Cite this: Phys. Chem. Chem. Phys., 2012, 14, 9092-9103

www.rsc.org/pccp

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[†]

Jian-Hao Li,^{ab} Jeng-Da Chai,*^{ac} Guang-Yu Guo^{acd} and Michitoshi Hayashi*^b

Received 22nd November 2011, Accepted 24th April 2012 DOI: 10.1039/c2cp23676a

We systematically investigate the possible complex transition origin of electronic excitations of giant molecular systems by using the recently proposed ONTO analysis [J.-H. Li, J.-D. Chai, G. Y. Guo and M. Hayashi, Chem. Phys. Lett., 2011, 514, 362.] combined with long-range corrected TDDFT calculations. Thymine (Thy) related excitations of a B-DNA biomolecule are then studied as examples, where the model systems have been constructed by extracting from the perfect or an X-ray crystal (PDB code 3BSE) B-DNA structure with at least one Thy included. In the first part, we consider the systems composed of a core molecular segment (e.g. Thy, or di-Thy) and a surrounding physical/chemical environment of interest (e.g. backbone, adjacent stacking nucleobases) in gas phase and examine how the excitation properties of the core vary in response to the environment. We find that the orbitals contributed by the DNA backbone and surrounding nucleobases often participate in a transition of Thy-related excitations affecting their composition, absorption energy, and oscillator strength. A vast number of strongly backboneorbital involved excitations are also found at an absorption wavelength below ~ 180 nm predicted by TD- ω B97X. In the second part, we take into account geometrically induced variation of the excitation properties of various B-DNA segments, e.g. di-Thy, dTpdT etc., obtained from different sources (ideal and 3BSE). It is found that the transition origin of several Thy-related excitations of these segments is sensitive to slight conformational variations, suggesting that DNA with thermal motions may from time to time exhibit very different photo-induced physical and/or chemical processes.

1. Introduction

The primary functions of DNA molecules to encode cellular genetic information and to engage in replication have readily rendered them to be the most important biomolecules.^{1,2} Investigation on them has long been a hot topic to date such that a great deal of understanding has been advanced on their mechanisms, physical properties, chemical properties, *etc.* and possible application in nanotechnology.^{3–6} Among all the important issues on DNA, one is related to their optical absorption spectroscopy.

It has been well-known that the optical absorption spectra of nucleobases – thymine, adenine, cytosine, guanine – or different combinations of them are dominated by bright ${}^{1}\pi\pi^{*}$ excitations interlaced with dark ${}^{1}n\pi^{*}$ ones in the ultraviolet range.^{7–35} Its related experimental and theoretical studies can be reviewed, for example, in an excellent collection;³⁶ several quantum chemistry methods, like configuration interaction singles (CIS),^{37–43} complete active space self-consistent-field with multiconfigurational second order perturbation theory (CASPT2/CASSCF),^{44–46} second-order approximate coupled-cluster theory with the resolution-of-the-identity approximation (RI-CC2),^{47,48} time-dependent DFT (TDDFT)^{49–57} etc. have been used in the literature.

For studying how nucleobases alter their excitation properties in response to the surrounding environment within a giant complex DNA molecule, however, it is still a formidable task nowadays employing high-level quantum chemistry approaches, like equations of motion coupled cluster singles and doubles (EOM-CCSD)^{58–60} *etc.*, to directly simulate a huge system composed of the core part and the interesting environment. TDDFT,^{61–64} on the other hand, may currently be the only

^a Department of Physics, Center for Theoretical Sciences,

National Taiwan University, Taipei 10617, Taiwan

^b Center for Condensed Matter Sciences, National Taiwan University, Taipei 10617, Taiwan. E-mail: atmyh@ntu.edu.tw

^c Center for Quantum Science and Engineering, National Taiwan

University, Taipei 10617, Taiwan. E-mail: jdchai@phys.ntu.edu.tw ^d Graduate Institute of Applied Physics, National Chengchi University,

Taipei 11605, Taiwan † Electronic supplementary information (ESI) available: The geometries

of 3BSE molecular segments (Fig. S1), internal coordinates of (non-)relaxed ideal/3BSE dTpdT (Table S1), effects of geometry relaxation, basis set, and solvent (Table S2 and Fig. S2), and projection information on QNTO analysis (Table S3). See DOI: 10.1039/c2cp23676a

choice that can deal with large size molecular systems at the same time with less computational difficulties, although the correctness of output can be a debate.

In practice, model potentials and/or embedded system schemes are often employed to divide the whole system into many sub-layers/sub-systems handled by separate theoretical treatments.^{65–80} In this way, interesting local excitations of a core subsystem embedded in an environment can be sophisticatedly described. However, a noticeable drawback is that orbital transitions from/to environments are not allowed. Fortunately, a direct simulation of giant molecular systems by (TD-)DFT has recently become feasible and reliable due to the great progress in the development of long-range corrected (LC) hybrid functionals.^{81–94} In this case, effects of such orbital transitions can be feasibly included.

In the LC hybrid scheme, the Coulomb operator is first split into a long-range (LR) operator $L(r_{12})/r_{12}$ and a complementary short-range (SR) operator $[1 - L(r_{12})]/r_{12}$,

$$\frac{1}{r_{12}} = \frac{L(r_{12})}{r_{12}} + \frac{[1 - L(r_{12})]}{r_{12}} \tag{1}$$

where $L(r_{12})$ is a function of inter-electronic separation $r_{12} \equiv |\mathbf{r}_{12}| = |\mathbf{r}_1 - \mathbf{r}_2|$, ranging from 0 to 1, and approaching 1 at large r_{12} . Currently, the most popular $L(r_{12})$ used in the LC hybrid scheme is the standard error function (erf), although there are several variants as well.

After the splitting, the LR part of exchange is treated exactly by Hartree–Fock (HF) theory (where $\gamma_{\sigma}(\mathbf{r}_1, \mathbf{r}_2)$ is the one-electron spin density matrix),

$$E_{\rm x}^{\rm LR-HF} = -\frac{1}{2} \sum_{\sigma} \iint L(r_{12}) \frac{|\gamma_{\sigma}(\mathbf{r}_1, \mathbf{r}_2)|^2}{r_{12}} \,\mathrm{d}\mathbf{r}_1 \,\mathrm{d}\mathbf{r}_2 \qquad (2)$$

the SR part of exchange is treated by density functional approximations (DFAs) (*e.g.* LDA or GGAs), and the correlation (also treated by DFAs) remains the same as that of the full Coulomb interaction,

$$E_{\rm xc}^{\rm LC-DFA} = E_{\rm x}^{\rm LR-HF} + E_{\rm x}^{\rm SR-DFA} + E_{\rm c}^{\rm DFA}$$
(3)

In recent years, LC hybrid functionals have gained increasing attention due to their success in several important applications.^{83–86}

In addition, we have lately proposed quantified natural transition orbital (QNTO) analysis⁹⁵ which is able to clearly exhibit the transition origin of electronic excitations. This framework should be especially useful for treating large biomolecules with high orbital energy density, where even the low-lying excitations are generally composed of several singly excited configurations (SECs) with similar weights pairing different occupied and unoccupied orbitals in the linear combination of SECs (LCSEC). This situation may be more commonplace particularly for DNA systems that consist of many similar molecular segments. Under such circumstances, QNTO analysis can often lead to one or two dominant SECs corresponding to NTO1 and NTO2 in the NTO-based LCSEC. Here NTO1(2) stands for the first (second) dominant NTO pair. The NTO1(2) can then be interpreted in terms of a standard-orbitals set chosen based on the type of systems or chemical interest; for example, if any important photo-induced chemical reaction takes place in a dimer system, orbitals from

monomer segment can be chosen as standard-orbitals. In other words, the transition origin of an electronic excitation of a dimer can be interpreted by the monomer orbitals.

Thus, in light of the above two developments, the present work is an effort by using TDDFT with LC hybrid functionals (LC-TDDFT) and QNTO analysis to systematically investigate the electronic excitations of giant molecular systems, here the DNA, in which orbitals exchange between the coresubsystem and the environment can be important.

We choose a Thy-appearing region of B-DNA (the commonest cellular form of DNA)⁹⁶ as an example and focus on its Thyrelated excitations in which hole- and electron-orbitals are both significantly contributed by Thy. We study how different surrounding environments can play a role in its transition origin, absorption energy and oscillator strength. This is accomplished by extracting various B-DNA segments (nucleobases or oligonucleotides) with different environmental factors existing in the vicinity of a Thy and by looking into their respective electronic excitation properties. Moreover, the excitation properties of B-DNA segments obtained from different sources (ideal and 3BSE) are compared with each other to investigate the conformational effect as well.

Recently, Kozak et al. have reported a thorough study⁹⁷ on the singlet electronic excitations of two π -stacked Thys by implementing a variety of quantum chemistry methods, including CIS, CIS(2), EOM-CCSD, TDDFT and CASSCF, based on molecular exciton theory.⁹⁸ It is found that the coupling of Thy monomer orbitals to form dimer orbitals strongly depends on the distance and the relative orientation of two Thys. However, we have shown by using QNTO analysis⁹⁵ that for a real dimer system with low or no symmetry, such as two π -stacked Thys extracted from B-DNA, two local orbital transitions can no longer equally contribute to an electronic excitation predicted by the molecular exciton theory. If the dimer separation is short enough, the additional charge transfer (CT) contribution can also exist. Under such circumstances, QNTO can be a powerful tool for studying the transition origins of electronic excitations that may consist of local or CT transition components in such a low-symmetry situation, which is commonplace in median and large bio-molecular systems. Lange and Herbert⁹⁹ have also applied TDDFT with LC hybrid functionals to obtain the accurate absorption energies of interbase CT electronic excitations that are underestimated by TDDFT calculation with pure (e.g. LDA) or global hybrid (e.g. B3LYP) density functionals. Since both of the reports focus on bare nucleobase model systems, a remaining question of interest is indeed to what extent the environmental factors like backbone affect the electronic excitations of nucleobases.

The paper is organized as follows. In Section 2, computational procedures are introduced, where the preparation of various B-DNA sliced systems and DFT/TDDFT calculation details are presented. In Section 3.1, the interpretation of the excitation properties of various systems calculated by TDDFT with QNTO analysis is introduced. Sections 3.2–3.6 compare the results of various systems where the influence of different environmental factors, *i.e.* the presence of another Thy, backbone, Watson–Crick (WC) pairing nucleobases, stacking nucleobases, and conformational variation to the Thy-related excitations is successively focused on. The examination of factors such Downloaded by National Taiwan University on 08 June 2012 Published on 24 April 2012 on http://pubs.rsc.org | doi:10.1039/C2CP23676A as geometry relaxation, calculation basis set, and surrounding water solvent of Polarizable Continuum Model (PCM)¹⁰⁰ on affecting electronic excitations is also available in ESI.† Section 3.7 demonstrates the overall backbone influence on several low-lying excitations (not particularly Thy-related ones) of the systems consisting of backbone structure. Conclusion is given in Section 4.

2. Computational details

2.1. Preparation and TDDFT calculations of Thy-comprised molecules

Various molecular systems including at least one Thy nucleobase have been prepared by extracting from ideal¹⁰¹ or X-ray determined (1.6 Å resolution; PDB code 3BSE)¹⁰² B-DNA crystal structure as shown in Fig. 1 and Fig. S1 (ESI⁺). To minimize the possible geometrical clashes or artificially unrealistic atom positions in the two data, a dTp(dApdTpdTpdG)pdC···AdTp(dApdA)pdCG for the ideal case and a dCp(dApdTpdTpdG)pdT...GdTp-(dApdA)pdCA for the 3BSE case have been taken first and their hydrogen and backbone atoms have been optimized by DFT/\u03c6B97X in water of the PCM model with the heavy atoms of other nucleobases being fixed. Note that both pieces are composed of two strands of 6 DNA nucleobases attached to each other by WC hydrogen bonds and their sugar (d) and phosphate (p) backbones are mostly retained. The (dApdTpdTpdG) and (dApdA) are the sections for constructing molecular segments shown in Fig. 1 and Fig. S1 (ESI[†]). Table S1 (ESI[†]) also shows the major internal coordinates of the DNA backbone for the (non-)relaxed ideal/3BSE dTpdT. The partial optimizations employed here can avoid generation of unreliable DNA structures since each of the environmental factors, e.g. base-stacking, WC-pairing forces, backbone, solvent, can play a subtle role in balancing DNA conformation; partial consideration of these factors or inappropriate use of force field in molecular mechanics or exchange-correlation functionals in DFT can, for instance, lead to crashed WC nucleobase pairs. With the flexible structure of nucleobases in DNA in terms of twist (Ω), tilt (τ), roll (ρ) etc. contingent to numerous environmental factors,¹⁰³ the two pieces have thus served as good examples to allow possible investigation

of (1) environmental and (2) conformational effect on electronic excitations. For (1) we study how the DNA backbone, WC-pairing bases, and stacking bases affect the Thy-related excitations of a core segment; for (2) we examine the variation of excitation properties of Thy-related excitations resulting from the geometry variation of a segment.

TDDFT calculations for the extracted systems (Fig. 1 and Fig. S1, ESI[†]) are carried out in gas phase. Although water solvent surrounding DNA molecules should also affect its electronic excitations, the inclusion of wandering H₂O molecules for these systems become too complex and therefore is not performed here. However, the PCM solvent field surrounding the optimized ideal dTp(dApdTpdTpdG)pdC···AdTp(dApdA)pdCG has also been included in the TDDFT calculation of its extracted dTpdT as a test, presented in Table S2 (ESI[†]), along with the other two test cases of geometry relaxation and basis set influence on electronic excitations. Note that the role of actual H₂O molecules in affecting electronic excitations cannot be fully replaced by the polarized solvent field. The H₂O molecules can, for instance, have CT contribution on electronic excitations of the solutes. Further studies on this topic should therefore be concentrated elsewhere.

All the DFT/TDDFT calculations are executed using Gaussian09¹⁰⁴ with an economic basis set $6-31G(d)^{105}$ given that up to 70 excitations of a system as large as dApdTpdTpdG will be computed. Moreover, since we focus on the solute form of the B-DNA molecule whose structure is stabilized by ionizing one proton for each phosphate, extra electrons are added to each sliced B-DNA segment to account for the charged phosphates. For instance, the dTpdT molecule, consisting of one phosphate, is set to have net charge – e. Therefore each molecule is a closed-shell electronic system and we pay attention to the singlet electronic excitations.

Various exchange–correlation functionals – B3LYP,^{106,107} CAM-B3LYP,⁸² ωB97X,⁸⁴ ωB97X-D⁸³ and LC-ωPBE⁸¹ – are implemented in the DFT/TDDFT calculations. After a TDDFT calculation, QNTO analysis is followed, whose code has been developed locally, allowing us to process TDDFT output results generated by Gaussian09.

In the previous report⁹⁵ we have found that $TD-\omega B97X$ among several tested theoretical methods most closely reproduces the transition origins of several low-lying excitations



Fig. 1 Various Thy-comprised systems extracted from partially relaxed ideal B-DNA (see the text): (a) Thy with the sugar backbone (dT), (b) two π -stacked thymines (di-Thy), (c) di-Thy with the sugar-phosphate backbone (dTpdT), (d) dTpdT with WC pairing nucleobases (dTpdT···dApdA), and (e) dTpdT with adjacent stacking nucleobases (dApdTpdTpdG). The order of bases is counted from the 5'-end to the 3'-end of the DNA strand. The broken chemical bonds upon extraction have been saturated with a hydrogen atom along the same bond direction.

that EOM-CCSD predicts for Thy and dT. Hence, dealing with larger DNA-sliced systems here, we focus on TD- ω B97X results. The same reason applies when using DFT/ ω B97X for B-DNA structural relaxation discussed above.

3. Results and discussion

3.1. The excitation energy, oscillator strength, and transition origin of several low-lying Thy-related excitations of ideal di-Thy, 3BSE dT, ideal/3BSE dTpdT…dApdA and ideal/3BSE dApdTpdTpdG

The detailed information on calculated absorption wavelength, oscillator strength and NTO1(2) transition origin of the first several selected Thy-related excitations of various systems are listed in Table 1. Since we are interested in the influence of the surrounding environment and conformational variation on Thy-related excitations, the standard-orbitals are chosen from the monomer orbitals of the ideal/3BSE Thy generated in its B3LYP-DFT calculations. The standard orbitals set can also be generated by other theoretical methods.

Incidentally, the orthogonality of DFT/B3LYP standard orbitals from different moieties of a DNA segment is maintained (mutual projection is below or in the same order of the second projection digit with the largest being ~ 0.02) for all the studied cases here. For cases where orthogonality is not fulfilled, such as the two moieties are very close to each other, the standard orbitals cannot be regarded as purely the monomer orbitals and orthogonalization has to be performed first or the inter-projection amount has to be borne in mind.

Moreover, the targeted core - Thy - is rather intact in different situations studied here so that the Thy stays in the same chemical species, not in other isomers with the same composition of atoms. Therefore, its ground state electronic structure remains the same. If the geometry of the target, Thy, is distorted strongly such that it, when let free, will be stabilized to other isomers, its ground state electronic structure is then changed, and the comparison of transition origins of two different species becomes meaningless. In such a case, the same set of standard-orbitals for electrons from unoccupied orbitals and for holes from occupied orbitals also cannot be well defined, since some of the occupied and unoccupied orbitals can be switched. In Table 1, the selected Thy-related excitations, denoted SO-hosted (standard-orbitals hosted) excitations thereafter, are those with their hole- (NTO1-H) and electron-orbital (NTO1-E) of NTO1 having overall more than 30% density contribution from the standard-orbitals for holes and for electrons, respectively. This is confirmed by first projecting the NTO1(2)-H(E) to the standard-orbitals and the density contribution of a standard-orbital is calculated by its square of projection coefficient. The panels a and b in Fig. 2 show the standard-orbitals of the 2nd ideal Thy for holes and for electrons, respectively. The other standard-orbitals of the 1st ideal, 1st 3BSE and 2nd 3BSE Thy for holes and for electrons are similar to those in Fig. 2. Therefore, for example, if the studied core molecular segment is the ideal di-Thy, the standard-orbitals used for projection of a NTO1(2)-H(E) will be the 6 orbitals of the 2nd ideal Thy shown in Fig. 2 plus the other 6 similar orbitals from the 1st ideal Thy.

The detailed projection coefficients of the NTO1(2)-H(E) of each SO-hosted excitation are listed in Table S3 (ESI†). Moreover, the transition expression of a NTO1(2) is determined according to these coefficients. To denote the transition expression, we introduce several notations: for example, P1t2 refers to P1 of the 2nd Thy with a positive coefficient in a system containing two Thys, *e.g.* di-Thy, dTpdT, while P1t2 corresponds to P1 with a negative coefficient. P1t12 is an abbreviation of P1t1 plus P1t2 both of which have density contribution greater than 0.3 in a NTO1(2)-H. Similarly, P1t12 corresponds to the case having a combination of P1t1 and P1t2. In addition, the parentheses "(·)" in P1t(1)2 indicates that it is made up of P1t1 and P1t2 but the coefficient of P1t1 is (0.3, 0.1].

The integration of the orbitals originating from backbone structure contributing to a NTO1(2)-H is denoted B in a transition expression, whereas the integration of Thy orbitals other than the ones shown in the panels a and b in Fig. 2 for NTO1(2)-H and for NTO1(2)-E, respectively, are both denoted O. A1 and A2 are also likewise introduced for dTpdT···dApdA in a transition expression indicating the integrated contribution of orbitals from the 1st Ade and 2nd Ade, while A and G are introduced for dApdTpdTpdG indicating orbital-contribution from Ade and Gua. If the contribution of the orbital-fraction (B, O, A1, A2, A or G) is (0.3, 0.1], it is shown in "(·)" in the transition expression.

Following these rules, the transition expression of the excitations of each system can be determined. As an example, Fig. 3 shows NTO1-H(E) of an excitation of dTpdT and the corresponding transition expression. In addition, a similar categorization (Type-A-Type-G) introduced in the study of Thy and dT⁹⁵ for SO-hosted electronic excitations is also adopted here to classify electronic excitations: N1-S1 (Type-A); P1-S1 (Type-B); N1 N2-(S1) S2 (Type-C); P2-S1 (Type-D); P1-S2 (Type-E); N2-S1 (Type-F); N1 (N2)-S2 (Type-G). Although this should only be regarded as a rough classification for excitations because their actual composition can be complex, it helps link similar excitations between two systems together. Since most of the studied segments here consist of two Thys, a SO-hosted excitation can be a local (on the same Thy) or CT (across the two Thys) one. An excitation denoted Type-A2 is therefore used to represent local Type-A excitation occurring on the 2nd Thy. Type-B12, on the other hand, stands for a Type-B excitation with the P1 orbital on the 1st Thy excited to the S1 orbital on the 2nd Thy. For simplicity, in the following discussions the term "local Type-A(s)" will be used to refer to "local Type-A excitation(s)" and so on. Furthermore, an excitation can also be combined with different types of transitions. For instance, the excitation with NTO1 transition expression "P1t2 P2t1-S1t1" is combined with Type-D1 local transition and Type-B21 CT transition with the opposite phase between P1t2 and P2t1 and is denoted Type-D1B21.

We also employ $\sigma_{\rm E}$ (environmentally resulted root mean square deviation of standard-orbital coefficients) for examining the environmental effect on NTO1 transition origins of excitations of a core molecular segment, while for comparing results of the same molecules extracted from ideal and 3BSE sources it is denoted $\sigma_{\rm G}$ (geometry resulted deviation). The reference systems used for calculating different σ are detailed in Table 1.

Table 1 Various excitation properties of (a) ideal Thy, dT; (b) ideal di-Thy, dTpdT, dTpdT, dTpdT, dApdA, and dApdTpdTpdG; (c) 3BSE dT, dTpdT, dTpdT···dApdA and dApdTpdTpdG in gas phase provided by TD- ω B97X calculation and QNTO analysis. *N* denotes excitation order, whereas P denotes the NTO2 phase for cases where NTO1 has less than 70% domination in the whole excitation. λ (nm) and *f* stand for absorption wavelength and oscillator strength, respectively. NTO1(2) and % record the expression of the transition origin of the first (second) NTO pair and its domination in the whole excitation. The type denotes the excitation classification based on the EOM-CCSD Thy NTO1 expressions used in ref. 95. In (b) and (c) the excitation of a system is referenced to a similar excitation of ideal di-Thy, while the 9th excitation (Type-A2) of ideal dTpdT is referenced to the 2nd excitation of ideal di-Thy, while the 9th excitation origin compared to the result of the referenced system

(a)																	
Ideal Thy									Ideal dT								
N	λ (nm)	f		NTO1		%	Тура	e N	λ	(nm)) f		NTO1		%	Туре	
1	242.48	0.0	0001	N1t2-S1t2		100	A2	1	2	41.11	0.	.0001	N1t2-S1t2 (S2t	2)	99	(A2)	
2	224.84	0.2	2011	P1t2–S1t2		96	B2	2	2	28.88	0.	.2877	P1t2–S1t2	_	96	B2	
3	190.96	0.0	0000	N1t2 N2t2–(S1t2) S2	2t2	99	C2	3	1	89.38	0.	.0004	N1t2 N2t2–S2t2	2	98	(C2)	
4	177.69	0.0)690 149	P2t2–S1t2		98	D2	4	1	78.50	0.	.1858	P1t2-S2t2		92	E2	
5	1/3.93	0.2	0003	P112-5212 N2+2 \$1+2		97	E2 E2	5	1	70.13 67 70	0.	0065	P212-5112 N2+2 B S1+2		92	(F2)	
7	158.23	0.0	0005	N1t2 N2t2–(S1t2) S2	2t2	93	(G2)) 7	1	58.12	0.	.0005	N1t2 N2t2-S1t2	2 S2t2	81	(G2)	
(b)																	
N; P	λ (nm)	f	NTO1((2)	$\sigma_{\rm E}$	%	Туре	N;	Ρλ(nm) j	f	NTO	1(2)	$\sigma_{\rm E}$	% Т	/pe	
Ideal	di-Thy	$(\sigma_{\rm E} \text{ ref.})$	to idea	l Thy)				Id	eal dT	pdT ($(\sigma_{\rm E} \text{ ref.})$	to ide	eal di-Thy)				
1	245.43	0.0001	Nlt1-S	Sltl		99	Al	1	237	7.59 (0.0002	N1t2-	-S1t2 (S2t2)	0.019 8	36 (A	.2)	
2	242.48	0.0001	N1t2-S	51t2	0.003	99	A2 D2	2	237	1.48 (0.0006	NItl-	-SItl (S2tl)	0.025 8	37 (A	.1)	
5 4	228.88	0.1270	P1t2-5	112	0.007	94 04	B2 B1	3	234	1.83 () 27 (0.0455	P1t2-	S112 S1t1	0.070 7	1 B. 72 B	2	
5	191.30	0.0201	N1t2 N	$V_{2t2} = (S_{1t2}) S_{2t2}$	0.010	99	C2		192	2.51 (0.0015	N1t1	N2t1-S2t1	0.007 7	6 (C	1	
6	190.76	0.0001	NItl N	$V_{2t2} = (S_{1t2}) + S_{2t2}$		98	(C1)	6	188	3.85 (0.0203	Plt1 ((B)-S1t2 (S2t1)	0.048 9)2 (E	1B12)	
7	183.11	0.0165	P1t1-S	1t2 S2t1	_	90	È1B12	7	188	8.26 (0.0009	N1t2	N2t2-(S1t2) S2t2	0.014 9	93 Č	2	
8	180.26	0.0093	P1t2 P2	$2t\underline{1}(2)-S1t1(2)$ (S2t2)		65	(D <u>1</u> B21 e	etc.) 9	183	8.96 (0.1209	P1t2-	S1t1 S2t2	— 8	39 (E	2B21)	
+	150.00		P2t2-S	1t(1)2	0.015	28		10	177	7.62 (0.0686	P2t1	B–S1t1	- 7	/2 (E)])	
9	179.02	0.1154	P2t2-S	1t2 2+1 \$1+1	0.015	70 96	D2	12	176	5.87 (0.0592	PItI($\underline{2}$)–(S1t2) S2t <u>1</u>	- :	94 (E	<u>1</u> B12E21)	
10	177.95	0.0854	FILZ F.	211-5111		80	DIDZI	_				FIU)2 (F21 <u>1</u>)–3111	-	10		
Ideal	dTpdT.	··dApd	A ($\sigma_{\rm F}$ r	ef. to ideal dTpdT)				Id	eal dA	pdTp	odTpdC	G (σ _F r	ef. to ideal dTpdT)			
1	237.62	0.0401	Plt2-S	1t2	0.033	72	B2	1	239	0.17 (0.0408	Pltl-	Sltl	0.107 8	80 B	l	
2	235.23	0.4234	P1t1-S	1t1	0.037	70	B1	3	235	5.85 (0.2410	P1t2-	S1t2	0.104 8	33 B2	2	
3	226.00	0.0011	Nlt1-S	Sltl	0.042	96	Al	4	233	3.13 (0.0006	Nltl-	-S1t1 (S2t1)	0.034 9	96 (A	.1)	
4	224.87	0.0003	NIt2-S	S1t2 (S2t2)	0.023	96	(A2)	5	232	2.75 (0.0004	NIt2-	-S1t2 (S2t2)	0.018 9	97 (A	.2)	
12	193.13	0.0008	(NItI) (N1t2)	$N_{2}t_{1} = S_{2}t_{1}$ $N_{2}t_{2} = (S_{1}t_{2}) = S_{2}t_{2}$	0.020	95	(C1)	13	19:	0.01	0.0003	(N I U D 1 + 2	$) N_2 t_1 - (S_1 t_1) S_2 t_1$ S_1+1 (S_2+2)	0.03/8	58 (C	.1) (2021)	
13	186.43	0.0010	(1N112) P1t2 (F	$1N_{212} = (S_{112}) S_{212}$	0.021	94 47	(C2) (D1E2B2)	14	191	1.20 (1.48 (0.0384	N1t2	$N2t2_{(S2t2)}$	0.085 0	55 (E 50 C	2 D 21)	
_	100.45	0.0050	P1t1 (A	$(32t_{1})$ $(32t_{2})$		42		20	187	7.00	0.0501	Plt1-	(S1t2) S2t1 (A)	0.125 7	76 (E	- 1B12)	
18	185.87	0.0211	Pltl (F	(A2)-S1t2 (S2t1)		74	(D2E1B1	2) 23	184	1.40 (0.0365	Plt2-	(S1t1) (S2t <u>2</u>) G	- 8	35 (E	<u>2</u> B21)	
19	185.10	0.0891	P1t(1)2	(A2)–S1t12 (S2t2)	—	49	(B2B21 e	etc.) 27	179	0.35 (0.0778	P1t2-	S2t2 G	— 7	73 (E	2)	
+			$P1t(\underline{1})2$	(A2)–S1t1 <u>2</u>		37											
23	182.27	0.0236	P2t1 A	2–Sltl Al	0.128	60	(D1)										
+			(P2t1)	A2-SITI AI		30											
(c)																	
N; P	λ (nm)	f	NT	01(2)	$\sigma_{ m G}$	%	6 Туре	N; P	λ (nn	1) f		NTO	1(2)	$\sigma_{ m G}$	%	Туре	
3BSE	$dT (\sigma_G)$	ref. to	ideal d	Г) Э. 51+2 (52+2)	0.012		(10)	3BSE	dTpd	$\Gamma(\sigma_G)$; ref. to	o ideal	dTpdT)	0.050	00	(1 1)	
1	238.70	0.000	0 $\mathbf{P}1 t'$	2-51t2(52t2)	0.013	9 1 0	9 (A2) 6 D2	1	238.0	2 0.	.0010	NIU-	-S1t1(S2t1) S1t2(S2t2)	0.059	98	(A1)	
23	224.38	0.291	9 PIL 15 N1t	2-5112 2 N2t2-82t2	0.004	1 9 2 9	9 (C2)	3	233.7	80. 60	.0002	P1t1_	-S112 (S212) S1t1	0.039	90 67	(A2) B1	
4	176.14	0.177	$^{\prime}8$ P1t2	2–S2t2	0.020	8	5 E2	_	231.7	0 0.	.0401	Plt2-	S1t2	0.145	29	DI	
5	174.97	0.123	5 P2t2	2–S1t2	0.029	8	5 D2	4	226.7	6 0.	.4176	P1t2-	S1t2	0.119	66	B2	
6	166.91	0.001	2 N2t	2 B–S1t2	0.047	9	7 (F2)	+				Pltl-	S1t1		29		
7	158.09	0.002	24 N1t	2 N2t2 (B)–S1t2 S2t2	0.075	5 7	5 (G2)	5	191.9	8 0.	.0035	Nltl	N2t1-S2t1	0.069	97	(C1)	
								6	189.6	0 0	.0006	Nlt2	N2t2–(S1t $\underline{2}$) S2t2	0.031	96	C2	
								8	185.0	1 0	.0024	PItI ((B)-S1t2		98	(B12)	
								10	178.1	9 0.	1079	PIII = D1+2	5211 (P2+1) R S1+1		70 81	(D1B21)	
								13	176.5	4 0	0580	$(P1t^{2})$	$P_{2t1} B_{S1t1}$		74	(D1B21) (D1R21)	
								15	170.5	r 0.	.0000	(1112)	, 12(1 D 5)(1		77	(1021)	
3BSE	E dTpdT	···dApo	iA (σ _G	ref. to ideal dTpdT···	dApdA	A)		3BSE	3BSE dApdTpdTpdG (σ_{G} ref. to ideal dApdTpdTpdG)								
1	233.55	0.084	8 P1t	1 (O)–S1t1	0.115	5 7	4 (B1)	2	236.4	4 0	.0370	Pltl-	S1t1	0.083	81	B1	
3	228.95	0.228	8 P1t2	2–S1t2	0.093	6	6 B2	3	233.7	0 0.	.0032	Nltl-	-S1t1	0.061	93	Al	

Table 1 (continued)

(c)													
N; P	λ (nm)	f	NTO1(2)	$\sigma_{ m G}$	%	Type	N; P	λ (nm)	f	NTO1(2)	$\sigma_{\rm G}$	%	Type
_			A1 (A2)-A1		15		4	230.11	0.2649	P1t2-S1t2	0.035	79	B2
4	227.76	0.0261	N1t1–S1t1	0.060	91	A1	5	227.65	0.0003	N1t2-S1t2 (S2t2)	0.026	96	(A2)
5	226.04	0.0038	N1t2-S1t2 (S2t2)	0.029	94	(A2)	13	192.96	0.0002	N1t1 N2t1-(S1t1) S2t1	0.056	80	C1
14	192.02	0.0027	N1t1 N2t1–S2t1	0.056	95	(C1)	17	188.75	0.0007	N1t2 N2t2–(S1t2) S2t2	0.026	70	C2
15	190.13	0.0005	N1t2 N2t2–S2t2	0.036	94	(C2)	24	183.57	0.0828	P1t2 (B)-S1t1		53	(B21)
19	184.59	0.0084	P2t1 (A1) A2–S1t1	0.117	79	(D1)	+			P1t1 A-(S2t1) A		25	
22	182.45	0.0366	P2t2 (A1) A2–S1t2		66	(D2)	26	182.41	0.0547	P1t2 (B)–(S1t1) S2t2 (G)	0.130	40	(E2B21)
+			A1 (A2)–A1 A2		17		+			(P2t1) B–S1t1 (S2t2)		23	
24	180.33	0.0042	P1t1 (B) (O)–S1t2		92	(B12)	27	181.67	0.0237	P1t2 B–S1t1		75	(B21)
25	180.05	0.0181	P1t2-S1t1	—	87	B21	28	181.06	0.0446	P1t1 (B)–S1t2	—	74	(B12)



Fig. 2 Standard-orbitals of Thy generated in the B3LYP-DFT calculation. Isovalue = 0.02 is adopted for the graphical representation. P1, N1, P2 and N2 correspond to the HOMO, HOMO – 1, HOMO – 2, and HOMO – 3, respectively, and S1 and S2 are the LUMO and LUMO + 1. P1 and P2 are from the components of ${}^{1}\pi\pi^*$ -character excitations while N1 and N2 are from those of ${}^{1}n\pi^*$ -character excitations.



Fig. 3 An example of SO-hosted excitations of the ideal dTpdT. The NTO1–H and NTO1–E for the 6th excitation (188.85 nm, f = 0.0203) of ideal dTpdT are plotted. NTO1 accounts for 92% of the whole excitation. The transition origin can be expressed in terms of standard-orbitals as "P1t1 (B)–S1t2 (S2t1)".

3.2. Ideal Thy vs. di-Thy

To detail the environmentally resulted variation of excitation properties of several chosen core segments consisting of Thy(s), we take a deeper look into each type of electronic excitation. We first compare the results of ideal Thy and di-Thy shown in Table 1a and b.

Local Type-B. Obviously the single bright local Type-B of Thy at 224.84 nm evolves into two bright excitations of di-Thy: one being a little weakened at 228.88 nm (Type-B2) and the other a little enhanced at 225.66 nm (Type-B1). Moreover, the two bright excitations of di-Thy are both dominated by single NTO1 instead of by two similar weighted local transitions given in molecular exciton theory.⁹⁸ Therefore the separation of the two Thys is short such that their interaction cannot be well described by dipole–dipole interaction.

Local Type-A/C. As for the local Type-A, its absorption wavelength of Thy at 242.48 nm results in the Type-A1 and Type-A2 of di-Thy at 245.43 nm and 242.48 nm; in other words, almost no energy splitting occurs. This is attributed to the weak transition dipole of Thy, which also leads to the small oscillator strengths. Similarly, no energy splitting occurs for the two weak local Type-C absorptions as well.

Local Type-D/E; CT Type-B. For even higher-lying excitations, Table 1b shows that they are mainly formed by local Type-D/E and CT Type-B transitions. Not like local Type-Bs, we see that local Type-D/E transitions are often mixed with CT Type-B transitions to form excitations, as the excitations at 183.11 nm (Type-E1B12), 180.26 nm (Type-D<u>1</u>B21 *etc.*), and 177.93 nm (Type-D1B21) have shown. This should be due to their nearly degenerate nature.

The σ_E shown for the Type-A2, B2, C2, and D2 indicates that the presence of the 1st Thy has a minor effect on the transition composition of these local excitations of the 2nd Thy.

3.3. Ideal di-Thy vs. dTpdT

We now move onto the effect of the link between two Thys, *i.e.* the DNA backbone, on electronic excitations.

Local Type-B. In the presence of the DNA backbone, the two bright absorption peaks of di-Thy become a brighter one and a darker one of dTpdT. The stronger one is predicted to be Type-B1 at 232.2 7 nm while the weaker one is Type-B2 at 234.83 nm, both have a 3% energy lowering. Moreover, their NTO1 contributions to excitation are both lowered, at 72% and 71%, respectively. Therefore, evidently the presence of the backbone not only shifts the excitation energies, but also mediates the transition origins, even if the backbone almost donates no orbital components to NTO1–H(E) (for details of transition coefficients see Table S3 (ESI†)).

Local Type-A/C. Next we deal with the local Type-A. In the presence of the backbone the two local Type-As instead have 2% and 3% energy arising, in contrast to the energy-lowering imposed on the local Type-Bs discussed above. Consequently these two types of excitations exhibit even closer absorption

energies. The presence of the backbone also results in a slight S2-component mixing into the Type-A1 and Type-A2.

The remaining two ${}^{1}n\pi^{*}$ -character excitations of di-Thy, *i.e.* local Type-Cs, though switching their relative order, retain their main features such as absorption energy, oscillator strength and transition origin of NTO1 in the presence of the backbone.

Local Type-D/E; CT Type-B. Finally we consider the excitations formed by the mixture of CT Type-B transition and local Type-D/E transition. The Type-E1B12 has an enhancement of 23% of oscillator strength (from 0.0165 to 0.0203) with a small shifted excitation energy (183.11 nm to 188.85 nm), while the excitations of di-Thy at 180.26 nm (Type-D1B21 *etc.*), 179.02 nm (Type-D2), and 177.93 nm (Type-D1B21) have no correspondences in the first 10 excitations of dTpdT and are replaced by excitations at 183.96 nm (Type-E2B21), 177.62 nm (Type-D1), and 176.87 nm (Type-E1B12E21). Furthermore, the backbone orbital component is involved in the 188.85 nm and 177.62 nm excitations. These results suggest that the larger average distance of electron transition which encompasses wider spatial area for these higher-lying excitations results in the high backbone-sensitivity.

3.4. Ideal dTpdT vs. dTpdT···dApdA

Local Type-B. For even a larger system with dApdA on the opposite strand in charge of the WC hydrogen bonding effect, the TD- ω B97X calculated absorption spectrum (Fig. 4) shows that there are two strong peaks. However, after examining their NTO1 transition origins only the lower energy one at 235.23 nm corresponds to a SO-hosted (Type-B1) excitation. It corresponds well to the brightest peak, also a Type-B1, of dTpdT with a minor energy shift (-1%). A similar situation also occurs for the weaker Type-B2, now switched to be the lowest singlet excitation (Singlet-1).

Local Type-A/C. The NTO1 transition origins of the two local Type-As are retained as well, but again their excitation energies arise (5% and 6%). For the two local Type-Cs, their absorption energies have minor deviation accompanied with a modest $\sigma_{\rm E}$ (0.020 and 0.021).

Local Type-D/E; CT Type-B. For the CT Type-B involved excitations, except the Type-D1, the excitations of dTpdT at 183.96 nm (Type-E2B21), 177.62 nm (Type-D1), and 176.87 nm (Type-E1B12E21) are replaced by those with different transition origins at 186.43 nm (Type-D1E2B21), 185.87 nm (Type-D2E1B12), and 185.10 nm (Type-E2B21 *etc.*). Moreover, all the 4 excitations mix in orbital components from the opposing adenines to their NTO1(2). Therefore again the environment effect is more pronounced in these higher-lying excitations.

3.5. Ideal dTpdT vs. dApdTpdTpdG

Local Type-B. In the presence of stacking nucleobases next to dTpdT and the backbone linking them, *i.e.* dApdTpdTpdG, the two local Type-Bs switch their positions such that the brighter absorption changes from the Type-B1 to the Type-B2, even if their excitation energies are not altered much. The oscillator strength of the brighter one is also dampened to almost one half. Hence it can be concluded that the local Type-Bs

are sensitive to environmental surroundings of the backbone and stacking nucleobases; that is, the NTO1 contribution to the whole excitation can be reduced, the absorption energy can be red-shifted, and the relative strength between Type-B1 and Type-B2 can be variable.

Local Type-A/C. On the other hand, the two local Type-As again have a 2% energy arising and a minor change of the transition origin reflected in $\sigma_{\rm E}$ compared to the results of dTpdT.

As for the two local Type-Cs, they have a minor deviation of the transition origin as well but have even a smaller energy shift compared to those imposed on local Type-As.

Local Type-D/E; CT Type-B. The remaining excitations related to local Type-E and/or CT Type-B transitions of dTpdT again are affected quite intensively. The Type-E1B12, although retains similar excitation energy, has more than doubled its oscillator strength and its NTO1 transition expression mixes in A with the involvement of S2t1 increased whereas B and S1t2 diminished at the same time. On the other hand, the Type-E2B21 has more notable energy lowering (4%) but with an oscillator strength reduction (52%). The remaining Type-D1 and Type-E1B12E21 of dTpdT are replaced by the Type-E2B21 and Type-E2, both of which have noticeable G mixed in the NTO1-E.

3.6. Ideal vs. 3BSE (X-ray crystal) structures

We examine the effect of conformational variation on electronic excitation properties of dT, dTpdT, dTpdT···dApdA and dApdTpdTpdG. The transition origins of excitations of geometrically different systems can be compared with each other, given that the segment stays in the chemically same species. The detailed information on the first 10 (7 for dT) SO-hosted excitations of each studied 3BSE system is listed in Table 1b.

dT. A one-to-one mapping can be made for all the 7 SO-hosted excitations and their absorption energies, oscillator strengths and NTO1 transition origins are much retained.

dTpdT. Analysis of dTpdT shows that the conformational variation is quite influential on the NTO1 transition origin of the two local Type-Bs such that the $\sigma_{\rm G}$ value for Type-B1 is 0.145 and for Type-B2 is 0.119. The stronger absorption also transforms from the Type-B1 to the Type-B2, even if the relative magnitude of absorption energy and oscillator strength both have little changes.

In contrast, the local Type-As and Type-Cs are rather sustainable in the NTO1 transition origin and absorption energy to conformational variation, and their oscillator strengths remain small. Moreover, the Singlet-1 becomes the Type-A1.

The remaining listed 4 higher-lying excitations have no corresponding ones in their ideal counterpart system, although they are still at around the absorption wavelength of 175 nm–190 nm; this can also be observed in the results of ideal di-Thy, dTpdT, dTpdT…dApdA and dApdTpdTpdG, that is, CT Type-B involved excitations of the first 10 SO-hosted excitations are all located in this energy range, though their composition of the NTO1 transition origin is very sensitive to the surrounding environment.



Fig. 4 Percentage of backbone density contribution to the NTO1–H (red and black columns) and calculated absorption cross sections (curves) of the first several electronic excitations of backbone-included systems. Red and black columns correspond to the SO-hosted and non-SO-hosted excitations, respectively. The absorption cross sections are plotted by broadening the oscillator strength of each excitation with 0.03 eV of half-width at half-height. (a) The first 16, 25, 50, and 70 electronic excitations of ideal dT, dTpdT, dTpdT···dApdA, and dApdTpdTpdG, respectively, provided by TDDFT with a global hybrid functional B3LYP and four LC hybrid functionals. Wavelength ranges for plotting the B3LYP results of dTpdT and larger systems are adjusted independently reflecting the considerably smaller calculated absorption energies. (b) A magnified plot for the TD- ω B97X results of all 4 systems at the wavelength range in which the backbone intensely participates. For more discussion see the text.

dTpdT···**dApdA** and **dApdTpdTpdG**. For much larger systems, dTpdT····dApdA and dApdTpdTpdG, excitations can still be well traced by local Type-B, local Type-A/C, local Type-D/E and CT Type-B. Local Type-As and Type-Cs again remain rather intact, while local Type-Bs are more sensitive to both environmental and conformational mediations. Overall, the flip of relative strength of Type-B1 and Type-B2 has occurred under conditions of (i) ideal dTpdT vs. ideal dApdTpdTpdG, (ii) ideal

dTpdT vs. 3BSE dTpdT, and (iii) ideal dTpdT····dApdA vs. 3BSE dTpdT····dApdA. Obviously case (iii) should be mainly due to case (ii), but the presence of dApdA further leads to a strongly dampened oscillator strength of the Type-B2 (from 0.4176 to 0.2288), not observed in the Type-B1 of the ideal case (from 0.4242 to 0.4234). Thus, both environmental and conformational effects here play a role and cooperate with each other in mediating the properties of the ${}^{1}\pi\pi^{*}$ -character local Type-Bs.

The remaining higher-lying excitations are similarly sensi- TD-ωB97X predicts

tive to conformational fluctuation such that their transition origins have different mixtures of local Type-D/E and CT Type-B transition components. Overall, it is observed that conformational influence can

be the same or even outstrip environmental effects on the excitation properties of the SO-hosted excitations.

3.7. Detailed backbone-orbital involvement in the first several excitations of B-DNA segments containing backbone

Fig. 4a shows the percentage of backbone-orbital density contribution to NTO1-H in the first several electronic excitations (not only SO-hosted ones) of the backbone-included systems extracted from ideal B-DNA predicted by TDDFT with different functionals. Evidently TD-B3LYP underestimates the absorption energy of backbone-to-base excitations crowding around the first several excitations. However, even LC hybrid functionals predict an impressive involvement of backbone-orbitals in several excitations. The diverse predictions of these functionals are mainly due to their different mixtures of Hartree–Fock (HF) exchange in their functional forms.

As shown in eqn (2), the $L(r_{12})$ in a LC hybrid functional can be interpreted as an r_{12} -dependent proportion of HF exchange. For example, both ω B97 and LC- ω PBE adopt $L(r_{12}) = \operatorname{erf}(\omega r_{12})$, with $\omega = 0.4$ Bohr⁻¹, while ω B97X adopts $L(r_{12}) = a_x \times [1 - \operatorname{erf}(\omega r_{12})] + \operatorname{erf}(\omega r_{12})$, with $\omega = 0.3$ Bohr⁻¹ and $a_x = 0.157706$. By contrast, a global hybrid functional, containing a fixed proportion a_x of HF exchange, adopts $L(r_{12}) = a_x$ and a pure DFA, containing no HF exchange, adopts $L(r_{12}) = 0$.

To look into more detail, we plot, in Fig. 5, the r_{12} dependent proportion of HF exchange $L(r_{12})$, for various functionals. Evidently we observe that there is a strong correlation between the $L(r_{12})$ in a LC hybrid functional and the appearance of significant backbone-to-base CT contributions to excitations. As the $L(r_{12})$ is larger for a functional in the range of inter-electronic separation $r_{12} > 1$ Bohr for the electron-hole interaction, the first excitation that has a significant contribution from the backbone to NTO1-H appears at higher absorption energy.



Fig. 5 $L(r_{12})$ vs. r_{12} for different functionals. For details, please see the text and eqn (1)–(3).

TD- ω B97X predicts that strongly backbone involved excitations are located below ~180 nm for all the 4 systems with distinctive sizes. A magnified plot for the results at the wavelength range in which backbone intensely participates is given in Fig. 4b. Clearly, except for dT, for the other 3 systems most excitations with noticeable absorption strength either have pronounced backbone involvement in them, or are embedded within the jungles of strongly backbone-involved dark absorptions that can facilitate energy transfer among them, capable of leading to, for instance, non-radiative energy dissipations.

Given the noticeable diversity among the results of different LC hybrid functionals, a comprehensive assessment of their performance for these larger electronic systems is important, for which more benchmarks are in need.

Nevertheless, up to this point, we could still conclude that the DNA backbone is quantum mechanically quite important upon UV light absorption or the dynamics afterwards.

4. Conclusion

We have examined several low-lying electronic excitations of various Thy-contained systems extracted from two sources of B-DNA, *i.e.* the ideal and an X-ray crystal structure. Various environmental and conformational effects on Thy-related excitations have been investigated quantum mechanically. The latter and former effects are studied by comparing the LC-TDDFT results of two systems that are different from each other in geometry and in a physical/chemical surrounding, respectively. Through the employment of QNTO analysis, the transition origin of up to hundred electronic excitations of different systems can readily be determined and therefore similar excitations can be identified and compared with each other.

We find that although many low-lying SO-hosted $\pi\pi\pi^*$ and ${}^{1}n\pi^{*}$ -character excitations (or combinations of them) of Thy-contained systems can be interpreted by single NTO1, there are still several cases that NTO1 has less than 70% domination, in which case NTO2 is also taken into account. Therefore the present study provides a more realistic description of the molecular exciton view in which the contribution of the local electronic transitions proportion can be different. Moreover, decomposition of NTO1(2)-H(E) of an excitation into standard-orbitals helps pinpoint its transition origin. It is possible that an excitation is a mixture of different local and CT transitions. It is also found that as the segment size grows to include the backbone or WC-pairing or adjacent stacking nucleobases, the NTO1 transition origin is often more complicated because various orbital-components are mixed; the NTO1 domination is also likely to decrease at the same time. The DNA backbone, even though not donating orbitals, is also found to be capable of mediating the transition origins of SO-hosted excitations. Our analysis provides a quantitative investigation foundation for these complex phenomena.

Both conformational and environmental effects on SO-hosted excitations are found to mediate not only the absorption energy and oscillator strength but also the NTO1 transition origin. This situation is significant especially for the higher-lying excitations. The higher-lying local Type-D, E and CT Type-B transitions are more sensitive to environmental surroundings as well as conformational variations than local Type-A, B and C transitions such that they often bear various mixtures to form the transition origin of excitations around 175 nm-190 nm. This situation probably results from their near-degenerate feature of transitions. In addition, even though local Type-Bs are relatively intact in their transition origins, their oscillator strengths can be easily mediated by environmental and conformational factors. By plotting backbone-orbital contribution for the first several. SO-hosted or non-SO-hosted, excitations of various backbone-included systems predicted by TD-ωB97X, we also find that a dense population of backbone-involved excitations is located at a wavelength below ~ 180 nm, irrespective of the economic basis set used, as shown in Table S2 and Fig. S2 (ESI[†]).

It should, however, be noted that the simulated excitation properties of these Thy-comprised systems are contingent to several factors such as the used calculation methods or other yet reflected molecular environments. For instance, a larger basis set than the current one is anticipated to cause some redshifts of simulated absorption energies, especially for ${}^{1}\pi\pi^{*}$ excitations. In Table S2 (ESI[†]), we have therefore tested the effect of the basis set, geometry relaxation, and PCM solvent on the excitation properties of ideal dTpdT. A general conclusion is that the lower-lying local Type-As, Type-Bs, and Type-Cs are less sensitive to the used basis set and geometry relaxation as opposed to the higher-lying excitations formed by local Type-D/E and CT Type-B transitions. On the other hand, the PCM solvent surrounding dTpdT is found to be capable of resulting in electronic excitations with qualitatively different transition origins where backbone-orbitals often strongly participate. However, because the solvent is only classically taken into account, further quantum mechanical studies on H₂O molecules explicitly present are still needed; the same situation applies for other molecular environments, e.g. ions.

Overall, the present study suggests that a DNA molecule continually under various interactions with itself or environmental molecules in solution can exhibit from time to time a slightly or very different transition origin of NTO1(2) depending on its experience at the instance of light-absorptions. It has also been reviewed in the chapter 14 of ref. 36 that various electronic properties of nucleobases are notably dependent on their microscopic environment and the present study helps further provide a systematic quantum chemistry framework to investigate them. Moreover, the origins of two electronic excitations of the same or different systems can be, as expressed by standard-orbitals, quite different even if their absorption energy and oscillator strength are similar to each other, which implies that a further determination of the correct transition origin of an excitation is quite important which governs later dynamics of an electronically excited molecule. Excitations can be distinguished in the present framework and therefore a series of potential energy surfaces of a gradually geometrically altered system within a complex environment is forthcoming.

numerous factors can affect experimentally observed absorption maxima¹⁰⁸ – the best estimate for the uracil ${}^{1}\pi\pi^{*}$ absorption energy from cross-checking highly accurate theoretical calculations has still predicted a higher value than the observed spectrum maximum by ~ 0.17 eV,¹⁰⁹ and is attributed by the authors to strong vibronic coupling with the nearby ${}^{1}n\pi^{*}$ excitation. Our present work is focused on TD- ω B97X, which has been shown to well reproduce the excitation properties of Thy and dT given by the highly accurate EOM-CCSD,⁹⁵ and therefore is believed to be highly reliable. For an even better prediction of excitations, however, calculation with a larger basis set and/or finer tuning of DFT/ TDDFT functionals are still needed, especially for higherlying excitations. This may be achieved from cross-referencing results from high-level theoretical methods and experiments; the latter may require a more realistic condition in calculation such as the inclusion of solvent molecules in order to draw a direct comparison. These labours naturally lie on the difficulties of studying large and complex electronic systems where many aspects have to be considered in practicable computations.

Acknowledgements

We would like to thank National Science Council (NSC), and National Center for Theoretical Sciences (NCTS) for their financial supports. We would also like to thank National Center for High-performance Computing (NCHC) for providing computational resources.

References

- D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, W. H. Freeman & Company, New York, 3rd edn, 2000.
- 2 R. F. Weaver, *Molecular Biology*, McGraw-Hill, New York, 4th edn, 2007.
- 3 P. W. Rothemund, Nature, 2006, 440, 297.
- 4 E. S. Andersen, M. Dong and M. M. Nielsen, *Nature*, 2009, 459(7243), 73.
- 5 Y. Ishitsuka and T. Ha, Nat. Nanotechnol., 2009, 4, 281.
- 6 F. A. Aldaye, A. L. Palmer and H. F. Sleiman, *Science*, 2008, 321, 1795.
- 7 C. E. Crespo-Hernandez, B. Cohen, P. M. Hare and B. Kohler, *Chem. Rev.*, 2004, **104**, 1977.
- 8 M. K. Shukla and J. Leszczynski, J. Biomol. Struct. Dyn., 2007, 25, 93.
- 9 T. Inagaki, A. Ito, K. Heida and T. Ho, *Photochem. Photobiol.*, 1986, 44, 303.
- 10 J. S. Novros and L. B. Clark, J. Phys. Chem., 1986, 90, 5666.
- 11 Y. Matsuoks and B. Norden, J. Phys. Chem., 1982, 86, 1378.
- 12 A. Holmen, A. Broo and B. Albinsson, J. Phys. Chem., 1994, 98, 4998.
- 13 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *J. Chem. Phys.*, 2001, **115**, 4604.
- 14 M. Mons, I. Dimicoli, F. Piuzzi, B. Tardivel and M. Elhanine, J. Phys. Chem. A, 2002, 106, 5088.
- 15 M. Mons, F. Piuzzi, I. Dimicoli, L. Gorb and J. Leszczynki, J. Phys. Chem. A, 2006, 110, 10921.
- 16 M. K. Shukla and J. Leszczynski, Chem. Phys. Lett., 2006, 429, 261.
- 17 C. A. Sprecher and W. C. Johnson Jr., *Biopolymers*, 1977, 16, 2243.
- 18 L. B. Clark, J. Am. Chem. Soc., 1977, 99, 3934.
- 19 L. B. Clark, J. Am. Chem. Soc., 1994, 116, 5265.
- 20 C. Santhosh and P. C. Mishra, J. Mol. Struct., 1989, 198, 327.

- 21 D. W. Miles, S. J. Hann, R. K. Robins and H. Eyring, J. Phys. Chem., 1968, 72, 1483.
- 22 L. B. Clark, J. Phys. Chem., 1989, 93, 5345.
- 23 L. B. Clark, J. Phys. Chem., 1990, 94, 2873.
- 24 A. Holmen, A. Broo, B. Albinsson and B. Norden, J. Am. Chem. Soc., 1997, 119, 12240.
- 25 L. B. Clark, J. Phys. Chem., 1995, 99, 4466.
- 26 N. J. Kim, G. Jeong, Y. S. Kim, J. Sung, S. K. Kim and Y. D. Park, J. Chem. Phys., 2000, 113, 10051.
- 27 D. C. Luhrs, J. Viallon and I. Fischer, *Phys. Chem. Chem. Phys.*, 2001, **3**, 1827.
- 28 E. Nir, K. Kleinermanns, L. Grace and M. S. de Vries, J. Phys. Chem. A, 2001, 105, 5106.
- 29 D. Voet, W. B. Gratzer, R. A. Cox and P. Doty, *Biopolymers*, 1963, 1, 193.
- 30 L. B. Clark and I. Tinoco Jr., J. Am. Chem. Soc., 1965, 87, 11.
- 31 T. Yamada and H. Fukutome, *Biopolymers*, 1968, 6, 43.
- 32 F. Zaloudek, J. S. Novros and L. B. Clark, J. Am. Chem. Soc., 1985, 107, 7344.
- 33 W. C. Johnson Jr., P. M. Vipond and J. C. Girod, *Biopolymers*, 1971, **10**, 923.
- 34 A. Kaito, M. Hatano, T. Ueda and S. Shibuya, Bull. Chem. Soc. Jpn., 1980, 53, 3073.
- 35 K. Raksany and I. Foldvary, Biopolymers, 1978, 17, 887.
- 36 Radiation Induced Molecular Phenomena in Nucleic Acids, ed. M. K. Shukla and J. Leszczynski, Springer-Verlag, 2008.
- 37 M. K. Shukla and J. Leszczynski, *Computational Chemistry: Reviews of Current Trends*, ed. J. Leszczynski, World Scientific, Singapore, 2003, vol. 8, p. 249.
- 38 B. Mennucci, A. Toniolo and J. Tomasi, J. Phys. Chem. A, 2001, 105, 4749.
- 39 M. K. Shukla and P. C. Mishra, Chem. Phys., 1999, 240, 319.
- 40 M. K. Shukla, S. K. Mishra, A. Kumar and P. C. Mishra, J. Comput. Chem., 2000, 21, 826.
- 41 S. K. Mishra, M. K. Shukla and P. C. Mishra, *Spectrochim. Acta*, *Part A*, 2000, 56, 1355.
- 42 A. Broo, J. Phys. Chem. A, 1998, 102, 526.
- 43 M. K. Shukla and J. Leszczynski, J. Phys. Chem. A, 2002, 106, 11338.
- 44 M. P. Fulscher, L. Serrano-Andres and B. O. Roos, J. Am. Chem. Soc., 1997, 119, 6168.
- 45 J. Lorentzon, M. P. Fulscher and B. O. Roos, J. Am. Chem. Soc., 1995, 117, 9265.
- 46 M. P. Fulscher and B. O. Roos, J. Am. Chem. Soc., 1995, 117, 2089.
- 47 T. Fleig, S. Knecht and C. Hattig, J. Phys. Chem. A, 2007, 111, 5482.
- 48 H. H. Ritze, P. Hobza and D. Nachtigallova, *Phys. Chem. Chem. Phys.*, 2007, 9, 1672.
- 49 B. Mennucci, A. Toniolo and J. Tomasi, J. Phys. Chem. A, 2001, 105, 4749.
- 50 B. Mennucci, A. Toniolo and J. Tomasi, J. Phys. Chem. A, 2001, 105, 7126–7134.
- 51 M. K. Shukla and J. Leszczynski, J. Comput. Chem., 2004, 25, 768.
- 52 A. Tsolakidis and E. Kaxiras, J. Phys. Chem. A, 2005, 109, 2373.
- 53 D. Varsano, R. D. Felice, M. A. L. Marques and A. Rubio, *J. Phys. Chem. B*, 2006, **110**, 7129.
- 54 J. Cerny, V. Spirko, M. Mons, P. Hobza and D. Nachtigallova, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3059.
- 55 R. Improta and V. Barone, J. Am. Chem. Soc., 2004, 126, 14320.
- 56 T. Gustavsson, A. Banyasz, E. Lazzarotto, D. Markovitsi, G. Scalmani, M. J. Frisch, V. Barone and R. Improta, J. Am. Chem. Soc., 2006, 128, 607.
- 57 R. So and S. Alavi, J. Comput. Chem., 2007, 28, 1776.
- 58 R. J. Bartlett and M. Musial, Rev. Mod. Phys., 2007, 79, 291.
- 59 H. Sekino and R. J. Bartlett, Int. J. Quantum Chem., Quantum Chem. Symp., 1984, 18, 255.
- 60 J. F. Stanton and R. J. Bartlett, J. Chem. Phys., 1993, 98, 7029.
- 61 E. Runge and E. K. U. Gross, Phys. Rev. Lett., 1984, 52, 997.
- 62 Time-Dependent Density-Functional Theory, ed. M. A. L. Marques, C. A. Ullrich, F. Nogueira, A. Rubio, K. Burke and E. K. U. Gross, Springer-Verlag, 2006.
- 63 M. E. Casida, in *Recent Advances in Density Functional Methods*, ed. D. P. Chong, World Scientific, Singapore, 1995, vol. 1.

- 64 A. Dreuw and M. Head-Gordon, Chem. Rev., 2005, 105, 4009.
- 65 J. Neugebauer, J. Phys. Chem. B, 2008, 112, 2207.
- 66 J. Neugebauer, J. Chem. Phys., 2007, 126, 134116.
- 67 M. Dulak, J. W. Kamiński and T. A. Wesolowski, J. Chem. Theory Comput., 2007, 3, 735.
- 68 J. W. Kaminski, S. Gusarov, T. A. Wesolowski and A. Kovalenko, J. Phys. Chem. A, 2010, 114, 6082.
- 69 M. Dulak, J. W. Kamiński and T. A. Wesolowski, Int. J. Quantum Chem., 2009, 109, 1886.
- 70 A. W. Götz, S. M. Beyhan and L. Visscher, J. Chem. Theory Comput., 2009, 5, 3161.
- 71 J. Neugebauer, J. Chem. Phys., 2009, 131, 084104.
- 72 S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher and M. Reiher, J. Chem. Phys., 2010, **132**, 164101.
- 73 K. Kiewisch, G. Eickerling, M. Reiher and J. Neugebauer, J. Chem. Phys., 2008, 128, 044114.
- 74 J. Neugebauer, ChemPhysChem, 2009, 10, 3148.
- 75 J. Neugebauer, C. Curutchet, A. Muñoz-Losa and B. Mennucci, J. Chem. Theory Comput., 2010, 6, 1843.
- 76 C. R. Jacob and L. Visscher, J. Chem. Phys., 2008, 128, 155102.
- 77 T. A. Wesolowski, J. Am. Chem. Soc., 2004, 126, 11444.
- 78 G. Fradelos, J. W. Kaminski, T. A. Wesolowski and S. Leutwyler, J. Phys. Chem. A, 2009, 113, 9766.
- 79 C. R. Jacob, J. Neugebauer and L. Visscher, J. Comput. Chem., 2008, 29, 1011.
- 80 J. Neugebauer, Phys. Rep., 2010, 489, 1.
- 81 O. A. Vydrov, J. Heyd, A. V. Krukau and G. E. Scuseria, J. Chem. Phys., 2006, 125, 074106.
- 82 T. Yanai, D. Tew and N. Handy, Chem. Phys. Lett., 2004, 393, 51.
- 83 J.-D. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys., 2008, 10, 6615.
- 84 J.-D. Chai and M. Head-Gordon, J. Chem. Phys., 2008, 128, 084106.
- 85 J.-D. Chai and M. Head-Gordon, Chem. Phys. Lett., 2008, 467, 176.
- 86 J.-D. Chai and M. Head-Gordon, J. Chem. Phys., 2009, 131, 174105.
- 87 A. Savin, in *Recent Developments and Applications of Modern Density Functional Theory*, ed. J. M. Seminario, Elsevier, Amsterdam, 1996, p. 327.
- 88 H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, J. Chem. Phys., 2001, 115, 3540.
- 89 Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai and K. Hirao, J. Chem. Phys., 2004, 120, 8425.
- 90 J. A. Parkhill, J.-D. Chai, A. D. Dutoi and M. Head-Gordon, *Chem. Phys. Lett.*, 2009, **478**, 283.
- 91 J.-W. Song, T. Hirosawa, T. Tsuneda and K. Hirao, J. Chem. Phys., 2007, 126, 154105.
- 92 M. A. Rohrdanz and J. M. Herbert, J. Chem. Phys., 2008, 129, 034107.
- 93 D. Jacquemin, E. A. Perpète, I. Ciofini and C. Adamo, *Theor. Chem. Acc.*, 2011, **128**, 127.
- 94 M. Caricato, G. W. Trucks, M. J. Frisch and K. B. Wiberg, J. Chem. Theory Comput., 2011, 7, 456.
- 95 J.-H. Li, J.-D. Chai, G. Y. Guo and M. Hayashi, *Chem. Phys. Lett.*, 2011, **514**, 362.
- 96 A. G. Leslie, S. Arnott, R. Chandrasekaran and R. L. Ratliff, J. Mol. Biol., 1980, 143, 49.
- 97 C. R. Kozak, K. A. Kistler, Z. Lu and S. Matsika, J. Phys. Chem. B, 2010, 114, 1674.
- 98 M. Kasha, Radiat. Res., 1963, 20, 55.
- 99 A. W. Lange and J. M. Herbert, J. Am. Chem. Soc., 2009, 131, 3913.
- 100 J. Tomasi, B. Mennucci and R. Cammi, Chem. Rev., 2005, 105, 2999.
- 101 Ideal B-DNA PDB file adopted from "http://www.biochem.umd. edu/biochem/kahn/teach_res/dna_tutorial/".
- 102 N. Narayana and M. A. Weiss, J. Mol. Biol., 2009, 385, 469.
- 103 T. Schlick, Molecular Modeling and Simulation, Springer, New York, 2002.
- 104 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda,

J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman,

J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision* A.01, Gaussian, Inc., Wallingford, CT, 2009.

- 105 R. Ditchfield, W. J. Hehre and J. A. Pople, J. Chem. Phys., 1971, 54, 724.
- 106 A. D. Becke, J. Chem. Phys., 1993, 98, 5648.
- 107 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785.
- 108 M. Schreiber, M. R. Silva-Junior, S. P. A. Sauer and W. Thiel, J. Chem. Phys., 2008, 128, 134110.
- 109 E. Epifanovsky, K. Kowalski, P.-D. Fan, M. Valiev, S. Matsika and A. I. Krylov, J. Phys. Chem. A, 2008, **112**, 9983.