

# SUPPORTING INFORMATION

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

Aleksandr B. Sahakyan,<sup>#,†,‡</sup> Areeb Mahtey,<sup>#,†</sup> Fumiko Kawasaki,<sup>†,§</sup> and Shankar Balasubramanian<sup>\*,†,§,†</sup>

<sup>†</sup>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

<sup>§</sup>Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, U.K.

<sup>†</sup>School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, U.K.

<sup>#</sup>These authors contributed equally to this work.

\*Correspondence to sb10031@cam.ac.uk (S.B.).

<sup>‡</sup>Present address: MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DS, U.K.

<sup>§</sup>Present address: RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan.

### CONTENT

General Remarks .....	2
Note S1 .....	3
Note S2 .....	8
Note S3 .....	10
Note S4 .....	16
Note S5 .....	19
Note S6 .....	22
Scheme S1-S2 .....	8
Figure S1 .....	3
Figure S2 .....	9
Figure S3 .....	12
Figure S4 .....	13
Figure S5 .....	14
Figure S6 .....	15
Figure S7 .....	16
Figure S8 .....	17
Figure S9 .....	19-20
Figure S10 .....	21
Figure S11 .....	23
Table S1 .....	10
Table S2 .....	11
Table S3 .....	13
Table S4 .....	18
Spectral Data .....	25-33
Supporting Information References .....	34

## General Remarks

**Computations.** All the quantum mechanical computations were performed using Gaussian 03 suit of programs (Gaussian Inc., <http://gaussian.com>) on a Linux workstation with  $2 \times 2.93$  GHz 6-core X5670 Intel Xeon processors and 64 GB random access memory, housed in the Department of Chemistry, University of Cambridge.

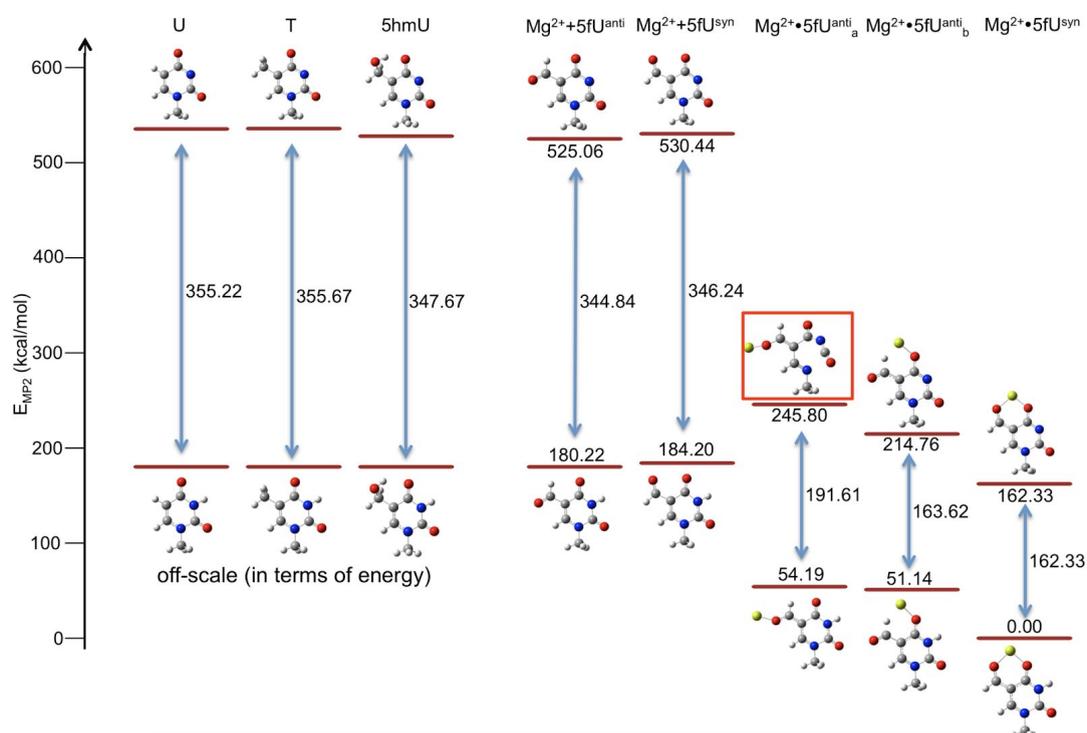
**Reagents and solvents.** All reagents were purchased from commercial suppliers (Sigma Aldrich® unless stated otherwise) and used as received. Extraction solvents and solvents used in column chromatography were used as received at HPLC grade. All ODNs were synthesised and sourced from ATD Bio with HPLC purification. Genomic DNA was extracted from HEK293T, HeLa, U2OS and mESCs (E14 strain) cell pellets cultured in the Balasubramanian group.

**Experimental techniques.** All reaction temperatures other than room temperature were recorded as the heating mantle temperature. Volatiles and compounds were removed/concentrated *in vacuo*. Merck Kieselgel 60 F254 pre-coated aluminium-backed plates were used for analytical Thin Layer Chromatography (TLC) and spots were visualized using UV light (254 nm). LC-MS was performed on an amaZon X ESI-MS (Bruker) connected to Ultimate 3000 LC (Dionex) (Agilent Technologies, Santa Clara, CA). Reverse phase flash column chromatography was performed using puriFlash (Interchim) with a puriFlash prep C18HQ LC column (15  $\mu$ m particle size, 150  $\times$  21.2 mm, Interchim). All oligonucleotide reactions were performed in technical, experimental triplicates.

**Characterisation.**  $^1\text{H}$  NMR (500 MHz) spectra were recorded in  $d_6$ -DMSO at 298 K, on a Bruker® DRX-500 instrument. Signals are quoted as  $\delta$  values in ppm and coupling constants ( $J$ ) are quoted in Hertz and approximated to the nearest 0.1. Notations used for  $^1\text{H}$  NMR spectral splitting patterns include: s – singlet, d – doublet, t – triplet, m – multiplet. Data analysis for the NMR spectra was performed using Bruker's TopSpin® 3.6 software. Accurate mass spectra were recorded on a Waters LCT Premier (ESI) spectrometer. DNA concentration was measured using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher). Genomic DNA fragments were assessed using TapeStation with Genomic DNA screentapes (Agilent). HRMS was obtained by direct infusion on a ThermoFinnigan Orbitrap Classic equipped with an ESI probe. IR spectra were obtained using a PerkinElmer Spectrum One FT-IR Spectrometer.

## Note S1. Details on the Performed Calculations

The main set of calculations have been done using Møller-Plesset<sup>[1]</sup> perturbation theory at MP2<sup>[2]</sup> level with 6-311++G(d,p) basis set<sup>[3]</sup> including polarisation and diffuse terms. This followed the earlier calculations at a relatively lower, RHF/6-31G, level of theory, where we did observe the ring opening phenomenon described below. The selected, or very similar, theory and basis set combination had overall been successfully used for nucleobases and other biologically active molecules.<sup>[4,5]</sup> For the initial intention, different uracil modifications, including T (5mU), were studied to infer gas phase N<sub>3</sub>-deprotonation energies, where the sugar substituent at the N<sub>1</sub> position was replaced with a methyl group. Taking into account the overall presence of Mg<sup>2+</sup> around nucleic acids<sup>[6]</sup> as the major counterbalancing ion, the original calculations started with varying arrangements of Mg<sup>2+</sup> ion around the studied bases. The geometries were optimised with and without hydrogens at N<sub>3</sub> position. Geometry optimisations of the deprotonated states started from optimised structures of the intact bases, by first removing the H<sub>3</sub> hydrogen and assigning one less electron to the system. The reached stationary structures were then verified to be potential energy surface (PES) minima, rather than first or higher order transition states, *via* frequency calculations to check the absence of imaginary frequencies. Some of the results of the performed calculations are brought in **Figure S1**.

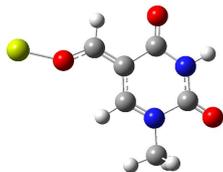


**Figure S1.** The outcome of the MP2/6-311++G(d,p) calculations for deprotonation of various U modifications, including T (i.e. 5mU). The serendipitous observation of the C<sub>1</sub>-N<sub>2</sub> ring opening for one of the geometric configurations of 5fU is outlined in a red box. See the text for further details.

There (**Figure S1**), the y-axis represents gas phase MP2 energy differences upon deprotonation. For the first 3 bases on the left (U, T and 5hmU), the energy of the initial states are brought in an arbitrarily equalised level (the absolute energies are different due to the different molecular compositions of the molecules), so that the deprotonation  $\Delta E_{\text{MP2}}$  values can be compared. As for the *anti* and *syn* conformations of 5fU, their energies are summed with  $E_{\text{MP2}}$  of  $\text{Mg}^{2+}$  ion (hence the notations  $\text{Mg}^{2+}+5\text{fU}^{\text{anti}}$  and  $\text{Mg}^{2+}+5\text{fU}^{\text{syn}}$  in **Figure S1**), to bring their scale into a comparable one with 5fU configurations, where  $\text{Mg}^{2+}$  actually interacts with 5fU (see  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti}}_{\text{a}}$ ,  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti}}_{\text{b}}$  and  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{syn}}$  in **Figure S1** and **Figure 1A**). The energies for the rightmost five columns of structures in **Figure S1** are referenced *via* the most stable *in vacuo* configuration,  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{syn}}$ , setting the energy of the latter to 0 kcal/mol. Please note, that the transition between the chelated structure  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{syn}}$  and the other configurations may be seamless in water, facilitated by specific hydrogen bonding interactions with surrounding water molecules throughout the transition, and by fast dynamics and multiple occupancy of  $\text{Mg}^{2+}$  ions around nucleic acids.<sup>[6]</sup> The gas phase transition between *syn* and *anti* forms of formyl group in 5fU has  $\sim 4$  kcal/mol energy difference, deeming the CHO *syn* $\leftrightarrow$ *anti* flips possible, potentially even easier in water through specific interactions of water molecules. Such extra stabilising interactions are not possible to capture *via* implicit solvation models, hence explicit solvent molecules and more comprehensive search for stationary points in the potential energy surface of the {base + water} system would be required to describe the transitions in water, a study that is outside of the scope of the present work. Interestingly, the ion- (and potentially hydrogen bonded  $\text{H}_2\text{O}$ ) bound configurations have notably lower gas-phase (g) energy barriers of deprotonation, with  $\Delta E_{\text{MP2}}$  being 191.61, 163.62 and 162.33 kcal/mol for the calculated three  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti}}_{\text{a}}$ ,  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti}}_{\text{b}}$  and  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{syn}}$  configurations respectively (**Figure S1**). These are even lower when we calculate Gibbs free energy differences, with  $\Delta G_{(\text{g})}^{\circ} = 179.47$  ( $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti}}_{\text{a}}$ ), 155.82 ( $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti}}_{\text{b}}$ ) and 153.02 ( $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{syn}}$ ) kcal/mol (**Figure 1A**). Free energies were derived from thermal calculations done at the same level of theory, by performing zero point vibrational energy (ZPVE) correction, rotational/translational correction, and corrections for the enthalpic and entropic contributions at 298.15 K. Frequency scaling factors for ZPVE correction,  $\Delta H_{\text{vib}}^{\text{T}=298.15\text{K}}$  enthalpic and  $S_{\text{vib}}^{\text{T}=298.15\text{K}}$  entropic contributions for the MP2/6-311++G(d,p) model chemistry were taken as 0.9768, 1.0071 and 1.0158 respectively.<sup>[7]</sup> The obtained deprotonation free energies were significantly lower and comparable to the compensating  $\text{H}^+$  proton solvation free energy in water ( $\Delta G_{(\text{g})}^{\circ}(\text{H}^+) \approx -265.9$  kcal/mol),<sup>[8,9]</sup> demonstrating the relative ease of deprotonation. 5fU alone was already demonstrated to have lower  $\text{pK}_{\text{a}}$  value (8.2 in nucleosides, 8.6 in nucleotides, as compared to 9.3 and 9.5 in the same context for U, and to 9.7 and 10.0 for T), owing to the electron withdrawing 5-formyl group.<sup>[10]</sup>

Interestingly, the deprotonated state of  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti}}_{\text{a}}$  was never observed, since upon the *in silico* deprotonation, the 5fU uracil ring immediately opens up through the breakage of the  $\text{N}_1\text{-C}_2$  bond (**Figure S1** and **Figure 1A**), reaching a ring-opened stationary intermediate (outlined with a red box in **Figure S1**). This phenomenon was also observed at lower, RHF/6-31G, level of theory. If forcing a ring-closed state after  $\text{N}_3$  deprotonation, we can see that the ring-opened state, which is a PES minimum, has significantly lower free energy ( $\Delta G_{(\text{g})}^{\circ} \text{ring opening} = -25.24 \text{ kcal/mol}$ ), as compared to the ring-closed variant, with the ring opening thus happening spontaneously upon deprotonation, without an extra barrier. The key optimised geometries are brought below. The stationary points (A) and (B) were characterised with 0 imaginary frequencies. The energies listed at the top of the Cartesian coordinates are brought in atomic units (a.u., hartree ( $E_h$ )) for energy,  $1 E_h = 627.50947 \text{ kcal/mol} = 2625.49962 \text{ kJ/mol}$ ).

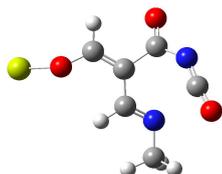
**A)  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti}}_{\text{a}}$ .** The result of a full geometry optimisation. PES minimum with 0 imaginary frequencies.



$$E_{\text{MP2}} = -765.14536288779 E_h, G^{\circ} = -765.06159296619 E_h$$

N	7	4.795578	-0.394594	-0.210478
C	6	3.686132	-1.296400	0.182358
O	8	3.656619	-2.404994	-0.262712
N	7	2.765983	-0.709511	1.018520
C	6	2.863403	0.514294	1.691694
O	8	2.064769	0.858111	2.529732
C	6	4.016323	1.343493	1.229975
C	6	4.938450	0.802784	0.292692
C	6	5.749846	-0.981393	-1.177226
C	6	4.119435	2.597591	1.777850
O	8	5.064329	3.478570	1.510855
H	1	2.011444	-1.325941	1.321766
H	1	5.794696	1.382737	-0.045999
H	1	5.204966	-1.248040	-2.083091
H	1	6.182417	-1.879740	-0.736427
H	1	6.519345	-0.240378	-1.389602
H	1	3.341162	2.876152	2.492419
Mg	12	5.679652	5.162670	1.883835

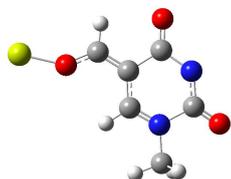
**B)  $\text{Mg}^{2+}\cdot 5\text{fU}^{\text{anti-}}$** . The result of a full geometry optimisation of the  $\text{N}_3$ -deprotonated analogue. The starting geometry was taken from (A), removing the  $\text{H}_3$  hydrogen and reducing the charge of the system by 1. The 5fU ring opened spontaneously upon geometry optimisation, reaching a PES minimum with 0 imaginary frequencies.



$$E_{\text{MP2}} = -764.84001985515 E_h, G^{\circ} = -764.77558575015 E_h$$

N	7	5.200010	-0.472686	0.054880
C	6	2.628599	-1.198559	0.103717
O	8	2.665818	-1.782704	-0.907067
N	7	2.420928	-0.674103	1.195631
C	6	3.102573	0.275788	1.956098
O	8	2.894032	0.403876	3.142998
C	6	4.024420	1.226103	1.231476
C	6	4.977326	0.783652	0.219269
C	6	6.159393	-0.859373	-0.965696
C	6	3.919523	2.522935	1.637371
O	8	4.628171	3.533045	1.091710
H	1	5.490119	1.554000	-0.375447
H	1	5.644366	-1.487701	-1.698365
H	1	6.935391	-1.470780	-0.497433
H	1	6.617395	0.000902	-1.474578
H	1	3.209289	2.796615	2.415923
Mg	12	5.925753	4.710342	1.573909

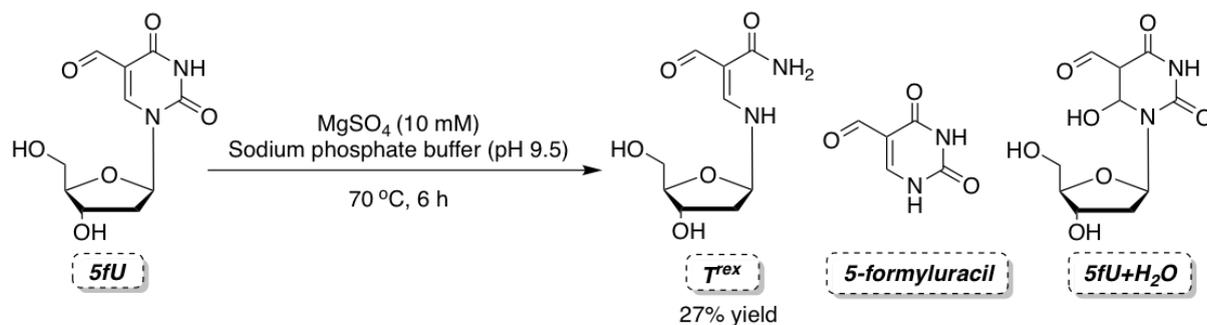
C)  $\text{Mg}^{2+} \cdot 5\text{fU}_{\text{a ring\_enforced}}^{\text{anti}}$ . The result of a single-point calculation. The structure is  $\text{N}_3$ -deprotonated, but artificially ring imposed. It is not a PES minimum, but a saddle point with 1 imaginary frequency. Geometry-wise, this is the same structure as the one from the full optimisation run (A), but with  $\text{H}_3$  removed and charge adjusted. The calculation was then proceeded as a single-point run without geometry optimisation, mimicking an instantaneous deprotonation with optimal rearrangement in the electronic structure, but without geometry rearrangement.



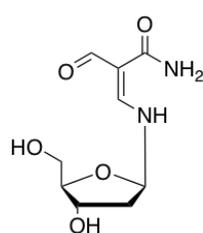
$$E_{\text{MP2}} = -764.80447944186 E_h, G^{\circ} = -764.73536188826 E_h$$

N	7	4.795578	-0.394594	-0.210478
C	6	3.686132	-1.296400	0.182358
O	8	3.656619	-2.404994	-0.262712
N	7	2.765983	-0.709511	1.018520
C	6	2.863403	0.514294	1.691694
O	8	2.064769	0.858111	2.529732
C	6	4.016323	1.343493	1.229975
C	6	4.938450	0.802784	0.292692
C	6	5.749846	-0.981393	-1.177226
C	6	4.119435	2.597591	1.777850
O	8	5.064329	3.478570	1.510855
H	1	5.794696	1.382737	-0.045999
H	1	5.204966	-1.248040	-2.083091
H	1	6.182417	-1.879740	-0.736427
H	1	6.519345	-0.240378	-1.389602
H	1	3.341162	2.876152	2.492419
Mg	12	5.679652	5.162670	1.883835

## Note S2. Experimental Characterisation of dT<sup>rex</sup>

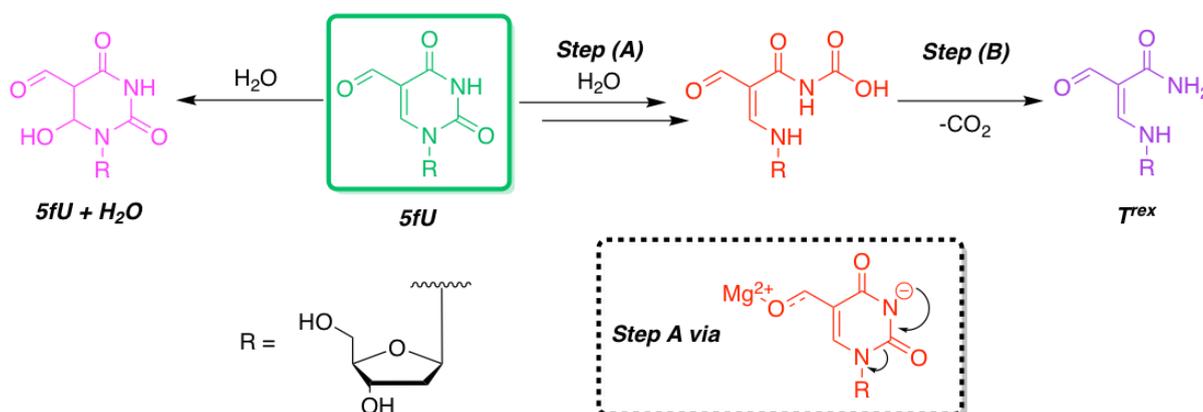


**Scheme S1.** Synthesis of dT<sup>rex</sup> synthetic standard.



5-formyl-2'-deoxyuridine (113 mg, 0.44 mmol) was dissolved in pH 9.5 sodium phosphate buffer (10 mL, 0.5 M). An aqueous MgSO<sub>4</sub> solution (0.11 mL, 1 M) was added and a resulting cloudy suspension was heated at 70 °C for 6 h, allowed to cool to room temperature, and filtered through a 0.22 μm Millex® filter unit.

The resulting filtrate was purified by HPLC using an isocratic elution gradient of 100% H<sub>2</sub>O (0.1% formic acid) over 30 min. The fraction containing ring-opened 5-formyl uridine (T<sup>rex</sup>) was freeze-dried to yield a white amorphous solid (11.3 mg, 0.05 mmol, 26.6% yield), with 46% of starting material recovered – yield based on reacted 5-formyl-2'-deoxyuridine (5fU). The purified T<sup>rex</sup> nucleoside was stable at room temperature for several weeks and at -20 °C for several months (assessed by <sup>1</sup>H, <sup>13</sup>C NMR and LC-MS), suggesting that T<sup>rex</sup> is not prone to further rearrangement or polymerization. Concern for its stability arose because of its analogous structure to acrylamides, which are known to polymerize.

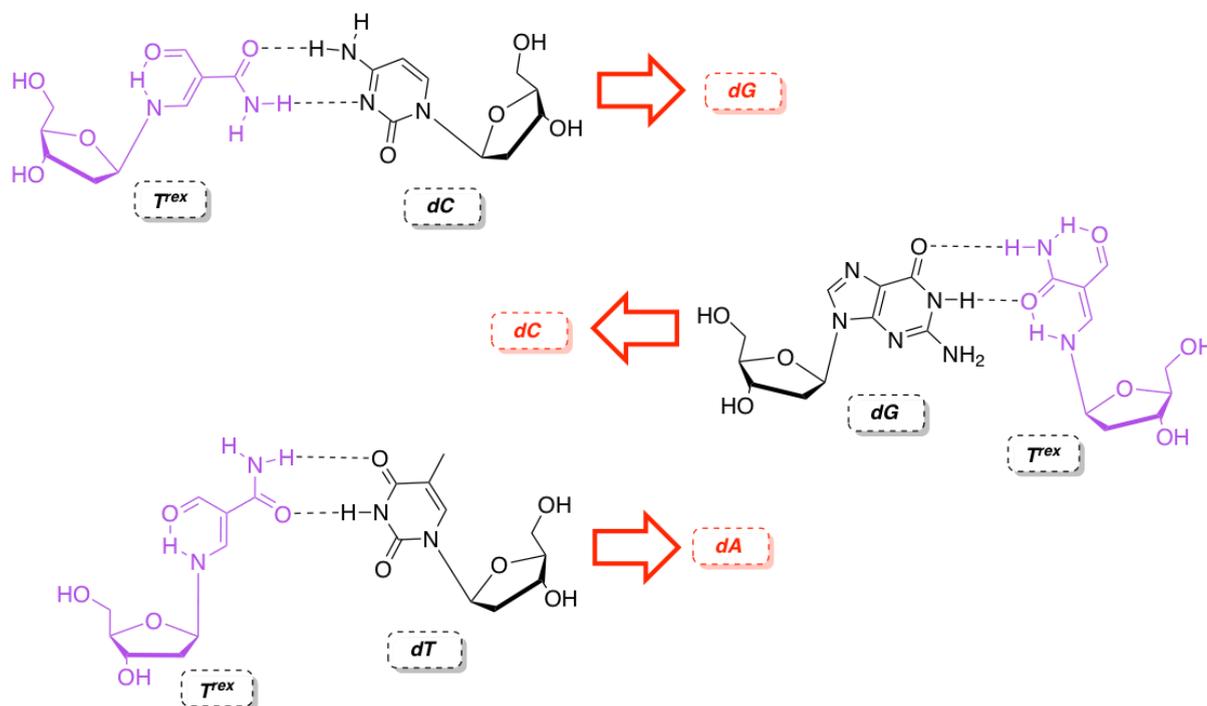


**Scheme S2.** A hypothesised schematic depicting the formation of T<sup>rex</sup> from 5fU (blue) via hydrolysis of an isocyanate intermediate to form a carbamate derivative which subsequently undergoes decarboxylation. In red shows the formation of hydrated 5fU across the C<sub>5</sub>-C<sub>6</sub> position. “R” depicts the 2'-deoxyribose.

$^1\text{H NMR}$  (500 MHz,  $d_6$ -DMSO, 2.50 ppm ref.)  $\delta$  = 10.47-10.40 (m, close to d-of-ds,  $\sim$ 13.6 Hz,  $\sim$ 9.3 Hz, 1H, -NH-), 8.92 (s, 1H, -CHO), 8.23 (s, br,  $\sim$ 2.8 Hz spl. differentiable, 1H, -CONH<sup>a</sup>H<sup>b</sup>), 7.87 (d, 13.3 Hz, 1H, -NH-CH=), 7.15 (s, br,  $\sim$ 2.8 Hz spl. differentiable, 1H, -CONH<sup>a</sup>H<sup>b</sup>), 5.39-5.33 (m, close to d-of-ts,  $\sim$ 6.6 Hz,  $\sim$ 8.8 Hz, 1H, -CH<sub>2</sub>-CH(O-)-NH- i.e. H1'), 5.14 (s, br, 1H, -CH(O-)-CH(OH)-CH<sub>2</sub>- i.e. HO3'), 4.91 (s, br, 1H, -CH<sub>2</sub>-OH i.e. HO5'), 4.21-4.17 (m, 1H, -CH(O-)-CH(OH)-CH<sub>2</sub>- i.e. H3'), 3.77-3.74 (m, close to t-of-ds, 4.1 Hz, 2.2 Hz, 1H, HO-CH<sub>2</sub>-CH(O-)-CH(OH)- i.e. H4'), 3.43 (d,  $\sim$ 3.4 Hz, 2H, -CH<sub>2</sub>-OH, i.e. H5' & H5''), 2.15-2.04 (m, non-1<sup>st</sup>-order, 2H, -CH<sub>2</sub>-CH(O-)-NH- i.e. H2' & H2'').

$^{13}\text{C NMR}$  (125 MHz,  $d_6$ -DMSO, 39.5 ppm ref.)  $\delta$  = 188.8 (-CHO,  $\sim$ 169 Hz  $^1J_{\text{CH}}$  is observed, unfiltered from HMBC), 168.5 (-CONH<sub>2</sub>), 164.7 (-NH-CH=), 104.2 (=C(CHO)CONH<sub>2</sub>), 89.3 (-CH<sub>2</sub>-CH(O-)-NH- i.e. C1'), 87.5 (HO-CH<sub>2</sub>-CH(O-)-CH(OH)- i.e. C4'), 71.3 (-CH(O-)-CH(OH)-CH<sub>2</sub>- i.e. C3'), 61.9 (-CH<sub>2</sub>-OH, i.e. C5'), 41.0 (-CH<sub>2</sub>-CH(O-)-NH- i.e. C2').

Microanalysis identified the compound as the monoformate salt, yielding %C, 41.80; %H, 5.90; %N, 10.60, in accordance with the theoretical values (%C, 43.48; %H, 5.84; %N, 10.14). HRMS  $\text{C}_9\text{H}_{14}\text{N}_2\text{O}_5\text{Na}^+ [\text{M}+\text{Na}]^+$  calculated 253.0795, found 253.0785.



**Figure S2.** Potential base pair formation of the canonical bases with T<sup>rex</sup> and thus an illustration of its potential to introduce mutations into the genome during DNA replication.

### Note S3. Reaction Condition Screening and Optimisation on 5fU-ODN

**Table S1:** Sequences of ODNs used for this study (positions in bold highlight modifications). AP corresponds to an abasic site lesion. Cleaved 5fU-ODN corresponds to cleavage of the 5fU-ODN II at the 5fU site.

ODNs	Sequences
<b>5fU-ODN I</b>	5'-TACA <b>5fU</b> GCGAT-3'  Complementary sequence: 5' <b>Phos</b> -ATCGCATGTA-3'
<b>T<sup>rex</sup>-ODN I</b>	5'-TACA <b>T<sup>rex</sup></b> GCGAT-3'
<b>AP-ODN I</b>	5'-TACA <b>AP</b> GCGAT-3'
<b>5fU-ODN II</b>	5'-ATCGAGAATCGGTGC <b>5fU</b> AGTAACTAAGTCTCCAGG-3'
<b>T<sup>rex</sup>-ODN II</b>	5'-ATCGAGAATCGGTGC <b>T<sup>rex</sup></b> AGTAACTAAGTCTCCAGG-3'
<b>AP-ODN II</b>	5'-ATCGAGAATCGGTGC <b>AP</b> AGTAACTAAGTCTCCAGG-3'
<b>Cleaved 5fU-ODN II</b>	5'-ATCGAGAATCGGTGC-3'

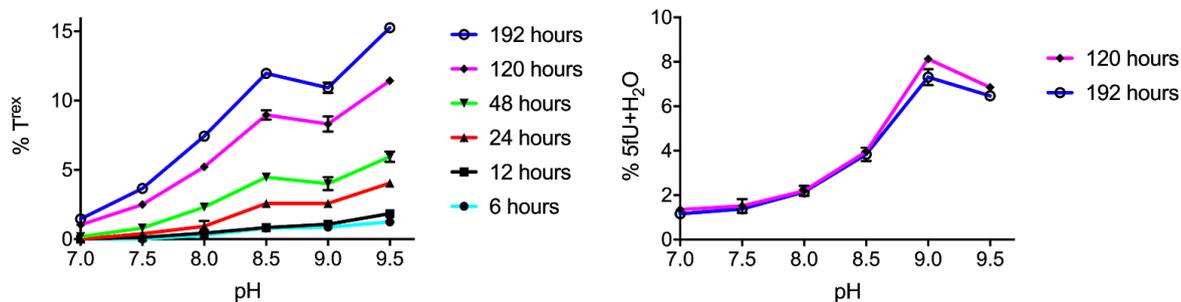
**LC-MS analysis of ODNs (10 mer).** LC-MS was performed on a Bruker amaZon X system, using a XTerra® MS C18 column (2.1 × 50 mm, 2.5 µm particle size), with 100 mM 1,1,1,3,3,3-hexafluoro-2-propanol; 10 mM NEt<sub>3</sub> (solvent A) and MeOH (solvent B) at a flow-rate of 0.2 mL min<sup>-1</sup>. The column was pre-equilibrated with 10% B for 10 min and followed by a gradual increase to 20% B over a duration of 25 min. Reaction conversion was calculated by integration of UV signals of the starting material and product(s) at 260 nm. Identity of the product(s) was confirmed by ESI-MS with negative polarity in ultra-scan mode. Data was acquired between 1000-2800 *m/z*. Observed ESI<sup>-</sup> signals and their retention times are listed in **Table S2**.

**Table S2:** LC-MS data for 10 mer ODNs used in this study and their reaction products.

ODN	MW	ESI-MS	Retention Time (min)
5fU-ODN I	3041.0	$[M-2H]^{-2} = 1519$	11.1
T <sup>rex</sup> -ODN I	3015.0	$[M-2H]^{-2} = 1506$	10.6
5fU-ODN I + H <sub>2</sub> O	3059.0	$[M-2H]^{-2} = 1528$	12.6
5fU-ODN I Complement	3107.0	$[M-2H]^{-2} = 1552$	14.5
AP-ODN I	2918.9	$[M-2H]^{-2} = 1458$	10.2

**Salt screen on ring opening of 5fU-ODN I.** 5fU-ODN (2.5  $\mu$ L, 100  $\mu$ M) was added to inorganic salt (1 mM) (**Figure 2B**) in either pH 7.0, pH 8.0 or pH 9.0 sodium phosphate buffer (50 mM) in a final reaction volume of 50  $\mu$ L. The reactions were incubated at 37 °C for 48 h, before being purified through a mini quick spin oligo column (Roche), which was pre-washed with water (3  $\times$  300  $\mu$ L). Samples were diluted 2-fold to a final concentration of 2.5  $\mu$ M and traced by general oligomer LC-MS analysis.

**pH screen and time-course analysis on ring opening of 5fU-ODN I.** 5fU-ODN (2.5  $\mu$ L, 100  $\mu$ M) was added to MgSO<sub>4</sub> (0.5  $\mu$ L, 100 mM) at different pH (7.0, 7.5, 8.0, 8.5, 9.0, 9.5) sodium phosphate buffer (50 mM) in a final reaction volume of 50  $\mu$ L. The reactions were incubated at 37 °C for a maximum duration of 192 h, with regular time increments. These were subsequently purified through a mini quick spin oligo column (Roche), which was pre-washed with water (3  $\times$  300  $\mu$ L). Samples were diluted 2-fold to a final concentration of 2.5  $\mu$ M and traced by general oligomer LC-MS analysis.



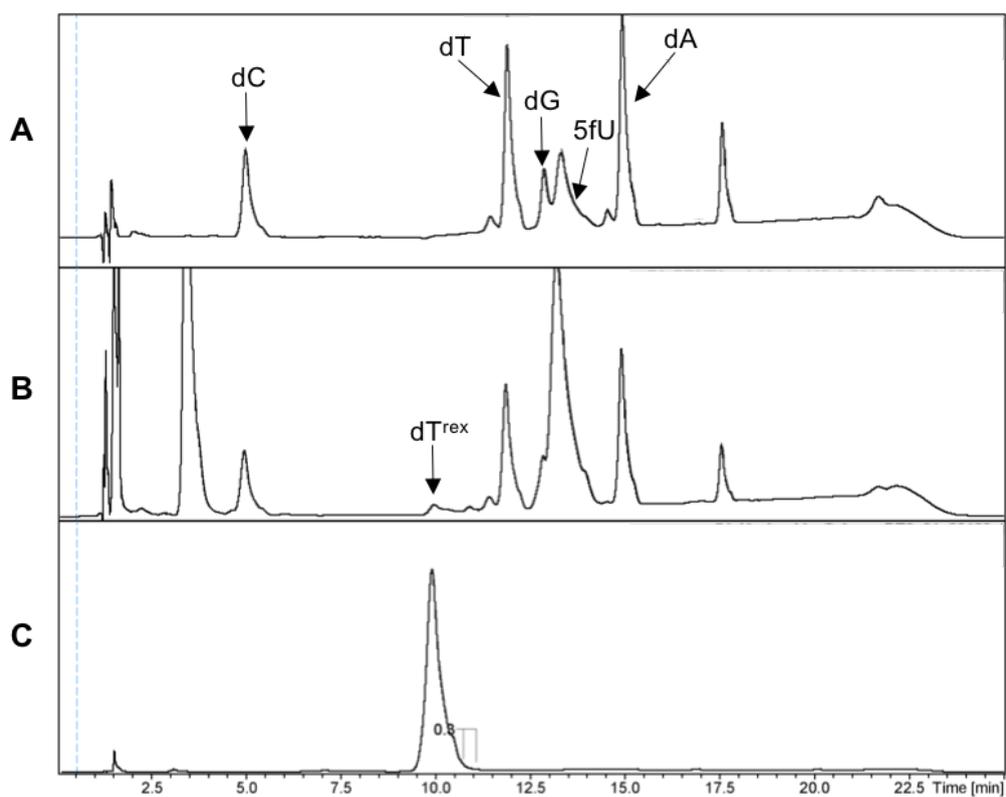
**Figure S3.** Dependence of 5fU ring opening on pH (left), with a local minimum observed at pH 9.0, due to maximal production of 5fU+H<sub>2</sub>O as seen in (right).

**Validation of T<sup>rex</sup> formation in 5fU-ODN I.** Ca. 18  $\mu\text{g}$  of 5fU-ODN I was subjected to ring opening conditions on a 1  $\mu\text{g}$  reaction scale. 5fU-ODN (3.3  $\mu\text{L}$ , 100  $\mu\text{M}$ ) was added to MgSO<sub>4</sub> (0.5  $\mu\text{L}$ , 100 mM) and pH 9.5 sodium phosphate buffer (50 mM) in a final reaction volume of 50  $\mu\text{L}$ . The reaction was incubated at 70 °C for 6 h, before being pooled and purified through a mini quick spin oligo column (Roche), which was pre-washed with water (3  $\times$  300  $\mu\text{L}$ ). The samples was concentrated *in vacuo* and reconstituted in 80  $\mu\text{L}$  ultrapure HPLC-grade water, with a small amount traced by general oligomer LC-MS analysis to confirm reaction progression.

Subsequently, the 5fU-ODN I (2  $\mu\text{g}$ ) and 5fU/T<sup>rex</sup>-ODN I mixture (generated above) were digested separately for 1 hour, into their constituent mononucleosides, using 4 units of Nuclease P1 (Sigma-Aldrich) and 10 units of Alkaline Phosphatase (Promega) at 37 °C in 50  $\mu\text{L}$  of reaction buffer containing 5 mM ZnCl<sub>2</sub> and 50 mM NH<sub>4</sub>OAc. The digestions for the latter (5fU/T<sup>rex</sup>-ODN I) were then pooled and purified through a 10 KDa Amicon® Ultra-0.5 mL centrifugal size exclusion filter (Merck Millipore), and the filtrate was concentrated *in vacuo*; reconstituted in 80  $\mu\text{L}$  ultrapure HPLC-grade water. The digested DNA was then analysed by LC-MS on a Bruker amaZon X system, using an Agilent Pursuit 5 C18 column (4.6  $\times$  150 mm, 5  $\mu\text{m}$  particle size), with 10 mM ammonium formate (solvent A) and 0.1% formic acid in MeCN (solvent B) at a flow-rate of 1.5 mL min<sup>-1</sup>. The column was pre-equilibrated with 100% A for 6.5 min, followed by a gradual increase to 10% B over a duration of 13.5 min, then returned to 100% A over 1 min and maintained isocratic for 19 min. Identity of the product(s) was confirmed by ESI-MS and co-elution with the synthetic standard. Observed ESI<sup>+</sup> signals and their retention times are listed in **Table S3** (see **Figure S4** for LC-MS traces of digests).

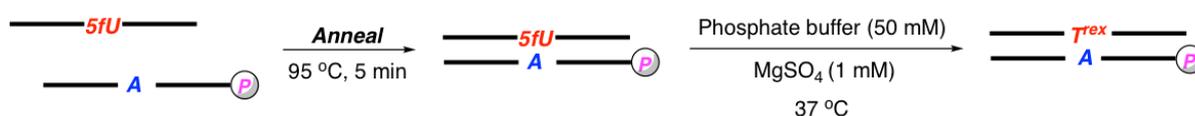
**Table S3:** LC-MS data for 2'-deoxyribonucleosides, including ESI-MS and LC retention times.

dN	ESI-MS	Retention Time (min)
C	$[M+H]^+ = 228$	5.0
T	$[M+H]^+ = 243$	12.7
A	$[M+H]^+ = 252$	14.9
G	$[M+H]^+ = 268$	11.9
5fU	$[M+H]^+ = 257$	13.8
T <sup>rex</sup>	$[M+H]^+ = 231$	10.0



**Figure S4.** (A) LC trace at 260 nm of digested 5fU-ODN I (10 mer), highlighting key peaks corresponding to the different 2'-deoxyribonucleosides. (B) LC trace at 260 nm of digested T<sup>rex</sup>- and 5fU-ODN I (10 mer) mixture obtained from ring opening of 5fU-ODN I, revealing the presence of a peak for dT<sup>rex</sup> at 10.0 min. (C) LC trace at 260 nm of dT<sup>rex</sup> synthetic standard and its co-elution with the T<sup>rex</sup>-ODN digest.

**Ring-opening of dsDNA 5fU-ODN I.** 5fU-ODN I (12.5  $\mu\text{L}$ , 100  $\mu\text{M}$ ) and its complement strand (12.5  $\mu\text{L}$ , 100  $\mu\text{M}$ ) were annealed in a pH 7.0 solution (50  $\mu\text{L}$ ) containing Tris-HCl (20 mM) and NaCl (50 mM), by heating at 95  $^{\circ}\text{C}$  for 5 min and slowly cooling the mixture to 4  $^{\circ}\text{C}$  at a rate of 0.1  $^{\circ}\text{C}/\text{s}$ . Subsequently, the duplex ODN was buffer-exchanged with ultrapure HPLC-grade water through a Bio-Spin® P-6 column (Bio-Rad). Duplex 5fU-ODN (10  $\mu\text{L}$ , 25  $\mu\text{M}$ ) was added to  $\text{MgSO}_4$  (0.5  $\mu\text{L}$ , 100 mM) with either pH 7.0 or 9.5 sodium phosphate buffer (50 mM) in a final reaction volume of 50  $\mu\text{L}$ . The reaction was incubated at 37  $^{\circ}\text{C}$  for a maximum duration of 192 h, with regular time increments, before being purified through a mini quick spin oligo column (Roche), which was pre-washed with water ( $3 \times 300 \mu\text{L}$ ). Samples were diluted 2-fold to a final concentration of 2.5  $\mu\text{M}$  and traced by general oligomer LC-MS analysis. Stability of the duplex was tested using LC-MS analysis. It is reasonable to question the stability of the duplex under such reaction conditions. However, the  $T_m$  of the duplex is 45  $^{\circ}\text{C}$  under these conditions and any duplex that does dissociate will almost immediately re-anneal as there will be a propensity for the ODN to exist in duplex form.

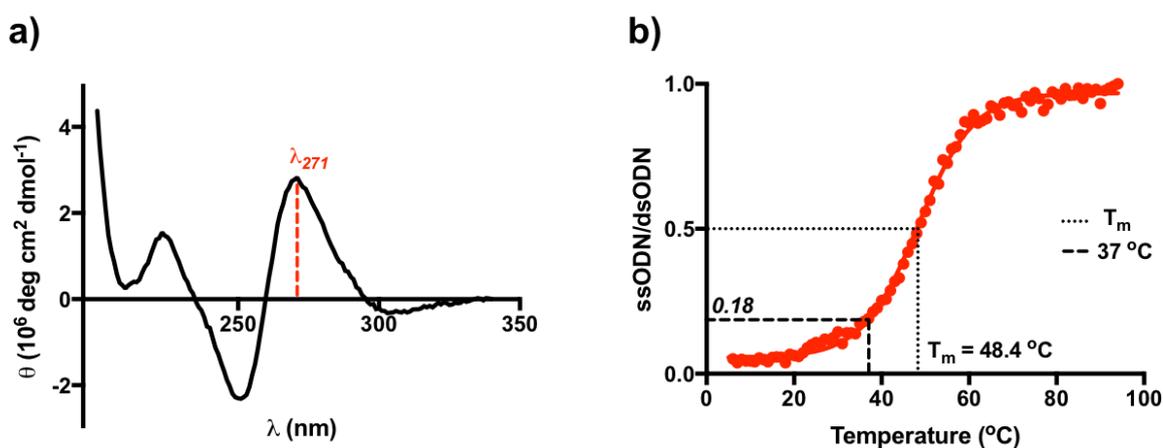


**Figure S5.** A schematic depicting the generation of a duplex ODN, containing a single 5fU site. The duplex subsequently being subjected to ring opening conditions to generate a duplex ODN containing  $T^{\text{rex}}$ .

**Biophysics of dsDNA 5fU-ODN and circular dichroism melting.** 5fU-ODN I (12.5  $\mu\text{L}$ , 100  $\mu\text{M}$ ) and its complement strand (12.5  $\mu\text{L}$ , 100  $\mu\text{M}$ ) were annealed in a pH 7.0 solution (50  $\mu\text{L}$ ) containing Tris-HCl (20 mM) and NaCl (50 mM), by heating at 95  $^{\circ}\text{C}$  for 5 min and slowly cooling the mixture to 4  $^{\circ}\text{C}$  at a rate of 0.1  $^{\circ}\text{C}/\text{s}$ . Subsequently, the duplex ODN was buffer-exchanged with pH 7.0 sodium phosphate buffer (50 mM) containing  $\text{MgSO}_4$  (1 mM) through a Bio-Spin® P-6 column (Bio-Rad). The circular dichroism (CD) spectrum of the duplex (150  $\mu\text{L}$ , 25  $\mu\text{M}$ ) was recorded between 200-340 nm in a cuvette with a path length of 10 mm on an Applied Photophysics Chirascan™ instrument, fitted with a Quantum Northwest TC 1 temperature controller unit. Data was acquired on a Pro-Data Chirascan software at 5  $^{\circ}\text{C}$  with 3 technical replicates, a bandwidth of 1 nm and a sampling rate of one acquisition per second. The CD spectrum was subsequently buffer (blank) subtracted and zero-corrected at 340 nm to give a  $\lambda_{\text{max}}$  of 271 nm and a  $\lambda_{\text{min}}$  of 251 nm.

The CD melting curve was recorded on the 5fU-ODN duplex as above in reaction buffer at 271 nm in a cuvette with a path length of 10 mm on an Applied Photophysics Chirascan™ instrument, fitted with a Quantum Northwest TC 1 temperature controller unit. Data was acquired on the Pro-Data Chirascan software using a smooth temperature ramping between 5  $^{\circ}\text{C}$  and 94  $^{\circ}\text{C}$  at a rate of 1  $^{\circ}\text{C}/\text{min}$  with 3

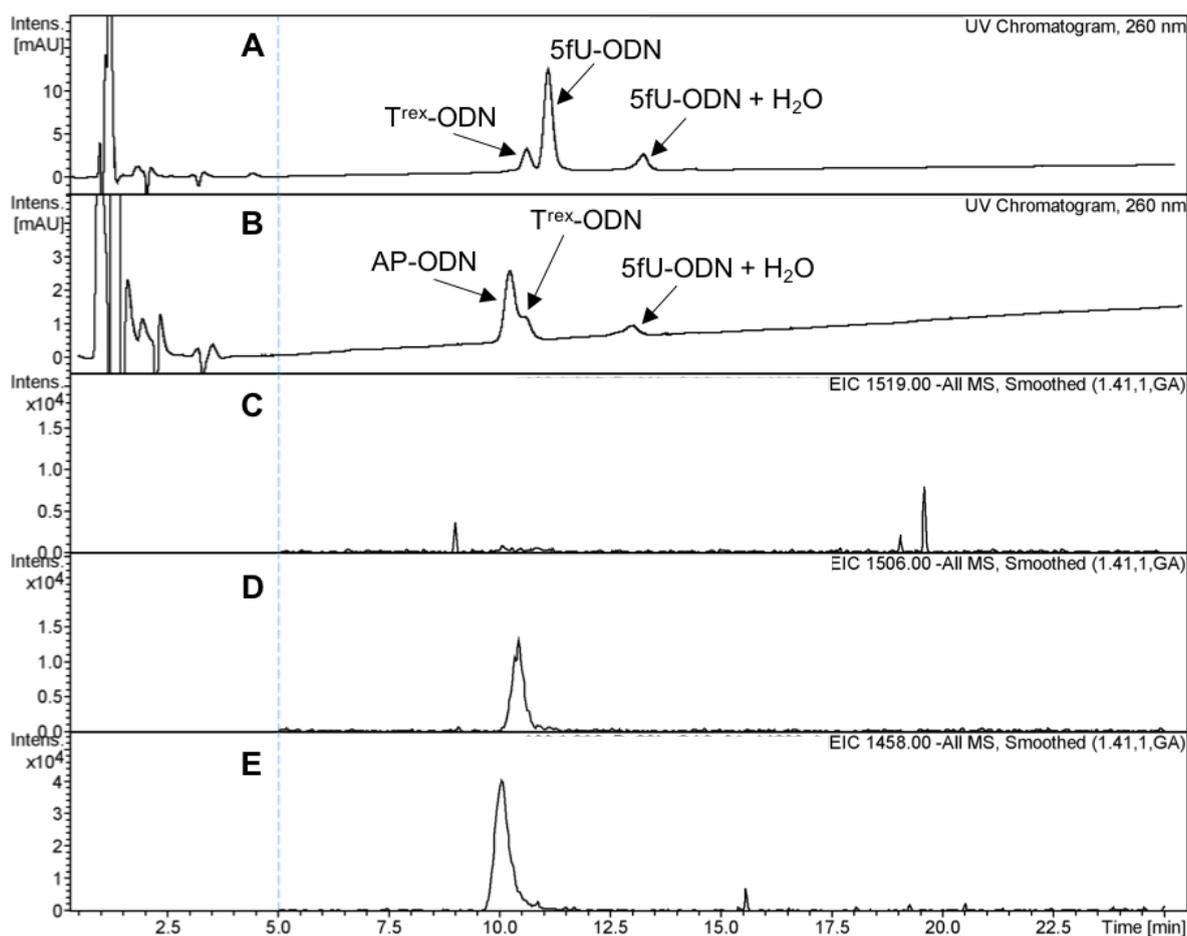
technical replicates (average shown in the plots below), a bandwidth of 1 nm and a sampling rate of one acquisition per second. The data was normalised to a range between 0 and 1 and the fraction of ssODN (single-stranded DNA) over dsODN (duplex DNA) was plotted against temperature. Fitting the curve using the Boltzmann sigmoidal model gave a  $T_m$  of 48.4 °C, at which 50% of the duplex was in the denatured state with a  $R^2$  of 0.9973. It can be seen that under the reaction conditions of 37 °C, 18% of the dsODN exists in a denatured state and we deem this value to be acceptable for us to perform the study on the duplex 5fU ODN.



**Figure S6. (a)** Shown is the CD spectrum of duplex 5fU ODN I (10 base pairs) in pH 7.0 sodium phosphate (50 mM) reaction buffer, containing  $\text{MgSO}_4$  (1 mM). The CD spectrum is reported as an average of 3 technical replicates and was buffer subtracted and zero-corrected at 340 nm. **(b)** Shown is the normalised CD melting curve of 5fU ODN I duplex in the reaction buffer mentioned above, recorded at 271 nm. The data (an average of 3 technical replicates) is fitted using a Boltzmann sigmoidal model to give a  $T_m$  of 48.4 °C. The majority of the duplex (82%) remains intact at 37 °C (reaction conditions) and is thus likely to be stable during the course of the reaction study on the duplex.

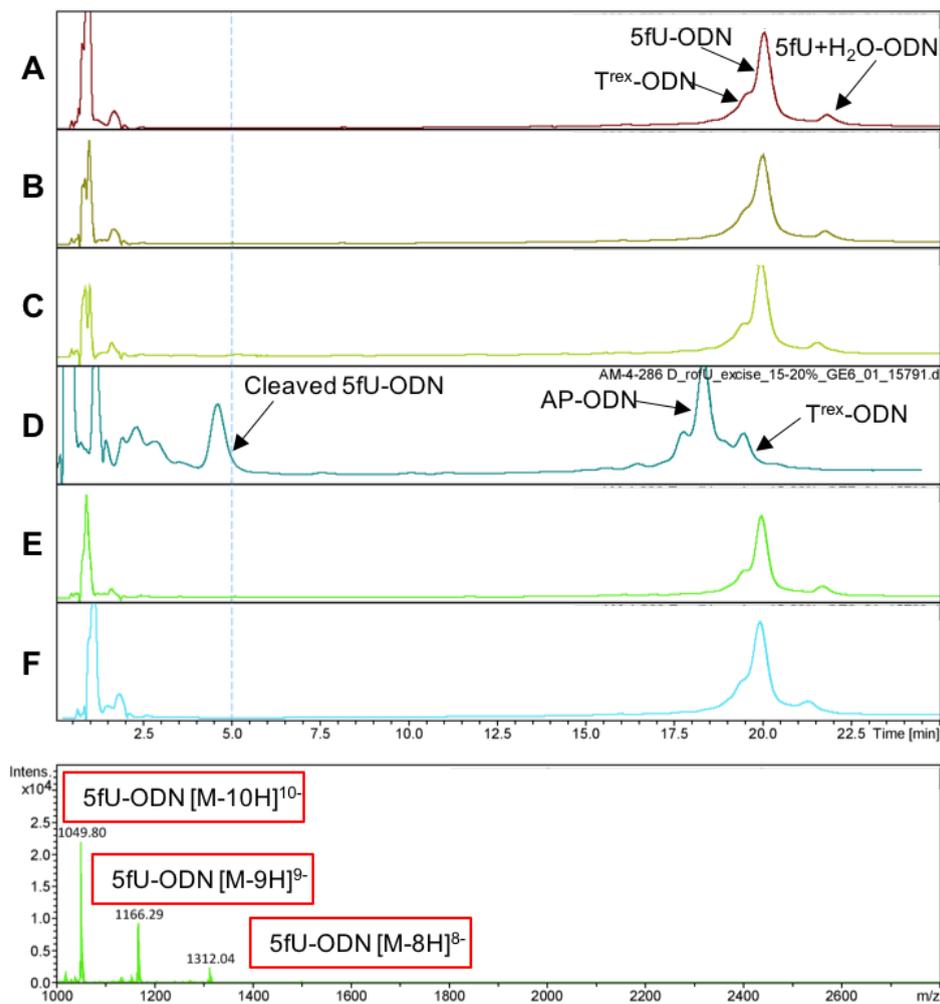
## Note S4. Investigation of T<sup>rex</sup> Susceptibility to Repair by Glycosylases

**hSMUG1 treatment of 5fU/T<sup>rex</sup>-ODN I.** The mixture of 5fU-ODN I and T<sup>rex</sup>-ODN I (20  $\mu$ L, 5  $\mu$ M) was incubated at 37  $^{\circ}$ C for 22 h with 15 units of hSMUG1 (NEB), supplemented with 5  $\mu$ g of purified BSA (NEB) in NEBuffer<sup>TM</sup> 1 at a final reaction volume of 50  $\mu$ L. The reactions were subsequently purified through an Oligo Clean & Concentrator<sup>TM</sup> kit (Zymo), following manufacturer's guidelines. Samples were eluted into 50  $\mu$ L of ultrapure HPLC-grade water and traced by general oligomer LC-MS analysis (for ESI-MS see **Table S2**).



**Figure S7.** (A) LC trace at 260 nm of T<sup>rex</sup>- and 5fU-ODN I (10 mer) mixture, after subjection to ring opening conditions. (B) LC trace at 260 nm of T<sup>rex</sup>- and 5fU-ODN I (10 mer) mixture, after incubation with hSMUG1 for 22 hours, highlighting the presence of T<sup>rex</sup>-ODN and quantitative excision of 5fU-ODN. (C-E) EIC tracts of 5fU-ODN ( $m/z$  1519), T<sup>rex</sup>-ODN ( $m/z$  1506) and 5fU-ODN+H<sub>2</sub>O ( $m/z$  1528) respectively.

**Glycosylase screen on 5fU/T<sup>rex</sup>-ODN II.** To generate the T<sup>rex</sup>-ODN II. 5fU-ODN II (2.5  $\mu$ L, 100  $\mu$ M) was added to MgSO<sub>4</sub> (0.5  $\mu$ L, 100 mM) and pH 9.5 sodium phosphate buffer (50 mM) in a final reaction volume of 50  $\mu$ L. The reaction was incubated at 70 °C for 6 h, before being purified through a mini quick spin oligo column (Roche), which was pre-washed with water (3  $\times$  300  $\mu$ L). Sample was traced by general oligomer LC-MS analysis (for ESI-MS see **Table S4**) prior to incubation with a panel of glycosylases. The 5fU/T<sup>rex</sup>-ODN II mixture (10  $\mu$ L, 5  $\mu$ M) was incubated according to the manufacturer's guidelines for 22 h with 5 units of either UDG, *Afu*-UDG, Fpg, Endo III (Nth), T4 PDG or *Tma* Endo III glycosylases (NEB), in the appropriate reaction buffers at a final reaction volume of 50  $\mu$ L. The reactions were subsequently purified through an Oligo Clean & Concentrator™ kit (Zymo), following manufacturer's guidelines. Samples were eluted into 50  $\mu$ L of ultrapure HPLC-grade water and traced by oligomer LC-MS analysis (for ESI-MS see **Table S4**).



**Figure S8.** Glycosylase screen for 22 hours on T<sup>rex</sup>- and 5fU-ODN II (34 mer) mixture, with LC traces recorded at 260 nm. Incubation with UDG (A), *Afu*-UDG (B), Endonuclease III (Nth) (C), T4 PDG (E) and *Tma* Endonuclease III (F) showed no excision activity. Incubation with Fpg (D) revealed quantitative excision of 5fU-ODN, into a 1:1 ratio of AP-ODN and the cleaved ODN. ESI-MS trace shows the molecular ions for the 5fU-ODN II.

**LC-MS analysis of ODNs (34 mer).** LC-MS was performed on a Bruker amaZon X system, using a XTerra® MS C18 column (2.1 × 50 mm, 2.5 μm particle size), with 100 mM 1,1,1,3,3,3-hexafluoro-2-propanol; 10 mM NEt<sub>3</sub> (solvent A) and MeOH (solvent B) at a flow-rate of 0.2 mL min<sup>-1</sup>. The column was pre-equilibrated at 15% B for 10 min and followed by a gradual increase to 20% B over a duration of 25 min. Reaction conversion was calculated by integration of UV signals of the starting material and product(s) at 260 nm. Identity of the product(s) was confirmed by ESI-MS with negative polarity in ultra-scan mode. Data was acquired between 1000-2800 *m/z*. Observed ESI<sup>-</sup> signals and their retention times are listed in **Table S4**.

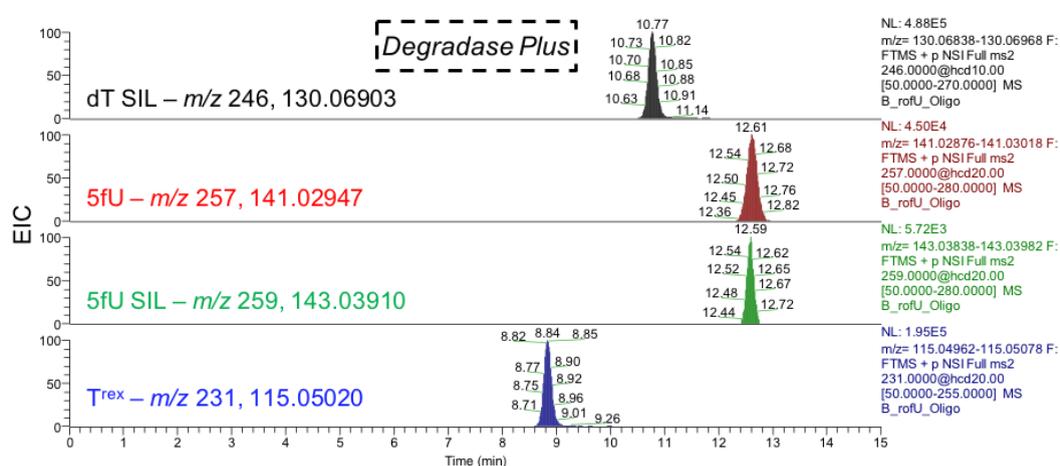
**Table S4:** LC-MS data for 34 mer ODNs used in this study and their reaction products.

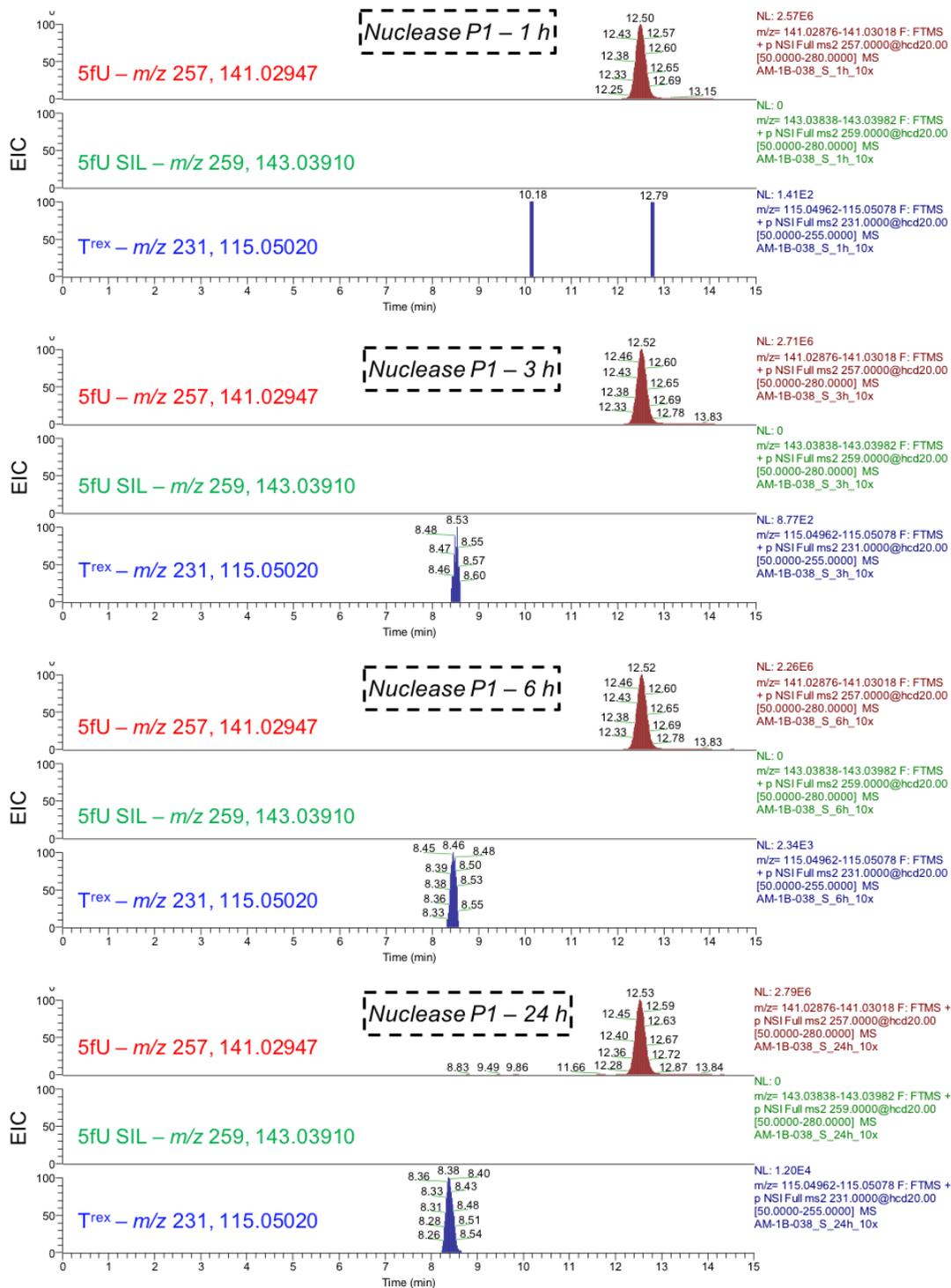
<b>ODN</b>	<b>MW</b>	<b>ESI-MS</b>	<b>Retention Time (min)</b>
<b>5fU-ODN II</b>	10504	[M-10H] <sup>-10</sup> = 1049	20.0
<b>T<sup>rex</sup>-ODN II</b>	10478	[M-10H] <sup>-10</sup> = 1047	19.5
<b>5fU-ODN + H<sub>2</sub>O</b>	10522	[M-10H] <sup>-10</sup> = 1051	21.7
<b>AP-ODN II</b>	10365	[M-10H] <sup>-10</sup> = 1035	18.3
<b>5fU-ODN II Cleaved</b>	4617.1	[M-4H] <sup>-4</sup> = 1538	4.6

## Note S5. Controls for *in vivo* Detection of T<sup>rex</sup> in Human Cell Lines

**Optimisation of enzymatic digestion on 5fU-ODN I.** 5fU-ODN (1 µg) was digested into its constituent mononucleosides using 5 units of DNA Degradase Plus<sup>TM</sup> (Zymo) with 10x Degradase buffer at 37 °C for 6 hours in 50 µL reaction volume. The digestion was purified through a 10 KDa Amicon® Ultra-0.5 mL centrifugal size exclusion filter (Merck Millipore), and the filtrate was then enriched offline for T<sup>rex</sup> (3.9 - 5.6 min) and 5fU (5.6 - 6.8 min) on a Bruker amaZon X HPLC system, using a Waters ACQUITY UPLC® HSS T3 column (2.1 × 100 mm, 1.8 µm particle size), at a flow rate of 0.35 mL min<sup>-1</sup> in a 0.1% formic acid in H<sub>2</sub>O (solvent A) and 0.1% formic acid in MeCN (solvent B) solvent system. After sample injection, this gradient was maintained for 4.0 min at 2% B, after which the gradient was ramped to 100% B over 1.0 min, maintained for 1.5 min and then immediately returned to 2% B and maintained for 7.5 min. The enriched fractions were lyophilised and reconstituted in 40 µL ultrapure HPLC-grade water, spiked with an internal standard of dT SIL (D<sub>3</sub>-dT, +3 Da) and 5fU SIL (<sup>13</sup>CD-5fU, +2 Da), at a final concentration of 25 nM and 2.5 nM respectively. The enriched fractions were subjected to NanoHPLC-MS/MS analysis on a Q-Exactive<sup>TM</sup> Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific).

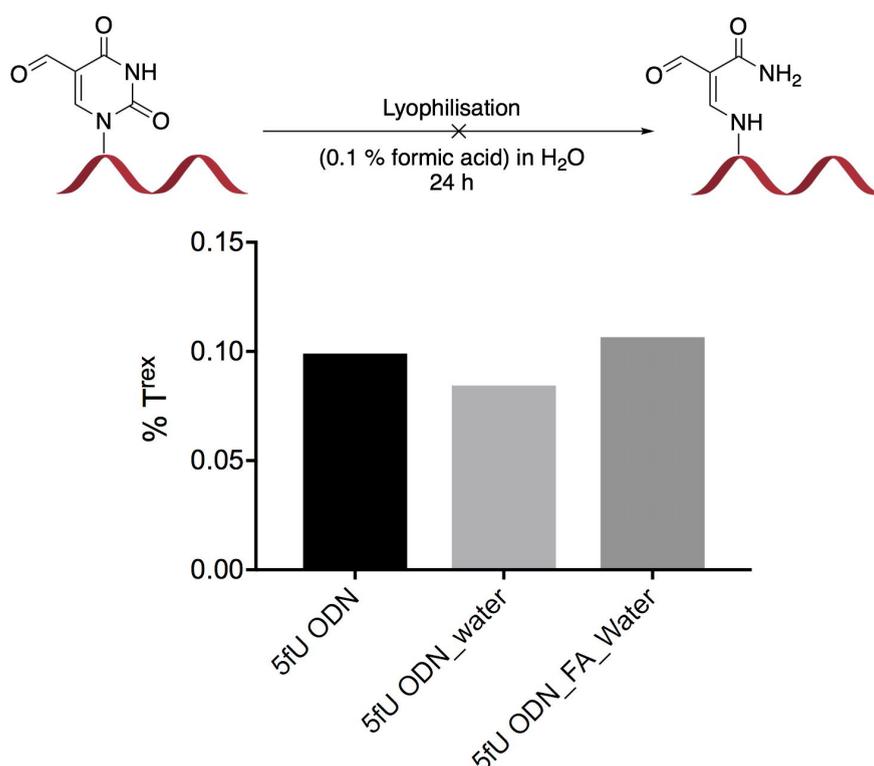
This however, resulted in the artificial formation of T<sup>rex</sup>, at a qualitative level based on signal intensity of ca. 1% of 5fU (**Figure S9**). Using 4 units of Nuclease P1 (Sigma-Aldrich) and 10 units of Alkaline Phosphatase (Promega) at 37 °C for 1 hour in 50 µL of reaction buffer containing 5 mM ZnCl<sub>2</sub> and 50 mM NH<sub>4</sub>OAc eliminated this artificial formation of T<sup>rex</sup>, as seen from the NanoHPLC-MS/MS traces of the Nuclease P1 digestion time-course over 24 hours (**Figure S9**).





**Figure S9.** NanoHPLC-MS/MS traces for 5fU-ODN (10 mer) digestion using Degradase Plus and a timecourse of Nuclease P1 and Alkaline Phosphatase treatment over 24 hours. dT-SIL (black), 5fU (maroon), 5fU-SIL (green) and  $T^{\text{rex}}$  (blue) was visualised by targeting for  $m/z$  130.06903 as a fragment of  $m/z$  246;  $m/z$  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. NL represents the ion count (signal intensity) of the corresponding targeted mass. It is clear that digestion using Degradase Plus produces significant amounts of  $T^{\text{rex}}$ , whereas incubation of the 5fU-ODN with Nuclease P1 drastically minimised  $T^{\text{rex}}$  formation, with no signal being detected until 3 hours of incubation.

**Effect of lyophilisation on 5fU-ODN I.** 5fU-ODN (1  $\mu\text{g}$ ) was suspended in either ultrapure HPLC-grade water or 0.1% formic acid (FA) in  $\text{H}_2\text{O}$  (elution solvent on enrichment LC-MS platform) and lyophilised for a duration of 24 hours. The samples were subsequently reconstituted in 50  $\mu\text{l}$  ultrapure HPLC-grade water and analysed for ring opening by LC-MS on a Bruker amaZon X system, using a XTerra® MS C18 column ( $2.1 \times 50$  mm, 2.5  $\mu\text{m}$  particle size), with 100 mM 1,1,1,3,3,3-hexafluoro-2-propanol; 10 mM  $\text{NEt}_3$  (solvent A) and MeOH (solvent B) at a flow-rate of 0.2  $\text{mL min}^{-1}$ . The column was pre-equilibrated with 10% B for 10 min, followed by a gradual increase to 20% B over a duration of 25 min. Reaction conversion was calculated by integration of UV signals of the starting material and product(s) at 260 nm. Identity of the product(s) was confirmed by ESI-MS with negative polarity in ultra-scan mode. Data was acquired between 1000-2800  $m/z$ . Observed ESI<sup>-</sup> signals and their retention times are listed in **Table S2**.

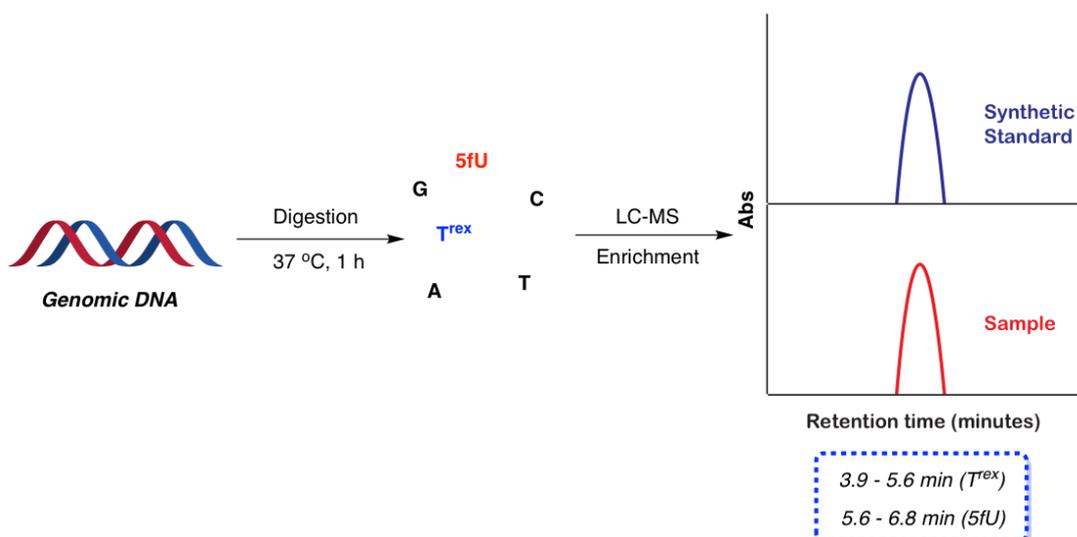


**Figure S10.** Schematic of lyophilisation control on 5fU-ODN I (10 mer), with analyses performed by general oligomer LC-MS analysis, recorded at 260 nm. % $T^{\text{rex}}$  determined by integration of UV signals. 5fU-ODN represents the 10 mer oligomer with no lyophilisation, thus background levels of  $T^{\text{rex}}$ -ODN. 5fU-ODN\_water represents the 5fU-ODN lyophilised in ultrapure HPLC-grade water. 5fU-ODN\_FA\_water represents the 5fU-ODN lyophilised in ultrapure HPLC-grade water containing 0.1% formic acid (FA).

## Note S6. Identification of T<sup>rex</sup> in Human-Derived Cell Lines

**Genomic DNA extraction.** U2OS cells ( $1.5 \times 10^7$ ) were suspended in a 1:1 ratio of PBS pH 7.2 (Gibco®) and Cell Lysis Solution (Qiagen) with a total volume of 2 mL and this was supplemented with 750 mg mL<sup>-1</sup> RNase A (Qiagen) and inverted for 3 h at room temperature. Following this, 24 U mL<sup>-1</sup> Proteinase K (NEB) was added and the mixture was incubated at room temperature for 18 h. DNA was subsequently purified using phenol:chloroform:isoamyl alcohol, saturated with 10 mM Tris, pH 8.0 and 1 mM EDTA (25:24:1 Sigma) and Phase Lock Gel (5 Prime); precipitated from 75% ethanol, stored at -80 °C for 2 h, washed 3 × with ice-cold 70% ethanol, air-dried for 5 min and then resuspended in ultrapure HPLC-grade water. DNA was quantified and assessed for purity by Nanodrop ND-1000 Spectrophotometer.

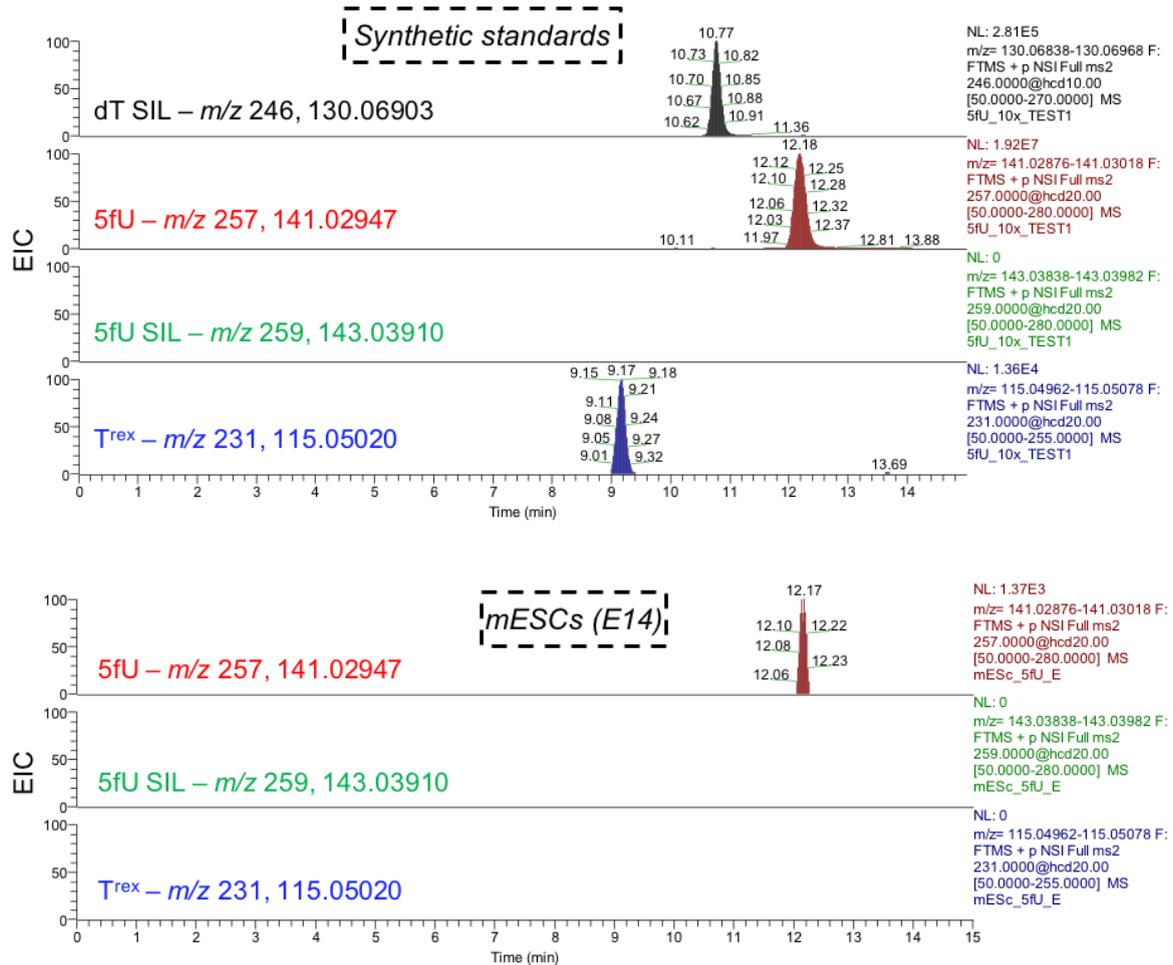
**In-vivo detection of T<sup>rex</sup> in U2OS and HeLa DNA.** 50 - 100 µg of the isolated DNA was digested on a (5 µg reaction scale, i.e. 10-20 individual digestions) into its constituent mononucleosides using 4 units of Nuclease P1 (Sigma-Aldrich) and 10 units of Alkaline Phosphatase (Promega) at 37 °C for 1 hour in 50 µL of reaction buffer containing 5 mM ZnCl<sub>2</sub> and 50 mM NH<sub>4</sub>OAc. The digestions were then pooled and purified through a 10 KDa Amicon® Ultra-0.5 mL centrifugal size exclusion filter (Merck Millipore); the filtrate was concentrated *in vacuo* and reconstituted in 80 µL ultrapure HPLC-grade water. The digested DNA was then enriched offline (**Figure S11**) for T<sup>rex</sup> (3.9 - 5.6 min) and 5fU (5.6 - 6.8 min) on a Bruker amaZon X HPLC system, using a Waters ACQUITY UPLC® HSS T3 column (2.1 × 100 mm, 1.8 µm particle size), at a flow rate of 0.35 mL min<sup>-1</sup> in a 0.1% formic acid in H<sub>2</sub>O (solvent A) and 0.1% formic acid in MeCN (solvent B) solvent system. After sample injection, the gradient was maintained for 4.0 min at 2% B, after which it was ramped to 100% B over 1.0 min, maintained for 1.5 min and then immediately returned to 2% B and maintained for 7.5 min. The enriched fractions were lyophilised and reconstituted in 40 µL ultrapure HPLC-grade water, spiked with an internal standard of 5fU SIL (<sup>13</sup>CD-5fU, +2 Da), at a final concentration of 2.5 nM.



**Figure S11.** Workflow for offline nucleoside enrichment (against synthetic standards) of DNA digests, to enhance signal-to-noise ratio and thus detection of low abundance DNA modifications by HPLC-MS/MS.

Nucleobase detection of digested (and enriched) DNA was analysed by subjection to NanoHPLC-MS/MS analysis on a Q-Exactive<sup>TM</sup> Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific). This was equipped with a nanospray ionisation source, coupled with an Ultimate 3000 RSLCnano HPLC system (Dionex), installed with 2 × nanoviper connectors (75 μm × 150 mm) and a commercially sourced Hypercarb KAPPA column (30 mm × 0.18 mm ID, 5 μm particle size, Thermo Fisher). Samples were injected onto the column at a volume of 5 μL *via* the loading pump in 95:5 0.1% formic acid H<sub>2</sub>O:MeCN with a flow rate of 2 μL min<sup>-1</sup>. A valve switch to the NC pump followed after 5 min; the flow rate was set to 1.5 μL min<sup>-1</sup> and run with a gradient of 95:5 to 0:100 0.1% formic acid H<sub>2</sub>O:MeCN with a run-time of 19 min. Parent ions of *m/z* 231, 257 and 259 were fragmented in positive ion mode with 10% normalised collision energy in parallel-reaction monitoring (PRM) mode. MS<sup>2</sup> resolution was 35,000 with an AGC target of 2e<sup>5</sup>, a maximum injection time of 100 ms and an isolation window of 1.0 *m/z*. Extracted ion chromatograms (±5 ppm) of base fragments 115.05020, 141.02947 and 143.03910 corresponding to T<sup>rex</sup>, 5fU and 5fU SIL respectively were used for the detection. Spectra were processed using the XCalibur QuanBrowser software (Thermo Fisher).

An additional example to HEK293T cells, shown in **Figure 3**, lacking a signal for T<sup>rex</sup> by NanoHPLC-MS/MS is shown below, along with the synthetic standards of dT-SIL, 5fU, 5fU-SIL and T<sup>rex</sup>.



NanoHPLC-MS/MS traces for synthetic standards of dT-SIL (black), 5fU (maroon), 5fU-SIL (green) and T<sup>rex</sup> (blue) by targeting for  $m/z$  130.06903 as a fragment of  $m/z$  246;  $m/z$  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. Also shown is the NanoHPLC-MS/MS trace for genomic DNA digested, enriched and analysed from mESCs (E14), highlighting the absence of a signal for T<sup>rex</sup>. NL represents the ion count (signal intensity) of the corresponding targeted mass.

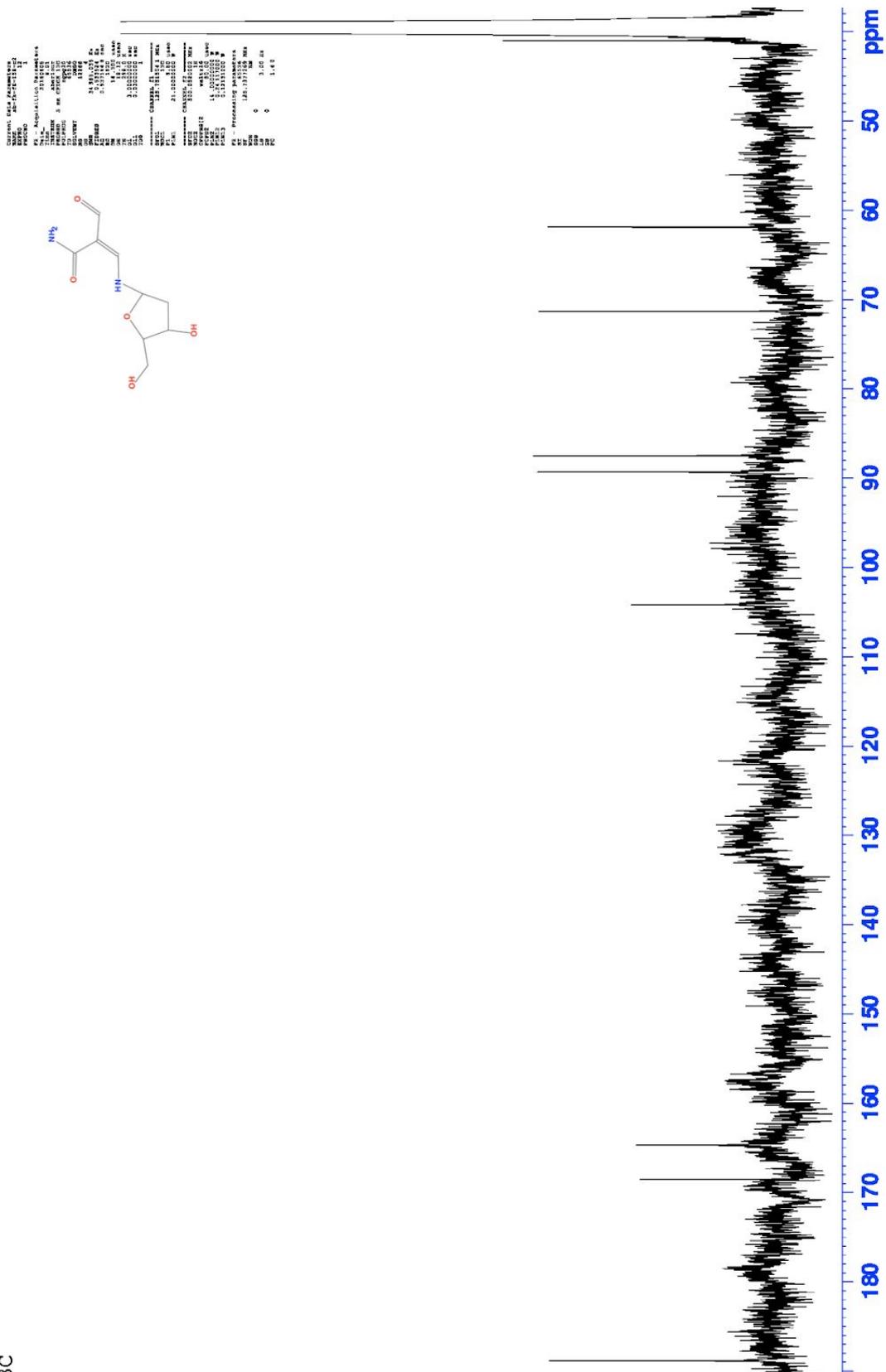
The T<sup>rex</sup> peaks observed in the U2OS and HeLa NanoHPLC-MS/MS traces are not a result of artificial T<sup>rex</sup> formation during digestion of the genomic DNA. The EIC ratio of T<sup>rex</sup> to 5fU in these genomic samples is much higher than the theoretical ratio predicted from the oligonucleotide digestion under the same time conditions.



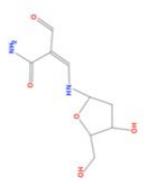


<sup>13</sup>C

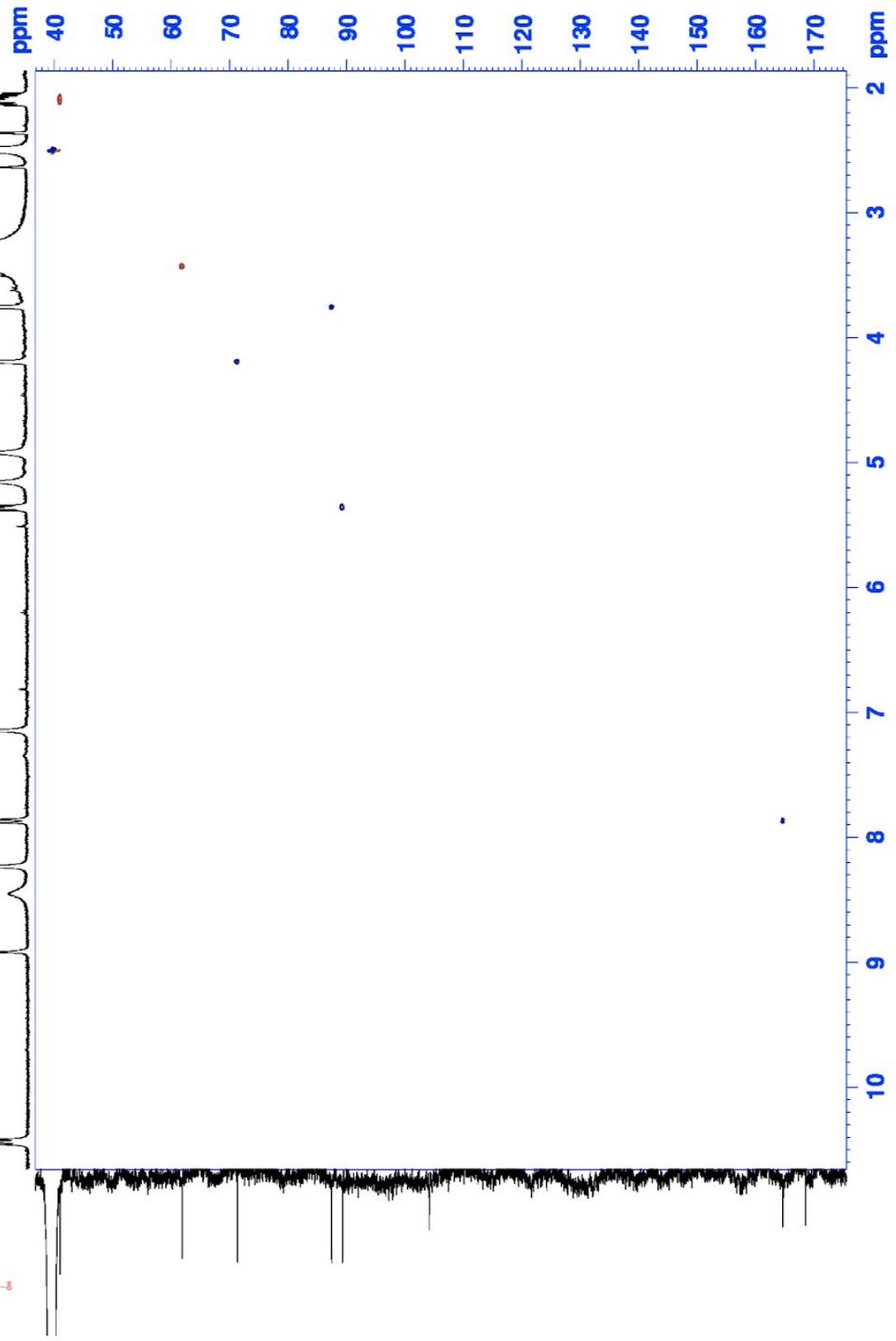
<sup>13</sup>C NMR spectrum of dT<sup>rex</sup> nucleoside. The baseline fluctuation is because of cryo-probe wobble correction.



HSQC



Current File Parameters  
NAME: hsqc-131117  
PROC: hsqc-131117  
P1 - Acquisition Parameters  
Date\_ Time: 21.11.2016 11:54:00  
PROBHD: 5 mm QNP 1H/1  
PULPROG: zgpg30  
WALTZ16: waltz16  
NUC1: 13C  
NUC2: 1H  
SFO: 125.760133 MHz  
SFO2: 500.136447 MHz  
AQ: 0.199500000 sec  
RG: 655.000000  
RG2: 327.500000  
SI: 655  
SF: 125.760133 MHz  
WDW: EM  
SSB: 0  
LB: 3.000000 Hz  
GB: 0  
PC: 1.000000  
P2 - Processing parameters  
SI: 655  
SF: 125.760133 MHz  
WDW: EM  
SSB: 0  
LB: 3.000000 Hz  
GB: 0  
PC: 1.000000  
P3 - Acquisition Parameters  
Date\_ Time: 21.11.2016 11:54:00  
PROBHD: 5 mm QNP 1H/1  
PULPROG: zgpg30  
WALTZ16: waltz16  
NUC1: 13C  
NUC2: 1H  
SFO: 125.760133 MHz  
SFO2: 500.136447 MHz  
AQ: 0.199500000 sec  
RG: 655.000000  
RG2: 327.500000  
SI: 655  
SF: 125.760133 MHz  
WDW: EM  
SSB: 0  
LB: 3.000000 Hz  
GB: 0  
PC: 1.000000  
P4 - Processing parameters  
SI: 655  
SF: 125.760133 MHz  
WDW: EM  
SSB: 0  
LB: 3.000000 Hz  
GB: 0  
PC: 1.000000

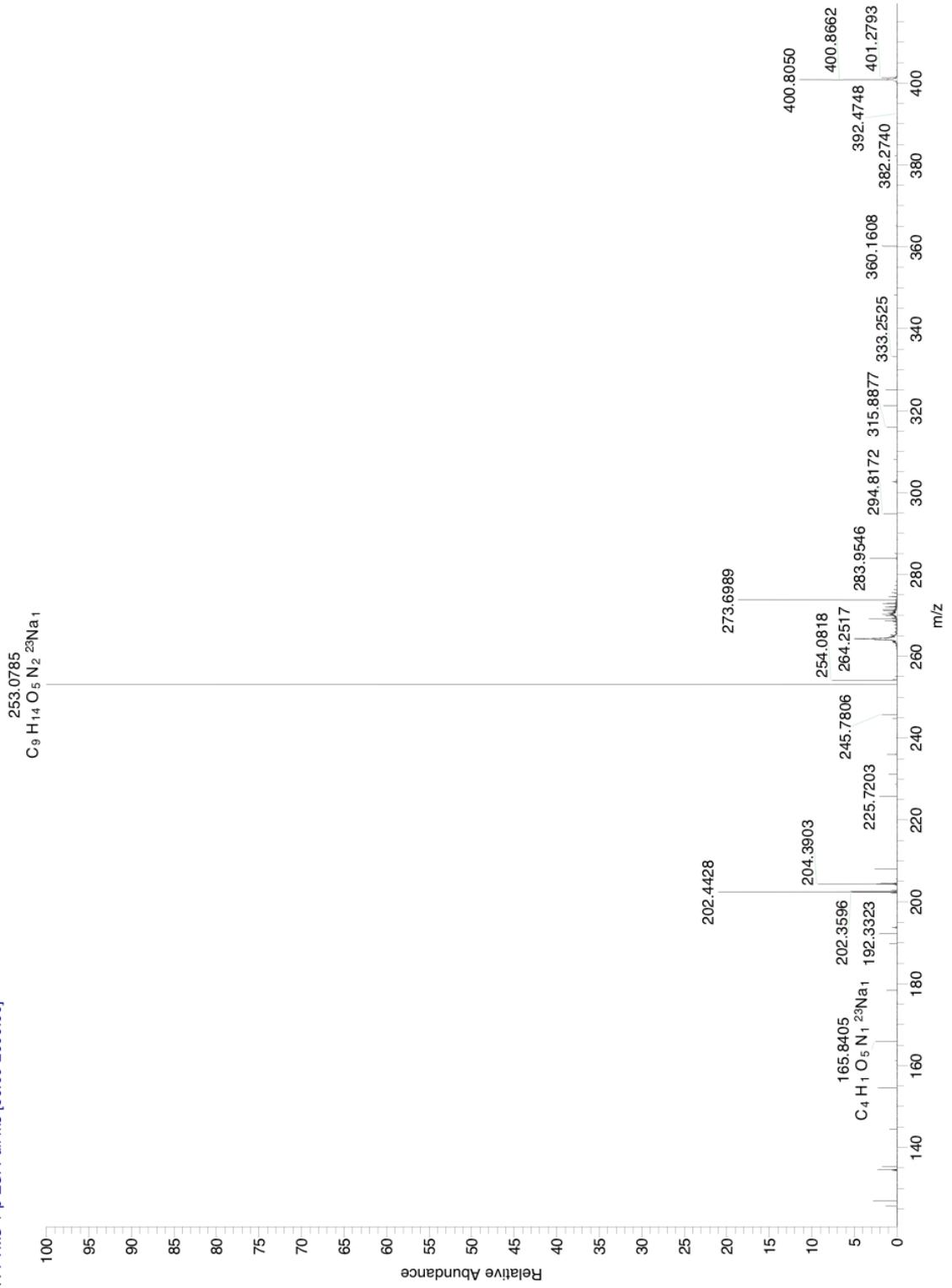


NMR HSQC spectrum of dT<sup>rex</sup> nucleoside.

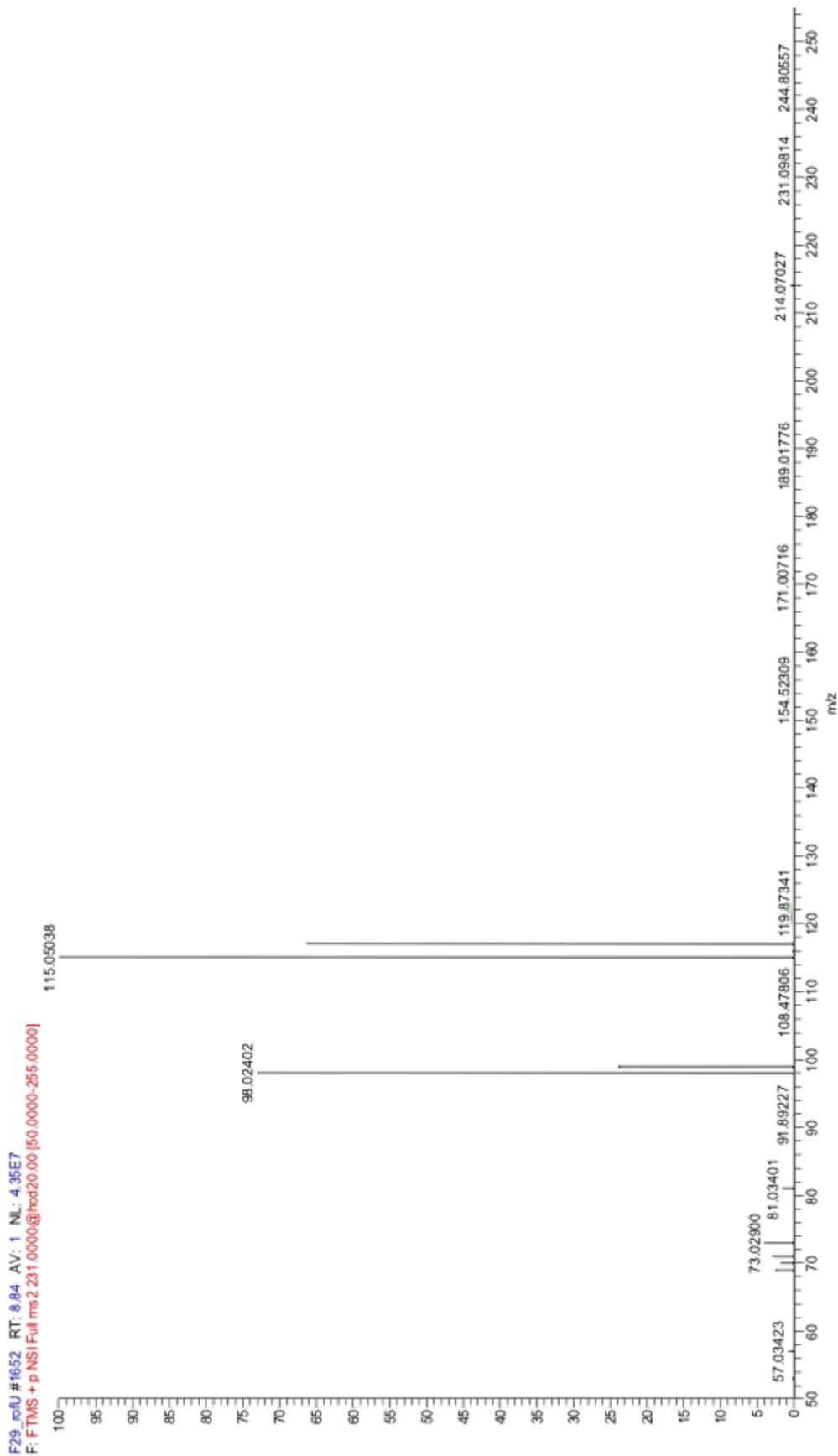


C:\Xcalibur\...SB\_39653\_181026121314 19 Nov 18 16:30:37 SB\_39653

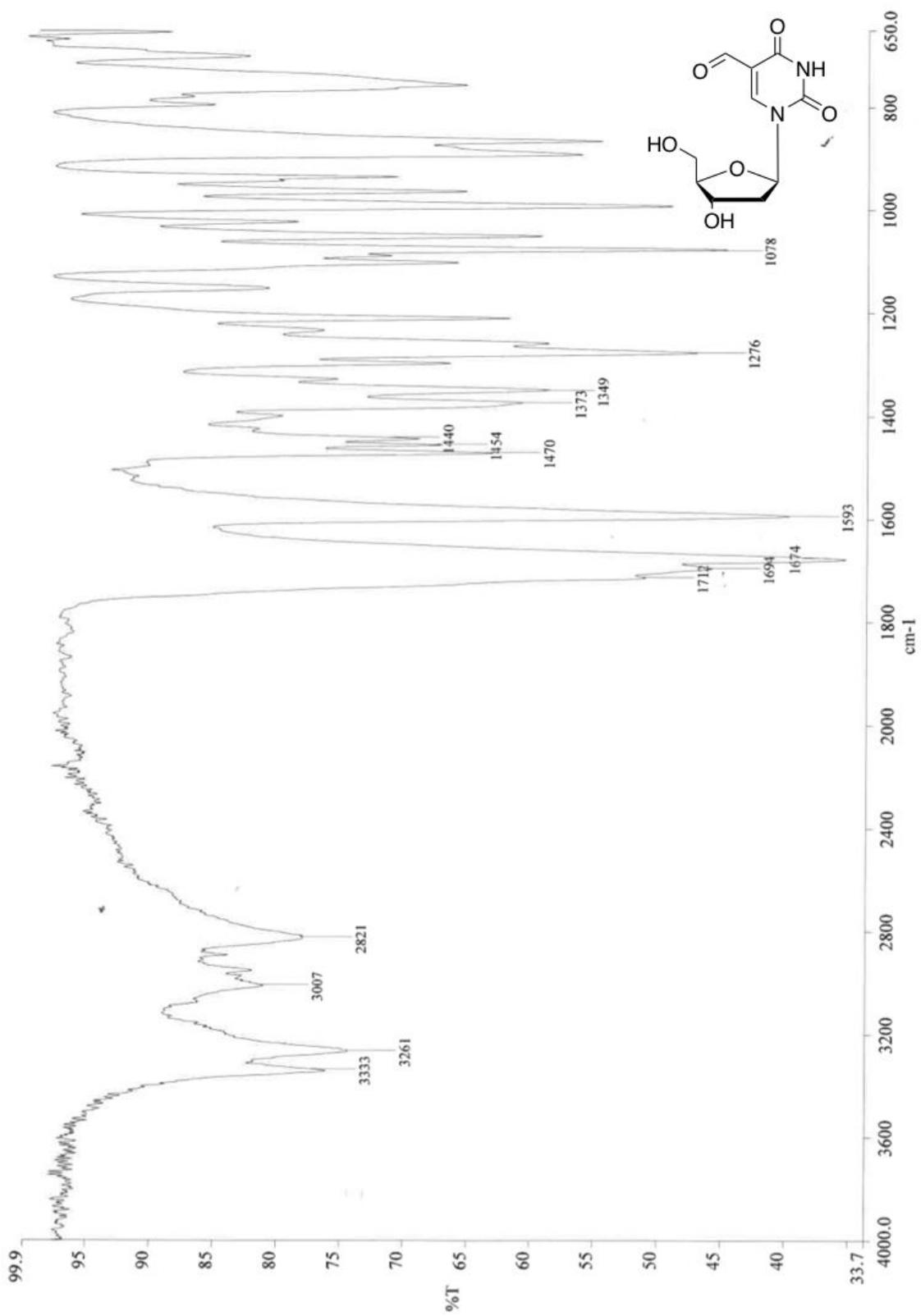
SB\_39653\_181026121314.#3-37 RT: 0.06-0.91 AV: 35 NL: 4.48E5  
T: FTMS + p ESI Full ms [50.00-2000.00]



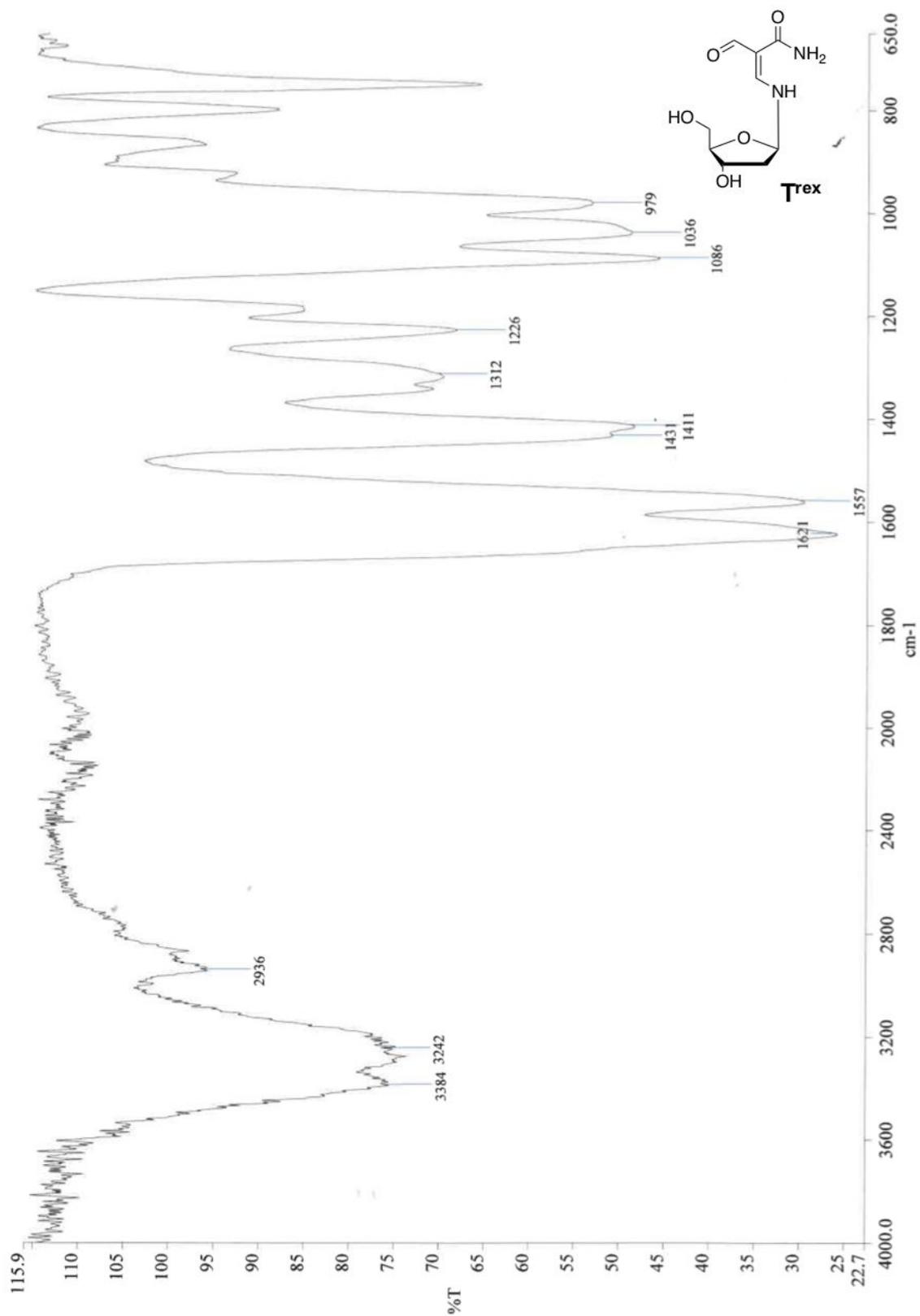
HRMS spectrum in positive ion mode of dT<sup>tex</sup>.



HRMS/MS spectrum in positive ion mode of dT<sup>rex</sup>.



FT-IR spectrum of 5-formyl-2'-deoxyuridine.



FT-IR spectrum of dT<sup>rex</sup>.

## References

- [1] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618-622.
- [2] M. Head-Gordon, J. A. Pople, M. J. Frisch, *Chem. Phys. Lett.* **1988**, *153*, 503-506.
- [3] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650-654.
- [4] H. A. De Abreu, W. B. De Almeida, H. A. Duarte, *Chem. Phys. Lett.* **2004**, *383*, 47-52.
- [5] Z. Q. Chen, C. H. Zhang, C. K. Kim, Y. Xue, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6471-6483.
- [6] N. Russo, M. Toscano, A. Grand, *J. Phys. Chem. A* **2003**, *107*, 11533-11538.
- [7] J. P. Merrick, D. Moran, L. Radom, *J. Phys. Chem. A* **2007**, *111*, 11683-11700.
- [8] M. D. Tissandier, K. A. Cowen, W. Y. Feng, et al., *J. Phys. Chem. A* **1998**, *102*, 7787-7794.
- [9] V. S. Bryantsev, M. S. Diallo, W. A. Goddard, *J. Phys. Chem. A* **2007**, *111*, 4422-4430.
- [10] M. Yoshida, K. Makino, H. Morita, et al., *Nucleic Acids Res.* **1997**, *25*, 1570-1577.

## Author Contributions

ABS conceived the computational study and performed the calculations. AM planned and carried out the experimental studies, including mass spectrometry analysis. FK conducted the initial small-scale synthesis of the synthetic standard. SB supervised the research. ABS, AM and SB, wrote the manuscript with the contributions from all authors. All authors have given approval to the final version of the manuscript.