Quantum Surprises from the Watson-Crick and Hoogsteen G·C Nucleobase Pairs: A Comprehensive QM/QTAIM Investigation

Ol’ha O. Brovarets (o.o.brovarets@gmail.com)
Institute of Molecular Biology and Genetics National Academy of Sciences of Ukraine
https://orcid.org/0000-0002-8929-293X

Alona Muradova
Taras Shevchenko National University of Kyiv

Dmytro M. Hovorun
Institute of Molecular Biology and Genetics National Academy of Sciences of Ukraine

Research Article

Keywords: DNA, RNA, G-C nucleobase pair, tautomerization, transition state, mutagenic tautomer, tight ion pair, quantum-chemical calculations, mutations

Posted Date: December 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1109763/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Quantum surprises from the Watson-Crick and Hoogsteen G·C nucleobase pairs: A comprehensive QM/QTAIM investigation

Ol’ha O. Brovarets1, Alona Muradova2 & Dmytro M. Hovorun1,2

1Department of Molecular and Quantum Biophysics, Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, 150 Akademika Zabolotnoho Street, 03680, Kyiv, Ukraine
2Department of Molecular Biotechnology and Bioinformatics, Institute of High Technologies, Taras Shevchenko National University of Kyiv, 2-h Akademika Hlushkova Avenue, 03022, Kyiv, Ukraine

Corresponding author. E-mail: o.o.brovarets@gmail.com

Abstract. In this study at the MP2/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory in the isolated state it was revealed 14 novel physico-chemical mechanisms of the tautomerization of the G·C nucleotide base pairs in the Watson-Crick G·C(WC) / G*·C*(WC), reverse Watson-Crick G*·C*(rWC) / G·C*(rWC), Hoogsteen G*·C*(H) / G*N7·C(H) or reverse Hoogsteen G*N7·C*(H) / G*·C* (rH) configurations into the wobble (wWC, wH) and reverse wobble (rwWC, rwH) base pairs: 1. G·C(WC)↔G·C*(rwWC), 2./3. G*·C*(WC)↔G·C*(rwWC)/G*N2·C*(rwWC), 4. G*·C*(rWC)↔G*·C*(wWC), 5. G·C*O2(rWC)↔G·C*(wWC); 6./7./8./9. G*·C*(H)↔G*·C(rwH)/G*N1·O2(wH)/G*N1·O2(rwH)/G*N7·C*(rwH)↔G*N1·C*(rwH), 10. G*N7·C(H)↔G*N7·C*(wH) amino, 11./12. G*N1·C*(rH)↔G*N7·C*(wH)/G*N1·C*(wH), 13. G*N7·C*(rH)↔G*N7·C*(wH)↔G*N7·C*(wH) and 14. G*N7·C*(rwH)↔G*N7·C*(rwH) perp↔G*·C*(wH)↔G*N7·C(rwH) reaction pathways. It was established that the presence in the base pair of the two anti-parallel neighboring H-bonds is a necessary and sufficient condition for the implementation of such transformations, since it enables intermolecular proton transfer between the bases inside the base pair. It was found out that these tautomeric transitions are controlled by the TSs with quasi-orthogonal structure, which are tight G*·C/G·C* ion pairs, joined by at least two parallel intermolecular H-bonds, connected on a common negatively charged endocyclic N/C atoms – proton acceptor. All reaction pathways have been reliably confirmed. These transitions are accompanied by the changing of the mutual cis-orientation of the N9H and N1H glycosidic bonds of the bases on the trans-orientation and vice versa. These data complement the reported earlier mechanisms of the tautomerisations of the classical A·T and G·C DNA base pairs. Experimental verification of the novel G·C nucleobase pairs is looking as an attractive task for the future research.

Key words: DNA; RNA; G·C nucleobase pair; tautomerization; transition state; mutagenic tautomer; tight ion pair; quantum-chemical calculations.
**INTRODUCTION.**

The topic of tautomerism is of paramount importance nowadays [1-7], since, from the one side, it enables to explain the chemical structure of the biomolecules, and, from the other side – their functioning in the living cell.

In general, this topic attracted active researchers’ attention in different areas of research – drug design, physics of crystals, in NMR spectroscopy and biologically important molecules [8-13]. The point of view, that in biological molecules tautomeric transformations are inseparable from the conformational transformations, is becoming more and more popular last time [3, 14]. This enables to open new possibilities for the understanding of the subtle intimate mechanisms of the functioning of the biomolecules in a living cell.

Tautomerism represents especial interest in nucleic acids [15], since transfer of single proton inside the nucleobase pairs leads to the groundbreaking changes in their structure and functioning [4]. Also, tautomerism is usually associated with the mutagenic properties of the molecules [3, 16]. Thus, recently it was found [3, 18] that sequential single proton transfer inside classical Watson-Crick A-T/G-C and unusual DNA base pairs leads to the Watson-Crick↔wobble transitions and further formation of the wobble base pairs involving rare tautomers, which cause spontaneous point mutations.

In view of all presented in the literature data [2, 7, 18, 19] on the possible mechanisms of tautomerization, it arises quite logical question – “How complete they are?” Just answer on this biologically important question enables to establish which of them are responsible for the spontaneous point mutations and which – for the other biological roles.

This study is a further development of the previous works [20-22], devoted to the tautomerically-conformational transformations of the classical G·C DNA base pairs, leading to the formation of the mutagenic tautomers of the G and C DNA bases.

So, *aim of this study* – to reveal physico-chemical mechanisms, which define the tautomerization of the G·C nucleobase pair in its four biologically important configurations. These investigations have been performed at the MP2/6-311++G(2df,pd)/B3LYP/6-311++G(d,p) level of QM theory. Obtained data significantly extend existing ideas about the possible biological role of the prototropic tautomerism of the pairs of nucleotide bases in the processes of the functioning of the nucleic acids.

**COMPUTATIONAL METHODS.**

**Density functional theory calculations of the geometry and vibrational frequencies.**

Equilibrium geometries of the investigated G·C nucleobase pairs and transition states (TSs) of their mutual conformationally-tautomeric transformations, as well as their harmonic vibrational frequencies have been calculated at the B3LYP/6-311++G(d,p) level of theory [23-27], using
Gaussian’09 program package [28]. Applied level of theory has proved itself successful for the calculations of the similar systems [29, 30]. A scaling factor that is equal to 0.9668 has been applied in the present work for the correction of the harmonic frequencies of all complexes and TSs of their tautomeric and conformational transitions [31, 32]. We have confirmed the local minima and TSs, localized by Synchronous Transit-guided Quasi-Newton method [33], on the potential energy surface by the absence or presence, respectively, of the imaginary frequency in their vibrational spectra. Further, reaction pathways of the conformationally-tautomeric transformations have been confirmed by using the Intrinsic Reaction Coordinate (IRC) procedure [20, 21], moving from each TS in the reverse and forward directions.

All calculations have been carried out in the continuum with \( \varepsilon=1 \) under normal conditions (T=298.15 K) [34, 35], that adequately reflects the processes occurring in real biological systems without reduction of the intrinsically inherent structurally-functional properties of the base pairs in the composition of DNA. Moreover, this environment (\( \varepsilon=1 \)) satisfactorily models base pair recognition pocket of the DNA-polymerase machinery, which is substantially hydrophobic.

**Single point energy calculations.** Geometry optimizations have been followed by the electronic energy calculations as the single point calculations for the optimized geometries of the G·C nucleobase pairs and TSs of their conformationally-tautomeric transformations at the MP2/6-311++G(2df,pd) level of theory [36, 37].

**QTAIM analysis.** Bader’s Quantum Theory of Atoms in Molecules (QTAIM) [38-43] has been applied by using program package AIMAll [44] in order to analyze the electron density distribution. The presence of the bond critical point (BCP), namely the so-called \((3,-1)\) BCP, and a bond path between hydrogen donor and acceptor, as well as the positive value of the Laplacian at this BCP (\( \Delta \rho > 0 \)), have been considered as criteria for the H-bond formation [38-43]. Wave functions have been obtained at the B3LYP/6-311++G(d,p) level of theory.

**OBTAINED RESULTS AND THEIR DISCUSSION.**

Obtained results are presented in Table 1 and on Figure 1. Their careful analysis revealed the data, which are analyzed in more details below.

First, it would be considered novel pathways of the tautomerizations for the G·C nucleobase pairs, which have Watson-Crick geometry with \( \text{cis-} \) oriented N9H and N1H glycosidic bonds and reverse Watson-Crick geometry with \( \text{trans-} \) oriented N9H and N1H glycosidic bonds.

1. It was established that classical Watson-Crick G·C(WC) pair of the nucleotide bases tautomerizes into the reverse wobble Watson-Crick G·C*(rw\(_{\text{WC}}\)) nucleobase pair with \( \text{trans-} \) oriented N9H and N1H glycosidic bonds through two different pathways as from the topological, so from the energetical points of view: 1. G·C(WC)\( \leftrightarrow \)G·C*(rw\(_{\text{WC}}\)) and 2. G*·C*(WC)\( \leftrightarrow \)G·C*(rw\(_{\text{WC}}\)) (Fig. 1, 1 and 2). First of them – G·C(WC)\( \leftrightarrow \)G·C*(rw\(_{\text{WC}}\)) – occurs
directly through the quasi-orthogonal \( \text{TS}^{G^*\cdot C^+(WC)\leftrightarrow G^*\cdot C^+(rWWC)} \) transition state, which is tight ion pair – deprotonated \( G^* \cdot C^+ \), deprotonated by the N1 site of the G base and covered by three intermolecular specific contacts – two \((G)\text{N}3^+\text{H}...\text{N}1^+(G)\) and \((G)\text{N}3^+\text{H}...\text{N}1^+(G)\) H-bonds, which are locked on the joint N1° nitrogen atom of the \( G \), and one attractive \((G)\text{O}6^-...\text{N}3^+(C)\) van der Waals contact (Fig. 1, I). Second of these mechanisms – \( G^*\cdot C^*(WC)\leftrightarrow G^*\cdot C^*(rWWC) \) – occurs through the short-lived intermediate – Löwdin’s \( G^*\cdot C^*(WC) \) base pair and \( \text{TS}^{G^*\cdot C^*(WC)\leftrightarrow G^*\cdot C^*(rWWC)} \) transition state, tight \( G^*\cdot C^- \) ion pair, which is protonated by the O6 site of the G base and joined by two intermolecular \((G)\text{O}6^-\text{H}...\text{N}3^+(C)\) and \((G)\text{N}1^+\text{H}...\text{N}3^+(C)\) H-bonds, which are locked on the joint nitrogen N3° atom of the \( C^- \) (Fig. 1, 2). Obviously, that tautomeric transformation of the G-C(WC) base pair through the intermediate \( G^*\cdot C^*(WC) \) nucleobase pair is more favorable from the energetical point of view (relative Gibbs free energy of the TS \( \Delta G_{TS}=25.67 \text{ kcal}\cdot\text{mol}^{-1} \)) and thus – it is faster from the kinetical point of view (Fig. 1, 2, Table 1).

2. Moreover, transition of the classical G-C(WC) base pair into the Löwdin’s \( G^*\cdot C^*(WC) \) base pair [2] – \( G\cdot C(WC)\leftrightarrow G^*\cdot C^*(WC) \) – also provides a tautomeric transition of the Löwdin’s \( G^*\cdot C^*(WC) \) base pair into the reverse wobble Watson-Crick \( G^*\cdot C^*(rWWC) \) base pair: \( 3. G^*\cdot C^*(WC)\leftrightarrow G^*\cdot C^*(rWWC) \) reaction pathway (Fig. 1, 3). It is realized through the \( \text{TS}^{G^*\cdot C^+(WC)\leftrightarrow G^*\cdot C^+(rWWC)} \) – tight \( G^-\cdot C^+ \) ion pair, which is deprotonated by the N1 and N2 sites, protonated by the O2 site of the G base and joined by three intermolecular \((C)\text{O}2^-\text{H}...\text{N}1^+(G)\), \((C)\text{O}2^-\text{H}...\text{N}1^+(G)\) and \((C)\text{N}3^+\text{H}...\text{N}1^+(G)\) H-bonds. This is the slowest tautomerization of the classical G-C(WC) base pair among all base pairs, which are considered in this work \( (\Delta G_{TS}=62.00 \text{ kcal}\cdot\text{mol}^{-1}) \) (Table 1).

3. Quite intriguing results have been obtained for the tautomerization mechanisms of the reverse Watson-Crick \( G^*\cdot C^*(rWC) \) and \( G\cdot C^*\cdot O_2(rWC) \) nucleobase pairs with \textit{trans}-oriented N9H and N1H glycosidic bonds, the last of which is high-energetical tautomer (relative Gibbs free \( \Delta G / \Delta E \) energy=3.41 / 3.70 kcal-mol\(^{-1}\)) of the reverse Watson-Crick \( G^*\cdot C^*(rWC) \) base pair, leading to the wobble Watson-Crick \( G^*\cdot C^*(wWC) \) and \( G\cdot C^*(wWC) \) nucleobase pairs: \( 4. G^*\cdot C^*(rWC)\leftrightarrow G^*\cdot C^*(wWC) \) and \( 5. G\cdot C^*\cdot O_2(rWC)\leftrightarrow G\cdot C^*(wWC) \) reaction pathways (Fig. 1, 4 and 5). First of these base pairs – \( G^*\cdot C^*(rWC) \) – tautomerically transforms into the wobble Watson-Crick \( G^*\cdot C^*(wWC) \) base pair – \( G^*\cdot C^*(rWC)\leftrightarrow G^*\cdot C^*(wWC) \) – through the \( \text{TS}^{G^*\cdot C^+(rWC)\leftrightarrow G^*\cdot C^+(wWC)} \) – quasi-orthogonal tight \( G^-\cdot C^+ \) ion pair, which is deprotonated by the N1 and N2 sites, protonated by the O6 site of the G base and joined by four specific intermolecular contacts – two \((C)\text{N}4^+\text{H}...\text{N}1^+(G)\) and \((C)\text{N}3^+\text{H}...\text{N}1^+(G)\) H-bonds, which are locked on the joint N1° atom, and two attractive \((C)\text{N}4^-...\text{N}2^+(G)\) and \((C)\text{N}3^-...\text{N}2^+(G)\) van der Waals contacts, which are locked on the joint N2° atom of the \( C^- \) (Fig. 1, 4). Second of the afore mentioned base pairs – \( G^*\cdot C^*\cdot O_2(rWC) \).
tautomERICALLY transforms into the wobble Watson-Crick G·C*(wWC) nucleobase pair, for which pyrimidine C base is exposed into the major groove of DNA accordingly the purine G base. This G·C*O2(rWC)→G·C*(wWC) transition is controlled by the TS

G·C*O2(rWC)→G·C*(wWC) – tight G·C* ion pair with quasi-orthogonal geometry, which is deprotonated by the N1 site of the G base and joined by two intermolecular (C)O2H...N1(G) and (C)N3H...N1(G) H-bonds (Fig. 1, 5).

Notably, that tautomeric transition of the G*·C*(rWC) base pair into the G*·C*(wWC) base pair is slower (ΔΔGTS=43.13 kcal/mol1), than transition of the G·C*O2(rWC) base pair into the G·C*(wWC) base pair (ΔΔGTS=30.67 kcal/mol1) since relative Gibbs free energy barrier ΔΔGTS of the G*·C*(rWC)→G*·C*(wWC) reaction is higher on 12.46 kcal/mol1 in comparison with the G·C*O2(rWC)→G·C*(wWC) reaction (Table 1).

Further, we have considered results, which are attractive from the biological point of view and are concerning novel pathways of the tautomerization of the Hoogsteen and reverse Hoogsteen G·C nucleobase pairs with trans-oriented N9H and N1H glycosidic bonds.

4. It was shown that Hoogsteen G*4·C*(H) nucleobase pair with cis-oriented N9H and N1H glycosidic bonds and trans-oriented O6H hydroxyl group of the G* base tautomerises into the non-planar reverse wobble Hoogsteen G*4·C(rWH) base pair with trans-oriented N9H and N1H glycosidic bonds – 6. G*4·C*(H)→G*4·C(rWH) (Fig. 1, 6). This tautomeric transition is controlled by the TS

G*4·C(rWH)→G*4·C*(rWH), which is tight G*·C+ ion pair with quasi-orthogonal geometry, protonated by the N7 site of the G base and stabilized by three intermolecular H-bonds – (G)O6H...N3(C) and (G)N7H...N3(C), which are focused on the common N3 atom, and (G)N7H...N4(G). Two last of them are bifurcated from the common N7H group of the G base.

5. At the same time, the Hoogsteen G*4·C*(H) nucleobase pair demonstrates high ability to tautomerise into the wobble Hoogsteen G*4·C*O2(wWH) nucleobase pair through the 7. G*4·C*(H)→G*4·C*O2(wWH) reaction pathway (Fig. 1, 7) and TS

G*4·C*O2(wWH) which is tight G*·C+ ion pair with quasi-orthogonal geometry, protonated by the N7 site of the G base. This structure is stabilized by three intermolecular H-bonds – (G)O6H...N3(C) and (G)N7H...N3(C), which are focused on the common N3 atom, and (G)N7H...O2'(G). Two last of them are bifurcation (Fig. 1, 7).

6. Quite unexpected mechanism defines the tautomerization of the Hoogsteen G*4·C*(H) nucleobase pair into the reverse wobble Hoogsteen G*4·C*O2(rWH) structure – 8. G*4·C*(H)→G*4·C*O2(rWH) (Fig. 1, 8). This reaction is controlled by the TS

G*4·C*O2(rWH) which is tight G*·C+ ion pair with quasi-orthogonal geometry by the participation of the carbanione of the G base, deprotonated by the C8 site of the G base (Fig. 1, 8). This structure is joined by two unexpected intermolecular (C)O2H...C8(G) and (C)N3H...C8(G) H-bonds, which are focused on the common C8 atom (Fig. 1, 8). This one-stage
8. $G^*\cdot C^*(H)\leftrightarrow G^*\cdot C^*(O_2(rwH))$ reaction is quite more rapid ($\Delta \Delta G_{TS} = 59.69$ kcal·mol$^{-1}$), than two-stage $9. G^*\cdot C^*(H)\leftrightarrow G^*\cdot N_7^*\cdot C^*(rwH)\leftrightarrow G^*\cdot C^*(O_2(rwH))$ reaction ($\Delta \Delta G_{TS} = 70.71$ kcal·mol$^{-1}$) (Fig. 1, 9), which is controlled by two TSs – $TS^{G^*-C^*G^*\cdot C^*(H)\rightarrow G^*\cdot N_7^*\cdot C^*(rwH)}$, which is quasi-orthogonal tight G$^+$$\cdot$C$^+$ ion pair, deprotonated by the C8 site of the G base and joined by two (C)O2$^+$H...N7(G) and (C)N3$^+$H...N7(G) H-bonds, which are connected on the common N7$^-$ atom, and covalently-bonded $TS^{G^*\cdot N_7^*\cdot C^*(rwH)\rightarrow G^*\cdot C^*O_2(rwH)}$, which is joined by the (G)N7H...O2(C) H-bond and (G)C8-H-N3(C) covalent bridge.

7. Quite intriguing situation is observed in the case of the tautomerization of the Hoogsteen G$^*\cdot C^*(H)$ base pair, which is high-energy tautomer of the G$^*\cdot C^*(H)$ pair ($\Delta G / \Delta E = 3.92 / 3.79$ kcal·mol$^{-1}$), through the 10. G$^*\cdot N_7^*\cdot C^*(H)\leftrightarrow G^*\cdot C^*(wH)_\text{amino}$ reaction pathway (Fig. 1, 10). This reaction proceeds through the quasi-orthogonal $TS^{G^+\cdot C^*-G^*\cdot N_7^*\cdot C^*(H)\rightarrow G^*\cdot C^*(wH)_\text{amino}}$, which is tight G$^+\cdot$C$^+$ ion pair, involving G$^+$ base, protonated at the N7 site. This structure is stabilized by two intermolecular (G)O6$^+$H...N4(C) and (G)N7$^+$H...N4(C) H-bonds, which are focused on the common N4$^-$ atom of the C base. Product of this reaction – wobble Hoogsteen G$^*\cdot C^*(wH)_\text{amino}$ base pair is significantly non-planar and has cis-orientation of the N9H and N1H glycosidic bonds, similarly to the starting G$^*\cdot N_7^*\cdot C^*(H)$ base pair (Fig. 1, 10).

8. The G$^*\cdot C^*(rH)$ nucleobase pair, which has reverse Hoogsteen configuration, tautomeres into the G$^*\cdot N_7^*\cdot C^*(wH)$ and G$^*\cdot N_7^*\cdot C^*(wH)_\text{amino}$ base pairs with wobble Hoogsteen configuration through the quasi-orthogonal $TS^{G^+\cdot C^*-G^*\cdot N_7^*\cdot C^*(rH)\rightarrow G^*\cdot N_7^*\cdot C^*(wH)}$ and $TS^{G^+\cdot C^*-G^*\cdot N_7^*\cdot C^*(rH)\rightarrow G^*\cdot C^*(wH)_\text{amino}}$, which contains G$^+$, deprotonated by the O6 site, and is stabilized by the participation of the four specific intermolecular interactions – two (C)O2$^+$H...N7(G) and (C)N3$^+$H...N7(G) H-bonds, focused on the common N7$^-$ atom of the G base, and two attractive van der Waals contacts – (G)O6$^-$...O2$^+(C)$ and (G)O6$^-$...N3$^+(C)$, focused on the common O6$^-$ atom of the G base. First of the van der Waals contacts – (G)O6$^-$...O2$^+(C)$ – is the shortest (2.676 Å) among all others considered here (Fig. 1, 11). Second 12. G$^*\cdot C^*(rH)\rightarrow G^*\cdot C^*(wH)$ reaction is controlled by the $TS^{G^+\cdot C^*-G^*\cdot N_7^*\cdot C^*(rH)\rightarrow G^*\cdot N_7^*\cdot C^*(wH)}$, which is tight G$^+\cdot$C$^+$ ion pair by the participation of the G$^+$, deprotonated at the C8 site and containing carbanione of the G base. This structure is stabilized by two unusual intermolecular (C)N4$^+$H...C8(G) and (C)N3$^+$H...C8(G) H-bonds, which are focused on the common C8$^-$ atom of the G base (Fig. 1, 12).

9. Reverse Hoogsteen G$^*\cdot N_7^*\cdot C^*(rH)$ nucleobase pair, which is high energy tautomer of the reverse Hoogsteen G$^*\cdot C^*(rH)$ nucleobase pair ($\Delta G / \Delta E = 23.70 / 23.36$ kcal·mol$^{-1}$), tautomeres
into the wobble Hoogsteen $G^{\ast\cdot}C(w_H)$ nucleobase pair, involving carbanione of the G base, by two-stage reaction – 13. $G^{\ast\cdot}N^7C(r_H)\leftrightarrow G^{\ast\cdot}N^7C(w_H)\leftrightarrow G^{\ast\cdot}C(w_H)$ (Fig. 1, 13).

First stage 13. $G^{\ast\cdot}N^7C(r_H)\leftrightarrow G^{\ast\cdot}N^7C(w_H)$ is controlled by the quasi-orthogonal $TS^{G-\cdot C^+G^*\cdot N^7\cdot C(r_H)\leftrightarrow G^*\cdot N^7\cdot C(w_H)}$, which is tight $G\cdot C^+$ ion pair by the participation of the carbanione of the G base, deprotonated by the C8 site and joined by the participation of two intermolecular (C)N4\( ^+\)H…N7\( ^-\)(G) and (C)N3\( ^+\)H…N7\( ^-\)(G) H-bonds, which are focused on the common N7 atom of the G base.

Second stage 13. $G^{\ast\cdot}N^7\cdot C^*(w_H)\leftrightarrow G^{\ast\cdot}C(w_H)$ represents double proton transfer in the wobble Hoogsteen $G^{\ast\cdot}N^7\cdot C^*(w_H)$ base pair, which occurs through the covalently-bonded $TS^{G^*\cdot C^*(w_H)\leftrightarrow G^*\cdot C^*(r_H)}$, stabilized by the (G)C8-H-N3(C) covalent bridge and (G)N7H…N4(C) H-bond, leading to the wobble Hoogsteen $G^{\ast\cdot}C(w_H)$ nucleobase pair (Fig. 1, 13). Formed $G^{\ast\cdot}C(w_H)$ nucleobase pair is stabilized by the participation of two intermolecular (C)N4\( ^+\)H…N7(G) and (G)C8\( ^+\)H…N3(C) H-bonds.

10. Finally, for the first time it was observed quite complicated mechanism of the tautomeric transformation of the reverse wobble Hoogsteen $G^{\ast\cdot}N^7\cdot C^*(r_w_H)$ nucleobase pair into the reverse wobble Hoogsteen $G^{\ast\cdot}C(r_w_H)$ nucleobase pair – 14. $G^{\ast\cdot}N^7\cdot C^*(r_w_H)\leftrightarrow G^{\ast\cdot}N^7\cdot C^*(r_p_r_H)$ perp $\leftrightarrow G^{\ast\cdot}C^*(r_H)\leftrightarrow G^{\ast\cdot}C(w_H)$ (Fig. 1, 14). This reaction is not accompanied by the changing of the mutual orientation of the N9H and N1H glycosidic bonds.

This transition occurs through three TSs: $TS^{G^*\cdot N^7\cdot C^*(r_w_H)\leftrightarrow G^*\cdot N^7\cdot C^*(r_p_r_H)}$ perp and two dynamically stable intermediates: reverse wobble Hoogsteen $G^{\ast\cdot}N^7\cdot C^*(r_p_r_H)$ perp nucleobase pair and reverse wobble Hoogsteen $G^{\ast\cdot}C^*(r_w_H)$ ion pair, which are significantly non-planar structures (Fig. 1, 14). First among TSs – $TS^{G^*\cdot N^7\cdot C^*(r_w_H)\leftrightarrow G^*\cdot N^7\cdot C^*(r_p_r_H)}$ perp is responsible for the conformational $G^{\ast\cdot}N^7\cdot C^*(r_w_H)\leftrightarrow G^{\ast\cdot}N^7\cdot C^*(r_p_r_H)$ perp transition, second among TSs – $TS^{G^*\cdot N^7\cdot C^*(r_p_r_H)\leftrightarrow G^{\ast\cdot}C^*(r_w_H)}$ is responsible for the $G^{\ast\cdot}N^7\cdot C^*(r_p_r_H)$ perp $\leftrightarrow G^{\ast\cdot}C^*(r_w_H)$ single proton transfer and third among TSs – $TS^{G^*\cdot C^*G^{\ast\cdot}C^*(r_w_H)\leftrightarrow G^{\ast\cdot}N^7\cdot C^*(r_p_r_H)}$ leads to the formation of the reverse wobble Hoogsteen $G^{\ast\cdot}C(r_p_r_H)$ nucleobase pair (Fig. 1, 14). At this, $TS^{G^*\cdot N^7\cdot C^*(r_H)\leftrightarrow G^*\cdot N^7\cdot C^*(r_p_r_H)}$ perp is stabilized by the participation of the intermolecular (G)N7H…N4(C) H-bond and attractive (G)O6…N3(C) van der Waals contact; $TS^{G^*\cdot C^*(r_w_H)\leftrightarrow G^{\ast\cdot}C^*(r_p_r_H)}$ is stabilized by the attractive (G)O6…C4(C) van der Waals contact and (G)N7-H-N4(C) covalent bridge; $TS^{G^*\cdot C^*G^{\ast}\cdot C^*(r_w_H)\leftrightarrow G^{\ast\cdot}C^*(r_p_r_H)}$ is a tight $G^{\cdot}C^+$ ion pair, deprotonated by the N7 site of the G base, and stabilized by two intermolecular (C)N4\( ^+\)H…N7\( ^-\)(G) and (C)N3\( ^+\)H…N7\( ^-\)(G) H-bonds, focused on the common N7 atom, and attractive (G)O6…N4\( ^+\)(C) van der Waals contact. The $G^{\cdot}C^+(r_w_H)$ ion pair with quasi-orthogonal geometry is stabilized by the participation of two intermolecular (C)N4\( ^+\)H…N7\( ^-\)(G) and (C)N3\( ^+\)H…N7\( ^-\)(G) H-bonds, focused on the common N7 atom, and attractive (G)O6…C4\( ^+\)(C) van der Waals contact.
Now it would be shortly considered non-planar deformation of the G and C nucleotide bases, accompanying investigated here processes. Despite the structural softness of the G and C bases for bending [20, 21], cycles of the G and C bases remain planar even at the TSs of the transitions. The largest orientational deformation occurs in the exocyclic fragments – NH$_2$ amino groups and OH hydroxyl groups. Thus, in particular, piramidality of the NH$_2$ amino group of the G base significantly increases at the transformation into the G$^+$·C$^+$ ion pair – it could be explained by the weakness of the electronic conjugation between the N2 amino atom and π-electron system of the ring of the base at its deprotonation. Also, it is observed pyramidalization of the NH$_2$ amino group of the protonated C base. The largest changes in the orientation of the exocyclic groups were observed at the TS$^{G$·C$}\rightarrow G^+\cdot C^+(\text{rH})\rightarrow G^N7\cdot C^*(\text{wH})$ (Fig. 1, 11).

Last, but not least. Joining results, which have been obtained in this work, with the data, which have been received in our previous works [20, 21], we came to the conclusion, that classical Watson-Crick G·C(WC) nucleobase pair with cis-oriented glycosidic bonds transforms to two reverse Watson-Crick G$^t$·C$(rWC)$ and G·C$^*$O$_2$(rWC) nucleobase pairs with trans-oriented glycosidic bonds, which are in tautomeric equilibrium between each other, by four different reaction pathways: 1) G·C(WC)$\leftrightarrow$G·C$(rWC)$↔G·C$^*$O$_2$(rWC);
2) G·C(WC)$\leftrightarrow$G$^t$·C$(rWC)$↔G·C$(rWC)$↔G·C$^*$O$_2$(rWC);
3) G·C(WC)$\leftrightarrow$G$^t$·C$(wWC)$↔G$^t$·C$(rWC)$; 4) G·C(WC)$\leftrightarrow$G·C$^*$O$_2$(rWC)$\leftrightarrow$G·C$^*$O$_2$(rWC).

Similarly, it has been proven that Hoogsteen G$^t$·C$(rH)$ and G$^t$N7·C$(rH)$ nucleobase pairs transform to three reverse Hoogsteen G$^t$·C$(rH)$, G$^t$·C$^*$O$_2$(rH) and G$^t$N7·C(rH) nucleobase pairs through six different pathways: 1) G$^t$·C$(rH)$↔G$^t$·C$(rH)$↔G$^t$·C$(wH)$↔G$^t$·C$^*$O$_2$(rH);
2) G$^t$·C$(rH)$↔G$^t$·C$(rH)$↔G$^t$·C$(wH)$↔G$^t$·C$^*$O$_2$(rH); 3) G$^t$·C$(rH)$↔G$^t$N7·C$(wH)$↔G$^t$·C$^*$O$_2$(rH)↔G$^t$·C$^*$O$_2$(rH); 4) G$^t$N7·C$(rH)$↔G$^t$N7·C$(rH)$↔G$^t$·C$(wH)$↔G$^t$·C$(rH)$; 5) G$^t$i·C$(rH)$↔G$^t$i·C$(wH)$↔G$^t$i·C$(rH)$; 6) G$^t$i·C$(rH)$↔G$^t$iN7·C$(wH)$↔G$^t$iN7·C$(rH)$.

Attractiveness of these tautomeric transitions, which are accompanied by the changing of the mutual orientation of the glycosidic bonds of the bases from cis- on trans-orientation, consists in the fact that they have quite high energy of interaction at the TSs (ΔG$_{\text{int}}$ > 100 kcal·mol$^{-1}$) and thus can reorganize stacking of the neighboring base pairs and significantly change the conformation of the sugar-phosphate residues. In other words, they are perfect pretendents on the role of the drivers of the transition of the DNA and RNA molecules from the states with anti-parallel strands into the duplexes with parallel strands [45].
CONCLUSIONS.

In this study it was performed careful QM/QTAIM investigation, aimed to identify all possible quantum mechanisms of the tautomerization of the classical G·C nucleobase pairs as their intrinsic property, which enable to make the following conclusions.

It was revealed for the first time novel physico-chemical mechanisms of the conformationally-tautomeric transitions of the four biologically-important configurations of the G·C nucleobase pairs – Watson-Crick G·C(WC), reverse Watson-Crick G·C(rWC), Hoogsteen G·C(H) and reverse Hoogsteen G·C(rH), occurring through the quasi-orthogonal TSs, which represent tight G⁺·C/G⁻·C⁺ ion pairs, joined by at least two anti-parallel H-bonds, closed on a common N/C endocyclic atom carrying negative charge.

Existence of the two neighboring anti-parallel intermolecular H-bonds in the base pair is a necessary and sufficient condition of the existence of such tautomeric transformations. In this case tautomerization realises exclusively by the participation of the atoms, involved in these above mentioned H-bonds.

Revealed pathways of the quantum tautomerization are controlled by the quite complicated potential energy hypersurface, which have been obtained by the base pairs in the course of the long-term evolution. However, these transitions could not be predicted by the qualitative stereochemical analysis without thorough QM/QTAIM research. Every of each transition has very unique behavior and could not be generalized with the aim of their extension on others pairs of nucleotide bases.

Biological attractiveness of the detected tautomeric transformations, which are accompanied by the changing of the mutual orientation of the glycosidic bonds of the bases on the opposite (cis↔trans) consist in the fact that they can determine tautomically-conformational transitions both in DNA, so in RNA molecules, leading to the changing of the mutual orientation of their sugar-phosphate residues – from the anti-parallel on the parallel and vice versa. Moreover, they could be responsible for the supporting of the unique spatial structures of these biological macromolecules in the process of their functioning.

Declarations.

Funding: Not applicable.
Conflicts of interest/Competing interests: Not applicable.
Availability of data and material: Not applicable.
Authors’ contributions: OB – idea formulation, setting of the task, calculation of the data, building of the graphs, data extrapolation, preparing, and proofreading of the draft of the manuscript. AM – idea formulation, calculation of the data, preparing, and proofreading of the draft of the manuscript. DH – idea formulation, preparing, and proofreading of the draft of the manuscript. All authors contributed to the article and approved the submitted version.
REFERENCES.

Table 1. Vibrational and energetic characteristics of the tautomeric-conformational transformations of the G·C nucleobase pairs via the single (SPT) or double proton (DPT) transfer accompanied with the cis→trans conformational transformations through the quasi-orthogonal TSs as tight G·C/G·C+ ion pairs obtained at the MP2/6-311++G(2df,pd) // B3LYP/6-311++G(d,p) level of QM theory in the isolated state (ε=1) under normal conditions (T=298.15 K) (see Fig. 1).

<table>
<thead>
<tr>
<th>TautomERICally-CONFORMATIONal transformations</th>
<th>(v_i^{TS})</th>
<th>(\Delta \Delta G_{TS}^b)</th>
<th>(\Delta \Delta E_{TS}^c)</th>
<th>(\Delta G^d)</th>
<th>(\Delta E^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. G*·C*(WC)→G·C*(rwWC)</td>
<td>118.5</td>
<td>25.67</td>
<td>28.73</td>
<td>3.31</td>
<td>4.25</td>
</tr>
<tr>
<td>3. G*·C*(WC)→G*\text{N2}·C*(rwWC)</td>
<td>447.7</td>
<td>62.00</td>
<td>64.07</td>
<td>27.94</td>
<td>29.24</td>
</tr>
<tr>
<td>4. G*·C*(rWC)→G*·C(wWC)</td>
<td>34.7</td>
<td>43.13</td>
<td>43.60</td>
<td>4.04</td>
<td>5.74</td>
</tr>
<tr>
<td>5. G·C*(O2(rWC)→G·C*(wWC)</td>
<td>677.5</td>
<td>30.67</td>
<td>33.10</td>
<td>-0.74</td>
<td>0.03</td>
</tr>
<tr>
<td>6. G*·C*(H)→G*·C*(rwH)</td>
<td>187.4</td>
<td>27.70</td>
<td>28.77</td>
<td>2.29</td>
<td>3.04</td>
</tr>
<tr>
<td>7. G*·C*(H)→G*·C*(O2(rW))</td>
<td>195.4</td>
<td>29.03</td>
<td>30.13</td>
<td>16.63</td>
<td>17.20</td>
</tr>
<tr>
<td>8. G*·C*(H)→G*·C*(O2(wH))</td>
<td>1574.2</td>
<td>59.69</td>
<td>64.75</td>
<td>20.57</td>
<td>21.88</td>
</tr>
<tr>
<td>9. G*·C*(H)→G*·C*(O2(rW))</td>
<td>478.3</td>
<td>70.71</td>
<td>73.95</td>
<td>37.52</td>
<td>38.54</td>
</tr>
<tr>
<td>10. G*·C*(H)→G*·C*(wH)</td>
<td>1135.3</td>
<td>0.56</td>
<td>3.20</td>
<td>-16.96</td>
<td>-16.66</td>
</tr>
<tr>
<td>12. G*·C*(rH)→G*·C*(wH)</td>
<td>304.0</td>
<td>38.88</td>
<td>39.16</td>
<td>14.35</td>
<td>15.11</td>
</tr>
<tr>
<td>13. G*·C*(rH)→G*·C*(wH)</td>
<td>303.8</td>
<td>46.77</td>
<td>49.74</td>
<td>2.43</td>
<td>4.56</td>
</tr>
<tr>
<td>14. G*·C*(rH)→G*·C*(wH)</td>
<td>57.3</td>
<td>29.98</td>
<td>30.83</td>
<td>8.98</td>
<td>10.17</td>
</tr>
<tr>
<td>15. G*·C*(rH)→G*·C*(wH)</td>
<td>1141.3</td>
<td>-0.23</td>
<td>2.42</td>
<td>-30.25</td>
<td>-28.98</td>
</tr>
<tr>
<td>16. G*·C*(rH)→G*·C*(wH)</td>
<td>32.8</td>
<td>8.25</td>
<td>7.09</td>
<td>5.95</td>
<td>5.52</td>
</tr>
<tr>
<td>17. G*·C*(rH)→G*·C*(wH)</td>
<td>898.7</td>
<td>1.29</td>
<td>3.03</td>
<td>0.60</td>
<td>-0.31</td>
</tr>
<tr>
<td>18. G*·C*(rH)→G*·C*(wH)</td>
<td>39.1</td>
<td>3.39</td>
<td>3.11</td>
<td>5.13</td>
<td>0.16</td>
</tr>
</tbody>
</table>

\(^a\)Imaginary frequency at the TS of the tautomERICally-CONFORMATIONal transformation, cm\(^{-1}\);
\(^b\)Relative Gibbs free energy of the TS of the tautomERICally-CONFORMATIONal transformation (T=298.15 K), kcal·mol\(^{-1}\);
\(^c\)Relative electronic energy of the TSs of the tautomERICally-CONFORMATIONal transformation, kcal·mol\(^{-1}\);
\(^d\)Relative Gibbs free energy of the formed nucleobase pair (T=298.15 K), kcal·mol\(^{-1}\);
\(^e\)Relative electronic energy of the formed nucleobase pair, kcal·mol\(^{-1}\).
1. $G \cdot C (WC) \leftrightarrow G \cdot C^* (rw_{WC})$

$G \cdot C (WC)$
$(\Delta G = 0.00 / \Delta E = 0.00 / \mu = 6.22)$

$G \cdot C^* (rw_{WC})$
$(\Delta G = 29.86 / \Delta E = 29.77 / \mu = 10.23)$

2. $G^* \cdot C^*(WC) \leftrightarrow G^* \cdot C^*(rw_{WC})$

$G^* \cdot C^*(WC)$
$(\Delta G = 0.00 / \Delta E = 0.00 / \mu = 6.09)$

$G^* \cdot C^*(rw_{WC})$
$(\Delta G = 25.67 / \Delta E = 28.73 / \mu = 5.00)$

3. $G^* \cdot C^*(WC) \leftrightarrow G^* \cdot N_2 \cdot C^*(rw_{WC})$

$G^* \cdot C^*(WC)$

$G^* \cdot C^*(rw_{WC})$

$G^* \cdot N_2 \cdot C^*(rw_{WC})$
4. \( G^* \cdot C^* (r_{WC}) \leftrightarrow G^* \cdot C(w_{WC}) \)

\[
\begin{align*}
(\Delta G = 0.00 / \Delta E = 0.00 / \mu = 6.09) \\
(\nu_i = 447.7i)
\end{align*}
\]

\[
\begin{align*}
(\Delta G = 62.00 / \Delta E = 64.07 / \mu = 11.51)
\end{align*}
\]

\[
(\Delta G = 27.94 / \Delta E = 29.24 / \mu = 4.00)
\]

5. \( G \cdot C^* O_2 (r_{WC}) \leftrightarrow G \cdot C^* (w_{WC}) \)

\[
\begin{align*}
(\Delta G = 0.00 / \Delta E = 0.00 / \mu = 5.43)
\end{align*}
\]

\[
\begin{align*}
(\Delta G = 30.67 / \Delta E = 33.10 / \mu = 8.14)
\end{align*}
\]

\[
(\Delta G = -0.74 / \Delta E = 0.03 / \mu = 7.29)
\]

6. \( G^t \cdot C^* (H) \leftrightarrow G^t \cdot C(r_{wH}) \)

\[
\begin{align*}
(\Delta G = 0.00 / \Delta E = 0.00 / \mu = 6.54)
\end{align*}
\]
7. $ G^* \cdot C^*(H) \leftrightarrow G^* \cdot C^* \cdot O_2 (w_H)$

$$ \Delta G = 0.00 \quad \Delta E = 0.00 \quad \mu = 5.16 $$

$$ \nu_i = 187.4i $$

$$ \Delta G = 27.70 \quad \Delta E = 28.77 \quad \mu = 7.00 $$

$$ \Delta G = 2.29 \quad \Delta E = 3.04 \quad \mu = 8.25 $$
\[
\begin{align*}
G^*\cdot C^*(H) & \quad (\Delta G=0.00 / \Delta E=0.00 / \mu=5.16) \\
G^*\cdot C^*(rH) & \quad (\Delta G=3.56 / \Delta E=4.72 / \mu=4.29)
\end{align*}
\]

\[
\begin{align*}
G^*\cdot C^*(H) & \quad (\Delta G=0.00 / \Delta E=0.00 / \mu=7.32) \\
G^*\cdot C^*(wH) & \quad (\Delta G=3.56 / \Delta E=4.72 / \mu=4.29)
\end{align*}
\]

\[
\begin{align*}
TS_{G^*\cdot C^*(H)\rightarrow G^*\cdot C^*(wH)} & \quad (\mu=7.32) \\
TS_{G^*\cdot C^*(rH)\rightarrow G^*\cdot C^*(wH)} & \quad (\mu=4.29)
\end{align*}
\]
12. $G^\ddagger_t \cdot C^\ddagger(rH) \leftrightarrow G^\ddagger_t \cdot C^\ddagger(wH)$

$G^\ddagger_t \cdot C^\ddagger(rH)$
$(\Delta G=0.00 / \Delta E=0.00 / \mu=6.14)$

TS$^{G^\ddagger_t \cdot C^\ddagger(rH) \rightarrow G^\ddagger_t \cdot C^\ddagger(wH)}$
$(v_i=304.0i)$
$(\Delta G=38.88 / \Delta E=39.16 / \mu=8.63)$

$G^\ddagger_t \cdot C^\ddagger(wH)$
$(\Delta G=14.35 / \Delta E=15.11 / \mu=9.46)$

13. $G^\ddagger_t \cdot C^\ddagger(rH) \leftrightarrow G^\ddagger_t \cdot C^\ddagger(wH) \leftrightarrow G^\ddagger_t \cdot C^\ddagger(wH)$

$G^\ddagger_t \cdot C^\ddagger(rH)$
$(\Delta G=0.00 / \Delta E=0.00 / \mu=6.14)$

TS$^{G^\ddagger_t \cdot C^\ddagger(rH) \rightarrow G^\ddagger_t \cdot C^\ddagger(wH)}$
$(v_i=303.8i)$
$(\Delta G=46.77 / \Delta E=49.74 / \mu=13.64)$

$G^\ddagger_t \cdot C^\ddagger(wH)$
$(\Delta G=2.43 / \Delta E=4.56 / \mu=4.66)$

$G^\ddagger_t \cdot C^\ddagger(wH)$
$(\Delta G=8.98 / \Delta E=10.17 / \mu=4.81)$
14. \(G^*_{N7}\cdot C^*(rw_H) \leftrightarrow G^*_{N7}\cdot C^*(rw_H)_{perp} \leftrightarrow G^*\cdot C^+(w_H) \leftrightarrow G^*\cdot C(rw_H)\)
Fig. 1. Reaction pathways of the tautomerically-conformational transformations of the biologically important G·C nucleobase pairs via the quasi-orthogonal TSs as tight G$^+$·C$^-$ / G$^-$$^\cdot$C$^+$ ion pairs, obtained at the MP2/6-311++G(2df,pd)//B3LYP/6-311++G(d,p) level of QM theory in the isolated state ($\nu_i$ – imaginary frequency in cm$^{-1}$, $\Delta G$ – relative Gibbs free energy in kcal·mol$^{-1}$, $\Delta E$ – electronic energy in kcal·mol$^{-1}$). Intermolecular $\text{AH} \cdots \text{B}$ H-bonds and $\text{A} \cdots \text{B}$ attractive van der Waals contacts are designated by the dotted lines, their lengths $\text{H} \cdots \text{B}$ and $\text{A} \cdots \text{B}$ are presented in angstroms.