Supporting Information: Excimers Intermediates en route to Long-Lived Charge Transfer States in Single-Stranded Adenine DNA Revealed by Nonadiabatic Dynamics

Lea M. Ibele,[†] Pedro A. Sánchez-Murcia,[‡] Sebastian Mai,[*,‡,⊥] Juan J. Nogueira,[*,¶,§] and Leticia González[*,‡,∥]

†*Department of Chemistry, Durham University, Durham, DH1 3LE, UK*

‡*Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Str. 17, 1090 Vienna, Austria*

¶*Chemistry Department, Universidad Autónoma de Madrid, Calle Francisco Tomás y Valiente, 7, 28049 Madrid, Spain*

§*IADCHEM, Institute for Advanced Research in Chemistry, Universidad Autónoma de Madrid, Calle Francisco Tomás y Valiente, 7, 28049 Madrid, Spain*

∥*Vienna Research Platform on Accelerating Photoreaction Discovery, University of Vienna, Währinger Str. 17, 1090 Vienna, Austria*

⊥*Present Address: Photonics Institute, Vienna University of Technology, Gußhausstraße 27-29, 1040 Vienna*

E-mail: sebastian.mai@tuwien.ac.at; juan.nogueira@uam.es; leticia.gonzalez@univie.ac.at

S1. Further computational details

S1.1. Classical molecular dynamics sampling

In order to sample the Franck-Condon region in the electronic ground state classical molecular dynamics simulations of solvated $(dA)_{20}$ in a water box (truncated octahedral extended to 10 Å from the solute atoms) were performed. The single strand was neutralized with Na^+ ions, yielding a total number of 46924 atoms, 639 atoms composing the DNA strand, 19 Na^+ atoms and 15422 water molecules. The Coulomb interactions were computed with the particle mesh Ewald (PME) method and the direct space cutoff in the PME method was set to 10 Å. All classical molecular dynamics were performed with Amber14,[S1] using the ff14SB set of parameters[S2] to describe the DNA and the TIP3P model[S3] for the water molecules and the Amber ff99 parameters for the sodium ions. The system was first minimized using the steepest descent method for 10000 steps, followed by 10000 steps using the conjugate gradient method. Subsequently, it was heated to 300 K at constant volume for 200 ps, employing a Langevin thermostat with a collision frequency of 1 ps^{-1}, with a time step of 2 fs. For the first 100 ps the atoms of the DNA strand were constrained with a force constant of 40 kcal/(mol$Å^2$), which was removed for the last 100 ps. Afterwards, it was equilibrated for 600 ps where the restraints applied during the heating were gradually removed and for the last 200 ps it was changed to the isothermal-isobaric ensemble. The time step was reduced from 2 fs to 0.5 fs to have a better resolution of the vibrational motion of the strand and to use the same time step as in the subsequent surface hopping simulations, which require the use of small time steps due to the excess of energy in the vibrational degrees of freedom. The Berendsen barostat with a pressure relaxtion time of 5 ps and a reference pressure of 1 bar was used. Then, a production run of 60 ns was evolved in the isothermal-isobaric ensemble, using identical settings as described above. From the last 35 ns of the dynamics an ensemble of 100 equidistant snapshots was selected for the static calculations, and an ensemble of 300 snapshots to sample initial conditions for the dynamics.

S1.2. Absorption spectra

For each snapshot from the molecular dynamics simulations, electronic excitations were computed employing an electrostatic embedding quantum mechanics/molecular mechanics (QM/MM) scheme. Two different sized QM regions were used: six or four adenines from the middle of the $(dA)_{20}$ strand. Depending on the size of the QM region, the lowest 30 (six QM adenines) or 20 (four QM adenines) singlet states were computed. The atoms in the QM region are described by time-dependent density functional theory (TDDFT) employing the CAM-B3LYP functional[S4] and the Ahlrichs def2-SV(P) basis set.[S5] The quantum mechanical computations were carried out with the ORCA program package.[S6] Due to the large system size, the resolution of the identity approximation, implemented as RIJ-COSX[S7,S8] in ORCA, was used in combination with the def2/j auxiliary basis set. The influence of the MM region was simulated by inclusion of point charges obtained using the same force fields as in the classical MD sampling with the molecular mechanics program Tinker[S9] and bonds between the QM and MM region were treated according to the hydrogen link atom approach.[S10] The point charge of the MM atom involved in the bond is redistributed to all neighboring MM atoms and set to zero on the bonded MM atom. In the analysis, all excited states that were mainly localized on either of the edge nucleobases was excluded as the proximity of the MM region is assumed to hinder an accurate description of said states. As observed in previous work,[S11] taking into account the edge nucleobases in the analysis, increases the contribution of monomer-like states while decreases the contribution of delocalized states. This is an artifact of the fact that nucleobases at the end of the QM region cannot be delocalized towards the MM nucleobases and thus, their delocalization length is underestimated. Indeed, this artificial increase of monomer-like states, is also observed in our calculations for QM sizes four and six, if the contributions of edge nucleobases are included. The comparison of the spectra shown in Figure S1 with the spectra without edge nucleobases in the main text (Figure 2) confirms thta monomer-like contributions are increased, so they should be considered an upper value. The contributions of other states, exciton, excimer and charge

transfer states, are unaffected.

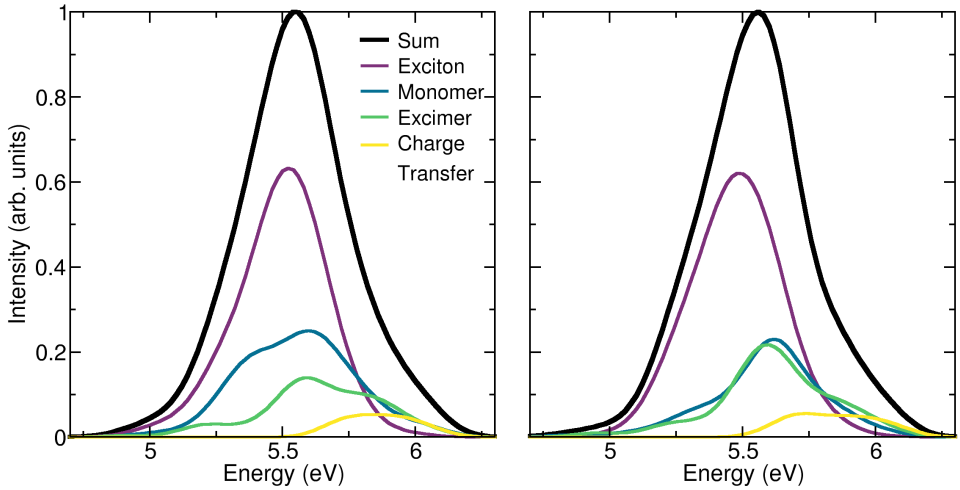

Figure S1: Absorption spectra of $(dA)_{20}$ with 4 (left plot) and 6 (right plot) adenines included in the QM region, including all contributions also those mainly localized on the nucleobases on the edge of the QM region.

The computed absorption spectrum is blue-shifted by 0.7 eV with respect to the experimental spectrum. As already shown,[S11] increasing the size of the basis set improves the agreement with the experimental spectrum. Figure S2a shows the absorption spectrum using a QM region including four adenines, obtained employing the def2-TZVP basis set. The spectrum is shifted to lower energies by about 0.25 eV, however the composition of the spectrum with respect to the different types of excited states does not change significantly. As this investigation focuses on the characterization of the excited states, we consider that the smaller basis set provides a sufficiently accurate description to be used throughout the dynamics and avoids a substantial increase in computational cost.

In TDDFT, a very commonly applied approximation is the so-called Tamm-Dancoff Approximation,[S12] that reduces the perturbation density matrix by neglecting de-excitations in the Casida formulation.[S13] While this approximation usually does not significantly alter TDDFT results, for our case the composition of the absorption spectrum changes significantly, as shown in Figure S2b. The contribution of excimer states increases, contributing almost 50 % of the absorption spectrum, while the amount of excitonic states decreases.

Hence, it can be assumed, that the charge transfer character of the excitations is largely overestimated when applying the Tamm-Dancoff approximation and thus, it was not used for the excited-state dynamics.

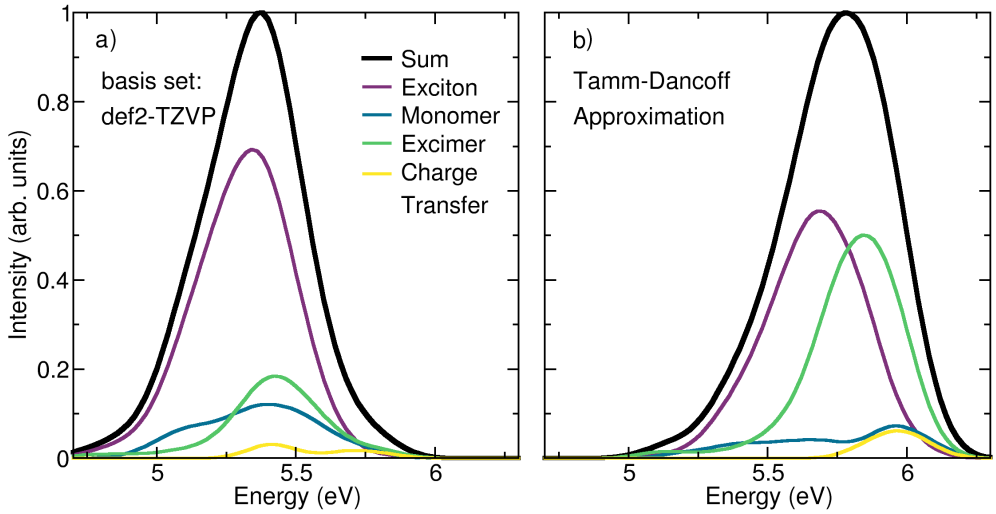

Figure S2: Absorption spectra of $(dA)_{20}$ with 4 adenines included in the QM region, employing the larger def2-TZVP basis set (a) and employing the Tamm-Dancoff Approximation (b).

S1.3. Nonadiabatic Excited state dynamics

For the nonadiabatic excited state dynamics, a QM region including four nucleobases was employed. 100 initial conditions were sampled from an absorption spectrum, computed from 300 initial conditions (equidistant snapshots from the last 35 ns of the classical molecular dynamics production run) within an excitation window between 5.29 and 5.41 eV. Since the computed absorption spectrum is blue shifted by 0.7 eV with respect to the experimental spectrum, this excitation window corresponds to the photoexcitation at 4.65 eV in the different time-resolved experiments.[S14–S16] The system is schematically shown in Fig. S3, the QM region including four adenines is located in the center of the single strand and highlighted in palatinate.

The electronic structure quantities were computed at the CAM-B3LYP/def2-SV(P) level of theory, as specified above for the static calculations. For the dynamics, only the lowest ten

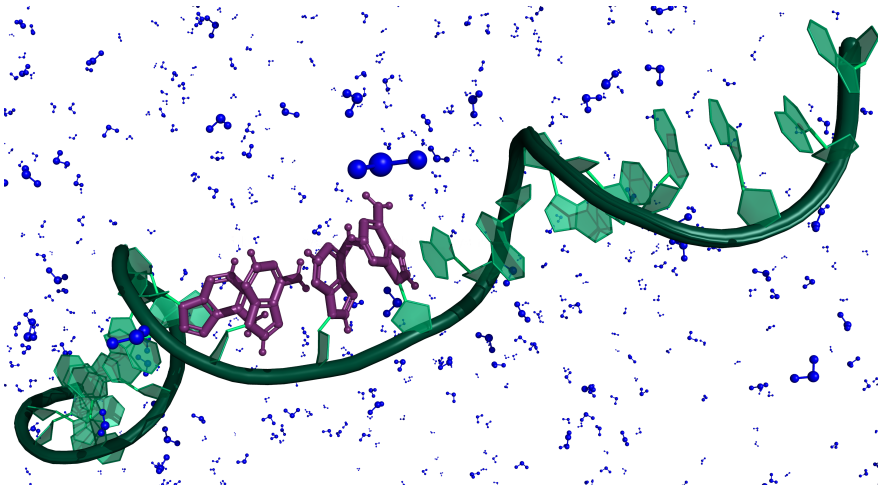

Figure S3: Illustrative representation of solvated $(dA)_{20}$, the four adenines forming the QM region are highlighted in palatinate, the MM region includes the water molecules (blue), the backbone and eight adenines on each side of the QM region (green).

excited states were computed because higher states do not significantly impact the dynamical behavior but have a dramatic impact in the computational cost. The trajectories are initialized in the lowest 8 singlet excited states according to their corresponding oscillator strength (10 in S_1, 21 in S_2, 22 in S_3, 13 in S_4, 20 in S_5, 9 in S_6, 4 in S_7, 1 in S_8) and are propagated according to the surface hopping approach according to the molecular dynamics suite SHARC 2.1.[S17,S18] The dynamics are simulated for 400 fs with a time step of 0.5 fs for the nuclear dynamics and 0.02 fs for the electronic wavefunction. The electronic wavefunction is propagated in the local diabatization approach[S19] using wavefunction overlaps to estimate nonadiabatic couplings; the overlaps are computed from truncated wavefunction that use 99 % of the wavefunction norm.[S20] During the hops the full velocity vector of the QM region without the hydrogen link atoms is rescaled. The decoherence is corrected according to a energy based scheme,[S21] where instead of the full kinetic energy of the system, only the kinetic energy of the QM region is taken into account. The decoherence parameter was set to 0.1 Hartree. We note, that the original implementation in SHARC was used where the amplitudes of the non-running states are damped,[S21] instead of the surface hopping populations (modulus squared of the amplitudes).[S22] Hence, in the employed algorithm, the decoherence time is effectively half with respect to the model.

S1.4. Characterization of the electronic excited states

All excited states were characterized with respect to the charge transfer character and average delocalization length of their excitations, classified according to the rules explained below. The values for these properties were obtained using the TheoDORE wavefunction analysis package.[S23,S24] To this end, the charge transfer numbers are computed from a two-dimensional population analysis of the transition density matrix, that divide each excitation into the contributions from excitations from orbitals on fragment A to orbitals on fragment B:

$$\Omega_{AB} = \frac{1}{2} \sum_{\mu \in A} \sum_{\nu \in B} ((\mathbf{S}^{1/2}\mathbf{D}\mathbf{S}^{1/2})_{\mu\nu})^2 \tag{1}$$

where $\mathbf{D}$ and $\mathbf{S}$ are the transition density and overlap matrices, respectively. From these charge transfer numbers, a descriptor for the charge transfer character (CT) can be obtained by normalized summation over the off-diagonal elements:

$$\mathrm{CT} = \Omega^{-1} \sum_{A,B \neq A} \Omega_{AB} \tag{2}$$

Additionally, the averaged delocalization length ($\mathrm{DL_{av}}$) is computed as the arithmetic average between the hole and electron participation ratios ($\mathrm{PR_h}$, $\mathrm{PR_e}$), defined as:

$$\mathrm{PR_h} = \frac{\Omega^2}{\sum_B (\sum_A \Omega_{AB})^2} \text{ and } \mathrm{PR_e} = \frac{\Omega^2}{\sum_A (\sum_B \Omega_{AB})^2} \tag{3}$$

The adenines included in the QM region are each defined as a separate fragment in the wavefunction analysis. The two descriptors above, CT and $\mathrm{DL_{av}}$, are sufficient to distinguish four types of excited states formed in stacked DNA bases: exciton, monomer-like excitation, excimer and charge transfer states. These states are schematically shown in Figure 2 of the main text. Excitonic and monomer-like states are both characterized by very low CT, hence a threshold of 0.2 was used to define them. Excitons are delocalized over more than one nucleobase, hence every excitation with low CT where more than a quarter of the total

density is delocalized ($DL_{av} > 1.25$) are attributed to excitons. Charge transfer states are dominated by excitations between different fragments and only show minor contributions of local excitations. Hence, all excitations with high CT (>0.8) are classified as charge transfer states. Although in the strict sense, they would have DL_{av} of close to 1.0, there was no restriction formally applied, as only very few states with significantly higher DL_{av} were observed. The fourth excited state considered in the analysis are so-called excimer states. They show a combination of both, local as well as charge transfer excitations. While they are often observed as an intermediate between monomer or exciton states to charge transfer states, the situation when both hole and electron are delocalized over the same two fragments and both charge transfer and local excitations are similarly contributing can also be a more stabilized state. All states with a CT between 0.2 and 0.8 were classified as excimers, if their DL_{av} was larger than 1.25 (which has been fulfilled in all occurring cases).

The characterization of these different states was used in the static calculations as well as during the dynamics. The computed absorption spectra were decomposed into the contributions of the respecting states to allow comparison of the respective initialization conditions. During the surface hopping dynamics, the running states of the trajectories are classified according to this scheme and the populations were analyzed with respect to those diabatic-like states instead of the widely used adiabatic picture. This allowed for an estimation of the nature of the long lived states formed.

S2. Decay to the electronic ground state

In total five trajectories were observed where the energy gap between the S_0 and S_1 state falls below 0.1 eV in the course of the dynamics. LR-TDDFT within the adiabatic approximation of the exchange-correlation kernel fails to describe the correct topology of the S_0/S_1 conical intersection between the ground and first excited state. Hence, a decay to the ground state cannot be simulated directly, and the trajectories are enforced to deactivate to the ground

state. Due to the small number of trajectories showing this behavior, a time constant in the kinetic model can only obtained with a large error and thus, it was not deemed useful, to assign a definite value to it. However, with certainty, the time constant of this decay can be assumed to be larger than 2 ps, as given in the schematic model in Fig. 4 of the main text. The exact value of the experimental findings[S16] (300 fs) thus cannot be reproduced, but given that it is on the same order of magnitude, qualitatively the model agrees with the experiment.

Analysing the structure of the five trajectories around the time of the decay to the ground state showed for all of them that one of the nucleobases undergoes an out-of-plane distortion of the C_2 atom on the six-membered ring. An exemplary structure is shown in Fig. S4. This structure is very similar to the S_1/S_0 conical intersection calculated for isolated adenine by Barbatti et al.[S25] Looking at the localization of the excitation in these trajectories prior to forming this structure, we observe it to be mainly localized in a monomer-like manner on the respective nucleobase undergoing the C_2 puckering. Thus, the deactivation to the ground state occurs in the comparable manner as in the isolated nucleobases.

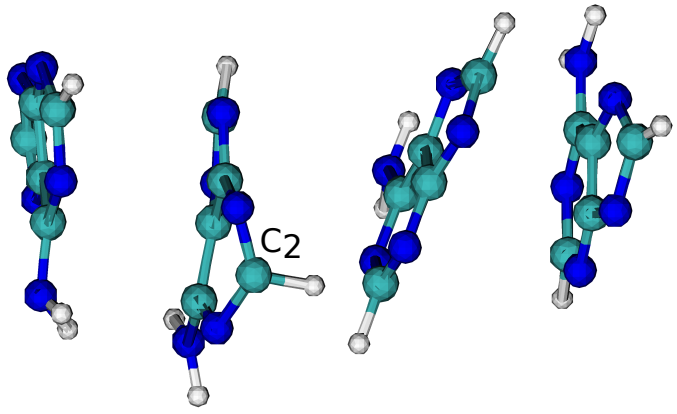

Figure S4: Exemplary geometry of the 4 adenines of the QM region from one trajectory of the dynamics, with an the S_0/S_1 energy gap below 0.1 eV, indicating a C_2 puckered conical intersection.

S3. Nucleobase separation in the ground and the excited state

To investigate the impact of the change of level of theory after the classical molecular dynamics sampling and the effect of the excitation, we propagated 50 initial conditions in the ground state QM/MM, describing the QM region using the Density-Functional Tight-Binding method (DFTB3).[?] The average distances between all nucleobases during the ground state dynamics is shown in the right plot of figure S5 and compared to those during the excited state dynamics (left plot). Of course, this cannot be seen as a one-to-one comparison, as fewer trajectories have been propagated on the ground state and as the QM region is described with the semi-empirical DFTB3 method for computational efficiency. However, it can be seen, that while both dynamics show changes of the inter base separation of the same order of magnitude, in the ground state dynamics less changes occur within the same time frame.

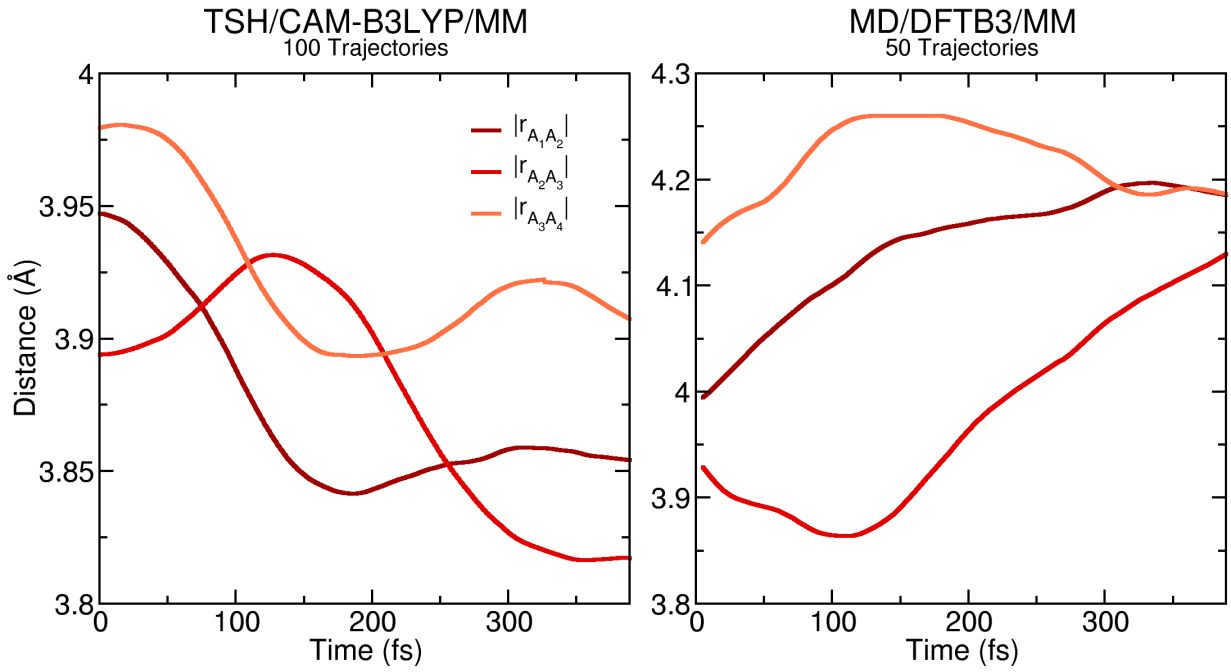

Figure S5: Average spatial separation between the four nucleobases in the QM region during the excited state trajectory surface hopping dynamics, CAM-B3LYP/MM, (left) and during ground state dynamics, DFTB3/MM, (right).

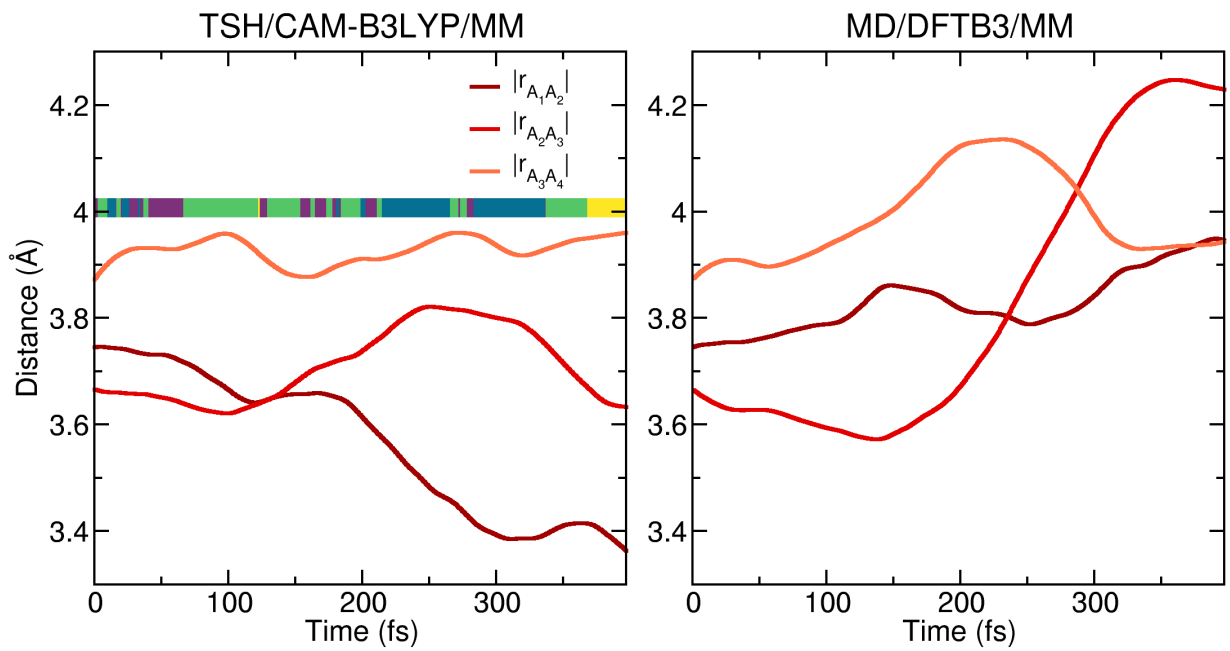

Figure S6: Spatial separation of the four QM nucleobases along a single trajectory upon excitation with trajectory surface hopping (left) and in the ground state. The color bar in the left plot indicates the state occupied by the trajectory (blue – monomer-like, palatinate – exciton, green – excimer, yellow charge transfer state).

A more direct comparison can be obtained looking at the inter base separation of a TSH trajectory versus the identical trajectory propagated in the ground state, see Fig. S6. Looking at the color bar in the left plot, it can be seen that towards the end of the run, the trajectory transitions into a charge transfer state (yellow). This occurs once the inter base distance between adenine 1 and 2 (involved in the charge transfer) is significantly reduced. Comparing the evolution of the spatial separation in the ground and excited state, it can be observed, that during the first 100 fs, both dynamics behave almost identically. However, subsequently the inter base distances in the TSH are reduced or stay approximately the same, while in the ground state dynamics they tend to increase. Thus, we can conclude, that at the earliest times after the initiation of the TSH dynamics, geometrical rearragements occur as a result of the change in level of theory from the sampling, while at later times, we see the effect of the excited state dynamics on the nuclear configuration.

Indeed, this is in line with the observations discussed in the main text, where strong

fluctuations in the base separations (of bases involved in the excited states) are seen during the early 100-150 fs of the dynamics, while at later times, the average separations of the different excited states seem to plateau and stabilise.

References

(S1) Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Gohlke, H.; Goetz, A.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C.; Smith, W.; Swails, J.; Walker, R.; Wang, J.; Wolf, R.; Wu, X.; Kollman, P. A. AMBER 14. 2014.

(S2) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713, PMID: 26574453.

(S3) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(S4) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **2004**, *393*, 51–57.

(S5) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *The Journal of Chemical Physics* **1992**, *97*, 2571–2577.

(S6) Neese, F. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *2*, 73–78.

(S7) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chemical Physics* **2009**, *356*, 98–109.

(S8) Petrenko, T.; Kossmann, S.; Neese, F. Efficient time-dependent density functional theory approximations for hybrid density functionals: analytical gradients and parallelization. *The Journal of Chemical Physics* **2011**, *134*, 054116.

(S9) Harger, M.; Li, D.; Wang, Z.; Dalby, K.; Lagardère, L.; Piquemal, J.-P.; Ponder, J.; Ren, P. Tinker-OpenMM: Absolute and relative alchemical free energies using AMOEBA on GPUs. *Journal of Computational Chemistry* **2017**, *38*, 2047–2055.

(S10) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition* **2009**, *48*, 1198–1229.

(S11) Nogueira, J. J.; Plasser, F.; González, L. Electronic delocalization, charge transfer and hypochromism in the UV absorption spectrum of polyadenine unravelled by multiscale computations and quantitative wavefunction analysis. *Chemical Science* **2017**, *8*, 5682–5691.

(S12) Hirata, S.; Head-Gordon, M. Time-dependent density functional theory within the Tamm–Dancoff approximation. *Chemical Physics Letters* **1999**, *314*, 291 – 299.

(S13) Casida, M. E. *Recent Advances in Density Functional Methods*; World Scientific: Singapore, 1995; pp 155–192.

(S14) Crespo-Hernández, C.; Cohen, B.; Kohler, B. Base stacking controls excited-state dynamics in A·T DNA. *Nature* **2005**, *436*, 1141–4.

(S15) Takaya, T.; Su, C.; Harpe, K. d. L.; Crespo-Hernández, C. E.; Kohler, B. UV excitation of single DNA and RNA strands produces high yields of exciplex states between two

stacked bases. *Proceedings of the National Academy of Sciences* **2008**, *105*, 10285–10290.

(S16) Borrego-Varillas, R.; Cerullo, G.; Markovitsi, D. Exciton Trapping Dynamics in DNA Multimers. *The Journal of Physical Chemistry Letters* **2019**, *10*, 1639–1643.

(S17) Richter, M.; Marquetand, P.; González-Vázquez, J.; Sola, I.; González, L. SHARC: ab Initio Molecular Dynamics with Surface Hopping in the Adiabatic Representation Including Arbitrary Couplings. *Journal of Chemical Theory and Computation* **2011**, *7*, 1253–1258.

(S18) Mai, S.; Marquetand, P.; González, L. A general method to describe intersystem crossing dynamics in trajectory surface hopping. *International Journal of Quantum Chemistry* **2015**, *115*, 1215–1231.

(S19) Plasser, F.; Granucci, G.; Pittner, J.; Barbatti, M.; Persico, M.; Lischka, H. Surface hopping dynamics using a locally diabatic formalism: Charge transfer in the ethylene dimer cation and excited state dynamics in the 2-pyridone dimer. *The Journal of Chemical Physics* **2012**, *137*, 22A514.

(S20) Plasser, F.; Ruckenbauer, M.; Mai, S.; Oppel, M.; Marquetand, P.; González, L. Efficient and Flexible Computation of Many-Electron Wave Function Overlaps. *Journal of Chemical Theory and Computation* **2016**, *12*, 1207–1219.

(S21) Granucci, G.; Persico, M.; Zoccante, A. Including quantum decoherence in surface hopping. *The Journal of Chemical Physics* **2010**, *133*, 134111.

(S22) Prlj, A. NEWTON-X forum: Implementation of decoherence correction (2019).

(S23) Plasser, F.; Aquino, A. J. A.; Hase, W. L.; Lischka, H. UV Absorption Spectrum of Alternating DNA Duplexes. Analysis of Excitonic and Charge Transfer Interactions. *The Journal of Physical Chemistry A* **2012**, *116*, 11151–11160.

(S24) Mewes, S. A.; Plasser, F.; Dreuw, A. Communication: Exciton analysis in time-dependent density functional theory: How functionals shape excited-state characters. *The Journal of Chemical Physics* **2015**, *143*, 171101.

(S25) Barbatti, M.; Aquino, A. J. A.; Szymczak, J. J.; Nachtigallová, D.; Hobza, P.; Lischka, H. Relaxation mechanisms of UV-photoexcited DNA and RNA nucleobases. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 21453–21458.