the G·G*↔G*·G DPT tautomerisation obtained at the B3LYP/6-311++G(d,p) level of QM theory in vacuum

Complex	AH···B H-bond	ρ	$\Delta\rho$	$100\cdot\varepsilon$	$d_{A\cdots B}$	$d_{H\cdots B}$	∠AH···B	μ
Key point 1 (-25.13 Bohr): G·G*	O6···HO6	0.050	0.138	2.48	2.646	1.649	171.2	8.45
	N1H···N1	0.035	0.091	6.64	2.920	1.893	172.4	
	N2H···N2	0.016	0.048	6.71	3.231	2.244	163.9	
Key point 2 (-0.29 Bohr): $\Delta\rho_{H\cdots N1}=0$	O6···HO6	0.092	0.149	1.77	2.469	1.426	174.8	8.29
	N1H···N1	0.115	0.000	4.31	2.597	1.409	174.9	
	N2···HN2	0.022	0.066	6.14	3.080	2.107	160.1	
Key point 3 (-0.08 Bohr): $\rho_{N1-H}=\rho_{H-N1}$	O6···HO6	0.097	0.132	1.80	2.465	1.406	174.8	8.28
	N1-H-N1	0.155	-0.221	3.93	2.596	1.294	175.6	
	N2···HN2	0.022	0.066	6.02	3.079	2.105	160.1	
Key point 4 (0.00 Bohr): $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$	O6···HO6	0.100	0.124	1.81	2.464	1.398	174.7	8.31
	N2···HN2	0.022	0.066	5.97	3.079	2.104	160.1	
Key point 5 (0.17 Bohr): $\Delta\rho_{N1\cdots H}=0$	O6···HO6	0.106	0.101	1.82	2.461	1.377	174.7	8.43
	N1···HN1	0.112	0.000	4.52	2.600	1.419	176.0	
	N2···HN2	0.022	0.066	5.86	3.078	2.102	160.0	
Key point 6 (0.42 Bohr): $\Delta\rho_{O6\cdots H}=0$	O6···HO6	0.126	0.011	1.74	2.450	1.314	174.5	8.46
	N1···HN1	0.089	0.071	5.04	2.617	1.511	176.0	
	N2···HN2	0.022	0.066	5.75	3.075	2.098	160.1	
Key point 7 (0.62 Bohr): $\rho_{O6-H}=\rho_{H-O6}$	O6-H-O6	0.161	-0.234	1.59	2.439	1.227	174.2	8.30
	N1···HN1	0.081	0.088	5.20	2.630	1.545	175.8	
	N2···HN2	0.022	0.066	5.78	3.075	2.097	160.4	
Key point 8 (0.79 Bohr): $\Delta\rho_{H\cdots O6}=0$	O6H···O6	0.133	0.000	1.76	2.435	1.290	173.9	8.23
	N1···HN1	0.077	0.098	5.27	2.636	1.564	175.7	
	N2···HN2	0.022	0.067	5.86	3.075	2.097	160.5	
Key point 9 (7.35 Bohr): G*·G	O6H···O6	0.050	0.138	2.48	2.646	1.649	171.2	8.45
	N1···HN1	0.035	0.091	6.64	2.920	1.893	172.4	
	N2···HN2	0.016	0.048	6.71	3.231	2.244	163.9	

Notes: For footnote definitions see Table 1. μ – the dipole moment of the complex, D.

Table 3 Energetic and kinetic characteristics of the G·G*↔G*·G DPT tautomerisation obtained at the different levels of QM theory for the geometry calculated at the B3LYP/6-311++G(d,p) level of QM theory in vacuum

Level of QM theory	$\Delta\Delta G_{TS}$[a]	$\Delta\Delta E_{TS}$[b]		τ[c]	$\tau_{99.9\%}$[d]
		kcal·mol^{-1}	cm^{-1}		
MP2/6-311++G(2df,pd)	4.97	7.79	2724.2	$3.29\cdot10^{-10}$	$1.52\cdot10^{-9}$
MP2/6-311++G(3df,2pd)	5.46	8.28	2896.9	$7.56\cdot10^{-10}$	$3.48\cdot10^{-9}$
MP2/cc-pVTZ	5.21	8.03	2807.7	$4.92\cdot10^{-10}$	$2.27\cdot10^{-9}$
MP2/cc-pVQZ	5.51	8.33	2914.2	$8.22\cdot10^{-10}$	$3.79\cdot10^{-9}$

[a]The Gibbs free energy of activation for the forward and reverse reactions of the G·G*↔G*·G DPT tautomerisation ($\Delta G_{G\cdot G^*/G^*\cdot G}$=0; T=298.15 K), kcal·mol^{-1}

[b]The activation electronic energy for the forward and reverse reactions of the G·G*↔G*·G DPT tautomerisation ($\Delta E_{G\cdot G^*/G^*\cdot G}$=0)

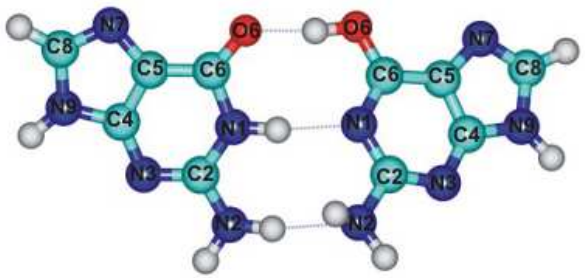
[c]The lifetime of the G·G*/G*·G DNA base mispair, s

[d]The time necessary to reach 99.9% of the equilibrium concentration between the reactant G·G* and the product G*·G of the G·G*↔G*·G tautomerisation reaction *via* the DPT, s

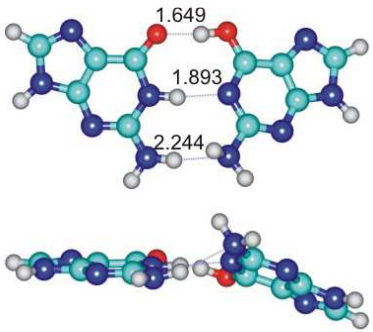
The frequency of the vibrational mode in the G·G*/G*·G DNA base mispair, which becomes imaginary in the $TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$ of the G·G*↔G*·G DPT tautomerisation, is equal to 3195.3 cm^{-1} and zero-point vibrational energy associated with this normal mode is equal to 4.57 kcal·mol^{-1} or 1597.6 cm^{-1} obtained at the B3LYP/6-311++G(d,p) level of QM theory.

Table 4 Selected physico-chemical characteristics of the incorrect purine·purine nucleobase pairs of biological importance possessing Watson-Crick geometry

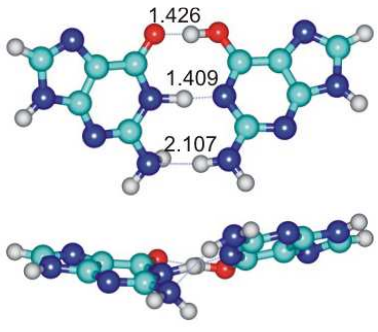
Base mispairs	References	Geometric parameters				Energetic parameters			
		Symmetry	R(H_{N9}-H_{N9}), Å	α_1, degree	α_2, degree	H-bonds	E_{HB}, kcal·mol^{-1}	-ΔE_{int}, kcal·mol^{-1}	-ΔG_{int}, kcal·mol^{-1}
G·G*	Present investigation	C_1	12.047	45.4	43.6	O6H···O6	8.39	23.40	9.65
						N1H···N1	6.14		
						N2H···N2	2.73		
A·A*	[32]	C_s	12.324	45.6	46.7	N6H···N6	7.01	17.89	4.32
						N1H···N1	6.88		
A·G	[33]	C_1	12.375	43.4	45.1	N6H···O6	5.68	17.54	3.57
						N1H···N1	6.51		
						N2H···HC2	0.68		
H·H*	[17]	C_s	12.345	42.2	42.6	O6H···O6	7.95	18.91	6.84
						N1H···N1	6.97		
A·H	[94]	C_s	12.335	44.6	45.7	N6H···O6	5.40	24.72	12.92
						N1H···N1	6.99		



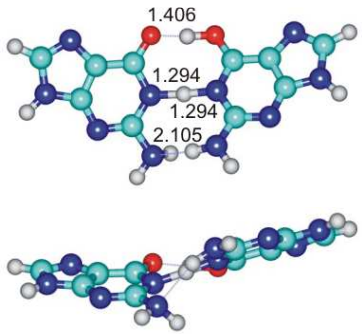
Scheme 1 Geometrical structure of the G·G* DNA base mispair (C_1). The numeration of atoms is generally accepted [88]. The intermolecular H-bonds are marked by dashed lines.



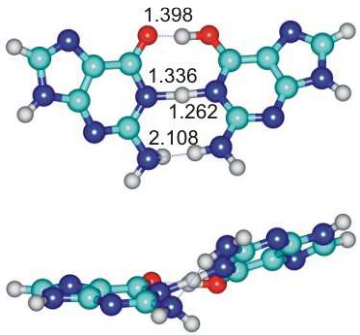
Key point 1
G·G*
IRC=-25.13 Bohr



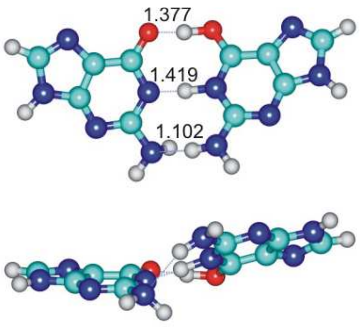
Key point 2
$\Delta\rho_{H\cdots N1}=0$
IRC=-0.29 Bohr



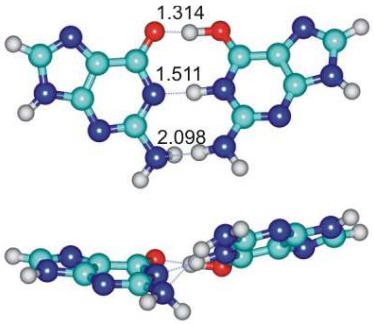
Key point 3
$\rho_{N1-H}=\rho_{H-N1}$
IRC=-0.08 Bohr



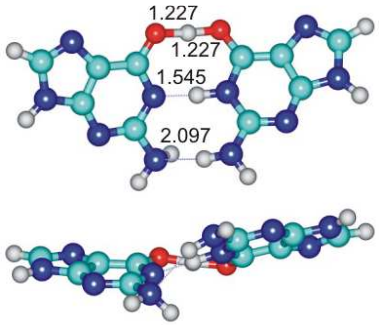
Key point 4
$TS_{G\cdot G^*\leftrightarrow G^*\cdot G}$
IRC=0.00 Bohr



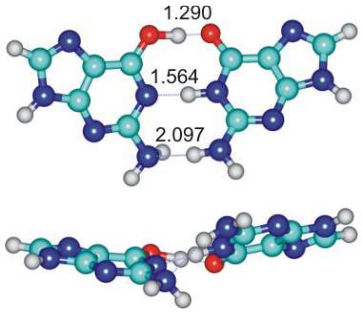
Key point 5
$\Delta\rho_{N1\cdots H}=0$
IRC=0.17 Bohr



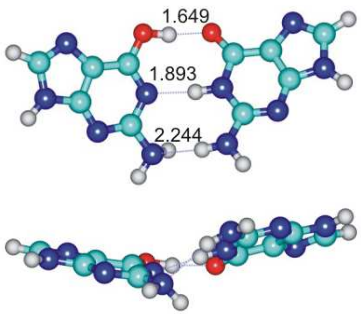
Key point 6
$\Delta\rho_{O6\cdots H}=0$
IRC=0.42 Bohr



Key point 7
$\rho_{O6-H}=\rho_{H-O6}$
IRC=0.62 Bohr

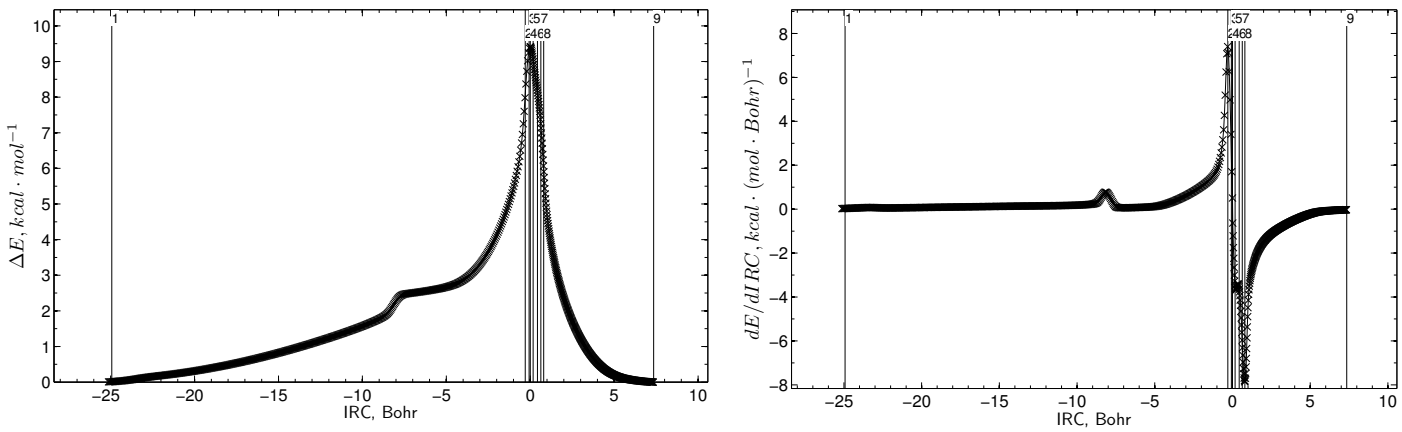


Key point 8
$\Delta\rho_{H\cdots O6}=0$
IRC=0.79 Bohr



Key point 9
G·G*
IRC=7.35 Bohr

Fig. 1 Geometric structures of the 9 key points describing the evolution of the G·G*↔G*·G tautomerisation *via* the DPT along the IRC obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*. Coordinates of the 9 key points are presented below each structure. The dotted lines indicate the AH···B H-bonds, while continuous lines show covalent bonds (their lengths are presented in angstroms). Carbon atoms are in light-blue, nitrogen in dark-blue, hydrogen in grey and oxygen in red.



a **b**

Fig. 2 Profiles of: (**a**) the electronic energy E and (**b**) the first derivative of the electronic energy with respect to the IRC (dE/dIRC) along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*

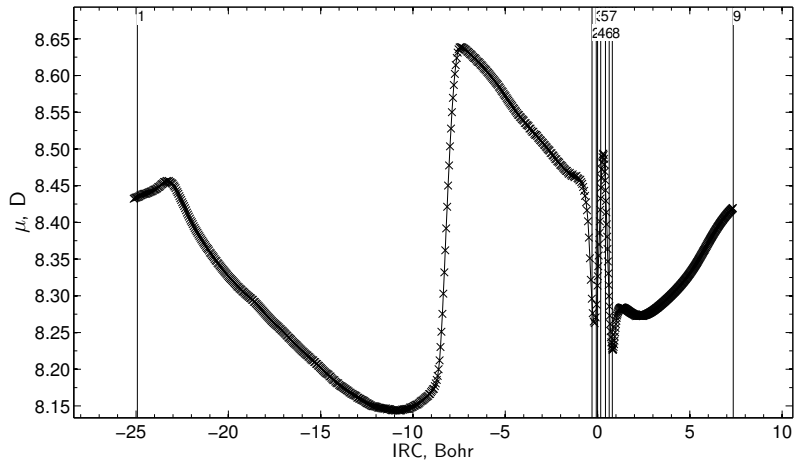


Fig. 3 Profile of the dipole moment μ along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*

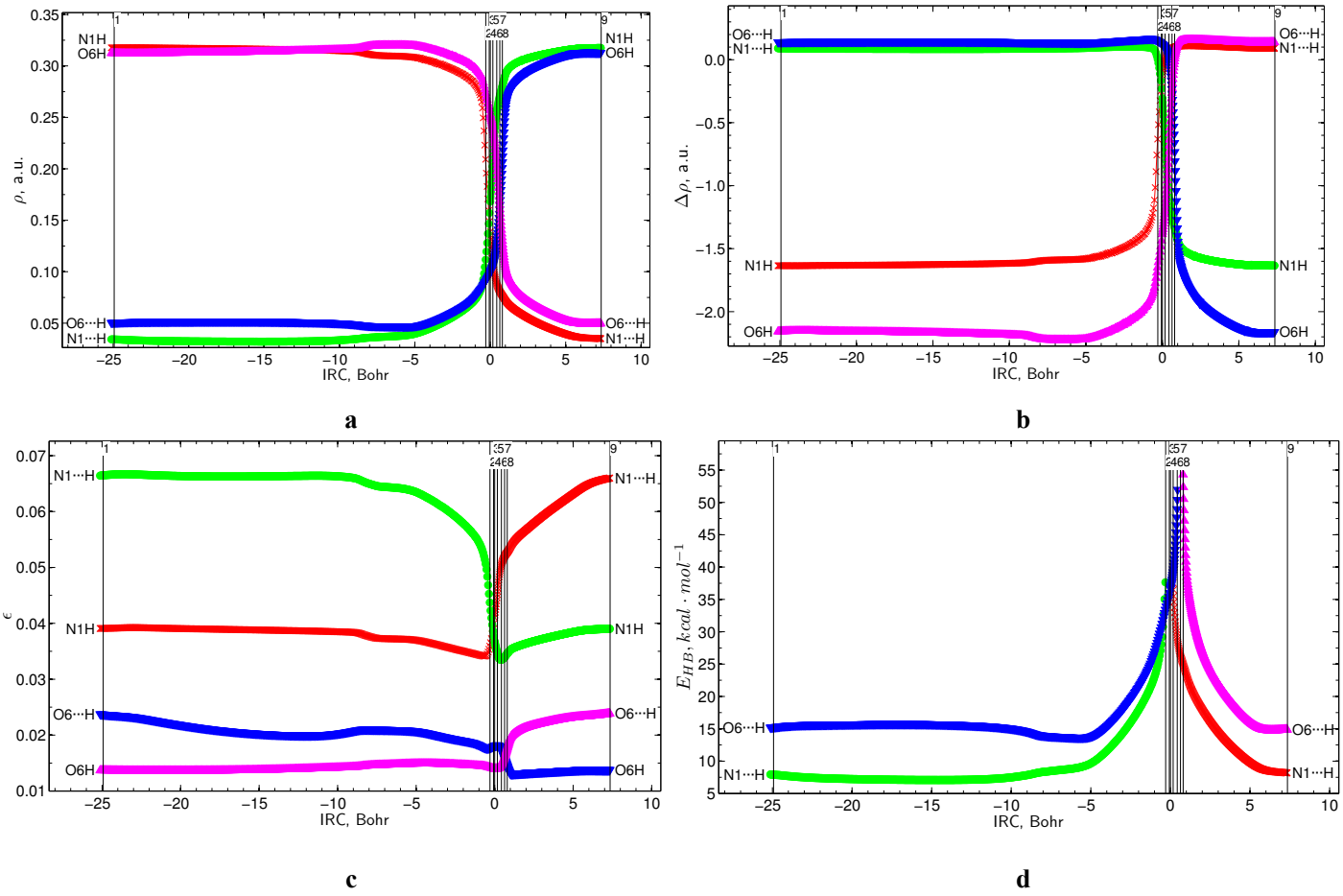


Fig. 4 Profiles of: (**a**) the electron density ρ; (**b**) the Laplacian of the electron density Δρ; (**c**) the ellipticity ε and (**d**) the energy of the H-bond E_{HB}, estimated by the EML formula [85,86], at the (3,-1) BCPs of the covalent and hydrogen bonds along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*

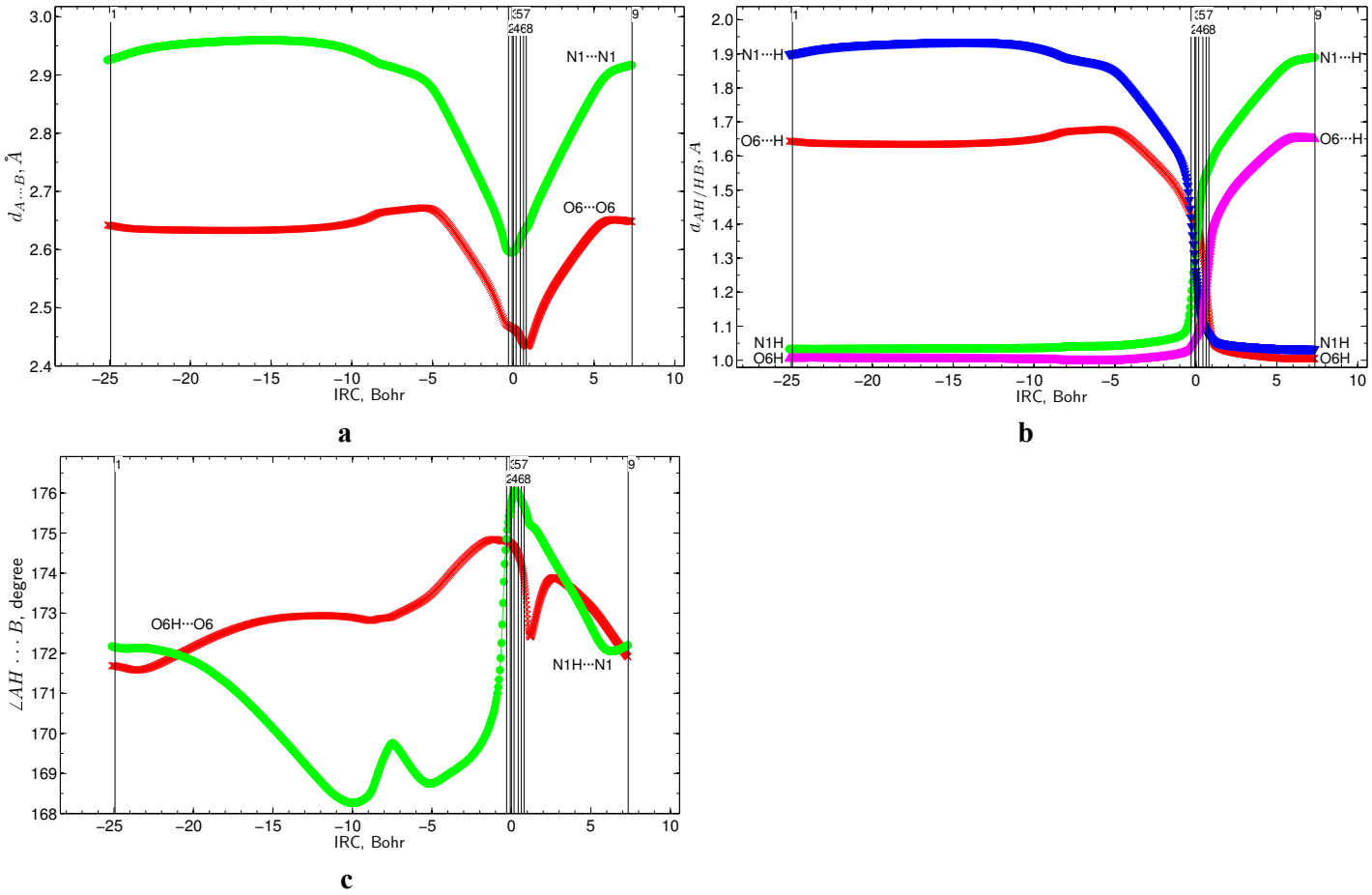


Fig. 5 Profiles of: (**a**) the distance $d_{A\cdots B}$ between the electronegative A and B atoms; (**b**) the distance $d_{AH/HB}$ between the hydrogen and electronegative A or B atoms and (**c**) the angle ∠AH···B of the AH···B H-bonds along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*

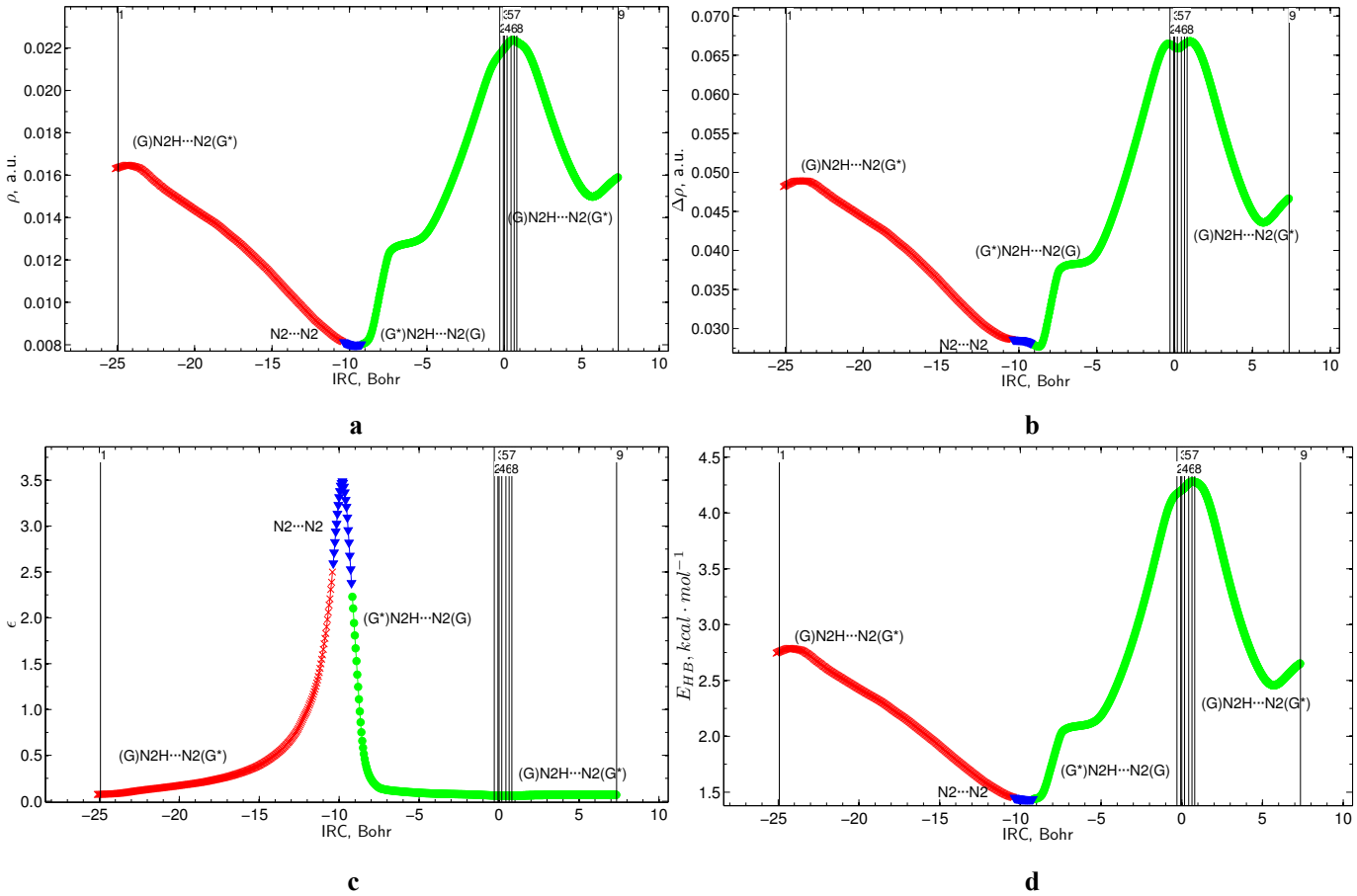


Fig. 6 Profiles of: (**a**) the electron density ρ; (**b**) the Laplacian of the electron density Δρ; (**c**) the ellipticity ε and (**d**) the energy, estimated by the EML formula [85,86], at the (3,-1) BCPs of the (G)N2H···N2(G*) / (G*)N2H···N2(G) H-bonds and the N2···N2 van der Waals contact along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*

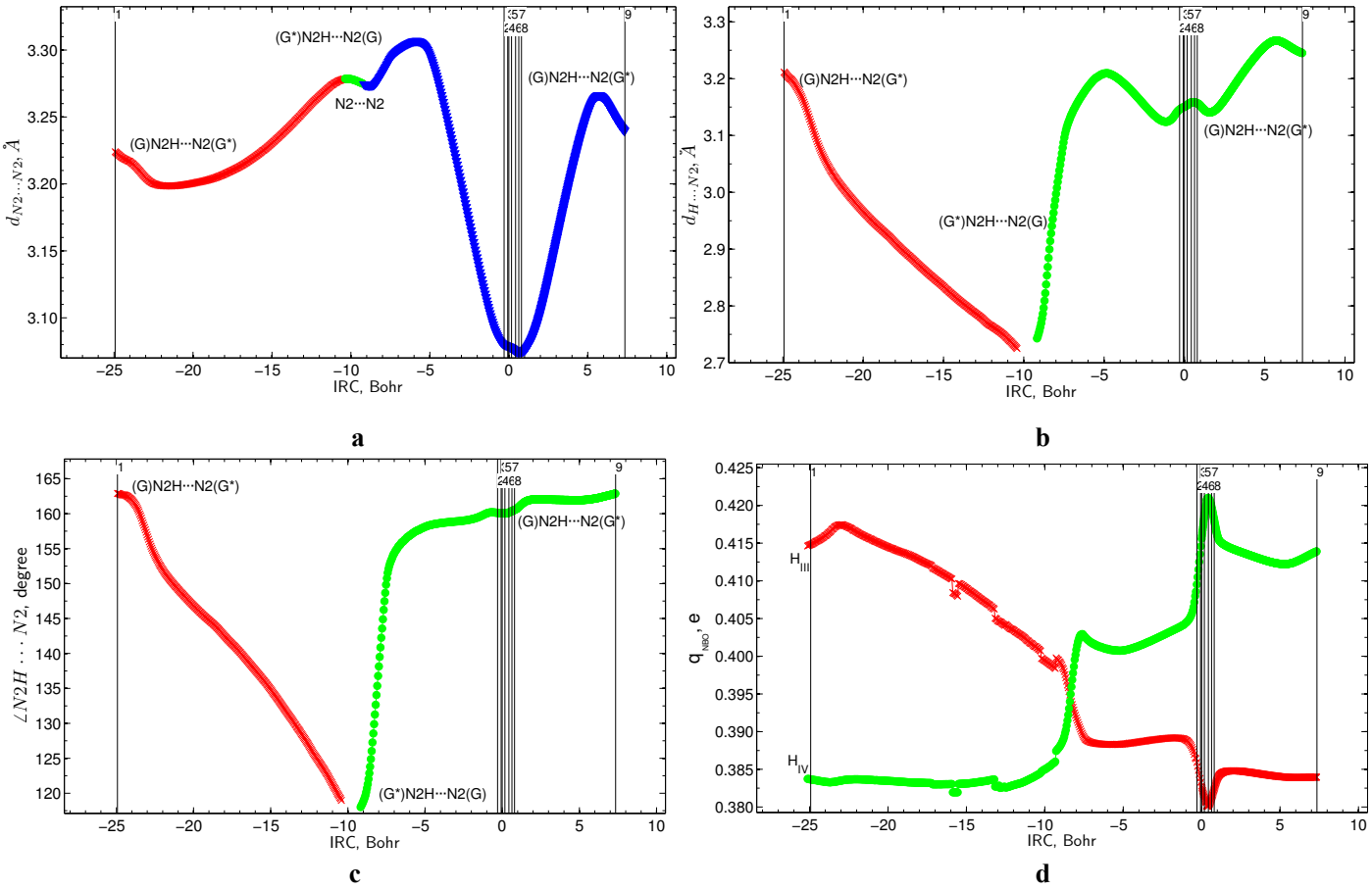


Fig. 7 Profiles of: (**a**) the distance $d_{N2\cdots N2}$ between the N2 nitrogen atoms; (**b**) the distance $d_{H\cdots N2}$ between the hydrogen and N2 nitrogen atoms; (**c**) the angle ∠N2H···N2 and (**d**) the NBO charges of the hydrogens of the (G)N2H_{III}···N2(G*) / (G*)N2H_{IV}···N2(G) H-bonds and the N2···N2 van der Waals contact along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*

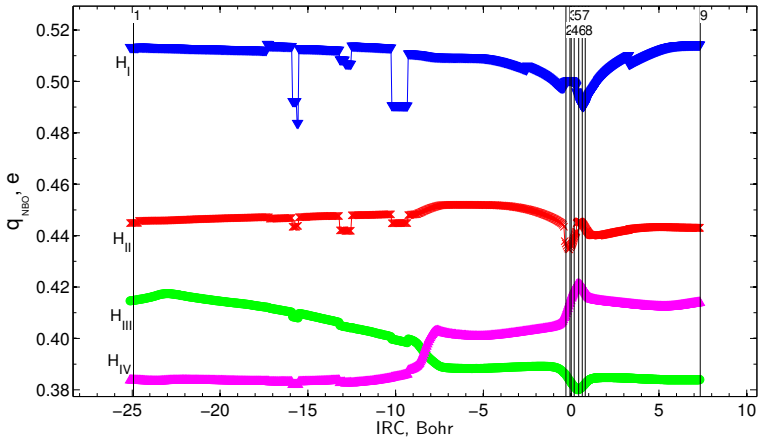
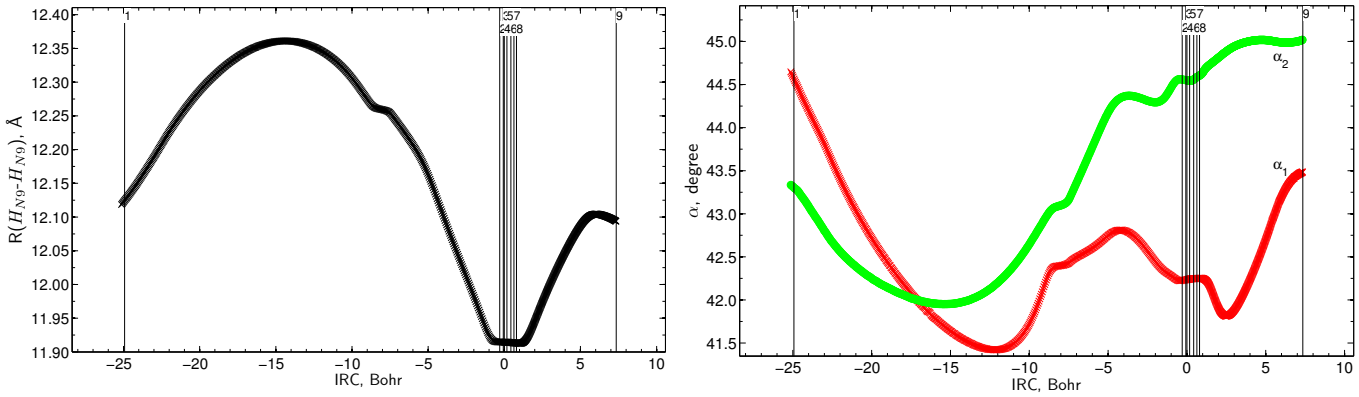
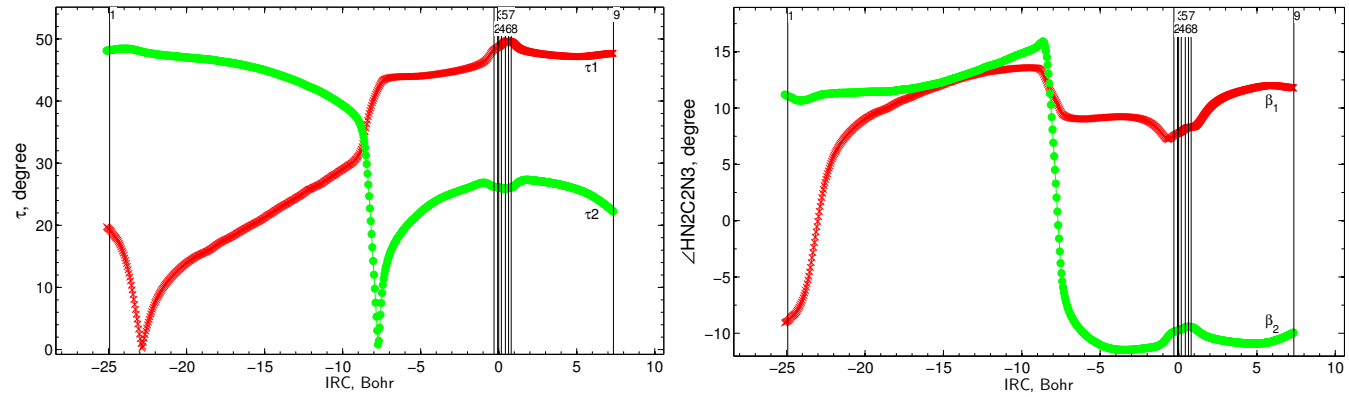


Fig. 8 Profiles of the NBO charges of the hydrogen atoms involved in the $N1H_I\cdots N1$, $O6H_{II}\cdots O6$ and $N2H_{III}\cdots N2$ H-bonds and in the $N2H_{IV}$ (G*) imino group of the G* base along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*



a **b**

Fig. 9 Profiles of: (**a**) the distance R(H_{N9}-H_{N9}) between the glycosidic hydrogens and (**b**) the α_1 (∠N9H(G)H(G*)) and α_2 (∠N9H(G*)H(G)) glycosidic angles along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*



a **b**

Fig. 10 Profiles of: (**a**) the angles between the HN2H plane and the C2N2 exocyclic bond of the amino fragments in the G (τ_1) and G* (τ_2) bases and (**b**) the β_1 (∠HN2C2N3(G)) and β_2 (∠HN2C2N3(G*)) dihedral angles along the IRC of the G·G*↔G*·G tautomerisation *via* the DPT obtained at the B3LYP/6-311++G(d,p) level of theory *in vacuo*

Graphical Abstract.

The G·G*↔G*·G DPT tautomerisation assisted by the third specific contact proceeds through the asynchronous concerted mechanism.