This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
A novel conception for spontaneous transversions caused by homo-pyrimidine DNA mismatches: A QM/QTAIM highlight

Ol’ha O. Brovarets* a,b & Dmytro M. Hovorun a,b,□

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. E-mail: dhovorun@imbg.org.ua

Abstract. We have firstly shown that the T·T(w) and C·C(w) DNA mismatches with wobble (w) geometry stay in the slow tautomeric equilibrium with short T·T*(WC) and C·C*(WC) Watson-Crick (WC) mispairs. These non-dissociative tautomeric rearrangements are controlled by the plane-symmetric, highly stable, highly polar and zwitterionic transition states. Obtained results allow us to understand, in what way the T·T(w) and C·C(w) mismatches acquire enzymatically competent T·T*(WC) and C·C*(WC) conformations directly in the hydrophobic recognition pocket of the high-fidelity DNA-polymerase, thereby producing thermodynamically non-equilibrium spontaneous transversions. The simplest numerical estimation of the frequency ratio of the TT to CC spontaneous transversions satisfactorily agrees with experimental data.

Keywords: DNA biosynthesis · Incorporation error · Spontaneous transversion · Wobble conformation · Watson-Crick mispair · Homo-pyrimidine mismatch · Tautomeric transition

† Electronic supplementary information (ESI) available: (i) Computational details; (ii) Physico-chemical parameters of the specific intermolecular contacts; (iii) Energetic and kinetic characteristics of the mispairs tautomerisation; (iv) Interaction energies for the investigated DNA mispairs; (v) Structures of the stationary points of the tautomeric conversions; (vi) Profiles of the glycosidic parameters of the mispairs along the IRC of the tautomerisation; (vii) Ranges of the existence of the obtained patterns of the specific intermolecular interactions along the IRC of the tautomerisation. See DOI: 10.1039/x0xx00000x.
**Introduction.** An intriguing theme of the origin of the spontaneous point mutations has excited researches’ mind during several decades suggesting DNA mispairs as their source [1-6]. Characterizing current state of this biologically important area of knowledge, it can be certainty stated that one of the stumbling stones in the theory of spontaneous point mutagenesis, which slowly gets on its feet, is understanding of the nature of the spontaneous transversions [7-10] caused by the homo-pyrimidine DNA base mispairs. Thorough analysis of the literature evidences that progress in this biologically important area is constrained by the lack of experimental data, especially of the X-ray crystallographic analysis of the enzymatically competent conformation of the incorrect C·C and T·T nucleobase pairs in the active center of the high-fidelity replicative DNA-polymerase at the site of incorporation [11-13], and, from the other hand, with the absence of new model conceptions and approaches.

The rare tautomer hypothesis of spontaneous point mutagenesis formulated by Watson and Crick [1] has been recently challenged by computational [14,15] and experimental [11,12,16-18] studies demonstrating that displacement of one single proton alters H-bonding scheme of the mispair, making it indistinguishable from the canonical Watson-Crick base pairs, that give impetus to further studies. It is logical to assume within the framework of this hypothesis that exceptionally the C·C*(WC) [19-21] and T·T*(WC) [20-22] base mispairs possessing Watson-Crick-like (WC) geometry and containing rare tautomers [23,24] of the C and T DNA bases (here and below they are marked with an asterisk) can be enzymatically incorporated into the structure of the DNA double helix and so are implicated in the origin of spontaneous transversions. Albeit wobble (w) [25-27] C·C(w) and T·T(w) base mispairs cause mild structural and dynamic distortions in the composition of the DNA double helix [28], we have suggested that the C·C*(WC) [19] and T·T*(WC) [20] mismatches with cis-oriented glycosidic bonds [20,21] are their enzymatically competent configurations causing spontaneous transversions.

In this study we have examined in details all possible tautomerisation ways connecting the homo-pyrimidine T·T(w) and C·C(w) mispairs into the C·C*(WC) and T·T*(WC) DNA mismatches and occurring without breakage of the mutual cis-orientation of their glycosidic bonds. Formed C·C*(WC) and T·T*(WC) mismatches are able to accommodate into the hydrophobic recognition pocket of the high-fidelity replicative DNA-polymerase and thus to be responsible for the origin of transversion mutations.

All geometric, energetic and vibrational calculations of the considered base mispairs and transition states (TSs) of their conversion have been performed by Gaussian’09 package [29] using B3LYP [30] and MP2 [31] levels of theory combined with a wide variety of basis sets. Bader's quantum theory of Atoms in Molecules was applied to analyse the electron density distribution [32-34]. IRC calculations have been performed to further confirm the proposed mechanisms of conversion [35] and to obtain profiles of the energetic and geometric characteristics of the base pairs and H-bonds in them along the reaction pathway [36-38] (for more details see ESI†).
Thus, we have shown for the first time that plane-symmetric ($C_s$) T·T(w) mismatch, that in fact is observed in the DNA double helix under physiological conditions [25-27], stays in slow tautomeric equilibrium with two T·T*(WC) [20] and T*O2·T(WC) mispairs representing itself propeller-like and structurally non-rigid complexes [39,40] (Figs. 1, S1 and Tables 1, S1, S4). Notably, the easiness of their acquisition of the plane-symmetric architecture ($\Delta \Delta G_{TS}=1.68$ and $1.54 \text{ kcal} \cdot \text{mol}^{-1}$, respectively, T=298.15 K [39]) enables them to be effectively incorporated into the structure of the DNA double helix.

Both T·T(w)↔T·T*(WC) and T·T(w)↔T*O2·T(WC) tautomerisation processes are controlled by the plane-symmetric, highly polar, highly stable and zwitterionic TS$^{+}$·T·T(w)↔T·T*(WC) and TS$^{+}$·T·T(w)↔T*O2·T(WC) transition states, respectively (Figs. 1, S1 and Tables S1-S5). Their structures are stabilized by strong electrostatic interactions ($\sim 100 \text{ kcal} \cdot \text{mol}^{-1}$) and four intermolecular H-bonds [41], whose contribution to the interaction energy is not decisive (Tables S1, S3). High stability of these TSs excludes direct participation in these tautomerisation processes of the endogenous water molecules as active participants of these events [42] and allows to consider the influence of the stacking interactions with neighboring Watson-Crick base pairs [5,6,43,44] on the course of these tautomerisation reactions as slight perturbation. It is important that both T·T(w)↔T·T*(WC) and T·T(w)↔T*O2·T(WC) tautomerisation processes occur by the non-dissociative mechanism (Figs. S2, S3) and are controlled by the 12 and 10 unique patterns of the specific intermolecular contacts, respectively, that successively replace each other along the IRC of tautomerisation (Fig. 2 and Tables S1, S5).

Another important finding is that the T·T*O2(WC) mispair is in the rapid tautomeric equilibrium with the T*O2·T(WC) mismatch ($\tau_{99.9\%}=3.88 \cdot 10^{10} \text{ s}$), that is implemented by the mechanism of the concerted synchronous double proton transfer (DPT) [45,46] along neighboring H-bonds (Figs. 1b, S1 and Tables 1, S1-S4).

At the same time, the DPT tautomerisation of the T·T(w) DNA mismatch into the T·T*O2(w) mispair with $C_s$ symmetry actually does not happen, since the terminal T·T*O2(w) mismatch is short-lived ($\tau=5.61 \cdot 10^{-15} \text{ s}$), dynamically unstable structure, for which low-frequency intermolecular vibrations can’t develop (Figs. 1b, S1 and Tables 1, S1-S4).

By comparison of the energetic and kinetic characteristics for the T·T(w)↔T·T*(WC) and T·T(w)↔T*O2(WC) tautomerisation processes, we have established that exactly first of them is responsible for the acquisition by the T·T(w) mispair of the enzymatically competent T·T*(WC) conformation in the hydrophobic recognition pocket of the high-fidelity replicative DNA-polymerase, that is a necessary and sufficient condition for the occurrence of the corresponding spontaneous transversions [11-13].

In the isolated state the T·T(w) mispair has a pseudo-two-fold symmetry. Calculations show that the transition between these pseudo-symmetrical states is provided by the non-dissociative mechanism
of the tautomeric nature (Fig. 3): exactly it, in our opinion, adequately explains existing experimental data [47] that have been left without any microstructural interpretation for a long period of time.

We have also established for the first time (Figs. 4, S4 and Tables 1, S1-S3, S6) that propeller-like and non-rigid C·C*(WC) mismatch with C\(_{\text{I}}\) symmetry [19] slowly tautomerises into the C·C(w) (ā\(\angle\)C4N3(C)N3C4(C*)=9.0\(^o\)) and C·C*(w) (C\(_{\text{I}}\)) mispairs through the highly polar, highly stable and zwitterionic \(\text{TS}^{\text{C-C*}}_{\text{C}}\rightarrow\text{C·C*(WC)}\) (ā\(\angle\)C4N3(C)N3C4(C*)=4.3\(^o\)) and \(\text{TS}^{\text{C·C*(WC)}\rightarrow\text{C·C*(w)}\right\}}\) (C\(_{\text{I}}\)) transition states, respectively. These TSs are stabilized by strong electrostatic interactions (~ 100 kcal·mol\(^{-1}\) (Table S3)) involving intermolecular H-bonds of medium strength that, similarly to the transformations of the T·T(w) mispair, excludes direct participation of the water molecules in these tautomerisation processes and indicates the negligible impact of the \(\pi\)-\(\pi\) stacking interactions [5,43,44] and sequence [6] on the considered tautomerisation processes.

Our results show that the C·C(w)↔C·C*(WC) and C·C*(WC)↔C*·C*(w) reactions occur by the non-dissociative mechanism (Figs. S5, S6) and are accompanied by 11 and 10 unique patterns of the specific intermolecular contacts including H-bonds, attractive van der Waals contacts and loosened covalent bridges, respectively, that consistently change each other along the IRC of tautomerisation (Fig. 5 and Tables S1, S7).

Moreover, we have also tested the C*·C*(w) mispair for the possibility of its tautomerisation via the DPT along intermolecular H-bonds according to Löwdin’s mechanism [48,49]. Consequently, it was established that the C*·C*(w)↔C*\(_{\text{O2}}\)·C(w) DPT tautomerisation does not occur, since the C*\(_{\text{O2}}\)·C(w) mismatch is dynamically unstable structure (\(\tau=2.45\times10^{-13}\) s), for which low-frequency intermolecular vibrations can not develop (Figs. 4, S4 and Tables 1, S6).

Biological importance of the C·C(w)→C·C*(WC) tautomeric conversion relies in the fact that it is kinetically controlled pathway for the formation directly in the essentially hydrophobic recognition pocket of the high-fidelity DNA-polymerase of the enzymatically competent C·C*(WC) mispair responsible for the occurrence of spontaneous point CC incorporation errors.

Our results raise at least two interesting issues representing a challenge for the experiment. Firstly, – in what proportions coexist in double helical DNA highly polar C·C(w), C·C*(WC), which in the free state is a global minimum, and plane symmetric C*\(_{\text{O2}}\)·C(w) mispairs? Secondly, since all these three mispairs have pseudo-two-fold symmetry, so the question arises about the mechanisms of their mutual transformation in the composition of the double helical DNA. Calculations indicate that these processes proceed via the non-dissociative mechanisms of the tautomeric origin (Fig. 6).

Our data are in a good agreement with other theoretical calculations [28]. Thus, fixation in the DNA duplex of the C·C(w) mispair, stabilized by the N4H···N3 H-bond and N3···O2 van der Waals contact, and the T·T(w) mismatch, joined by the N3H···O4 and N3H···O2 H-bonds (Table S1), can be attributed to the fact that they are transformed into the C·C*(WC) and T·T*(WC) mismatches, respectively, through the tautomeric rearrangement accompanied with structural rebuilding of the base.
mispairs (Figs. 3 and 6). This finding can’t be reflected by the MD calculations operating solely with the canonical tautomers [28].

Obtained data shed light on the microstructural mechanisms of the occurrence of the spontaneous transversions caused by the homo-pyrimidine DNA mismatches and recognition of the latest by the reparation systems [50-52].

Numerical evaluation of the probability ratio of the acquisition of the enzymatically competent conformation by the T·T(w) mispair relatively to the analogical probability for the C·C(w) mismatch (2.3·10²) is consistent with experimental data showing values of the extending efficiencies that are higher by several orders for the T·T, than for the C·C transversion mispairs [53-55], and also significantly lower frequencies of spontaneous transversions in comparison with spontaneous transitions [56,57]. At this we have applied known formulas of physico-chemical kinetics [58,59]: thus, we have suggested that

\[ k_f = 1.09 \cdot 10^{-10} / k_r = 4.19 \cdot 10^{-4} \text{ s}^{-1} \quad (T\cdotT(w) \leftrightarrow T\cdotT*(WC)), \]
\[ k_f = 1.59 \cdot 10^{-6} / k_r = 4.71 \cdot 10^{-13} \text{ s}^{-1} \quad (C\cdotC(w) \leftrightarrow C\cdotC*(WC)) \]

and high-fidelity DNA-polymerase spends time \( \Delta t_{\text{wrong pol}} = 8.3 \cdot 10^{-3} \text{ s} \) [60,61] for the incorporation of one incoming incorrect nucleotide. Notably, we do not compare here absolute values of the mentioned above probabilities with corresponding literature data, since we are convinced that for their correct estimation it should be necessarily taken into account quantum tunneling of protons [62], which significantly accelerates tautomerisation processes.

Our findings also add substantially to the understanding, why the pyrimidine-pyrimidine T·T(w) mismatches are much better deleted from the genome by the reparation systems “sharpened” for the wobble architecture than the C·C(w) mispairs (see works [28, 63] and refs. therein): the point is that the T·T(w)↔T·T*(WC) tautomeric equilibrium is much more shifted to the left than the C·C(w)↔C·C*(WC) tautomeric balance.

Acknowledgments. O.O.B. was supported by the Grant of the President of Ukraine to support scientific research of young scientists for 2015 year from the State Fund for Fundamental Research of Ukraine (project № GP/F61/028) and by the grant of the NAS of Ukraine for the realization of the scientific and research work of the young scientists of the NAS of Ukraine for 2015-2016 years. 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. The authors sincerely thank Dr. Ivan S. Voiteshenko (Institute of High Technologies, Taras Shevchenko National University of Kyiv) and Dr. Fernando R. Clemente (Gaussian, Inc.) for their technical support of the work.

References.


Fig. 1. Stationary structures on the reaction pathways of the (a) T·T(w)↔T·T*(WC)/T*·O2·T(WC)/T*·T*·O2(w) and (b) T*·O2·T(w)↔T·T*·O2(WC) tautomerisations via the sequential DPT accompanied by the structural rearrangement of the base pair obtained at the B3LYP/6-311++G(d,p) level of theory. ΔE - electronic energy of the base mispairs and TSs. Dotted lines indicate AH···B H-bonds and attractive A···B van der Waals contacts, while continuous – covalent bonds; carbon atoms are in light-blue, nitrogen – in dark-blue, hydrogen – in grey and oxygen – in red.
Fig. 2. Exchange of the patterns of the intermolecular AH⋯B H-bonds and attractive O/N⋯O van der Waals contacts (their energies $E_{AH⋯B/O/N⋯O}$ are estimated by the EML formula at the (3,-1) BCPs) along the IRC of the biologically important (a) T·T(w)↔T·T*(WC) and (b) T·T(w)↔T*$_{O2}$·T(WC) tautomerisations via the sequential DPT accompanied by the structural rearrangement of the base pair obtained at the B3LYP/6-311++G(d,p) level of theory (see Figs. 1, S1 and Tables S1, S5).

Fig. 3. Interconversion of the pseudo-twofold symmetrical T·T(w) DNA mismatches via the tautomeric transition. Presented isoenergetic structures are linked with each other by the pseudo-dyadic axis of symmetry. For the designations see Fig. 1.
Table 1. Energetic and kinetic characteristics of the $T\cdot T(w)\leftrightarrow T\cdot T^*(WC)$, $T\cdot T(w)\leftrightarrow T^*\cdot T^*_{O2}(WC)$, $T^*_{O2}\cdot T(w)\leftrightarrow T^*\cdot T^*_{O2}(WC)$, $T\cdot T(w)\leftrightarrow T^*\cdot T^*_{O2}(w)$, $C\cdot C(w)\leftrightarrow C\cdot C^*(WC)$, $C\cdot C^*(WC)\leftrightarrow C^*\cdot C^*(w)$ and $C^*\cdot C^*(w)\leftrightarrow C^*\cdot T_{O2}\cdot C(w)$ tautomerisations via the sequential DPT accompanied by the structural rearrangement of the base pair obtained at the MP2/cc-pVQZ//B3LYP/6-311++G(d,p) level of theory (see also Figs. 1, 4, S1 and S4).

<table>
<thead>
<tr>
<th>Tautomerisation</th>
<th>$\Delta G^a$</th>
<th>$\Delta E^b$</th>
<th>$\Delta G^c_{TS}$</th>
<th>$\Delta E^d_{TS}$</th>
<th>$\Delta G^e$</th>
<th>$\Delta E^f$</th>
<th>$\tau_{99.9%}^g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T\cdot T(w)\leftrightarrow T\cdot T^*(WC)$</td>
<td>8.98</td>
<td>8.64</td>
<td>31.06</td>
<td>31.90</td>
<td>22.09</td>
<td>23.26</td>
<td>1.65·$10^4$</td>
</tr>
<tr>
<td>$T\cdot T(w)\leftrightarrow T^<em>\cdot T^</em>_{O2}(WC)$</td>
<td>12.72</td>
<td>12.81</td>
<td>32.82</td>
<td>33.96</td>
<td>20.11</td>
<td>21.15</td>
<td>5.82·$10^2$</td>
</tr>
<tr>
<td>$T^<em>_{O2}\cdot T(w)\leftrightarrow T\cdot T^</em>_{O2}(WC)$</td>
<td>0.00</td>
<td>0.00</td>
<td>4.40</td>
<td>7.56</td>
<td>4.40</td>
<td>7.56</td>
<td>3.88·$10^{-10}$</td>
</tr>
<tr>
<td>$T\cdot T(w)\leftrightarrow T^<em>\cdot T^</em>_{O2}(w)$</td>
<td>15.61</td>
<td>15.99</td>
<td>13.82</td>
<td>17.03</td>
<td>-1.79</td>
<td>1.03</td>
<td>3.88·$10^{-14}$</td>
</tr>
<tr>
<td>$C\cdot C(w)\leftrightarrow C\cdot C^*(WC)$</td>
<td>-8.90</td>
<td>-10.73</td>
<td>25.38</td>
<td>24.32</td>
<td>34.28</td>
<td>35.05</td>
<td>4.35·$10^9$</td>
</tr>
<tr>
<td>$C\cdot C^<em>(WC)\leftrightarrow C^</em>\cdot C^*(w)$</td>
<td>1.34</td>
<td>1.50</td>
<td>18.34</td>
<td>19.44</td>
<td>17.00</td>
<td>17.94</td>
<td>2.90</td>
</tr>
<tr>
<td>$C^<em>\cdot C^</em>(w)\leftrightarrow C^*\cdot T_{O2}\cdot C(w)$</td>
<td>6.07</td>
<td>6.36</td>
<td>6.59</td>
<td>9.27</td>
<td>0.53</td>
<td>2.91</td>
<td>1.69·$10^{-12}$</td>
</tr>
</tbody>
</table>

$^a$The Gibbs free energy of the product relatively the reactant of the tautomerisation reaction (T=298.15 K), kcal·mol$^{-1}$

$^b$The electronic energy of the product relatively the reactant of the tautomerisation reaction, kcal·mol$^{-1}$

$^c$The Gibbs free energy barrier for the forward reaction of tautomerisation, kcal·mol$^{-1}$

$^d$The electronic energy barrier for the forward reaction of tautomerisation, kcal·mol$^{-1}$

$^e$The Gibbs free energy barrier for the reverse reaction of tautomerisation, kcal·mol$^{-1}$

$^f$The electronic energy barrier for the reverse reaction of tautomerisation, kcal·mol$^{-1}$

$^g$The time necessary to reach 99.9% of the equilibrium concentration between the reactant and the product of the tautomerisation reaction, s

See also summary Table S2 for the Gibbs and electronic energies of the mispairs and TSs relatively the global minima – the $T\cdot T(w)$ and $C\cdot C^*(WC)$ mispairs.
Fig. 4. Stationary structures on the reaction pathways of the (a) C·C(w)↔C·C*(WC), (b) C·C*(WC)↔C*·C*(w) and (c) C*·C*(w)↔C*O2·C(w) tautomerisations via the sequential DPT accompanied by the structural rearrangement of the base pair obtained at the B3LYP/6-311++G(d,p) level of theory. For the designations see Fig. 1.
Fig. 5. Exchange of the patterns of the intermolecular AH···B, AH···HB H-bonds and attractive O/N···O van der Waals contacts (their energies $E_{AH···B/AH···HB/O/N···O}$ are estimated by the EML formula at the (3,-1) BCPs) along the IRC of the biologically important (a) C·C(w)↔C·C*(WC) and (b) C·C*(WC)↔C·C*(w) tautomerisations via the sequential DPT accompanied by the structural rearrangement of the base pair obtained at the B3LYP/6-311++G(d,p) level of theory (see Figs. 4, S4 and Tables S1, S7).

Fig. 6. Interconversion of the pseudo-twofold symmetrical (a) C·C(w) and (b) C·C*(w) DNA mismatches via the tautomeric transition. Presented isoenergetic structures are linked with each other by the pseudo-dyadic axis of symmetry. For the designations see Fig. 1.