Photophysics and Photochemistry of DNA Molecules - Electronic Excited States Leading to Thymine Dimerization

Jian-Hao Li,[†,‡] T. J. Zuehlsdorff,[¶,‡] M. C. Payne,[‡] and Nicholas D. M. Hine[§,‡]

[†]*Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

[‡]*TCM Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

[¶]*School of Natural Sciences, University of California Merced, N. Lake Road, CA 95344, USA*

[§]*Department of Physics, University of Warwick, Coventry, CV4 7AL, United Kingdom*

Abstract

We combine quantified natural transition orbital (QNTO) analysis with large-scale linear response time-dependent DFT (TDDFT) to investigate the concerted [2 + 2] thymine dimerisation reaction. This reaction is a main cause of UV-light induced damage to DNA, but its mechanism has remained poorly understood. QNTO analysis enables the electronic excitations of a molecule to be identified on the basis of their transition origins across a wide range of molecular geometries, allowing the participating excited states to be identified relatively straightforwardly. We identify a barrierless funnel that is responsible for the ultrafast reaction previously indicated in experiments. The reactive state is found to have crossings with several bright excited states, revealing how the initially populated bright states can decay rapidly to the reactive state. We also examine the contribution of environmental factors such as inclusion of the DNA backbone, which can affect the conformation of the potential energy surfaces of the relevant states.

1 INTRODUCTION

Extensive previous research addressing the singlet channel of the thymine dimerisation has focused on the potential energy surface of the first singlet excited state (S_1) of a thymine complex (denoted di-Thy).[1–5] For example, a barrierless path from the S_1 excitation of di-Thy to the S_0 state of a cis-syn thymine dimer (denoted Th<>Th) has been identified.[1–4] However, the photo-induced cycloaddition reaction must be initiated on a bright electronic excited state, whereas the S_1 state is dark: the process by which this bright state is depopulated to the dimerisable S_1 state is still unclear. Furthermore, the (changeable) character of the dimerisable S_1 requires further clarification.

The properties of energetically adjacent excited states, as well as the ground state itself, thus clearly need to be investigated in order to obtain a thorough picture of the reaction process. It is important to know how the initially populated bright state(s) proceed through

the configuration space interacting with other close-lying states with different character, so that the dimerisable S_1 state can be formed.

In the present work, we approach this problem by sampling a series of molecular structures and calculating their vertical excitation properties using TDDFT. We then use the QNTO approach[6,7] to analyse the character of multiple excited states of these structures, from which we aim to identify the relevant reaction pathway(s).

2 METHOD

All the calculations are carried out using the linear scaling (LS) DFT/TDDFT package ONETEP,[8,9] which uses an efficient representation of the density matrix in terms of in-situ optimised localised support functions known as Nonorthogonal Generalised Wannier Functions (NGWFs). Separate valence and conduction NGWF representations can be used to represent the occupied and unoccupied subspaces respectively.[10] This allows a minimal number of functions to be used while maintaining an overall accuracy equivalent to plane wave methods. In the linear response[11] (LR) TDDFT calculations, we employ the Tamm-Dancoff approximation (TDA) throughout this work. While this introduces a further approximation, it has the advantage, as well as considerably simplifying the formalism, of avoiding the singlet instability issue[12] which occurs for full LR-TDDFT when the ground state (S_0) and the first singlet excited state (S_1) are nearly degenerate. The LR-TDDFT/TDA (denoted TDDFT) results are analysed by the QNTO method[6,7] that quantifies the characters of electronic excitations and can be used to identify similar excitations across a variety of molecular systems.[7] We employ a cut-off Coulomb method[13] to exclude the effect of periodic images from neighbouring simulation cells. Norm-conserving pseudopotentials are used to remove core electrons from the calculation. The in-situ optimisation of the NGWFs[8] ensures a minimal set (1 NGWF for H and 4 for C and O atoms) is suitable for the present study. We use a cutoff of 10 bohr radii for valence NGWFs and 16 bohr for conduction NGWFs, and a 1000

eV kinetic energy cutoff for the psinc grid. These settings ensure that the DFT/TDDFT results are well-converged. The (A)LDA functional in the Perdew-Zunger parameterisation[14] for DFT/TDDFT calculations is used throughout. While a wide variety of more advanced functionals are available, we expect that (A)LDA is able to qualitatively describe low-lying excitations of small systems such as two stacked thymines. For large system calculations, we employ transition density kernels[15] confined in the region of the two thymines that are able to exclude long-ranged charge transfer contributions that are incorrectly represented within LR-TDDFT. Molecules and molecular orbitals are visualised using VMD.[16]

2.1 Construction of Initial di-Thy

We first perform a linear synchronous transit calculation followed by conjugate gradient minimisation[17,18] (LST-CG) between an ideal-B-DNA extracted di-Thy and a geometrically optimised Th<>Th to build the initial structure (Fig. 1) for subsequent constrained optimisation scans. Two main coordinates will be used in the constrained optimisations: the distances between the two C5 carbon atoms on the two thymine units (C5 and C5'), and the angle between the C5'-C5 and C5-C6 bonds: these are proxies for the separation and alignment of the two rings.

These optimisation scans for the C5-C5' distance and the C5'-C5-C6 angle close to the LST-CG structure (Table S1) show a range of structures, from those resembling Th<>Th, as reflected by a shortened C6-C6' bond, to those resembling the thymine complex. Furthermore, we note that the two groups of molecular structures have distinct ground state characters as verified by QNTO analysis. Therefore, there must be a crossing of S_0, the ground state, and S_1, the first singlet excited state, between the two groups of structures.

2.2 Ground State Optimisation Scans of di-Thy

Our next target is to find the specific excited state (or states) of di-Thy that are reaction candidates leading to the Th<>Th structure. We first perform finer ground state optimisation

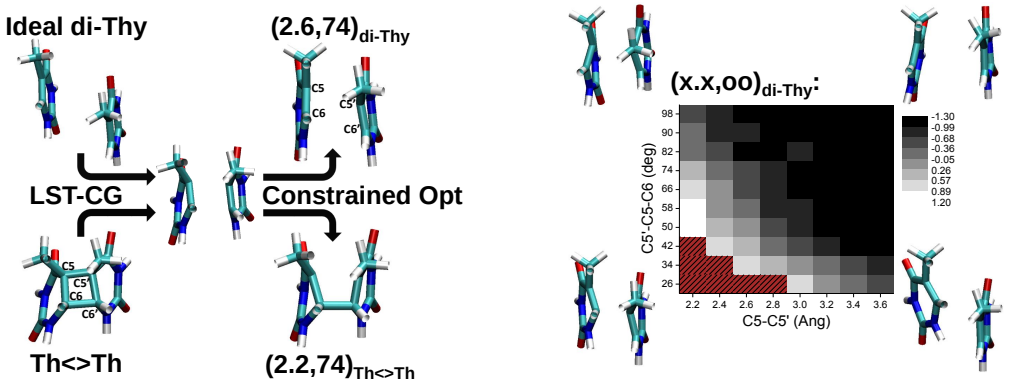

Figure 1: Flow chart (left) of calculations and the energy map (right) for the di-Thy ground state optimisation scans. We first obtain an intermediate structure from an LST-CG calculation between an ideal-B-DNA extracted di-Thy and a geometrically optimised cis-syn thymine dimer. Constrained ground state optimisation scans of C5-C5' and C5'-C5-C6 for the intermediate structure (Table S1) then generate structures tending towards Th<>Th and di-Thy, respectively. Next, finer ground state optimisation scans are followed, starting from $(2.6, 74)_{\mathrm{di-Thy}}$, namely, di-Thy with C5-C5'=2.6 Å and C5'-C5-C6=74°. These are used to construct a di-Thy ground state energy map. The colour of each cell denotes its energy relative to the energy of the $(2.2, 74)_{\mathrm{Th<>Th}}$ structure. Some cells at the left-hand corner of the map are left blank (shaded red region) due to their highly distorted structures. For more discussions see the text.

scans in the di-Thy region, with C5-C5'=2.2−3.6 Å in steps of 0.2 Å and C5'-C5-C6=26°−98° in steps of 8°. The ground state energies of these scanned structures are shown in Fig. 1. Some parameter sets of di-Thy at the left-hand corner of the map are left blank because the geometries are considered to be highly distorted, reflected also by the relatively large ground state energies at this border. As expected, a repulsive force is built up pushing the two thymines apart, as indicated by the higher ground state energies for those structures with shorter C5-C5' distances. A smaller C5'-C5-C6 angle also leads to a higher ground state energy, reflected by the more distorted base planes that can be seen in the structure to the bottom-left of the map.

2.3 Identification of Similar-character States

For all structures in the map, we calculate the first 8 electronic excitations and compare their transition origins of the first natural transition orbital pair (NTO1) for identifying excitations

with a similar character. In order to examine the effect of chemical environment, i.e. DNA backbone, in the later part of this work, we define the core region[7] of di-Thy by excluding the two hydrogen atoms (H1 atom of each thymine) such that the core region is identical to the ones for larger systems containing the DNA backbone. We then renormalise the core sections of the hole and electron orbitals of NTO1 (NTO1-H and NTO1-E, respectively), and focus on those excitations which have a $\geq 50\%$ density contribution from the core region. This condition may be expressed as

$$|\langle \mathrm{sc}|\mathrm{s}\rangle| \geq 1/\sqrt{2} \tag{1}$$

for both the NTO1-H and NTO1-E, where $|\mathrm{s}\rangle$ and $|\mathrm{sc}\rangle$ denote the NTO and the renormalised core section of NTO (cNTO), respectively.[7]

A specific di-Thy structure, C5-C5'=2.6 Å and C5'-C5-C6=74° (denoted as $(2.6, 74)_{\text{di-Thy}}$), which should lie reasonably close to the S_0/S_1 crossing (see Table S1), is selected as a standard. The excitations of $(2.6, 74)_{\text{di-Thy}}$ that satisfy Eq. 1 (the first 5 excitations, plus the 7th and 8th excitations) are referred to as standard excitations. Each of these standard excitations is given an identification label which is used for the purpose of consistently labelling similar excitations across the whole di-Thy map shown in Fig. 1.

Note that the ranges covered in the C5-C5' and C5'-C5-C6 scans are quite wide, such that the NTO1 transition origins of some geometries can have little overlap with any of the excitations for the $(2.6, 74)_{\text{di-Thy}}$ geometry used as a standard. This could be due to: (1) evolution of transition origins along the molecular coordinates, (2) mixture of different transition origins, and (3) significant difference of molecular geometries in comparison to the standard geometry. Hence, in such cases, we identify the transition origins of a targeted excited state by comparing to other intermediate structures. This propagation method was also employed in ref 7. Note, however, that in the present case the situation is more complicated as the system size is larger and there are two, instead of one, dimensions of the

scanned molecular coordinates. For robustness we also employ a self-consistent scheme in the present work that is detailed in the Supporting Information.

3 RESULTS AND DISCUSSION

3.1 Candidate Reactive States

The self-consistently identified similar excitations of all the structures in the di-Thy map (Fig. 1) are plotted in Fig. 2. There are three states which remain bright across the map, which we call b1 (marked by a rhombus), b2 (star), and b3 (plus+circle), respectively. Their oscillator strengths are shown in the corresponding panel below the potential energy profiles. From the plot of the NTO1-H and NTO1-E orbitals in Fig. 3 for the three excitations, it can be seen that they all have a $^1(\pi\pi^*)$ feature. Note also that b2 and b3 are defined on $(2.8, 42)_{\text{di-Thy}}$ instead of $(2.6, 74)_{\text{di-Thy}}$ because they cannot be derived, and are thus distinct from any of the standard excitations of $(2.6, 74)_{\text{di-Thy}}$. It can be seen that the first eight excitations of $(2.6, 74)_{\text{di-Thy}}$ respectively correspond to the 5th, 8th, N/A, 2nd, 3rd, N/A, 1st, N/A excitation of $(2.8, 42)_{\text{di-Thy}}$, where N/A denotes that no matching standard excitation of $(2.6, 74)_{\text{di-Thy}}$ is found. Similarly, the 6th (b2) and 7th (b3) excitation of $(2.8, 42)_{\text{di-Thy}}$ have no corresponding excitation in the first 8 excitations of $(2.6, 74)_{\text{di-Thy}}$. Therefore, we add b2 and b3 to the list of standard excitations in addition to those defined on $(2.6, 74)_{\text{di-Thy}}$ in order to include all possible (low-lying) excitations that could participate in the thymine dimerisation reaction. The substantial reordering of excitations suggests that $(2.6, 74)_{\text{di-Thy}}$ and $(2.8, 42)_{\text{di-Thy}}$ are considerably far apart in the configuration space. Moreover, it can be seen that the further away a structure moves from the standard, $(2.6, 74)_{\text{di-Thy}}$, or $(2.8, 42)_{\text{di-Thy}}$ for b2 and b3, the more likely it is that multiple (mostly two) excitations are derived to have the same labels. This suggests that as the structure evolves, some states have strong couplings with each other and their transition origins mix strongly such that each of them

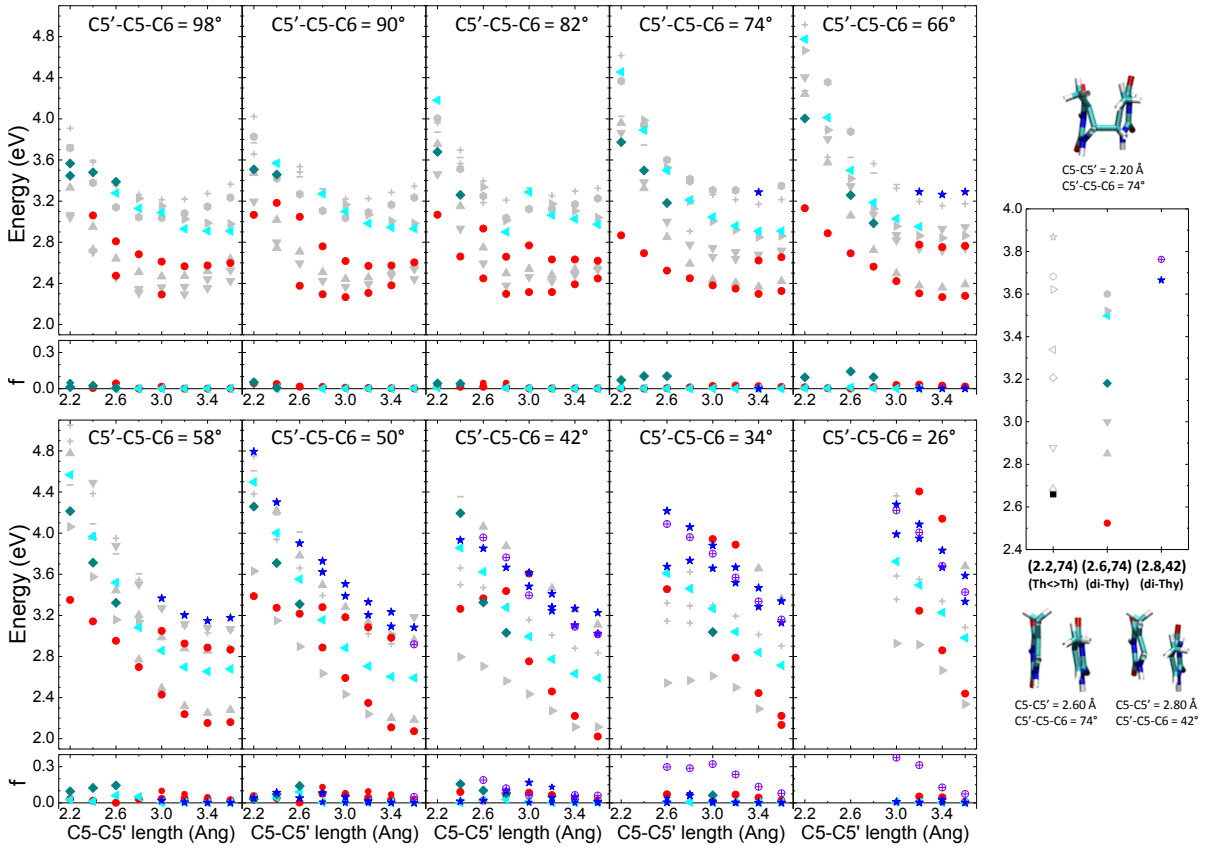

Figure 2: Potential energy profiles and oscillator strengths for the first 8 excitations of the map of di-Thy structures shown in Fig. 1. Similar-character excitations are marked by the same label defined at $(2.6, 74)_{\text{di-Thy}}$ or $(2.8, 42)_{\text{di-Thy}}$. The labels for $(2.2, 74)_{\text{Th<>Th}}$ are used in other figures for linear interpolation/TDDFT calculations. The five types of excitation: d1(●), d2(◀), b1(◆), b2(★), b3(⊕) are the focus of the present work. Different sizes of label are used in the oscillator strength panel when there are multiple excitations that are assigned the same label (the smaller size the label, the higher the excitation rank). Crosses (+) denote 'no label' for those excitations that have a character distinct from any of the standard excitations. Dashes (-) denote those excitations that do not satisfy Eq. 1 for both NTO1-H and NTO1-E. Because of the evolution and mixing of transition origins as the molecular structure changes, there can be multiple excitations assigned to the same label whereas some others no label. See the text for further discussion.

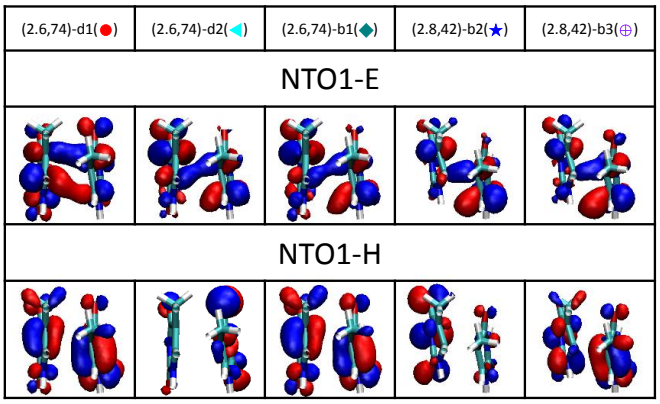

Figure 3: The NTO1-H and NTO1-E of the five types of standard excitation. Isovalue=0.04 e/bohr^3. The d1 and d2 (b1, b2, and b3) are in general dark (bright) excitations for the map of structures in Fig. 1. From the orbital shapes of these NTOs, it can be seen that d1, b1, b2, and b3 have a $^1(\pi\pi^*)$ feature, whereas d2 has a $^1(n\pi^*)$ feature.

evolves eventually to have a substantial component of the same standard excitation. Note also that the oscillator strengths of the three bright states vary strongly as the structure changes.

Crucially, the state we call d1 (filled circle), which is generally a dark state, can be seen to be a leading candidate for providing a route to dimerisation, as will be detailed in later discussions. In spite of the fact that the d1-state is generally dark, it has a $^1(\pi\pi^*)$ feature as can be seen in Fig. 3. Hence, it shows that a dark excitation is not necessarily associated with a $^1(n\pi^*)$ feature, and the oscillator strength of the $^1(\pi\pi^*)$ state can vary substantially due to the geometry of molecules. It is therefore of paramount importance to examine the transition origins of excitations alongside the excitation energy and oscillator strength before a specific excited state can be correctly identified. We also find that a generally dark state which we call d2 (left-pointed triangle) plays a role in fostering the [2 + 2] thymine dimerisation reaction. The plot of NTO1-H and NTO1-E orbitals for the b2-state (Fig. 3) shows that it has a $^1(\pi\pi^*)$ feature. Thus, the five types of excitations, d1, d2, b1, b2, and b3, will be our primary focus in the following discussion; other types of excitation are simply marked in light grey.

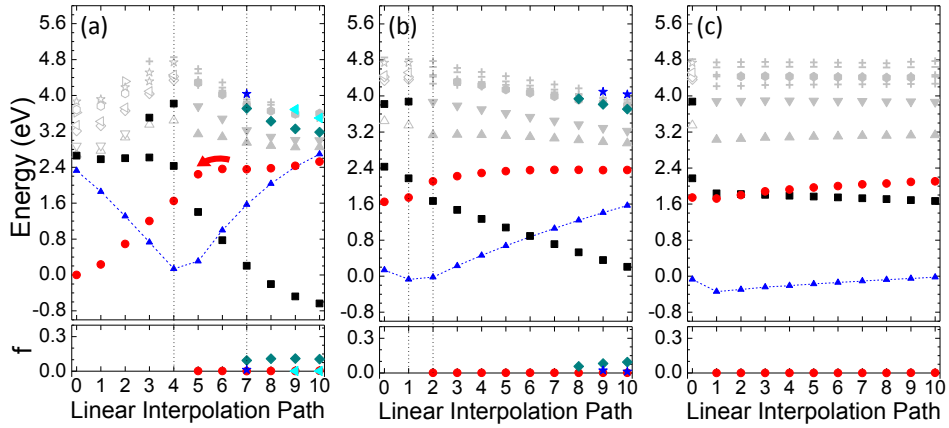

Figure 4: (a) Linear interpolation between $(2.2, 74)_{\text{Th<>Th}}$ and $(2.6, 74)_{\text{di-Thy}}$ with each LI-structure being calculated by TDDFT. (b) LI/TDDFT calculations performed between LIP=4 and LIP=7 of Fig. 4a. (c) LI/TDDFT calculations performed between LIP=1 and LIP=2 of Fig. 4b. Molecular structures corresponding to the two end-points of the linear interpolation are displayed above the potential energy diagrams. Triangles linked by dash lines denote HOMO-LUMO gap. A d1-funnel, as indicated by the red arrow in (a), can be seen linking S_1 (red filled circles) of the di-Thy to S_0 of the Th<>Th. Similarly, the S_0 of di-Thy (black squares) is linked to S_1 of Th<>Th. More discussions are given in Sec. SIII.

3.2 Dimerisable State

We now define the structure which has C5-C5'=2.2 Å and C5'-C5-C6=74° (denoted as $(2.2, 74)_{\text{Th<>Th}}$) as the standard for the Th<>Th structure. $(2.2, 74)_{\text{Th<>Th}}$ is expected to lie reasonably close to the S_0/S_1 crossing (see Table S1). We then perform a linear interpolation (LI) with 10 steps between the $(2.2, 74)_{\text{Th<>Th}}$ and $(2.6, 74)_{\text{di-Thy}}$ geometries in order to identify the exact location of the S_0/S_1 crossing between them. TDDFT calculations are then carried out for these LI-derived conformations for the purpose of identifying the reaction pathway(s). Since we have already mapped the space of di-Thy structures (Fig. 1) using excitation labels that are determined self-consistently (Fig. 2), we can now derive corresponding excitation labels for those LI conformations with a di-Thy ground state character by simply referencing to all the structures in the di-Thy map. For those LI structures with a Th<>Th ground state character, on the other hand, the excitation labels are determined self-consistently among themselves. The results are plotted in Fig. 4a. It can be seen that the first singlet excited state (S_1) with d1-character decreases monotonically in energy towards

the Th<>Th geometry. We discuss this in more detail in Sec. SIII demonstrating that the S_1 states at LI path (LIP)$=5-10$, which all have the d1-character, evolve to become the ground state, S_0, of the Th<>Th.

In addition to predicting a crossing between the S_0 and S_1, the current results also demonstrate that it is a d1-character state of di-Thy, which can be traced to different excitation order for different di-Thy conformations, that evolves to become the ground state of Th<>Th. Thus, the d1-state forms a barrierless funnel proceeding from a thymine complex to the dimerised configuration. A barrierless path on S_1 of a d1-state character leading to the thymine dimer has also been predicted by CASSCF/CASPT2 calculations.[19] This can in fact be understood by considering the bonding indicated by the form of the excitations in Fig. 3. The NTO1-E of the d1-state clearly suggests the formation of C5-C5' and C6-C6' bonds as the electron density intensifies between them. It is not hard to imagine that the consequent rearrangement of electrons promotes the formation of C5-C5' and C6-C6' bonds of Th<>Th as the electronic structure of di-Thy proceeds through the d1-funnel.[19,20] Note also that the d1-state is not always the first singlet excited state S_1. Therefore, the S_1 state of some structures has a distinct character from d1 and cannot result in the dimerisation reaction. In the following we will examine how the d1-state can be populated and under what conditions the barrierless funnel is present.

3.3 Bright States that Cross with the Dimerisable State

It can be seen in Fig. 2 that the energies of the excitations of most similar character vary strongly with the scanning coordinates, and these states cross or become nearly degenerate with other states in several places. In particular, a b1-state can be seen to cross a d1-state at a structure where the b1-state is bright. The d1-state in question is slightly lower in energy (< 0.3 eV) than the b1-state at $(2.6, 50)_{\text{di-Thy}}$ and slightly higher than the b1-state at $(2.6, 42)_{\text{di-Thy}}$. Similarly, a crossing between a d1 and a b2 or b3 occurs in a structure where

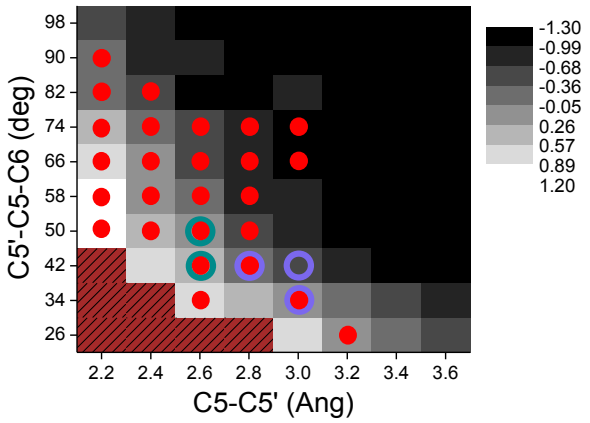

Figure 5: Ground state optimisation scans map for di-Thy with the positions of d1-funnel and several crossings indicated. The cells marked by red filled circles have an underlying barrierless d1-funnel (see main text) linking di-Thy to Th<>Th. Two types of crossing are marked, one between a b1- and a d1-state (green circle), the other between a d1- and a b2- or b3-state (blue circle).

the b2 or b3 has a substantial absorption. The d1 in question is slightly lower (higher) in energy than the b2 or b3 at $(2.8, 42)_{\text{di-Thy}}$ ($(3.0, 42)_{\text{di-Thy}}$ and $(3.0, 34)_{\text{di-Thy}}$). These structures are marked by circles in Fig. 5. Moreover, despite the fact that the d2-state is dark in many geometries, it has a substantial oscillator strength at $(2.6, 50)_{\text{di-Thy}}$ and a crossing of the d2 and d1 can be found, for example, between $(2.6, 50)_{\text{di-Thy}}$ and $(2.8, 42)_{\text{di-Thy}}$. Thus, we have identified three types of crossing that can funnel the energy of a bright state to a d1-character dark state via internal conversion, at structures for which the d1-state is not yet the lowest energy singlet S_1.

3.4 Identification of Ultrafast Dimerisable Structures

To investigate further whether a reaction can proceed after the system crosses to the d1-state, we can investigate other parts of the phase space of possible configurations. We perform LI calculations between $(2.2, 74)_{\text{Th<>Th}}$ and the structures discussed above. The results for $(2.6, 50)_{\text{di-Thy}}$ and $(2.8, 42)_{\text{di-Thy}}$ are plotted in Fig. 6. It can be seen again that the d1-

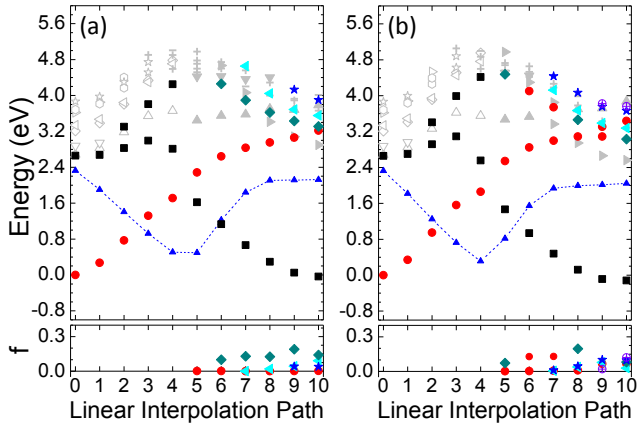

Figure 6: TDDFT calculations at LI geometries between $(2.2, 74)_{\text{Th<>Th}}$ and (a) $(2.6, 50)_{\text{di-Thy}}$, (b) $(2.8, 42)_{\text{di-Thy}}$. A crossing between d1-state and a bright b1-state occurs near the former case whereas between d1-state and bright b2- or b3-state in the latter case. Molecular structures corresponding to the linear interpolation end-points are plotted above the potential energy diagrams. As can be seen in both diagrams, the crossing links to a barrierless d1-funnel (red filled circles), facilitating the ultrafast decay from the bright states to the ground state of thymine dimer.

funnel is present near these crossings. We further explore the whole di-Thy map of Fig. 1, performing similar LI/TDDFT calculations to identify the d1-funnel. The results are shown in Fig. 5, where the cells marked by red filled circles indicate that a barrierless d1-funnel is present. Thus, once the system is de-excited to the d1-state at these structures, either via the three types of crossing discussed above or other internal conversion mechanisms, it will readily result in the [2 + 2] cycloaddition reaction. This is consistent with the finding that the ground state configuration of the thymine pair determines if the dimerisation can occur.[21,22]

Our results also indicate that the reactive d1-state has higher energies than other excitations at some structures, corroborating the CASSCF/CASPT2 results[3] that indicate dimerisation requires appropriate conformations and environments for the reactive state to have a low enough energy to favour its population. In ref 22, Law *et al.*, introduced a 'dimerisable' condition by studying the results of molecular dynamics simulations for thymidylyl-(3'-5')-thymidine molecule in aqueous cosolvent mixtures. This condition can be expressed as d < 3.63 Å and $|\eta| < 48.2°$, where d is the distance between the midpoints of the C5-C6 and

C5'-C6' double bonds and η is the C5-C6-C6'-C5' dihedral angle. We note that except for $(3.2, 26)_{\text{di-Thy}}$, all the conformations where the d1-funnel is present satisfy the 'dimerisable' condition.

Furthermore, we have demonstrated that, of all the configurations studied (Fig. 5), the d1-state is generally dark at the configurations in which the d1-funnel is present. Thus, it would require internal conversion to populate the d1-state from one of the bright state(s). The 'peaked' topology[23] of the S_0/S_1 crossing for the d1-funnel suggests that the system would also be able to evolve back to the di-Thy ground state at the crossing. It is significant that the regions where the d1-funnel is present are regions of relatively high ground state energy. This must play a role in suppressing the reaction and preventing it from occurring with excessive frequency, hence the low quantum yield ($\sim 3\%$) of the reaction.

3.5 Thymine Dimerisation in DNA Environment

3.5.1 dTpdT

Now we examine the environmental effects of DNA on the thymine dimerisation reaction. We first look at a system, dTpdT, that is formed by a thymine pair linked by the deoxyribose and phosphate group of the DNA backbone. In particular, the presence of the backbone can result in conformational constraints that make some structures highly unfavourable. The system is neutralised by adding a proton next to the phosphate group; the neutralisation widens the band gap as the HOMO orbitals located at the phosphate are stabilised by the proton next to it.

Note that in B-DNA the phosphate groups are stabilised by metal ions such as Mg(II),[24] which is partially modelled here by the proton atom next to the phosphate group. The stabilisation of the phosphate group not only widens the fundamental band gap but also improves the energy underestimation of the charge transfer (CT) transitions which would occur for TDDFT with (semi-)local density functionals such as LDA.[25] Nevertheless, the correct positions of the CT states remain an open question: the use of different HF-exchange

mixture can result in the backbone-to-base CT transitions having significantly different energies.[26] Here, we employ a scheme of confining the transition density kernel[15] in the TDDFT calculations for dTpdT in order to suppress backbone-to-base charge transfer states.

First, we perform ground state optimisations with constrained values of C5-C5' distance and C5'-C5-C6 angle for several dTpdT structures, whose ground state energies are plotted in Fig. 7. We then calculate the first 8 excitations of these structures (Fig. S3) with the transition density kernel, $K^{\alpha\beta}_{\{1\}I}$, being built in the localised basis functions centered on the di-Thy core atoms. Thus, $\langle \mathrm{sc}|\mathrm{s}\rangle$ for both the NTO1-H and NTO1-E of an excitation is consequently equal to 1. This can help filter out the energetically underestimated CT transitions.

We examine the effect of using a localised transition density kernel on the calculated excitation properties in Sec. SIV, where we also show that the presence of backbone is able to substantially change the excitation energies and oscillator strengths of several excitations. However, by examining the dTpdT structures for which the C5-C5' and C5'-C5-C6 values lie close to the region where the two types of crossing of di-Thy are located (Fig. 5), we are still able to identify the same types of crossing as in the case of isolated thymine pair, with the positions being slightly shifted in the map (Fig. 7). As can be seen in Fig. S3, the crossing of d1 and b1 move to structures lying between $(2.8, 50)_{\mathrm{dTpdT}}$ and $(2.8, 42)_{\mathrm{dTpdT}}$. For the crossing of d1 and b2 or b3, on the other hand, the structure(s) where the former has slightly lower (higher) energy than the latter move to $(3.0, 34)_{\mathrm{dTpdT}}$ ($(3.2, 34)_{\mathrm{dTpdT}}$ and $(3.2, 26)_{\mathrm{dTpdT}}$). We also see that the d2-state, which is mostly dark in the di-Thy map (Fig. 2), shows substantial absorption at several dTpdT structures (see Fig. S3). Moreover, a crossing of a d1- and d2-state can be seen to occur when the d2-state has a significant absorption: A d1-state is slightly lower (higher) in energy than a d2-state at $(2.8, 50)_{\mathrm{dTpdT}}$ ($(2.8, 34)_{\mathrm{dTpdT}}$ and $(3.0, 42)_{\mathrm{dTpdT}}$).

We also perform LI calculations between $(2.2, 74)_{\mathrm{T<p>T}}$, where T<p>T denotes cis-syn dimerised dTpdT, and each of the structures mentioned above. The excitations of those

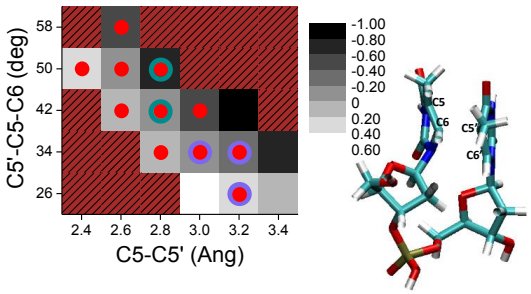

Figure 7: Ground state optimisation scans for dTpdT with C5-C5' and C5'-C5-C6 set close to the region where the two types of crossing of di-Thy are located in Fig. 5. The colour of each cell denotes its ground state energy relative to that of $(2.2, 74)_{\mathrm{dTpdT}}$. The cells marked by green (blue) circles denote the structures at which a crossing between the d1-funnel and a b1- (b2- or b3-)state is located, e.g. $(2.8, 42)_{\mathrm{dTpdT}}$, whose molecular structure is plotted to the right of the map. The d1-funnel is present at those structures indicated by a red filled circle.

structures with an undimerised ground state character are characterised by directly referencing the di-Thy structures of Fig. 5. The excitations of those with a T<p>T ground state character are self-consistently characterised among themselves. The results for $(2.8, 50)_{\mathrm{dTpdT}}$ and $(3.0, 34)_{\mathrm{dTpdT}}$ are plotted in Fig. S5. Again, the d1-funnel can be clearly seen, as the energy of the d1-state decays monotonically towards the T<p>T configuration.

Thus, the bright states discussed above all provide a pathway for the [2 + 2] dimerisation reaction when the system absorbs light to populate the bright state(s), decays to the d1-state via crossing(s), and proceeds through the d1-funnel. The widely present d1-funnel can gather energy from any states that happen to have significant absorption due to geometrical or environmental factors and have nearby crossing(s) with the d1-state.

3.5.2 TTTT/AAAA

We have also performed TDDFT calculations for two thymines embedded within two shortened complementary B-DNA strands to study the base stacking and base pairing effects on the electronic excitations of a thymine pair. The two strands are in the ideal B-DNA geometry and are neutralised by adding protons adjacent to the phosphate groups as in the

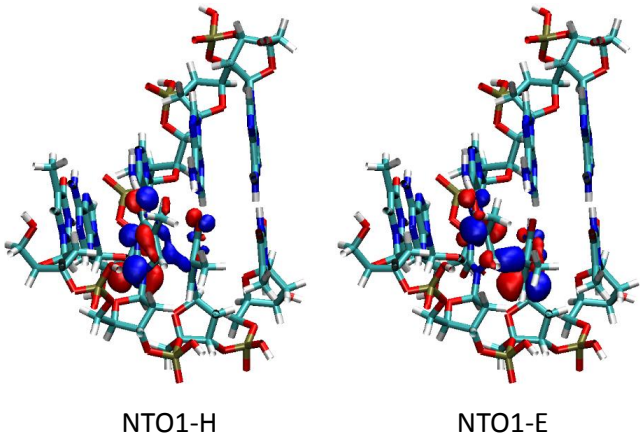

Figure 8: The NTO1-H and NTO1-E of an electronic excitation of two shortened complementary B-DNA strands with TTTT/AAAA sequence. The transition density kernel for the TDDFT calculation has been locally confined to the middle two thymine bases. By referencing the excitations of all the di-Thy map structures in Fig. 7, the excitation is characterised as a b1-state. The similarity of transition origins between this excitation and the standard b1-state (Fig. 2) can clearly be seen by comparing the plots of their NTO1-H and NTO1-E orbitals, although noticeable differences between them also exist due to the environmental and geometrical factors for the embedded thymine pair, which in turn have altered the orbital shapes of the NTO1 electron-hole pair.

previous dTpdT study. The middle two thymines are then set to C5-C5'=2.8 Å and C5'-C5-C6=42° and, along with the local backbone, are locally optimised by fixing the positions of heavy atoms of the other DNA backbone and nucleobases. The system here is used as an example to demonstrate that the TDDFT/QNTO can be employed to investigate much larger molecular systems.

The transition density kernel for the TDDFT calculation of this system is confined locally to the two central thymine bases as in the dTpdT study. The NTO1-H and NTO1-E orbitals of the 2nd electronic excitation calculated by TDDFT are plotted in Fig. 8. The NTO1 transition vector component,[7] $\sqrt{\lambda_1}$, is 0.99, showing that the excitation is dominated by the NTO1 electron-hole pair. Thus, we have been able to use the present TDDFT/QNTO method to analyse local excitations of a very large molecular system such as the two shortened DNA strands here. Because of the subtle balance of forces in realistic DNA strands, we leave a more sophisticated environmental setting of the DNA systems, e.g. including the effects of

solvent, counterions and geometry optimisation, to future work.

4 CONCLUSIONS

We have applied the TDDFT/QNTO approach to study the concerted [2 + 2] thymine dimerisation reaction, which is the most common type of UV-light induced damage to DNA, and thus a leading cause of skin cancer. We are able to show that a barrierless funnel can be identified that is responsible for the ultrafast reaction previously indicated in experiments. The reactive state is also found to have crossings with several bright excited states, revealing how the initially populated bright states can decay rapidly to the reactive state.

We have also studied a system of two thymines linked by the DNA backbone to look into the effect of the DNA environment in mediating the properties of electronic excitations of the two thymines. Although the presence of backbone is able to substantially change the excitation energies and oscillator strengths of several excitations, we are still able to identify the d1-funnel and observe bright states that cross with the d1-state. This suggests that the ultrafast dimerisation can occur in a similar scenario as in the case of two pure thymines as discussed above.

We have also shown that a very large system of two thymines embedded within two shortened complementary DNA strands can also be studied using the QNTO method and local excitations on the two thymines can be compared with those of di-Thy. Thus, it can be anticipated that in future we will be able to investigate, based on the approaches presented in this work, various interesting large and/or complex biomolecular systems for their photophysical/photochemical properties involving multiple electronic excited states.

ASSOCIATED CONTENT

The Supporting Information is available free of charge on the ACS Publications website at DOI:

Supporting information including ground state optimisation scans of C5-C5' and C5'-C5-C6 for a thymine pair near the LST-CG structure (Fig. 1), procedures of self-consistent labelling for excitations across a variety of structures, demonstration of d1-state evolving to the S_0 of Th<>Th, effect of using a localised transition density kernel, and LI/TDDFT calculations between T<p>T and dTpdT. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

AUTHOR INFORMATION

E-mail: jian-hao.li@epfl.ch; N.D.M.Hine@warwick.ac.uk. Tel: +4424 7657 4415. Fax: +4424 7652 3376.

ACKNOWLEDGEMENTS

All calculations in this work were carried out using the Cambridge HPC Service under EPSRC Grant EP/J017639/1. JHL acknowledges the support of Taiwan Cambridge Scholarship. TJZ acknowledges the support of EPSRC Grant EP/J017639/1 and the ARCHER eCSE programme. NDMH acknowledges the support of the Winton Programme for the Physics of Sustainability, and EPSRC Grant EP/P02209X/1.

REFERENCES

(1) Barbatti, M. Computational Reference Data for the Photochemistry of Cyclobutane Pyrimidine Dimers. *ChemPhysChem* **2014**, *15*, 3342–3354.

(2) Boggio-Pasqua, M.; Groenhof, G.; Schäfer, L. V.; Grubmüller, H.; Robb, M. A. Ultrafast Deactivation Channel for Thymine Dimerization. *Journal of the American Chemical Society* **2007**, *129*, 10996–10997.

(3) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics* **2007**, *79*, 291–352.

(4) González-Ramírez, I.; Roca-Sanjuán, D.; Climent, T.; Serrano-Pérez, J. J.; Merchán, M.; Serrano-Andrés, L. On the photoproduction of DNA/RNA cyclobutane pyrimidine dimers. *Theoretical Chemistry Accounts* **2011**, *128*, 705–711.

(5) Zhao, H.; Liu, K.; Song, D.; Su, H. Physical quenching in competition with the formation of cyclobutane pyrimidine dimers in DNA photolesion. *The Journal of Physical Chemistry. A* **2014**, *118*, 9105–9112.

(6) Li, J.-H.; Chai, J.-D.; Guo, G.-Y.; Hayashi, M. The quantified NTO analysis for the electronic excitations of molecular many-body systems. *Chemical Physics Letters* **2011**, *514*, 362–367.

(7) Li, J.-H.; Zuehlsdorff, T. J.; Payne, M. C.; Hine, N. D. M. Identifying and tracing potential energy surfaces of electronic excitations with specific character via their transition origins: application to oxirane. *Physical Chemistry Chemical Physics* **2015**, *17*, 12065–12079.

(8) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *The Journal of Chemical Physics* **2005**, *122*, 084119.

(9) Zuehlsdorff, T. J.; Hine, N. D. M.; Spencer, J. S.; Harrison, N. M.; Riley, D. J.; Haynes, P. D. Linear-scaling time-dependent density-functional theory in the linear response formalism. *The Journal of Chemical Physics* **2013**, *139*, 064104.

(10) Ratcliff, L. E.; Hine, N. D. M.; Haynes, P. D. Calculating optical absorption spectra for large systems using linear-scaling density functional theory. *Physical Review B* **2011**, *84*, 165131.

(11) Casida, M. E. In *Recent Advances in Density Functional Methods (Part I)*; Chong, D. P., Ed.; World Scientific: Singapore, 1995.

(12) Casida, M. E.; Gutierrez, F.; Guan, J.; Gadea, F.-X.; Salahub, D.; Daudey, J.-P. Charge-transfer correction for improved time-dependent local density approximation excited-state potential energy curves: Analysis within the two-level model with illustration for H2 and LiH. *The Journal of Chemical Physics* **2000**, *113*, 7062–7071.

(13) Hine, N. D. M.; Dziedzic, J.; Haynes, P. D.; Skylaris, C.-K. Electrostatic interactions in finite systems treated with periodic boundary conditions: Application to linear-scaling density functional theory. *The Journal of Chemical Physics* **2011**, *135*, 204103.

(14) Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B* **1981**, *23*, 5048–5079.

(15) Zuehlsdorff, T. J. *Computing the Optical Properties of Large Systems*; Springer International Publishing: Switzerland, 2015.

(16) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(17) Lever, G.; Cole, D. J.; Lonsdale, R.; Ranaghan, K. E.; Wales, D. J.; Mulholland, A. J.; Skylaris, C.-K.; Payne, M. C. Large-Scale Density Functional Theory Transition State Searching in Enzymes. *The Journal of Physical Chemistry Letters* **2014**, *5*, 3614–3619.

(18) Govind, N.; Petersen, M.; Fitzgerald, G.; King-Smith, D.; Andzelm, J. A generalized synchronous transit method for transition state location. *Computational Materials Science* **2003**, *28*, 250–258.

(19) Conti, I.; Martínez-Fernández, L.; Esposito, L.; Hofinger, S.; Nenov, A.; Garavelli, M.; Improta, R. Multiple Electronic and Structural Factors Control Cyclobutane Pyrimi-

dine Dimer and 6-4 Thymine-Thymine Photodimerization in a DNA Duplex. *Chemistry - A European Journal 23*, 15177–15188.

(20) 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.

(21) Schreier, W. J.; Schrader, T. E.; Koller, F. O.; Gilch, P.; Crespo-HernÃ¡ndez, C. E.; Swaminathan, V. N.; Carell, T.; Zinth, W.; Kohler, B. Thymine Dimerization in DNA Is an Ultrafast Photoreaction. *Science* **2007**, *315*, 625–629.

(22) Law, Y. K.; Azadi, J.; Crespo-HernÃ¡ndez, C. E.; Olmon, E.; Kohler, B. Predicting Thymine Dimerization Yields from Molecular Dynamics Simulations. *Biophysical Journal* **2008**, *94*, 3590–3600.

(23) Atchity, G. J.; Xantheas, S. S.; Ruedenberg, K. Potential energy surfaces near intersections. *The Journal of Chemical Physics* **1991**, *95*, 1862–1876.

(24) Hackl, E. V.; Kornilova, S. V.; Blagoi, Y. P. DNA structural transitions induced by divalent metal ions in aqueous solutions. *International Journal of Biological Macromolecules* **2005**, *35*, 175–191.

(25) Aquino, A. J. A.; Nachtigallova, D.; Hobza, P.; Truhlar, D. G.; Hättig, C.; Lischka, H. The charge-transfer states in a stacked nucleobase dimer complex: A benchmark study. *Journal of Computational Chemistry 32*, 1217–1227.

(26) Li, J.-H.; Chai, J.-D.; Guo, G.-Y.; Hayashi, M. Significant role of the DNA backbone in mediating the transition origin of electronic excitations of B-DNA - implication from long range corrected TDDFT and quantified NTO analysis - Physical Chemistry Chemical Physics (RSC Publishing). *Physical Chemistry Chemical Physics* **2012**, *14*, 9092–9103.

Graphical TOC Entry