# Electronic Supplementary Information (ESI) for: Sequential Electron Transfer Governs the UV-Induced Self-Repair of DNA Photolesions†

Rafał Szabla,*[a,b] Holger Kruse,[b] Petr Stadlbauer,[b,c] Jiří Šponer,[b] and Andrzej L. Sobolewski[a]

*February 15, 2018*

## Computational methods.

### MD Simulations.

Point charges for the nucleobase of the T=T dimer were derived using the ESP calculation at the HF/6-31G* level using Gaussian09 followed by two-stage RESP approach. Point charges for the sugar-phosphate backbone were adapted from standard nucleotides. Other force field parameters (e.g. bonding, torsional) were adapted directly from the parmOL15 force field, as our test simulations of the dimer (bases only) with these parameters gave satisfatory results (puckering of the cyclobutane ring in the T=T dimer was similar to known values and distribution for cyclobutane). The geometry of the tetranucleotide was derived directly from a GATT sequence that was cut from the experimental structure of a double helical DNA (PDB ID: 5OMX). Conformation of the T=T dimer in the tetranucleotide was obtained by manual formation of the cyclobutane ring and its optimization.

Simulations of the tetranucleotide solvated in a truncated octahedral water box with ions was carried out using the parmOL15 force field. After standard equilibration of the system (described in detail elsewhere), the production phase was started. The time time was set to 4 fs with application of the hydrogen mass repartitioning and the SHAKE algorithm. The temperature and pressure were set at 300 K and 1 atm, respectively, using Berendsen weak-coupling thermostat and barostat. 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.[1]

### Trajectory clustering.

The eRMSD metric measures the similarity of two structures based on relative orientation of its nucleobases. Thus, it is a suitable tool for distinguishing stacked and unstacked conformations or syn and anti oriented nucleotides. Our two trajectories contained 10 and 12 clusters, respectively, with the population of at least 1%. The cluster with guanosine in the syn conformation stacked on adenosine and T=T dimer was dominant in both simulations - 31.22% with OPC water model (Figure S1) and 47.02% with SPC/E water model (Figure S2). The stacked conformation with guanosine in the anti orientation was less populated because of the formation of intramolecular 5'OH...N3 hydrogen bond. This bond can be formed only at the 5'terminus, so we expect that the stacked conformation with the anti orientation would dominate in the context of longer nucleotide chains and in double stranded helices. The clustering results are shown pictorially in Fig. S1.

## Pre-optimizations of the local-minima.

Location of specific diabatic states and local minima on the $S_1$ hypersurface is a considerable difficulty in determining the photoreactivity of oligonucleotides. This is usually caused by high complexity of the $S_1$ surface and multiple degrees of freedom involved in each stage of the process. The most effective strategy for finding the crucial intermediate $S_1$ minima in the GAT=T tetramer was to perform pre-optimizations with significantly smaller QM region and the rest of the strand and environment frozen. For example, to locate the minimum of the $\pi_G\pi_G^*$ LE state, we limited the QM region only to the G base, imposed constraints on the rest of the system and performed a rough optimization (∼50 optimization steps) of the $S_1$ geometry at the ADC(2)/SV(P) level. The final optimization of the $\pi_G\pi_G^*$ LE geometry was further performed using the $QM_{bases}$/MM setup, the ADC(2)/def2-SVP method and the XOPT external optimizer.[2] In the case of the charge-transfer and exciplex states, we adapted a similar approach, but the pre-optimizations often consisted of two steps. For instance, to locate the $G^{+\cdot}A^{-\cdot}$ minimum, we first limited the QM region to the G base and optimized a radical cation on G at the MP2/SV(P) level. The second step involved the QM region limited to the A base and a optimization of a radical anion on A at the MP2/SV(P), keeping the whole surroundings frozen (including the G base). The final optimization of the $G^{+\cdot}A^{-\cdot}$ minimum was performed using the $QM_{bases}$/MM setup and the ADC(2)/def2-SVP method. While this process sometimes proved problematic in selected cases, we expect that the application of such procedures to other oligonucleotides could significantly simplify further computational studies of DNA and RNA photochemistry.

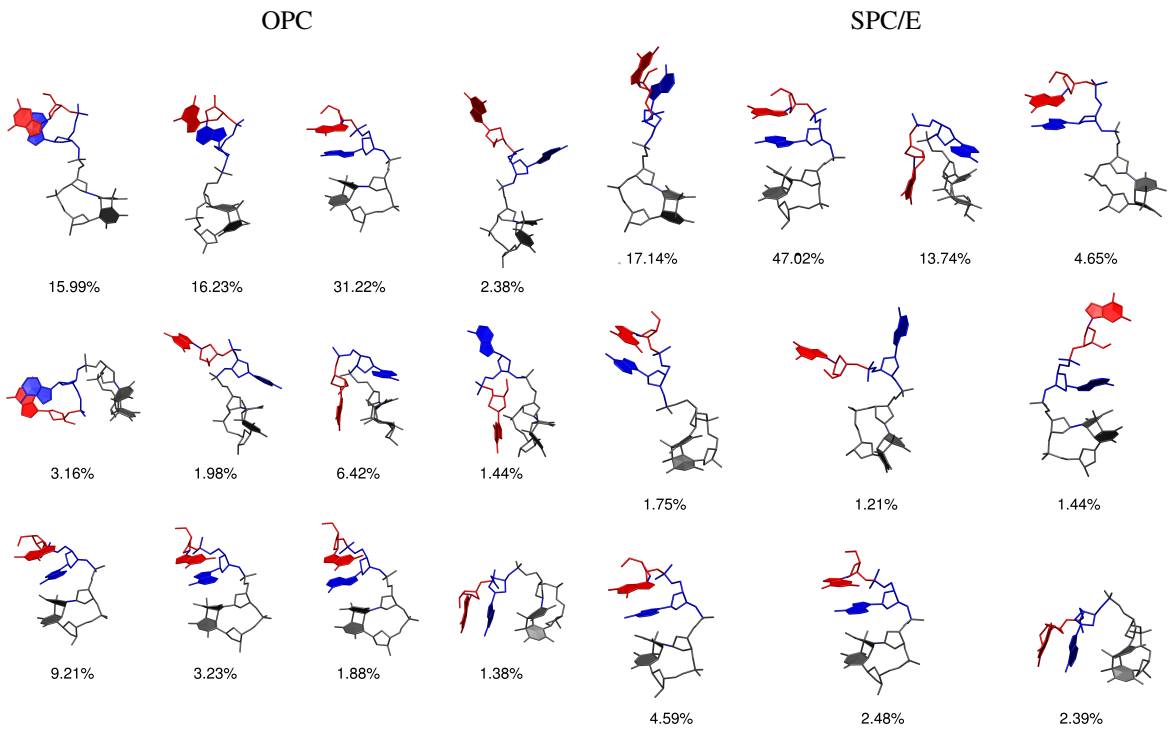


**Fig. S1** Clusters and their populations in the OPC and SPC/E water models. Guanosine is red, adenosine is blue and the T=T dimer is grey.

It is important to emphasize that, in the current work we were not able to clearly describe many reaction coordinates connecting the consecutive stationary points on the $S_1$ hypersurface. Even though, the local minima $S_1$ are very often characterized by different interbase distance, simple relaxed scans along one or more specific interatomic distances between neighbouring bases were insufficient to find the correct transition paths. This additionally emphasizes the complexity of the $S_1$ surface and the fact that these transition paths constitute multiple interbase and intrabase nuclear degrees of freedom. We are currently working on an efficient approach to tackle this problem.

## Vertical excitation energies of the GA-*syn* conformer.

The vertical excitation energies of the GA-*syn* conformer are presented in Table 1. Similarly as in the case of the GA-*anti* conformer the $\pi_G\pi^*_G$ is among the lowest-energy UV-absorbing states. It is worth noting though, that in this particular conformer it is nearly isoenergetic with the $\pi_A\pi^*_A$ state. While, the exact determination of the position of UV absorption bands would require better sampling of the configurational space and more accurate description of the solvation, the results presented in Table 1 suggest that the $\pi\pi^*$ excitations associated with the adenine base might be somewhat red-shifted in the GA-*syn* conformer. Similarly the lowest-energy CT state identified as the $\pi_A\pi^*_{TT}$ excitation has a $\sim$0.5 eV lower excitation energy when compared to the GA-*anti* conformer.

**Table 1** Vertical excitation energies (in eV) of the GA(-*syn*) conformer, computed assuming the QM$_{bases}$/MM setup at the ADC(2)/TZVP level based on the PBEh-3c/MM ground-state geometry.

| state / transition | | $E_{exc}$/[eV] | $f_{osc}$ | $\lambda$/[nm] |
|---|---|---|---|---|
| GA(-*anti*)T=T conformer | | | | |
| $S_1$(LE) | $n_{TT}\pi^*_{TT}$ | 4.69 | $7.33\cdot10^{-4}$ | 264.4 |
| $S_2$(LE) | $n_{TT}\pi^*_{TT}$ | 4.84 | $1.12\cdot10^{-3}$ | 256.2 |
| $S_3$(LE) | $\pi_A\pi^*_A$ | 4.96 | $13.4\cdot10^{-2}$ | 250.0 |
| $S_4$(LE) | $\pi_G\pi^*_G$ | 5.00 | $5.61\cdot10^{-2}$ | 248.0 |
| $S_5$(LE) | $\pi_A\pi^*_A$ | 5.06 | $6.75\cdot10^{-2}$ | 245.0 |
| $S_7$(CT) | $\pi_A\pi^*_{TT}$ | 5.33 | $4.26\cdot10^{-4}$ | 232.6 |

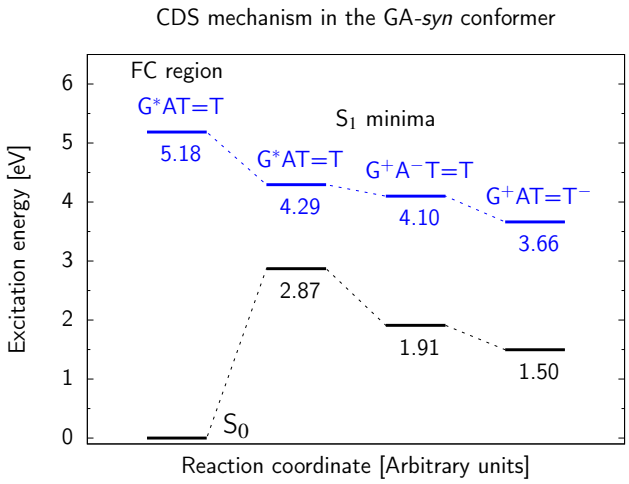


**Fig. S2** The sequential electron transfer (SET) mechanism initiated in the $\pi_G\pi^*_G$ LE state of the GA-*syn* conformer of the GAT=T tetramer. The geometry optimizations and energy calculations were performed using QM$_{bases}$/MM setup and the ADC(2)/def2-SVP method.

## Sequential electron transfer in the GA-*syn* conformer.

Fig. S2 presents the SET mechanism found in the GA-*syn* conformer. To show the similarities to the GA-*anti* conformer discussed in the main article we focused on locating the key intermediate stationary points, namely the ground-state, $\pi_G\pi^*_G$ LE, $G^{+\cdot}A^{-\cdot}$ and $G^{+\cdot}ATT^{-\cdot}$ minima. We anticipate that similarly as in the GA-*anti* conformer, the availability of the $G^{+\cdot}ATT^{-\cdot}$ minimum is sufficient to initiate the self-repair activity of the GAT=T tetranucleotide.

## Calculation of diabatic couplings in the GA-*anti* conformer.

The probability of electron transfer can be evaluated based on the computation of electronic coupling matrix elements between the different states considered above. For this purpose we employed the Generalized Mulliken Hush approach (GMH),[3,4] which provides reliable electronic coupling values for geometries which lie far from the degeneracy point of the two considered diabatic states. The GMH diabatic coupling matrix elements were computed in each of the $S_1$ minima along the SET pathway for the GA-*anti* conformer (see Fig. 3 in the main article). The values presentend in Table 2 confirm our initial assumption that the $\pi_G\pi^*_A$ CT state can be readily populated from the minimum of the $\pi_G\pi^*_G$ LE state, owing to the highest diabatic coupling value of 0.76 eV. In contrast, the coupling between the $\pi_G\pi^*_G$ LE excitation and the electronic state engaged in the opposite electron transfer process ($\pi_A\pi^*_G$) is ∼3 times lower. The subsequent electron transfer from the negatively charged adenine in the $\pi_G\pi^*_A$ CT state to the T=T dimer in the $\pi_G\pi^*_{TT}$ CT state is associated with somewhat weaker but still appreciable diabatic coupling of 0.182 eV. Our calculations suggest that two GA-exciplex states exhibit stronger coupling with the $\pi_G\pi^*_A$ CT state, albeit their respective minima lie higher in energy than the $G^{+\cdot}A^{-\cdot}$ minimum. This indicates that the GA-exciplex states are inaccessible in the SET pathway. Lastly, sizeable diabatic coupling between the $\pi_G\pi^*_{TT}$ and $\pi_A\pi^*_{TT}$ CT states indicate that the latter electronic state may be involved in the T=T repair mechanism. In summary, the GMH diabatic coupling matrix elements support the validity of the deduced SET mechanism and the proposed ordering of the local minima.

**Table 2** Diabatic coupling matrix elements calculated between the considered electronic states using the Generalized Mulliken Hush method, assuming QM$_{bases}$/MM setup and the ADC(2)/TZVP method for electronic structure calculations. Semi-quantitatively consistent values were obtained using the QM$_{DNA}$/MM setup and the ADC(2)/def2-SVP method

| Initial state | Final state | GMH [eV] | Boys localization [eV] |
|---|---|---|---|
| $\pi_G\pi^*_G$ LE minimum | | | |
| $\pi_G\pi^*_G$ LE | $\pi_G\pi^*_A$ CT | 0.760 | 0.803 |
| $\pi_G\pi^*_G$ LE | $\pi_A\pi^*_G$ CT | 0.271 | 0.268 |
| $G^{+\cdot}A^{-\cdot}$ minimum | | | |
| $\pi_G\pi^*_A$ CT | $\pi_G\pi^*_{TT}$ CT | 0.182 | 0.180 |
| $\pi_G\pi^*_A$ CT | $\pi\pi^*_{GA}$ exciplex | 0.302 | 0.272 |
| $G^{+\cdot}AT=T^{-\cdot}$ minimum | | | |
| $\pi_G\pi^*_{TT}$ CT | $\pi_A\pi^*_{TT}$ CT | 0.476 | 0.477 |

## References


1 D. Case, J. Berryman, R. Betz, D. Cerutti, T. Cheatham III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, S. Legrand, P. Li, T. Luchko, R. Luo, B. Madej, K. Merz, G. Monard, P. Needham, H. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, D. Roe, A. Roitberg, R. Salomon-Ferrer, C. Simmerling, W. Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, D. York and P. Kollman, *AMBER 14*, 2015.

2 H. Kruse, *https://github.com/hokru/xopt; local development version.*, 2016, Institute of Biophysics, Brno.

3 R. J. Cave and M. D. Newton, *J. Chem. Phys.*, 1997, **106**,

9213–9226.
4 J. E. Subotnik, S. Yeganeh, R. J. Cave and M. A. Ratner, *J. Chem. Phys.*, 2008, **129**, 244101.