Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2018

Supporting Information for:

Dynamics of the excited-state hydrogen transfer in a (dG)•(dC) homopolymer: Intrinsic photostability of DNA

Antonio Francés-Monerris,[a]* Hugo Gattuso,[b] Daniel Roca-Sanjuán,[c] Iñaki Tuñón,[d] Marco Marazzi,[e] Elise Dumont,[f] and Antonio Monari[a]*

*E-mails: Antonio.Frances@univ-lorraine.fr, Antonio.Monari@univ-lorraine.fr

a. Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques, Boulevard des Aiguillettes, BP 70239, 54506, Vandoeuvre-lès-Nancy (France).

b. Theoretical Physical Chemistry, Research Unit Molecular Systems (UR MOLSYS) University of Liège, 4000 Liège, Belgium

c. Instituto de Ciencia Molecular, Universitat de València, Apartado 22085, ES-46071 Paterna (Spain).

d. Departamento de Química Física, Universitat de València, 46100 Burjassot (Spain).

e. Photochemistry Group, Universidad de la Rioja, Madre de Dios, 51, E-26006, Logroño (Spain).

f. Laboratoire de Chimie, UMR 5182 CNRS et Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07 (France).

Table of Contents

Computational Details

Classical molecular dynamic simulations

The starting structure of the 14 base-pairs long (dG)·(dC) B-DNA homopolymer (see Figure 1) was built using the NAB utility[1] available in AmberTools.[2] In order to correctly explore the conformational space of the ground state via classic molecular dynamics (MD), the system was then placed at the center of a truncated octahedron containing 7309 TIP3P[3] water molecules and 26 Na^+ cations to obtain charge neutralization. The bonded and non-bonded parameters were taken from the parm99 force field[4] including bsc1 corrections for DNA[5] and the simulations were ran using the GPU cuda version of Amber16.[2] After a preliminary minimization of the energy of the system composed of 4000 steps steepest descent and 4000 steps conjugated gradient algorithms, a thermalization step was used to increase the temperature from 0 to 300 K in 200 ps in the NVT ensemble. This was followed by a 400 ps equilibration step in the NPT ensemble and the simulation was finally carried on for 100 ns in the NPT ensemble. The MD simulations were carried out under periodic boundary conditions and using particle mesh-ewald (PME) with a cutoff of 9.0 Å and a time step of 2 fs. The pressure was set to 1 atm and maintained constant using the Monte Carlo barostat, while temperature conservation was assured using Langevin dynamics. Moreover, the Shake algorithm[6] was applied to bonds involving hydrogen atoms.

QM/MM characterization of the Franck-Condon region

Two electronic structure methods have been used to study the Franck-Condon region of the (dG)·(dC) homopolymer by means of hybrid quantum mechanics/molecular mechanics (QM/MM) approach. The influence of the surroundings' electrostatic potential was taken into account by electrostatic embedding procedures, i.e. by modifying the QM Fock matrix to include the interaction with the environment MM point charges. In particular, the *ab initio* multiconfigurational complete-active-space second-order perturbation theory (CASPT2)[7] and the time-dependent density functional theory (TD-DFT)/CAMB3LYP methods (hereafter, CASPT2/MM and TD-CAMB3LYP/MM methods, respectively) have been used. 11 equidistant snapshots separated by 10 ns were obtained from the MM simulations of the (dG)·(dC) homopolymer as random representations of the full conformational space of the B-DNA double strand (see Figure S1). Excited states from Franck-Condon region were studied considering a QM partition involving a tetramer composed by two guanine and the two paired cytosine moieties. The backbone, *i.e.* sugars and phosphate, have not been included in the QM partition and dangling covalent bonds at the QM/MM frontier have been treated with the Link Atom (LA) approach.

Due to computational cost reasons, in the CASPT2/MM calculations of each snapshot, the DNA tetramer super-system was broken down into two sets of subsystems (see Figure S2). One describes the excitations in a guanine-guanine/cytosine trimer [$(G/CC)_{QM}$, the slash denoting the π-stacked C/C subsystem], whereas the other one includes two π-stacked guanine molecules and an opposite cytosine [$(GG/C)_{QM}$]. See Table S1 for more details about the partitioning schemes. The multiconfigurational wave functions were built using the complete-active-space self-consistent field (CASSCF) method distributing 12 electrons among the most relevant two occupied π orbitals and the most important two π* orbitals of each nucleobase, *i.e.* CASSCF(12,12), as calibrated in previous studies.[8] In some snapshots using the $(G/CC)_{QM}$ partitioning scheme, only one π* orbital of guanine was included in the active space. Extra calculations including two π* orbitals of guanine showed

negligible energy differences of ~0.2 eV. Six roots were demanded in the state-average (SA)-CASSCF procedure, and the double-ξ plus polarization basis set ANO-S-VDZP was used throughout. CASPT2 calculations were performed on top of the multiconfigurational SA-CASSCF wave functions in order to account for the missing dynamic electron correlation. All the core electrons were maintained frozen during the perturbation step. The original zeroth-order Hamiltonian was used by setting the ionization-potential electron-affinity (IPEA) parameter to 0.0 a.u.[9] The recommended value of 0.2 a.u. for the imaginary shift was used to minimize the presence of weakly-interacting intruder states.[10] Details on the determination of the oscillator strengths (*f*) can be found elsewhere.[8,11] All multiconfigurational calculations have been conducted with the MOLCAS 8 software package.[12]

For the TD-CAMB3LYP/MM calculations, the four nucleobases were included in the QM part, *i.e.* $(GG/CC)_{QM}$ partitioning scheme (see Table S1). A total of 10 singlet excited states were computed in the linear-response formulations calculations. The 6-31G basis set was used throughout as an acceptable compromise between accuracy and computational cost, given the lack of significant differences when comparing the results with larger basis sets (see Figure S3 for calibration studies). Moreover, the use of diffuse functions in QM/MM framework can lead to unphysical overpolarization due to a too strong coupling of the atomic orbitals with the MM point charges. All TD-DFT calculations were performed using the GAUSSIAN 09 suite of programs.[13]

Adiabatic QM/MM molecular dynamics

Time evolution of the S_0 and S_1 states of the selected snapshots have been explored by means of electrostatic embedding QM/MM molecular dynamics simulations of the (dG)·(dC) B-DNA homopolymer in water solution (see Figure 1). One, two, and four consecutive base pairs have been included in the QM partition [hereafter, $(G\text{-}C)_{QM}$, $(GG/CC)_{QM}$, $(GGGG/CCCC)_{QM}$ partitioning respectively, see Table S1] while the remaining atoms have been treated at MM level. Newton's equations of motion were solved using a time step of 1 fs. The chosen QM methods for the simulations were the CAMB3LYP/MM for S_0 and the TD- CAMB3LYP/MM for the S_1 state as implemented in the GAUSSIAN 09 suite of programs.[13] The force field was consistent with the one used for the classical MD, i.e. parm99[4] with bsc1 corrections for DNA.[5] Nonpolarizable force fields typically account satisfactorily for equilibrium and dynamical properties of sodium ions in water solution.[14–16] The AMBER/GAUSSIAN interface[17] was used to perform the QM/MM dynamics. Thereby, energy and the molecular gradients for the QM part were solved by GAUSSIAN 09[13] while nuclei propagations were performed by AMBER 16[2] combining both QM and MM gradients. A cutoff of 9 Å was used for all QM-MM interactions. The 6-31G basis set was used throughout.

A total of 45 trajectories were run (22 in S_0 and 23 in S_1, results compiled in Tables 1 and 3) starting from the 11 different initial DNA arrangements. The initial velocities for the QM/MM trajectories were randomly generated by the AMBER 16 program at the first step of the simulation from a distribution corresponding to a temperature of 300 K. No external force was applied to drive the decay paths.

Snapshots of the trajectories

$(G\text{-}C)_{QM}$ trajectories. TD-CAMB3LYP/MM single-point calculations of the snapshots extracted from the $(G\text{-}C)_{QM}$ trajectories were computed using the TERACHEM[18–20] program interfaced with AMBER 16. Energy differences at the TD-DFT level must be considered as upper-bound estimations of the actual energy gap due to the limited accuracy of this method to represent the correct dimensionality of CI seams.[21] Following the report by Levine et al[22] the S_1/S_0 crossings at the TD-DFT level will be called intersections. To assess the description of the ESHT processes by the

TD-CAMB3LYP/MM method, some energy differences of selected QM/MM snapshots of the $(G\text{-}C)_{QM}$ trajectories were recomputed with the CASPT2/MM method. The active space for the $(G\text{-}C)_{QM}$ scheme comprised four π and three π* molecular orbitals (MOs) of the guanine moiety plus two π and three π* MOs localized on the cytosine molecule, *i.e.* 12 electrons distributed into 12 MOs, as reported in a previous work.[11] Four roots were demanded in the state-average (SA)-CASSCF procedure due to the convergence of the S_1 energy upon increasing the number of states (see Table S2). The atomic natural orbital (ANO) L-type[23] with the contraction scheme C, N, O [4s3p1d]/H [2ps1p] (hereafter, ANO-L 431/21) has been used in these calculations as an acceptable compromise between accuracy and computational cost. The smaller ANO-S with a double-ξ plus polarization basis set (ANO-S-VDZP) was also calibrated (see Table S2).

Results on the TD-CAMB3LYP/MM calibration are shown in Figures S4 and S5, and confirm that while TD-CAMB3LYP/MM, as expected, tends to slightly overestimate the energy gap, both methods provide, in general, a coherent representation of the energy differences. No surface hopping method was used, the decay times were estimated identifying the intersection points accessed during the S_1 dynamics. Thus, the present work does not aim to quantify the yield of occurrence of the different photoprocesses displayed by the (dG)·(dC) homopolymer but to provide a detailed description of them, shedding light in the fundamental aspects of photostability. Structures with an energy difference between the S_0 and the S_1 states larger than 0.8 eV at the TD-CAMB3LYP/MM level were not considered energy degeneracy points. On the other hand, energy differences comprised between 0.5 and 0.8 eV were recomputed using the CASPT2/MM method. Thus, only those structures below 0.5 eV (either at the TD-CAMB3LYP or the CASPT2/MM levels) were used to estimate the decay times of the $(G\text{-}C)_{QM}$ trajectories (see Figures S6-S16). Such threshold allow to account those near-degeneracy regions where the vibronic couplings might give rise to hop probabilities. It shall be noted that the dynamical effects studied in the present work, which provide the chemical and biological relevance of the paper, are unambiguously shown at the timescales yielded by the 0.5 eV threshold, even though the approach likely underestimates the provided lifetimes.

$(GG/CC)_{QM}$ and $(GGGG/CCCC)_{QM}$ trajectories. The TD-CAMB3LYP/MM method was used to compute the single-point energies of the $(GG/CC)_{QM}$ and the $(GGGG/CCCC)_{QM}$ snapshots. However, multiconfigurational calculations were required to confirm the FPT mechanism observed in the $(GGGG/CCCC)_{QM}$ trajectory. Since the *ab initio* description of eight DNA nucleobases is simply prohibitive, only the four relevant DNA nucleobases for the FPT process were included in the multiconfigurational computations. Thus, the CASPT2/MM method was actually applied using a '$(GG/CC)_{QM}$' partition scheme. Notwithstanding this approximation, these calculations are challenging due to the high number of atoms and electrons to treat and the exponential scaling of the calculations' complexity. Since only the S_0 and S_1 electronic states are of interest, three roots were demanded in the SA-CASSCF procedure in order to minimize the computational cost. The less computationally demanding ANO-S-VDZP was chosen given the negligible differences with respect to the larger ANO-L 431/21 basis set in the $(G\text{-}C)_{QM}$ calibration (see Table S2). Selection of the active space was based on including only the MOs necessary to describe the CT G→C transitions, with a size of 12 electrons distributed into 12 MOs. The SA-CASSCF natural orbitals that compose the active space are displayed in Figure S17. Surprisingly, at the majority of the computed structures, only one π* MO of the cytosine significantly contributed in the converged SA-CASSCF wave function of the S_0, S_1, and S_2 states. The remaining orbitals that compose the active space are thus three π MOs of each guanine molecule (total of six π MOs) and five π∗ MOs also localized over the guanine moieties. It is remarkably to mention that one of the latter π∗ MO has diffuse character and an occupation number of 0.01, which means that is not participating in the construction of the wave

function and that the size of the active space is large and flexible enough to describe the low-lying states at different geometries. Moreover, additional attempts to include a second π* MO of the other cytosine by setting an active space of 12 electrons distributed into 13 MOs, were not successful. Therefore, one can conclude that the π→π* (G→C) transitions dominate the electronic structure of the system, followed by small contributions of the π→π* excitations localized over the deprotonated/dehydrogenated guanine molecules, whereas the π→π* transitions of the protonated/hydrogenated cytosine moieties remain higher in energy. In particular, the S_1 wave function of the CI structure shown in Figure 4b (t = 87 fs) is dominated by the $\pi_1 \rightarrow \pi_1^*$ transition (weight *ca.* 80%), whereas the S_2 state, which lies only 0.34 eV higher in energy, is dominated by the $\pi_2 \rightarrow \pi_1^*$ one-electron excitation (weight *ca.* 90%). For these reasons, all the MOs but one localized over the cytosine molecules remain out of the active space, as repeatedly observed for the set of structures reported in Figure 4b.

Results on the Franck-Condon region

The main discrepancy between both TD-CAMB3LYP and the CASPT2/MM methods arises at the gas-phase optimized Franck-Condon geometry, where the TD-CAMB3LYP method tends to invert the order of the guanine locally excited and charge transfer state as compared to CASPT2 (see Table S3)[11] and equation-of-motion coupled-cluster singles, doubles, and perturbative triples [EOM-CCSD(T)] results.[24] However, since here we are specifically interested in the H-transfer process, this feature can be used to facilitate the population of the G→C CT state and therefore increase the accumulation of statistical data on the ESTH process, while making the dynamical study affordable.

The vertical absorption energies of the S_1 state for the 11 snapshots are summarised in Table S5. In the *g* snapshot, the energy of the G→C CT state does not significantly vary when comparing both $(G/CC)_{QM}$ and $(GG/C)_{QM}$ partitioning schemes, however, the inclusion of the π-stacked guanine in the latter stabilises the local excitation over guanine inducing a slight mixing of the CT state with guanine excitonic states (see figure S2). As a consequence, the charge separation decreases and the oscillator strength increases (see Table S5). The large charge separation between the stacked guanine nucleobases and the cytosine moiety confirm the dominant inter-strand character of the state. It is important to remark that, for the *g* snapshot, the bright π,π* state localised in guanine is determined at ~4.25 eV (λ = 292 nm, UVB), lying only ~0.3-0.4 eV above S_1 (see Table S3). Thus, it is reasonable to think that the G→C CT state will be accessible after light absorption, especially considering that the electron donation to the adjacent cytosine [S_2 state of the $(G/CC)_{QM}$ results, see Table S4] is located in between the S_1 G→C state and the aforementioned bright local state in guanine.

Regarding the *i* snapshot, the two guanine nucleobases participate in the G→C CT excitation (see Figure S2). Obviously, this phenomenon is missing in the $(G/CC)_{QM}$ system, in which the energy of the G→C CT state is underestimated. Nevertheless, both G-C/C and G/G-C partitioning schemes indicate that the inter-strand CT states are the lowest excited states for this particular DNA arrangements. The present study is very far from providing definitive statistical data about the energetics of the inter-strand CT state in DNA. However, taking into account that the experimental absorption spectrum of a (dG)·(dC) homopolymer composed by ~1200 base pairs evidenced a band maximum at ~255 nm (~4.86 eV),[25] it is reasonable to conclude that those particular situations in which the G→C states lie below the absorption peak are indeed frequent and hence deserve a particular attention.

We also noted that the delocalization of the different CT states changes between the different snapshots. As revealed by the analysis of the molecular charges shown in Supplementary Information, at the TD-CAMB3LYP/MM level, the inter-strand G→C CT state involves a guanine nucleobase and the cytosine of the adjacent WC base pair in snapshots *a*, *b*, *g*, and *i*. In other words, the inter-strand transfer of electron density involves two consecutive WC base pairs. On the contrary, the snapshot *c* implies the electron donation from the Frenkel exciton delocalized over the two guanine nucleobases, whereas the snapshot *k* represents the typical G→C CT state over a single WC base pair.

FPT transfer video details

The video showing the FPT transfer process (available at the journal's website) has been made combining both S_1 and S_0 dynamics of the *g* trajectory, having a total duration of 140 fs. The first 87 fs belong to the simulation in the S_1 state until the INT_2 species is formed and the $(S_1/S_0)_{CI}$ area at the CASPT2/MM level is reached (see Figure 4b). The atomic coordinates and velocities were stored and used to continue the simulation on the S_0 surface. The WC canonical DNA structure is recovered after the process.

References

1 T. J. Macke and D. A. Case, in *Molecular Modeling of Nucleic Acids*, American Chemical Society, 1997, vol. 682, pp. 24–379.

2 D. A. Case, R. M. Betz, D. S. Cerutti, I. T.E. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, *AMBER 2016, Univ. California, San Fr.*

3 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

4 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

5 I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case and M. Orozco, *Nat. Methods*, 2016, **13**, 55–58.

6 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

7 K. Andersson, P. A. Malmqvist and B. O. Roos, *J. Chem. Phys.*, 1992, **96**, 1218–1226.

8 A. Frances-Monerris, J. Segarra-Marti, M. Merchan and D. Roca-Sanjuan, *Theor. Chem. Acc.*, 2016, **135**, 31.

9 G. Ghigo, B. O. Roos and P. A. Malmqvist, *Chem. Phys. Lett.*, 2004, **396**, 142–149.

10 N. Forsberg and P. A. Malmqvist, *Chem. Phys. Lett.*, 1997, **274**, 196–204.

11 V. Sauri, J. P. Gobbo, J. J. Serrano-Perez, M. Lundberg, P. B. Coto, L. Serrano-Andres, A. C. Borin, R. Lindh, M. Merchan and D. Roca-Sanjuan, *J. Chem. Theory Comput.*, 2013, **9**, 481–496.

12 F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata and R. Lindh, *J. Comput. Chem.*, 2015, **37**, 506–541.

13 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, R. Janesko, R. Gomperts, B. Menucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, D. Sonnenberg, F. Williams-Young, F. Ding, F. Lipparini, F. Edigi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, W. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, M. Hasegawa, T. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, M. J. J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 09, Revis. D.01, Gaussian, Inc., Wallingford CT, 2013.*

14 S. Mamatkulov, M. Fyta and R. R. Netz, *J. Chem. Phys.*, 2013, **138**, 24505.

15 J. Kahlen, L. Salimi, M. Sulpizi, C. Peter and D. Donadio, *J. Phys. Chem. B*, 2014, **118**, 3960–3972.

16 E. Duboué-Dijon, P. Delcroix, H. Martinez-Seara, J. Hladílková, P. Coufal, T. Křížek and P. Jungwirth, *J. Phys. Chem. B*, 2018, **122**, 5640–5648.

17 A. W. Götz, M. A. Clark and R. C. Walker, *J. Comput. Chem.*, 2014, **35**, 95–108.

18 I. S. Ufimtsev and T. J. Martínez, *J. Chem. Theory Comput.*, 2008, **4**, 222–231.

19 I. S. Ufimtsev and T. J. Martinez, *J. Chem. Theory Comput.*, 2009, **5**, 1004–1015.

20 I. S. Ufimtsev and T. J. Martinez, *J. Chem. Theory Comput.*, 2009, **5**, 2619–2628.

21 M. Huix-Rotllant, A. Nikiforov, W. Thiel and M. Filatov, *Top. Curr. Chem.*, 2016, **368**, 445–476.

22 B. G. Levine, C. Ko, J. Quenneville and T. J. MartÍnez, *Mol. Phys.*, 2006, **104**, 1039–1051.

23 P. O. Widmark, P. A. Malmqvist and B. O. Roos, *Theor. Chim. Acta*, 1990, **77**, 291–306.

24 H. Sun, S. Zhang, C. Zhong and Z. Sun, *J. Comput. Chem.*, 2016, **37**, 684–693.

25 I. Vaya, P. Changenet-Barret, T. Gustavsson, D. Zikich, A. B. Kotlyar and D. Markovitsi, *Photochem. Photobiol. Sci.*, 2010, **9**, 1193–1195.

Tables

Table S1. Details on the QM/MM partition schemes and numbering of the nucleobases.

Partition scheme	Residues that include the nucleobases of the QM part	Total number of atoms in the QM part
$(G/CC)_{QM}$	#7 (Gua), #22 (Cyt), #21 (Cyt)	42
$(GG/C)_{QM}$	#7 (Gua), #8 (Gua), #22 (Cyt)	45
$(G\text{-}C)_{QM}$	#7 (Gua), #22 (Cyt)	29
$(GG/CC)_{QM}$	#7 (Gua), #8 (Gua), #21 (Cyt), #22 (Cyt)	58
$(GGGG/CCCC)_{QM}$	#6 (Gua), #7 (Gua), #8 (Gua), #9 (Gua), #20 (Cyt), #21 (Cyt), #22 (Cyt), #23 (Cyt)	116

(dG)·(dC) homopolymer sequence

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
5’-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	-3’
3’-	C	C	C	C	C	C	C	C	C	C	C	C	C	C	-5’
	28	27	26	25	24	23	22	21	20	19	18	17	16	15	

Table S2. CASPT2/MM vertical excitation energies (eV) of a random structure obtained from the QM/MM MD snapshots as a function of the basis set and number of states computed in the CASSCF procedure. Unless otherwise specified, results correspond to the ANO-L 431/21 basis set.

State / Number of roots	3[a]	3	4	5	10
S_1	1.29	1.27	1.25	1.03	1.20
S_2	3.00	2.97	3.07	2.52	2.64
S_3			3.37	3.61	2.97
S_4				3.62	3.90
S_5					4.34
S_6					4.52
S_7					4.62
S_8					4.68
S_9					5.30

[a]ANO-S-VDZP basis set.

Table S3. Comparison between the TD-CAMB3LYP and the CASPT2 descriptions of the Franck-Condon region of the G-C base pair in the gas phase optimized with the CASSCF method.[11]

Adiabatic State	**TD-CAMB3LYP/6-31G**		**CASPT2/ANO-S-VDZP[a]**	
	Nature	Excitation energy (*f*)	Nature	Excitation energy (*f*)
S_1	G→C	5.24 (0.0034)	G→G	4.74 (0.383)
S_2	C→C	5.41 (0.0924)	C→C	5.10 (0.207)
S_3	G→G	5.49 (0.0833)	G→C	5.14 (0.085)
S_4	n,π* (C)	5.60 (0.0016)		
S_5	n,π* (G-C)	5.87 (0.0011)		
S_6	G→G	5.96 (0.4316)		
S_7	C→C	6.03 (0.1164)		

[a]Reference [11] of ESI.

Table S4. Lowest-lying excited states of the snapshot *g* at the FC region. Mulliken charges correspond to the respective CASSCF/MM wave functions. Bright states are highlighted in bold. Note the partial mixing of the π,π* state of guanine with the G→C states.

$(G/CC)_{QM}$ CASPT2/MM				**$(GG/C)_{QM}$ CASPT2/MM**			
State	**Nature**	**Excitation energy (*f*)**	**Charges G | C C**	**State**	**Nature**	**Excitation energy (*f*)**	**Charges G | C G**
S_1	G→C	3.86 (0.0009)	+0.99 | -0.96 -0.03	S_1	G→C	3.94 (0.0512)	+0.76 | -0.76 +0.00
S_2	G→C	3.94 (0.1094)	+0.73 | -0.65 -0.08	S_2	G→G	**4.25 (0.1881)**	+0.32 | -0.24 -0.08
S_3	G→G	**4.23 (0.2753)**	+0.27 | -0.23 -0.03				

Table S5. Study of the S_1 state at the Franck-Condon region of the (dG)·(dC) homopolymer with the TD-CAMB3LYP/MM and the CASPT2/MM methods. The excitation energies (in eV), oscillator strengths (*f*) and Mulliken charges of the guanine and cytosine nucleobases correspond to the 11 snapshots obtained from the MD simulation. Mulliken charges for the CASPT2/MM results correspond to the respective CASSCF/MM wave functions.

Snapshot	**$(GG/CC)_{QM}$ TD-CAMB3LYP/MM**			**$(G/CC)_{QM}$ CASPT2/MM**			**$(GG/C)_{QM}$ CASPT2/MM**		
	Main nature of S_1	**Excitation energy (*f*)**	**Charges G | C G | C**	**Main nature of S_1**	**Excitation energy (*f*)**	**Charges G | C C**	**Main nature of S_1**	**Excitation energy (*f*)**	**Charges G | C G**
a	G→C	4.57 (0.0000)	-0.22 | -0.75 +0.81 | +0.16	G→G	4.85 (0.2341)	+0.02 | -0.01 +0.01	G→G	4.81 (0.0937)	+0.03 | -0.01 -0.01
b	G→C	4.27 (0.0002)	+0.81 | +0.17 -0.20 | -0.78	C→C	4.20 (0.0950)	+0.01 | +0.50 -0.50	G→G	3.80 (0.0232)	-0.10 | -0.03 +0.13
c	G→C	4.48 (0.0034)	+0.47 | -0.79 +0.23 | +0.09	C→C	4.27 (0.1324)	+0.01 | -0.02 +0.01	G→G	4.25 (0.1952)	+0.04 | -0.02 -0.01
d	G→G	4.86 (0.0358)	-0.01 | +0.03 -0.09 | +0.07	C→C	4.43 (0.1299)	+0.02 | -0.01 -0.01	G→G	4.41 (0.3029)	+0.01 | -0.03 +0.02
e	G→G	5.05 (0.0075)	-0.06 | +0.07 -0.10 | +0.09	G→G	4.31 (0.4699)	+0.01 | +0.00 -0.01	G→G	4.31 (0.1168)	+0.02 | -0.02 +0.00
f	G→G	4.64 (0.0015)	-0.96 | +0.01 +0.83 | +0.12	C→C	4.20 (0.0834)	+0.03 | -0.01 -0.02	G→G	4.25 (0.3018)	+0.06 | -0.03 -0.03
g	G→C	3.50 (0.0000)	-0.12 | -0.81 +0.78 | +0.15	G→C	3.86 (0.0009)	+0.99 | -0.96 -0.03	G→C	3.94 (0.0512)	+0.76 | -0.76 +0.00
h	G→G	4.21 (0.0093)	-0.93 | -0.03 +0.85 | +0.11	C→C	4.42 (0.1596)	+0.02 | -0.02 +0.00	G→G	3.87 (0.0205)	+0.99 | -0.04 -0.95
i	G→C	2.75 (0.0000)	-0.10 | -0.83 +0.87 | +0.07	G→C	2.49 (0.0000)	+1.04 | -1.00 -0.04	G→C	3.19 (0.0050)	+1.04 | -1.02 -0.02
j	G→G	5.10 (0.0305)	-0.09 | +0.09 +0.07 | -0.07	C→C	4.53 (0.2440)	+0.01 | -0.01 +0.00	G→G	4.92 (0.1013)	+0.01 | -0.03 +0.02
k	G→C	4.27 (0.0001)	+0.66 | -0.80 +0.08 | +0.06	C→C	4.48 (0.1306)	+0.03 | -0.04 +0.01	G→G	4.70 (0.3010)	+0.04 | -0.04 +0.00

Table S6. CASSCF/MM Mulliken charges for the $(GG/CC)_{QM}$ tetramer involved in the FPT mechanism displayed in Figure 4b of main article (run *g*). Results correspond only to some selected snapshots. Nucleobase labels are shown in Figure 3, and the states of interest are highlighted in bold.

Frame (fs)	**S_0**		**S_1**		**S_2**	
	GUA B	CYT B	GUA B	CYT B	GUA B	CYT B
	GUA A	CYT A	GUA A	CYT A	GUA A	CYT A
74	-0.94	0.94	**-0.92**	**-0.06**	0.03	-0.05
	-0.83	0.83	**0.15**	**0.83**	-0.80	0.83
79	-0.99	0.98	**-0.99**	**-0.01**	0.00	-0.01
	-1.06	1.06	**-0.06**	**1.06**	-1.05	1.06
84	0.03	-0.04	**-0.95**	**-0.05**	0.03	-0.04
	-0.86	0.88	**0.12**	**0.88**	-0.86	0.88
85	-0.95	0.95	**-0.94**	**-0.05**	0.02	-0.04
	-0.88	0.88	**0.11**	**0.88**	-0.86	0.88
86	-0.96	0.95	**-0.97**	**-0.04**	0.01	-0.02
	-0.90	0.91	**0.10**	**0.91**	-0.89	0.91
87	**-0.85**	**0.84**	**-0.13**	**0.12**	-0.98	-0.02
	-0.94	**0.94**	**-0.93**	**0.94**	0.06	0.94
88	**-0.97**	**0.96**	-0.06	0.05	-1.00	0.00
	-0.98	**0.99**	-0.97	0.99	0.01	0.99
89	**-1.01**	**1.00**	-0.93	0.02	-0.14	0.03
	-1.02	**1.03**	-0.11	1.03	-0.92	1.03
94	**-1.00**	**1.00**	-0.01	0.00	-1.01	0.00
	-0.90	**0.90**	-0.90	0.90	0.10	0.90
99	**-0.68**	**0.70**	-0.35	0.36	-1.01	0.03
	-0.78	**0.76**	-0.77	0.76	0.22	0.76

Table S7. TD-CAMB3LYP/MM Mulliken charges for the $(GGGG/CCCC)_{QM}$ system involved in the FPT mechanism displayed in Figure 4b of main article (run *g*). Results correspond only to some selected snapshots. Nucleobase labels are shown in Figure 3, and the states of interest are highlighted in bold.

Frame (fs)	**S_0**		**S_1**	
	GUA GUA B GUA A GUA	CYT CYT B CYT A CYT	GUA GUA B GUA A GUA	CYT CYT B CYT A CYT
19	-0.04 -0.06 -0.10 -0.10	0.04 0.05 0.11 0.09	**-0.06** **-0.15** **0.74** **-0.10**	**0.02** **-0.77** **0.19** **0.11**
39	-0.02 -0.12 -0.12 -0.07	0.01 0.13 0.13 0.06	**-0.02** **-0.24** **0.73** **-0.06**	**0.01** **-0.69** **0.21** **0.07**
59	-0.05 -0.77 -0.10 -0.12	0.04 0.77 0.11 0.10	**0.07** **-0.14** **-0.13** **-0.12**	**0.07** **0.00** **0.15** **0.11**
79	-0.02 -0.74 -0.78 -0.10	0.02 0.75 0.80 0.08	**0.03** **-0.85** **0.03** **-0.09**	**0.03** **-0.09** **0.91** **0.09**
87	**-0.03** **-0.78** **-0.74** **-0.09**	**0.04** **0.79** **0.75** **0.07**	0.02 -0.09 -0.77 -0.09	0.06 0.01 0.78 0.08
99	**-0.04** **-0.80** **-0.67** **-0.10**	**0.05** **0.80** **0.68** **0.08**	-0.02 -0.06 -0.71 -0.10	0.07 0.02 0.72 0.09
105	**-0.05** **-0.77** **-0.66** **-0.11**	**0.05** **0.78** **0.67** **0.09**	-0.04 -0.17 -0.69 -0.11	0.06 0.15 0.70 0.10
119	**-0.07** **-0.78** **-0.67** **-0.06**	**0.07** **0.78** **0.67** **0.04**	-0.05 -0.03 -0.71 -0.06	0.09 0.00 0.71 0.05

Table S8. TD-CAMB3LYP/MM energy differences ($\Delta E_{S2\text{-}S1}$) between the S_1 and the S_2 vertical excitation energies (E) for the $(GGGG/CCCC)_{QM}$ system involved in the FPT mechanism displayed in Figure 4b of main article (run *g*). All energies in eV.

Frame (fs)	E (S_1)	f (S_1)	E (S_2)	f (S_2)	$\Delta E_{S2\text{-}S1}$
0	3.743	0.0002	4.024	0.0001	0.28
4	3.568	0.0008	3.777	0.0008	0.21
9	2.859	0.0010	3.374	0.0003	0.52
14	2.642	0.0014	3.035	0.0000	0.39
19	2.312	0.0000	2.942	0.0000	0.63
24	1.423	0.0000	1.884	0.0001	0.46
29	1.016	0.0000	1.791	0.0002	0.78
34	1.756	0.0000	2.572	0.0001	0.82
39	1.749	0.0000	2.570	0.0001	0.82
44	1.551	0.0000	2.022	0.0001	0.47
49	2.210	0.0000	2.428	0.0000	0.22
54	1.993	0.0000	2.634	0.0000	0.64
59	2.075	0.0001	2.577	0.0000	0.50
64	1.158	0.0012	1.485	0.0000	0.33
69	0.481	0.0000	0.774	0.0023	0.29
74	1.202	0.0001	1.799	0.0005	0.60
79	1.058	0.0001	1.440	0.0002	0.38
84	1.495	0.0002	1.543	0.0003	0.05
87	0.934	0.0010	1.324	0.0001	0.39
89	0.826	0.0013	1.293	0.0001	0.47
94	1.346	0.0008	2.145	0.0005	0.80
99	0.720	0.0008	1.629	0.0003	0.91
104	0.389	0.0010	1.500	0.0002	1.11
105	0.248	0.0014	1.409	0.0001	1.16

Table S9. CASPT2/MM energy differences ($\Delta E_{S2\text{-}S1}$) between the S_1 and the S_2 vertical excitation energies (E) of the $(GGGG/CCCC)_{QM}$ system involved in the FPT mechanism displayed in Figure 4b of main article (run *g*). All energies in eV.

Frame (fs)	E (S_1)	E (S_2)	$\Delta E_{S2\text{-}S1}$
74	1.625	1.897	0.27
79	0.776	1.120	0.34
84	0.808	0.857	0.05
87	0.082	0.424	0.34
89	0.247	0.410	0.16

Figures

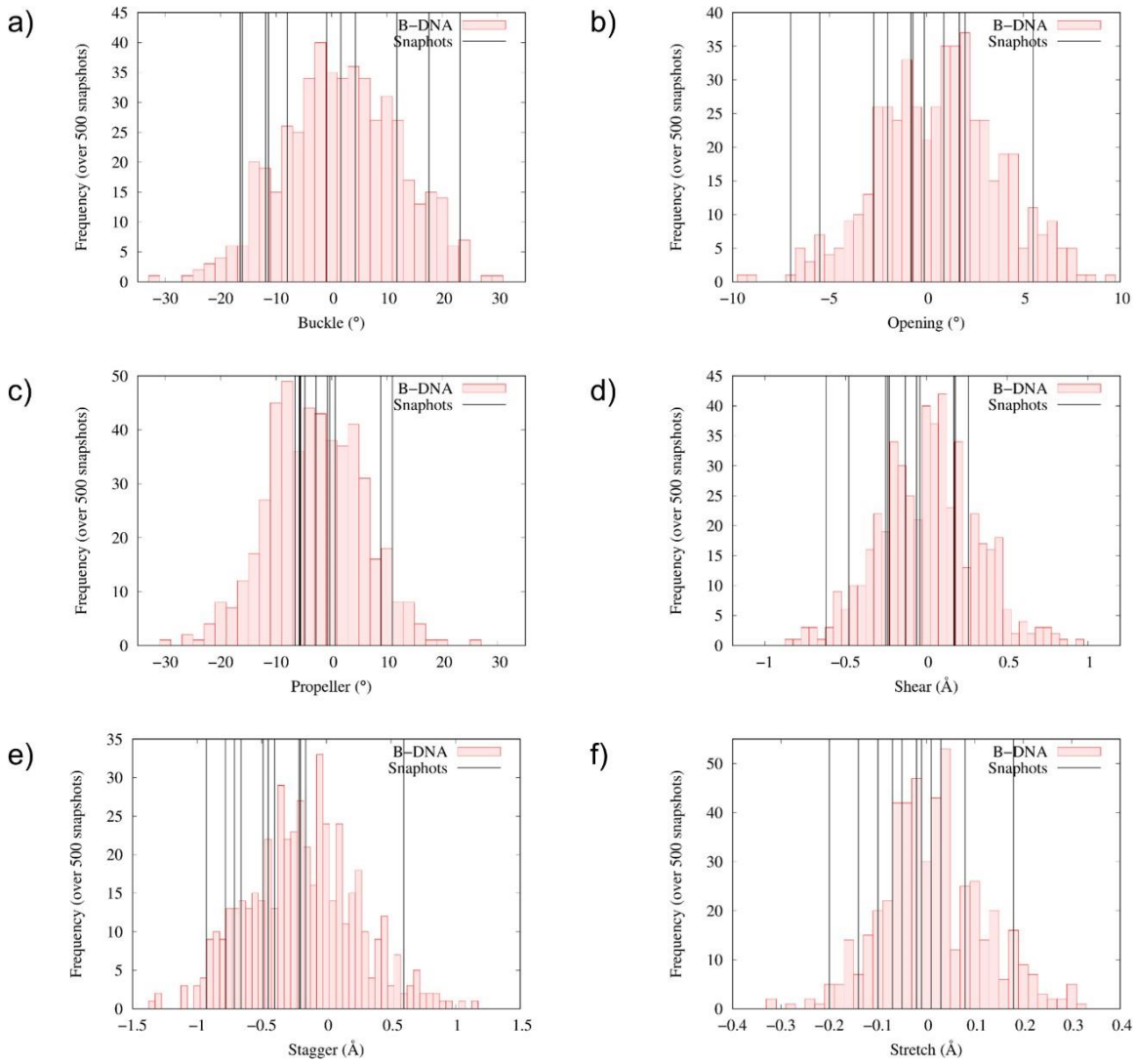

Figure S1. Comparison of characteristic-DNA geometrical parameters between the selected 11 snapshots for the present QM/MM study (black lines) and the B-DNA full conformational space (pink thick bars).

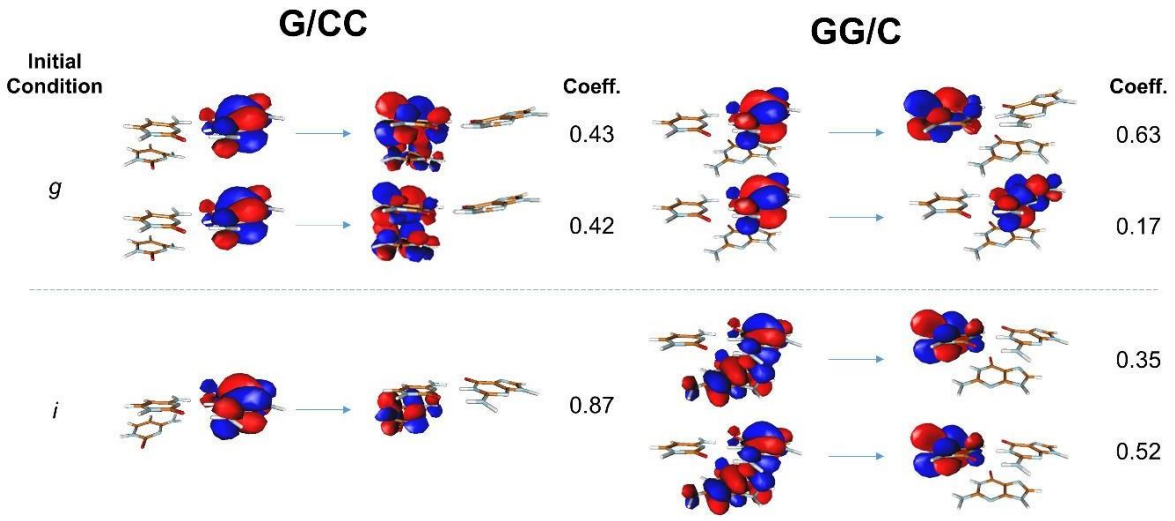

Figure S2. S_1 electronic excitations of the $(G/CC)_{QM}$ and the $(GG/C)_{QM}$ partitioning schemes at the *g* and *i* snapshots.

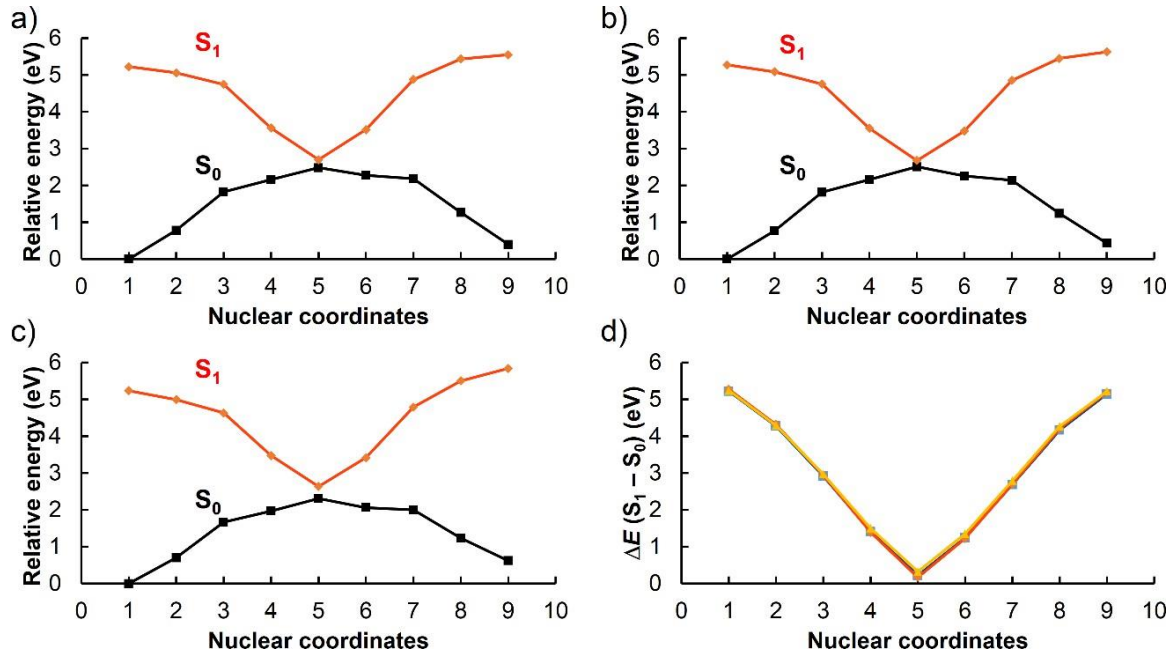

Figure S3. TD-CAMB3LYP description of the G-C ground (S_0) and charge-transfer (S_1) states using the a) 6-311G**, b) 6-31G*, c) 6-31G. Plot d) jointly displays the S_1 determinations with the three previous basis sets. Nuclear coordinates 1, 5, and 9 correspond to the WC, INT, and tautomer geometries as displayed in Figure 6 of Supporting Information's reference [11].

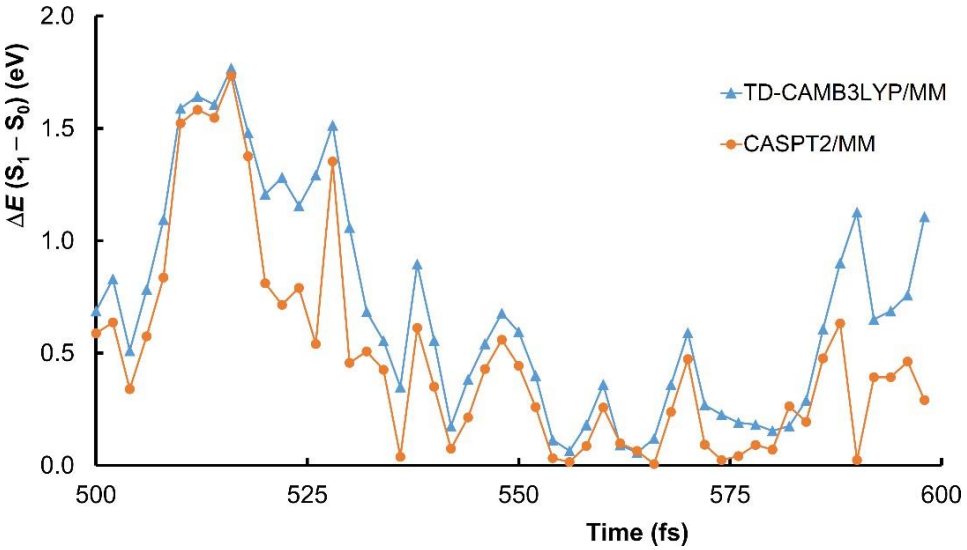

Figure S4. S_1-S_0 energy difference with both TD-CAMB3LYP/MM and CASPT2/MM corresponding to the run *g* of the $(G\text{-}C)_{QM}$ partitioning scheme.

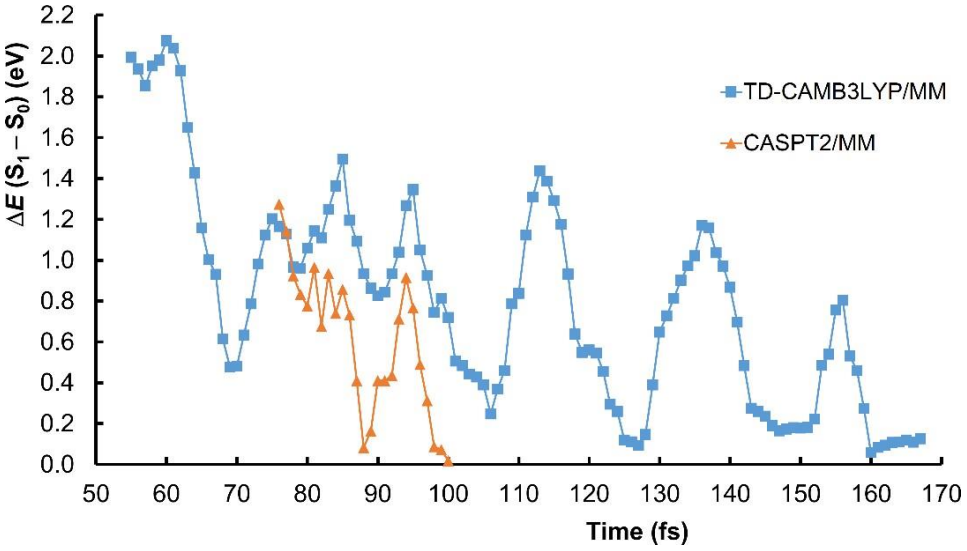

Figure S5. S_1-S_0 energy difference with both TD-CAMB3LYP/MM and CASPT2/MM corresponding to the run *g* of the $(GGGG/CCCC)_{QM}$ partitioning scheme. Note that the CASPT2/MM energies are computed including only the two adjacent G-C pairs involved in the FPT mechanism as described in pages 5-6 of this Supplementary Information.

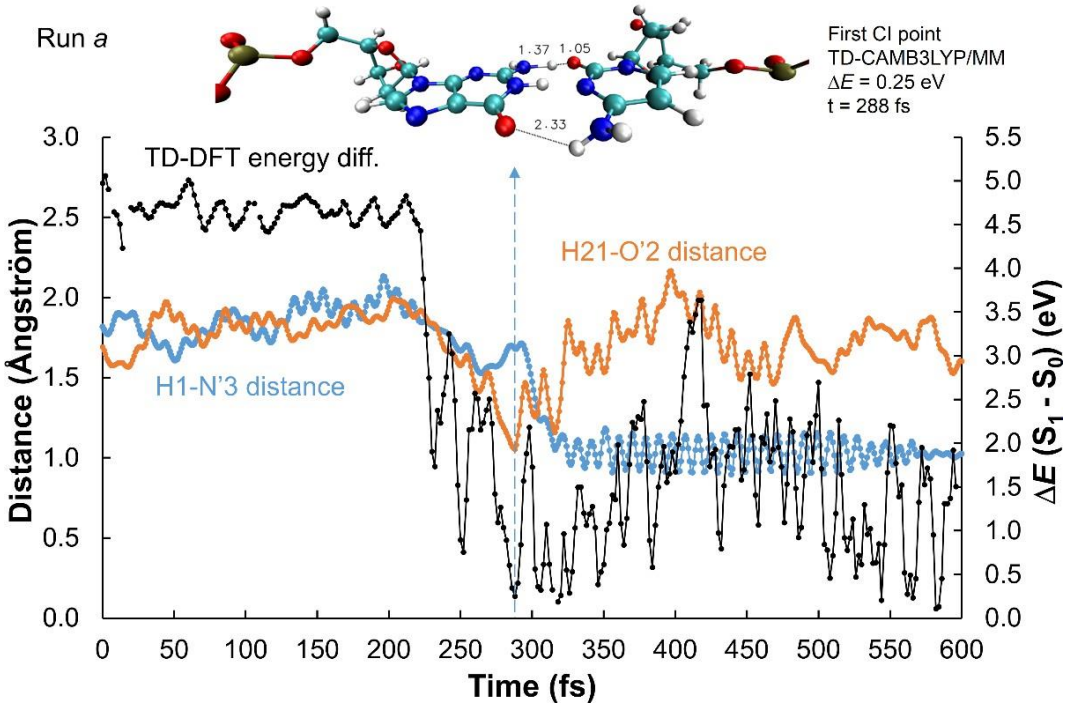

Figure S6. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *a*. The first intersection point (ΔE < 0.5 eV) is shown.

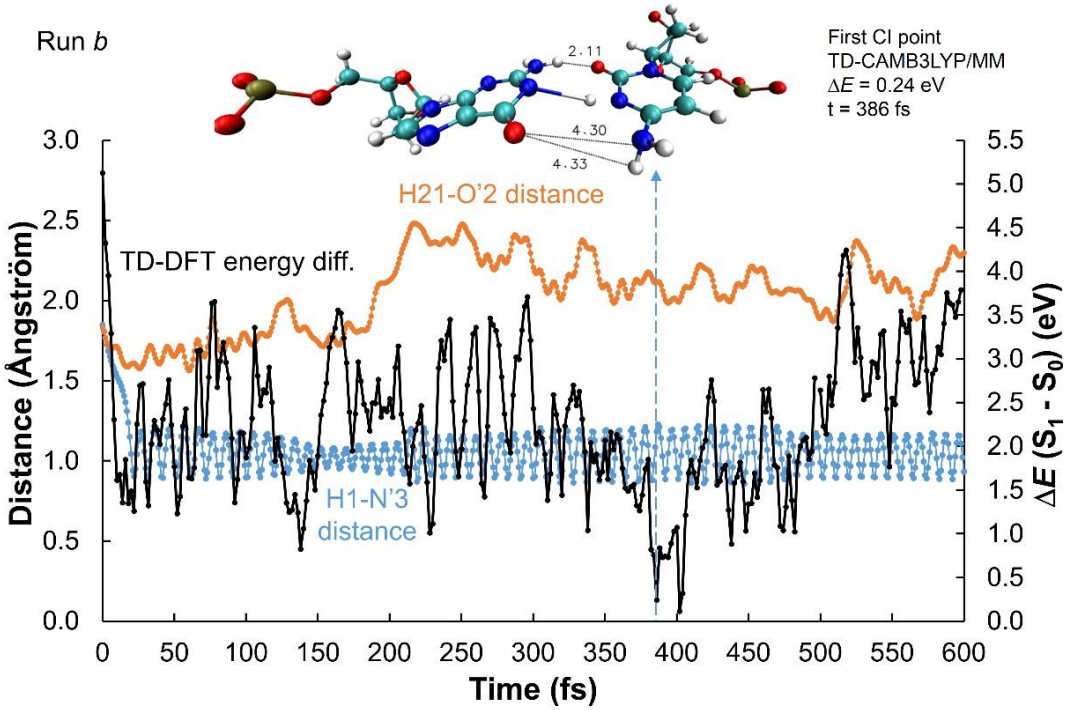

Figure S7. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *b*. The first intersection point (ΔE < 0.5 eV) is shown.

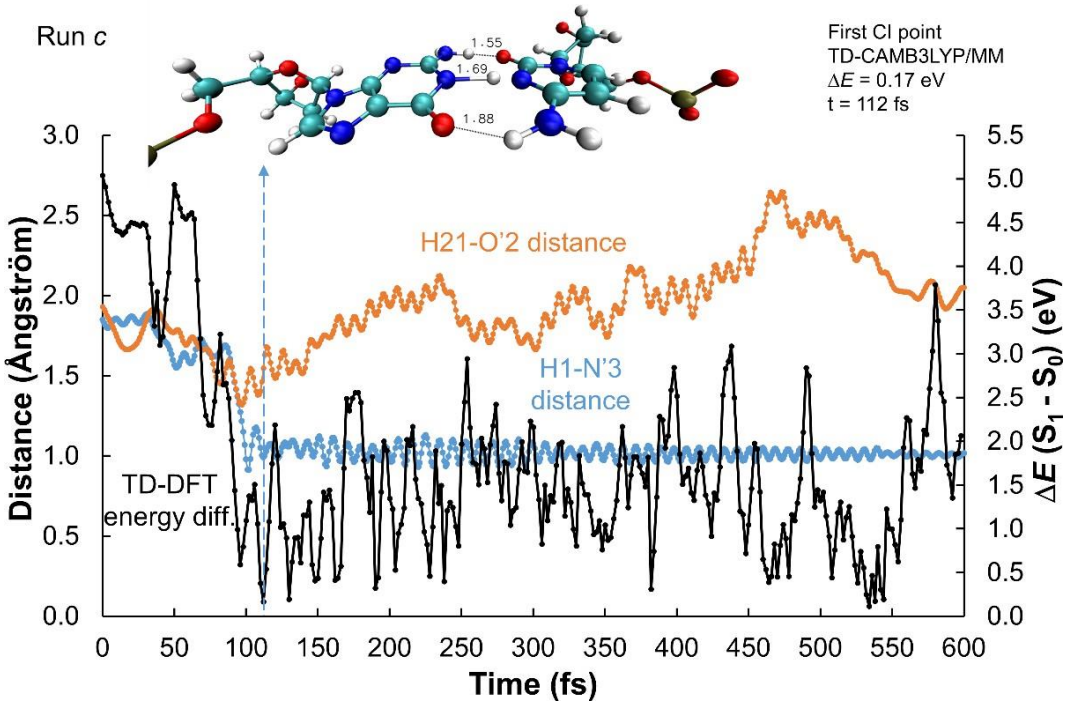

Figure S8. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *c*. The first intersection point (ΔE < 0.5 eV) is shown.

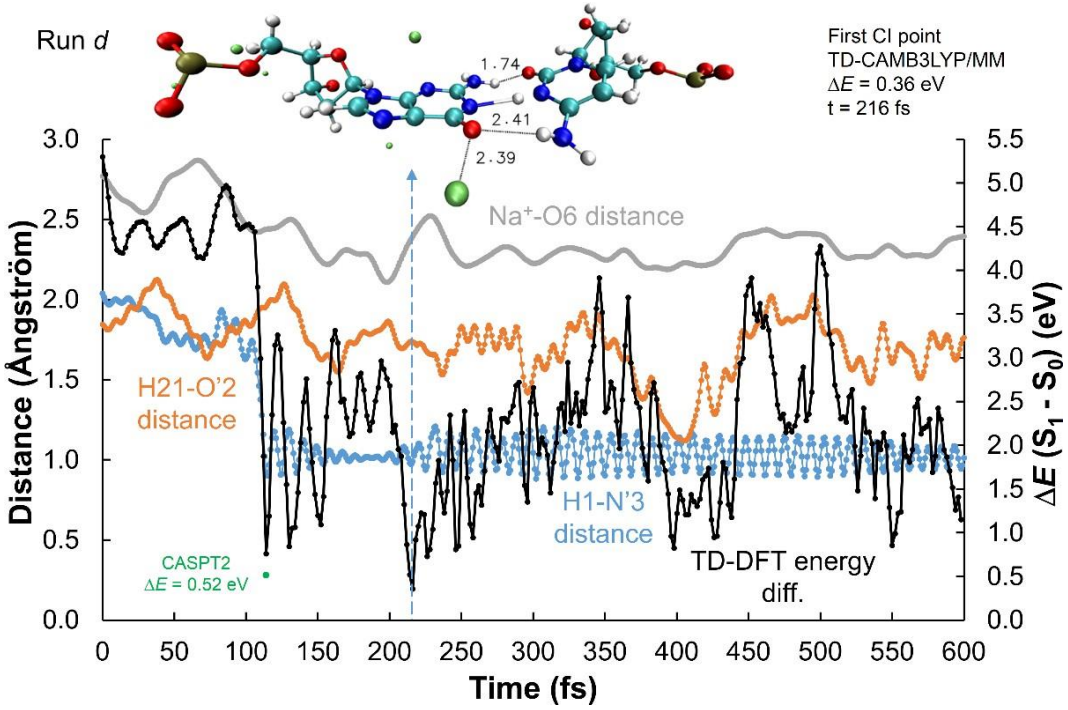

Figure S9. H1-N'3, H21-O'2 and Na^+-O3 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *d*. The first intersection point (ΔE < 0.5 eV) is shown. The Na^+ atom is shown in green.

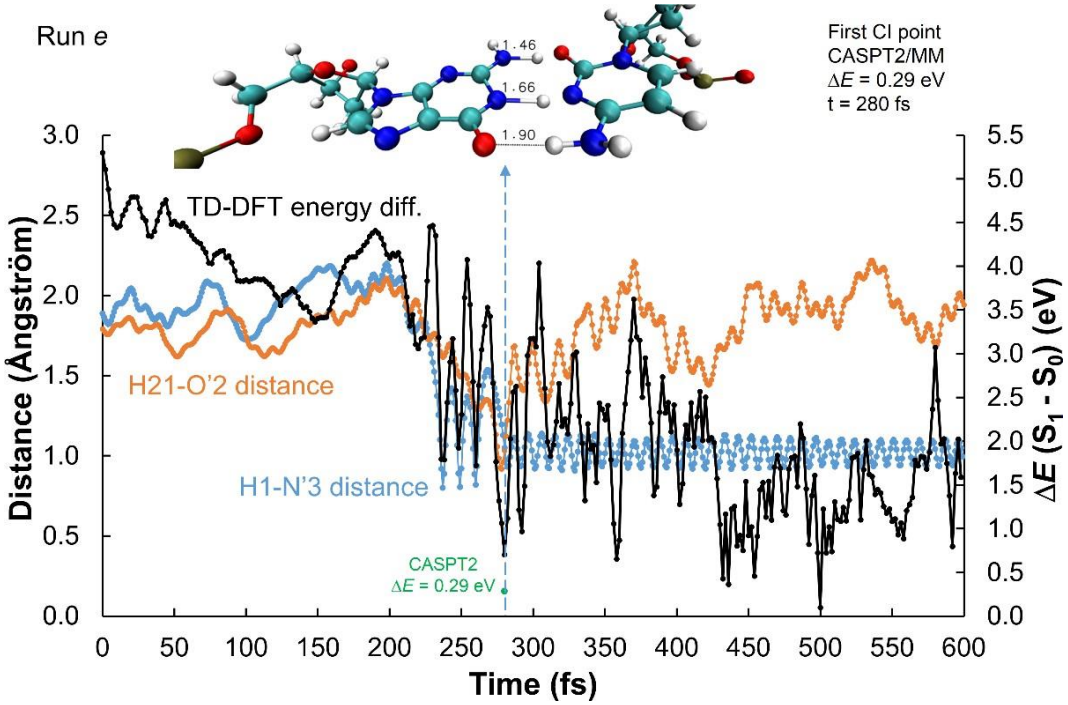

Figure S10. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *e*. The first CI point (ΔE < 0.5 eV) is shown.

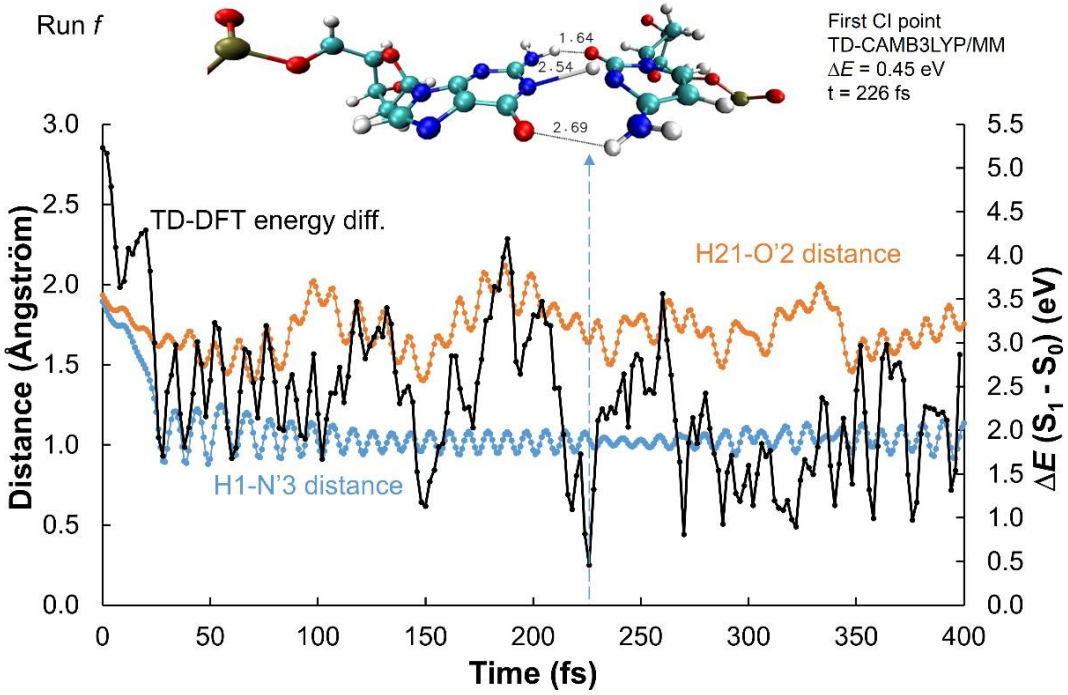

Figure S11. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *f*. The first intersection point (ΔE < 0.5 eV) is shown.

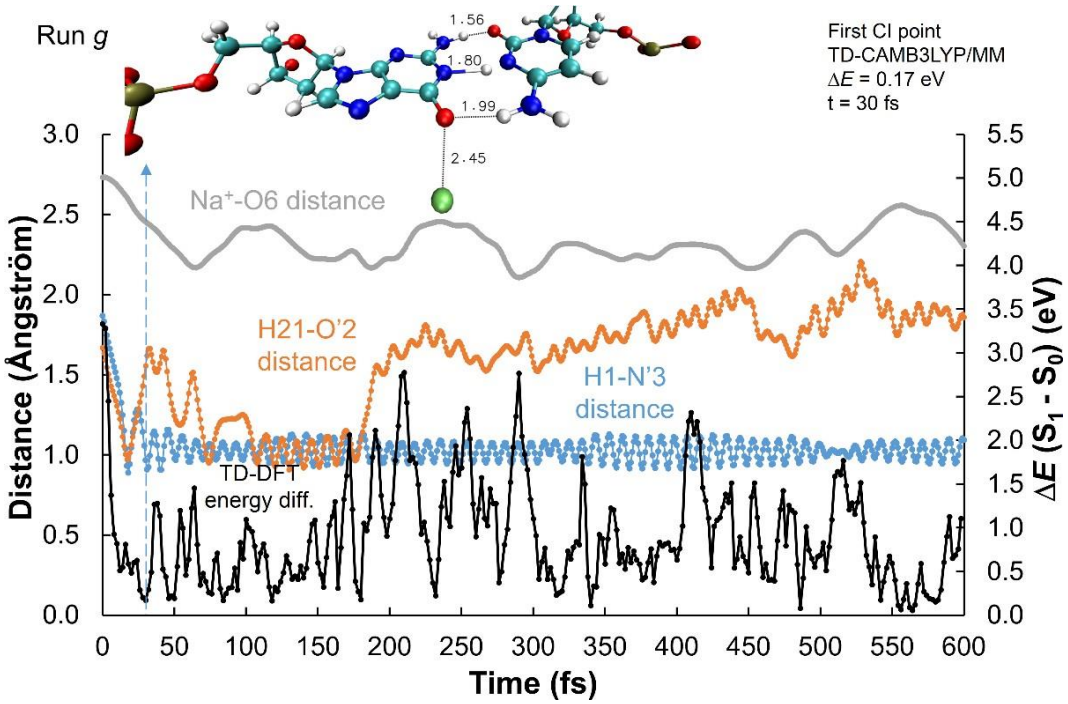

Figure S12. H1-N'3, H21-O'2 and Na^+-O3 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *g*. The first intersection point (ΔE < 0.5 eV) is shown. The Na^+ atom is shown in green.

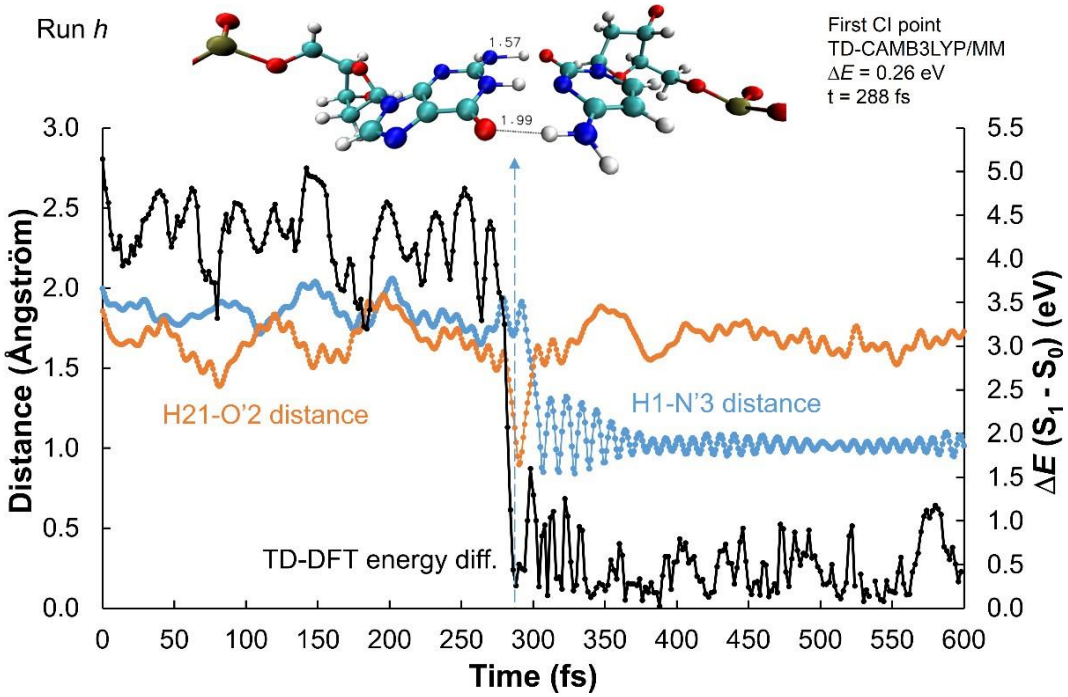

Figure S13. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *h*. The first intersection point (ΔE < 0.5 eV) is shown.

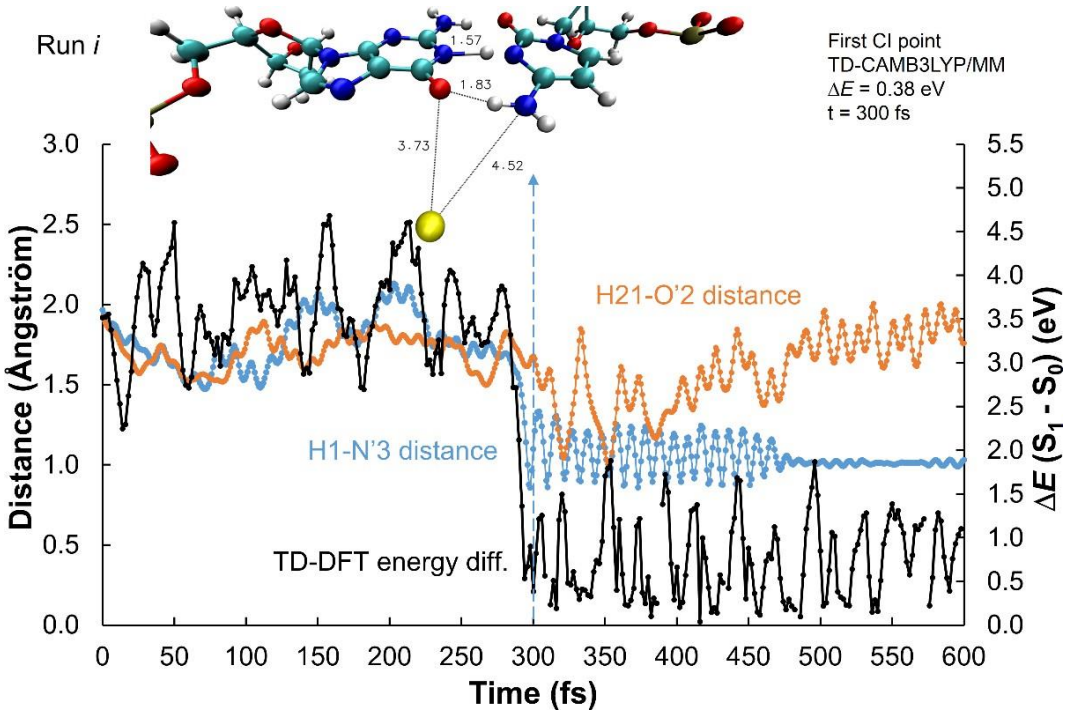

Figure S14. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *i*. The first intersection point ($\Delta E < 0.5$ eV) is shown. The yellow atom corresponds to a Na^+ cation.

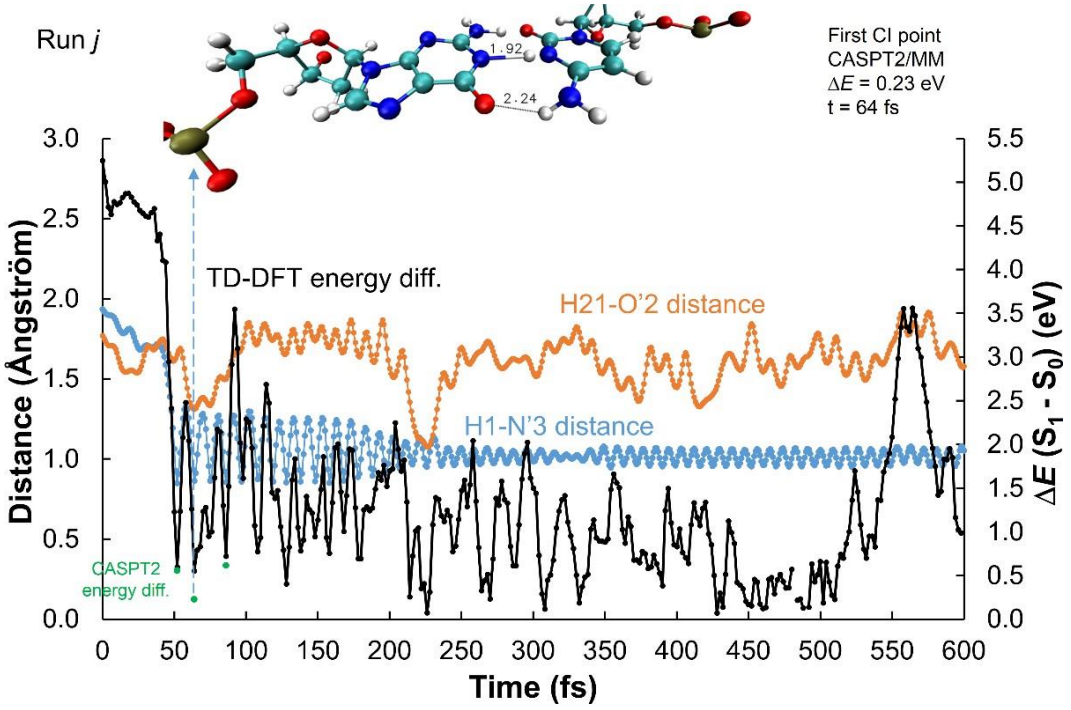

Figure S15. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *j*. The first CI point ($\Delta E < 0.5$ eV) is shown.

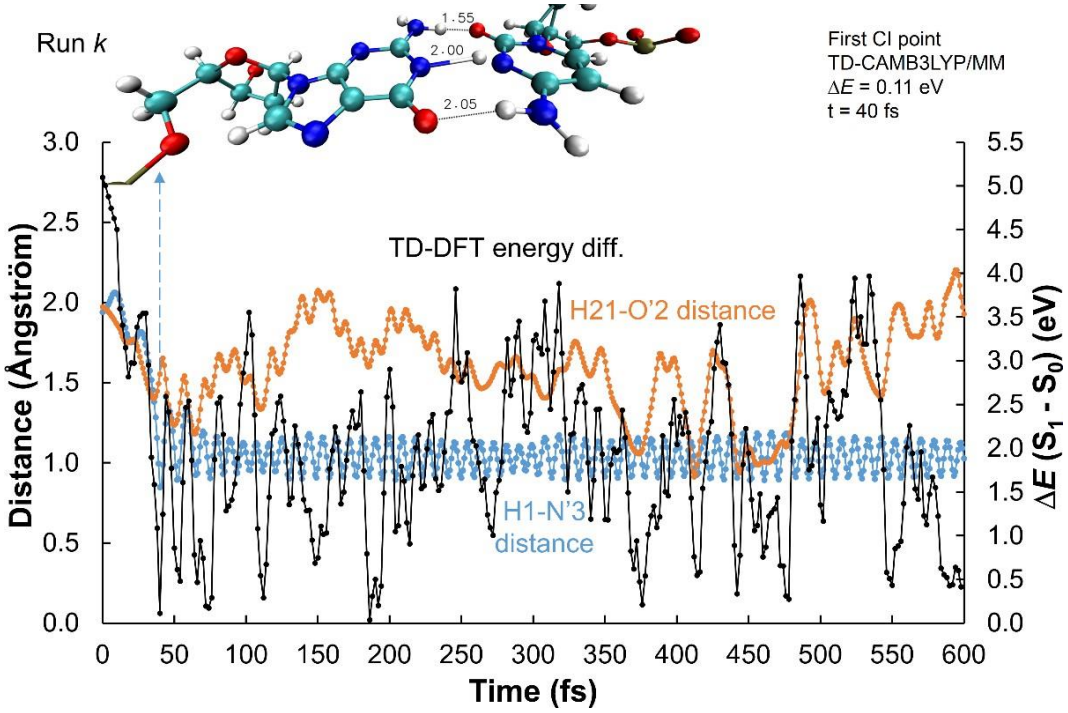

Figure S16. H1-N'3 and H21-O'2 distances and energy differences between the S_1 and the S_0 states of the $(G\text{-}C)_{QM}$ run *k*. The first intersection point (ΔE < 0.5 eV) is shown.

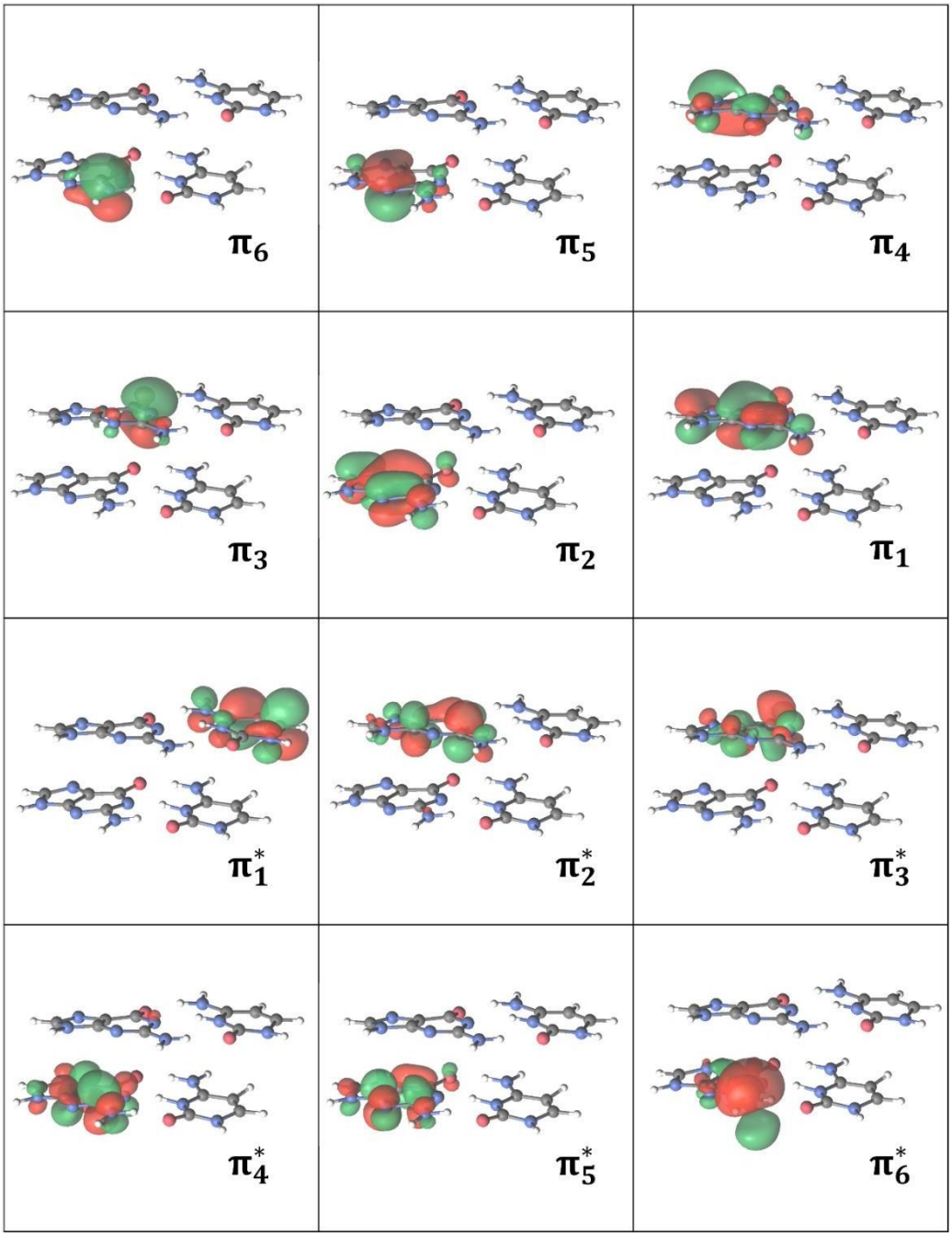

Figure S17. SA-CASSCF natural orbitals included in the active space for the CI point (t = 87 fs) that mediates the deactivation of the $(GGGG/CCCC)_{QM}$ run shown in Figure 4b.

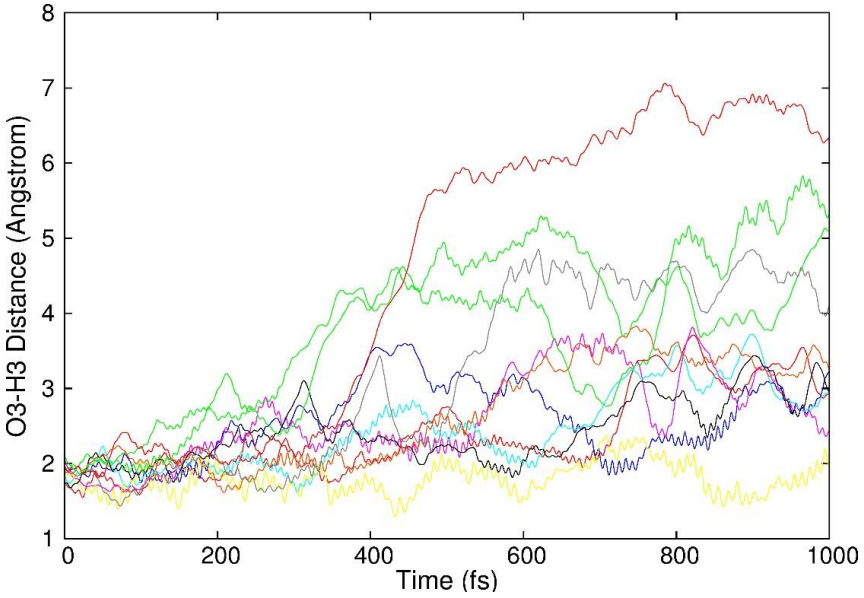

Figure S18. O6-H'41 distances of all $(G\text{-}C)_{QM}$ trajectories on the S_1 surface.

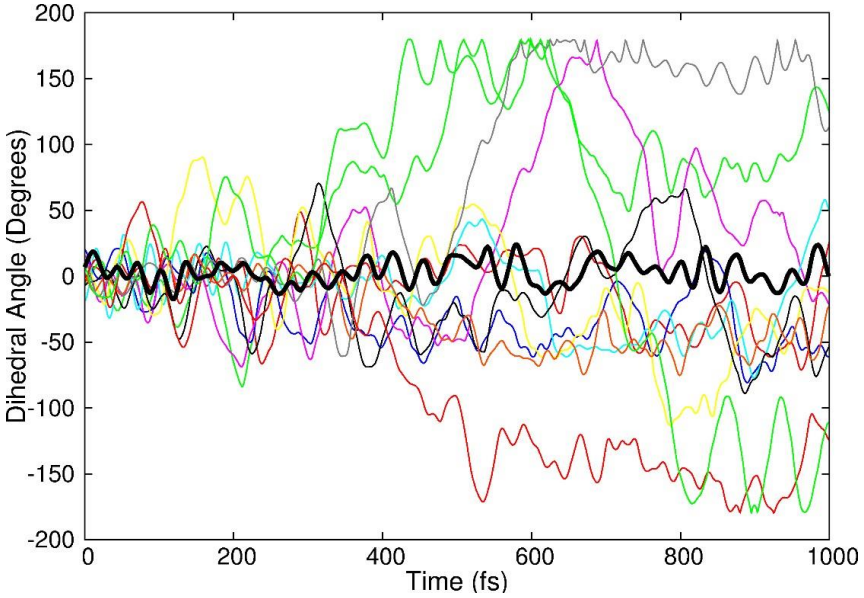

Figure S19. H'41-N'4-C'4-N'3 dihedral angles of all $(G\text{-}C)_{QM}$ trajectories on the S_1 state surface. The black thick curve placed at the center corresponds to a run in the ground state.

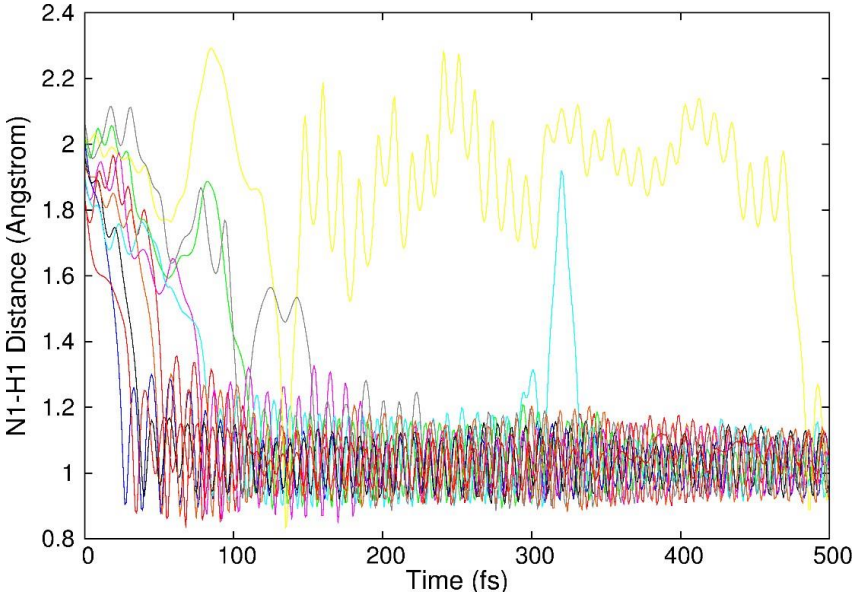

Figure S20. H1-N'3 distances corresponding to the transferred H atom of all $(GG/CC)_{QM}$ trajectories on the S_1 surface.

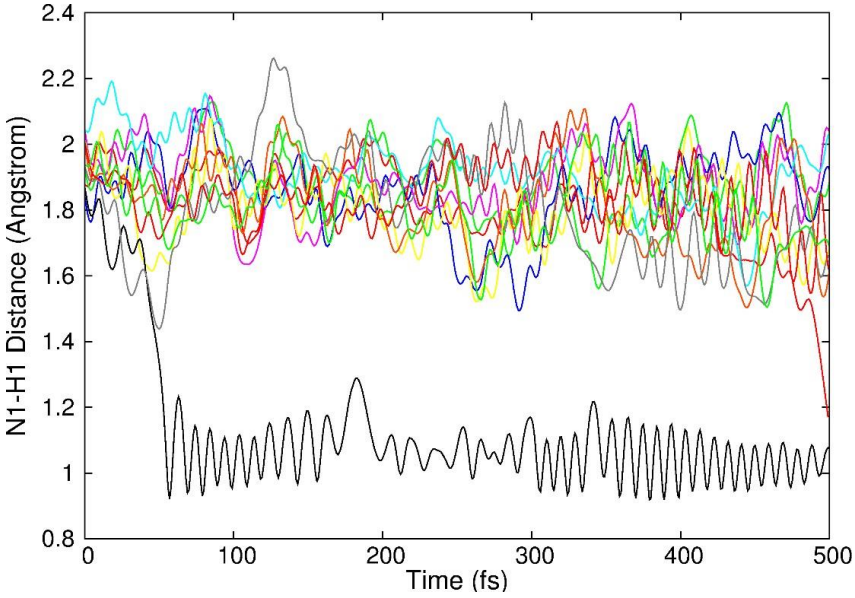

Figure S21. H1-N'3 distances corresponding to the NOT transferred H atom of all $(GG/CC)_{QM}$ trajectories on the S_1 surface (except *g* run).

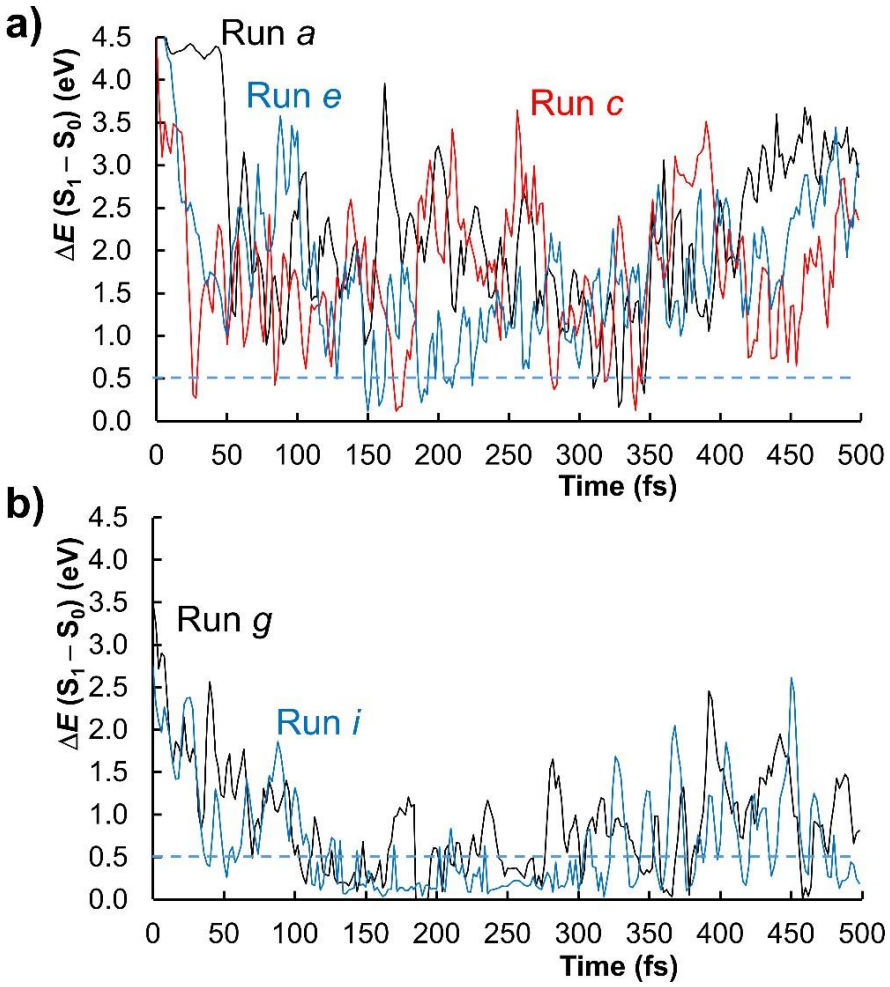

Figure S22. TD-CAMB3LYP/MM energy difference between the S_1 and the S_0 states for the $(GG/CC)_{QM}$ system. a) runs *a*, *c*, and *e* and b) runs *g* and *i*. Blue dashed lines at 0.5 eV represent an energy threshold to estimate intersection points.

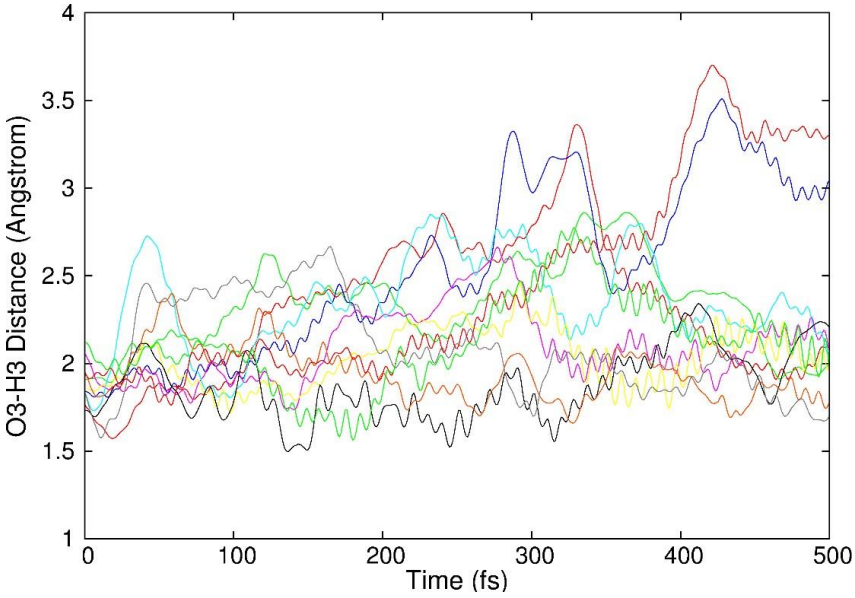

Figure S23. O6-H'41 distances of all $(GG/CC)_{QM}$ trajectories on the S_1 surface.

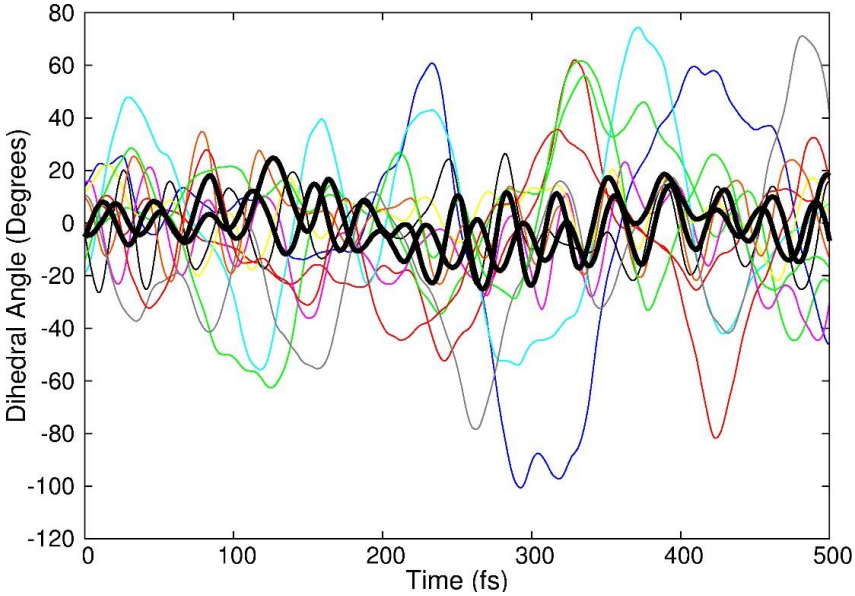

Figure S24. H'41-N'4-C'4-N'3 dihedral angles of all $(GG/CC)_{QM}$ trajectories on the S_1 state surface. The black thick curves placed at the center corresponds to a run in the ground state.

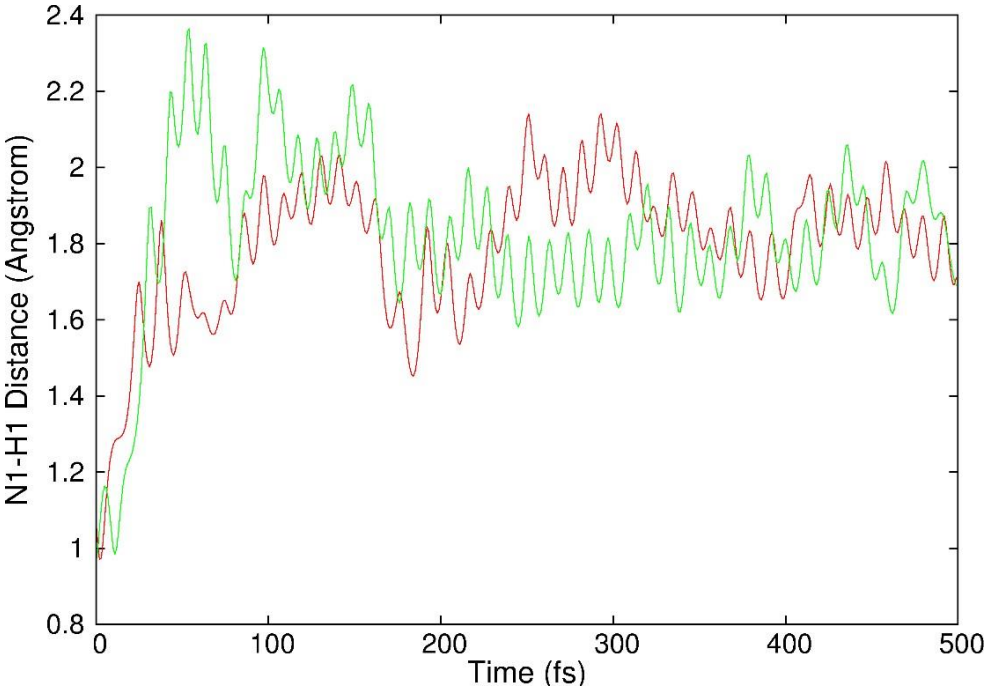

Figure S25. H1-N'3 distances corresponding the two G-C base pairs involved in the FPT mechanism of the $(GGGG/CCCC)_{QM}$ trajectory run on the S_0 surface. The initial conditions (t = 0 fs) of this run corresponds to the CI point at the CASPT2/MM and CASPT2 levels displayed in Figure 4b (t = 87 fs in the excited state run).

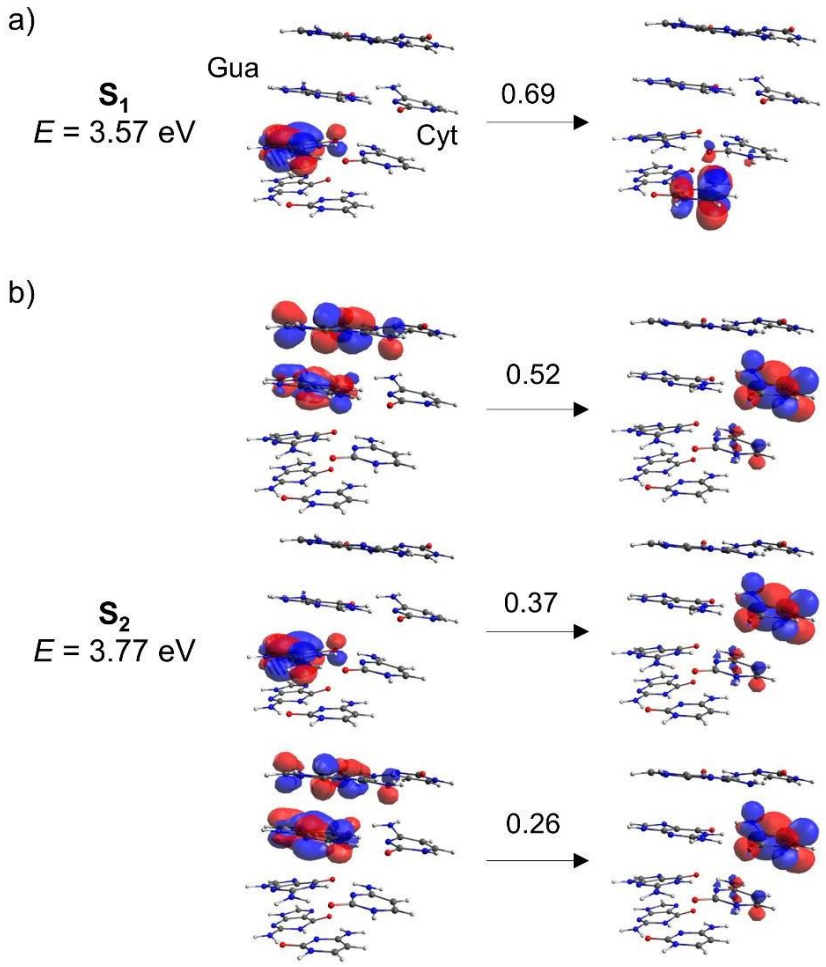

Figure S26. Nature of the excited wave functions corresponding to the snapshot t = 4 fs of the run *g* using the $(GGGG/CCCC)_{QM}$ partition (Figure 4b of the main article), using the TD-CAMB3LYP/MM method. S_1 (a) and S_2 (b) excitations are shown, the coefficients of the monoelectronic transitions are displayed above the arrows.

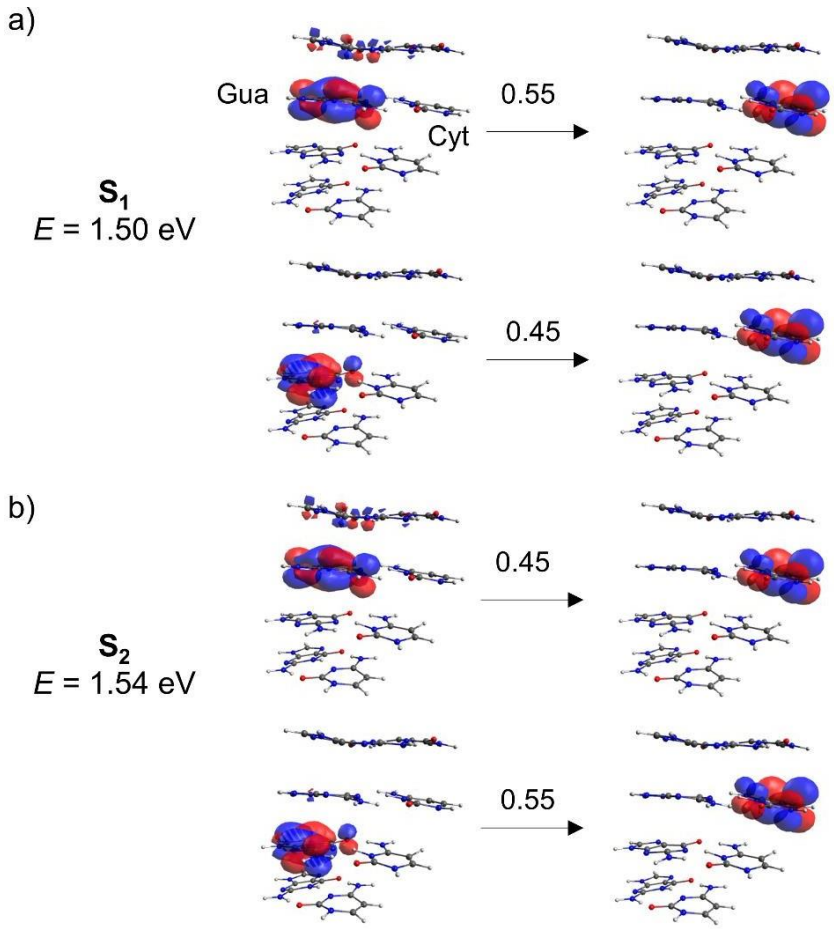

Figure S27. Nature of the excited wave functions corresponding to the snapshot t = 84 fs of the run *g* using the $(GGGG/CCCC)_{QM}$ partition (Figure 4b of the main article), using the TD-CAMB3LYP/MM method. S_1 (a) and S_2 (b) excitations are shown, the coefficients of the monoelectronic transitions are displayed above the arrows.