Photochemistry of thymine in solution and DNA revealed by an electrostatic embedding QM/MM combined with mixed-reference spin-flip TDDFT

Miquel Huix-Rotllant, Karno Schwinn, Vladimir Pomogaev, Maryam Farmani, Nicolas Ferré, Seunghoon Lee, Cheol Ho Choi

Photochemistry of thymine in solution and DNA revealed by an electrostatic embedding QM/MM combined with mixed-reference spin-flip TDDFT

Miquel Huix-Rotllant,*,† Karno Schwinn,† Vladimir Pomogaev,‡ Maryam Farmani,‡ Nicolas Ferré,† Seunghoon Lee,¶ and Cheol Ho Choi*,‡

†*Aix-Marseille Univ, CNRS, ICR, Marseille, France*
‡*Department of Chemistry, Kyungpook National University, Daegu 41566, South Korea*
¶*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA*

E-mail: miquel.huix-rotllant@cnrs.fr; cchoi@knu.ac.kr

Abstract

The photochemistry of nucleobases, important for their role as building blocks of DNA, is largely affected by the electrostatic environment in which they are soaked. For example, despite the numerous studies of thymine in solution and DNA, there is still a debate on the photochemical deactivation pathways after UV absorption. Many theoretical models are oversimplified due to the lack of computationally accurate and efficient electronic structure methodologies that capture excited state electron correlation effects when nucleobases are embedded in large electrostatic media. Here, we combine mixed-reference spin-flip time-dependent density-functional theory (MRSF-TDDFT) with electrostatic embedding QM/MM using electrostatic potential fitting (ESPF) atomic charges, as a strategy to accurately and efficiently describe the electronic structure of chromophores polarized by an electrostatic medium. In particular, we develop analytic expressions for the energy and gradient of MRSF/MM based on the ESPF coupling using atom-centered grids and total charge conservation. We apply this methodology to the study of solvation effects on thymine photochemistry in water and thymine dimers in DNA. In the former, the combination of trajectory surface hopping (TSH) non-adiabatic molecular dynamics (NAMD) with MRSF/MM remarkably revealed the accelerated deactivation decay pathways, which is consistent with the experimental decay time of $\sim$ 400 fs. The enhanced hopping rate can be explained by the preferential stabilization of corresponding conical interactions due to their increased dipole moments. Structurally, it is a consequence of characteristic methyl puckered geometries near the conical intersection region. For the thymine dimer in B-DNA, we found new photochemical pathways through conical intersections that could explain the formation of cyclobutadiene dimers and 6-4 photoproducts.

1 Introduction

Photobiology, the study of photoactive biological macromolecules, has become an active field of research in recent years. DNA UV photodamage, the mechanism of vision by rhodopsins or the photosystems I and II involved in photosynthesis, are just a few well-known examples of the objects under study in this field of research. Other important photoactive proteins under intensive investigation are flavoproteins, luciferases, hemoglobins, photoactive yel-

low proteins, and green fluorescent proteins, just to mention a few. Many photoactive macromolecules are characterized by rapid photochemical events (few picoseconds) implying electronic state crossings like those frequently found in rhodopsins, cryptochromes, myoglobin or DNA nucleobases.[1]

DNA nucleobases are perhaps one of the clearest and best-studied examples of nonradiative decays through conical intersections, which offer natural protection of monomers against UV mutations. Despite this, depending on the DNA sequence, mutations can be formed between stacked nucleobases. This is the case for the well-known thymine dimers,[2–4] however, its detailed mechanism is still elusive due to the lack of realistic modelings. Understanding the mechanism by which DNA disposes of this excess energy is a crucial element of photobiology. Due to the complexity of DNA strands, many studies have been investigating smaller nucleic acid (NA) constituents such as nucleobases, nucleosides, and nucleotides.[5] While many of these studies have been performed on isolated gas phase molecules, the ability of NA constituents to dissipate excess energy into the surrounding solvent environment (namely, water) is of particular interest, which may provide a clue to the understanding of dynamics in much more complex DNA strands.

Computationally, various QM/MM schemes have been the main workforce for biological systems.[6] The ESPF method is a way of accounting for the QM and MM electrostatic interactions that can be formulated in any of the embedding strategies, from mechanical to *polarizable*, which is an important requirement to cope with complex biological environments.

Photobiology additionally requires efficient and accurate descriptions of both ground and excited states. Recently, mixed-reference spin-flip (MRSF)-TDDFT[7,8] ingeniously combines the reduced density matrices (RDMs) of the two $M_S = +1$ and -1 triplet-ground references within the linear response theory, providing additional nondynamic types of electron correlation – the missing piece of conventional LR theories. The balanced dynamic and nondynamic electron correlations of MRSF-TDDFT enable accurate computation of the excitation energies[9] and the real (and avoided) conical intersections overcoming the limitation of TDDFT.[10] Simultaneously, MRSF-TDDFT eliminates the major drawback of the spin-flip TDDFT methods,[11] the spin-contamination of the ground and excited states. The introduction of the tensor equation-of-motion (TEOM) formalism yielded the spin-adopted (SA) SF-TDDFT that is also free of spin contamination.[12] However, the matrix elements of TEOM are evaluated using the Wigner-Eckart theorem, which is not satisfied by the approximate density functionals. Hence, the SA-SF-DFT formalism was derived by introducing an *aposteriori* DFT correction into the equations. In addition, due to the complexity of TEOM, the analytic energy gradient for the SA-SF-DFT has yet to be derived.

On the other hand, the efficient analytical gradient of MRSF-TDDFT has been developed allowing current molecular dynamics studies.[8] In the recent studies[13] of the S_2-to-S_1 internal conversions (ICs) of s-*trans*-butadiene and s-*trans*-hexatriene, the results of MRSF-TDDFT are quantitatively consistent with the high-level equation-of-motion coupled-cluster (EOMCC) theories as well as multi-state many-body perturbation theory methods, where both CASSCF and TDDFT fail to describe them due the missing of either dynamic correlation or the doubly excited configurations. In a series of recent non-adiabatic molecular dynamics (NAMD) simulations[14–16] in gas phase, it has been demonstrated that the MRSF-TDDFT can outperform other popular theories such as CASSCF and TDDFT and can reveal new nonadiabatic processes. The particular combination of ESPF and MRSF-TDDFT in a QM/MM scheme, thus, potentially can provide the most accurate yet efficient tool in the field of photobiology.

In the current paper, we first present the hybrid formulations of ESPF and MRSF-TDDFT in the context of QM/MM. Applications of the new QM/MM protocol on the photochemistry of solvated thymine as well as thymine dimers in DNA shall be followed to demonstrate its performance.

2 Methodology

We use the notation in capitals $\mathbf{R}_A = X_A, Y_A, Z_A$ and $\tilde{\mathbf{R}}_A = \tilde{X}_A, \tilde{Y}_A, \tilde{Z}_A$ to symbolize the positions of QM and MM atom A, respectively. The partial derivatives with respect to the coordinate X of atom C will be symbolically expressed as

$$E^{X_C} = \frac{\partial E}{\partial X_C}\,. \tag{1}$$

The derivative notation in parenthesis (X_C) refers to the derivative of operators, keeping the basis functions fixed.

Labels $i, j, \cdots$, $x, y, \cdots$ and $a, b, \cdots$ respectively refer to indices for doubly-, singly- occupied and virtual KS spin molecular orbitals (MOs) of reference states, whereas p, q, r, s, t, u, refer to generic MO indices. Greek indices μ,ν,κ, and δ refer to atomic orbitals while σ and τ index the spin function of the MO.

2.1 MRSF-TDDFT/MM

Here we formulate an electrostatic embedding QM/MM method for excited states, combining the strengths of MRSF-TDDFT and electrostatic embedding QM/MM using electrostatic potential fitted (ESPF) charges. We recently developed the ESPF QM/MM method including total charge conservation and grid derivative effects up to the second derivative of the energy for ground-state single-determinant wavefunctions solutions of a set of Fock equations, showing an overall linear scaling with respect to MM subsystem size.[17–19] A particular transformation to spinor-like mixed spin functions combines $|\alpha\alpha\rangle$ ($M_S = +1$) and $|\beta\beta\rangle$ ($M_S = -1$) of a restricted open Hartree-Fock (ROHF) triplet into a *hypothetical* single reference[7,8] in the MRSF-TDDFT formulation.

2.1.1 Energy

The total energy of an excited state I with a total spin S at the QM/MM level is given by,

$$E_I^{[S]} = E_{I,\mathrm{QM}}^{[S]} + E_{\mathrm{MM}}\,. \tag{2}$$

in which the QM energy is $E_{I,QM}^{[S]} = E_{\mathrm{ref}} + \Omega_I^{[S]}$ is the sum of the reference state energy and the MRSF excitation energy $\Omega_I^{[S]}$. In an electrostatic embedding QM/MM formulation these energies are obtained from the solution of a hamiltonian containing the interaction term between the QM and MM systems. Using electrostatic potential fitting charges, we can define such interaction as a one-electron operator

$$h_{\mu\nu}^{\mathrm{ESPF}} = \sum_A^{N_{\mathrm{QM}}} \left(\Phi^{\mathrm{av}} - \phi_A\right) Q_{A,\mu\nu} - \Phi^{\mathrm{av}} S_{\mu\nu}\,, \tag{3}$$

in which ϕ_A is the external electrostatic field generated by the MM charges on QM center A, the average external field is defined as $\Phi^{\mathrm{av}} = N_{\mathrm{QM}}^{-1} \sum_A \phi_A$, $\mathbf{S}$ is the overlap matrix and $\mathbf{Q}_A$ is the net charge population of QM center A, which is defined as

$$Q_{A,\mu\nu} = \sum_k^{N_{\mathrm{grid}}} T_{A,k}^{+} \langle\chi_\mu| \frac{1}{|\mathbf{r} - \mathbf{r}_k|} |\chi_\nu\rangle\,. \tag{4}$$

Here, $\mathbf{r}_k = \mathbf{R}_A + \alpha \cdot \mathbf{r}_l$ refers to a Lebedev grid point, constructed on the QM centers by adding a Lebedev points $\mathbf{r}_l$ multiplied by a factor α that ensures that the grid points lay outside the Van der Waals surface of the molecule. The matrix $\mathbf{T}^+$ is defined as $\mathbf{T}^+ = (\mathbf{T}^\dagger\mathbf{T})^{-1}\mathbf{T}^\dagger$ and corresponds to the pseudoinverse of the electrostatic kernel $\mathbf{T}_{k,A} = |\mathbf{r}_k - \mathbf{R}_A|^{-1}$. The electrostatic one-electron integrals are constructed on a grid around the QM atoms. To respect the symmetry properties of the Hamiltonian in an electrostatic embedding formulation, it is advantageous that the grid is atom-centred.[19]

The MRSF excitation energies are obtained from poles of the response function of the mixed reference (within the Tamm-Dancoff approximation)

$$\sum_{rs} A_{pq,rs}^{[S]} X_{rs,I}^{[S]} = X_{pq,I}^{[S]} \Omega_I^{[S]}\,, \tag{5}$$

for the singlet ($S = 0$) and the triplet ($S = 1$) states. The orbital Hessian matrix consists of two matrices, i.e. $\mathbf{A}^{[S]} = \mathbf{A}^{[S](0)} + \mathbf{A}'^{[S]}$. On the one hand, the spin-pairing coupling matrix ($\mathbf{A}'^{[S]}$) is a *posteriori* coupling between the re-

sponses from the two references. This coupling is given by a series of two-electron integrals, and thus, remains equivalent in the ESPF QM/MM formulation. On the other hand, $\mathbf{A}^{[S](0)}$ is given by

$$\begin{aligned} A^{[S](0)}_{pq,rs} &= U^{[S]}_{pq} \left[\delta_{pr} F^{\mathrm{p}}_{qs} \right. \\ &- \left. \delta_{qs} F^{\mathrm{h}}_{pr} - c_H (pr|sq)\right] U^{[S]}_{rs} \end{aligned} \quad (6)$$

in which

$$\begin{aligned} F^{\mathrm{h}}_{pr} &= h^0_{pq} + h^{\mathrm{ESPF}}_{pq} \quad (7) \\ &+ \sum_i \left[2(pq|ii) - c_H (pi|iq)\right] \\ &+ \sum_x \left[(pq|xx) - c_H (px|xq)\right] \\ F^{\mathrm{p}}_{pr} &= h^0_{pq} + h^{\mathrm{ESPF}}_{pq} \\ &+ \sum_i \left[2(pq|ii) - c_H (pi|iq)\right] \\ &+ \sum_x (pq|xx), \quad (8) \end{aligned}$$

h^0_{pq} and $(pq|rs)$ are the one- and two-electron integrals, respectively, c_H is the mixing coefficient for the exact Hartree-Fock exchange, and $U^{[S]}_{pq}$ is the dimensional-transformation matrix introduced in Ref. 8 in such a manner that a majority of existing SF-TDDFT codes can be utilized for the implementation of MRSF-TDDFT. We can find the ESPF contribution in the Fock matrix as an additional mean-field contribution.

2.1.2 Analytic excited state energy gradient

The gradient of the total energy defined in Eq. 2 with respect to a QM and MM atom is respectively given by

$$\begin{aligned} E^{[S],X_C}_I &= E^{[S],X_C}_{I,\mathrm{QM}} + E^{X_C}_{\mathrm{MM}} \\ E^{[S],\tilde{X}_C}_I &= E^{[S],\tilde{X}_C}_{I,\mathrm{QM}} + E^{\tilde{X}_C}_{\mathrm{MM}}\,, \end{aligned} \quad (9)$$

in which the gradient of the excited state energy with respect to QM and MM atoms are given by $E^{[S],X_C}_{I,\mathrm{QM}} = E^{X_C}_{\mathrm{ref}} + \Omega^{[S],X_C}_I$ and $E^{[S],\tilde{X}_C}_{I,\mathrm{QM}} = E^{\tilde{X}_C}_{\mathrm{ref}} + \Omega^{[S],\tilde{X}_C}_I$. The expressions for the reference state gradients $E^{X_C}_{\mathrm{ref}}$ and $E^{\tilde{X}_C}_{\mathrm{ref}}$ are given in Ref. 19 for RHF wavefunctions and Ref. 18 for ROHF/UHF wavefunctions, whereas the $E^{X_C}_{\mathrm{MM}}$ and $E^{\tilde{X}_C}_{\mathrm{MM}}$ have trivial analytic expressions of the classical energy terms. Since the QM and MM atoms are treated at different levels, their respective gradients have different mathematical expressions. The first derivative of the ESPF hamiltonian (Eq. 3) with respect to QM and MM atom perturbations is respectively given by

$$\begin{aligned} h^{\mathrm{ESPF},X_C}_{\mu\nu} &= \sum_A^{N_{\mathrm{QM}}} \left(\Phi^{\mathrm{av},X_C} - \phi^{X_C}_A\right) Q_{A,\mu\nu} \\ &+ \sum_A^{N_{\mathrm{QM}}} \left(\Phi^{\mathrm{av}} - \phi_A\right) Q^{X_C}_{A,\mu\nu} \\ &- \Phi^{\mathrm{av}} S^{X_C}_{\mu\nu} - \Phi^{\mathrm{av},X_C} S_{\mu\nu} \\ h^{\mathrm{ESPF},\tilde{X}_C}_{\mu\nu} &= \sum_A^{N_{\mathrm{QM}}} \left(\Phi^{\mathrm{av},\tilde{X}_C} - \phi^{\tilde{X}_C}_A\right) Q_{A,\mu\nu} \\ &- \Phi^{\mathrm{av},\tilde{X}_C} S_{\mu\nu}\,, \end{aligned} \quad (10)$$

in which the first derivative of the ESPF operator with respect to QM atom perturbations is respectively given by

$$\begin{aligned} Q^{X_C}_{A,\mu\nu} &= \sum_k^{N_{\mathrm{grid}}} T^{+,X_C}_{A,k} \langle\chi_\mu| \frac{1}{|\mathbf{r} - \mathbf{r}_k|} |\chi_\nu\rangle \\ &+ \sum_k^{N_{\mathrm{grid}}} T^{+}_{A,k} \left[\langle\chi_\mu| \frac{1}{|\mathbf{r} - \mathbf{r}_k|} |\chi_\nu\rangle\right]^{X_C}. \end{aligned}$$

The derivative of the pseudo-inverse kernel can be expressed as,

$$\begin{aligned} T^{+,X_C}_{A,k} &= \sum_B^{N_{\mathrm{QM}}} \left[(\mathbf{T}^\dagger \mathbf{T})\right]^{-1,X_C}_{A,B} T_{B,k} \\ &+ \sum_B^{N_{\mathrm{QM}}} \left[(\mathbf{T}^\dagger \mathbf{T})\right]^{-1}_{A,B} T^{X_C}_{B,k}\,. \end{aligned} \quad (11)$$

The inverse derivative in the first term can be constructed from the derivatives of $\mathbf{T}$ using the expression $(\mathbf{T}^\dagger\mathbf{T})^{-1,X_C} = -(\mathbf{T}^\dagger\mathbf{T})^{-1}(\mathbf{T}^\dagger\mathbf{T})^{X_C}(\mathbf{T}^\dagger\mathbf{T})^{-1}$. The derivative

of the inverse of the square of the electrostatic kernel can be overall avoided by an equivalent expression involving only derivatives of $\mathbf{T}$.[17,19]

Combining these results with the derivative of the MRSF-TDDFT/MM excitation energy (Eq. 5) leads to an analytic gradient expression with respect to QM and MM atoms given respectively by

$$\begin{aligned}
\Omega_I^{[S],X_C} &= \sum_{\mu\nu\sigma} \left(h_{\mu\nu}^{0,X_C} + h_{\mu\nu}^{\mathrm{ESPF},X_C}\right) P_{\mu\nu,I}^{[S]} \\
&- \sum_{\mu\nu\sigma} S_{\mu\nu}^{X_C} W_{\mu\nu,I}^{[S]} \\
&+ \sum_{\mu\nu\sigma,\kappa\lambda\tau} (\mu\nu|\kappa\tau)^{X_C} \Gamma_{\mu\nu\sigma,\kappa\lambda\tau,I}^{[S]} \\
\Omega^{\tilde{X}_C} &= \sum_{\mu\nu\sigma} h_{\mu\nu}^{\mathrm{ESPF},\tilde{X}_C} P_{\mu\nu,I}^{[S]}
\end{aligned} \tag{12}$$

, where $\mathbf{P}_I^{[S]}$ and $\mathbf{\Gamma}_I^{[S]}$ are the relaxed difference one- and two-particle density matrices, respectively, and $\mathbf{W}_I^{[S]}$ is the relaxed Lagrange multiplier for the orthonormality of MOs.

2.1.3 Branching update method for approximate QM/MM conical intersection search

In the branching update method,[20] the optimization is performed by minimizing the energy $E = (E_I + E_J)/2 + (E_J - E_I)^2$, using the projected gradient

$$\nabla E = B\nabla E_{\mathrm{av}} + 4(E_I - E_J)\nabla E_{\mathrm{d}}\,, \tag{13}$$

in which the average gradient is given by $\nabla E_{\mathrm{av}} = (\nabla E_I + \nabla E_J)/2$ and the gradient difference is given by $\nabla E_{\mathrm{d}} = (\nabla E_I - \nabla E_2)/2$. The projector B projects out the branching plane vectors from the average gradient, and is defined as,

$$B = 1 - \nabla E'_{\mathrm{d}} \nabla E'^{\mathrm{T}}_{\mathrm{d}} - \nabla E'_{\mathrm{o}} \nabla E'^{\mathrm{T}}_{\mathrm{o}}\,, \tag{14}$$

in which ∇E_{o} is an vector orthogonal to ∇E_{d}. The prime indicates that these vectors are normalized. The ESPF QM/MM, difference gradient vector between two excited states is defined as

$$\nabla\left(E_J - E_I\right) = \begin{bmatrix} \Omega_J^{X_C} - \Omega_I^{X_C} \\ \\ \Omega_J^{\tilde{X}_C} - \Omega_I^{\tilde{X}_C} \end{bmatrix}, \tag{15}$$

As for the orthogonal gradient, it is estimated by a simple update procedure,

$$\begin{aligned}
\nabla E'_{\mathrm{orth},k} &= \frac{1}{N}\left[\left(\nabla E'_{\mathrm{o},k-1}\nabla E'_{\mathrm{d},k}\right)\nabla E'_{\mathrm{d},k-1}\right. \\
&- \left.\left(\nabla E'_{\mathrm{d},k-1}\nabla E'_{\mathrm{d},k}\right)\nabla E'_{\mathrm{o},k-1}\right]\,,
\end{aligned} \tag{16}$$

in which the normalization factor is defined as

$$N = \sqrt{\left(\nabla E'_{\mathrm{d},k-1}\nabla E'_{\mathrm{d},k}\right)^2 + \left(\nabla E'_{\mathrm{o},k-1}\nabla E'_{\mathrm{o},k}\right)^2} \tag{17}$$

and k indicates the index for the optimization step.

For conical intersection search with branching update method using QM/MM optimization methods, one has to carefully equilibrate the MM environment during the optimization projecting out the branching plane in order not to lift up the degeneracy. In QM/MM structures are optimized in the so-called microiteration approach,[21] that is, at each optimization step first the MM subsystem is optimized with a steepest-descent algorithm with the QM fragment frozen and second the QM fragment is optimized in the frozen MM environment with a quadratically convergent algorithm. This is done until both systems' gradients are less than a given threshold. For each of the two states I and J defining the conical intersection, a set of partial ESPF charges are computed for the QM atoms like $p_A^I = Z_A - Tr[P^{(I)}Q_A]$ and $p_A^J = Z_A - Tr[P^{(J)}Q_A]$ for $P^{(I)}$ and $P^{(J)}$ being the relaxed density matrices of the corresponding states. These QM charges are then used to compute the interaction with the rest of MM subsystem, leading to two different classical contributions, namely E_I^{MM} E_J^{MM}. Then, the optimization step in the MM environment optimizes the average energy $E_{IJ}^{\mathrm{MM}} = (E_I^{\mathrm{MM}} + E_J^{\mathrm{MM}})/2$ with the corresponding average gradient projecting out the branch-

ing plane $\nabla E_{IJ}^{\mathrm{MM}} = B(\nabla E_I^{\mathrm{MM}} + \nabla E_J^{\mathrm{MM}})/2$.

2.1.4 Nonadiabatic Couplings with QM/MM

The multi-state NAMD (Nonadiabatic Molecular Dynamics) in combination with QM/MM scheme was developed based on the fewest-switches surface-hopping algorithm.[22] Nonadiabatic coupling vectors are computed numerically by using a fast overlap calculation[23,24] with a sub time-step size 10^{-5} fs for the propagation of electronic degrees of freedom.[25]

2.2 Computational details

The MRSF-TDDFT/MM with ESPF coupling has been implemented in a local development version of GAMESS-US (Version 30 SEP 2021 R2 Patch 1) interfaced with the Tinker 8.10.1,[26,27] modified to implement the QM/MM interaction model of Ferré and coworkers.[21] For all calculations, the BH&HLYP/6-31G* has been used for QM calculations. Besides ESPF-based electrostatics, QM/MM interactions include van der Waals terms between pairs of atoms as long as both are not QM. When the frontier between the QM and MM subsystems involves a covalent bond, additional terms are taken into account: stretching, bonding and torsions when at least one atom is MM. The link-atom scheme is applied to the frontier QM–MM bond, using an hydrogen center whose position is constrained using the Morokuma approach.[28] The solvated thymine was modeled with a QM/MM scheme, where the QM thymine is surrounded by 659 TIP3P waters as shown in Fig. 1. A spherical boundary condition with 17 Å radius was applied for NVT ensemble, which was initially equilibrated with all MM parameters for 47 ps. The number of water molecules was determined to reproduce the density of water at ambient conditions and the temperature was controlled at 300K by the Nose-Hoover thermostat. Two sets of 50 snapshots with S_1 and S_2 bright geometries from the equilibrated trajectories were adopted for subsequent NVE QM/MM-NAMD simulations, respectively, where the MRSF and TIP3P were utilized for QM and MM regions. A desoxyribonucleic acid (DNA) decamer was taken from experimental crystallographic data (sequence, 5'-D(GCTTAATTCG)-3', PDBID: 1N4E), which is known to form thymine dimers upon UV excitation. The DNA crystal structure was soaked in a water 61 Å^3 cube, with the addition of 37 KCl ions to neutralize the system, using CHARMM GUI.[29] The model for the MRSF-TDDFT/MM calculations is depicted in Fig. 5. The system was first minimized at the MM level using Tinker 8.10.1 in Ewald PBC conditions, then equilibrated using MD in NVT ensemble (125 ps, with 1 fs timesteps) and finally a production run in NPT (5 ns, with 2 fs timesteps). Parameters were taken from CHARMM36m force field.[30–33] This structure was then used for performing the subsequent simulations. The QM/MM calculations were done for the unit cell in a vacuum, with the box of waters frozen to the equilibrated structure, and the molecules around the DNA strands in a radius of 20 Å were let to be mobile (ca. 21200 atoms).

2.3 Results and discussion

2.3.1 Photochemistry of Solvated Thymine

In the gas phase, many nucleobases display unusually long excited-state lifetimes on the order of several ps.[34–45] In particular, thymine displays the longest gas phase decay constants. The controversy over the relaxation mechanism of photoexcited thymine such as S_1-trapping, S_2-trapping, and S_1&S_2-trapping has been recently resolved by our recent NAMD study using MRSF-TDDFT.[14] The study supports the S_1-trapping mechanism with two lifetimes, τ_1=30±1 fs and τ_2=6.1±0.035 ps, quantitatively consistent with the recent time-resolved experiments.[46] Accordingly, upon photo-excitation to the S_2 ($\pi\pi^*$) state, thymine undergoes an ultrafast (*ca.* 30 fs) $S_2 \rightarrow S_1$ internal conversion through a conical intersection between S_2 and S_1 ($CI_{21,BLA}$) and resides around the minimum on the $S_{1,min}$ ($n_O\pi^*$) surface, slowly decaying to the ground state (*ca.*

6.1 ps). Therefore, it is clear that the $S_{1,min}$ is the most frequently found structure during the overall lifetime of photoexcited thymine in gas phase, accounting for its long decay constants.

Apart from the singlet models, the CASSCF NAMD simulations[47] predicted 0.9±0.1 ps for the $S_1 \rightarrow T_2$ intersystem crossing (ISC) followed by an ultrafast relaxation to T_1, which however is much shorter than the recent experimental timescale of 3.5±0.3 ps.[46] Since we focus on the investigation of solvation effects, the competition between internal conversion and ISC shall be explored in the future.

On the other hand, the excess thermal energy of the excited state in a solution can be well dissipated into the surrounding solvents, changing its entire photo-dynamics. Thymine and its derivatives show strong absorption features at $\sim$ 265 and $\sim$ 205 nm (4.68 and 6.05 eV) in water.[48–50] These bands exhibit virtually no shifts in nucleobase and nucleotide.[51] The assignment of the absorption has been considered in numerous studies on the electronic structure of thymine[52–55] that are well reviewed by Improta et al,[44] where, in isolated Thymine, the lower energy absorption band corresponds to a $\pi\pi^*$ highest occupied molecular orbital (HOMO)→lowest unoccupied molecular orbital (LUMO) transition while the optically dark $n\pi^*$ (HOMO-1)→LUMO transition lies lower in energy. In water, the $\pi\pi^*$ state is stabilized preferentially as compared to $n\pi^*$ state,[56] causing the $\pi\pi^*$ and $n\pi^*$ states to be near–degenerate. The higher lying absorption band at $\sim$ 205 nm also appears to be of $\pi\pi^*$ character based on comparison to gas phase results,[44] and recent calculations indicate that multiple $\pi\pi^*$ transitions contribute to this band in liquid water.[50]

MRSF-QM/MM calculations for four excited states were performed on the 478 snapshots from the QM/MM-MD trajectories at S_0 state, which was utilized to produce a simulated UV spectrum and the results are presented in Fig. 1, where distinct two bands of near $\sim$ 220 and $\sim$ 260 nm are observed, faithfully reproducing previous studies of two bands in aqueous conditions. (It is noted that the resulting values are red-shifted by 40 nm.)

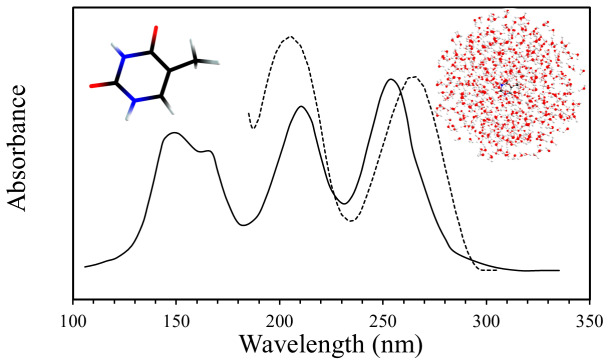


Figure 1: The simulated UV-absorption spectrum of solvated thymine in aqueous condition. MRSF-TDDFT calculations were done on 478 snapshots of ground state QM/MM-MD simulations. TIP3P was used for the MM. The spectrum is red-shifted by 40nm. The experimental spectrum in dotted line is taken from ref. 51. (Copyright permissions are obtained from John Wiley and Sons).

In the case of thymine in gas phase, the S_1 and S_2 are dark $n\pi^*$ and bright $\pi\pi^*$ states, respectively.[14] The atomic numbering and the $S_{0,min}$ geometric data by MRSF-TDDFT with BH&HLYP/6-31G(d) in gas phase are presented in the first figure of Fig. 4, where it is seen that the ground state geometries of gas phase are quantitatively consistent with the crystal values.[57] On the other hand, the relative order of bright and dark states tends to be reversed in QM/MM scheme with the bright S_1 and bright S_2 ratio of 7:3. The VEEs of bright S_1 and S_2 snapshots that are utilized for QM/MM-NAMD are shown in Table S1 and S2, respectively. The lowering bright state energy in a polar solution can be understood by its polar zwitterionic character. Therefore, contrasting to the NAMD simulation in gas phase,[14] the three state (S_0, S_1 and S_2) TSH QM/MM-NAMD simulations with MRSF/BH&HLYP were initiated in the bright S_1 state on the 30 randomly selected snapshots of ground state QM/MM-MD. This particular bright S_1 state corresponds to the lower energy bright state of the two distinct bands. The time evolution of the three state populations up to 500 fs are presented in the Fig. 2, where the corresponding NAMD results

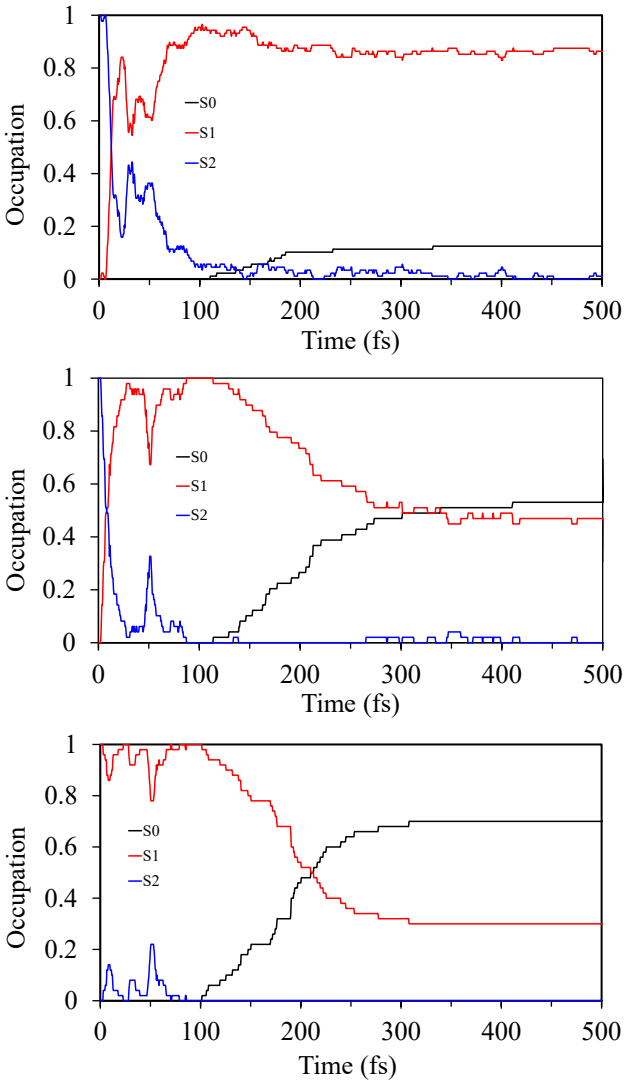


Figure 2: Time evolution of thymine populations from gas phase NAMD from initial excitation to S_2 state (top). The middle and bottom are the same population changes of solution QM/MM-NAMD. The former and the latter are initially excited to S_2 and S_1, respectively.

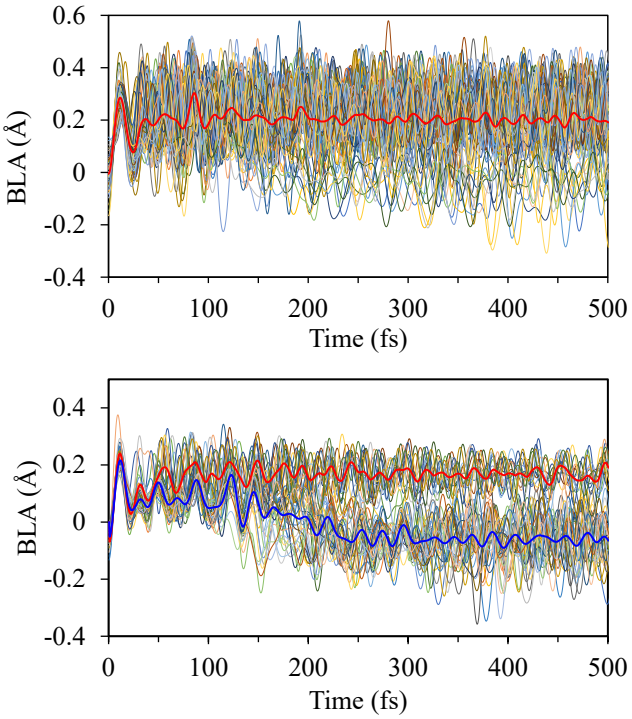


Figure 3: The BLA evolution of thymine along all NAMD trajectories for the first 500fs (top) from gas phase NAMD and (bottom) from solution QM/MM-NAMD starting from S_1 exitation. The red and blue curves in the bottom represent the averaged BLA over the two reaction channels, while the single red curve in the top is the averaged BLA over entire trajectories. The BLA values of FC, $CI_{21,BLA}$ and S_1 minimum in gas phase are 0, 0.135 and 0.215 Å, respectively.

in gas phase are also shown. The initial S_2 population is quickly transferred to S_1 in the gas phase by the characteristic BLA (bond-length alternation) vibrations in about $\sim$ 100 fs. The BLAs is defined as,

$$\mathrm{BLA} = \frac{1}{2}(\Delta R_{C_4=O_8} + \Delta R_{C_5=C_6}) - \Delta R_{C_4-C_5}, \quad (18)$$

where ΔRs are displacements with respect to the S_0 equilibrium geometry. On the other hand, the same change is not observed in the solution phase except for minor $S_1 \rightarrow S_2$ internal conversions, since the NAMD was started from S_1 state. In both cases, the ground state populations start to appear after $\sim$ 100 fs. Contrasting to the slow S_0 population increase in gas phase, an abrupt changes are seen in solution reaching 50% $\sim$ 70% of the total population in 300 fs, regardless of initial excitation of S_2 or S_1. This QM/MM-NAMD result is consistent with recent liquid jet time-resolved photoelectron spectroscopy,[58] where the lower lying $\pi\pi^*$ (S_1) state is found to decay in $\sim$ 400 fs independent of pump photon energy.

It can be expected that the ultrafast dynamics lasting $\sim$ 100 fs in solution correspond to $\pi\pi^* \rightarrow n\pi*$ internal conversion of gas phase, even though the population plot does not represent it. In order to verify it, the bond length alternations (BLAs) as defined by Eq.18 of all trajectories were calculated and the results are presented in Fig. 3 along with those of gas phase simulations taken from our previous study,[14] where ΔRs are displacements with respect to the S_0 equilibrium geometry. Consequently, the BLA values of FC, $CI_{21,BLA}$ (the conical intersection between S_2 and S_1) and S_1 minimum in gas phase are 0, 0.135 and 0.215 Å, respectively. In the of gas phase, the BLAs are oscillating in the range of 0 $\sim$ 0.4 with an average value of 0.2 (the BLA of $S_{1,min}$ structure). The BLAs below zero are relatively minor. On the other hand, characteristic two groups of BLA values are seen in the solution simulations. The first group is oscillating with a much narrower range of 0.1 $\sim$ 0.2 as compared to those of gas phase, which indicates that the amplitudes of BLA vibration are damped by the solvents right from

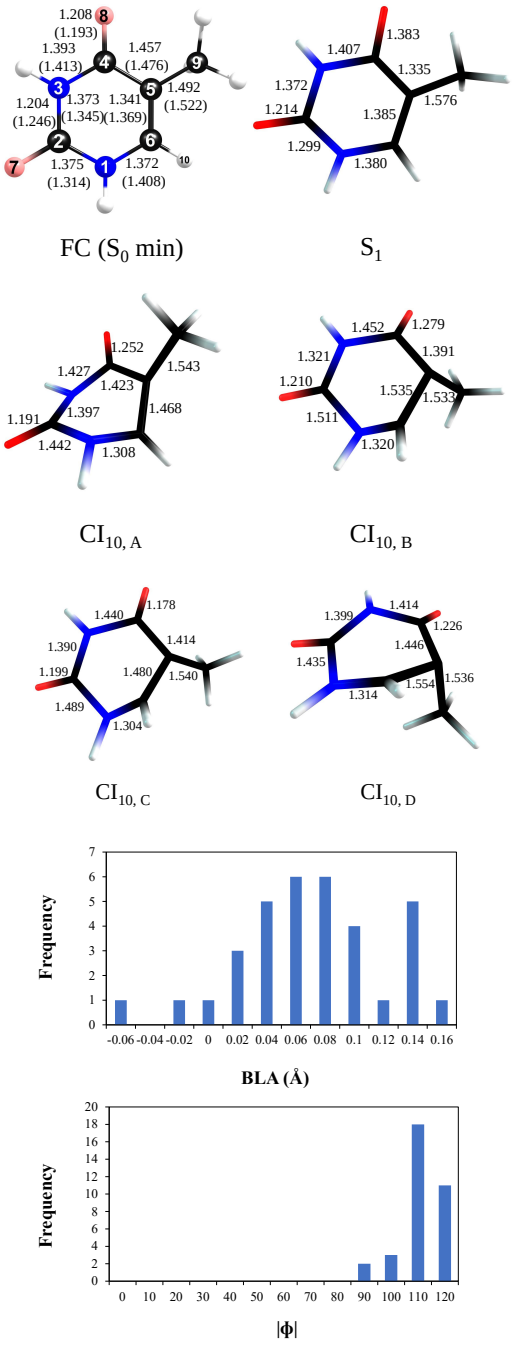


Figure 4: FC ($S_{0,min}$) is optimized geometry of ground state in gas-phase with atom numbering and crystallographic data[57] given in parenthesis, while S_1 is the last snapshot that remains in S_1 state. All CI_{10} corresponds to the structures at the hopping point. The (BLA/Å, ϕ/°) values are; S_1 (0.167, 20.53), $CI_{10,A}$ (0.055, -100.57), $CI_{10,B}$ (0.134, 110.07), $CI_{10,C}$ (0.033, 102.06), $CI_{10,D}$ (0.062, 109.41). ϕ is defined as the $H_{10}C_6C_5C_9$ dihedral angle. The histograms of BLA and ϕ of geometries at 34 hopping points are also presented.

the beginning of BLA vibration. Remarkably, a distinctive second group with a significantly larger population oscillates with the negative values of BLA, which correspond to the quick rise of the ground state populations.

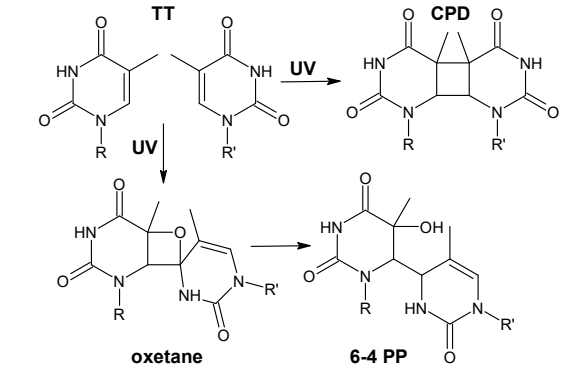
The four representative structures (designated as $CI_{10,A} \sim CI_{10,D}$) at the nonadiabatic transitions to ground states in the trajectories of solution simulations are presented in Fig. 4. In all four cases, puckered methyl groups as measured by ϕ ($H_{10}C_6C_5C_9$ dihedral angle) are larger than 100 °, which are consistent with the CI_{10} structures of isolated thymine in gas phase.[14] The histograms of BLA and ϕ are also shown on the basis of 34 structures at hopping points, where the characterisic puckering angle (ϕ) of $100 \sim 120°$ is clearly seen. In comparison, the corresponding torsion and BLA of the last snapshot that remains in S_1 state, are 20.53° and 0.167 Å, respectively, partially representing the $S_{1,min}$ structure of gas phase.

Therefore, it is clear that the solvation as shown in Fig. S1 stabilizes the CI structures facilitating $S_1 \rightarrow S_0$ nonadiabatic transitions. The dipole moments of FC, S_1, $CI_{10,A}$, $CI_{10,B}$ $CI_{10,C}$ and $CI_{10,D}$ were calculated without MM part to present the net values of QM region and the results are presented in Table 1. It is clearly seen the enhanced dipole moments of $CI_{10,A} \sim CI_{10,D}$ (6.24 ∼ 7.15 Debye) in S_1 state, which would be largely stabilized by solvation causing the quick rise of the ground state population. On the other hand, the dipole moment of the last snapshot that remains in S_1 state is a much smaller value of 3.28 Debye.

Table 1: Dipole Moments (in Debye) of Ground and Excited States at CIs, $S_{0,min}$ and S_1 Structures (the last snapshot that remained in S_1 state), where the large dipole moments of each structure are in *italic*.

Geometries	S_0	S_1	S_2
FC ($S_{0,min}$)	4.33	2.40	*5.47*
$CI_{10,A}$	3.69	*7.15*	3.64
$CI_{10,B}$	3.23	*6.54*	4.30
$CI_{10,C}$	3.56	*6.26*	1.05
$CI_{10,D}$	3.27	*6.46*	0.50
S_1 (last snapshot)	6.14	3.28	*8.03*

2.3.2 Photochemical mechanism of Thymine dimers formation in DNA



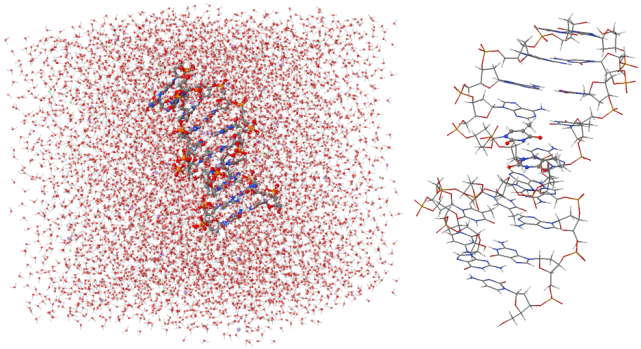


Figure 5: (top) Representation of the main mutagenic photochemical reactions of a thymine dimer (TT) after UV excitation, namely cyclobutane pyrimidine dimer (CPD) and the 6-4 photoproduct (6-4 PP) formed via an oxetane intermediate. (bottom) Depiction of the DNA dodecamer 5'-D(GCTTAATTCG)-3' used in the current study. On the left, DNA double-strand soaked in water boxed with counterions. On the right, close up of the central thymine dimer treated at the QM level.

UV light absorption of two adjacent thymines in the same strand of DNA is known to produce mutations that, unless repaired, can ultimately lead to the development of melanomas.[59] Although the mutated photoproducts are either cyclobutane pyrimidine dimers (CPD), or 6-4 (PP) configuration (See Fig. 5, top), their corresponding formation mechanisms are still under intense debate.[60] It has been suggested that the particular dark excited state with a

double excitation character is responsible for the formation of CPD.[3] However, it is not entirely clear how such an energetic dark state can be populated. On the other hand, the 6-4 PP may be formed by way of an oxetane intermediate,[61,62] without the involvement of triplet states. So far, significantly simplified models either of just a thymine pair in gas phase[3] or a small DNA fragment have been mostly adopted in the study of thymine photo-dimerization. As demonstrated in the previous section of the solvation effect, the photo-dynamics of thymine can be dramatically altered by polar media. Perhaps, the same principle can be applied to the thymine in DNA. We designed a realistic model of a dodecamer BDNA surrounded by a box of water molecules and counter ions (see Fig. 5, bottom) to sufficiently take the DNA environmental effect into account. As a consequence, the total system is composed of ca. 21200 atoms. Two central thymines are treated at the MRSF-TDDFT level, whereas the rest of the system is kept at the MM level. A sphere of 20 Å around the central thymine is active, whereas the rest of the system is kept at a fixed position. No periodic boundary conditions are considered in the simulation.

In the current study, alternative pathways of CPD and 6-4 PP formations are explored. The results of MRSF/MM calculations are summarized in Fig. 6, The first three steps of the photochemistry are commonly shared by the CPD and the 6-4 PP formation. At the Franck-Condon (FC) with S_0 minimum geometry, the S_1 ($\pi\pi^*$) and S_2 ($n\pi^*$) states are almost degenerate. The S_2 undergo a rapid internal conversion to the lower S_1 state of $n\pi^*$ character after reaching the S_1/S_2 conical intersection that is accessed in a barrierless reaction.[63] As discussed in the previous section, this ultrafast $\pi\pi^* \rightarrow n\pi^*$ population transfer completes within about 100 fs in both gas and aqueous condition leading to the metastable S_1 ($n\pi^*$) state. From the S_1 minimum, there exist several conical intersections (CI_{10}) with the ground state. There is the usual intra-monomer internal conversion (not considered in this study), competing with the formation of CPD and the oxetane intermediate, that forms back thymine S_0 in a hot vibrational state and is part of the natural protection of DNA against UV radiation. For the CPD CI_{10} intersection, it is 1.21 eV above the S_1 minimum, explaining the slow time and small quantum yield for the formation of CPD dimers. In the case of thymine monomer in gas phase, the barriers of CI_{10} are $0.25 \sim 1.53$ eV with the characteristic methyl puckered geometry (see S2 in the supporting information). Similar puckerings from the two thymine dimer in CPD CI_{10} is also seen. Interestingly, this conical intersection does not involve a doubly excited state, but rather, a delocalized $\pi\pi^*$ excitation over the two thymine units, contrasting to the one proposed in Ref. 3 In the case of oxetane intermediate formation, the conical intersection is almost degenerate to the S_1 minimum, although probably an energetic barrier needs to be crossed to reach the intersection region. In the conical intersection geometry, a covalent bond between the oxygen and the carbon is formed again with the methyl puckered conformation, with the involvement of a charge transfer between the π system of one thymine to the p_z on carbon of the other. In DNA, the 6-4 PP is less stable than the CPD, due to its larger geometric distortion, since it induces a larger steric repulsion with the rest of the DNA strand. Thus, stabilization of 6-4 PP will be strongly dependent on the conformation of DNA strand.

2.4 Conclusions

Here, we present the basic equations and implementation of an electrostatic embedding coupling (ESPF) between MRSF-TDDFT and classical MM force fields for the accurate representation of electrostatic interactions in a QM/MM scheme.

According to the MRSF-TDDFT/MM calculations on the snapshots of ground state QM/MM-MD, the bright S_2 ($\pi\pi^*$) tends to become the lowest excited state (S_1) in aqueous condition. It indicates that the expected stabilization of zwitterionic $\pi\pi^*$ by the polar solvent is well represented by our QM/MM scheme. In addition, MRSF QM/MM-NAMD setup yielded much faster thymine deactivation

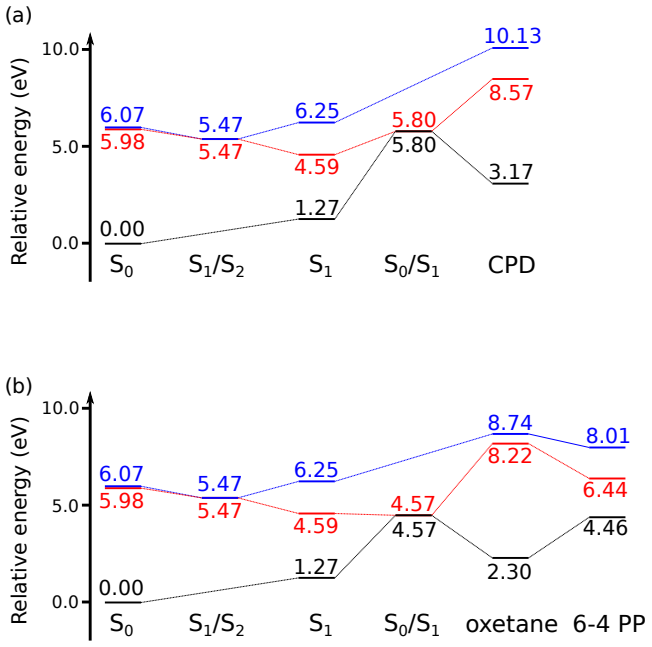


Figure 6: Schematic representation of the mechanism of (a) CPD and (b) 6-4 PP formation, depicting the ground and the first two singlet excited states. For more detailed geometries, see Fig. S2 in the supporting information.

as compared to that of gas phase, which is consistent with the experimentally found decay constant of $\sim$ 400 fs. According to our population changes, the initial absorption to $\pi\pi^*$ undergoes $\pi\pi^* \rightarrow n\pi^*$ internal conversions in 100 fs, which is followed by the $n\pi^* \rightarrow$ ground states in $\sim$ 300 fs, respectively. The structures of the latter hoppings typically have a puckered methyl group, which is the characteristic feature of S_1 and S_0 conical intersections in the case of gas phase. Significantly increased dipole moments at the puckered hopping geometries are also observed, which would imply its preferential stabilization in the more polar environment. Therefore, the accelerated deactivation of thymine in solution can be explained by the lowered barriers to conical intersections.

We have also shown that MRSF/MM can efficiently explore the potential energy surface of excited states and determine the photochemical mechanisms when combined with the branching update method for a conical intersection search. In the case of thymine dimers in BDNA, we found two formation mechanisms of cyclobutadiene thymine dimers and 6-4 photoproducts. For CPD, we found a new type of conical intersection that does not invoke the need for a double excitation state as previously stated in the literature. In the case of 6-4 PP, we confirm the previously found observation that an oxetane intermediate is formed before the 6-4 PP. In the BDNA sequence, the CPD photoproduct is more stable than the 6-4 PP, due to the larger steric repulsion of the latter.

In short, it is clearly demonstrated that the particular ESPF electrostatic embedding QM/MM in combination with MRSF-TDDFT faithfully reproduces the experimental results of the quick thymine deactivation in aqueous conditions. In the case of photo-dimerization of thymine dimer in DNA, two plausible pathways were located. These illustrate that electrostatic embedding MRSF/MM is an accurate yet practical protocol in the study of complex condensed phase photochemistry.

Acknowledgement MHR and KS acknowledge financial support by the "Agence Nationale pour la Recherche" through the project BIOMAGNET (ANR-16-CE29-0008-01). Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high performance computing resources. All authors thank the financing through the program PHC STAR 2019 granted by the "ministère de l'Europe et des Affaires étrangères" (MEAE), the "ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation" (MESRI) and the National Research Foundation of Korea (NRF). This work was supported by the Samsung Science and Technology Foundations (SSTF-BA1701-12) for the fundamental theory developments and the NRF funded by the Ministry of Science and ICT (2020R1A2C2008246 and 2020R1A5A1019141).

References

(1) Perun, S.; Sobolewski, A. L.; Domcke, W. Conical Intersections in Thymine. *J. Phys. Chem. A* **2006**, *110*, 13238–13244.

(2) Schreier, W. J.; Schrader, T. E.; Koller, F. O.; Gilch, P.; Crespo-

Hernández, C. E.; Swaminathan, V. N.; Carell, T.; Zinth, W.; Kohler, B. Thymine Dimerization in DNA Is an Ultrafast Photoreaction. *Science* **2007**, *315*, 625–629.

(3) Rauer, C.; Nogueira, J. J.; Marquetand, P.; González, L. Cyclobutane thymine photodimerization mechanism revealed by nonadiabatic molecular dynamics. *Journal of the American Chemical Society* **2016**, *138*, 15911–15916.

(4) Faraji, S.; Dreuw, A. Insights into light-driven DNA repair by photolyases: challenges and opportunities for electronic structure theory. *Photochemistry and photobiology* **2017**, *93*, 37–50.

(5) Stavros, V. G.; Verlet, J. R. Gas-phase femtosecond particle spectroscopy: a bottom-up approach to nucleotide dynamics. *Annual Review of Physical Chemistry* **2016**, *67*, 211–232.

(6) Cofer-Shabica, D. V.; Menger, M. F.; Ou, Q.; Shao, Y.; Subotnik, J. E.; Faraji, S. INAQS, a generic interface for non-adiabatic QM/MM dynamics: Design, implementation, and validation for GROMACS/Q-CHEM simulations. *arXiv preprint arXiv:2203.00225* **2022**,

(7) Lee, S.; Filatov, M.; Lee, S.; Choi, C. H. Eliminating spin-contamination of spin-flip time dependent density functional theory within linear response formalism by the use of zeroth-order mixed-reference (MR) reduced density matrix. *The Journal of chemical physics* **2018**, *149*, 104101.

(8) Lee, S.; Kim, E. E.; Nakata, H.; Lee, S.; Choi, C. H. Efficient implementations of analytic energy gradient for mixed-reference spin-flip time-dependent density functional theory (MRSF-TDDFT). *J. Chem. Phys.* **2019**, *150*, 184111.

(9) Horbatenko, Y.; Lee, S.; Filatov, M.; Choi, C. H. Performance Analysis and Optimization of Mixed-Reference Spin-Flip Time-Dependent Density Functional Theory (MRSF-TDDFT) for Vertical Excitation Energies and Singlet–Triplet Energy Gaps. *J. Phys. Chem. A* **2019**, *123*, 7991.

(10) Lee, S.; Shostak, S.; Filatov, M.; Choi, C. H. Conical Intersections in Organic Molecules: Benchmarking Mixed-Reference Spin–Flip Time-Dependent DFT (MRSF-TD-DFT) vs Spin–Flip TD-DFT. *The Journal of Physical Chemistry A* **2019**, *123*, 6455–6462.

(11) Salazar, E.; Faraji, S. Theoretical study of cyclohexadiene/hexatriene photochemical interconversion using spin-flip time-dependent density functional theory. *Molecular Physics* **2020**, *118*, e1764120.

(12) Zhang, X.; Herbert, J. M. Spin-flip, tensor equation-of-motion configuration interaction with a density-functional correction: A spin-complete method for exploring excited-state potential energy surfaces. *The Journal of Chemical Physics* **2015**, *143*, 234107.

(13) Park, W.; Shen, J.; Lee, S.; Piecuch, P.; Filatov, M.; Choi, C. H. Internal Conversion between Bright (1^1B_{u+}) and Dark (2^1A_{g-}) States in s-trans-Butadiene and s-trans-Hexatriene. *The Journal of Physical Chemistry Letters* **2021**, *12*, 9720–9729.

(14) Park, W.; Lee, S.; Huix-Rotllant, M.; Filatov, M.; Choi, C. H. Impact of the Dynamic Electron Correlation on the Unusually Long Excited-State Lifetime of Thymine. *J. Phys. Chem. Lett.* **2021**, *12*, 4339–4346.

(15) Park, W.; Shen, J.; Lee, S.; Piecuch, P.; Joo, T.; Filatov, M.; Choi, C. H. Dual Fluorescence of Octatetraene Hints at a Novel Type of Singlet-to-Singlet Thermally Activated Delayed Fluorescence Process. *J. Phys. Chem. C* **2022**,

(16) Park, W.; Filatov, M.; Sadiq, S.; Gerasimov, I.; Lee, S.; Joo, T.; Choi, C. H. A

Plausible Mechanism of Uracil Photohydration Involves an Unusual Intermediate. *J. Phys. Chem. Lett.* **2022**, *13*, 7072–7080.

(17) Schwinn, K.; Ferré, N.; Huix-Rotllant, M. Efficient Analytic Second Derivative of Electrostatic Embedding QM/MM Energy: Normal Mode Analysis of Plant Cryptochrome. *J. Chem. Theor. Comput.* **2020**, *16*, 3816–3824.

(18) Schwinn, K.; Ferré, N.; Huix-Rotllant, M. UV-visible absorption spectrum of FAD and its reduced forms embedded in a cryptochrome protein. *Phys. Chem. Chem. Phys.* **2020**, *22*, 12447–12455.

(19) Huix-Rotllant, M.; Ferré, N. Analytic Energy, Gradient, and Hessian of Electrostatic Embedding QM/MM Based on Electrostatic Potential-Fitted Atomic Charges Scaling Linearly with the MM Subsystem Size. *J. Chem. Theor. Comput.* **2021**, *17*, 538–548.

(20) Maeda, S.; Ohno, K.; Morokuma, K. Updated Branching Plane for Finding Conical Intersections without Coupling Derivative Vectors. *J. Chem. Theory Comput.* **2010**, *6*, 1538–1545.

(21) Melaccio, F.; Olivucci, M.; Lindh, R.; Ferré, N. Unique QM/MM potential energy surface exploration using microiterations. *Int. J. Quant. Chem.* **2011**, *111*, 3339–3346.

(22) Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990**, *93*, 1061.

(23) Lee, S.; Kim, E.; Lee, S.; Choi, C. H. Fast Overlap Evaluations for Nonadiabatic Molecular Dynamics Simulations: Applications to SF-TDDFT and TDDFT. *J. Chem. Theory Comput.* **2019**, *15*, 882.

(24) Lee, S.; Horbatenko, Y.; Filatov, M.; Choi, C. H. Fast and Accurate Computation of Nonadiabatic Coupling Matrix Elements Using the Truncated Leibniz Formula and Mixed-Reference Spin-Flip Time-Dependent Density Functional Theory. *J. Phys. Chem. Lett.* **2021**, *12*, 4722–4728.

(25) Mitrić, R.; Werner, U.; Bonačić-Koutecký, V. Nonadiabatic dynamics and simulation of time resolved photoelectron spectra within time-dependent density functional theory: Ultrafast photoswitching in benzylideneaniline. *J. Chem. Phys.* **2008**, *129*, 164118.

(26) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery Jr, J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(27) Ponder, J. W. TINKER: Software tools for molecular design, version 8.10.1, 2021; `http://dasher.wustl.edu/tinker/` (accessed 2022-11-25).

(28) Dapprich, S.; Komáromi, I.; Byun, K.; Morokuma, K.; J, F. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Molec. Struc.: THEOCHEM* **1999**, *461-462*, 1–21.

(29) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM.

(30) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(31) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.

(32) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaus-

sian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(33) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2016**, *14*, 71–73.

(34) Yu, H.; Sanchez-Rodriguez, J. A.; Pollum, M.; Crespo-Hernández, C. E.; Mai, S.; Marquetand, P.; González, L.; Ullrich, S. Internal conversion and intersystem crossing pathways in UV excited, isolated uracils and their implications in prebiotic chemistry. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20168–20176.

(35) Kang, H.; Lee, K. T.; Jung, B.; Ko, Y. J.; Kim, S. K. Intrinsic lifetimes of the excited state of DNA and RNA bases. *J. Am. Chem. Soc.* **2002**, *124*, 12958–12959.

(36) Ullrich, S.; Schultz, T.; Zgierski, M. Z.; Stolow, A. Electronic relaxation dynamics in DNA and RNA bases studied by time-resolved photoelectron spectroscopy. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2796–2801.

(37) McFarland, B. K.; Farrell, J. P.; Miyabe, S.; Tarantelli, F.; Aguilar, A.; Berrah, N.; Bostedt, C.; Bozek, J. D.; Bucksbaum, P. H.; Castagna, J. C.; Coffee, R. N.; Cryan, J. P.; Fang, L.; Feifel, R.; Gaffney, K. J.; Glownia, J. M.; Martinez, T. J.; Mucke, M.; Murphy, B.; Natan, A.; Osipov, T.; Petrović, V. S.; Schorb, S.; Schultz, T.; Spector, L. S.; Swiggers, M.; Tenney, I.; Wang, S.; White, J. L.; White, W.; Gühr, M. Ultrafast X-ray Auger probing of photoexcited molecular dynamics. *Nat. Commun.* **2014**, *5*, 4235.

(38) González-Vázquez, J.; González, L.; Samoylova, E.; Schultz, T. Thymine relaxation after UV irradiation: the role of tautomerization and $\pi\sigma^*$ states. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3927–3934.

(39) Gador, N.; Samoylova, E.; Smith, V. R.; Stolow, A.; Rayner, D. M.; Radloff, W.; Hertel, I. V.; Schultz, T. Electronic structure of adenine and thymine base pairs studied by femtosecond electron- ion coincidence spectroscopy. *J. Phys. Chem. A* **2007**, *111*, 11743–11749.

(40) Canuel, C.; Mons, M.; Piuzzi, F.; Tardivel, B.; Dimicoli, I.; Elhanine, M. Excited states dynamics of DNA and RNA bases: Characterization of a stepwise deactivation pathway in the gas phase. *J. Chem. Phys.* **2005**, *122*, 074316.

(41) Samoylova, E.; Schultz, T.; Hertel, I. V.; Radloff, W. Analysis of ultrafast relaxation in photoexcited DNA base pairs of adenine and thymine. *Chem. Phys.* **2008**, *347*, 376–382.

(42) Ligare, M.; Siouri, F.; Bludsky, O.; Nachtigallova, D.; de Vries, M. S. Characterizing the dark state in thymine and uracil by double resonant spectroscopy and quantum computation. *Phys. Chem. Chem. Phys.* **2015**, *17*, 24336–24341.

(43) Samoylova, E.; Lippert, H.; Ullrich, S.; Hertel, I. V.; Radloff, W.; Schultz, T. Dynamics of photoinduced processes in adenine and thymine base pairs. *J. Am. Chem. Soc.* **2005**, *127*, 1782–1786.

(44) Improta, R.; Santoro, F.; Blancafort, L. Quantum mechanical studies on the photophysics and the photochemistry of nucleic acids and nucleobases. *Chemical reviews* **2016**, *116*, 3540–3593.

(45) Wolf, T. J. A.; Myhre, R. H.; Cryan, J. P.; Coriani, S.; Squibb, R. J.; Battistoni, A.; Berrah, N.; Bostedt, C.; Bucksbaum, P.; Coslovich, G.; Feifel, R.; Gaffney, K. J.; Grilj, J.; Martinez, T. J.; Miyabe, S.; Moeller, S. P.; Mucke, M.; Natan, A.; Obaid, R.; Osipov, T.; Plekan, O.;

Wang, S.; Koch, H.; Gühr, M. Probing ultrafast $\pi\pi^*/n\pi^*$ internal conversion in organic chromophores via K-edge resonant absorption. *Nat. Commun.* **2017**, *8*, 29.

(46) Wolf, T. J. A.; Parrish, R. M.; Myhre, R. H.; Martinez, T. J.; Koch, H.; Gühr, M. Observation of Ultrafast Intersystem Crossing in Thymine by Extreme Ultraviolet Time-Resolved Photoelectron Spectroscopy. *J. Phys. Chem. A* **2019**, *123*, 6897–6903.

(47) Mai, S.; Richter, M.; Marquetand, P.; González, L. The DNA nucleobase thymine in motion - Intersystem crossing simulated with surface hopping. *Chemi. Phys.* **2017**, *482*, 9–15.

(48) Fox, J. J.; Shugar, D. Spectrophotometric studies on nucleic acid derivatives and related compounds as a function of pH: II. Natural and synthetic pyrimidine nucleosides. *Biochimica et biophysica acta* **1952**, *9*, 369–384.

(49) Wilson, W. 229. Some 2: 4-diamino-5-acylamido-6-hydroxypyrimidines. *Journal of the Chemical Society (Resumed)* **1948**, 1157–1161.

(50) Pepino, A. J.; Segarra-Martí, J.; Nenov, A.; Improta, R.; Garavelli, M. Resolving ultrafast photoinduced deactivations in water-solvated pyrimidine nucleosides. *The Journal of Physical Chemistry Letters* **2017**, *8*, 1777–1783.

(51) Voet, D.; Gratzer, W.; Cox, R.; Doty, P. Absorption spectra of nucleotides, polynucleotides, and nucleic acids in the far ultraviolet. *Biopolymers: Original Research on Biomolecules* **1963**, *1*, 193–208.

(52) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P.; Thiel, W. Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *The Journal of chemical physics* **2008**, *128*, 134110.

(53) Zechmann, G.; Barbatti, M. Photophysics and deactivation pathways of thymine. *The Journal of Physical Chemistry A* **2008**, *112*, 8273–8279.

(54) Barbatti, M.; Aquino, A. J.; Lischka, H. The UV absorption of nucleobases: semi-classical ab initio spectra simulations. *Physical Chemistry Chemical Physics* **2010**, *12*, 4959–4967.

(55) Szalay, P. G.; Watson, T.; Perera, A.; Lotrich, V. F.; Bartlett, R. J. Benchmark studies on the building blocks of DNA. 1. Superiority of coupled cluster methods in describing the excited states of nucleobases in the Franck–Condon region. *The Journal of Physical Chemistry A* **2012**, *116*, 6702–6710.

(56) Pepino, A. J.; Segarra-Martí, J.; Nenov, A.; Rivalta, I.; Improta, R.; Garavelli, M. UV-induced long-lived decays in solvated pyrimidine nucleosides resolved at the MS-CASPT2/MM level. *Physical Chemistry Chemical Physics* **2018**, *20*, 6877–6890.

(57) Ozeki, K.; Sakabe, N.; Tanaka, J. The crystal structure of thymine. *Acta Crystallogr. B: Structural Crystallography and Crystal Chemistry* **1969**, *25*, 1038–1045.

(58) Erickson, B. A.; Heim, Z. N.; Pieri, E.; Liu, E.; Martinez, T. J.; Neumark, D. M. Relaxation dynamics of hydrated thymine, thymidine, and thymidine monophosphate probed by liquid jet time-resolved photoelectron spectroscopy. *The Journal of Physical Chemistry A* **2019**, *123*, 10676–10684.

(59) Pfeifer, G. P.; Besaratinia, A. UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. *Photochem. Photobiol. Sci.* **2012**, *11*, 90–97.

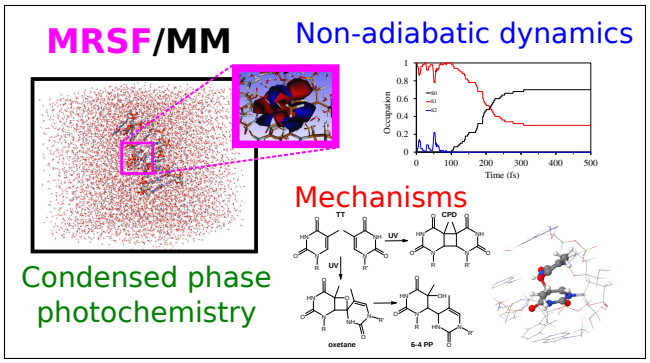
(60) Cadet, J.; Grand, A.; Douki, T. In *Photoinduced Phenomena in Nucleic Acids II: DNA Fragments and Phenomenological Aspects*; Barbatti, A. C., Marioand Borin, Ullrich, S., Eds.; 2015; pp 249–275.

(61) Labet, V.; Jorge, N.; Morell, C.; Douki, T.; Grand, A.; Cadet, J.; Eriksson, L. A. UV-induced formation of the thymine–thymine pyrimidine (6-4) pyrimidone photoproduct – a DFT study of the oxetane intermediate ring opening. *Photochem. Photobiol. Sci.* **2013**, *12*, 1509–1516.

(62) Giussani, A.; Conti, I.; Nenov, A.; Garavelli, M. Photoinduced formation mechanism of the thymine–thymine (6–4) adduct in DNA; a QM(CASPT2//CASSCF):MM(AMBER) study. *Faraday Discuss.* **2018**, *207*, 375–387.

(63) Huix-Rotllant, M. Photochemistry of Thymine in Protic Polar Nanomeric Droplets Using Electrostatic Embeding TD-DFT/MM. *Molecules* **2021**, *26*.

TOC Graphic


MRSF/MM
Non-adiabatic dynamics
Occupation
Time (fs)
Mechanisms
TT
CPD
UV
oxetane
6-4 PP
Condensed phase
photochemistry