Photoinduced Formation Mechanism of the Thymine-Thymine (6-4) Adduct in DNA; a QM(CASPT2//CASSCF):MM(AMBER) study

Angelo Giussani,[a,*‡] Irene Conti,[a,b] Artur Nenov,[a,b] Marco Garavelli[a,b,*]
[a] Dipartimento di Chimica "G. Ciamician", Università di Bologna, Via F. Selmi 2, 40126 Bologna, Italy
[b] Dipartimento di Chimica Industriale "Toso Montanari", Università di Bologna, Viale del Risorgimento , 40136 Bologna, Italy
[‡] Current address: Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK.

Abstract

The UVB-induced photomechanism leading the carbonyl group of a thymine nucleobase to react with the carbon-carbon double bond of consecutive thymine nucleobase in a DNA strand to form the thymine-thymine (6-4) adduct photodamage remains nowadays poorly understood. Key questions remain open, concerning both the intrinsic features of the photoreaction (as the contribution (or not) of triplet states, the nature of the involved states, the time-scale of the photoprocess), and the role play by the non-reactive surrounding of the two reactive pyrimidine nucleobases (as the nature of the flanked nucleobases and the flexibility of the whole DNA molecule). A reduced number of theoretical studies have been conducted on the title photoreaction, most of which have been performed on reduced model systems of DNA, consequently neglecting potential key parameters for the photoreaction as the constrains due to the double strain structure and the presence of pairs and stacked nucleobases. In the present contribution the photoactivated step of the title reaction has been studied in a DNA system, and in particular for a specific DNA hairpin for which the quantum yield of formation of the photodamage has been recently experimentally measured. The reaction has been characterized performing high-level QM/MM computations, combining the CASPT2//CASSCF approach for the study of the reactive part (i.e. the two thymine molecules) with an MM-Amber treatment of the surrounding environment. The possibility of a reaction path along both the singlet and triplet manifolds have been characterized, the nature of the reactive states have been analyzed, and the role played by the flexibility of the whole system, which in turn determines the initial accessible geometrical conformations, has been evaluated, consequently providing a substantial contribution towards the elucidation of the photoreaction mechanism. On the basis of the obtained results, a charge-transfer state can decay from a pro-reactive initial structure toward a region of energy degeneracy with the ground state, from which the subsequent decay along the ground state hypersurface can lead to the photoreaction.

INTRODUCTION

Two major classes of UV-induced damages can be form between a pair of consecutive pyrimidine nucleobases in a DNA strand: the cyclobutane pyrimidine dimers, and the pyrimidine-pyrimidone (6-4) adducts.[1,2] The latter can further evolve into their Dewar isomers through a subsequent photoreaction.[3] While a large amount of theoretical[4-12] and experimental[13-20] works have been published on the process leading to cyclobutane pyrimidine dimers, only few researches have been devoted to the elucidation of the mechanism operating in the pyrimidine-pyrimidone (6-4) adducts production.[5,11,18-24] The latter

reaction is believed to take place through the initial formation of an oxetane or an azedine ring, which constitutes the actual photoproduct, followed by thermal ring opening into the pyrimidine−pyrimidone (6-4) adduct.[1] Experimentally it is known that the reaction is inefficient (the quantum yields is of the order of 10^{-4}),[25-27] it occurs under UVB (280-315 nm) radiation but not under UVA (315-400 nm) exposure,[28,29] and the actual efficiency of the process depends on the nature of the involved pyrimidine nucleobases and of the flanked moieties.[26,30] The timescale for the formation of the thymine-thymine (6-4) adducts has been determined through transient absorption experiments performed by Marguet and Markovitsi to be in the ms regime.[20] It must be notice that such a timescale refers to the production of the damage (i.e. the pyrimidine-pyrimidone (6-4) adduct) and not to the photoactivated part of the process (i.e. the oxetane formation) whose timescale remains to be determined. Regarding the mechanism of formation, most of the efforts aimed at its elucidation have been performed studying the reaction of two thymine molecules leading to the thymine-thymine (6-4) adduct.[5,11,19,21-23] In such a case, the photoactivated part of the process leads to an oxetane ring between the carbonyl group of the 3-ended nucleobase and the carbon-carbon double bond of the second moiety. The photochemical addition of carbonyl compounds to alkenes to form an oxetane ring is known as the Paternò−Büchi reaction,[31] whose mechanism, although is normally accepted to proceed via the triplet $^3(n\pi^*)$ excited state of the carbonyl compound,[32] still constitutes a matter of debate in the scientific community.[33-35]

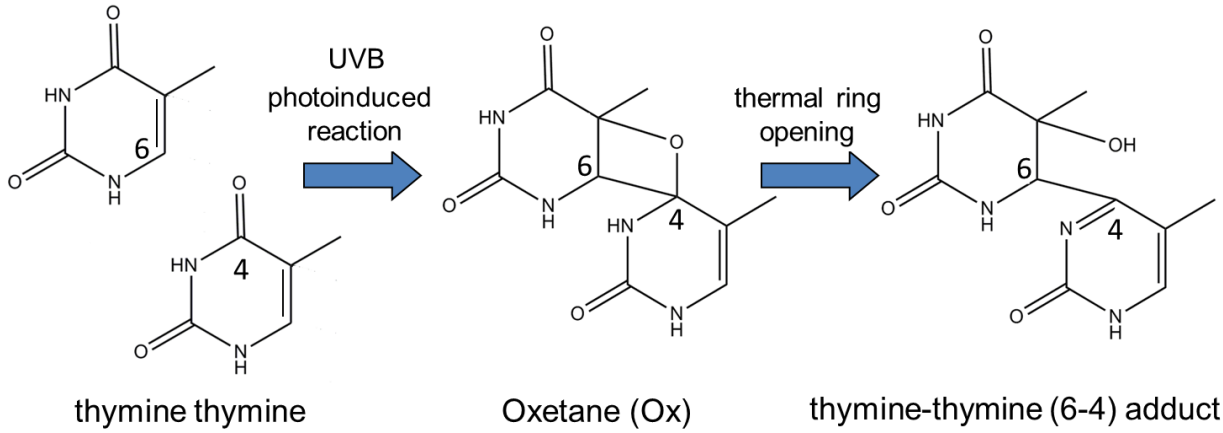

Figure 1. Scheme of the reaction leading to the thymine-thymine (6-4) adduct.

Two main explanations have been given in order to rationalize the photoinduced mechanism of formation of the thymine-thymine (6-4) adduct. On the basis of TD-M052X computations on dithymidine monophosphate in water and steady-state and time-resolved optical spectroscopy experiments on single strand $(dT)_{20}$, Markovitsi, Improta and co-workers concluded that the reaction proceeds on the potential energy surface of a charge-transfer state and it requires the surmounting of an energy barrier, which can explain the increase of the photoreaction's quantum yield with the increase of the incident radiation energy.[11,19] A related mechanism involving the population of a charge-transfer state has been also recently documented performing QM/MM computations on a double strand B-fragment.[24] Performing ab-initio CASPT2//CASSCF quantum chemical calculations on a model system compose by two thymine molecules in vacuum, Giussani et al have characterized a mechanism involving the initial

decay into the nπ* state that, leading to an elongation of the reactive carbonyl group, can subsequently reacts on the triplet manifold with the carbon-carbon double bond of the second nucleobase.[23]

In both works, a key geometry, previously characterized by Blancafort et al,[5] in which the new carbon-oxygen bond present in the final oxetane ring is almost form, has been indicated as the gate between the excited state of the reactant and the ground state of the photoproduct, acting as conical intersection and/or intersystem crossing region with the ground state.

While both mechanisms provide a plausible explanation of the photoprocess, the mechanism of the photoreaction in DNA systems, where constrains due to the double strain structure and the presence of pairs and stacked nucleobases can significantly alter the process, remains to be better characterized. In the present contribution the title reaction has been studied in a DNA system, and in particular for a specific DNA hairpin for which the quantum yield of formation of both classes of photodamages have been experimentally measured.[27] The mechanism of the reaction has been characterized performing high-level QM/MM computations, combining the well-known CASPT2//CASSCF approach for the study of the reactive part (i.e. the two involved thymine molecules) with an MM-Amber treatment of the surrounding environment.

On the basis of the obtained results, from a pro-reactive ground state geometry characterized by a reduced distance between the $C_{4'}O$ atoms, an high-lying charge-transfer state can evolve toward a region of energy degeneracy with the ground state, from which the system can evolve on the ground state potential energy hypersurface (PEH) to either the separated thymine molecules (non-reactive path), or toward the formation of an oxetane ring (reactive path), depending on the geometrical parameters of the crossing structure or the direction and velocity through which the ground state is repopulated. On the other hand, the possibility of a reaction along the triplet manifold is not here confirmed.

COMPUTATIONAL DETAILS

The system here under study is the $H_4$2 hairpin depicted in Figure 2, forms by adenine and thymine nucleobases. Each of the two strands composing the system includes four nucleobases. As shown in Figure 2, the two strands are covalently linked through a dodecane (C12) linker, the presence of which prevents the breakdown of the double helix structure. The two reactive thymine molecules will be labeled as shown in Figure 2c: the 5-ended thymine molecule, which contains the reactive carbon-carbon double bond, is label as T, while the 3-ended thymine nucleobase, where the reactive carbonyl group is located, is label as T'. The atoms of each moiety are enumerated as normally done in thymine, adding a ' mark for the atoms of the T' nucleobase.

The $H_4$2 system has been chosen in the present study for the following main reasons: experimental data for the quantum yield of formation of both the thymine-thymine (6-4) adduct and the cyclobutane thymine dimer have been provided; the $H_4$2 hairpin is characterized by an unusual high quantum yield of formation for the thymine-thymine (6-4) adduct (of the order of 10^{-3}, while most of the available data reporte values of 10^{-4});[25-27] despite the shortness of oligonucleotide, the reactive dimer is inside a DNA

double strand (i.e. both thymine molecules are paired with two adenine nucleobases of the second strand, and each thymine is flanked with an adenine molecule).[27]

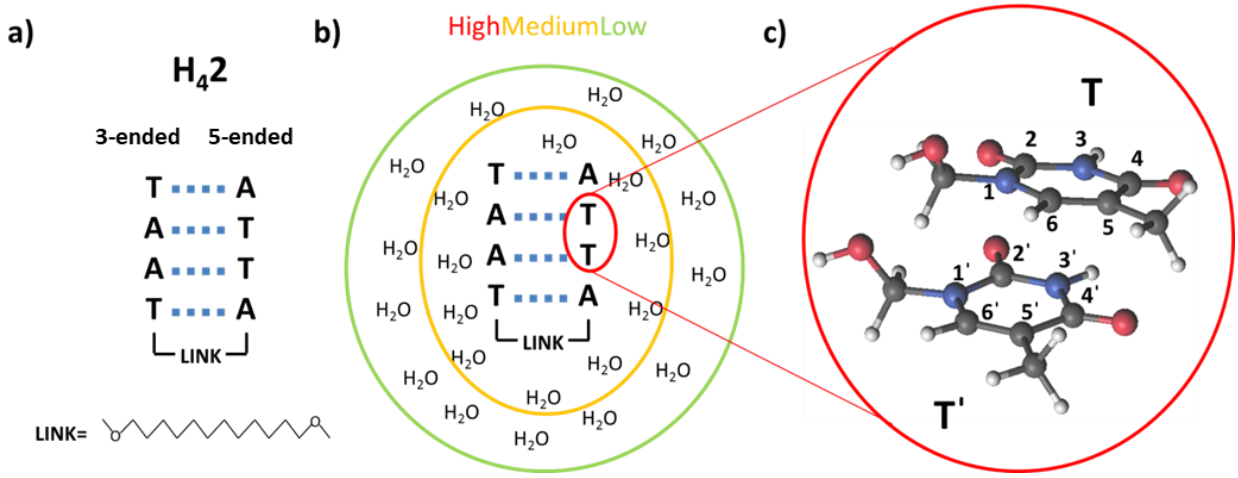

Figure 2. a) Structure of the $H_4$2 hairpin; b) High-Medium-Low subdivision adopted in the QM/MM computations; c) Atoms and nucleobases labeling

A classical 500 ns long molecular dynamics (using NTP condition at 383 K, 1 bar, and setting 2 fs integration timestep) has been performed, using the AMBER12 program and the ff12SB+gaff force fields.[36,37] Explicit solvation by TIP3P water molecules (4664 waters) with eighteen Na^+ ions for neutralization has been considered. The simulation has been performed with periodic boundary conditions and Ewald sum cutoff of 10 Å.

The QM/MM computations have been performed using the COBRAM interface developed in our group, whose details have been presented elsewhere.[38] The system (i.e. the $H_4$2 immersed in a box of water molecules) has been divided in three regions (see Figure 2b): the high region, composed by the two reactive thymine molecules, including both the sugar ring carbon directly linked to the nucleobase and the adjacent oxygen atom; the medium region, form by the remaining parts of the $H_4$2 system and by the water molecules whose H-atom is within a distance of 2.5 Å from a heteroatom of the high region or whose O-atom is within a distance of 2.5 Å from a H-atom of the high layer, covalently bound to a heteroatom; and the low region, composed by the water molecules not included in the high and medium parts. The high region is described at the QM level by means of CASPT2//CASSCF calculations,[39-43] while both the medium and low regions are treated at the MM level through Amber 12.[36,37] At the QM MM interface, a hydrogen link-atom approach has been adopted.

Ground and excited states geometry optimizations have been performed using HF and CASSCF,[39] (with Gaussian 03 and Molcas8, respectively)[44,45] and Amber 12 level of theory for respectively the high and medium regions, employing the optimization algorithm implement in Gaussian 03.[45] During optimizations both the high and medium regions were free to move, while the low part was kept fixed

into its initial conformation. Different CASSCF active space have been employed accordingly to the specific electronic state whose PEH was explored, a strategy normally employed in CASPT2//CASSCF photochemical studies,[46] while all the obtained critical points (i.e. minima, conical intersection, etc.) have been characterized by means of CASPT2 computations including in the active space 12 active electrons distributed in 10 active orbitals (two lone pairs, four π orbitals and 4 π^* orbitals, see Figure S1). In CASSCF and CASPT2 calculations the basis sets of the atomic natural orbital (ANO) of L-type contracted to C,N,O[4s,3p,1d]/H[2s1p] have been employed,[47,48] while the 6-31G* basis set has been used for ground state HF optimization. Within the CASPT2 scheme, an imaginary level-shift of 0.2 au has been used in order to avoid the presence of possible intruder states. The CASPT2 standard zeroth-order Hamiltonian has been employed as originally implemented.[41] The core orbitals have been frozen in the CASPT2 calculations. The Cholesky decomposition has been adopted to speed up the evaluation of two-electron integrals.[49] CASPT2 computations were performed using the Molcas 8 software.[44] The method employed for the calculation of the spin−orbit coupling (SOC) is described elsewhere.[23]

RESULTS

In order to sample the ground state accessible conformations of the system, a 500 ns MM dynamics of the $H_4$2 hairpin in water at 383 K has been performed. Since the main goal of the present contribution is to elucidate the photomechanism of the title reaction, we have conducted our study starting from the MM snapshot whose structure displays the shortest internucleobases distance between the $O_{4'}$ and the C_5 atoms, assuming that such a geometrical feature (i.e. the proximity of the two mentioned atoms) will strongly favor the formation of the thymine-thymine (6-4) adduct. The main reason behind such an assumption is that in different previous theoretical works a conical intersection and/or intersystem crossing region characterized by the formation of a new bond between the $O_{4'}$ and the C_5 atoms has been indicated as the key geometry connecting the excited states of two thymine molecules with the oxetane photoproduct.[5,23] It has been proposed a model in which the photoreactivity towards thymine-thymine (6-4) adduct formation is inferred by the $O_{4'}$-C_5 distance at the initial geometry under irradiation.[27]

At the selected snapshot (hereafter, snapCO) the $O_{4'}$-C_5 distance is equal to 2.810 Å (see Figure S7), and the double-strand structure of the $H_4$2 hairpin is essentially preserved, but the interstrand adenine-thymine pair next to the $T^{'}$ thymine, is no longer forming the usual interstrand hydrogen bonds (see Figure S6). This structural feature can compromise the extrapolation of the following results for longer double-strand DNA structures, although it has been proposed that thymine-thymine (6-4) adduct formation indeed passes through a disruption of the double-strand structure and that the high quantum yield of formation of thymine-thymine (6-4) adduct experimentally measured for the $H_4$2 hairpin can be due to its flexibility and consequent ability to adopt double-strand-broken conformations.[27]

The geometrical parameters describing the reactive thymine pair in the snapCO structure are the result of an MM description. In order to have instead QM parameters, a refinement of the initial structure has been operated by means of a ground state HF QM/MM optimization. In the latter optimization the $O_{4'}$-C_5 distance has been kept frozen, so to retain the main geometrical feature through which the selection of the snapshot was operated. The resulting optimized geometry (hereafter $^1(gs)_{min}$) displays bond

lengths comparable with the one obtained out of a CASSCF optimization of the isolated nucleobase thymine (see Figure S7).[23]

From the $^1(gs)_{min}$ structure the accessible excited states of the system have been characterized performing CASPT2(12,10):AMBER state-averaged computations over ten singlet and ten triplet states. As discussed in the Computational Details section, the (12,10) active space comprises two π, one n, and two π^* orbitals on each thymine (see Figure S1). The resulting energies, oscillator strengths, and dipole moments are present in Table 1, where the states are reported in order of increasing energy and the nature of the different states is specified when the analysis of the corresponding CASSCF(12,10) wave-functions allows a clear identification of it. As expected, among the lowest excited states are recognizable two $\pi\pi^*$ states localized on the C_6C_5 double bond of each thymine (states $^1\pi\pi^*$-T$'$ and $^1\pi\pi^*$-T, see Figure S2) and two $n\pi^*$ states localized on the C_4O carbonyl of each nucleobase (states $^1n\pi^*$-T$'$ and $^1n\pi^*$-T, see Figure S3). The corresponding triplet states (i.e. states $^3\pi\pi^*$-T$'$, $^3\pi\pi^*$-T, $^3n\pi^*$-T$'$ and $^3n\pi^*$-T) are also present. Additionally, $\pi\pi^*$ states describing an excitation form the occupied π orbital mainly localized on the C_4O carbonyl (orbitals π_O-T$'$ and π_O-T, see Figure S1), a surprisingly low singlet double excited state (state S5 in Table 1, hereafter state $2\pi2\pi^*$, involving the π-T$'$, π-T, $\pi1^*$-T$'$, and $\pi1^*$-T orbitals, see Figure S5), and a singlet $\pi\pi^*$ charge-transfer state form the T to the T$'$ nucleobase (hereafter ct-$\pi\pi^*$-TT$'$ state, see Figure S4), are among the computed roots. On the basis of the obtained oscillator strengths (see Table 1), the $^1\pi\pi^*$-T$'$, $^1\pi\pi^*$-T, and $2\pi2\pi^*$ states will be the most populated after light interaction, although others, and in particular the ct-$\pi\pi^*$-TT$'$ state, have also non-negligible probability to be populated. The two $\pi\pi^*$ states localized on the carbon-carbon double bond on each thymine (i.e. the $^1\pi\pi^*$-T$'$ and $^1\pi\pi^*$-T states) are separated by 0.31 eV, reflecting a significant splitting in the otherwise degenerate $\pi\pi^*$ states, pointing out to a strong interaction among them in the $^1(gs)_{min}$ structure.

As discussed in the Introduction section, two main mechanisms have been formulated regarding the photoinduced formation mechanism of the thymine-thymine (6-4) adduct: one, passing through the initial population of $n\pi^*$ states localized on the reactive carbonyl, as the above presented $^1n\pi^*$-T$'$ and $^3n\pi^*$-T$'$ states; the second involving the population of a charge-transfer state from the T to the T$'$ moieties, as the here characterized ct-$\pi\pi^*$-TT$'$ state. For that reason in the present contribution we have focused on the optimized points reached by these three states ($^1n\pi^*$-T$'$, $^3n\pi^*$-T$'$, ct-$\pi\pi^*$-TT$'$) and also on the evolution of the $^1\pi\pi^*$-T$'$ state, being the latter the lowest bright excited state of the system.

Regarding their initial population after light absorption, the bright and low-lying $^1\pi\pi^*$-T$'$ state is indeed populated. The $^1n\pi^*$-T$'$ and $^3n\pi^*$-T$'$ states could gain population due to their energy proximity with the bright $^1\pi\pi^*$-T state, being the three states placed in a 0.06 eV energy range (see Table 1). The ct-$\pi\pi^*$-TT$'$ has a non-zero oscillator strength (equal to 0.00690), consequently indicating its possible direct population, but its energetic position, 7.05 eV above the ground state, points out to the need of high-energy radiations in order to reach it. A possible implication of the ct-$\pi\pi^*$-TT$'$ state in the photomechanism will explain the reported increase quantum yield of photoreaction at increased excitation energies and the absence of photoproduction under UVA radiation on the basis of the high energy position of such a charge transfer state. [11,19,28,29] It would be also interesting to verify if in the

present system the population of the ct-ππ*-TT' state can occur from a minimum of the ^{1}ππ*-T state though oscillations stabilizing the former state, as recently documented for a double-strand B-fragment.[24] Further works in these respect are currently in progress.

The study of the mentioned states have been operated performing, unless otherwise specified, CASSCF(8,6):AMBER optimizations, using the reduced (8,6) active space as described in the Computational Details section.

The optimization of the ^{1}nπ*-T' and ^{3}nπ*-T' states has led in both cases to geometries characterized by a significant elongation of the $C_{4'}O$ bond, increased by 0.158 and 0.172 Å with respect to the $^1(gs)_{min}$ structure, respectively (see Figure S7). At the minimum of the ^{3}nπ*-T' state, the $C_{4'}O$ has also undertaken a sizeable bending out of the plane of the ring and towards the T thymine molecule, as reflected by the $C_{6'}C_{5'}C_{4'}O$ dihedral angle, equal to 171.46° at the $^1(gs)_{min}$ structure, while equal to 148.24° at the ^{3}nπ*-T' minimum. The electronic states at both ^{1}nπ*-T' and ^{3}nπ*-T' minima have been characterized performing CASPT2(12,10):AMBER single point computations, state averaging over six singlet and six triplet states. As reported in Table S2, at the ^{1}nπ*-T' minimum the ^{1}nπ*-T' and ^{3}nπ*-T' states are almost degenerate in energy (energy gap equal to 0.09 eV) while no other singlet state is close in energy to the ^{1}nπ*-T'. This makes suppose that, despite the predictable small SOC between the ^{1}nπ*-T' and ^{3}nπ*-T' states due to their common nature, part of the ^{1}nπ*-T' population will be able to decay into the ^{3}nπ*-T' PEH, from which it can reach the corresponding minimum. In such a geometry, the ^{3}nπ*-T' state is the lowest triplet state, T1, separated from the T2 ^{3}ππ*-T' by a small energy gap of 0.32 eV (see Table S3). These facts suggest the possibility of the presence, here not verified, of a nearby conical intersection between the ^{3}nπ*-T' and ^{3}ππ*-T' states, as documented for the isolated thymine molecule.[23] Accordingly to the first suggested mechanism described above and passing through the nπ* population, from the ^{3}nπ*-T' minimum the population should decay to a triplet ππ* state who can evolve toward a reactive singlet-triplet crossing region with the ground state, along a path characterized by the shortening of the $O_{4'}$-C_5 distance and the gradual change of the nature of the triplet state from a localized ππ* to a charge-transfer TT' state, being the latter the one crossing with the ground state. Although at the ^{3}nπ*-T' minimum the ^{3}nπ*-T' and ^{3}ππ*-T' states are close in energy (separated by 0.32 eV), a possible decay toward the ^{3}ππ*-T' state and in particular the possibility that the system will then evolve in the direction of $O_{4'}$-C_5 approaching, starting from the ^{3}nπ*-T' minimum where the $O_{4'}$-C_5 distance has the significant value of 3.141 Å, seems unfavorable, as a preliminary optimization of ^{3}ππ*-T' state, which kept the two molecule separated, appears to indicate. In order to evaluate the possibility of population of a charge transfer state form the ^{3}nπ*-T' minimum, the latter geometry has been further characterized performing a CASPT2(12,10):AMBER state-averaged computation over 15 triplet roots. The computation has allowed the localization of a triplet ct-ππ*-TT' state (hereafter ^{3}ct-ππ*-TT') placed at 6.57 eV with respect to the ground state at the $^1(gs)_{min}$ minimum, and consequently 2.11 eV above the ^{3}nπ*-T' state at the ^{3}nπ*-T' minimum. The possibility of a decay from the ^{3}nπ*-T' toward the ^{3}ct-ππ*-TT' state, which in turn would be expected to favorably decay toward a reactive singlet-triplet crossing region, seems consequently remote.

On the basis of previous theoretical results on thymine and related systems,[23,51] the optimization of the $^1\pi\pi^*$-T$'$ state was expected to provide a minimum whose main deformation would have been a localized and significant elongation of the $C_{5'}C_{6'}$ double bond. Surprisingly, the obtained structure displays only a modest elongation of the mentioned carbon-carbon double bond (which increases by 0.029 Å with respect to the $^1(gs)_{min}$ structure), together with a non-negligible elongation of the $C_{4'}O$ bond (+0.036 Å with respect to the $^1(gs)_{min}$), and a bending of the $C_{4'}O$ carbonyl group that closely reminds the one characterizing the $^3n\pi^*$-T$'$ minimum ($C_{6'}C_{5'}C_{4'}O$ dihedral angle equal to 144.87°), see Figure S7. It must be noticed that such a minimum is the result of a QM/MM optimization in which the reduced CASSCF(4,4) active space (including the π-T$'$, π-T, π1*-T$'$, and π1*-T orbitals) has been employed, since the analogous computation with the larger CASSCF(8,6) active space did not converge. The CASPT2(12,10):AMBER vertical computation at the optimized $^1\pi\pi^*$-T$'$ structure reveals that the $^1\pi\pi^*$-T$'$ state is stabilized by 0.21 eV with respect to its energy at the $^1(gs)_{min}$ geometry (see Table S4). Interesting, at the obtained $^1\pi\pi^*$-T$'$ minimum, the $^1\pi\pi^*$-T$'$ state is degenerate in energy with the $^3n\pi^*$-T$'$ state (energy gap equal to 0.01) and consequently part of the $^1\pi\pi^*$-T$'$ population will be in principle able to populate the $^3n\pi^*$-T$'$ state. The SOC among the latter states is equal to 15.97 cm^{-1}.

From the $^1(gs)_{min}$ structure the decay of the ct-$\pi\pi^*$-TT$'$ state has been studied performing the corresponding CASSCF(8,6):AMBER optimization. The latter optimization did not reach a minimum, since the ct-$\pi\pi^*$-TT$'$ evolves toward a region of the potential energy hypersurface where it is energetically close to the ground state. In order to characterize such a degenerate region, different points along the performed optimization have been extracted and analyzed. In particular three structures have resulted of significance. In the first one, hereafter $(ct/\pi\pi^*)_{CI}$, the ct-$\pi\pi^*$-TT$'$ state is degenerate in energy with the $^1\pi\pi^*$-T state, displaying an energy gap of only 0.02 eV (see Table S5). However the energy of the ct-$\pi\pi^*$-TT$'$ state is here actually above its energy at the $^1(gs)_{min}$ of around 0.5 eV, indicating that the system will not favorably reach this region from the initial geometry. The second important structure characterized along the ct-$\pi\pi^*$-TT$'$ optimization is a conical intersection between the latter state and the ground state, hereafter $(ct/gs)_{CI\text{-}noreact}$. The two states are in fact separated by 0.06 eV, and actually also the T1 state, of ^{3}ct-$\pi\pi^*$-TT$'$ nature, is only 0.04 and 0.10 eV below the ct-$\pi\pi^*$-TT$'$ and the ground state, respectively (see Table S6) consequently making of the $(ct/gs)_{CI\text{-}noreact}$ also a singlet-triplet crossing region. The $(ct/gs)_{CI\text{-}noreact}$ structure closely resembles the conical intersection previously reported as the key geometry leading to oxetane formation, being in fact characterized by the formation of a new carbon-oxygen bond between the $O_{4'}$ and the C_5 atoms .[5,23] Despite its geometry, the optimization of the ground state from the $(ct/gs)_{CI\text{-}noreact}$ point evolves back to the two separated thymine molecules. As point out by the characterization of the last structure extracted from the ct-$\pi\pi^*$-TT$'$ optimization, hereafter $(ct/gs)_{CI\text{-}react}$, the formation of a new carbon-oxygen bond between the $O_{4'}$ and the C_5 atoms seems to be a required but not sufficient condition for the pro-reactivity of a conical intersection. The $(ct/gs)_{CI\text{-}react}$ geometry is characterized by both the formation of the mentioned new bond and by a sizeable puckering of the $C_{4'}$ atom toward the T nucleobase, as reflected by the value of the $N_{1'}C_{6'}C_{5'}C_{4'}$ dihedral angle, equal to 3.51° and 12.66° at the $(ct/gs)_{CI\text{-}noreact}$ and $(ct/gs)_{CI\text{-}react}$ points, respectively (see Figure S7). This additional geometrical feature appears to be key for a reactive evolution of the ground state, as proven by the ground state optimization from the $(ct/gs)_{CI\text{-}react}$ geometry, which indeed evolves to an oxetane structure,

hereafter oxe (see Figure S7). It should be noted that the $(ct/gs)_{CI\text{-}noreact}$ is not strictly speaking a conical intersection, being the ct-ππ*-TT' and the ground state separated by a 0.32 eV energy gap (see Table S7), while the degeneracy between T1 ^{3}ct-ππ*-TT' and the ground state, of only 0.05 eV, is sufficient to consider the structure a proper singlet-triplet crossing region. It is plausible to believe that a similar pro-reactive conical intersection should be located in the nearby regions, although in the present contribution it has not been identified. Another possibility, in agreement with the mentioned fact that only $(ct/gs)_{CI\text{-}noreact}$ is strictly speaking a conical intersection, is that actually both $(ct/gs)_{CI\text{-}noreact}$ and $(ct/gs)_{CI\text{-}react}$ structures belong to the same seam of intersection and that the $(ct/gs)_{CI\text{-}react}$ geometry represents the direction along which the decay through the $(ct/gs)_{CI\text{-}noreact}$ conical intersection will evolve toward oxetane formation. This in turn means that a reactive or non-reactive decay will be determined by the direction and velocity with which the system passes through the $(ct/gs)_{CI\text{-}noreact}$ conical intersection: reactive, if evolving towards the $(ct/gs)_{CI\text{-}react}$ geometry, non-reactive in the other case.

On the basis of the described results, the charge-transfer ct-ππ*-TT' state can evolve from the $^1(gs)_{min}$, characterized by the relatively short $O_{4'}$-C_5 distance of 2.810 Å, toward a region of energy degeneracy with the ground state where both pro-reactive structures (as the $(ct/gs)_{CI\text{-}react}$) and no-reactive structures (as the $(ct/gs)_{CI\text{-}noreact}$) can be reached.

It is difficult to predict the percentage of the ct-ππ*-TT' population that will follow the reactive or the non-reactive path, since a dynamics treatment of the reaction will be required in order to answer such a question. What it possible to notice nevertheless, is that the non-reactive $(ct/gs)_{CI\text{-}noreact}$ conical intersection is placed around 0.5 eV below the reactive $(ct/gs)_{CI\text{-}react}$ structure (see table S6 and S7). It should however be kept in mind that the title photoreaction is an inefficient photoreaction, and in fact for the particular system here under study, characterized by displaying a relatively high quantum yield of the photoreaction, the latter is equal to the low value of $0.75\ 10^{-3}$.[27]

The obtained main results and energies are reported schematically in Figure 3 and in Table 2.

Table 1. Calculated CASPT2(12,10):AMBER Vertical Excitation Energies at the $^1(gs)_{min}$ Geometry (E_{VA}, eV) for the Lowest Valence Singlet and Spin Forbidden Triplet Excited States.[a]

State	E_{VA} (eV)	*f*	μ (D)
S0 (gs)	0.00	-	12.58
T1 ($^3\pi\pi^*$-T')	3.93	-	12.73
T2 ($^3\pi\pi^*$-T)	3.97	-	12.58
S1 ($^1\pi\pi^*$-T')	4.90	0.18487	13.81
S2 ($^1\pi\pi^*$-T)	5.21	0.93961	14.61
T3 ($^3n\pi^*$-T)	5.21	-	9.95
T4 ($^3n\pi^*$-T')	5.23	-	10.93
S3 ($^1n\pi^*$-T)	5.25	0.00075	9.64
S4 ($^1n\pi^*$-T')	5.27	0.00023	10.73
T5 ($^3\pi_O\pi^*$-T')	6.18	-	10.68
S5 ($2\pi2\pi^*$)	6.24	0.32562	13.22
T6 ($^3\pi_O\pi^*$-T)	6.30	-	10.38
S6 (ct-$\pi\pi^*$-TT')	7.05	0.00690	9.46
S7 ($^1\pi_O\pi^*$-T')	7.17	0.01275	10.92
S8	7.59	0.01781	9.25
T7	7.76	-	12.55
S9 ($^1\pi_O\pi^*$-T)	7.99	0.00512	10.76
T8	8.28	-	9.95
T9	8.33	-	10.96
T10	8.99	-	10.03

[a] The computed dipole moments (μ, D) and the oscillator strengths (*f*) for the singlet−singlet transitions are also included.

Table 2. Calculated CASPT2(12,10):AMBER energies (eV) of the Main Characterized States and Various Optimized Structures.[a]

State Geometry	gs	$^1n\pi^*$-T'	$^3n\pi^*$-T'	$^1\pi\pi^*$-T'	$^3\pi\pi^*$-T'	ct-$\pi\pi^*$-TT'	^{3}ct-$\pi\pi^*$-TT'
$^1(gs)_{min}$	0.00	5.27	5.23	4.90	3.93	7.05	-
$^1n\pi^*$-T' minimum	1.59	4.80	4.71	5.87	4.25	-	-
$^3n\pi^*$-T' minimum	1.93	4.73	4.46	5.87	4.78	-	6.57
$^1\pi\pi^*$-T' minimum	1.83	5.33	4.68	4.69	4.98	-	-
$(ct/gs)_{CI\text{-}noreact}$	4.53	-	-	-	-	4.47	4.43
$(ct/gs)_{CI\text{-}react}$	5.32	-	-	-	-	5.02	5.27

[a] All the reported values are referred to the ground state energy at the $^1(gs)_{min}$ geometry.

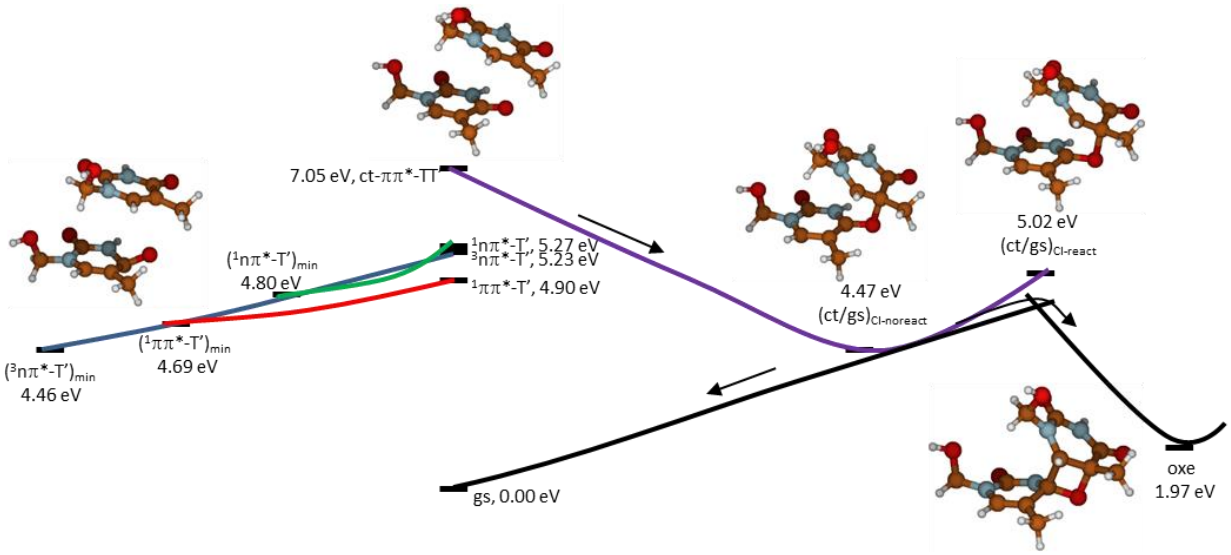

Figure 3. Schematic representation of the characterized paths. Red, blue, green, purple, and black solid lines indicate the ^{1}ππ*-T', ^{3}nπ*-T', ^{1}nπ*-T', ct-ππ*-TT', and ground state, respectively.

CONCLUSIONS

In the present contribution QM(CASPT2//CASSCF):MM(AMBER) calculations have been performed in order to elucidate the photoinduced formation mechanism of the thymine-thymine (6-4) adduct in the specific hairpin $H_4$2 for which the experimental quantum yield of the photoreaction have been measured.[27] Starting from a presumably pro-reactive ground state minimum, characterized by a relatively short $O_{4'}$-C_5 distance (2.810 Å), various excited states have been optimized and the corresponding minima analyzed.

The main conclusion here obtained is that from the characterized starting geometry a high-lying charge transfer state ct-ππ*-TT' can decay toward a region of energy degeneracy with the ground state, where both reactive and non-reactive geometries have been documented. Two types of regions have been in fact described along the ct-ππ*-TT' PEH. The first, exemplified by the (ct/gs)$_{CI-noreact}$ structure, is characterized by the formation of a new carbon-oxygen $O_{4'}$-C_5 bond, and from such a geometry the ground state decays back to the two separated thymine molecules. The second, exemplified by the (ct/gs)$_{CI-react}$ structure, is again characterized by the formation of a new carbon-oxygen $O_{4'}$-C_5 bond, but it also displays a significant puckering of the $C_{4'}$ atoms toward the C_6 atom, that are the atoms among which a new carbon-carbon bond needs to be formed in order to create the oxetane ring. This latter geometrical feature seems to be fundamental for the pro-reactivity, since, as proven by the corresponding optimization, from the (ct/gs)$_{CI-react}$ structure the ground state evolves towards oxetane formation. The involvement of the here characterized ct-ππ*-TT' charge-transfer state is in agreement with the recent work of Conti at al.[24] The identification of a reactive path form the characterized starting ground state geometry, ^{1}(gs)$_{min}$, consequently points out to the importance of a short $O_{4'}$-C_5 distance for the formation of the thymine-thymine (6-4) adduct photodamage.

The possibility of a reaction mechanism passing through the initial population of nπ* states have been here also partially explored. It appears that the first part of the proposed mechanism, leading to the population of the minima of singlet and triplet nπ* states localized on the reactive carbonyl group (here the ^{1}nπ*-T' and ^{3}nπ*-T' minima) is indeed possible, although it is not clear what will be the driving force of the second part of the process, leading from a minimum in which the reactive $O_{4'}$-C_5 atoms are separated by more than 3 Å (as in the ^{1}nπ*-T' and ^{3}nπ*-T' minima, see Figure S7) towards a reactive singlet-triplet crossing region with the ground state where a new carbon-oxygen $O_{4'}$-C_5 bond is formed. Further computational studies are needed in order to better clarify this point.

ACKNOWLEDGEMENTS

M.G acknowledges support by the European Research Council Advanced Grant STRATUS (ERC-2011-AdG No. 291198). A. G. acknowledges the support form the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 658173.

REFERENCES

(1) J. Cadet and P. Vigny The Photochemistry of Nucleic Acids. In *Bioorganic Photochemistry*; Morrison, H., Ed.; John Wiley & Sons, Inc.:New York, 1990; Vol. 1, pp 1-272.

(2) R. P. Sinha and D. P. Hader *Photochem. Photobiol. Sci.* 2002**, 1**, 225-236.

(3) J. S. Taylor and M. P. Cohrs *J. Am. Chem. Soc.* 1987, **109**, 2834–2835.

(4) M. Boggio-Pasqua, G. Groenhof, L. V. Schäfer, H. Grubmüller and M. Robb *J. Am. Chem. Soc.* **2007**, **129**, 10996-10997.

(5) L. Blancafort and A. Migani *J. Am. Chem. Soc.* 2007**, 129**, 14540-14541.

(6) D. Roca-Sanjuán, G. Olaso-González, I. González-Ramírez, L. Serrano-Andrés and M. Merchán *J. Am. Chem. Soc.* 2008**, 130**, 10768-10779.

(7) J. J. Serrano-Pérez, I. González-Ramírez, P. B. Coto, M. Merchán and L. Serrano-Andrés *J. Phys. Chem. B* 2008**, 112**, 14096-14098.

(8) T. Climent, I. González-Ramírez, R. González-Luque, M. Merchán and L. Serrano-Andrés *J. Phys. Chem. Lett.* 2010**, 1**, 2072-2076.

(9) F. Manson, T. Laino, I. Tavernelli, U. Rothlisberger and J. Hutter *J. Am. Chem. Soc.* 2008**, 130**, 3443-3450.

(10) M. Barbatti *ChemPhysChem* 2014, **15**, 3342-3354.

(11) R. Improta *J. Phys. Chem. B* 2012, **116**, 14261-14274.

(12) V. B. Delcheva and W. Domcke *Journal of Photochemistry and Photobiology A: Chemistry* 2013, **271**, 1-7.

(13) S. Mouret, C. Baudouin, M. Charveron, A. Favier, J. Cadet and T. Douki *Proc. Natl. Acad. Sci. USA* 2006**, 103**, 13765-13776.

(14) W. J. Schreier, T. E. Schrader, F. O. Koller, P. Gilch, C. E. Crespo-Hernández, V. N. Swaminathan, T. Carell, W. Zinth and B. Kohler. *Science* 2007**, 315**, 625-629.

(15) W. M. Kwok, C. Ma and D. L. Phillips *J. Am. Chem. Soc.* 2008**, 130**, 5131-5139.

(16) T. Douki, I. Bérard, A. Wack and S. Andrä *Chem. Eur. J.* 2014, **19**, 5787-5794.

(17) R. R. Ramazanov, D. A. Maksimov and A. I. Kononov *J. Am. Chem. Soc.* 2015, **137**, 11656-11665.

(18) M. C. Cuquerella, V. Lhiaubet-Vallet, F. Bosca and M. A. Miranda *Chem. Sci.* 2011**, 2**, 1219-1232.

(19) A. Banyasz, T. Douki, R. Improta, T. Gustavsson, D. Onidas, I. Vayá, M. Perron and D. Markovitsi *J. Am. Chem. Soc.* 2012, **134**, 14834-14845.

(20) S. Marguet and D. Markovitsi *J. Am. Chem. Soc.* 2005**, 127**, 5780-5781.

(21) Z. B. Yang, R. B. Zhang and L. A. Eriksson *Phys. Chem. Chem. Phys.* 2011**, 13**, 8961-8966.

(22) Z. B. Yang, L. A. Eriksson and R. B. Zhang *J. Phys. Chem. B* 2011**, 115**, 9681-9686.

(23) A. Giussani, L. Serrano-Andrés, M. Merchán, D. Roca-Sanjuán and M. Garavelli *J. Phys. Chem. B* 2013, **117**, 1999-2004.

(24) I Conti, L. Martínez-Fernández, L. Esposito, S. Hofinger, A. Nenov, M. Garavelli, and R. Improta *Chem. Eur. J.* 2017, **23**, 1-13.

(25) T. Douki *Photochem. Photobiol. Sci*. 2013, **12**, 1286

(26) Z. Pan, M. McCullagh, G. C. Schatz and F. D. Lewis *J. Phys. Chem. Lett.* 2011, **2**, 1432-1438.

(27) M. Hariharan, K. Siegmund, C. Saurel, M. McCullagh, G. C. Schatz and F. D. Lewis *Photochem. Photobiol. Sci.* 2014, **13**, 266.

(28) J. Cadet, E. Sageb and T. Douki *Mutation Research* 2005, **571**, 3-17.

(29) J. Cadet, S. Mouret, J.-L. Ravanat and T. Douki *Photochemistry and Photobiology*, 2012, **88**, 1048-1065.

(30) T. Douki and J. Cadet *Biochemistry*, 2001, **40**, 2495.

(31) E. Paterno and G. Chiefi *Gazz. Chim. Ital.* 1909, **39**, 431.

(32) I. J. Palmer, I. N. Ragazos, F. Bernardi, M Olivucci and M. A. Robb *J. Am. Chem. Soc.* 1994**, 116**, 2121-2132.

(33) B. S. Maeda, T. Taketsugu and K. Morokuma K. *Z. Phys. Chem.* 2013, **227**, 1421-1433.

(34) S. J. Harris, D. Murdock, M. P. Grubb, I. P. Clark, G. M. Greetham, M. Towrie and M. N. R. Ashfold *J. Phys. Chem. A* 2014, **118**, 10240-10245.

(35) R. Y. Brogaard, O. Schalk, A. E. Boguslavskiy, G. D. Enright, H. Hopf, V. Raev, E. Tarcoveanu, T. I. Sølling and A. Stolow *Phys. Chem. Chem. Phys.*, 2012,**14**, 8572-8580.

(36) D. Case, AMBER 12, University of California, San Francisco, 2012.

(37) R. Salomon-Ferrer, D. A. Case and R. C. Walker *WIREs Comput. Mol. Sci.* 2013, **3**, 198-210.

(38) P. Altoè, M. Stenta, A. Bottoni and M. Garavelli *Theor. Chem. Acc*. 2007, **118**, 219.

(39) B. O. Roos, P. R. Taylor and P. E. M. Siegbahn *Chem. Phys.* 1980, **48**, 157.

(40) K. Andersson, P.-Å Malmqvist, B. O. Roos, A. J. Sadlej and K. Wolinski *J. Phys. Chem*. 1990, **94**, 5483.

(41) K. Andersson, P.-Å. Malmqvist and B. O. Roos, *J. Chem. Phys.* 1992, **96**, 1218.

(42) B. O. Roos, K. Andersson, M. P. Fülscher, P.-Å. Malmqvist, L. Serrano-Andrés, K. Pierloot and M. Merchán *Adv. Chem. Phys.* 1996, **93**, 219.

(43) R. Montero, A. Longarte, Á. Peralta Conde, C. Redondo, F. Castaño, I. Gonzalez-Ramírez, A. Giussani, L. Serrano-Andrés and M. Merchán *J. Phys. Chem. A* 2009, **113**, 13509.

(44) F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez Galván, N. Ferre´, L. M. Frutos, L. Gagliardi, et al. *J. Comput. Chem.* 2016, **37**, 506.

(45) M. L. Frisch, et al., Gaussian Inc, Wallingford CT, 2004.

(46) A. Giussani *J. Chem. Theory Comput.* 2014, **10**, 3987-3995.

(47) P.-O. Widmark, P.-Å. Malmqvist and B. O. Roos *Theor. Chem. Acc*. 1990, **77**, 291.

(48) K. Pierloot, B. Dumez, P.-O. Widmark and B. O.Roos *Theor. Chem. Acta* 1995, **90**, 87.

(49) F. Aquilante, T. B. Pedersen and R. Lindh *Theor. Chem. Acc*. 2009, **124**, 1.

(50) J. J. Serrano-Pérez, R. González-Luque, M. Merchán and L. Serrano-Andrés *J. Phys. Chem. B* 2007, **111**, 11880-11883.

(51) A. Giussani, J. Segarra-Martí, D. Roca-Sanjuán and M. Merchán *Top. Curr. Chem*. 2015, **355**, 57-97.

(52) C. R. Kozak, K. A. Kistler, Z. Lu and S. Matsika *J. Phys. Chem. B* 2010, **114**, 1674-1683.