



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## 2,6-diaminopurine promotes repair of DNA lesions under prebiotic conditions

**Citation for published version:**

Szabla, R, Zdrowowicz, M, Spisz, P, Green, NJ, Stadlbauer, P, Kruse, H, Šponer, J & Rak, J 2021, '2,6-diaminopurine promotes repair of DNA lesions under prebiotic conditions', *Nature Communications*, vol. 12, no. 1, 3018. <https://doi.org/10.1038/s41467-021-23300-y>

**Digital Object Identifier (DOI):**

[10.1038/s41467-021-23300-y](https://doi.org/10.1038/s41467-021-23300-y)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Nature Communications

**General rights**










Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# 2,6-diaminopurine promotes repair of DNA lesions under prebiotic conditions

Rafał Szabla <sup>1,2</sup>, Magdalena Zdrowowicz <sup>3</sup>, Paulina Spisz <sup>3</sup>, Nicholas J. Green <sup>4</sup>, Petr Stadlbauer<sup>5</sup>, Holger Kruse <sup>5</sup>, Jiří Šponer <sup>5</sup> & Janusz Rak <sup>3</sup>

High-yielding and selective prebiotic syntheses of RNA and DNA nucleotides involve UV irradiation to promote the key reaction steps and eradicate biologically irrelevant isomers. While these syntheses were likely enabled by UV-rich prebiotic environment, UV-induced formation of photodamages in polymeric nucleic acids, such as cyclobutane pyrimidine dimers (CPDs), remains the key unresolved issue for the origins of RNA and DNA on Earth. Here, we demonstrate that substitution of adenine with 2,6-diaminopurine enables repair of CPDs with yields reaching 92%. This substantial self-repairing activity originates from excellent electron donating properties of 2,6-diaminopurine in nucleic acid strands. We also show that the deoxyribonucleosides of 2,6-diaminopurine and adenine can be formed under the same prebiotic conditions. Considering that 2,6-diaminopurine was previously shown to increase the rate of nonenzymatic RNA replication, this nucleobase could have played critical roles in the formation of functional and photostable RNA/DNA oligomers in UV-rich prebiotic environments.

<sup>1</sup>EaStCHEM School of Chemistry, University of Edinburgh, Edinburgh, UK. <sup>2</sup>Institute of Physics, Polish Academy of Sciences, Warsaw, Poland. <sup>3</sup>Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland. <sup>4</sup>MRC Laboratory of Molecular Biology, Cambridge, UK. <sup>5</sup>Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic. ✉email: [rafal.szabla@ed.ac.uk](mailto:rafal.szabla@ed.ac.uk); [magdalena.zdrowowicz@ug.edu.pl](mailto:magdalena.zdrowowicz@ug.edu.pl)

Exposure of nucleic acid strands to ultraviolet (UV) irradiation results in the formation of lesions which effectively suppress the key biological functions of DNA and may result in mutagenesis, immunosuppression, and skin cancer<sup>1–3</sup>. Cyclobutane pyrimidine dimers (CPDs) are among the most frequent photolesions and are formed in a [2 + 2] photocycloaddition reaction between the C=C bonds of adjacent pyrimidine bases<sup>4–6</sup>. While CPDs have a significant effect on the structure of the strand and its recognition, Nature developed complex machinery, i.e., photolyase enzymes, which locate the lesion and initiate the repair with photoinduced electron transfer from a flavin chromophore (see Fig. 1 and Supplementary Fig. 1 in the SI)<sup>7,8</sup>. However, photolyases evolved only after early living organisms emerged and, therefore, could not counteract photo-damage formation in the very first oligonucleotides on Earth<sup>9</sup>. Therefore, more primitive molecular mechanisms of photo-damage repair are likely to have protected the earliest biological systems from the ever-present threat of excessive UV-induced genetic mutation.

Recent advances in nonenzymatic template copying of nucleic acids<sup>10–13</sup> and prebiotic formation of their components<sup>14–19</sup> have brought us near to proposing a credible scenario for the emergence of informational polymers on early Earth. Among several different geochemical scenarios for the formation of purine and pyrimidine nucleosides<sup>14–19</sup>, involvement of UV light offers highest selectivity by driving the key chemical transformations and destroying biologically irrelevant stereoisomers<sup>20–24</sup>. While high prebiotic UV fluxes could pass through the anoxic Archean atmosphere and support the synthesis and selection of key DNA and RNA building blocks<sup>25</sup>, intense UV irradiation posed threat to the integrity of their oligomers. The resulting lack of compatibility between UV-assisted formation of nucleosides and the longevity of their oligomers remains one of the key unresolved issues for the origins of informational polymers on Earth. This led to the suggestion that specific DNA sequences could act as potent electron transmitters and allow photochemical charge separation leading to photoreversal of CPDs<sup>26</sup>. Nevertheless, the most photostable sequences like GATT, were demonstrated to repair only up to 25% of thymine CPDs (denoted as T=T)<sup>26</sup>. In addition, 8-oxo-purine bases, which could act as good electron donors and promote efficient photoinduced CPD repair<sup>27</sup>, were shown to nearly completely inhibit nonenzymatic template copying of nucleic acids<sup>13</sup>.

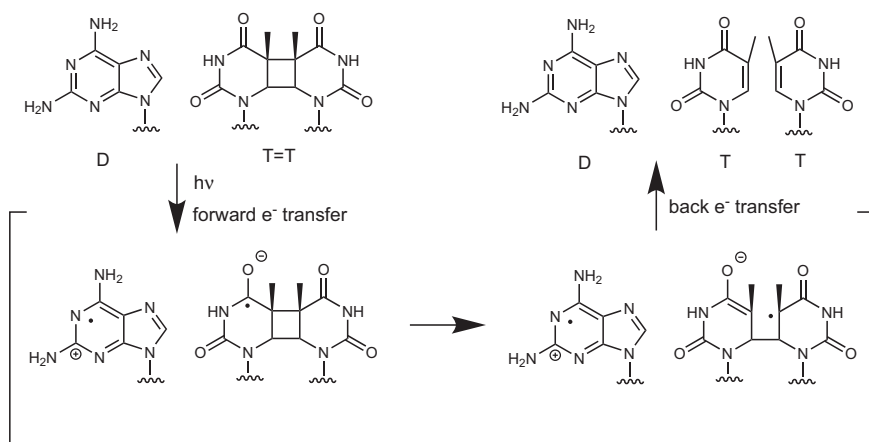
Previous theoretical works reported that charge-transfer (CT) electronic states in the GAT=T sequence are characterized by high vertical excitation energies and populated only indirectly,

outside of the Franck–Condon region of the oligomer<sup>28,29</sup>. Therefore, involvement of an alternative nucleoside acting as a prebiotic flavin mimic could significantly lower the excitation energies of CT states and increase the self-repairing activity of oligonucleotides.

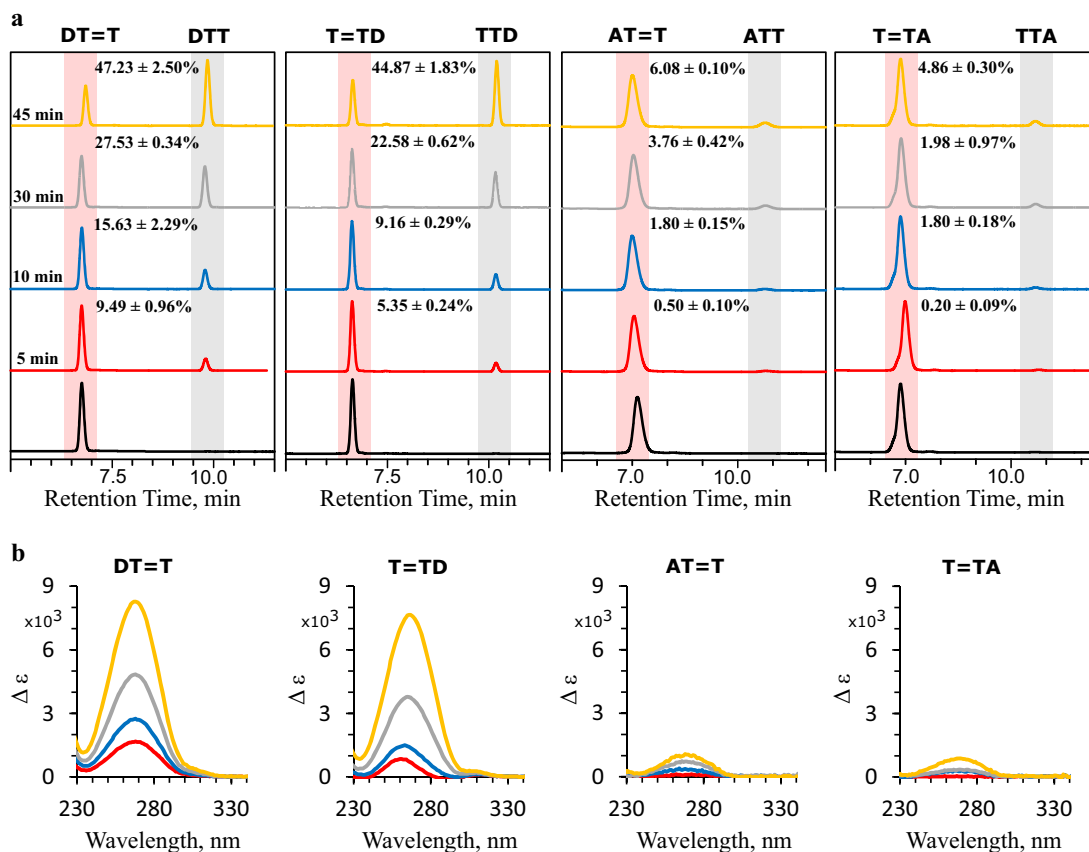
Here, we demonstrate that 2,6-diaminopurine (Dap) satisfies this criterion and enables highly efficient photoreversal of T=T CPDs. The subjects of our study were canonical and modified DNA trinucleotides containing a purine and two neighboring thymine (Thy) bases as model systems for photodamage repair. Consideration of trinucleotides consisting of deoxyribonucleosides is also consistent with the recent suggestion that DNA and RNA nucleosides could have emerged simultaneously on prebiotic Earth<sup>24</sup>. The self-repairing activity is manifested by higher T=T photoreversal quantum yields and electron transfer rates in the damaged trinucleotides containing Dap, when compared to canonical trimers containing adenine (Ade). It is worth noting that Dap is an excellent Watson–Crick base pairing partner for Thy and was shown to entirely replace Ade in the genome of phage S-2L of the cyanobacterium *S. elongatus*<sup>30,31</sup>. Furthermore, the deoxyriboside of Dap (D) was demonstrated to undergo prebiotic phosphorylation with diamidophosphate in the presence of 2-aminoimidazole, which could be followed by oligomerization with biological deoxyribosides<sup>32</sup>. Given that Dap could also enhance the rate of nonenzymatic RNA template copying<sup>33–35</sup>, we provide a possible bridge between the UV-assisted synthesis of nucleotides and formation of their first self-replicating oligomers.

## Results

**Low energies of charge-transfer states lead to high efficiency of self-repair.** Calculations of vertical and adiabatic ionization energies (VIE and AIE) provide a good first estimate of electron-donating properties of nucleobases<sup>36</sup>. Our MP2 calculations indicated that the deoxyriboside of 2,6-diaminopurine (D) has lower VIE and AIE by over 1.8 eV when compared to canonical deoxyribosides of adenine (A) and guanine (G) (see Supplementary Table 3 in the SI). Consequently, we decided to computationally investigate the photophysical properties of DNA trinucleotides containing a CPD with two dimerized thymine bases (T=T) and D or A either on the 5' or 3' terminus. For this purpose, we carried out 20- $\mu$ s long molecular dynamics (MD) simulations of these damaged trimers to identify their most representative stacked conformers and obtain their averaged structures (see Supplementary Figs. 21 and 22 in the SI). We further performed QM/MM calculations in an aqueous



**Fig. 1 Schematic representation of the self-repair mechanism of T=T dimers promoted by photoinduced electron transfer from 2,6-diaminopurine (D).** This mechanism is analogous to the enzymatic repair induced by the flavin chromophore acting as the electron donor (Supplementary Fig. 1 in the SI).



**Fig. 2 UV-irradiation experiments** ( $\lambda_{\text{exc}} = 280 \text{ nm}$ ,  $80 \mu\text{W cm}^{-2}$ ) of damaged DNA trinucleotides containing a T=T dimer and A or D. **a** Analytical HPLC analysis of irradiated DT=T, T=TD, AT=T, and T=TA performed after different irradiation times. The peaks at shorter retention times correspond to damaged trinucleotides, while the peaks at longer retentions correspond to the repaired photoproducts. **b** Differential absorption UV spectra demonstrating changes of the spectrum of the irradiated material at different irradiation times. Both trinucleotides containing D exhibit significant increase of absorption at 264 nm, which corresponds to the absorption maximum of repaired T chromophore. Only minor increase of absorption at 264 nm can be observed for the trinucleotides containing A.

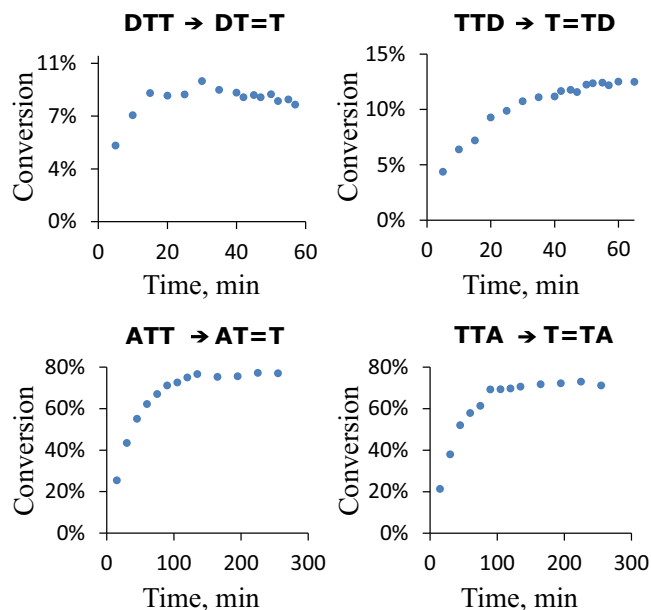
environment with the algebraic diagrammatic construction to the second-order [ADC(2)]<sup>37,38</sup> QM method, which was shown to provide accurate vertical excitation energies of the key photoactive states in DNA oligomers<sup>39</sup>, and the TZVP basis set (see the “Methods” section for more details). As expected, the repairing  $D^{\bullet+}T^{\bullet-}$  state has the excitation energy of 5.02 eV and exhibits configuration mixing with the bright (UV-absorbing)  $\pi\pi^*$  state on D that is lower in energy by merely 0.14 eV. Even though the repairing  $T^{\bullet-}D^{\bullet+}$  state has the vertical excitation energy of 5.62 eV, the bright  $S_1(\pi\pi^*)$  state on D also has a notable contribution from this CT configuration, which implies good self-repairing potential. In contrast, the analogous CT states of AT=T and T=TA have significantly higher vertical excitation energies (5.74 and 5.80 eV, respectively) and do not exhibit any clear mixing with the low-energy  $\pi\pi^*$  and  $n\pi^*$  states on A (see Supplementary Table 4 in the SI).

Encouraged by these predictions we performed UV-irradiations of the four damaged trinucleotides at 280 nm and analyzed the mixture at different time intervals with HPLC experiments and UV-vis spectroscopy (Fig. 2a). At this wavelength, over 95% of absorbed photons excite the purine base, whereas the extinction coefficients of the CPDs are very small<sup>40</sup>. As shown in Fig. 2a, irradiations of trimers containing D lead to rapid formation of a photoproduct with corresponding HPLC peaks visible at longer retention times than the damaged trimers. The accumulation of this photoproduct at different irradiation times is proportional to the depletion of the starting

material. The differential absorption spectra of the UV-irradiated samples generated with respect to the starting materials (Fig. 2b) demonstrate a gradual increase of absorption at 264 nm, which corresponds to the absorption maximum of recovered Thy chromophores (see Fig. 1 for schematic representation of the self-repair process). In addition, we have validated by LC-MS that the photoproduct corresponds to the repaired Thy bases within the trimers (see section S2.1 in the SI).

After 45 min of continuous irradiation, DT=T and T=TD underwent thymine recovery in 47.2% and 44.9% yields, respectively (Fig. 2a). In contrast, equivalent exposure of AT=T and T=TA to UV light allowed for self-repair of only 6.1% and 4.9% of the corresponding starting materials. Therefore, presence of the additional amino group in the 2 position of the purine ring significantly enhanced the self-repairing properties. This is also reflected by an order of magnitude higher self-repair quantum yields observed for damaged trimers containing D (see Supplementary Table 2 in the SI). In reality, the overall photostability of specific DNA oligomers is the resultant of rates of various interplaying photoinduced phenomena, including formation of CPDs, self-repair, and direct benign photorelaxation channels of nucleobases.

To provide a better estimate of the photostability of our trinucleotides we performed continuous irradiations of their undamaged forms in the UV-B spectral range to reach the photostationary equilibria (Fig. 3). The irradiation of DTT reached the photostationary state in <20 min with only 8% of



**Fig. 3** Continuous UV-irradiation experiments of undamaged DNA trinucleotides in the UV-B spectral range. Samples were exposed to continuous UV-B flux ( $1.95 \text{ mW cm}^{-2}$ ) to yield photostationary equilibria of T=T dimer formation and self-repair. Analysis of photoproducts was performed by HPLC and % of conversion of undamaged DNA trimers into the damaged trinucleotides was calculated by integration of the respective peak areas (see the SI and the Methods section for more details).

CPDs formed, leaving over 90% of the trinucleotide intact (see Fig. 3 and Supplementary Fig. 18 in the SI). TTD also exhibits very high photostability with CPDs formed in <13% yield after the photostationary state is reached. On the contrary, continuous irradiations of ATT and TTA resulted in the formation of T=T dimers in 72% and 76% yields, respectively. This indicates that majority of the material containing the ATT and TTA sequences would be prevented from performing key prebiotically relevant functions such as self-replication.

**Molecular mechanism of self-repair in modified and canonical trimers.** To scrutinize the above differences in photoreactivity of the damaged trimers, we performed further QM/MM calculations with ADC(2) as the QM method and the TZVP basis set. As shown in Fig. 4a, direct electron transfer from UV-excited D to the T=T dimer is energetically favorable and the corresponding  $S_1(\text{CT})$  minimum lies  $\sim 1.5 \text{ eV}$  lower in energy than the ring-puckered  $S_1$  minimum of photoexcited D. The relative energies of the  $D^+T=T^{\bullet-}$  and  $D^*T=T$  excited-state minima with respect to the ground-state minimum (adiabatic excitation energies) amount to 2.95 and 4.42 eV, respectively. Apart from low vertical and adiabatic excitation energies of the repairing  $D^+T=T^{\bullet-}$  state, electron transfer is facilitated by relatively high energy and sloped topography of the crossing between the  $D^*T=T$  state and the ground state, which lies 0.4 eV above the respective  $S_1$  minimum and hinders direct photorelaxation of D. Consequently, UV-excited D is much more likely to transfer an electron to the T=T dimer than undergo direct photorelaxation in DT=T. Similarly, the somewhat lower efficiency of self-repair observed for T=TD can be explained in terms of higher vertical excitation energy of the  $T=T^{\bullet-}D^+$  state and less sloped topography of the  $S_1/S_0$  state crossing<sup>41</sup> enabling smoother direct photorelaxation of D (Fig. 4b). The relatively high yields of self-repair in this trinucleotide are the consequence of a very low adiabatic excitation energy of the  $T=T^{\bullet-}D^+$  state (2.52 eV). Finally, once the

negative charge is transferred to the T=T dimer, it induces breaking of the C5–C5 bond in the  $S_1(\text{CT})/S_0$  state crossing. The remaining C6–C6 bond of the CPD is opened spontaneously in the vibrationally hot electronic ground state<sup>28</sup>, as schematically demonstrated in Fig. 1.

Negligible self-repairing activity of AT=T and T=TA trimers derives from high adiabatic and vertical excitation energies of the  $A^+T=T^{\bullet-}$  and  $T=T^{\bullet-}A^+$  states. In particular, the  $S_1 T=T^{\bullet-}A^+$  minimum is lower in energy than the  $S_1 T=TA^*$  minimum corresponding to photoexcited A base by merely 0.10 eV (see Fig. 4d). We were unable to locate the  $S_1$  minimum for the  $A^+T=T^{\bullet-}$  state, which is the result of an even higher adiabatic excitation energy for this state and lower energies of other valence excitations outside the Frank-Condon region of T=TA (see the SI for discussion). Furthermore, A may undergo much more efficient photorelaxation than D, which is manifested by a peaked  $S_1/S_0$  state crossing separated from the respective  $S_1$  minima by a modest energy barrier of merely 0.3 eV in both AT=T and T=TA trimers (Fig. 4c and d). It is worth noting that while our estimate is an upper bound to the barrier energy, this process is driven by ultrafast C2-H tilting motion, which can be facilitated by H atom tunneling.

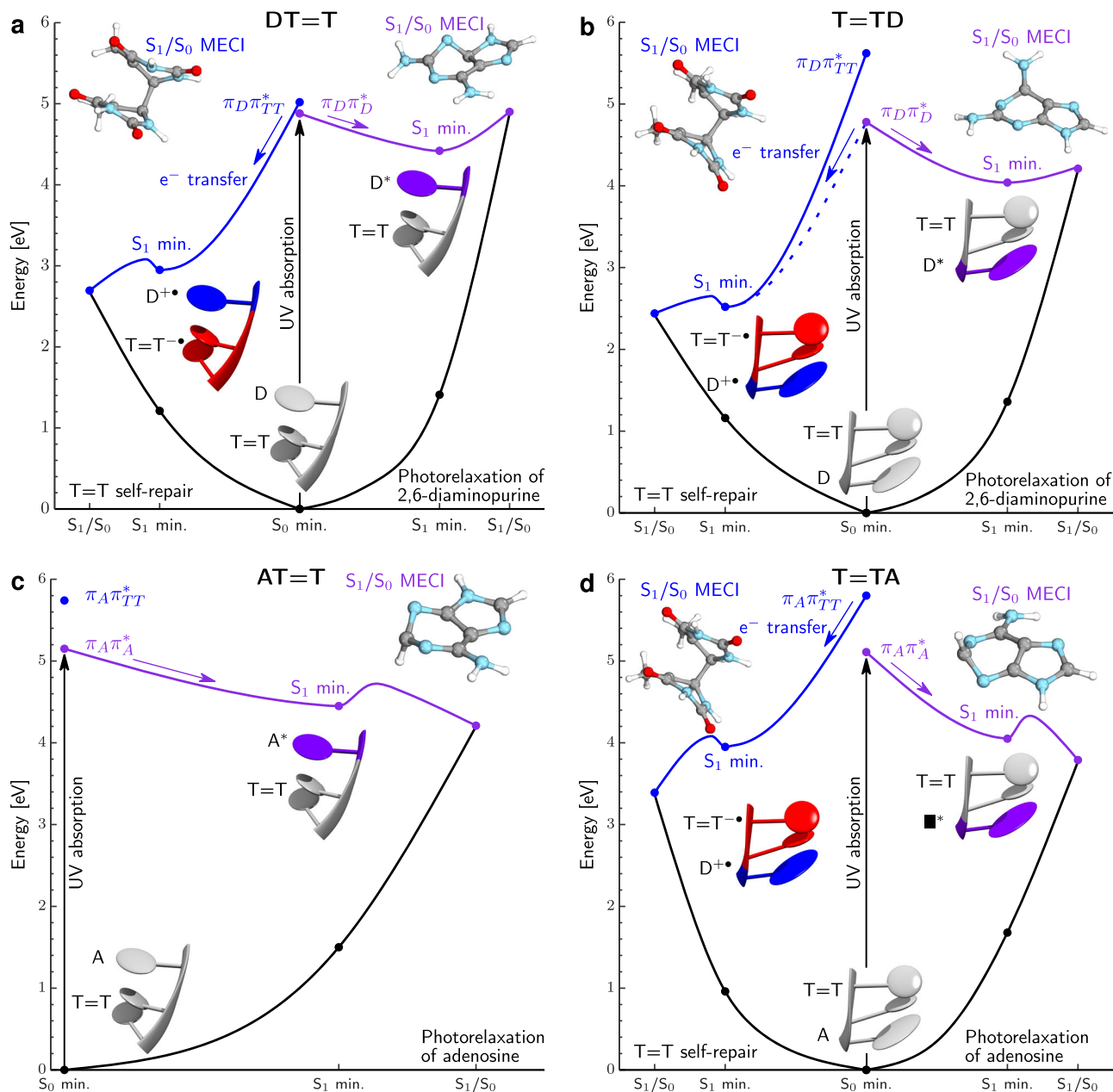
As observed for canonical and undamaged DNA trimers and the GAT=T tetramer, the energies of various CT states are strongly affected by the structural arrangement of the oligomer<sup>28,42</sup>. In particular, Lee and Matsika<sup>42</sup> showed that excitation energies of CT states are directly proportional to the distance between the donor–acceptor pair. Similarly, we assign the very low vertical excitation energy of the CT state in the DT=T trimer to a practically perfect spatial alignment of the internal Thy and Dap bases and the short distance between them (see section S2.2 in the SI for detailed discussion). Consequently, higher vertical excitation energy of the CT state in the T=TD trimer (by  $\sim 0.6 \text{ eV}$ ) is consistent with longer distance between the internal Thy and Dap bases. However, we do not see a straightforward correlation between the distance separating the donor and acceptor bases and the adiabatic excitation energies of the  $D^+T=T^{\bullet-}$  and  $T=T^{\bullet-}D^+$  states (CT minima).

We provide further interpretation of our experimental observations based on calculations of approximate charge-transfer rates within the Marcus theory of electron transfer and considering couplings between locally excited states on purine bases and CT states, derived from the Boys localization approach<sup>43</sup> (see Supplementary Table 5 and the associated discussion in the SI for more details). The estimated rate constants calculated using the geometries and energies from Fig. 4 indicate that charge separation in photoexcited D-containing trimers occurs on a timescale of hundreds of femtoseconds and could indeed outcompete the direct photorelaxation of D (see Table 1). In contrast, the estimated timescale of electron transfer from excited A to T=T is over an order of magnitude longer than in the case of trimers containing D ( $\sim 5.82 \text{ ps}$  for T=TA). Considering that UV-excited A undergoes direct photodeactivation to the electronic ground state within  $<2.0 \text{ ps}$ , its excited-state lifetime is too short to permit photoreversal of the T=T dimer<sup>44</sup>.

#### Self-repair of internal CPDs in longer nucleic acid oligomers.

The above results demonstrate that the photochemistry of damaged DNA trimers is characterized by the competition between purine photorelaxation and electron transfer processes and that the presence of D could induce efficient self-repair of CPDs on the termini of oligomers. To explore whether D could induce self-repair of internal CPDs, we performed analogous QM/MM simulations for the ADT=TA pentamer having an adenosine residue on each terminus. To examine the excited-state





**Fig. 4** Potential energy profiles calculated with the QM<sub>bases</sub>/MM approach for the key photorelaxation mechanisms of all four studied trinucleotides. **a** Blue curve corresponds to self-repair of the T=T dimer triggered by the electron transfer from D. The C5-C5 bond is broken in the D<sup>+</sup>T<sup>-</sup>/S<sub>0</sub> state crossing. Direct photorelaxation of D base for DT=T - purple curve. Self-repair is the energetically more favorable process. **b** Analogous photorelaxation channels in T=TD. Dashed blue line indicates mixing between the T=T<sup>-</sup>D<sup>+</sup> and T=TD<sup>\*</sup> states. **c** The T=T<sup>-</sup>A<sup>+</sup> configuration is not energetically available in the S<sub>1</sub> state and only the direct photorelaxation of A is presented. **d** High adiabatic excitation energy of the repairing A<sup>+</sup>T<sup>-</sup> state results in dominant role of direct photorelaxation of A in AT=T. The electronic-structure (QM) calculations were performed at the ADC(2) and MP2 levels with the TZVP basis set. The molecular structures depict the key structural changes for purine bases or the T=T in the S<sub>1</sub>/S<sub>0</sub> minimum-energy conical intersection (MECI).

**Table 1** Electron transfer rate constants and the corresponding timescales of electron transfer from the purine base to the T=T CPD estimated for three of the studied trinucleotides.

Trinucleotide	$k$ [s <sup>-1</sup> ]	$\tau$ [ps]
DT=T	$7.08 \times 10^{12}$	0.141
T=TD	$3.85 \times 10^{12}$	0.260
T=TA	$1.72 \times 10^{11}$	5.820

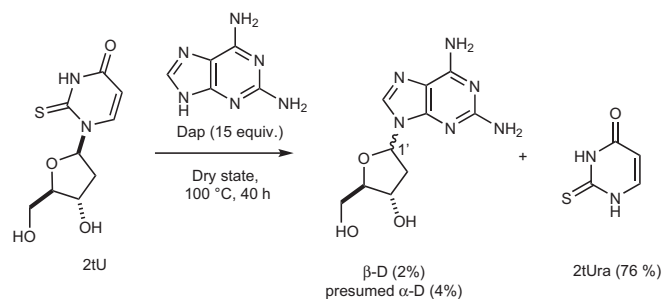
charge-transfer properties of ADT=TA we performed MD simulations and selected a conformer with all five nucleobases stacked, so that it mimics the structural arrangement of even longer single strands. We next performed calculations of vertical excitation energies and optimizations of the key minima of the S<sub>1</sub> excited state at the ADC(2)/def2-SVP level (note that a smaller basis set was applied because of increased computational costs). The lowest-lying excited state calculated for the optimized ground-state geometry of ADT=TA has the energy of 4.90 eV and can be characterized as the bright (UV-absorbing)  $\pi\pi^*$  excitation on D mixed with the AD<sup>+</sup>T<sup>-</sup>A electronic

configuration responsible for self-repair (see Supplementary Fig. 26 and the associated discussion in the SI for more details). In addition, the  $S_1(\text{CT})$  minimum of the  $\text{AD}^+\text{T}=\text{T}^-\text{A}$  state is lower in energy than the  $S_1$  minimum of the locally excited state on D by 0.72 eV. This value, referred to as the driving force of electron transfer, is approximately two times lower in magnitude than in the case of the  $\text{DT}=\text{T}$  and  $\text{T}=\text{TD}$  trimers. However, the CT minimum is still considerably lower in energy than the  $\text{AD}^*\text{T}=\text{TA}$  ( $\pi\pi^*$ ) minimum, which together with the low vertical excitation energy of the  $\text{AD}^+\text{T}=\text{T}^-\text{A}$  ( $\pi_D\pi_{\text{TT}}^*$ ) excitation shows that electron transfer from D to the  $\text{T}=\text{T}$  is also an energetically favorable process in the studied pentamer. Therefore, we expect the stacked conformer of  $\text{ADT}=\text{TA}$  to be involved in highly efficient self-repair of its internal CPD.

Our QM/MM simulations indicate that excited-state electron transfer between the D and  $\text{T}=\text{T}$  fragments is comparably as efficient inside a well-stacked DNA single strand as in the investigated DNA trimers. However, according to the classical MD simulations and conformational analysis, the  $\text{ADT}=\text{TA}$  oligomer spent <5% of time in the fully stacked arrangement and it sampled various partly or fully unstacked conformations for the remaining part of the simulation. In contrast, the stacked conformer of  $\text{AT}=\text{T}$  was populated for more than 50% of time during analogous simulations (see the SI for more details). Therefore, any potential irradiation experiments on the  $\text{ADT}=\text{TA}$  damaged pentamer would largely reflect the averaged photoreactivity of random coil structural arrangements. While such random coils would be present on the early Earth, the results would be less conclusive toward internal CPDs. In contrast, longer and more stable DNA double strands exhibit very complex photochemistry with multiple possible photoproducts<sup>26</sup>, owing to a large number of stacked chromophores. Consequently, we focused here only on short single-stranded oligomers, to clearly disentangle CPD self-repair from other competing photorelaxation channels. Self-repair in a flexible single-stranded DNA oligomer could also operate in partly unstacked conformations. We expect that conformers having a sufficiently close distance between the Dap base and the  $\text{T}=\text{T}$  dimer, and unstacked flanking bases, exhibit very similar photochemistry to the studied DNA trimers. Further experimental work is underway to test these predictions.

**Prebiotically plausible synthesis of  $\beta$ -2,6-diaminopurine 2'-deoxyriboside.** Considering its photostabilizing activity in DNA trimers, we sought to determine whether  $\beta$ -2,6-diaminopurine 2'-deoxyriboside ( $\beta$ -D) might be available on a prebiotic Earth for incorporation into nucleic acids. Recently, Xu et al.<sup>22</sup> reported a proof of principle synthesis of deoxyribosides, including deoxyadenosine, by transglycosylation of 2'-deoxy-2-thiouridine (2tU) with adenine, to provide 4% and 6% of  $\beta$ - and  $\alpha$ -A, respectively. The synthesis, although low yielding, demonstrates that a prebiotic world that contained RNA may also have contained small amounts of deoxyadenosine. Since Dap has previously been synthesized in a prebiotic fashion<sup>45</sup>, and has even been discovered on meteorites<sup>46</sup>, we reasoned that it may also have accumulated and participated in this transglycosylation reaction. Indeed, after evaporating an aqueous mixture of 2tU and Dap (15 equiv.) and heating to 100 °C for 40 h,  $\beta$ -D was formed (2%) along with at least one isomer (probably the  $\alpha$ -diastereomer, 4%) and other minor compounds (see Fig. 5). Thus, the reaction proceeds similarly to that of the reported synthesis of deoxyadenosine, providing similar proof of principle that DNA containing Dap may have indeed been present in a prebiotic world.

**Photostability of D vs A.** UV irradiation was suggested as the key environmental factor, which selected adenosine, guanosine,



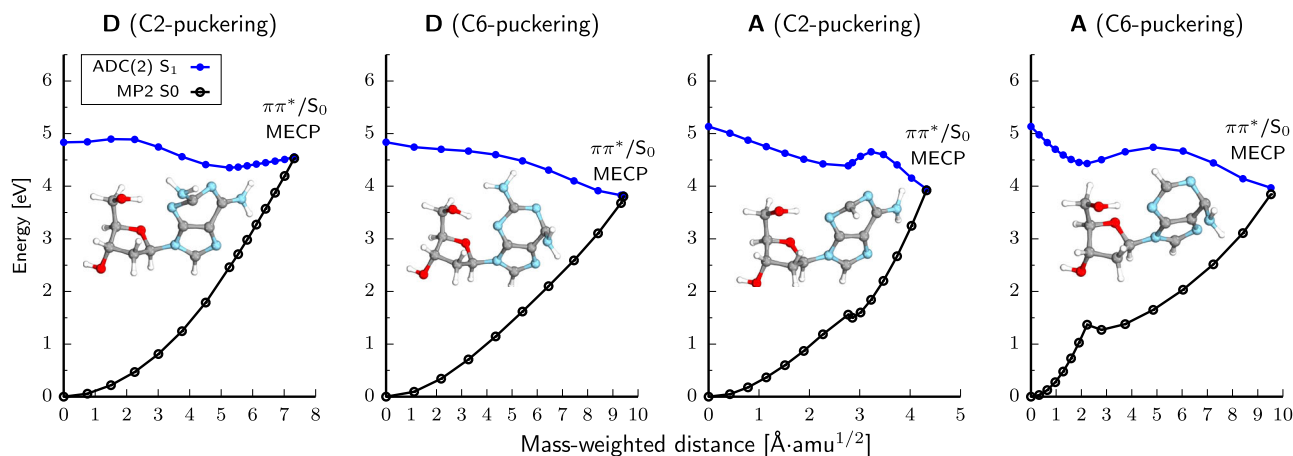
**Fig. 5 Transglycosylation reaction of 2'-deoxy-2-thiouridine (2tU) furnishing 2,6-diaminopurine 2'-deoxyriboside (D).** Transglycosylation of 2tU in the dry state with D gives a mixture of  $\alpha$  and  $\beta$  stereoisomers of the D deoxyriboside alongside loss of the 2tUra base.

cytidine, uridine, and thymidine as the most photostable canonical nucleosides<sup>47</sup>. Therefore, we performed comparative continuous irradiations of D and A within the UV-C spectral domain, to establish whether D could survive in the prebiotic environment as a potential building block. Our HPLC analysis (Supplementary Fig. 19 in the SI) performed for different irradiation times demonstrates that D undergoes photodegradation approximately two times more rapidly than A when exposed to UV-C radiation (254 nm). Approximately 11.6% and 5.9% of D and A were decomposed, respectively, after 120 min of continuous irradiation (Supplementary Fig. 19 in the SI). Even though D undergoes photodegradation more rapidly than A, the difference in photostability in the UV-C range is not overwhelming. Analogous irradiation experiments performed in the UV-B spectral range resulted in the photodegradation of D in merely 16.4% after 120 min, whereas A remained practically photostable under these conditions and its depletion was below the accuracy threshold of our analyses (Supplementary Fig. 20 in the SI). It worth noting that the measured photon flux in the UV-B region was an order of magnitude higher than in the UV-C range and several times higher than the overall UV flux estimated for the early Earth (see the following section for more details). Therefore, we conclude that D is sufficiently photostable to have been involved in the formation of the first informational polymer.

Based on our ADC(2) calculations, we ascribe the lower photostability of D to the fact that its key photorelaxation mechanisms are associated with out-of-plane rotation of amino groups that require longer timescales than the ultrafast C2-H bond rotation in A<sup>48</sup>. Thus, despite a modest energy barrier, rotation of the C2-H bond occurs more rapidly in aqueous environment than rotation of the C2-NH<sub>2</sub> group, which forms hydrogen bonds with nearby water molecules. Furthermore, the C2-NH<sub>2</sub> rotation in D leads to a sloped  $S_1/S_0$  state crossing that lies 0.3 eV above the C2-puckered  $S_1$  minimum (see Fig. 6). This additionally hinders damageless photodeactivation of D. Consequently, the anticipated longer lifetimes of the electronically excited bi-radical state in D are likely to entail destructive bimolecular photoreactions.

#### Comparison of irradiation conditions and prebiotic UV fluxes.

The integrated surface flux between 200 and 300 nm delivered to the surface of early Earth was estimated to be  $\sim 2700 \text{ erg s}^{-1} \text{ cm}^{-2}$ <sup>23</sup>. Therefore, we aimed to irradiate our samples with UV fluxes that would be consistent with this value within an order of magnitude in terms of intensity. Our tunable UV lamp used for the irradiation of damaged DNA trimers at 280 nm delivered flux of  $800 \text{ erg s}^{-1} \text{ cm}^{-2}$ , which is in good agreement with the UV environment expected for prebiotic Earth. In this experiment, we focused on irradiating our samples with the same wavelength applied previously by Pan et al.<sup>40</sup>



**Fig. 6** Potential energy profiles for the key photorelaxation mechanisms of 2,6-diaminopurine and adenine 2'-deoxyribosides (D and A). The electronic-structure calculations for ground and excited states were performed at the MP2 and ADC(2) levels, respectively, using the TZVP basis set. MECP minimum-energy crossing point.

in analogous irradiation experiments of canonical DNA trimers. Irradiation experiments aiming to determine the photostationary equilibria of undamaged DNA trimers and the photostability of D and A nucleosides delivered higher fluxes of  $17,000 \text{ erg s}^{-1} \text{ cm}^{-2}$  in the UV-B spectral range. It is important to emphasize that even though we did not perform broadband irradiation experiments that would fully cover the spectrum of the young Sun, our work provides a good comparative assessment of the self-repair properties and photostability of trinucleotides and nucleosides comprising Ade and Dap bases. Consequently, data presented in Fig. 3 and Supplementary Fig. 19 indicate that the D nucleosides and the studied trimers containing D could withstand days of continuous exposure to UV light on early Earth. Broadband irradiation experiments aiming to estimate the half-life of short D-containing DNA oligomers in UV-rich prebiotic environments are currently underway.

## Discussion

In summary, we have demonstrated that D might have served as an important component of the primordial genetic alphabet by significantly enhancing the photostability of DNA oligomers and increasing the self-repair rates of cyclobutane pyrimidine dimers (CPDs). These effects result from a decreased energy threshold for photoinduced charge separation and enhanced efficient intra-strand electron transfer. In particular, D readily inserts an electron into a CPD which is followed by self-repair of the damaged site. Thus, D can be considered as a prebiotic flavin mimic enabling DNA self-repair with  $\sim 90\%$  yield. To further reinforce its prebiotic relevance, we have shown that D could have been formed on early Earth alongside A through a transglycosylation reaction. Alternative syntheses of purine (deoxy)nucleosides raise the prospect that D might also be a product of these alternative chemistries, perhaps in more significant yields<sup>15,17,24,49,50</sup>. Given its only moderately lower photostability relative to A, D could survive long enough in a UV-rich environment to be incorporated into oligomers and play a further role in enzyme-free DNA self-replication. Ongoing questions remain as to the reason and mechanism by which D was eventually removed from the genetic alphabet. It is worth noting that the additional amino group in the 2 position of the purine ring could affect functional tertiary structures of RNA or hinder separation of duplexes. Consequently, if D had provided limited functionality compared to, for example, A, it might have not persisted after the emergence of the first CPD-repairing enzyme. The overarching and unintuitive result of this study is that the formation of highly photostable molecular assemblies could be achieved by joining moderately

photostable or even photounstable building blocks, here represented by D and pyrimidine-rich DNA fragments. We anticipate that these findings will shed further light on UV-induced prebiotic selection, as it has been widely postulated that Nature selected only the most photostable building blocks to protect their biomolecular polymers from the photodamage caused by continuous and intense UV irradiation<sup>47</sup>. The result of this work could be also used to support the design of self-healing macromolecules<sup>51</sup> or to enhance the electron transfer properties of DNA<sup>52</sup> through the design of photostable synthetic XNAs.

## Methods

### Irradiation experiments

**Samples.** All oligonucleotides (ATT, TTA, DTT, TTD) were purchased from Future Synthesis, Poland. The samples were HPLC-purified by the Future Synthesis. 2'-deoxyadenosine and 2,6-diaminopurine 2'-deoxyriboside were purchased from Merck and Alfa Aesar, respectively. All solutions were prepared using ultra-pure water produced by a Millipore purification system.

### Generation of AT=T, T=TA, DT=T, T=TD

**Illumination conditions.** The thymidine photodimerization was conducted in the pH 7.0 buffer solution of 10 mM phosphate containing 100 mM sodium chloride. All samples were illuminated with UV-B light in quartz tubes with the use of Irradiation Chamber BS-04 (Opsytec Dr. Gröbel, Ettlingen) at an average power of 1.95 mW measured by a Maestro11 power/energy meter, equipped with an 11PD photodetector (Standa, Lithuania). The accumulation of photoproduct was performed using HPLC (see Supplementary Fig. 4) followed by characterization with the use of mass spectrometry (LC-MS and LC-MS/MS, see Supplementary Figs. 5–16).

**HPLC analysis.** Damaged trinucleotides were isolated using reversed-phase HPLC method (Dionex Ultimate 3000 System with a Diode Array Detector, which was set at 260 nm). The mobile phases used in this study were 0.1% HCOOH and 80% ACN in water (gradient elution from 0 to 25% ACN in 15 min) with a flow rate of 1 mL/min. The separations were conducted on the Wakopak Handy ODS column ( $4.6 \times 150 \text{ mm}$ ,  $5 \mu\text{m}$  in particle size, and  $100 \text{ \AA}$  in pore size).

**Mass spectrometry.** Ultra-High Performance Liquid Chromatography system, Nexera X2 (Shimadzu) was coupled to ESI-TripleTOF 5600+ (AB SCIEX) mass spectrometer operated in negative ion mode. The separations were performed under the following conditions: Kinetex column (Phenomenex C18,  $100 \text{ \AA}$ ,  $2.1 \times 150 \text{ mm}$ ); flow rate  $0.3 \text{ mL/min}$ ; a gradient elution with 80% ACN and 0.1% HCOOH in water as mobile phases (from 0 to 25% ACN in 15 min). The MS experiments included the following parameters:  $-4.5 \text{ kV}$  ion spray voltage (ISV);  $-80 \text{ V}$  declustering potential (DP);  $-10 \text{ V}$  collision energy (CE);  $200 \text{ }^\circ\text{C}$  source temperature.

### Self-repair of T=T lesions

**Illumination conditions.** All samples were illuminated at 280 nm in quartz capillaries with the use of a xenon lamp. The 280-nm wavelength of incident light was selected using a monochromator (M250 Optel, Poland). The average incident light intensity for the above irradiation setup was  $80 \mu\text{W}$ . This value was determined with a Maestro11 laser power/energy meter, equipped with an 11PD photodetector



(Standa, Lithuania). To provide stable measuring conditions at pH 7, the oligonucleotides were dissolved in water and buffered by phosphate buffer at a concentration of 10 mM. For each illumination experiment, the concentration of trinucleotides was equal to 20  $\mu$ M. Illumination was performed in different time intervals up to 45 min. Between the illumination steps absorbance spectra from 230 to 340 nm were recorded. All experiments were performed at least in triplicate and followed by HPLC analysis and characterization of repair product by mass spectrometry (LC-MS and LC-MS/MS).

**HPLC analysis.** Irradiated and non-irradiated samples of trinucleotides were analyzed using the method described in the section “Generation of AT=T, T=TA, DT=T, T=TD”.

**Mass spectrometry.** All analyses were performed using the method described in the section “Generation of AT=T, T=TA, DT=T, T=TD”.

### Photodegradation of D and A

**Illumination conditions.** The water solution of A and D ( $10^{-4}$  M) containing phosphate buffer (10 mM, pH = 7.0) was illuminated at UV-C radiation in quartz capillaries for 30, 60, 90, and 120 min. All samples were deoxygenated before irradiation by purging with argon and illuminated in Irradiation Chamber BS-04 (Opsytec Dr. Gröbel, Ettlingen) used at an average power of 170  $\mu$ W (UV-C) or 1.7 mW (UV-B) measured by a Maestro11 power/energy meter, equipped with an 11PD photodetector (Standa, Lithuania). All experiments were performed at least in duplicate and followed by HPLC analysis. The percentage of photodegradation was calculated by integration of the respective peak areas.

**HPLC analysis.** Irradiated and non-irradiated samples of A and D were analyzed with the use of reversed-phase HPLC method on Dionex UltiMate 3000 System with Diode Array Detector. The detection of the effluents was carried out at 260 nm. Separations were performed using a Wakopak Handy ODS column (C18, 4.6  $\times$  150 mm, 5  $\mu$ m; 100 Å), 0.1% HCOOH and 80% ACN in water as mobile phases (gradient elution from 0 to 10% ACN in 10 min) and a flow rate of 1 mL/min.

### Synthetic procedures

**Solid-state glycosylation of 2,6-diaminopurine (Dap) with 2'-deoxy-2-thiouridine (2tU) to form  $\beta$ -2,6-diaminopurine 2'-deoxyriboside ( $\beta$ -D).** 0.20 mL of a standard solution (22 mM in MeOH) of 2'-deoxy-2-thiouridine 2tU (1.0 mg, 4.4  $\mu$ mol, 1.0 equiv.) was evaporated in a 10 mL round-bottom flask, leaving 2tU as a white solid distributed on the flask. 2,6-Diaminopurine Dap (9.3 mg, 66  $\mu$ mol, 15.0 equiv.), and milliQ-filtered water (pH 8 by addition of traces of NaOH) (0.9 mL) were added and the suspension was sonicated until all the 2'-deoxy-2-thiouridine had been dissolved (the diaminopurine remains insoluble). The pH at this stage was measured at 8. This mixture was then concentrated on a rotary evaporator at 45 °C and ~5 mbar until the solvent had been evaporated, leaving a white deposit on the inside of the flask. The flask was then immersed in an oil bath at 100 °C. After 40 h, the material was dissolved in D<sub>2</sub>O and <sup>1</sup>H NMR spectra recorded. The supernatant of this mixture was used for HPLC and LC-MS analysis, showing the presence of mainly diaminopurine Dap, 2-thiouracil 2tUra, and small quantities of  $\beta$ -2,6-diaminopurine 2'-deoxyriboside ( $\beta$ -D) and an isomer, presumably the  $\alpha$ -configured diastereomer (a standard of this material is not available, however, the reaction appears to have a similar outcome, and similar retention times, to the reported reaction between 2'-deoxy-2-thiouridine 2tU and adenine which forms  $\alpha$ - and  $\beta$ -configured diastereomers. The isomer has the same molecular mass and an almost identical UV-Vis spectrum to  $\beta$ -D). The identity of  $\beta$ -2,6-diaminopurine 2'-deoxyriboside  $\beta$ -D was confirmed by spiking experiments and measuring by both NMR and HPLC (Figs. S2, S3). The yield could not be determined accurately by NMR because of overlapping signals. Development of an analytical HPLC method (see below) gave conversions after 40 h of 2% and 4% for  $\beta$ -D and the presumed  $\alpha$ -diastereomer, respectively. Analytical HPLC was used to monitor the reaction, using a ThermoFisher Ultimate 3000 UPLC and Waters Atlantis T3 C18 column (5  $\mu$ m, 4.6 mm  $\times$  150 mm). UV absorption calibration curves for standards of 2'-deoxy-2-thiouridine 2tU, thiouracil 2tUra,  $\beta$ -2,6-diaminopurine 2'-deoxyriboside  $\beta$ -D were collected. UV absorption for known peaks (verified by spiking) was recorded and calibration curves used to determine relative amounts. Recovery was calibrated to aliquots from time zero at which the yield of starting material 2'-deoxy-2-thiouridine 2tU was set to 100%.

**Molecular dynamics simulations.** Force-field parameters for the T=T dimer were taken from our previous study<sup>28</sup>. Both the AT=T and T=TA trinucleotides were solvated in a truncated octahedral water box with the distance between solute and box border being at least 10 Å. For each trinucleotide we used two different water models, SPC/E<sup>53</sup> and OPC<sup>54</sup>. The box was first neutralized by 2 K<sup>+</sup> ions and then 0.15 M KCl was added, using Joung and Cheatham parameters for SPC/E and TIP4Pew, respectively<sup>55</sup>. The simulations were carried out with the parmOL15 force field<sup>56</sup>, which contains several dihedral reparametrizations<sup>57–59</sup> of the DNA force field developed by Cornell et al.<sup>60</sup>. Standard equilibration protocol was employed first. Five hundred steps of steepest descent minimization were followed by 500 steps of conjugate gradient minimization with 25 kcal mol<sup>-1</sup> Å<sup>-2</sup> position restraints on

DNA atoms. We then imposed position restraints of 25 kcal mol<sup>-1</sup> Å<sup>-2</sup> on the DNA fragment and heated the system from 0 to 300 K over 100 ps. Heating was performed at constant volume. The system was then minimized with 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> restraints on the DNA oligonucleotide using 500 steps of steepest descent method followed by 500 steps of the conjugate gradient optimization approach. The whole system was then equilibrated for 50 ps at a constant pressure of 1 bar and constant temperature of 300 K. We further performed an analogous series of consecutive minimizations and equilibrations applying decreasing restraints on the atomic position of the DNA fragment of 4, 3, 2, and 1 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The final equilibration was carried out using position restraints of 0.5 kcal mol<sup>-1</sup> Å<sup>-2</sup> and starting velocities from the previous equilibration step and was followed by free (unrestrained) molecular dynamics simulation for 50 ps. Temperature and pressure coupling constants were set to 0.2 ps during the equilibration, and to 5 ps during the final phase of molecular dynamics. Production phase was started afterward and each system was simulated for 10  $\mu$ s. The time step was set to 4 fs with application of the hydrogen mass repartitioning<sup>61</sup> and the SHAKE<sup>62</sup> and SETTLE<sup>63</sup> algorithms. The temperature and pressure were set at 300 K and 1 bar, respectively, using Berendsen weak-coupling thermostat and barostat<sup>64</sup>. Electrostatic interactions were treated using the Particle Mesh Ewald method (PME). The non-bonded cutoff was set to 9 Å. The calculations were performed with the CUDA accelerated pmemd module of AMBER 16<sup>65</sup>.

Each trajectory was clustered to find the important conformers. We applied a modified 15 algorithm of Rodriguez and Laio<sup>66</sup>, and used eRMSD as a metric<sup>67</sup>. It is a nucleic acid-specific measure that determines the similarity of two structures based on relative orientation of its nucleobases and is a suitable tool for distinguishing stacked and unstacked conformations or syn- and anti-oriented nucleotides. Only clusters with a population exceeding 1% were selected for further inspection.

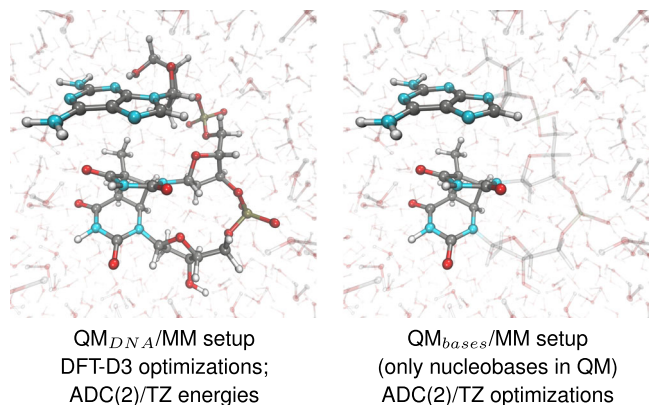
### QM/MM simulations

**Preparation of starting solvated oligonucleotide structures.** After analyzing and clustering the trajectories simulated for the AT=T and T=TA trimers we selected the most representative stacked conformers for the QM/MM simulations (see Supplementary Figs. 21 and 22 and the accompanying simulation). These conformers have the structural features of longer oligonucleotides and therefore our qualitative observations and results of the excited-state calculations are also relevant for longer DNA fragments. The stacked structural arrangement of AT=T is also the highest populated conformer in both MD simulations using the SPC/E and OPC solvent models. Even though the most populated conformer of T=TA is an unstacked structure, the stacked conformer is among the most populated structures of this trinucleotide (see Supplementary Figs. 21 and 22). In order to perform QM/MM calculations on equivalent structures of native trimers and those modified with 2,6-diaminopurine (Dap) we substituted the C2-H position of adenine with an amino group and generated the corresponding DT=T and T=TD stacked structures. Each trinucleotide was then solvated in a sphere of explicit water molecules with the radius of 24 Å. The solvent was then equilibrated for 10 ps with position constraints imposed on the solute molecule (trinucleotide). The resulting geometries were used in subsequent ground-state geometry optimizations.

**QM/MM setups.** The QM/MM calculations were performed in two different partitioning setups involving either the whole trinucleotide in the QM region (QM<sub>DNA</sub>/MM setup) or just the nucleobases treated at the QM level (QM<sub>bases</sub>/MM setup), as presented in Fig. 7. The remaining part of the system was described with the same force-field suite as in the MD simulations, including the SPC/E<sup>53</sup> force field for water. We applied the link hydrogen atom scheme in the QM<sub>bases</sub>/MM setup with the MM charge of the link H atom being set to 0. The boundary between the QM<sub>bases</sub> and MM regions bisected the N-glycosidic bonds. We employed the electrostatic embedding framework to all of the QM/MM calculations. We performed these simulations using our in-house modification of the QM/MM interface<sup>68</sup> of the AMBER suite of programs<sup>65</sup> to enable QM calculations employing the TURBOMOLE 7.3 program<sup>69</sup>.

**Geometry optimizations.** We optimized the minimum-energy geometries of the four selected structures of damaged DNA trimers using the cost-effective composite PBEh-3c<sup>70</sup> hybrid DFT scheme for the electronic-structure calculations and the QM<sub>DNA</sub>/MM setup. The PBEh-3c method was recently demonstrated to provide accurate geometries and energies of short nucleic acid fragments<sup>71</sup>. Initial optimizations of the ground-state geometries were performed for the whole solvated system using the AMBER limited-memory BFGS algorithm. We performed tighter optimization of the internal region of the sphere (within the radius of 12.5 Å from the most central atom of each conformer) using the open-source external geometry optimizer XOPT, which employs the rational function and approximate normal coordinates schemes<sup>72,73</sup>. The outer region of the sphere was kept frozen during this optimization procedure. This approach was used for all ground-state and excited-state geometry optimizations.

**Excited-state QM/MM calculations.** Calculations of excited-state energies and gradients within the QM/MM framework were performed using the algebraic diagrammatic construction to the second-order [ADC(2)]<sup>37,38,74</sup> method for electronic-structure calculations, whereas corresponding energies and gradients for



**Fig. 7 QM/MM setups used in the geometry optimizations and energy calculations.** Solid balls and sticks models show the two considered QM regions, while transparent residues (water molecules in both setups and the sugar-phosphate backbone on the right) correspond to the MM region described at the force-field level. The MM region interacts with the QM part via point charges (electrostatic embedding). ADC(2)/TZ denotes the ADC(2) method with the TZVP basis set.

the electronic ground-state were obtained at MP2 level. All of the ADC(2)/MP2 calculations were performed using the TZVP basis set. Vertical excitation energies contained in the Supplementary Table 4 were calculated using the QM<sub>DNA</sub>/MM setup on top of the geometries optimized at the PBEh-3c level. Optimizations of different minima on the  $S_1$  potential energy surface and  $S_1/S_0$  minimum-energy crossing points (MECPs) were performed using the QM<sub>bases</sub>/MM setup. The orbital characters of excited electronic states were determined based on analysis of molecular orbitals and key contributing configurations. In addition, we established the charge-transfer numbers and performed electron-hole population analysis using the TheoDOR 1.5.1 package<sup>75,76</sup>. MECP optimizations were performed using the algorithm of Levine, Coe, and Martínez<sup>77</sup>, which allows to locate highly representative geometries of  $S_1(\pi\pi^*)/S_0$  conical intersections. Even though the single-reference ADC(2)/MP2 approach does not provide a correct description of the entire topography of  $S_1/S_0$  conical intersections, it was demonstrated that it can accurately describe geometries of  $S_1(\pi\pi^*)/S_0$  MECPs in biomolecular systems with conjugated  $\pi$  bonds including qualitatively correct shapes of PE surfaces from the Frank-Condon region up to the state crossing<sup>78,79</sup>. While much less accurate results were obtained for  $S_1(n\pi^*)/S_0$  conical intersections<sup>79</sup>, in this work, we only considered  $S_1(\pi\pi^*)/S_0$  MECPs with only one dominant electronic configuration contributing to the excited-state character. Therefore, we expect that our approach provides at least qualitatively correct description of the mechanism. The energies of the key stationary points ( $S_0$  and  $S_1$  minima, MECPs, and vertical excitation energies) were used to construct illustrative PE surfaces in Fig. 3. The barrierless profiles between the Franck-Condon region of each trinucleotide and different  $S_1$  minima were confirmed by excited-state geometry optimizations. The barrier heights between the charge-transfer  $S_1$  minima and corresponding MECPs responsible for C5–C5 bond cleavage were estimated based on a relaxed potential energy scan along the C5–C5 distance performed for the whole trimer (see Supplementary Fig. 25 for more details). The barrier heights (presented in Fig. 3) between the  $S_1$  minima of the purine bases and the corresponding  $S_1/S_0$  MECPs were taken from the calculations performed for single nucleosides.

**Simulations of the ADT=TA pentamer.** Classical molecular dynamics simulations and subsequent ground-state geometry optimizations within the QM/MM framework for the ADT=TA pentamer were performed in an analogous way to the AT=T and T=TA trimers. The excited-state QM/MM calculations for this pentamer were also performed in an analogous way to the studied DNA trimers (using the ADC(2) method), except that we used the smaller def2-SVP basis set. See also the final section of the Supplementary Information and Supplementary Fig. 26, for more details.

**Excited-state QM simulations of A and D nucleosides.** The excited-state energies and geometries for adenosine and 2,6-diaminopurine nucleosides were calculated in the gas phase using the ADC(2) method for electronic excitations. The ground-state geometries and energies were obtained at the MP2 level. Similarly as in the case of the QM/MM calculations, for trinucleotides we used the TZVP basis set to keep the methodology consistent. MECPs for isolated nucleosides were optimized using the same algorithm as described above. The potential energy profiles presented in Fig. 5 in the main article were constructed by means of linear interpolation in internal coordinates between the respective  $S_0$  geometries (Frank-Condon region),  $S_1$  geometries, and  $S_1/S_0$  MECPs. All of the gas phase electronic-structure calculations were performed with the TURBOMOLE 7.3 program<sup>69</sup>.

## Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files.

## Code availability

The external optimizer XOPT is available under the link <https://github.com/hokru/xopt>. The modified interface between the TURBOMOLE 7.3 and AMBER 16 programs, which enabled performing QM/MM simulations with the ADC(2) electronic-structure method was deposited on the Figshare platform under the link <https://figshare.com/s/b8a5526e0c3335a270ab> with <https://doi.org/10.6084/m9.figshare.14268056>. Most recent revision of this code can be obtained upon reasonable request.

Received: 16 June 2020; Accepted: 20 April 2021;

Published online: 21 May 2021

## References

- Taylor, J. S. Unraveling the molecular pathway from sunlight to skin cancer. *Acc. Chem. Res.* **27**, 76–82 (1994).
- Pagès, V. & Fuchs, R. P. How DNA lesions are turned into mutations within cells? *Oncogene* **21**, 8957–8966 (2002).
- Kripke, M. L., Cox, P. A., Alas, L. G. & Yarosh, D. B. Pyrimidine dimers in DNA initiate systemic immunosuppression in UV-irradiated mice. *Proc. Natl Acad. Sci. USA* **89**, 7516–7520 (1992).
- Mees, W. J. et al. Thymine dimerization in DNA is an ultrafast photoreaction. *Science* **315**, 625–629 (2007).
- Mouret, S. et al. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proc. Natl Acad. Sci. USA* **103**, 13765–13770 (2006).
- Rauer, C., Nogueira, J. J., Marquetand, P. & González, L. Cyclobutane thymine photodimerization mechanism revealed by nonadiabatic molecular dynamics. *J. Am. Chem. Soc.* **138**, 15911–15916 (2016).
- Mees, A. et al. Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair. *Science* **306**, 1789–1793 (2004).
- Tan, C. et al. The molecular origin of high DNA-repair efficiency by photolyase. *Nat. Commun.* **6**, 7302 (2015).
- Gehring, W. & Rosbash, M. The coevolution of blue-light photoreception and circadian rhythms. *J. Mol. Evol.* **57**, S286–S289 (2003).
- Deck, C., Jauker, M. & Richert, C. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat. Chem.* **3**, 603–608 (2011).
- Mansy, S. S. et al. Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**, 122–125 (2008).
- Adamala, K. & Szostak, J. W. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **342**, 1098–1100 (2013).
- Kim, S. C., O’Flaherty, D. K., Zhou, L., Lelyveld, V. S. & Szostak, J. W. Inosine, but none of the 8-oxo-purines, is a plausible component of a primordial version of RNA. *Proc. Natl Acad. Sci. USA* **115**, 13318–13323 (2018).
- Saladino, R. et al. Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. *Proc. Natl Acad. Sci. USA* **112**, E2746–E2755 (2015).
- Becker, S. et al. A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).
- Saladino, R. et al. Proton irradiation: a key to the challenge of N-glycosidic bond formation in a prebiotic context. *Sci. Rep.* **7**, 14709 (2017).
- Becker, S. et al. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76–82 (2019).
- Teichert, J. S., Kruse, F. M. & Trapp, O. Direct prebiotic pathway to DNA nucleosides. *Angew. Chem. Int. Ed.* **58**, 9944–9947 (2019).
- Yadav, M., Kumar, R. & Krishnamurthy, R. Chemistry of abiotic nucleotide synthesis. *Chem. Rev.* **120**, 4766–4805 (2020).
- Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
- Xu, J. et al. A prebiotically plausible synthesis of pyrimidine  $\beta$ -ribonucleosides and their phosphate derivatives involving photoanomerization. *Nat. Chem.* **9**, 303–309 (2017).
- Xu, J., Green, N. J., Gibard, C., Krishnamurthy, R. & Sutherland, J. D. Prebiotic phosphorylation of 2-thiouridine provides either nucleotides or DNA building blocks via photoreduction. *Nat. Chem.* **11**, 457–462 (2019).
- Roberts, S. J. et al. Selective prebiotic conversion of pyrimidine and purine anhydronucleosides into Watson-Crick base-pairing arabino-furanosyl nucleosides in water. *Nat. Commun.* **9**, 4073 (2018).
- Xu, J. et al. Selective prebiotic formation of RNA pyrimidine and DNA purine nucleosides. *Nature* **582**, 60–66 (2020).

25. Ranjan, S. & Sasselov, D. D. Constraints on the early terrestrial surface UV environment relevant to prebiotic chemistry. *Astrobiology* **17**, 169–204 (2017).
26. Bucher, D. B., Kufner, C. L., Schlueter, A., Carell, T. & Zinth, W. UV-induced charge transfer states in DNA promote sequence selective self-repair. *J. Am. Chem. Soc.* **138**, 186–190 (2016).
27. Nguyen, K. V. & Burrows, C. J. A prebiotic role for 8-oxoguanosine as a flavin mimic in pyrimidine dimer photorepair. *J. Am. Chem. Soc.* **133**, 14586–14589 (2011).
28. Szabla, R., Kruse, H., Stadlbauer, P., Šponer, J. & Sobolewski, A. L. Sequential electron transfer governs the UV-induced self-repair of DNA photolesions. *Chem. Sci.* **9**, 3131–3140 (2018).
29. Piccinni, V., Reiter, S., Keefer, D. & de Vivie-Riedle, R. Multiscale conformational sampling reveals excited-state locality in DNA self-repair mechanism. *J. Phys. Chem. A* **124**, 9133–9140 (2020).
30. Khudyakov, I. Y. A., Kirnos, M. D., Alexandrushkina, N. I. & Vanyushin, B. F. Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine. *Virology* **88**, 8–18 (1978).
31. Kirnos, M. D., Khudyakov, I. Y., Alexandrushkina, N. I. & Vanyushin, B. F. 2-Amino adenine is an adenine substituting for a base in S-2L cyanophage DNA. *Nature* **270**, 369–370 (1977).
32. Krishnamurthy, R., Jiménez, E. I. & Gibard, C. Prebiotic phosphorylation and concomitant oligomerization of deoxynucleosides to form DNA. *Angew. Chem. Int. Ed.* **60**, 10775–10783 (2021).
33. Schrum, J. P., Ricardo, A., Krishnamurthy, M., Blain, J. C. & Szostak, J. W. Efficient and rapid template-directed nucleic acid copying using 2'-amino-2',3'-dideoxyribonucleoside–5'-phosphorimidazole monomers. *J. Am. Chem. Soc.* **131**, 14560–14570 (2009).
34. Hartel, C. & Göbel, M. W. Substitution of adenine by purine-2,6-diamine improves the nonenzymatic oligomerization of ribonucleotides on templates containing thymidine. *Helv. Chim. Acta* **83**, 2541–2549 (2000).
35. Kozlov, I. A. & Orgel, L. E. Nonenzymatic oligomerization reactions on templates containing inosinic acid or diaminopurine nucleotide residues. *Helv. Chim. Acta* **82**, 1799–1805 (1999).
36. Bucher, D. B., Pilles, B. M., Carell, T. & Zinth, W. Charge separation and charge delocalization identified in long-living states of photoexcited DNA. *PNAS* **111**, 4369–4374 (2014).
37. Hättig, C. Structure optimizations for excited states with correlated second-order methods: CC2 and ADC(2). *Adv. Quantum Chem.* **50**, 37–60 (2005).
38. Dreuw, A. & Wormit, M. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **5**, 82–95 (2015).
39. Spata, V. A., Lee, W. & Matsika, S. Excimers and exciplexes in photoinitiated processes of oligonucleotides. *J. Phys. Chem. Lett.* **7**, 976–984 (2016).
40. Pan, Z., Chen, J., Schreier, W. J., Kohler, B. & Lewis, F. D. Thymine dimer photoreversal in purine-containing trinucleotides. *J. Phys. Chem. B* **116**, 698–704 (2012).
41. Martinez, T. J. Seaming is believing. *Nature* **467**, 412–413 (2010).
42. Lee, W. & Matsika, S. Role of charge transfer states into the formation of cyclobutane pyrimidine dimers in DNA. *Faraday Discuss.* **216**, 507–519 (2019).
43. Subotnik, J. E., Yeganeh, S., Cave, R. J. & Ratner, M. A. Constructing diabatic states from adiabatic states: extending generalized Mulliken–Hush to multiple charge centers with Boys localization. *J. Chem. Phys.* **129**, 244101 (2008).
44. Pecourt, J.-M. L., Peon, J. & Kohler, B. DNA excited-state dynamics: ultrafast internal conversion and vibrational cooling in a series of nucleosides. *J. Am. Chem. Soc.* **123**, 10370–10378 (2001).
45. Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol.* **47**, 531–543 (1970).
46. Callahan, M. P. et al. Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl Acad. Sci. USA* **108**, 13995–13998 (2011).
47. Beckstead, A. A., Zhang, Y., Vries, M. S. & Kohler, B. Life in the light: nucleic acid photoproperties as a legacy of chemical evolution. *Phys. Chem. Chem. Phys.* **18**, 24228–24238 (2016).
48. Plasser, F. et al. Surface hopping dynamics with correlated single-reference methods: 9H-adenine as a case study. *J. Chem. Theory Comput.* **10**, 1395–1405 (2014).
49. Fialho, D. M., Roche, T. P. & Hud, N. V. Prebiotic syntheses of noncanonical nucleosides and nucleotides. *Chem. Rev.* **120**, 4806–4830 (2020).
50. Kim, H.-J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl Acad. Sci. USA* **114**, 11315–11320 (2017).
51. Urban, M. W. et al. Key-and-lock commodity self-healing copolymers. *Science* **362**, 220–225 (2018).
52. Crespo-Hernández, C. E., Cohen, B. & Kohler, B. Base stacking controls excited-state dynamics in A-T DNA. *Nature* **436**, 1141–1144 (2005).
53. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
54. Izadi, S., Anandkrishnan, R. & Onufriev, A. V. Building water models: a different approach. *J. Phys. Chem. Lett.* **5**, 3863–3871 (2014).
55. Joung, I. S. & Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
56. Zgarbová, M. et al. Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.* **11**, 5723–5736 (2015).
57. Pérez, A. et al. Refinement of the AMBER force field for nucleic acids: improving the description of  $\alpha/\gamma$  conformers. *Biophys. J.* **92**, 3817–3829 (2007).
58. Krepl, M. et al. Reference simulations of noncanonical nucleic acids with different  $\chi$  variants of the AMBER force field: quadruplex DNA, quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **8**, 2506–2520 (2012).
59. Zgarbová, M. et al. Toward improved description of DNA backbone: revisiting epsilon and zeta torsion force field parameters. *J. Chem. Theory Comput.* **9**, 2339–2354 (2013).
60. Cornell, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
61. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).
62. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
63. Miyamoto, S. & Kollman, P. A. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Computational Chem.* **13**, 952–962 (1992).
64. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
65. Case, D. A. et al. *AMBER 2016* (2016).
66. Rodriguez, A. & Laio, A. Clustering by fast search and find of density peaks. *Science* **344**, 1492–1496 (2014).
67. Bottaro, S., Di Palma, F. & Bussi, G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res.* **42**, 13306–13314 (2014).
68. Götz, A. W., Clark, M. A. & Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* **35**, 95–108 (2014).
69. Ahlrichs, R., Bär, M., Häser, M., Horn, H. & Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **162**, 165–169 (1989).
70. Grimme, S., Brandenburg, J. G., Bannwarth, C. & Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* **143**, 054107 (2015).
71. Szabla, R., Havrila, M., Kruse, H. & Šponer, J. Comparative assessment of different RNA tetranucleotides from the DFT-D3 and force field perspective. *J. Phys. Chem. B* **120**, 10635–10648 (2016).
72. Kruse, H. & Šponer, J. Towards biochemically relevant QM computations on nucleic acids: controlled electronic structure geometry optimization of nucleic acid structural motifs using penalty restraint functions. *Phys. Chem. Chem. Phys.* **17**, 1399–1410 (2014).
73. Eckert, F., Pulay, P. & Werner, H.-J. Ab initio geometry optimization for large molecules. *J. Comput. Chem.* **18**, 1473–1483 (1997).
74. Trofimov, A. B. & Schirmer, J. An efficient polarization propagator approach to valence electron excitation spectra. *J. Phys. B: Mol. Opt. Phys.* **28**, 2299–2324 (1995).
75. Plasser, F., Bäßler, S. A., Wormit, M. & Dreuw, A. New tools for the systematic analysis and visualization of electronic excitations. II. Applications. *J. Chem. Phys.* **141**, 024107 (2014).
76. Plasser, F., Wormit, M. & Dreuw, A. New tools for the systematic analysis and visualization of electronic excitations. I. Formalism. *J. Chem. Phys.* **141**, 024106 (2014).
77. Levine, B. G., Coe, J. D. & Martínez, T. J. Optimizing conical intersections without derivative coupling vectors: application to multistate multireference second-order perturbation theory (MS-CASPT2). *J. Phys. Chem. B* **112**, 405–413 (2008).
78. Tuna, D. et al. Assessment of approximate coupled-cluster and algebraic-diagrammatic-construction methods for ground- and excited-state reaction paths and the conical-intersection seam of a retinal-chromophore model. *J. Chem. Theory Comput.* **11**, 5758–5781 (2015).
79. Szabla, R., W. Góra, R. & Šponer, J. Ultrafast excited-state dynamics of isocytosine. *Phys. Chem. Chem. Phys.* **18**, 20208–20218 (2016).



## Acknowledgements

The authors thank Prof. John D. Sutherland, Prof. Dimitar D. Sasselov, and Dr. Corinna Kufner for helpful discussions. This research was supported by the Simons Foundation (494188 to R.S.), the Foundation for Polish Science (START Fellowship to R.S.), the Czech Science Foundation (project 21-23718S to J.S. and P.St.), the project SYMBIT (reg. number: CZ.02.1.01/0.0/0.0/15\_003/0000477 to J.S. and H.K.) financed by the ERDF and a grant from the Ministry of Science and Higher Education in Poland (DS/531-T080-D494-20 to J.R.).

## Author contributions

M.Z., P.S., and J.R. irradiation experiments; N.J.G. nucleoside synthesis; P.St. and J.S. molecular dynamics simulations; H.K. code development; R.S. idea, quantum chemical simulations, and coordination of the project. R.S. wrote the main article with input from all coauthors. M.Z., N.J.G., P.St., and R.S. assembled the SI and “Methods” section.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23300-y>.

**Correspondence** and requests for materials should be addressed to R.S. or M.Z.

**Peer review information** *Nature Communications* thanks Hannes Mutschler and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021