A Spontaneous Ring Opening Reaction Leads to a Repair-Resistant T Oxidation Product in Genomic DNA**

Aleksandr B. Sahakyan,[#] Areeb Mahtey,[#] Fumiko Kawasaki and Shankar Balasubramanian*

Abstract: The alphabet of modified DNA bases goes beyond the conventional four letters, with biological roles being found for many such modifications. Here we describe a novel observation for the thymine base that arises *via* spontaneous N_1-C_2 ring opening of its oxidation product 5-formyl uracil, after N_3-deprotonation. We first observed this phenomenon *in silico* though ab initio calculations, followed by *in vitro* experiments to verify its formation at a mononucleoside level and in a synthetic DNA oligonucleotide context. We show that the new base modification (T^{rex}, **T**hymine **r**ing **ex**punged) can form under physiological conditions, and is resistant to the action of common repair machineries. Furthermore, we found cases of natural existence of T^{rex} *in vivo*, while screening a number of human cell types and mESC (E14), suggesting potential biological relevance of this modification.

Beyond the four, A, T, G and C, canonical bases, DNA has been shown to contain a plethora of natural chemical modifications, many with important biological consequences.[1-3] The most widely studied chemical modifications are those arising through cytosine methylation and its stepwise oxidation, which forms part of the active demethylation pathway in mammalian cells.[3] Other DNA base modifications, such as 8-oxoguanine (8-oxoG), arise from the reaction of DNA with reactive oxygen species (ROS).[4] 8-oxoG in promoter sites were found to significantly reduce transcription of reporter genes due to 8-oxoG excision,[5,6] suggesting that even oxidative damage marks may have a degree of "epigenetic" character, affecting gene regulation. As well as being found in prokaryotic genomes, the oxidized T analogues, 5-hydroxymethyluracil (5hmU) and 5-formyl uracil (5fU), have been detected at a level of 0.5-5 per 10^6 bases in mammalian genomes.[7] Both of these modifications are often considered to be products of oxidative damage of T by ROS-mediated oxidation of the C_5-methyl group.[4] Due to the propensity of 5fU to exist in its enol tautomer, LC-MS/MS changes in its levels in differentiating mESCs have been shown to cluster with the oxidative damage marker 8-oxoG, highlighting its potential epigenetic nature.[8-10] As such, if this oxidized T variant is inefficiently repaired, it may contribute to T→C transitions and associated mutational burden in genomes. *In vitro* studies have also shown that 5fU can impede transcription factor binding.[11,12] However, these oxidative modifications of T are known to be excised and repaired from the mammalian genome *via* the base-excision repair pathway.[13-15]

In this study, we start from a serendipitous *in silico* finding of a novel ring-opening reaction of 5fU, which can lead to an alternative oxidized variant of T, denoted as T^{rex} (**T** **r**ing **ex**punged) hereafter. Our *in vitro* experiments indicate that T^{rex} does indeed form in near-physiological conditions. Moreover, we show that the modification is resistant to the common repair machineries, and demonstrate the natural existence of T^{rex} *in vivo*, using cell-based assays.

While calculating deprotonated states of DNA bases and their common epigenetic modifications (Note S1), we made an unexpected observation, whereby the deprotonated state of one particular 5fU configuration was never observed (Mg^{2+}•$5fU^{anti}_{a}$ in Figure 1A and Figure S1). Instead, upon the *in silico* deprotonation, the uracil ring in 5fU immediately opened through the breakage of the N_1-C_2 bond, reaching a ring-opened stationary intermediate. Even though the original calculations were done at a relatively lower, RHF/6-31G, level of theory, the same observation stayed true while recalculating the phenomenon with Møller-Plesset[16] perturbation theory at MP2[17] level, using a larger 6-311++G(d,p) basis set,[18] and with zero point vibrational energy, rotation-translational, enthalpic and entropic (298.15 K) corrections. When forcing a ring-closed state upon N_3 deprotonation, we observed that the ring-opened state has significantly lower free energy ($\Delta G_{(g)}{}^{o}{}_{ring\ opening}$ = -25.24 kcal/mol), with the ring opening thus happening spontaneously upon deprotonation, without an extra barrier.

The uracil ring was long known to be unstable, also exerting the least aromaticity out of all nucleic acid bases.[19,20] However, the major body of experimental and computational work has mainly focused on the behavior of uracil under harsh fragmentation and ionizing conditions in mass spectrometry or of astrochemical relevance.[21-24] 5fU was noted to undergo ring opening, but through the cleavage at N_1-C_6 bond, while reacting with primary alkyl amines,[25] and enamines.[26] For anionic U^-, a density functional theory investigation of enforced fragmentation

[*] Dr. A. B. Sahakyan,[#] A. Mahtey,[#] Dr. F. Kawasaki, Prof. S. Balasubramanian*
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW (UK)
E-mail: sb10031@cam.ac.uk

Dr. A. B. Sahakyan
Present address: MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DS (UK)

Dr. F. Kawasaki
Present address: RIKEN, Center for Advanced Intelligence Project, Tokyo 103-0027 (Japan)

Prof. S. Balasubramanian
Cancer Research UK, Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Robinson Way, Cambridge CB2 0RE (UK), and School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP (UK)

[#] These authors contributed equally to this work.

[**] A.M. is supported by The University of Cambridge, Vice Chancellor's Award (Cambridge Trust) and the Dudding & Stachulski Scholarship. The Balasubramanian group is supported by programme grant funding (C9681/A18618) and core-funding (C14303/A17197) from Cancer Research UK and by a Wellcome Trust Senior Investigator Award (209441/z/17/z).

Supporting information for this article can be found under: http://dx.doi.org/xxxxxxxxx

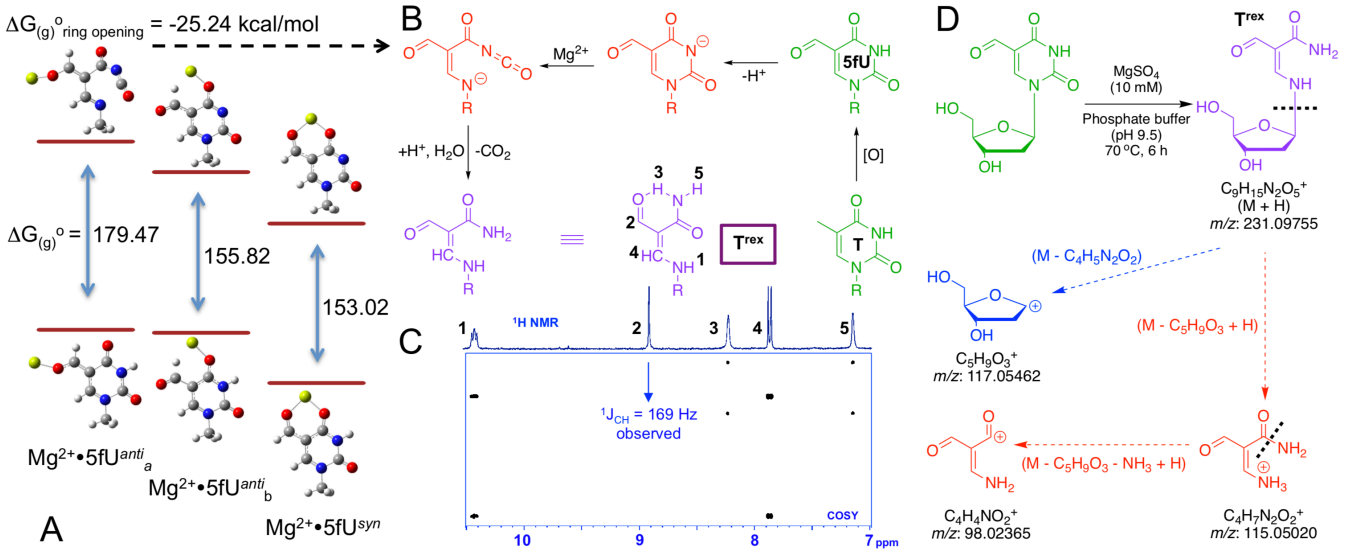


Figure 1. *In silico* observation of a spontaneous N_1-C_2 opening of 5fU ring upon deprotonation (A). The overall free energies of deprotonation are shown in gas phase. The free energy difference upon only the ring opening ($\Delta G_{(g)}^{\circ}{}_{\text{ring opening}}$) is shown for the configuration where spontaneous ring opening is observed. The proposed scheme for further modification towards the stable compound (T^{rex}, in purple) (B), where the part that represents a single step process revealed *in silico* is colored in red. Some elements of NMR characterization of T^{rex} (C). Generation of T^{rex} synthetic standard, along with key fragment ions detected by HPLC-MS/MS (D). The substituent denoted as R in (B) is methyl for the calculations illustrated in (A), and is 2'-deoxyriboside in the experiments reflected in (D) and (C).

revealed the relative strengths of N_3-C_4 over N_1-C_2 over C_5-C_6 bonds with 1.622, 1.710 and 5.459 eV (1 eV = 23.061 kcal/mol) barriers for cleavage.[21] Cole et al.[24] reported a 20.8 kcal/mol energy barrier for N_1-C_2 cleavage in N_3-deprotonated U^-. Our finding is therefore rather unique, showing that there is a barrier-less pathway ($\Delta G_{(g)}^{\circ}{}_{\text{ring opening}}$ = -25.24 kcal/mol) for the N_1-C_2 ring opening in $5fU^-$. Interestingly, this is observed only for 5fU (Figure S1), and only when the formyl group is flipped to an *anti* conformation with an additional weak interaction (Mg^{2+} in our calculations) to increase the electron-withdrawing properties of 5-formyl substituent. Furthermore, the overall deprotonation barriers are significantly lower for the examined 5fU configurations (Figure 1A and Figure S1, note the values are in gas phase, and $\Delta G_{(solv)}^{\circ}(H^+)$ in water is -265.9 kcal/mol[27]), contributing to the relatively low pK_a (8.2-8.6) at N_3 position in 5fU.[28] It would be interesting to check in future whether such ring opening can also happen in analogous synthetic bases of biomedical relevance, such as in 5-fluorouracil with electron withdrawing substituent.

To experimentally test the possibility of T^{rex} formation, we first tried a strongly basic condition to facilitate 5fU deprotonation, in the presence of Mg^{2+}. This resulted in the formation of the expected entity, 3-amino-2-formylacrylamide-2'-deoxyriboside (T^{rex} nucleoside, Figure 1B-D). The ring-opened species formed at a yield of ~26%, along with 5fU nucleobase (28%), trace amounts of C_5-C_6 hydrated 5fU (5fU+H_2O), and unreacted 5fU (46%). These non-physiological conditions were used to generate an isolatable quantity of T^{rex}, allowing for its full characterization (Figure 1C and Note S2) and use as a synthetic standard (Figure 1D) for NanoHPLC-MS/MS measurements on genomic DNA.

HRMS in positive ion mode, corroborates the molecular mass of *m/z* 253.0785 to the molecular formula $C_9H_{14}N_2O_5{}^{23}Na^+$ $[M+Na]^+$ (253.0795 calculated), with major fragment ions, shown in Figure 1D. NMR characterization of T^{rex} are shown in Note S2 and Figure 1C. These data collectively support the structure shown in Figure 1B and D, which forms *via* N_1-C_2 cleavage, to give an isocyanate, that undergoes rapid hydrolysis and decarboxylation to produce T^{rex} (Scheme S2 and Figure 1B). The experimental findings thus support those revealed theoretically, and the versatile donor/acceptor interfaces of T^{rex} suggest it may have mutagenic properties (Figure S2).

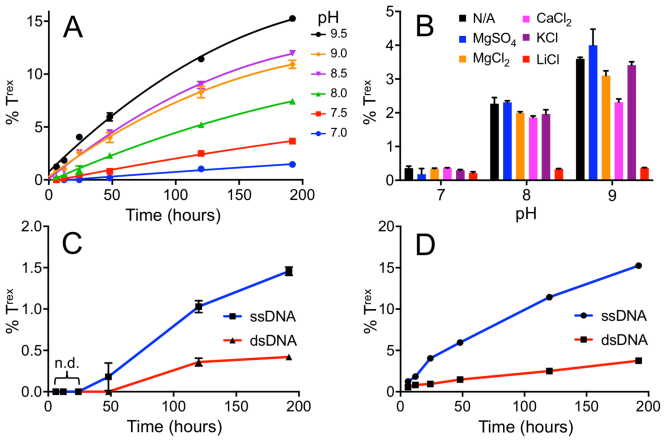


Figure 2. Screening various ring-opening conditions on 5fU-ODN (10 mer, see Note S3). All measurements are an average of experimental triplicates. A time-course study highlighting the kinetics of 5fU ring opening at varying pH (A). Mono- and divalent metal ion screen, in addition to the Na^+ present in the phosphate buffer (B). A time-course study at pH 7.0, showing the reduced formation of T^{rex} in a dsDNA context (C). Timepoints below 48 h were unquantifiable (n.d., below detection limit of the LC-MS) (C). A time-course study at pH 9.5 (D), showing a similar trend to that at pH 7.0.

Using a decameric oligonucleotide (ODN) containing a central, single 5fU site (Tables S1 and S2), we also demonstrate T^{rex} formation in a DNA context. Incubation of this ODN with $MgSO_4$ and phosphate buffer (pH 7.0-9.5) at 37°C over 8 days revealed that 5fU to T^{rex} conversion can be achieved at physiological conditions (Figure 2A). Interestingly, the slight shift in order of T^{rex} production for pH 8.5 and 9.0 in Figure 2A is associated with the competing production of 5fU+H_2O, reaching its maximum at pH 9.0 (Figure S3, Table S3). The formation of T^{rex} was validated through digestion of the corresponding ODN and cross-referencing against synthetic standards by LC-MS (Figure S4). It became apparent that T^{rex} formation had a stronger dependence on pH, than the type of mono- or divalent cation additionally supplemented into the Na^+-rich buffer (Figure 2B). This indicates that water molecules or Na^+ ions may be enough for the destabilization of the $5fU^-$ anion.

We next explored its formation in dsDNA, to more closely model the biological context. The decameric 5fU-ODN was annealed with its complementary strand, tagged with a 5'-phosphate to enhance its LC-MS separation (Figure S5 and S6). This duplex was then subjected to ring-opening conditions at pH 7.0 and 9.5. The yield of T^{rex} showed that the conversion efficiency was lower in dsDNA than ssDNA (Figure 2C and D), which is suggestive of a protective environment that duplex DNA provides, similar to its previously observed influence on cytosine deamination.[29]

Having demonstrated that T^{rex} can form spontaneously under physiological conditions, we next determined if T^{rex} can be excised from DNA, by glycosylases known to excise T modifications. A panel of enzymes (hSMUG1, UDG, *Afu*-UDG, T4 PDG, Fpg, Endonuclease III (Nth) and *Tma* Endonuclease III) possessing N-glycosylase and apyrimidinic-lyase activity for ring-opened and oxidized T/G modifications were tested on both 10 and 34 mer ODNs (Note S4, Table S4). All T^{rex} present in the biochemical assay was resistant to repair by these enzymes. hSMUG1 achieved quantitative excision of 5fU leaving an apyrimidinic site (AP-site), whilst keeping T^{rex} and 5fU+H_2O intact (Figure S7). UDG, *Afu*-UDG, Endonuclease III (Nth), T4 PDG and *Tma* Endonuclease III showed no activity on the 34 mer 5fU-ODN (Figure S8). Fpg gave quantitative excision of 5fU, with a 1:1 ratio of AP-ODN and cleaved ODN with no excision observed for T^{rex}. The lack of activity by the Fpg and Endo III (Nth) glycosylases for T^{rex} was interesting, considering they have been reported to remove αRT (ring-opened thymidine) from synthetic DNA.[30] The T^{rex} modification is therefore likely to reach steady-state levels in the genome. Although we have screened a large range of glycosylases, potentially, there still could be an enzyme capable of excising T^{rex} *in vivo*.

The resistance to repair of T^{rex} by common excision machinery led us to investigate the occurrence of this entity in natural genomic DNA. Necessary measures were taken to ensure any T^{rex} detected was not an artifact from sample preparation (Note S5). A digestion time-course was employed to ensure no T^{rex} formation was observed (by NanoHPLC-MS/MS) within the duration of complete digestion (Figure S9). To facilitate the detection, by enhancing the signal-to-noise ratio, digests were subjected to an offline HPLC enrichment for T^{rex} and 5fU (Note S6). We therefore additionally demonstrated that T^{rex} is not formed during lyophilisation (Figure S10).

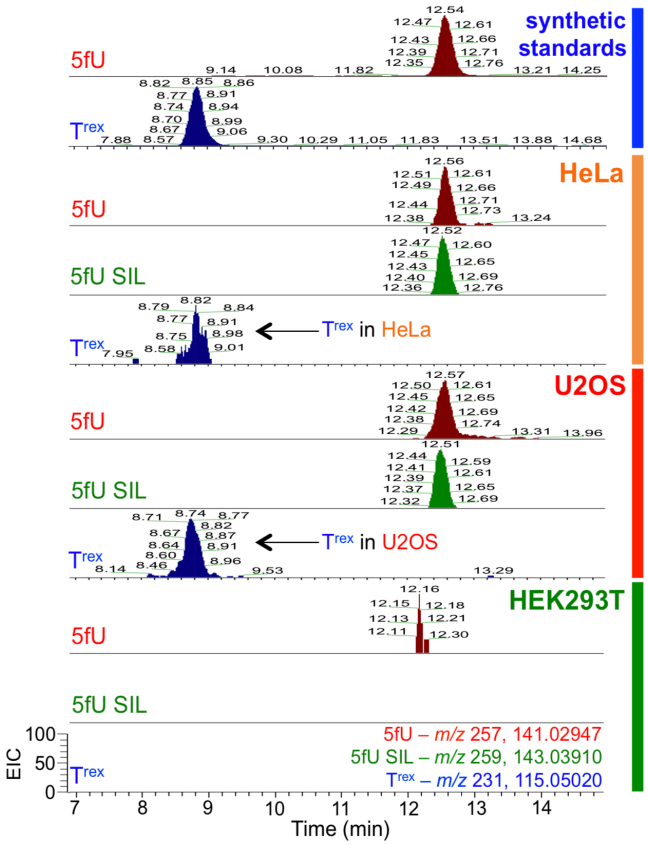


Figure 3. NanoHPLC-MS/MS traces of 5fU, (maroon), 5fU-SIL (green) and T^{rex} (blue) by targeting for *m/z* 141.02947 as a fragment of *m/z* 257; *m/z* 143.03910 as a fragment of *m/z* 259 and *m/z* 115.05020 as a fragment of *m/z* 231 respectively. The right-side vertical colored lines indicate the grouping of synthetic standards (blue), and cell lines HeLa (orange), U2OS (red), and HEK293T (green). mESC data are shown in Note S6. Arrows show the T^{rex} detection signals from HeLa and U2OS. The y axes are the extracted ion counts normalised to 100 for the signal of interest. The common normalized scale is brought for the bottom plot. The absolute counts (x10^3) for {5fU, 5fU SIL, T^{rex}} respectively are {8.00, 8.32, 1.69} for HeLa, {18.7, 9.31, 7.22} for U2OS and {0.69, na, na} for HEK293T.

DNA from HEK293T, HeLa, U2OS and mESCs (E14 strain) was digested in this screen, and we confirmed LC co-elution and identical MS/MS fragmentation as compared to a synthetic standard of T^{rex} (Figure S11). The NanoHPLC-MS/MS analysis confirmed the existence of this modification in the U2OS and HeLa cell lines (Figure 3). Interestingly, the level of ROS are known to be higher in intensively metabolizing cancer cells[31], which may explain the accumulation of the T^{rex} modification in cancer-derived U2OS and HeLa cell lines.

In conclusion, we have discovered new chemistry on 5fU starting with theoretical predictions leading to experimental validation of its N_1-C_2 ring opening phenomenon. The resulting T^{rex} forms under physiological conditions *in vitro* and may be repair resistant. The reaction may be influential at an evolutionary timescale, contributing to the spontaneous mutation rates. Finally, we report the first detection of such a modification in U2OS and HeLa, human cancer-derived cell lines, which hints at the possibility that the modification may accumulate and be of biological relevance.

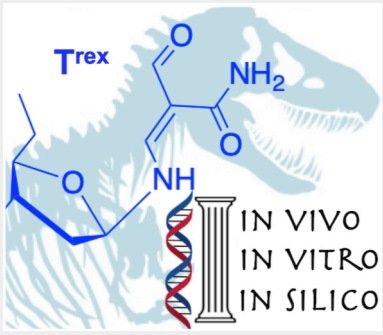
Keywords: base modification • *in vivo* detection • nucleic acids • oxidative damage • repair resistance • ring opening • thymine

[1] A. Breiling, F. Lyko, *Epigen. Chromatin* **2015**, *8*, 24.
[2] E.-A. Raiber, R. Hardisty, P. van Delft, S. Balasubramanian, *Nat. Rev. Chem.* **2017**, *1*, 0069.
[3] T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Angew. Chem. Int. Ed.* **2018**, *57*, 4296–4312.
[4] M. S. Cooke, M. D. Evans, M. Dizdaroglu, J. Lunec, *FASEB J.* **2003**, *17*, 1195–1214.
[5] M. Yoshihara, L. Jiang, S. Akatsuka, M. Suyama, S. Toyokuni, *DNA Res.* **2014**, *21*, 603-612.
[6] J. Allgayer, N. Kitsera, S. Bartelt, B. Epe, A. Khobta, *Nucl. Acids Res.* **2016**, *44*, 7267–7280.
[7] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.
[8] E. J. Privat, L. C. Sowers, *Mutat. Res.* **1996**, *354*, 151–156.
[9] A. Klungland, R. Paulsen, V. Rolseth, Y. Yamada, Y. Ueno, P. Wiik, A. Matsuda, E. Seeberg, S. Bjelland, *Toxicol. Lett.* **2001**, *119*, 71–78.
[10] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truß, J. Steinbacher, B. Hackner, O. Kotljarova, et al., *Nat. Chem. Biol.* **2014**, *10*, 574–581.
[11] A. Kittaka, C. Horii, T. Kuze, T. Asakura, K. Ito, K. T. Nakamura, T. Miyasaka, J.-I. Inoue, *Synlett* **1999**, *S1*, 869–872.
[12] D. K. Rogstad, J. Heo, N. Vaidehi, W. A. Goddard, A. Burdzy, L. C. Sowers, *Biochem.* **2004**, *43*, 5688–5697.
[13] A. Masaoka, M. Matsubara, R. Hasegawa, T. Tanaka, S. Kurisu, H. Terato, Y. Ohyama, N. Karino, A. Matsuda, H. Ide, *Biochem.* **2003**, *42*, 5003–5012.
[14] P. F. Liu, A. Burdzy, L. C. Sowers, *DNA Repair* **2003**, *2*, 199–210.
[15] N. C. Bauer, A. H. Corbett, P. W. Doetsch, *Nucl. Acids Res.* **2015**, *43*, 10083–10101.
[16] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618–622.
[17] M. Head-Gordon, J. A. Pople, M. J. Frisch, *Chem. Phys. Lett.* **1988**, *153*, 503–506.
[18] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650–654.
[19] P. Cysewski, *J. Mol. Struct. Theochem* **2005**, *714*, 29–34.
[20] A. B. Sahakyan, M. Vendruscolo, *J. Phys. Chem. B* **2013**, *117*, 1989–1998.
[21] L. S. Arani, P. Mignon, H. Abdoul-Carime, B. Farizon, M. Farizon, H. Chermette, *Phys. Chem. Chem. Phys.* **2012**, *14*, 9855–9870.
[22] B. Barc, M. Ryszka, J. Spurrell, M. Dampc, P. Limão-Vieira, R. Parajuli, N. J. Mason, S. Eden, *J. Chem. Phys.* **2013**, *139*, 244311.
[23] D. Almeida, M. C. Bacchus-Montabonel, F. F. da Silva, G. García, P. Limão-Vieira, *J. Phys. Chem. A* **2014**, *118*, 6547–6552.
[24] C. A. Cole, Z.-C. Wang, T. P. Snow, V. M. Bierbaum, *Phys. Chem. Chem. Phys.* **2014**, *16*, 17835–17844.
[25] E. Sochacka, D. Smuga, *Tetrahedron Letters* **2007**, *48*, 1363–1367.
[26] H. Singh, Dolly, S. S. Chimni, S. Kumar, *J. Chem. Res. (S)* **1998**, 352–353.
[27] M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe, T. R. Tuttle, *J. Phys. Chem. A* **1998**, *102*, 7787–7794.
[28] M. Yoshida, K. Makino, H. Morita, H. Terato, Y. Ohyama, H. Ide, *Nucl. Acids Res.* **1997**, *25*, 1570–1577.
[29] L. A. Frederico, T. A. Kunkel, B. R. Shaw, *Biochem.* **1990**, *29*, 2532–2537.
[30] J. Jurado, M. Saparbaev, T. J. Matray, M. M. Greenberg, J. Laval, *Biochem.* **1998**, *37*, 7757–7763.
[31] J. C. Puigvert, K. Sanjiv, T. Helleday, *FEBS J.* **2016**, *283*, 232–245.

Entry for the Table of Contents

COMMUNICATION

Prompted by serendipitous *in silico* finding, we describe an alternative oxidised variant of T, denoted as T^{rex} (**T r**ing **ex**punged). Our *in vitro* experiments show that T^{rex} does indeed form in physiological conditions. We show that the modification is resistant to common repair machineries, and demonstrate the natural existence of T^{rex} *in vivo*, in human cancer-derived cell lines.



*Aleksandr B. Sahakyan, Areeb Mahtey, Fumiko Kawasaki, and Shankar Balasubramanian**

Page No. – Page No.

A Spontaneous Ring Opening Reaction Leads to a Repair-Resistant T Oxidation Product in Genomic DNA