PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Characterization of the Excited States of DNA Building Blocks: a Coupled Cluster Computational Study†

Zsuzsanna Benda and Péter G. Szalay*

DNA building blocks consisting of up to four nucleobases are investigated with the EOM-CCSD and CC2-LR methods in two B-DNA-like arrangements of a poly-adenine:poly-thymine (poly-A:poly-T) system. Excitation energies and oscillator strengths are presented and the character of the excited states are discussed. Excited states of single-stranded poly-A systems are highly delocalized, especially the spectroscopically bright states, where delocalization over up to four fragments can be observed. In case of poly-T systems, the states are somewhat less delocalized, extending to maximally about three fragments. A single A:T Watson-Crick pair has highly localized states, while delocalization over base pairs can be observed for some excited states of the $(A)_2$:$(T)_2$ system, but intrastrand delocalization is more pronounced in this case, as well. As for the characteristics of the simulated UV absorption spectra, a significant decrease of intensity can be observed in case of single strands with increasing chain length; this is due to the stacking interactions and is in accordance with previous results. On the other hand, the breaking of H-bonds between the two strands does not alter the spectral intensity considerably, it only causes a redshift of the absorption band, thus it is unable to explain the experimentally observed DNA hyperchromism on its own, and stacking interactions need to be considered for the description of this effect as well.

1 Introduction

For undertsanding the photostability of the DNA molecule, one possible way to start is to study the excited states of its UV absorbing building blocks, the nucleobases, and observe the changes due to interaction of these chromophores. Nucleobases have an ultrafast decay mechanism, which prevents harmful changes in their structure. The protection of the DNA molecule might be achieved by nonradiative decay like in the case of nucleobases or by excited-state energy transfer of highly delocalized states[1]. For understanding these processes, the determination of the extent of delocalization and its temporal changes is essential.

When several sufficiently close-lying chromophores are present in a system, resulting from the interaction of the chromophores, combinations of excited states can be formed. Since nucleobase monomers have different excited-state properties, the excited states of a DNA molecule is expected to be dependent on the sequence as well as on the relative position of nucleobases. For the most recent overview of the photophysics of DNA constituents see Ref. 2.

Institute of Chemistry, Eötvös University, H-1518 Budapest, P.O.Box 32, Hungary; E-mail: szalay@chem.elte.hu

† Electronic Supplementary Information (ESI) available: Comparison of the two B-DNA arrangements, CC2 excitation energies and oscillator strengths, additional spectra and diagrams of the analysed quantities. See DOI: 10.1039/b000000x/

In this study, we investigate the excited states of a poly-adenine (A):poly-thymine (T) system in two different arrangements. Both structures correspond to the B-DNA form of poly-A:poly-T, which typically has a narrower minor groove and a wider major groove than mixed-sequence B-DNA[3,4]. Single-stranded poly-A systems are relatively easily prepared and studied experimentally,[5] however, as a single Watson-Crick pair is not stable in solution, the effect of H-bonding cannot be studied separately from the effect of stacking interactions. Thus for the interpretation of experimental findings theoretical calculations are needed.

Two effects that lead to the alteration of the UV spectra of DNA are discussed in this paper. The first one is the DNA hyperchromic effect, which can be observed when heating double-stranded DNA. The increase in UV absorbance is related to the dissociation of double strands into two single strands (see Ref. 6), which means, that the breaking of H-bonds between Watson-Crick pairs and/or conformational changes enhance the photoactivity. The other effect is the decrease of the intensity in the UV spectrum of single-stranded polynucleotides with respect to the spectrum of the nucleobase monomers[7,8], which can be attributed to the stacked interactions between adjacent bases of the chain[9]. As the interaction mainly affects the neighbouring bases, additional fragments decrease the intensity less and less, and at the length of about 15 bases the effect reaches the polynucleotide limit[10].

The presence of excited states delocalized on multiple chromophores (exciton states) in DNA was already discussed in the 1960s[7,9,11,12], however, it has been investigated thoroughly only in the last decade. The size of the system limits the range of methods that are suited for such studies, however, several approaches have been applied for this problem so far.

The QM/MM approach offers a way for treating extended systems and the effect of conformational changes can be excellently considered this way. However, since it is not possible to include more than a few fragments in the QM region with a reasonable accuracy[6,13–15], only delocalized states limited to this small region can be observed with this method. The exciton theory of Markovitsi et al. was applied to larger chains and was able to reproduce several features of the experimental spectra[16–19], but it uses an approximate description of electronic coupling, which can result in its inability to describe some effects, such as the DNA hyperchromic effect[19]. The TDDFT method[20–28] and *ab initio* methods like CC2[1,27–30] and ADC(2)[13,14,30] were applied for short oligonucleotides or for the QM part of a QM/MM calculation to understand changes in the character of excited states and to simulate absorption spectra. Some TDDFT functionals tend to overestimate the stability of charge transfer (CT) states[22,24,26–28], and thus give a false interpretation of excited state processes, whereas CC2 and ADC(2) seem to give a balanced description of Frenkel-type and CT states.

Previous studies have shown, that the formation of delocalized excited states does not result in a significant shift in the spectrum[18,24], and splitting of the absorption bands cannot be observed due to the large number of states in a small energy range[18], which is in accord with experimental findings[12]. It was also mentioned, that the highly delocalized states appear near the absorption maximum[19]. The decrease in intensity and the blueshift of the peak when stacked oligomers are formed can also be observed[24]. TDDFT calculations showed[20], that the simulated spectra change significantly when adding the first few bases to the system, but after 4 bases, it remains practically unchanged. This lead to the conclusion, that the excited states are delocalized over approximately 4 bases[20]. Similar results were presented using the QM/MM approach, where it was stated that it is sufficient to include only 3 bases in the QM region[6].

The measurement of a property which can be connected to the calculated extent of delocalization of the excited state is problematic due to the fast decay of excited states and the limited resolution of experimental techniques[31]. It was shown, that the broadening of the absorption bands are mainly caused by molecular vibrations and conformational changes, and solvation has only a small effect on the bandwidth[32], thus it might be possible to simulate spectra without considering solvent effects.

On the other hand, the hyperchromic effect cannot be modelled by most of the approaches usually applied for nucleobase oligomers, probably due to the limited length of the oligomers that can be investigated or the inadequate description of excited states. D'Abramo et al. were able to model the hyperchromic effect[6], and proposed that the greater extent of delocalization along the chain in single strands than in double strands of poly-A:poly-T systems can explain the spectral changes. CC2 calculations on double-stranded nucleobase tetramers[28] showed, that H-bonding interactions have a less pronounced effect on the absorbance spectrum than stacked interactions.

In this paper, we study systems consisting of a few nucleobase units with high-level coupled-cluster methods to see if a more accurate description of excited states gives more insight into the problem and if experimentally observed effects can be expalined by using a relatively small, static system.

Currently, the highest-level results available are EOM-CCSD and EOM-CCSD(T) calculations on stacked dimers and Watson-Crick pairs[33,34], however, building-blocks of a poly-A:poly-T B-DNA (like stacked AA, TT and the A:T Watson-Crick dimer) have not been investigated yet. The CC2 method has already been used for such dimers[1,27–30], and recently it has been applied to the $(A)_2$:$(T)_2$ tetramer as well[28]. However, longer single stranded oligomers has not been studied yet.

In this paper, based on CC2 and CCSD calculations, we present excitation energies, oscillator strengths and simulated spectra of single-stranded and double-stranded nucleobase oligomers of different lengths, as well as a quantitative study of the character of the excited states using the transition density matrix analysis of Plasser et al[35,36].

2 Methods and computational details

Electronic transitions can be either of Frenkel-type, if the hole and particle orbitals are on the same fragments, or of charge-separated type, if they are located on different fragments. Both types can be localized or delocalized, depending on the number of fragments contributing to the excitation.

The natural transition orbitals (NTOs)[37–39], calculated by the diagonalization of the transition density matrix, give a compact description of the excited state and are useful for understanding the character of the transition. However, for the description of multi-chromophore systems often several NTO pairs are needed, and extracting quantitative information becomes more complicated. For the quantitative description of the excited states the transition density analysis of Plasser et al. can be applied[35,36]. This procedure enables one to assign numerical values to the various properties, describing in what extent the different fragments contribute to the excitation, how large the charge transfer character is, etc. This analysis has already been applied to DNA model systems, like the poly(AT):poly(AT) duplex[13] and stacked dimers in various arrangements[30], to investigate the extent of delocalization and the contribution of CT states to the spectrum.

The central quantity of this analysis is the matrix Ω, which is obtained from the transition density matrix using a procedure similar to the Mulliken-type population analysis[35]. The elements of this matrix (Ω_{AB}) give the probability of the hole being on fragment A while the excited electron is on fragment B. This way the diagonal elements correspond to Frenkel-type excitations, while the off-diagonal elements give the contribution of charge separated states to the studied excited state. The so called electron-hole correlation plots (see later in Figure 2 and 9) are produced by simply plotting the magnitude of the Ω_{AB} matrix elements.

Several other quantities can be calculated from these matrix elements, here we give only a short description of the quantities

frequently used in this paper, to help the reader. For more details see the original publications (Ref. 35 and 36).

The charge transfer character (*CT*) gives the weight of charge-separated states in the excitation, thus it is 0 for Frenkel-type excitations, and 1 for completely charge-separated transitions. The total number of fragments participating in the excitation is described by the participation ratio (*PR*), which is 1 in the case of completely localized states and can be as large as the number of fragments for delocalized states. The coherence length (*COH*) describes the mixing of Frenkel-type and charge-separated states, giving 1 if the excitation is purely of one type. If the fragments are arranged linearly, the average position of the excitation can be given by a single quantity, the position (*POS*), which can take values between 1 and the number of the fragments. The number of NTO pairs needed to describe the excitation is given by the PR_{NTO} quantity. For the excited state analysis the TheoDORE[40] program was used.

Two poly-A:poly-T B-DNA structures were considered in this study; an X-ray crystallographic structure, 1PLY[41], and a DFT optimized structure, WW1[42]. Both systems consist of five A:T Watson-Crick pairs. The two systems and the numbering of the nucleobase fragments are shown in Figure 1. In case of the WW1 structure the monomers lie closer to each other and also their structure is more diverse, while nucleobase fragments have nearly identical geometries in the 1PLY structure.

The sugar-phosphate backbones were removed and the resulting nucleobase ends were capped with hydrogen atoms, thus only nucleobases were calculated. For the study of base pairing effects, a methyl-A:methyl-T Watson-Crick pair was also considered. In that case, the backbones were substituted by methyl groups. For the analysis of oligomers of different length, we built up the strands systematically, upwards from the fourth nucleobase (see Figure 2).

The EOM-CCSD[43] and CC2-LR[44] methods were applied for the study of the excited states of these systems. In all calculations frozen core and the cc-pVDZ[45] basis set was used. A double zeta basis is certainly appropriate for a qualitative analysis performed here. The missing diffuse functions prevent description of Rydberg-states, that we decided to avoid anyway due to the large number of states appearing even in case of dimers. For the RI approximation[46,47] the corresponding (cc-pVDZ) auxiliary basis set[48] was applied.

With the CCSD method only monomers and dimers were calculated, whereas with the CC2 method oligomers of maximally four fragments could be investigated. For adenine monomers three transitions (an $n-\pi^*$ and two $\pi-\pi^*$), for thymine monomers two transitions (an $n-\pi^*$ and a $\pi-\pi^*$) have been considered, as higher transitions either have low oscillator strength or are too high in energy to appear in the investigated range of the absorption spectra.

A simplified notation is used to characterize the transitions. This consists of three parts: the letters *n* and *p* correspond to $n-\pi^*$ and $\pi-\pi^*$ types, respectively, the first number refers to the serial number of the transition (eg. *1p* and *2p* distinguish the first and second $\pi-\pi^*$ transitions of adenine). Finally, the last number separated by hyphen shows on which fragment the state is localized when the monomers are at infinite distance. For example, *2p-1* is a $2\pi-\pi^*$ transition on the first fragment of the "A" system (see Figure 2).

For the study of the effect of base pairing, we considered the third A:T pair of the oligomers. In the 1PLY structure, the distance between the N3 atom of adenine and the H3 atom of thymine is 1.883 Å, while in the WW1 structure it is 1.647 Å. We increased this distance stepwise and investigated the excited states as the function of the distance. For the study of the $(A)_2$:$(T)_2$ tetramer, the second and third A:T pairs were considered.

In the simulated absorption spectra, the bands were modelled by a Gaussian (FWHM=0.3 eV) characterized by an area matching the oscillator strength of the transition. The stick spectra (oscillator strength vs. ΔE) and the modelled absorption bands (intensity vs. ΔE) are plotted simultaneously on the graphs.

Calculations were performed using the CFOUR[49] and the TURBOMOLE[50] program packages. The Molden[51] and the VMD[52] programs were used for the altering of DNA structures.

3 Results and Discussion

3.1 CCSD results for monomers and dimers

In Table 1 excitation energies and oscillator strengths calculated at the CCSD level are listed for the adenine (A) and thymine (T) monomers, for their stacked dimers and for the A:T Watson-Crick pair. The geometries are taken from both the 1PLY and WW1 structures; only the 3rd and 4th monomers are considered.

Let us start with the results on the adenine monomers. The first observation is that in case of the 1PLY structure the excitation energies and oscillator strengths for the two monomers are practically identical, but the difference is also very small in case of the WW1 structure. This means that the geometry of the nucleobases does not depend much on the position in the oligomers. The order of the states is $1\pi-\pi^*$, $n-\pi^*$, $2\pi-\pi^*$ which is consistent with previous results (see e.g.[53]). Comparing to the CC2 results (Table S1 in Electronic Supplementary Information (ESI)[†]), one observes that CC2 excitation energies are somewhat lower (by 0.2-0.3 eV) and the order of the states is different, with $n-\pi^*$ being the lowest. This also confirms earlier observations[53]. As well known, the $n-\pi^*$ and the $1\pi-\pi^*$ transitions are practically dark, while the $2\pi-\pi^*$ transition shows substantial oscillator strength, therefore will dominate the spectrum.

In case of thymine very similar conclusions can be drawn: excitation energies of different fragments are very close (practically identical in case of 1PLY), and CC2 excitation energies (see Figure 5) are again lower than CCSD ones. The order of the states is the same with the two methods, $n-\pi^*$ being the lowest. The $n-\pi^*$ transition is dark, the $\pi-\pi^*$ is almost as bright as the $2\pi-\pi^*$ transition of adenine.

In case of the dimers, as expected, there is a slight splitting of the transition energies. All transition energies are lower than those of the corresponding monomer transitions, except for one of the $2\pi-\pi^*$ states of adenine pair and one of the $\pi-\pi^*$ states of thymine pair, which are slightly higher than in the monomers (Table 1). The oscillator strength is distributed unevenly between the components; the highest energy $2\pi-\pi^*$ transition of adenine

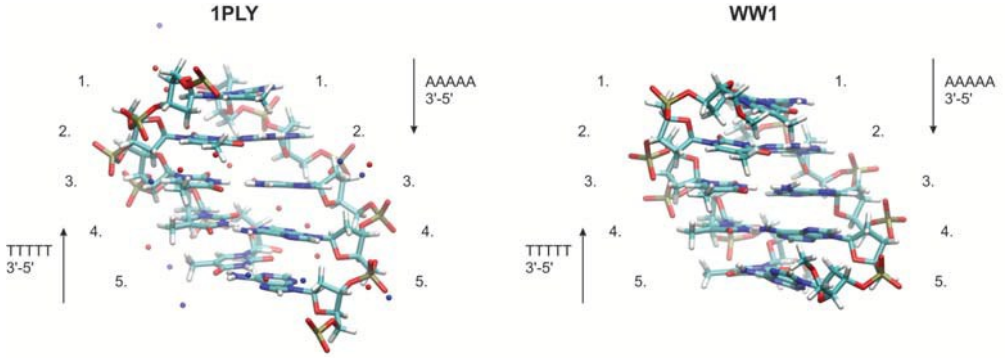

Fig. 1 The two poly-A:poly-T B-DNA structures used in this study along with the numbering of the nucleobase fragments.

Table 1 Excitation energies (ΔE /eV) and oscillator strengths (f) of the studied excited states of adenine and thymine monomers and dimers in different arrangements, calculated by the CCSD method. In case of monomers, the properties of the monomeric excited states are given for both the third and the fourth fragment, separated by a slash

		1PLY		WW1	
	State	ΔE /eV	f	ΔE /eV	f
A	$1\pi-\pi^*$	5.621/5.621	0.002/0.002	5.349/5.344	0.003/0.007
	$n-\pi^*$	5.672/5.672	0.001/0.001	5.372/5.393	0.004/0.001
	$2\pi-\pi^*$	5.928/5.927	0.299/0.299	5.631/5.641	0.308/0.308
$(A)_2$	$1\pi-\pi^*$	5.591	0.000	5.279	0.002
		5.599	0.001	5.301	0.004
	$n-\pi^*$	5.630	0.002	5.294	0.000
		5.639	0.001	5.350	0.001
	$2\pi-\pi^*$	5.759	0.036	5.442	0.044
		5.947	0.399	5.628	0.392
T	$n-\pi^*$	5.128/5.120	0.000/0.000	4.990/4.982	0.000/0.000
	$\pi-\pi^*$	5.714/5.716	0.214/0.214	5.591/5.600	0.220/0.219
$(T)_2$	$n-\pi^*$	5.072	0.000	4.945	0.000
		5.096	0.000	4.961	0.000
	$\pi-\pi^*$	5.573	0.115	5.352	0.092
		5.731	0.242	5.603	0.251
A:T	T $n-\pi^*$	5.346	0.000	5.244	0.001
	A $1\pi-\pi^*$	5.590	0.000	5.344	0.008
	T $\pi-\pi^*$ ‡	5.615	0.184	5.515	0.212
	A $2\pi-\pi^*$ ‡	5.862	0.316	5.599	0.292
	A $n-\pi^*$	5.923	0.002	5.654	0.002

‡ Some mixing between these two states can be observed in the WW1 arrangement.

is still expected to dominate the spectrum (see also later), while in the case of the thymine dimer, the oscillator strength (Table 1) is distributed roughly in a 2:1 ratio between the higher and the lower $\pi-\pi^*$ states. In case of the WW1 structure the $n-\pi^*$ and $1\pi-\pi^*$ transitions of the $(A)_2$ dimer lie very close to each other, even some mixing can be observed.

Base pairing in the A:T Watson-Crick pair has also some effect on the spectrum. As expected[33], $n-\pi^*$ transitions are affected more due to the involvement of the lone pairs in the hydrogen-bonds: the energies of both thymine and adenine $n-\pi^*$ transitions grow by about 0.2-0.3 eV. Energies of the $\pi-\pi^*$ transitions decrease by less then 0.1 eV; and we also observe some mixing between the $\pi-\pi^*$ transition of thymine and $2\pi-\pi^*$ transition of adenine in the WW1 structure. There is small change in the oscillator strengths, which - when considering both structures - does not seem to be systematic with respect to its direction.

3.2 Splitting of the excitation energies in dimers: comparison of CCSD and CC2 results

In Table 2 the magnitude of the splitting of the excitation energies in the dimers can be studied; listed are both CCSD and CC2 results. The first observation is that only the $\pi-\pi^*$ transition of thymine and $2\pi-\pi^*$ transition of adenine show noteworthy splitting. These are the brightest states, in agreement with the Förster theory.

The magnitude of the splitting depends on the structure, the state and slightly also on the method. CCSD and CC2 results are not exactly the same but seem to be systematic. Splitting in case of the WW1 structure is in most cases larger, which is probably due to the slightly shorter distance between the nucleobases (see Figure S1).

Previous CC2 studies on nucleobase dimers presented similar splittings of the excitation energy[1,27,30,34]. For the adenine dimer,

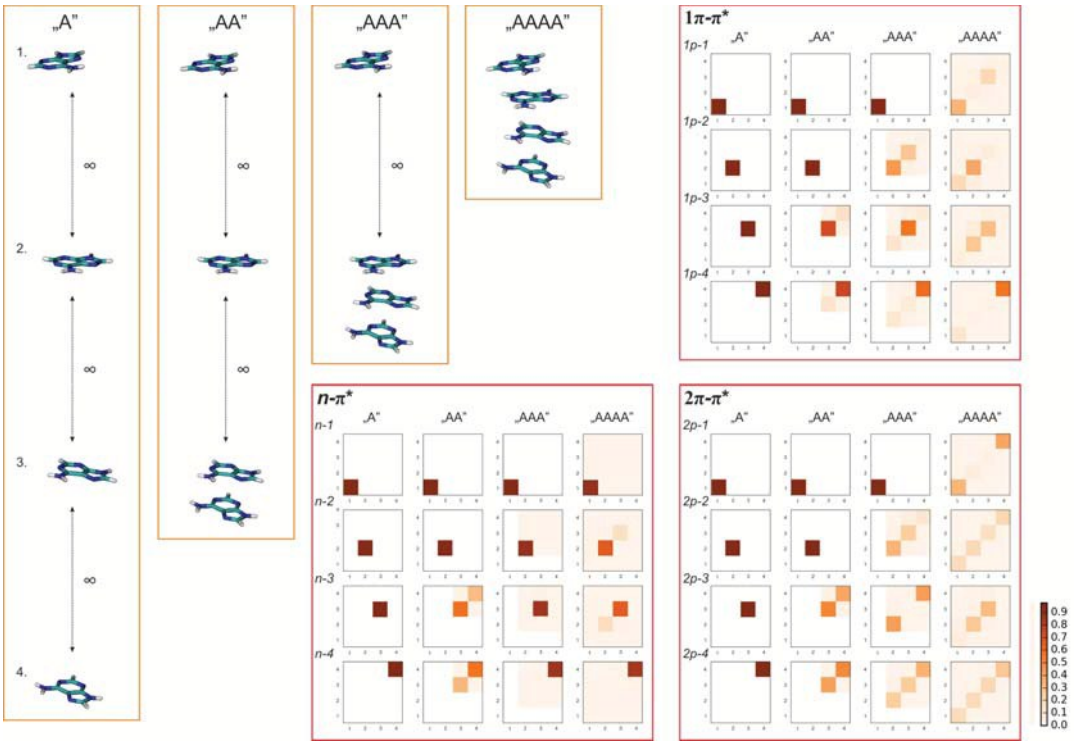

Fig. 2 The arrangement of adenine nucleobases in the investigated systems and the CC2 electron-hole correlation plots of the calculated excited states. The different oligomers were build upwards, starting from the fourth monomer. The system labelled "A" corresponds to four adenine monomers at infinite distance from each other, "AA" to a dimer and two monomers in infinite distance, and so on. The same labels are used in the electron-hole correlation plots and in the diagrams below. In the electron-hole correlation plots the coloring shows the magnitude of the corresponding matrix element. The matrix elements between two fragments in infinite distance is zero by definition, this is indicated by white color. Those elements which were calculated, but are close to zero have a light brown color (see colorbar).

Table 2 Splitting of the excitation energies in the adenine and thymine dimers in different arrangements

Dimer	State	Energy splitting /eV			
		1PLY		WW1	
		CC2	CCSD	CC2	CCSD
$(T)_2$	$n-\pi^*$	0.002	0.037	0.056	0.017
	$\pi-\pi^*$	0.149	0.155	0.242	0.251
$(A)_2$	$n-\pi^*$	0.010	0.009	0.041	0.056
	$1\pi-\pi^*$	0.022	0.007	0.074	0.022
	$2\pi-\pi^*$	0.152	0.188	0.124	0.185

the agreement is particularly good between our 1PLY results and the CC2 results of Lange and Herbert[27] (0.01 eV for the $n-\pi^*$, 0.02 and 0.13 eV for the $\pi-\pi^*$ states). In the case of the thymine dimer, CCSD values are in good agreement with the CC2 values for the $\pi-\pi^*$ state, but the differences are slightly larger for the $n-\pi^*$ state. The results of Ritze et al.[1] are similar to our 1PLY results for the thymine dimer, they calculated 0.021 eV and 0.151 eV splitting for the $n-\pi^*$ and $\pi-\pi^*$ state, respectively. The CC2/cc-pVDZ results of Ramazanov et al.[30] at a B-DNA arrangement show 0.005 eV and 0.136 eV splittings, which are closer to the 1PLY results than to the WW1 results.

3.3 CC2 results for single strands

3.3.1 Poly-A strand.

With the CC2 method the lowest excited states of the poly-A systems are the $n-\pi^*$ states. For the 1PLY structure, the electron-hole correlation plots (Figure 2) show Frenkel-type excitations in all cases, as the off-diagonal Ω matrix elements are negligible. The extent of delocalization can also be estimated by inspecting the structure of the matrices.

For the system "A" where there is no interaction between the monomers, the states are localized which is represented by one brown block at the appropriate position of the Ω matrix. For the dimer ("AA" system), $1\pi-\pi^*$ transitions are mostly localized while in case of $n-\pi^*$ and $2\pi-\pi^*$ transitions there are two significant diagonal elements (corresponding to the third and fourth fragment) in the dimer, which means that both fragments participate in the *n-3* and *n-4* or *2p-3* and *2p-4* excitations, although with different weights. In case of the "AAA" and "AAAA" systems the $n-\pi^*$ transitions are again localized, some delocalization can be observed in case of the $1\pi-\pi^*$ transitions, while the most delocalized are the $2\pi-\pi^*$ states. More detailed analysis can be performed by investigating the *PR* and *POS* values, which give quantitative information about the extent of the delocalization. In the ESI†, diagrams showing the change of these quantities with the chain length (for systems defined in Figure 2) can be seen for

both 1PLY and WW1 structures. In case of the $n-\pi^*$ states, many features of the analysed quantities are similar for the two structures (Figure S2 and S3), but high *PR* values appear at different strand lengths.

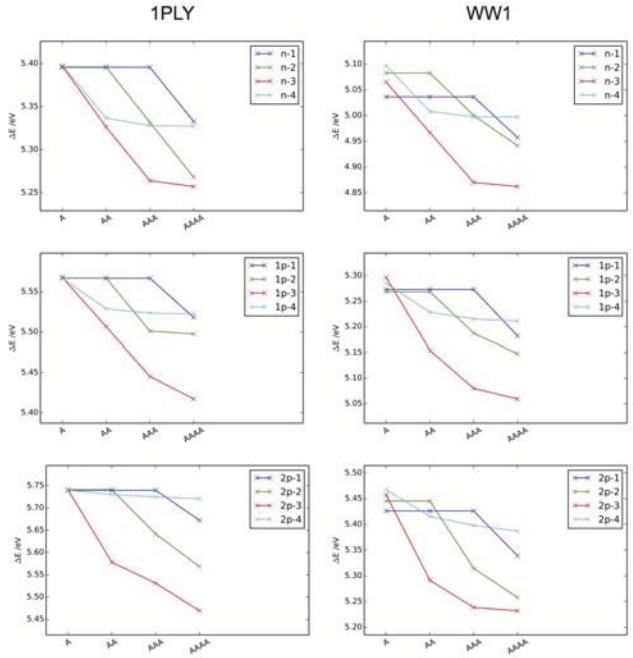

Fig. 3 Transition energies (ΔE /eV) of the excited states of the poly-A systems in the 1PLY (left) and WW1 (right) arrangements, calculated by the CC2 method.

In Figure 3 the variation of transition energies with respect to the number of interacting fragments is given for both structures. In case of 1PLY structure, the $n-\pi^*$ excitation energies form clearly three levels in the *dE* graph; the highest level belonging to the monomers, the middle level to fragments with one stacked neighbour (fragments at the end of the respective strand) and the lowest energy level to fragments with two stacked neighbours (middle ones in the strand). However, in the WW1 structure, the excitation energy graphs does not show the same three-level structure, the fragments with the same number of neighbours are not that close in energy. This may be caused by larger coupling between $n-\pi^*$ states as well as the somewhat more diverse structure of fragments. Nevertheless, it can be concluded, that the energy splitting of these states is rather small (see also Table 2), this implies small interaction between the fragments. Although an increase can be observed in the *COH* and *CT* values with increasing fragment number for both structures (see Figure S2 and S3), these quantities remain close to 1 and 0, respectively, which indicates pure Frenkel-type transitions.

While the $n-\pi^*$ states are separated from the $\pi-\pi^*$ states by an energy gap, the two types of $\pi-\pi^*$ states come close to each other if stacking interaction is present, causing a mixing of $1\pi-\pi^*$ and $2\pi-\pi^*$ states in some cases. Therefore the assignment of the excited states in different oligomers is sometimes not unique. In the problematic cases the assignment is such that the plotted quantities would change more smoothly with the oligomer length.

Even for the description of the $1\pi-\pi^*$ state of the monomers, more than one NTO pairs are necessary ($PR_{NTO} \approx 1.6$). The excited states of the longer oligomers in some cases need more than three NTO pairs for the correct description, and this number can even go up to 4.32 or 3.96 in the 1PLY or WW1 arrangement of the AAAA system (see Figure S4-S7). These states are mainly Frenkel excitonic states, but the contribution of charge separated transitions grow with the system size, the longer oligomers can have states with *CT* around 0.2 and *COH* 1.4-1.5.

The delocalization of the $2\pi-\pi^*$ states is generally higher than that of the the $1\pi-\pi^*$ states, but both grow with increasing fragment number (see Figure S4-S7). The energy splitting of the $1\pi-\pi^*$ states are larger than the splitting of the $n-\pi^*$ states, but still relatively small compared to the $2\pi-\pi^*$ states (Table 2). On the energy graphs (Figure 3) we cannot see the same structure that we saw for $n-\pi^*$ states, this could be attributed to the increased coupling and the mixing of different excitation types, and, in the WW1 structure, also to the diverse structure of fragments. The excited states lie extremely close in the range of 5.46-5.56 eV in the case of the 1PLY and 5.15-5.25 eV in the case of the WW1 structure.

Out of the three types only some of the $2\pi-\pi^*$ type states have considerable oscillator strength. The oscillator strength graphs of the two structures have a similar, branching shape (see Figure 4 for 1PLY results and Figure S7 for WW1 results). The bright states of the monomers combine into a dark and a bright state in the dimer, the latter having larger oscillator strength than the monomers. However, the sum of the oscillator strengths decreases. Adding the third adenine, the oscillator strength of the bright state keeps increasing, while the other states lose some of their intensity. In the tetramer we have two bright and two dark states, the two bright states being highly delocalized with *PR* values 2.65 and 3.87 (1PLY) or 2.45 and 2.97 (WW1).

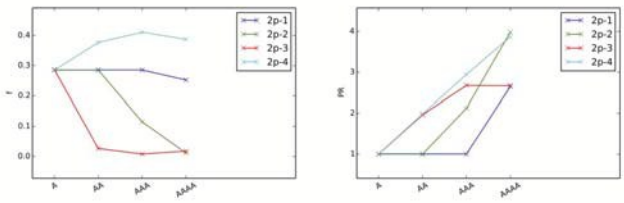

Fig. 4 Oscillator strength (f) and participation ratio (PR) of the $2\pi-\pi^*$ states of the poly-A systems at the 1PLY geometry, calculated by the CC2 method.

3.3.2 Poly-T strand.

The two type of states investigated in poly-T systems (an $n-\pi^*$ and a $\pi-\pi^*$) are separated from each other by a typically 0.4-0.6 eV wide gap, thus mixing of the two types cannot be observed.

The $n-\pi^*$ states are highly localized in all cases, with negligible contribution of the charge separated states (see Figure S8 and S9). In the 1PLY arrangement, the $n-\pi^*$ excitation energy graphs (Figure 5) show the same pattern that we saw in the case of ade-

nine's $n-\pi^*$ states in Figure 3, but in the WW1 arrangement the excitation energies does not follow this trend.

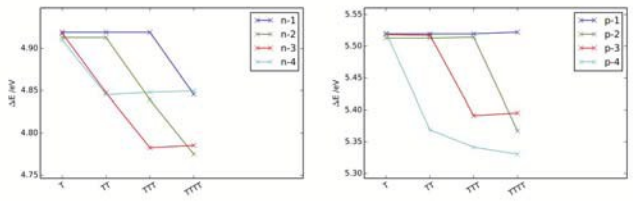

Fig. 5 Transition energies (ΔE /eV) of the excited states of the poly-T systems at the 1PLY geometry, calculated by the CC2 method.

The $\pi-\pi^*$ states have the same kind of branching pattern in the excitation energy graphs (Figure 5 right panel) as the $\pi-\pi^*$ states of adenine (Figure 3). There is an interesting property of the poly-T systems which seems to be independent of the structure, and can be observed in the *PR* graphs (Figure 6). Extending the size of the oligomer, delocalized states are formed, but there is always one state which remains somewhat localized. This state is always localized on the uppermost fragment, and has an excitation energy close to the monomeric excitation energies (5.52 eV in the 1PLY case, see Figure 5). As the serial number of the uppermost bonded fragment changes in every building step (from 4 to 1), the label of this localized state changes also from *p-4* to *p-1*. In the oligomers, this state has a *PR* value around 1.10 in the 1PLY case, and around 1.3 in the WW1 case, all other states have a larger (maximally 3) *PR* value.

The 1PLY oscillator strength graph (left panel of Figure 6) resembles the adenine $2\pi-\pi^*$ graph (Figure 4), but while the oscillator strength of the brightest state (*2p-4*) of adenine grew by a lesser extent in each step, and even dropped at the tetramer, in case of thymine (*p-3* state) it grows by a larger extent in each step. Another important difference is that here the energy of the brightest state decreases significantly for the trimer and the tetramer, which causes a notable redshift in the spectrum (see later in Figure 7).

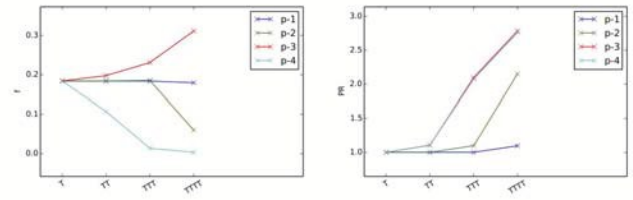

Fig. 6 Oscillator strength (f) and participation ratio (PR) of the $\pi-\pi^*$ states of the poly-T systems at the 1PLY geometry, calculated by the CC2 method.

3.3.3 Simulated spectra.

The calculated spectra of the poly-A and poly-T systems in the 1PLY arrangement can be seen in Figure 7. Since our aim is to model the change of the intensity as the monomers move into a stacked arrangement, the number of the nucleobases were kept constant in the simulated spectra, so the oscillator strengths were normalized to one nucleobase.

In the simulated spectra (Figure 7), a slight shift of the absorption maximum can be observed. It is more pronounced in the case of the poly-T strand, where both methods give a redshift, but in the case of the poly-A strand, there is a slight blueshift with CCSD and a slight redshift with CC2. The larger shift in thymine oligomers are due to the fact that both $\pi-\pi^*$ states of the dimer have large oscillator strength, and one of them stabilizes by a large extent. In the longer oligomers, the lower energy transitions have larger weight than the transitions around the monomeric excitation energies. Contrary to that, the spectrum of the adenine dimer is dominated by one transition, that has similar energy to the monomeric transitions. Since CCSD gives a higher excitation energy for this state, a small blueshift appears in the spectrum. This is consistent with previous computational and experimental results[10,14,24,54], but in larger oligomers the lower energy transitions might gain notable oscillator strength (as we saw for CC2 results), and this could eventually cause a redshift in the CCSD spectra, as well.

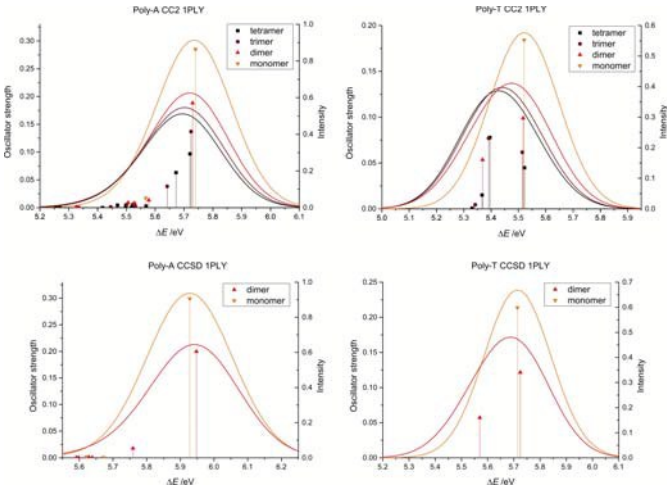

Fig. 7 Simulated absorption spectra of poly-A and poly-T systems of different length, calculated by CC2 and CCSD methods in the 1PLY arrangement of nucleobases. Oscillator strengths are normalized to one nucleobase and absorption bands are represented by FWHM=0.3 eV Gaussians having area equal to the sum of oscillator strengths of the given system.

A substantial decrease of intensity can be observed for both strands with the lengthening of the systems.The CC2 calculations, which could include also trimers and tetramers, show that the sum of oscillator strengths (which is equal to the area of the simulated bands) changes the most between the monomer and dimer, and the decrease seems to converge rapidly (see also Figure S12).

The sum of oscillator strengths calculated with CC2 and CCSD are in very good agreement for the poly-A systems, but for the poly-T system CCSD gives a much larger sum in both arrangements (see Figure S12). Despite this, the percentage of the decrease of the sum of oscillator strengths (Figure 8) is similar with

the two methods.

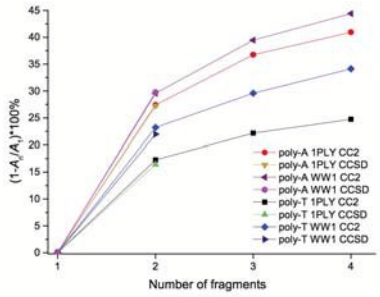

Fig. 8 Hypochromic effect in poly-A and poly-T systems of different length, calculated by CC2 and CCSD methods in the 1PLY and WW1 arrangements of nucleobases. The decrease of absorption intensity with respect to the absorption intensity of the nucleobase monomers is given in percent. For the calculation of the decrease in intensities, oscillator strengths normalized to one nucleobase were used.

The peak area decreased by 27.4% and 27.2% according to the CC2 and CCSD calculations in the case of the 1PLY adenine dimers, which is close to the 24% decrease calculated by Spata et al.[14] with the ADC(2) method, considering more than 200 geometries. This decrease is slightly larger, 29.6% (CC2) and 29.8% (CCSD) for the WW1 structure. For the tetramer, decrease as much as 40.9% (1PLY) and 44.4% (WW1) can be observed with the CC2 method. In the case of poly-T, the hypochromic effect is smaller, and the two structures behave quite differently. The decrease is much larger for the WW1 tetramer (34.1%), than for the 1PLY tetramer (25.0%).

3.4 Watson-Crick pairing

3.4.1 A:T pair.

Forming the A:T pair in the 1PLY Watson-Crick arrangement does not influence the delocalization of the states considerably, the *PR* values does not exceed 1.06 (CC2 result), i.e. the states remain highly localized. The first two states are localized on the T fragment, the next three on the A fragment (see POS values in Figure S13). In the WW1 structure the fragments are closer to each other, which enables the mixing of the A $2\pi-\pi^*$ and the T $\pi-\pi^*$ states. The *PR* values (see Table S2) reflect this mixing, the two states have *PR*=1.64 and 1.58 values, respectively, while for the other three states *PR* is lower than 1.07. In all cases, the *CT* values are close to zero, which indicates pure Frenkel-type states at both geometries.

The excitation energies of the $n-\pi^*$ states of adenine and thymine increase as the fragments move closer, while those of the $\pi-\pi^*$ states decrease. The destabilization of the $n-\pi^*$ states is only about 0.22-0.28 eV, while the stabilization of the $\pi-\pi^*$ states is about 0.03-0.10 eV in both arrangements with the CCSD method, and similar results are obtained with the CC2 method. The same trends were observed by Perun et al.[29] in an MP2 optimized planar arrangement of the nucleobases, and by Lange et al.[27] in a canonical B-DNA arrangement.

The methyl-A:methyl-T pair was also investigated in the 1PLY arrangement, and similar trends were observed for the excitation energies (see Figure S14 for more details). The energy curves of A $n-\pi^*$ and A $\pi-\pi^*$ transitions (plotted in Figure S13 and S14) cross in both A:T and mA:mT systems. At this point, the oscillator strengths and the other quantities from the analysis also become equal for the two states, and the NTOs are a mixture of those of the individual states. In the A:T system there is an additional crossing between the A $n-\pi^*$ and the T $\pi-\pi^*$ states, but *PR* values remain close to 1, showing that there is no delocalization between the fragments.

In the search for charge transfer states, two states, the eighth and ninth excited states of the 1PLY A:T pair, were found that are a nearly equal mixture of a T $\pi-\pi^*$ and an A $\rightarrow$ T charge transfer state (Figure 9). These lie 0.63 and 0.75 eV above the bright $2\pi-\pi^*$ state localized on adenine. Both partially CT states can be described by a single pair of NTOs (see PR_{NTO} values in Figure 9). The hole orbitals are the positive and negative combination of an A π and a T π orbital, and the final orbital is the same T π^* orbital. The electron-hole correlation plots (right bottom of Figure 9) show that the A-T and T-T matrix elements have similar magnitude, while the *CT* numbers give the exact weights (0.448 and 0.554) of charge transfer states in the transitions.

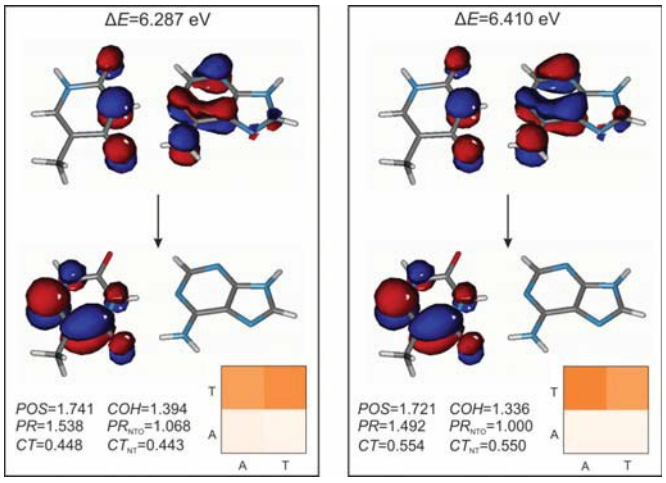

Fig. 9 Mixed locally excited-charge transfer states of the A:T Watson-Crick pair in the 1PLY arrangement, calculated by the CC2 method. In the figure the excitation energies, dominant NTO orbitals, electron-hole correlation matrices and some quantities from the analysis of the transition density matrix are shown.

Earlier calculations[13,14,25–29,33] have already shown that CT states appear at somewhat higher energies than the bright states. For the A:T pair with the CC2 method Lange et al.[27] found an A $\rightarrow$ T CT state 0.6 eV higher, while Perun et al.[29] and Sun et al.[28] found it 0.8 eV and 0.7 eV higher, respectively, than the A $2\pi-\pi^*$ state. These results are in good agreement with the results presented here. Some DFT functionals are not able to describe CT states properly, and give low lying CT states[25,27], while other functionals, like M05-2X and LRC-PBE0, eliminate this problem and give 0.5 eV[25] - 0.7 eV[27] energy differences, in good agreement with the CC2 results.

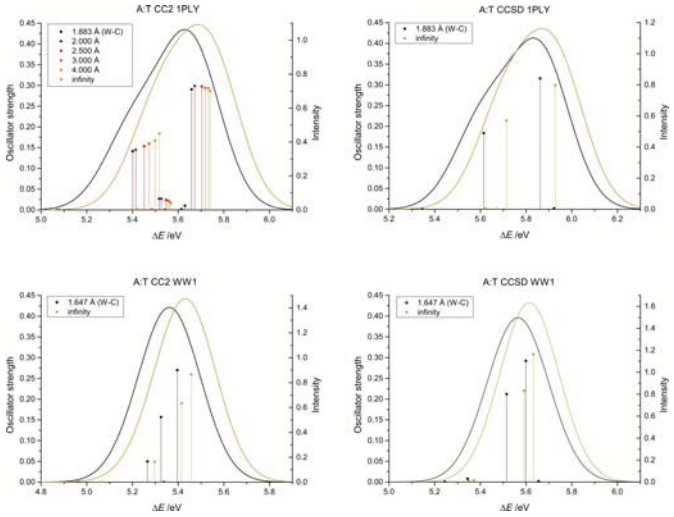

Fig. 10 Simulated absorption spectra of the A:T nucleobase pair for different intermolecular distances, calculated by CC2 and CCSD methods in the 1PLY and WW1 arrangements. The distance between the N3 atom of adenine and the H3 atom of thymine is changed without rotating the fragments. Absorption bands are represented by FWHM=0.3 eV Gaussians having area equal to the sum of oscillator strengths of the given system.

3.4.2 Simulated spectra.

It can be seen on the simulated spectra (Figure 10), that the oscillator strength of the A $2\pi-\pi^*$ state does not change significantly when interacting with thymine in the WC pair (see also Table 1 for CCSD results and Figure S13 for CC2 results). For the 1PLY Watson-Crick geometry it shows a slight increase (0.017 with the CCSD method), while the oscillator strength of the T $\pi-\pi^*$ state shows a somewhat larger decrease (0.030 with CCSD). In the WW1 Watson-Crick arrangement the two bright states are a mixture of the A $2\pi-\pi^*$ and the T $\pi-\pi^*$ states ($PR\approx 1.6$ with CC2), with oscillator strengths closer to each other. This causes a 0.016 and 0.008 decrease in the CCSD oscillator strength of the A $2\pi-\pi^*$ state and the T $\pi-\pi^*$ state, respectively, when moving the fragments to the Watson-Crick position. The investigation of the 1PLY methylated A:T pair (see Figure S14 and S15 for data and simulated spectrum) shows that the oscillator strengths behave similarly to A:T, although the A $1\pi-\pi^*$ state takes some of the oscillator strength of the A $2\pi-\pi^*$ state.

The excitation energy of both bright states decreases when the fragments get closer, but the stabilization is larger for the T $\pi-\pi^*$ state. This causes a broadening of the simulated peak (Figure 10), and a shoulder in the 1PLY spectra, that appears on the low-energy side of the peak[25]. However, as the sum of the oscillator strengths does not change significantly with the distance of the fragments in none of the investigated cases, these model systems cannot reproduce the DNA hyperchromic effect observed experimentally[6].

3.5 $(A)_2$:$(T)_2$ tetramer.

To investigate the effect of base pairing and base stacking simultaneously, we calculated the first ten excited states of the $(A)_2$:$(T)_2$ tetramer in the WW1 arrangement. The $n-\pi^*$ states of the thymine fragments are the lowest excited states here as well (Table 3), whereas the $n-\pi^*$ states of the adenine fragments are at higher energies. All $n-\pi^*$ states are highly localized and do not contribute to the absorption spectrum. However, the $\pi-\pi^*$ states are mostly delocalized over at least two monomers, and mixing of different monomeric excited states (for example A $1\pi-\pi^*$ with A $2\pi-\pi^*$ or T $\pi-\pi^*$) can be observed in many cases. It is hard to classify the states using the NTOs, so in Table 3 we only show which monomers participate in the transition, and whether it is an $n-\pi^*$ or a $\pi-\pi^*$ type transition. If more than one monomeric state can be associated with a state, the weight of the monomeric states decreases in the given order.

As the studied system is no longer linear, the *POS* values would not be very meaningful in this case. However, one can divide the system into two parts, and calculate *POS* and *PR* values considering the resulting two-fragment subsystems. This partition can be done in two ways, by applying either the 'W-C' or the 'dimer' partition. According to the 'W-C' partition, the first fragment is the upper A:T pair (which contains the 2^{nd} A and the 2^{nd} T monomers) and the second is the lower A:T pair (3^{rd} A and 3^{rd} T monomers). Following the 'dimer' partition, the $(A)_2$ dimer is considered as the first fragment and the $(T)_2$ dimer is considered as the second one. Using the previous definitions, the vertical and horizontal delocalization can be investigated separately.

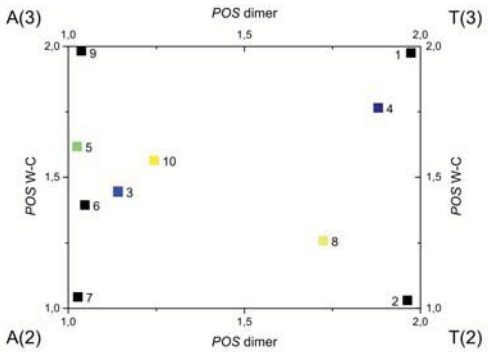

Fig. 11 The position of excitations in the $(A)_2$:$(T)_2$ tetramer determined by the *POS* dimer and *POS* W-C quantities in Table 3. The numbering of states is the same as in Table 3 and the color represents the oscillator strength of the excitation, changing from black for the dark states to yellow for the brightest state. At the corners of the diagram the nucleobase monomers that correspond to the given (minimal or maximal) values of the *POS* dimer and *POS* W-C quantities are also given.

As can be seen from the '*PR* W-C' values in Table 3, in many cases more than one W-C pair participates in the excitation, which means that many states are delocalized vertically (see also '*POS* W-C' in Figure 11). However, there are only a few states which are delocalized horizontally (with more than one stacked dimer participating in the excitation, see '*PR* dimer' and '*POS* dimer' values). Also, this horizontal delocalization is less pronounced,

Table 3 Excitation energy (ΔE /eV) and oscillator strength (f) of the first ten excited states of the $(A)_2$:$(T)_2$ tetramer in the WW1 geometry, calculated with the CC2 method. The position (*POS*) and participation ratio (*PR*) is also given for these states applying different partitions; in the 'W-C' partition, the upper A:T pair (from the 2^{nd} A and the 2^{nd} T monomers) is the first fragment and the lower A:T pair (3^{rd} A and 3^{rd} T monomers) is the second fragment, in the the 'dimer' partition the first fragment is the $(A)_2$ dimer and the second is the $(T)_2$ dimer. In the second column, those monomeric excited states (with the number of the fragment in parenthesis) are listed, that have a large contribution to the excitation

State	Type	ΔE /eV	f	*POS* W-C	*PR* W-C	*POS* dimer	*PR* dimer
1	T(3) $n-\pi^*$	4.825	0.000	1.975	1.052	1.973	1.055
2	T(2) $n-\pi^*$	4.995	0.000	1.031	1.065	1.963	1.078
3	A(2)+A(3)+T(2) $\pi-\pi^*$	5.095	0.062	1.446	1.976	1.141	1.319
4	T(3) $\pi-\pi^*$	5.138	0.016	1.765	1.561	1.880	1.268
5	A(3)+A(2) $\pi-\pi^*$	5.173	0.129	1.617	1.895	1.025	1.051
6	A(2)+A(3) $\pi-\pi^*$	5.186	0.008	1.396	1.916	1.047	1.099
7	A(2) $n-\pi^*$	5.223	0.001	1.044	1.092	1.027	1.056
8	T(2)+A(3)+A(2) $\pi-\pi^*$	5.242	0.233	1.259	1.623	1.724	1.665
9	A(3) $n-\pi^*$	5.305	0.003	1.983	1.034	1.037	1.079
10	A(3)+A(2)+T(2) $\pi-\pi^*$	5.336	0.261	1.564	1.968	1.244	1.584

only two states have '*PR* dimer' values around 1.6 and two around 1.3, while four states have '*PR* W-C' values around 1.9 and two around 1.6. It is interesting to see in Figure 11, that the two brightest states are the mostly delocalized horizontally.

In Figure 11 the $n-\pi^*$ states appear at the corners of the diagram, which again shows that they are highly localized. As for the $\pi-\pi^*$ states, the vertical delocalization is more significant over stacked adenines than over thymines (most states appear midway between the adenine fragments in Figure 11). The horizontal delocalization seems to be more significant over the second A:T pair than over the third A:T pair.

The absorption spectrum of the $(A)_2$:$(T)_2$ tetramer (Figure 12, red curve) is mainly formed by four transitions. These excited states are delocalized over the adenine fragments and the T(2) fragment to a different extent (see Table 3). The sum of the oscillator strengths for the tetramer is much smaller than the sum for the two A:T pairs in infinite distance (left side of Figure 12, black curve), showing the role of the stacking interactions in the hypochromic effect. The same observations were made for single strands (see Figure 8). The decrease of the peak area is 27.3% from the A:T pairs to the tetramer, and a large redshift of the band maximum can be observed.

The $(A)_2$:$(T)_2$ spectrum shows an even larger redshift with respect to the spectrum of $(A)_2$ and $(T)_2$ dimers at infinite distance (right side of Figure 12, black curve). However, the peak areas are nearly the same for the two systems, thus no hyperchromic effect can be observed when the hydrogen-bonds are broken and the strands are moved apart. These findings are similar to those for the separation of the A:T pair (Figure 10) and to the results of Sun et al.[28]. This indicates that the DNA hyperchromic effect cannot be attributed simply to the breaking of the hydrogen bonds, changes in the structure of the strands, thus the alteration of stacking arrangements might play the key role in this effect. We saw previously, that in the limit of free nucleobases the absorption intensity is larger than in B-DNA form oligonucleotides. This could give an explanation to the hyperchromic effect, as the bases are less restricted (thus closer to the free nucleobase model) in single-stranded DNA structures than in double-stranded structures.

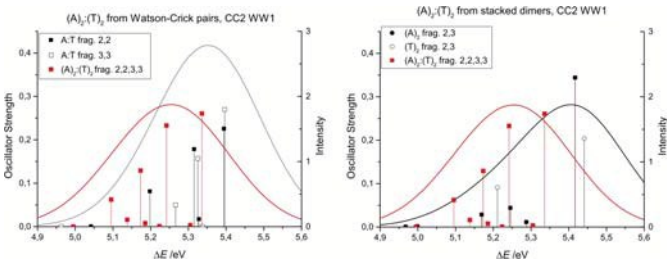

Fig. 12 Absorption spectrum of the $(A)_2$:$(T)_2$ tetramer in the WW1 geometry, calculated with the CC2 method (red curves), compared to the sum of the spectra of the two separate Watson-Crick pairs (left figure, black curve) and to the sum of the spectra of the two separate stacked base pairs (right figure, black curve).

4 Conclusions

The extent of delocalization of the excited states of DNA, as well as their charge transfer character influences the behaviour of the molecule after ultraviolet photoexcitation, thus the precise determination of these properties is of fundamental importance for understanding the photostability of DNA.

In the present study, we have examined a poly-adenine:polithymine double helix in an experimental and also in a quantum chemically optimized B-DNA arrangement by coupled-cluster methods. CCSD excitation energies and oscillator strengths for $(A)_2$, $(T)_2$ and A:T dimers in the B-DNA arrangement and CC2 results for single-stranded nucleobase trimers and tetramers were presented here for the first time. Quantities describing the extent of delocalization and the charge transfer character of the excited states were calculated by analysing the transition density matrices. The alteration of these quantities with the length of the single strands, and during the formation of Watson-Crick pairs was examined.

It was found, that in poly-A and poly-T single strands the energy splitting of the $n-\pi^*$ type states is rather small, and in most cases these states are localized on one base. The splitting of the spectroscopically bright $\pi-\pi^*$ states, on the other hand, is quite large, and these states are delocalized over up to all four bases. However, one can find localized states among the $\pi-\pi^*$ states as

well.

In the simulated spectra, a large decrease (hypochromic effect) in intensity can be observed while increasing the length of single strands, which is in agreement with previous experimental and computational findings both qualitatively and quantitatively, and is attributed to the interaction of stacked bases. There are only slight shifts in the absorption maxima, and the shift is more pronounced in the case of the poly-T strand. There is little interstrand delocalization between fragments of the A:T Watson-Crick pair, but intrastrand delocalization can be observed in the $(A)_2$:$(T)_2$ tetramer.

For the A:T system, two states could be located with partial charge transfer character 0.6-0.7 eV above the bright state of the A fragment. These two states are nearly equal mixtures of an A→T charge transfer state and a T $\pi - \pi^*$ state.

In some of the simulated spectra of the A:T pair, a shoulder can be observed on the low-energy side of the most intense band. This is caused by the larger stabilization of the $\pi - \pi^*$ state of thymine than the ones of adenine. When increasing the distance between the bases, this shoulder disappears and the absorption maximum is shifted to higher energies.

When moving two A:T pairs in a stacked position, a hypochromic effect can be observed similarly to single strands, demonstrating the effect of intrastrand interactions.

By breaking the H-bonds between the bases, the sum of oscillator strengths does not change considerably either in the case of A:T or in the case of $(A)_2$:$(T)_2$, thus the hyperchromism of DNA (the growth of intensity while separating the two strands) cannot be described by the vanishing H-bonding interactions alone, the alteration of stacking interactions also has to be considered.

Acknowledgement

This work has been supported by Orszagos Tudomanyos Kutatasi Alap (OTKA; Grant No. 104672).

References

1 H.-H. Ritze, P. Hobza and D. Nachtigallova, *Physical Chemistry Chemical Physics*, 2007, **9**, 1672–1675.

2 R. Improta, F. Santoro and L. Blancafort, *Chemical Reviews*, 2016, **116**, 3540–3593.

3 D. G. Alexeev, A. A. Lipanov and I. Y. Skuratovskii, *Nature*, 1987, **325**, 821–823.

4 S. Arnott, R. Chandrasekaran, I. Hall and L. Puigjaner, *Nucleic Acids Research*, 1983, **11**, 4141–4155.

5 I. Buchvarov, Q. Wang, M. Raytchev, A. Trifonov and T. Fiebig, *Proceedings of the National Academy of Sciences*, 2007, **104**, 4794–4797.

6 M. D'Abramo, C. L. Castellazzi, M. Orozco and A. Amadei, *The Journal of Physical Chemistry B*, 2013, **117**, 8697–8704.

7 A. Rich and I. Tinoco, *Journal of the American Chemical Society*, 1960, **82**, 6409–6411.

8 M. Leng and G. Felsenfeld, *Journal of Molecular Biology*, 1966, **15**, 455–466.

9 W. Rhodes, *Journal of the American Chemical Society*, 1961, **83**, 3609–3617.

10 J. Brahms, A. Michelson and K. V. Holde, *Journal of Molecular Biology*, 1966, **15**, 467–488.

11 I. Tinoco, *Journal of the American Chemical Society*, 1960, **82**, 4785–4790.

12 J. Eisinger and R. G. Shulman, *Science*, 1968, **161**, 1311–1319.

13 F. Plasser, A. J. A. Aquino, W. L. Hase and H. Lischka, *The Journal of Physical Chemistry A*, 2012, **116**, 11151–11160.

14 V. A. Spata and S. Matsika, *The Journal of Physical Chemistry A*, 2014, **118**, 12021–12030.

15 A. A. Voityuk, *Photochemical and Photobiological Sciences*, 2013, **12**, 1303–1309.

16 B. Bouvier, T. Gustavsson, D. Markovitsi and P. Millié, *Chemical Physics*, 2002, **275**, 75–92.

17 E. Emanuele, D. Markovitsi, P. Millié and K. Zakrzewska, *ChemPhysChem*, 2005, **6**, 1387–1392.

18 B. Bouvier, J.-P. Dognon, R. Lavery, D. Markovitsi, P. Millié, D. Onidas and K. Zakrzewska, *The Journal of Physical Chemistry B*, 2003, **107**, 13512–13522.

19 D. Markovitsi, T. Gustavsson and A. Banyasz, *Mutation Research/Reviews in Mutation Research*, 2010, **704**, 21–28.

20 A. Banyasz, T. Gustavsson, D. Onidas, P. Changenet-Barret, D. Markovitsi and R. Improta, *Chemistry – A European Journal*, 2013, **19**, 3762–3774.

21 D. Varsano, R. Di Felice, M. A. L. Marques and A. Rubio, *The Journal of Physical Chemistry B*, 2006, **110**, 7129–7138.

22 M. K. Shukla and J. Leszczynski, *Molecular Physics*, 2010, **108**, 3131–3146.

23 F. Santoro, V. Barone and R. Improta, *Proceedings of the National Academy of Sciences*, 2007, **104**, 9931–9936.

24 R. Improta, *Physical Chemistry Chemical Physics*, 2008, **10**, 2656–2664.

25 F. Santoro, V. Barone and R. Improta, *ChemPhysChem*, 2008, **9**, 2531–2537.

26 A. W. Lange, M. A. Rohrdanz and J. M. Herbert, *The Journal of Physical Chemistry B*, 2008, **112**, 6304–6308.

27 A. W. Lange and J. M. Herbert, *Journal of the American Chemical Society*, 2009, **131**, 3913–3922.

28 H. Sun, S. Zhang, C. Zhong and Z. Sun, *Journal of Computational Chemistry*, 2016, **37**, 684–693.

29 S. Perun, A. L. Sobolewski and W. Domcke, *The Journal of Physical Chemistry A*, 2006, **110**, 9031–9038.

30 R. R. Ramazanov, D. A. Maksimov and A. I. Kononov, *Journal of the American Chemical Society*, 2015, **137**, 11656–11665.

31 C. Su, C. T. Middleton and B. Kohler, *The Journal of Physical Chemistry B*, 2012, **116**, 10266–10274.

32 L. M. Nielsen, S. Ã. Pedersen, M.-B. S. Kirketerp and S. B. Nielsen, *The Journal of Chemical Physics*, 2012, **136**, 064302.

33 P. G. Szalay, T. J. Watson, Jr., A. Perera, V. F. Lotrich and R. J. Bartlett, *The Journal of Physical Chemistry A*, 2013, **117**, 3149–3157.

34 C. R. Kozak, K. A. Kistler, Z. Lu and S. Matsika, *The Journal of Physical Chemistry B*, 2010, **114**, 1674–1683.

35 F. Plasser and H. Lischka, *Journal of Chemical Theory and Computation*, 2012, **8**, 2777–2789.

36 F. Plasser, M. Wormit and A. Dreuw, *The Journal of Chemical Physics*, 2014, **141**, 024106.

37 A. Luzanov, A. Sukhorukov and V. Umanskii, *Theoretical and Experimental Chemistry*, 1976, **10**, 354–361.

38 R. Martin, *The Journal of Chemical Physics*, 2003, **118**, 4775–4777.

39 I. Mayer, *Chemical Physics Letters*, 2007, **437**, 284–286.

40 *TheoDORE: A package for Theoretical Density, Orbital Relaxation and Exciton analysis written by Plasser, F., V1.1 2014,* http://www.theodore-qc.sourceforge.net.

41 R. Chandrasekaran, A. Radha and H.-S. Park, *Acta Crystallographica Section D*, 1995, **51**, 1025–1035.

42 T. Zubatiuk, O. Shishkin, L. Gorb, D. Hovorun and J. Leszczynski, *The Journal of Physical Chemistry B*, 2015, **119**, 381–391.

43 J. F. Stanton and R. J. Bartlett, *The Journal of Chemical Physics*, 1993, **98**, 7029–7039.

44 O. Christiansen, H. Koch and P. Jörgensen, *Chemical Physics Letters*, 1995, **243**, 409–418.

45 T. H. Dunning, *The Journal of Chemical Physics*, 1989, **90**, 1007–1023.

46 C. Hättig and F. Weigend, *The Journal of Chemical Physics*, 2000, **113**, 5154–5161.

47 C. Hättig and A. Köhn, *The Journal of Chemical Physics*, 2002, **117**, 6939–6951.

48 F. Weigend, A. Köhn and C. Hättig, *The Journal of Chemical Physics*, 2002, **116**, 3175–3183.

49 *CFOUR, a quantum chemical program package written by Stanton, J.F.; Gauss, J.F.; Harding, M.E. and Szalay, P.G. with contributions from Auer, A.A.; Bartlett, R.J.; Benedikt, U.; Berger; Bernholdt, D.E.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jonsson, D.; Jusélius, J.; Klein, K.; Lauderdale, W.J.; Matthews, D.; Metzroth, T.; O'Neill, D.P.; Price, D.R.; Prochnow, E.; Ruud, K.; Schiffmann, F.; Stopkowicz, S.; Tajti, A.; Varner, M.E.; Vázquez, J.; Wang, F. and Watts, J.D. and the integral packages MOLECULE (Almlöf, J. and Taylor, P.R.), PROPS (Taylor, P.R.), ABACUS (Helgaker, T.; Jensen, H.J. Aa.; Jørgensen, P. and Olsen, J.), and ECP routines by Mitin, A.V. and van Wüllen, C.*

50 *TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from* http://www.turbomole.com.

51 G. Schaftenaar and J. Noordik, *Journal of Computer-Aided Molecular Design*, 2000, **14**, 123–134.

52 W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33–38.

53 Z. Benda and P. G. Szalay, *The Journal of Physical Chemistry A*, 2014, **118**, 6197–6207.

54 M. Riley, B. Maling and M. J. Chamberlin, *Journal of Molecular Biology*, 1966, **20**, 359–389.