Supporting Information for "Resolving the Benzophenone DNA-Photosensitization Mechanism at QM/MM Level"

Elise Dumont[a]*, Meilani Wibowo[b], Daniel Roca-Sanjuán[b], Marco Garavelli[a,c], Xavier Assfeld[d,e], Antonio Monari[d,e]*

a) Laboratoire de Chimie, UMR 5182 CNRS et Ecole Normale Supérieure de Lyon, 46 allée d'Italie 69364, Lyon Cedex 07, France. b) Instituto de Ciencia Molecular, Universitat de València, P.O. Box 22085, ES-46701 València, Spain. c) Dipartimento di Chimica G. Ciamician, Università di Bologna, via Selmi 2 Bologna, Italy. d) Université de Lorraine - Nancy, Theory-Modeling-Simulation SRSMC, Boulevard des Aiguillettes 54506 Vandoeuvre-lès-Nancy, France. e) CNRS, Theory-Modeling-Simulation SRSMC, Boulevard des Aiguillettes 54506 Vandoeuvre-lès-Nancy, France.

1. Methodological Details

Our calculations were started from two representative snapshots of the stable interaction modes of benzophenone interacting with a poly(dA-dT) decamer, either minor-groove binding or double-insertion. The formers were obtained from the 10 ns corresponding MD trajectories of the DNA/BP system, described with the Amber ff99/bsc0 force field and explicitly solvated in a TIP3P water box. They were chosen as the lowest RMSD structures, in order to obtain meaningful starting structures for the static exploration of the triplet energy transfer.

We first extracted the benzophenone and the most proximal thymine (T4), corresponding to a 5' position (see Figure 1), to build up a reference BP+T model in gas-phase (i.e., without any mechanical or electrostatic embedding). To define the geometry of the reactants ($Q(^3BP)$) and products ($Q(^3T)$) along the $^3BP \rightarrow ^3T$ path we performed a Time-Dependent DFT (TD-DFT) optimization of the first (BP centered) and third (T centered) triplet state of the BP+T system. In addition to the gas-phase BP+T systems we also optimized singlet and triplet state geometries for two isolated non-interacting BP and T to be used as a reference. Subsequently, to obtain the starting optimized $Q(^3BP)$ and $Q(^3T)$ geometries in the DNA environment, hybrid QM/MM calculations within the hydrogen link-atom and electrostatic embedding scheme were performed through our COBRAMM interface. Gaussian09 and AMBER packages for the quantum mechanics and the MM calculations were used respectively. The hybrid *meta* GGA M06-2X functional was chosen in line with its performance to describe the vertical excitations of benzophenone, and also due to its ability to reproduce singlet-triplet excitations [1]. It should be recalled that this functionals has the same general behavior as long-range corrected ones, in particular in correcting for the charge-transfer and Rydberg states. Note also that, accordingly to Zhao and Truhlar, dispersion may be considered to be implicitly included by using M06-2X. Additional calculations (Table S1) were performed using long-range and dispersion-corrected functionals on the vertical excitation of benzophenone (BP), thymine (T) and adenine (A) confirming a general excellent performance of M06-2X in recovering triplet energies.

To obtain the initial $Q(^3BP+^1T)$ and $Q(^3T+^1BP)$ geometries In a first calculation, the benzophenone was selected as the high-level QM subsystem, whereas the rest of the macromolecule was treated at the force field level. Geometry optimizations for this system were performed for the singlet ground state and the lowest triplet state. In a second calculation the T4 thymine was now treated at the QM level while BP as well as the rest of the DNA and its counterions were treated classically. Once again the geometries of the ground and the first triplet state have been optimized. Instead of placing a single QM/MM frontier along the N-glycosidic linkage, here we prefer to use a double partition along the deoxyribose moiety that is far more distant from the localization of the triplet-induced structural deviations (see Figure S1). In order to assess the influence of electrostatic embedding (EE), single point energies for the equilibrium geometries have also been calculated at mechanical embedding (ME) level.

Upon geometry optimization in DNA and in model system we may see that the geometry of [3]BP features only a slightly more pronounced lengthening of the carbonyl double bond (d(C=O)=1.34 Å), and the B-helix embedding results in a dihedral compression, evidenced from the decrease of the dihedral angle between the phenyl rings (τ) from 48.8 to 30.1 degrees. Triplet states energies were calculated as vertical excitation energies along the Q(ξ) coordinate using time dependent DFT (TD-DFT) methodology with the same basis and the same functional. In this case to model the energy transfer the QM partition was constituted of both BP and the proximal thymine (T4), again the two link atom scheme was used to cap the thymine dangling bonds.

In addition to TD-DFT the potential energy curve along the Q(ξ) coordinate was calculated at multiconfiguration perturbation theory level, using the Molcas 7.6 code. The multiconfigurational complete-active-space SCF and corresponding second-order perturbation theory (CASSCF/CASPT2)[3,4] and restricted-active-space SCF (RASSCF)[5,7] energies of the three lowest-lying triplet states of BP and T embedded in a hydrated DNA double helix have been computed. The lowest triplet state has been further characterized with the RASPT2 method. The ANO-S-VDZP basis set has been used. In the CASPT2 and RASPT2 computations, an imaginary level-shift of 0.2 a.u. has been employed to prevent the occurrence of intruder states.[8] The ionization potential electron affinity (IPEA) shift has been set to 0.00 in all the reported calculations.[9] Both the single-state (SS) and multistate (MS) CASPT2 procedures have been considered.[10]

The energy transfer process involves the lowest-lying triplet state of each monomer, which are the nπ* state of BP and the $\pi\pi$* state of T. These states are mainly characterized by an excitation from the oxygen lone pair orbital (n_O) to the π* LUMO orbital, in BP, and from the π HOMO to π* LUMO, in T. Therefore, these orbitals are expected to be the most relevant orbitals for describing the triplet energy transfer process. Nevertheless, other valence π and π* orbitals might contribute significantly to the states. Hence, in order to provide an accurate description, a systematic analysis has been performed by using four levels of theory in which new sets of correlated pairs of π and π* orbitals of the dimer are added in each step. The first level of theory corresponds to the inclusion in the CAS space of only the n_O and π* LUMO of BP and the π HOMO and π* LUMO of T, representing therefore a CASPT2(6,5) level. For the next levels, two and four couples of π and π* orbitals are added [CASPT2(10,9) CASPT2(14,13) levels, respectively]. Finally, the RASPT2 method has been employed to account for all the valence π and π* orbitals, adding the five most relevant orbitals in the RAS2 space and 7 π orbitals and 11 π* orbitals in the RAS1 and RAS3 spaces, respectively [RASPT2(20,2,2;7,5,11]. A maximum of two holes and two electrons are allowed in the RAS1 and RAS3 spaces, respectively, to generate the configuration state functions. As evidenced in the results, and despite the increased computational cost the inclusion of all the valence π-orbital via the RASPT2 procedures seems important to recover the main features of the energy transfer process and in particular the stabilization of the [3]T state.

Calibration of the M06-2X functional performance vs. CASPT2

	BP	T	A
M06-2X	3.03	3.08	3.60
ωB97X-D	2.97	2.96	3.43
CAM-B3LYP-D3BJ	2.91	2.90	3.36
CASPT2	3.11-2.85[10]	3.59-2.87[11]	4.00-3.36[11]

Table S1: Vertical excitation energy (eV) for the first triplet state centered on benzophenone (BP), thymine (T) and adenine (A) calculated with different functionals, and the range of vertical and adiabatic excitation energies (eV) obtained with the CASPT2 method (see Ref. 10.and 11). We may notice that the adenine triplet state lays systematically higher in energy compared to the one on BP, hence making the sensibilization of A energetically unfavorable.

Estimation of Dexter's exchange rate

The Dexter short-range charge transfer is the dominant transfer process between donor and acceptor at close distance. Typically it is estimated to dominate until distances of about 10 Å, and is basically determined by an exchange of electrons between the two chromophores. Therefore, the mechanism is quite different from the Förster resonance that is produced by the coupling between monomers' dipole moments. Hence, the key factor determining the Dexter transfer rate is the overlap between the donor and acceptor electronic densities.

However, phenomenological the Dexter transfer rate is expressed as

$$k_{et} \propto e^{-2R_{DA}/L}$$

where R_{DA} is the distance between the chromophore and L the sum of the van der Walls spheres.

The value of R_{DA} has been straightforwardly obtained from our molecular dynamic trajectory while the value of L has been taken from the force field values.

In order to give an estimation of the donor/acceptor wavefunctions overlap we relied on the description of the excited states provided by the NTOs. Indeed the Dexter mechanism can be simplified as an exchange between the virtual NTOs of the donor and acceptor and between the corresponding occupied NTOs. The total overlap S_{et} is therefore calculated as:

$$S_{et} = S_{occ}^{et} + S_{virt}^{et}$$

$$S_{occ/virt}^{et} = \int |\phi_{occ/virt}^{D}| \cdot |\phi_{occ/virt}^{A}| \cdot dxdydz$$

where the integral has been calculated on the real 3D space expressing the virtual and occupied NTOs into a grid of point following the Gaussian cube file format. S_{et} has been evaluated at the crossing point, i.e. at the top of the energetic barrier for the ^{3}BP to ^{3}T transfer.

Scheme S1. Two link atoms partition for the thymine T4, along the C5'-O4' and C1'-C2' covalent linkages. A link-atom scheme is used, and a total charge of 0.1893 is redistributed equally over the nucleoside monophosphate.
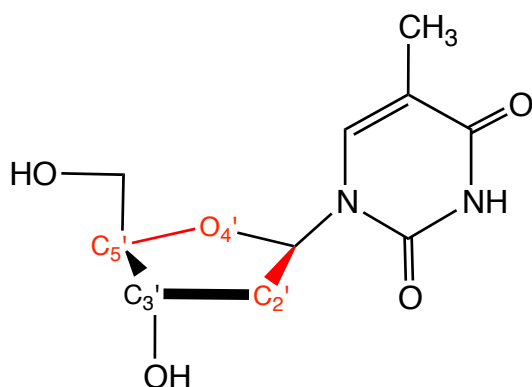
Scheme S2. Energy Diagram for the lowest-lying triplet states (double-insertion) calculated at the Q(^{3}T+^{1}BP) geometry. Values in DNA for both electrostatic and mechanical embedding schemes (EE and ME) are given for comparison.
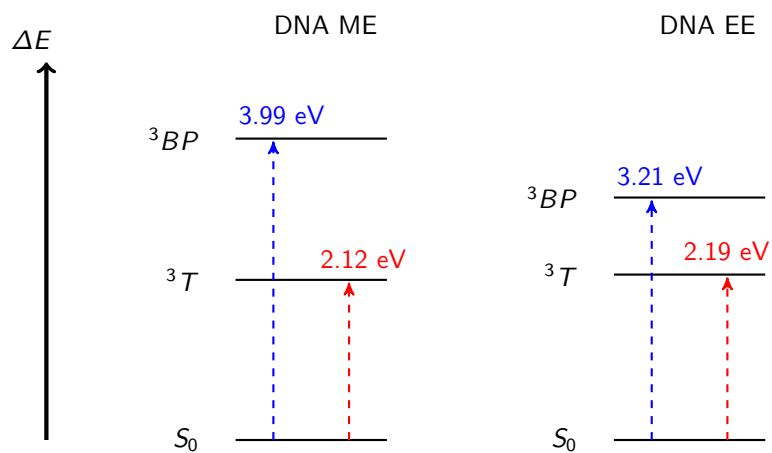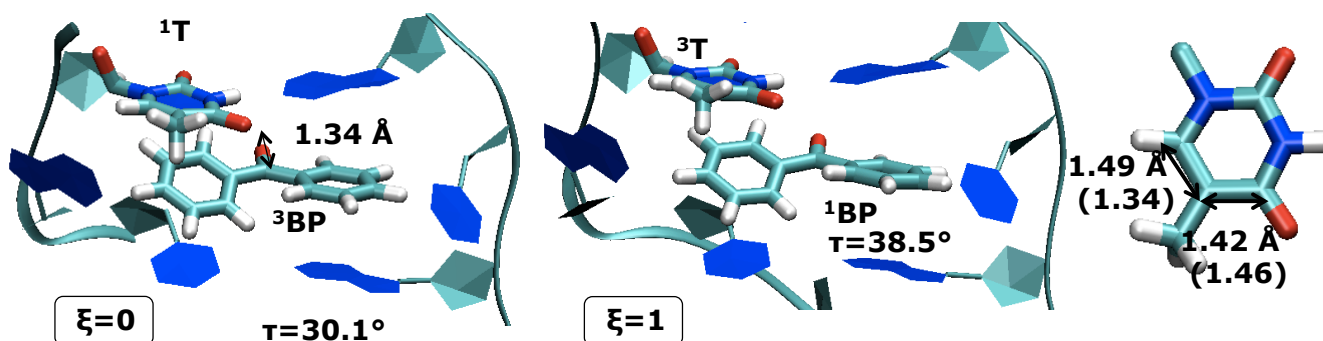
Scheme S3. Energy Diagram for the lowest-lying triplet states (double-insertion) calculated at the Q(^{1}T+^{3}BP) geometry. Values in DNA for both electrostatic and mechanical embedding schemes (EE and ME) are given for comparison.
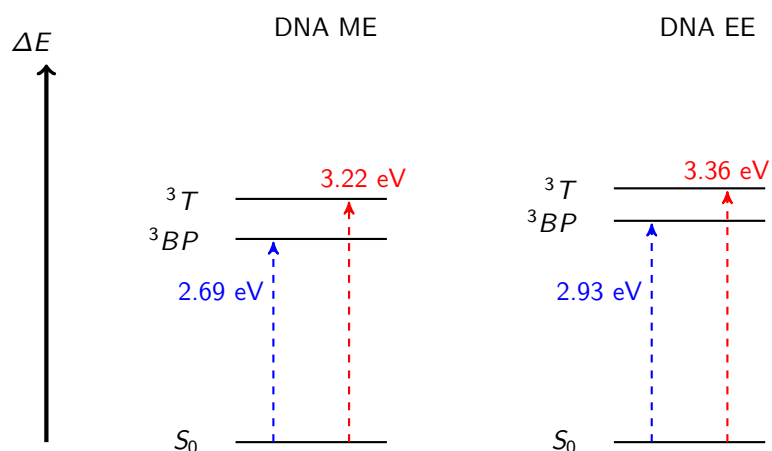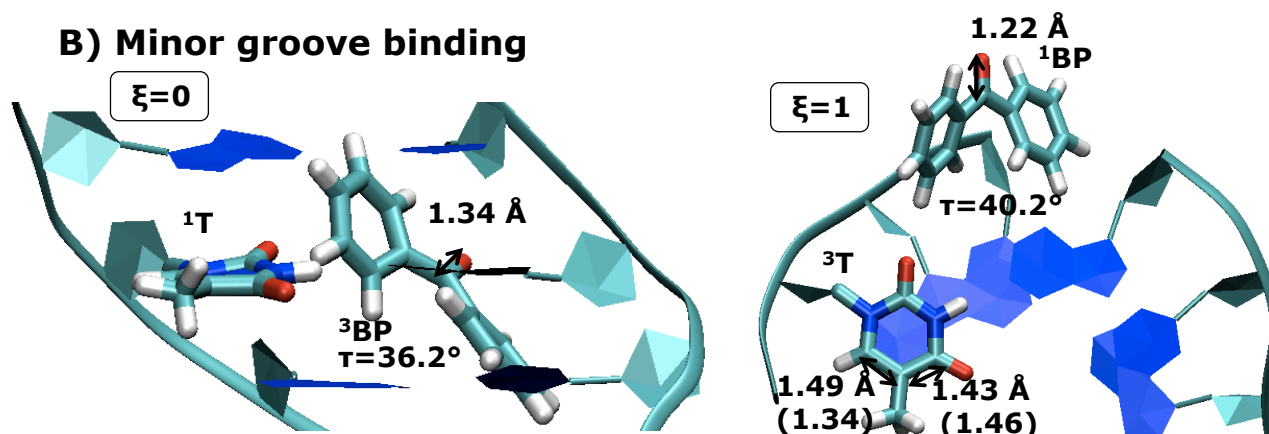

Scheme S4: Cartoon representation for the geometries at ξ=0.0 and ξ =1.0 for the two two stable interaction modes of BP-DNA.
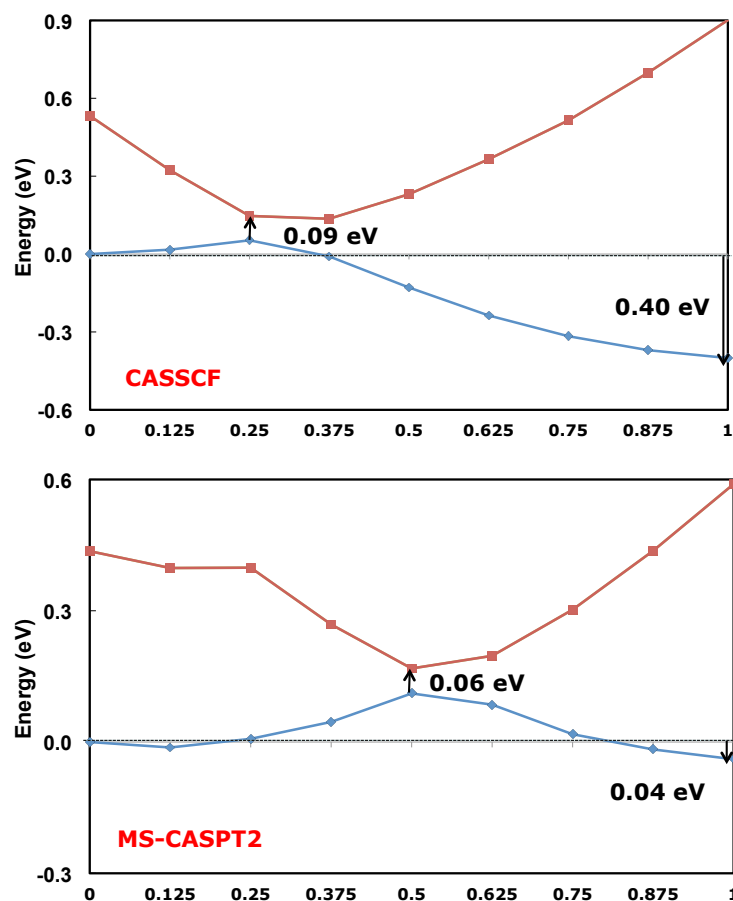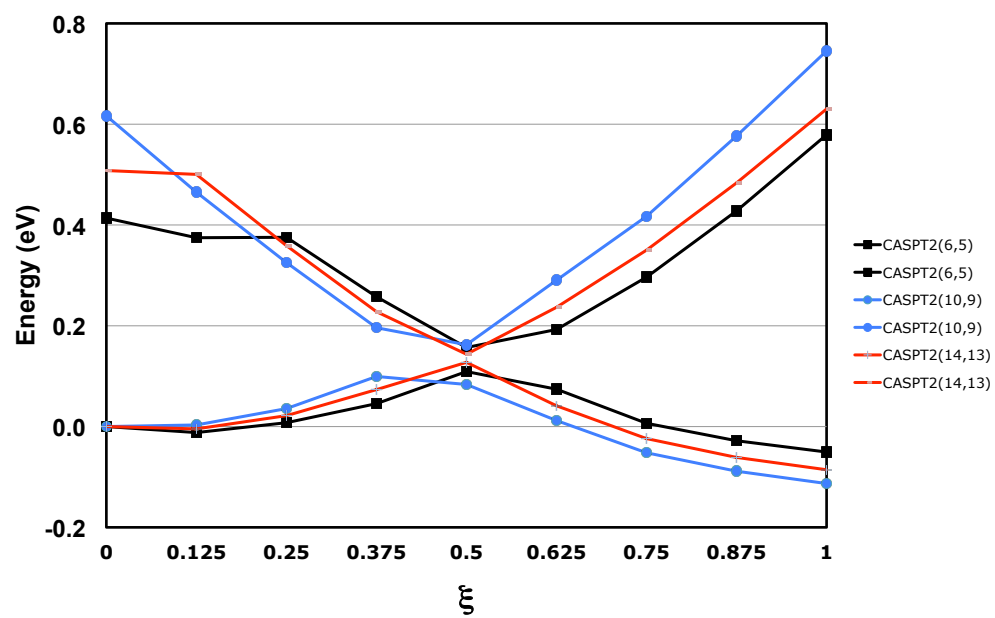
A) Double inserted mode

B) Minor groove binding

Scheme S5. Adiabatic potential energy evolution of the six first triplet excited states for $^3BP \rightarrow {}^3T$ transfer for a gas-phase {BP+T} system. A cartoon representation is given for the model system at ξ=1.0, where one visualizes the formation of a non Watson and Crick weak interaction at 2.06 Å. This feature is reproduced by M06-2X optimization but is also obtained with other density functionals [1]. In particular the gas phase models lacks the mechanical constrains imposed by the DNA macromolecular structures, and hence the mobility of the two interacting moieties is much higher, this biases the description of equilibrium geometries and artificially stabilizes ^3T.
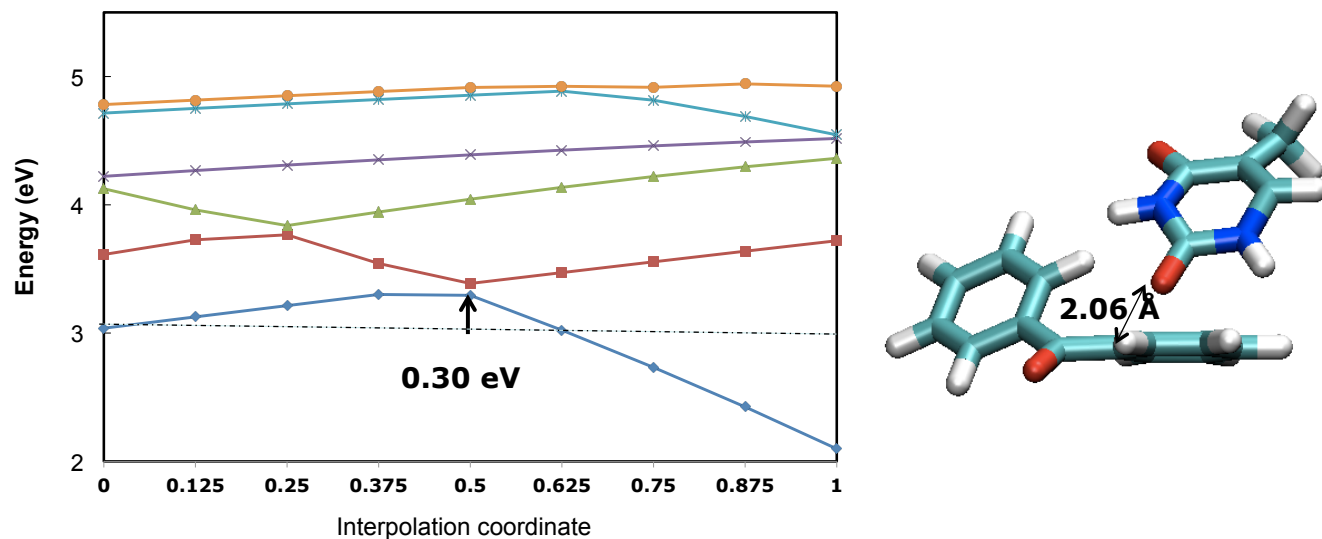
Scheme S6. CAS-SCF and MS-CASPT2 curves for the two first triplet states obtained with a (14,13) active space.

Scheme S7. CASPT2 curves obtained with increasing active spaces (see text for description).

Table S2-a: Values of the energies (in eV) along the TDDFT pathway for the double inserted mode for the six first excited states. The x,y,z components of the dipole moment are given in Debye at ξ=0.0.

ξ	T1	T2	T3	T4	T5	T6
0.000	0.0000	0.4375	0.4690	0.7538	1.1772	1.2344
	(1.806, -2.322, -1.901)	(2.092, -1.777, -1.416)	(1.833, -2.027, -1.532)	(0.486, -2.415, -1.486)	(3.356, -2.667, -2.440)	(1.945, -2.227, -1.741)
0.125	0.0426	0.3397	0.4832	0.7731	1.1898	1.2482
0.250	0.0730	0.2038	0.5135	0.7838	1.2017	1.2584
0.375	0.0598	0.1151	0.5574	0.8025	1.2142	1.2777
0.500	-0.0878	0.1450	0.5886	0.8193	1.2203	1.2862
0.625	-0.2407	0.1861	0.6296	0.8360	1.2335	1.2988
0.750	-0.4024	0.2182	0.6611	0.8533	1.2423	1.3141
0.875	-0.5684	0.2551	0.6947	0.8670	1.2552	1.3273
1.000	-0.7401	0.2881	0.7253	0.8809	1.2649	1.339

Table S2-b: Values of the energies (in eV) along the TDDFT pathway for the minor-groove binding mode for the six first excited states. The x,y,z components of the dipole moment are given in Debye at ξ=0.0.

ξ	T1	T2	T3	T4	T5	T6
0.000	0.0000	0.7894	1.2089	1.2621	1.3776	1.7864
	(2.242, -2.047, -2.634)	(2.566, -1.226, -2.347)	(2.466, -1.347, -3.030)	(2.404, -2.169, -2.546)	(2.363, -1.504, -3.783)	(2.353, -2.505, -2.288)
0.125	0.0314	0.7024	0.9185	1.1533	1.3216	1.6727
0.250	0.0972	0.6488	0.6645	1.0752	1.2990	1.5915
0.375	0.2033	0.4427	0.6524	1.0539	1.3205	1.5625
0.500	0.2700	0.3363	0.7031	1.0692	1.3826	1.5742
0.625	0.1210	0.4980	0.8127	1.1320	1.4840	1.6328
0.750	0.0115	0.6645	0.9722	1.2268	1.6093	1.7245
0.875	-0.0654	0.8509	1.1946	1.3698	1.7756	1.8605
1.000	-0.1022	1.0659	1.4575	1.5642	1.9802	2.0440

Table S2-c: Values of the energies (in eV) along the CASPT2(14,13) pathway for the minor-groove binding mode for the six first excited states. The dipole moment and its x,y,z components are given in Debye at ξ=0.0.

ξ	T1	T2
0.000	0.0000	0.4138
	7.84 (5.458, -4.214, -3.740)	7.69 (4.572, -4.825, -3.869)
0.125	-0.0120	0.3748
0.250	0.0077	0.3756
0.375	0.0455	0.2575
0.500	0.1091	0.1569
0.625	0.0742	0.1926
0.750	0.0066	0.2965
0.875	-0.0281	0.4281
1.000	-0.0505	0.5790

REFERENCES

(1)   Jacquemin D, Perpete E., A. Ciofini I, Adamo C. Assessment of Functionals for TD-DFT Calculations of Singlet-Triplet Transitions *J. Chem. Theor. Comput*. **2010**, *6*, 1532-1537

(2)   Dupont C.; Patel C.; Dumont, E. Improved DFT description of intrastrand cross-links formation by inclusion of London dispersion corrections. *J. Phys. Chem. B.* **2011**, *115*, 15138-15144

(3)   Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order Perturbation Theory with a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(4)   Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. Second-order Perturbation Theory with a Complete Active Space Self-consistent Field Reference Function. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(5)   Malmqvist, P. Å.; Pierloot, K.; Shahi, A. R. M.; Cramer, C. J.; Gagliardi, L. The Restricted Active Space Followed by Second-Order Perturbation Theory Method: Theory and Application to the Study of CuO_2 and Cu_2O_2 Systems. *J. Chem. Phys.* **2008**, *128*, 204109.

(6)   Sauri, V.; Serrano-Andrés, L.; Shahi, A. R. M.; Gagliardi, L.; Vancoille, S.; Pierloot, K. Multiconfigurational Second-Order Perturbation Theory Restricted Active Space (RASPT2) Method for Electronic Excited States: A Benchmark Study. *J. Chem. Theor. Comput.* **2011**, *7*, 153-168.

(7)   González-Luque, R.; Climent, T.; González-Ramírez, I.; Merchán, M.; Serrano-Andrés, L. Singlet-Triplet States Interaction Regions in DNA/RNA Nucleobase Hypersurfaces *J. Chem. Theor. Comput.* **2010**, *6*, 2103-2114

(8)   Forsberg, N.; Malmqvist, P.-Å. Multiconfigurational perturbation theory with imaginary level shift. *Chem. Phys. Lett.* **1997**, *274*, 196–204

(9)   Ghigo, G.; Roos, B. O.; Malmqvist, P.-Å. A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2) *Chem. Phys. Lett.* **2004**, *396*, 142–149

(10) Finley, J.; Malmqvist, P.-Å.; Rood, B. O.; Serrano-Andrés, L. *Chem. Phys. Lett.* **1998**, *288*, 299-206

(11) Sergentu, D.-C. ; Maurice, R.; Havenith R. W. A.; Broer, R.; Roca-Sanjuán, D. Computational Determination of the Dominant Triplet Population Mechanism in Photoexcited Benzophenone *Phys. Chem. Chem. Phys.* **2014**, *16*, 25393-25403