universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Der Einfluss von π-π-Wechselwirkungen und Wasserstoff-brückenbindungen auf die Delokalisation elektronischer Anregung in doppelsträngiger B-DNA“

„The role of π stacking and hydrogen bonding for delocalization of electronic excitation in double-stranded B-DNA“

verfasst von / submitted by

Maximilian Meixner

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (M Sc)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 863
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Biologische Chemie / Biological Chemistry
Betreut von / Supervisor:	Univ.-Prof. Dr. Leticia González

Contents

Acknowledgements

I thank Univ.-Prof. Dr. Leticia González for giving me the opportunity to work on my Master thesis in her group. Furthermore, I would like to thank Dr. Juan Jose Nogueira Perez and Dr. Felix Plasser for supervising this thesis. I could not wish for better partners. Thank you for your constant support, honest advice and help throughout the workflow that kept me on the right track.

Special thanks to the whole research group for helping me with theoretical, practical and administrative problems. I enjoyed our time and activities after work, which I will keep in good memory.

Last but not least, I thank my family and friends, who gave me undoubted emotional and moral support that guided my way to success.

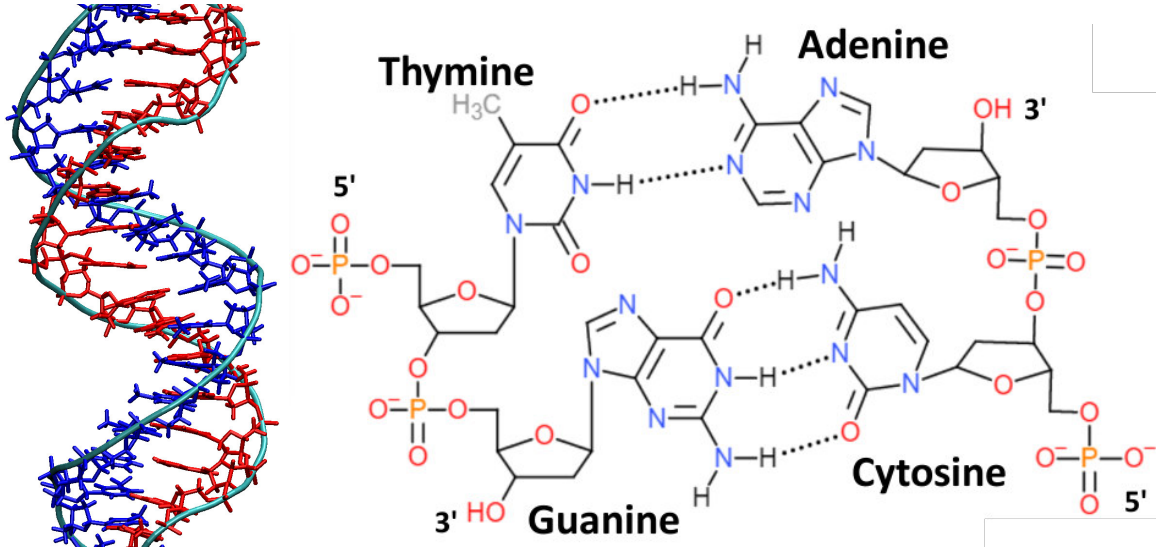
1 Motivation

This introductory chapter explains the constituents and structure of deoxyribonucleic acids (DNA) as well as replication and transcription as examples of fundamental processes maintaining life in order to elucidate the importance of this class of molecules. It is outlined that damaging these biomolecules can cause life threatening consequences. One of these threats is harmful ultra-violet radiation (UVR). The second part of this section concentrates on the photochemistry and photophysics of DNA and gives a description of electronic excitation in nucleobases as part of a double helix. It is pointed out how recent advances in experimental and theoretical approaches try to shed light on the complexity of the system and meet the challange of describing the UV absorption behavior of DNA.

The genetic code is probably the most valuable pattern in nature. Researchers in the early 1940s found that the code contains the neccessary information for the construction of proteins and is stored in genes in the cell nucleus in the form of chromosomes[1]. First advances revealed that chromosomes consist of DNA attached to proteins (chromatin), the former carrying the genetic code[2]. However, how DNA molecules could store the information to build proteins remained unknown. With the discovery of the molecular structure of DNA in 1953 by James Watson and Francis Crick this mystery was about to be solved[3]. Based on X-ray diffraction analysis the scientists assumed that DNA is a double helical structure. Figure 1a shows a cartoon of the three dimensional structure of the duplex. Apart from previous suggestions, they believed that two single strands wind around the same axis. Both chains consist of a sequence of nucleobases derived from nucleic acids and a sugar phosphate backbone. Four different nucleobases are incorporated in DNA, which belong either to the pyrimidine (thymine, T or cytosine, C) or purine (adenine, A or guanine, G) type. Each base is linked to β-D-deoxyribofuranose as shown in Figure 1b. Phosphate bridges are attached to the 3' and 5' oxygen atoms of the sugar residue via phosphodiester bonds and link one segment of the chain to the neighboring units. A unit of nucleobase, sugar and phosphate moiety is called nucleotide.

Each chain builds a right handed helix with the bases in the middle and their plane perpendicular to the helical axis. The backbone structure points outwards making it accessible to solvent molecules (see Figure 1a). When joining for a double helix, the single strands run in opposite directions. Both strands are held together via hydrogen bonds between nucleobases of opponent chains. These interactions are possible between bases of different type. This restricts pyrimidine bases to pair with those of the purine family. A schematic representation of two DNA base pairs is shown in Figure 1b. Additionally, adsorption chromatography of nucleic acids extracted from different microorganisms could prove that the ratio of the amount of adenine equals the one of thymine while guanine was found as abundant as cytosine[4]. This suggests that base pairing is only possible between

adenine-thymine and guanine-cytosine nucleotides. Since both members of a hydrogen bonded base pair belong to nucleobases on different strands, the base sequence of one strand determines the one of the second chain, i.e., they are complementary.



(a) (b)

Figure 1: Structure of DNA. (a) Three dimensional structure of a $(dA)_{20}\cdot(dT)_{20}$ double helix a the molecular dynamics simulation. Color code: adenine nucleotides in red, thymine nucleotides in blue, backbone represented as tubes. (b) Two dimensional structure of two base pairs of DNA showing the four bases (adenine, thymine, guanine, cytosine) and their sugar phosphate backbone. Dotted lines represent hydrogen bonds.

The double-stranded conformation allows transcription and replication by using one or both strands as templates, respectively. Replication is neccessary in order to copy and transfer the whole genome to daughter cells while transcription converts small parts of the code (genes) into ribonucleic acid chains (RNA), which can be further translated into a sequence of amino acids, the building blocks for proteins[1]. The sequence of three successive nucleobases (triplet or codon) determines the type of amino acid. Since 20 proteinogenic amino acids are known, several nucleobase triplets code for the same amino acid. The specific sequence of nucleobase triplets in coding DNA and the corresponding nucleobases in RNA molecules determine the constituents of the amino acid chain and, thus, the structure of the specific protein[5, 6].

These two fundamental functions of DNA are highly regulated as the accuracy of these processes is pivotal for the construction of macromolecules. In other words, protecting the value of the genetic code by maintaining the truth of its content is indespensable for nature. However, failures cannot be excluded completely and harmful modifications alter

the genetic code[1]. Not only mistakes in physiological processes but exogenous effects can cause pathological modifications to DNA. One of these threats is UVR from sunlight. Unrecognized and unrepaired photolesions block replication and transcription of DNA or cause mutations in the genetic code[7]. Long before the discovery of the structure of DNA, it was found that exposure to UVR causes cell death in bacteria due to absorption by specific substances in the cell[8], which could later on be identified as DNA[9]. With wavelengths smaller than visible light, UVR can be classified as UV-C ($<$ 280 nm), UV-B (280 - 315 nm) and UV-A (315 - 380 nm). The high-energy UV-C and large parts of UV-B wavelengths are mostly absorbed by ozone in the atmosphere[10]. The remaining sunlight that reaches earth's surface consists mainly of infrared radiation (55%) and visible light (40%). UV-A (4.7%) and UV-B (0.3%) make the smallest but most harmful fraction of light[11]. Photodamage by UV-A radiation is primarily induced indirectly via non-DNA sensitizers and endogenous oxygen in proximity to DNA[12] and comprises single and double strand breaks of the backbone as well as DNA-protein crosslinks and oxidative damage to nucleobases[13–15]. However, since assays based on high performance liquid chromatography and mass spectrometry could show that UV-A radiation is able to induce direct photodamage to DNA as well, the question of whether UV-A or UV-B is more dangerous has been open to discussion[16].

In the following paragraphs, the two main classes of photoproducts resulting from direct UV-B absorption are described and their biological significance is outlined. They both induce dimer reactions between adjacent pyrimidine bases while non-dimer base damage ocurrs less frequently. Lesions on purine bases are estimated to make up only 1% of the total amount of photoproducts, which is the reason why they are not extensively studied and, thus, together with non-dimer products are not further mentioned here[17].

Early investigations revealed that UV-B irradiated DNA forms cyclobutane pyrimidine dimers (CPDs)[18]. In this reaction the double bond between C5 and C6 of two neighbouring pyrimidine bases form a cyclobutane ring so that the bases dimerize covalently (see Figure 2a). The CPD reaction requires an efficient overlap of the electronic desity of the double bonds. As a consequence, DNA bending and unwinding of the helix are neccessary[19]. A study on curved DNA molecules could prove that CPD formation is enhanced in nucleosome like structures emphasizing that not only naked DNA but also highly condensed chromatin is target for direct photodamage[20]. With CPD specific enzymes and sequencing technology using gel electrophoresis it was found that 5′-TT sequences are most likely targets of photolesions after absorption of UV-B photons. Mixed 5′-CT and 5′-TC sites are more affected than 5′-CC, which is the least favored sequence for CPD formation[21].

Pyrimidine(6-4)pyrimidone adducts or (6-4)photoproducts (6-4PPs) are the second type of products that can be formed due to direct UV-absorption of DNA and are shown in Fig-

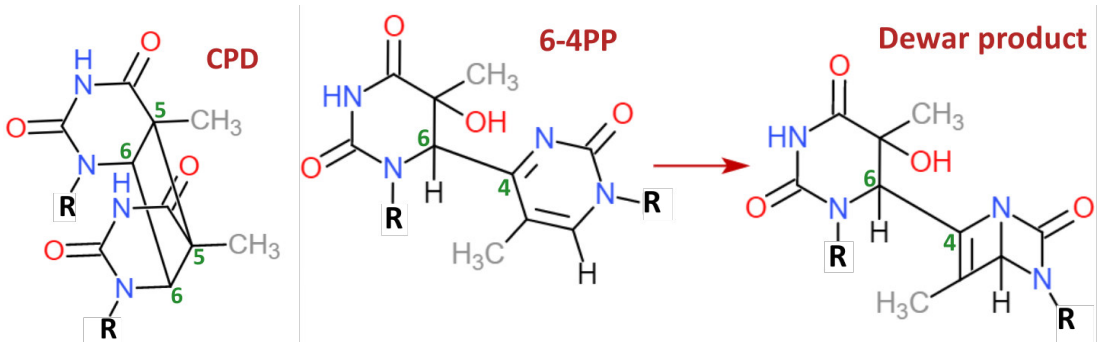


Figure 2: Structural formulars of two types of photoproducts formed upon direct UV-B absorption by DNA exemplary shown in the case of two stacked thymine bases. Atoms involved in the formation of pathological covalent bonds highlighted with green numbers; **R**: backbone. (a) Cyclobutane pyrimidine dimer (CPD). (b) (6-4)Photoproduct (6-4PP) that eventually converts to its Dewar isomer.

ure 2b. The underlying mechanism that result in a covalent bond between position 6 and 4 of adjacent pyrimidine bases is more complicated than CPD formation. The role of these damage products was highly discussed in the early 1990s since it was found that 6-4PPs are capable of photoisomerization to so called Dewar products. It has been shown that the majority of 6-4PPs indeed restructures into Dewar products when the cells are continously exposed to natural sunlight[22, 23]. Similar to CPDs, 6-4PPs require an unwinding of the helical struture of the DNA as well as base rotation. 6-4PPs cause severe distortions in the helical structure of DNA so that Watson-Crick hydrogen bonds cannot longer be formed because ring planes of 6-4PP-bases are almost perpendicular to each other. This makes it less difficult to detect these kind of lesions, which is the reason why they are more easily repaired[24]. Because of the more onerous reaction requirements, 6-4PPs are less abundant lesions compared to CPDs[25]. Unlike CPDs, the 6-4PPs and consequently their Dewar isomers are preferably formed at 5′-CT and 5′-CC sites[26].

Photolesions can be naturally repaired such that the physiological structures are restored. In general, two classes of repair mechanisms can be distinguished: reversal or excision of DNA damage. The former tries to ”undo” the damage whereas the latter cuts the damage out of the strand and replaces the neccessary parts. These mechanisms comprise nucleotide excision repair (NER) and base exision repair (BER) as well as enzymatic photoreactivation[27]. The latter is performed by the monomeric enzyme photolyase. Photoproducts induced by UV-B are mainly repaird via NER, although this mechanism is not restricted to this kind of damage. Additionally, in humans NER seems to be the only repair mechanism for UV-induced photoproducts[27]. Here, a whole sequence (oligonucleotide), that contains the damage site, is removed and replaced by an undamaged sequence. Sum-

marizing, UV-induced modifications of nucleic acids are likely to cause severe mutations of the genetic code. The most common photolesions induced by direct absorption of UVR are CPDs and 6-4PPs. The occurrence of these lesions is dependent on the environment, base sequence and efficiency of the repair and is proven to facilitate cancerogenesis[19].

Now that the most important consequences of photodamage in DNA are clarified, the last part of this introduction sheds light on the photophysics that lead to the formation of these alterations and tries to explain important effects of the absorption behavior of nucleic acids. UV light absorption promotes the chromophore from the electronic ground state to higher energy levels resulting in an electronic excited state by transfering an electron from an occupied to an unoccupied orbital[28]. As repair mechanisms are energetically onerous, nucleic acids in DNA and RNA possess intrinsic photostability, which means that they are naturally showing ultrafast vibrational relaxation and internal conversion pathways to convert excessive photoenergy into thermal energy. The barrier- and radiationless deactivation mechanisms comprise internal conversion via conical intersections, which requires strongly distorted puckering conformations for all natural DNA/RNA bases[29]. Nevertheless, excited state relaxation due to photoemission cannot be excluded as fluorescence of single nucleobases (at very low quantum yields) was detected at room temperature[30].

In general, the UV absorption spectrum of a double stranded DNA fragment reflects the weighted sum of the absorption spectra of its constituent nucleobases regarding the range of absorption energy. This observation gave rise to the initial assumption that weak interbase interactions do not influence the absorption behavior and, thus, single photons are absorbed by single nucleobases. However, nowadays it is known that this is not the case. In their electronic excited states bases in a DNA stack develop stronger electronic interactions accouting for cooperative behavior[31]. As mentioned before, chemical lesions resulting from UV absorption have hotspots depending on the DNA sequence, which implies that electronic coupling between nucleobases in a duplex is indeed important for absorption of UVR[32]. The role of interbase interactions (Watson-Crick hydrogen bonding and π stacking between neighbouring bases) in redistribution of the initial electronic excitation along the stack of base pairs is poorly understood, but it is essential for understanding subsequent DNA damage mechanisms[33]. Different research groups claim that either base pairing or stacking is the predominant effect involved in excited state relaxation processes. Reasons for the former were provided by the assumption that protons could tunnel across hydrogen bonds[34]. This phenomenon has been recently observed experimentally, proving a stronger photostability of Watson-Crick base pairs compared to individual bases[35]. On the other hand, π stacking efficiently enables the formation of intrastrand excimers in high quantum yields. These excited dimers, which reveal an increased life-time, eventually relax to the ground state, demonstrating a decay channel absent in individual bases and an increase in photostability[36]. This behavior highlights the importance of base stacking in

excited state dynamics of DNA double strands and could be a reason why most prominent photolesions involve two neighbouring bases.

The cooperative effects between nucleobases in a strand generate the formation of electronically excited states that are absent in individual nucleobases: Frenkel excitons and charge transfer (CT) states. In a Frenkel exciton two or more local transitions, in which an electron is promoted from and to orbitals located on the same nucleobase, are coupled. This state is spatially delocalized. However, the electron- and the hole orbitals involved in one electron transition are located on the same nucleobase[37]. Exciton formation is facilitated due to dipolar coupling of individual $\pi\pi^*$ transitions of different chromophores. Frenkel excitons do not include charge separation. In contrast to that, when the initial and final orbitals of a specific transition belong to different chromophores, it is a CT state. The resulting excited nucleobases are charged radicals and the excitation involves at least two bases. CT states are attributed to orbital overlap of the affected nucleobases[38]. Both exciton and CT character can have a specific contribution to the same excited state resulting in a mixed state. The spatial extent and the mixing of both delocalized excitons and CT transitions is dependent on the conformation and alignment of nucleobases[39]. Additionally, CT states are even more sensitive to the environment since interactions with water molecules or ions in the vicinity can influence charge separation[40]. Moreover, base pairing and stacking strongly influence the formation of excitons and CT states. Therefore, a clear characterization of excited states of a DNA strand is a challenging task.

To give an example, base stacking is believed to play the predominant role in terms of excited state formation and relaxation in single and double strands[36]. However, computational results revealed that orbital overlap of hydrogen bond donor and acceptor atoms contribute to the stability of nucleobase pairs and facilitate the formation of interstrand CT states[41]. Furthermore, according to different theoretical studies, excited states with significant CT contribution are believed to be located either in the red part[42] of the spectrum or neither above nor below the bright $\pi\pi^*$ transitions[43]. Another topic researchers disagree about in the literature is the spatial extent of electronic excitation in DNA. While it was initially belived that excited states remain localized on single nucleobases[31], further investigations could show that due to electronic coupling between nucleobases excited states spread over three chromophores[44], exceed four nucleobases[45] or are even delocalized over the whole length of the DNA molecule[46].

Another phenomenon in the photophysics of DNA is an observed hypochromic effect of the absorption of UVR of stacked polymers compared to single nucleotides and nucleobases[47]. This decrease in absorption of up to 40% has been known since the 1960s. First, this effect has been attributed to exciton interactions between stacked nucleobases in a helical arrangement. However, it has been shown that a zeroth-order exciton model is insufficient to explain hypochromism because important effects such as orbital overlap

and the environment are neglected[48]. This conclusion stresses the importance of orbital overlap effects and thus CT states for hypochromism. Indeed, quite recent experimental results suggest the mixing of exciton and CT states resulting in long-lived high energy states to be responsible for hypochromism[49]. In contrast, theoretical studies on a stacked adenine dimer at different separation distances showed that short-range effects responsible for delocalization and CT interactions do not play a significant role. This implies that hypochromism arises mainly due to monomer-like excitations perturbed by long-range effects[50].

Research in the past decades investigating the importance of interbase interactions (hydrogen bonds and π stacking) and quantum effects accounting for the spatial extent of delocalized excitons (dipole interactions) and CT states (orbital overlap) as well as hypochromism revealed different, sometimes contrary results as outlined above. This is comprehensible since experimentalists and theoreticians study different model systems using a variety of methods. From the computational point of view, the accurate simulation of excited states of DNA molecules is a challenge for modern quantum chemistry techniques[51]. In general, low-accuracy methods, such as semi-empirical models[52] for excited state calculations enable a description of large systems and their dynamics. This requires tedious parameterization, which can be a source of error. Time-dependent density functional theory (TDDFT)[53], generally more accurate and, thus, a better choice for excited state calculations, but faces the challenge of choosing an appropriate functional. Wavefunction based methods[54] can be more precise in predicting excited states than DFT. More importantly, when excited state dynamics are considered, the accurate description of intersections beween different electronic states requires the use of multi-reference methods[55]. In general, the higher the accuracy of a given electronic structure method, the smaller the feasible system size in terms of computational costs. A major disadvantage of a small system size is that the environment often cannot be considered at quantum mechanical (QM) level. However, for a realisitc description of DNA those effects are crucial[51]. For these reasons, hybrid methods of QM and molecular mechanics (QM/MM) have been developed[56]. From an experimental point of view, it has been very hard to observe any of the relevant processes in UV absorption because of their ultrafast nature. Since the beginning of the century and with the development of femto-second spectroscopy the primary processes after UV exposure of DNA could be studied in more detail[57]. However, diverse spectroscopical experiments are able to describe population dynamics only indirectly since, for example, relaxation features are measured via photoemission processes[51].

This computational study is devoted to the evaluation of the electronic structure of $(dA)_{20}\cdot(dT)_{20}$ B-DNA in aqueous solution. Investigations on electronic excited states as part of the first band of the UV absorption spectrum are carried out. Therefore, 100 geometries are selected from a molecular dynamics (MD) simulation. For each geometry, the first

60 electronic excited states are calculated using a QM/MM scheme. The relevant excited states are convoluted with Gaussians according to their energies and oscillator strengths in order to obtain the absorption spectrum. With a wavefunction analysis of electronic excited states we like to contribute to the ongoing research and give an insight on the importance of interbase interactions (hydrogen bonds vs. π stacking) for excited state formation by distinguishing between intra- and interstrand excitations. A special focus of this work is to assess the spatial extent of excited states and differentiate between exciton and CT states to show their relevance for the spectrum. Before discussing the results, the theory used in this work is explained.

2 Theory

For the evaluation of electronic excitations in a double-stranded DNA fragment, the following procedure was applied: a classical **MD** simulation of a $(dA)_{20}\cdot(dT)_{20}$ double strand in explicit solvent was performed. 100 geometries were selected from the simulation that represent a thermal ensemble of the DNA fragment. The first 60 excited states were calculated for each geometry with a **QM/MM** scheme. **Time-dependent density functional theory (TDDFT)** applied for the QM subsystem. The resulting excited states were analysed with TheoDORE[58][54][59], a **wavefunction analysis** code that extracts the neccessary information for the evaluation of the position of the excitations in the stack, the delocalization length (DL) and the character of each excited state. The theories behind these methods are explained in the following sections according to this workflow.

2.1 Molecular Dynamics

Molecules move in a three dimensional space, which means they are able to translate, rotate and vibrate. The constant motion of a system is key to chemical reactions and the diversity of compounds as we know them today. Thus, in order to simulate a system towards a realistic behavior, the challenge of describing the motion correctly has to be met. This chapter concentrates on the time propagation that facilitates the evolution of motion of a molecular system. The procedure of molecular dynamics includes a previous energy minimization of the system which is explained first[60].

2.1.1 Energy minimization

Energy minimization has become a common concept in MD since only a realistic geometry (with the lowest potential energy) describes the evolution of the molecular motion in the right way. However, after building the system *in silico*, e.g. bond lengths and angles are unrefined and sterical problems between atoms or molecules occur, which can be corrected

by minimizing the energy of the system[60]. In classical MD the structure of a molecule is described with potential functions that depend on the relative coordinates of the atoms and are based on MM methods (as described in the next section). Several algorithms have been developed for finding a minimum of this potential energy starting with a given geometry. Here, two first-order steepest descent[61] and conjugate gradient[62] methods are employed that make use of the first-order derivative of the potential energy V. For example, in the steepest descent method the gradient (first-order derivative) of the potential function is evaluated.

$$F(\vec{r}) = -\frac{\partial}{\partial \vec{r}} V(\vec{r}) \tag{1}$$

Equation 1 shows that the negative gradient of the potential energy is directly connected to the force vector, which enables a tool for searching along the potential energy surface for a (local) minimum where the gradient is zero.

$$\vec{r}(k) = \vec{r}(k-1) + \lambda(k) F(k) \tag{2}$$

In an iterative manner, a new atomic configuration $\vec{r}$ for step k can be generated using the actual three dimensional coordinates in a vector representation $\vec{r}(k-1)$ and the force $F(k)$ weighted with the step size $\lambda(k)$ like in equation 2. After each iteration the potential energy is compared to that of the previous configuration. If the potential is larger, the step size is reduced since it is assumed that the minimum was skipped. If the potential is indeed smaller than for the previous step, the step size is increased. This way, the conformation of the molecule will get closer to the next minimum on the potential energy surface without crossing local barriers. In contrast to that, the conjugate gradient method takes into account the gradients of the potential energy from previous steps in a linear manner by weighting them with the ratio of the squares of the magnitudes of these gradients. This is slightly more costly than the steepest descent method but is therefore most likely to converge faster. Once the structure has converged to the conformation of minimal energy, it can be used as the starting geometry for a following MD simulation.

2.1.2 Equations of motion

To model the dynamics of a system, simulation of time is indespensable enabling movement and, thus, evolution with time. For this, quantum effects are neglected and only the nuclei are considered representing point masses. Newton's classical equations of motion provide the neccessary information to model the behavior of the molecule with time[60]. In the form of Hamiltonian mechanics (exemplary for a single particle) they are shown in equation 3[63].

$$F = -\frac{dH}{dr} \qquad v = \frac{dH}{dp}. \tag{3}$$

The Hamiltonian H is the sum of the kinetic ($T = \frac{p^2}{2m}$) and the potential energy ($V = V(r)$). Note that T and V depend only on the momentum p and the coordinates r, respectively. From Newtons' equations the time-derivative of the momentum p is the force F, which can be transferred to the first Hamiltonian equation as the negative gradient of the potential energy (see equation 1). Additionally, in Newtons formulation the time-derivative of the coordinates is the velocity v, which leads to the second Hamiltonian equation

$$v = \frac{\partial T}{\partial p} \tag{4}$$

The equations of motions are integrated for example via the velocity verlet algorithm that calculates the new position r_{n+1} and velocity v_{n+1} of a molecule according to the ones from the previous step r_n and v_n. Equations 5 and 6 show the final form of this algorithm[64] using the Hamiltonian equations from above where Δt is the time step.

$$r_{n+1} = r_n + (\frac{\partial T}{\partial p})_n \, \Delta t + (\frac{F}{m})_n \, \frac{\Delta t^2}{2} \tag{5}$$

$$v_{n+1} = v_n + [(\frac{F}{m})_n + (\frac{F}{m})_{n+1}] \, \frac{\Delta t}{2} \tag{6}$$

This way it is clear that the velocity and acceleration ($\frac{F}{m} = a$) or kinetic and potential energy affect the current coordinates and thereby determine the new ones. According to how the potential energy is calculated, different MD methods have developed and comprise completely classical, *ab initio* or hybrid QM/MM calculations. Here, a fully classical MD simulation has been performed, where the potential energy is calculated with classical molecular mechanics (MM) equations employing force fields (FF).

2.1.3 Force Fields

FF are used for a classical calculation of the potential energy of the system[60]. They use MM equations that neglect electronic movement and are empirically parameterized. Therefore, they are less accurate than QM methods. A particular set of functions associated with a specific set of parameters is called a FF. MM, in general, describes the constitution of the molecule using potential functions, which are valid under Born-Oppenheimer restrictions[65]. Usually, the total energy is split into bonded and non-bonded interactions. The total potential of the molecule is simply achieved through adding the single contributions of bonded and non-bonded terms, thereby assuming independence among contributions.

The functions that describe these interactions include parameters obtained from QM calculations or experiments.

$$E_{total} = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} V_n[1 + cos(n\phi - \gamma)] + \tag{7}$$

$$+ \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 4\varepsilon_{i,j}[(\frac{\sigma_{i,j}}{r_{i,j}})^{12} - (\frac{\sigma_{i,j}}{r_{i,j}})^{6}] + \frac{q_i q_j}{\epsilon r_{i,j}}.$$

Equation 7 shows the potential energy function that describes the total potential energy E_{total} of the FF that is used for this work obtained from the Amber16 manual[66]. The aforementioned parameters are k_b, r_0, k_θ, θ_0, V_n, γ, $\sigma_{i,j}$, $\varepsilon_{i,j}$, q_i and q_j. The first term runs over all covalently bound pairs of atoms and describes the bond stretching according to a harmonic expression where the force constant is k_b and the equilibrium distance is r_0. Angles are calculated in the same way using the force constant k_θ and the equilibrum angle θ_0. Torsions are represented through the third term on the right side of equation 7. V_n is the torsion barrier, γ the phase and n the periodicity. Improper dihedrals (out-of-plane terms) are calculated in the same way. The remaining non-bonded terms of the potential energy function contain van-der-Waals (vdW) and electrostatic interactions. The former are expressed with the standard Lennard-Jones (LJ) potential. $r_{i,j}$ is the distance between two atoms i and j, $\varepsilon_{i,j}$ is the depth of the potential for the interaction of these atoms and $\sigma_{i,j}$ is the distance where the potential is zero. The partial charges q_i and q_j on atoms i and j, respectively, neccessary to describe the electrostatic interaction, are derived from high level *ab initio* calculations while ϵ is the electrostatic constant[66].

2.1.4 Periodic boundary conditions

For macromolecular systems, e.g. a DNA double strand in aqueous solution, the use of periodic boundary conditions (PBC) is required. For example, PBC enables the description of an infinite solvent environment employing a finite number of solvent molecules[67]. To do so, a small part, the so-called *unit cell*, is selected that represents the system. In three dimensional space this unit cell can be a cube with equal edge lengths. This cube is copied infinite times in every direction around the original unit cell to generate a perfectly filled space. The copies of the unit cell are called *images*. However, the shape of a cube includes a lot of solvent molecules, e.g. at the edges of the cube, that are too far away from the macromolecule in the center. This causes unneccessary computational costs. Thus, a truncated octahedron, as employed in this study, is a suitable alternative. Only the original unit cell is propagated throughout the simulation and when a molecule or atom passes through the boundaries of the box, it is displayed at the opposite side preserving its velocity. Since DNA is a negatively charged molecule, the solvent has to include ions to neutralize the

system. Otherwise, an inifinte sum of charges would be generated by the use of PBC[68]. The concept of PBC includes the *minimum image convention*, which states that an object of the unit cell interacts with the closest image of all other objects. This means that a spherical cutoff radius for non-bonded interactions can at most be half of the edge length of the unit cell. Another important note, when employing PBC is, that the size of the unit cell must not be too small or otherwise the macromolecule can interact with an image of itself from a neighboring box and thereby generate periodic artifacts.

2.1.5 Equilibration of the system

After constructing and minimizing the system, a proper equilibration guarantees realistic dynamics. Therefore, the system will be heated to 300 K by applying constraints on the DNA fragment. This way the solvation shell is able to relax and adapt to the changes in temperature. After that, the constraints are reduced to DNA bonds involving hydrogen atoms only, in order to enable a similar relaxation of the DNA strand. During the heating steps the canonical *NVT* ensemble is applied, which keeps the number of particles N, the volume V and the temperature T constant. The canonical ensemble represents possible states of a mechanical system differing in total energy. This is achieved through thermal equilibrium with a virtual heat bath and the temperature is controlled by exchanging heat with this heat bath[69]. In particular, the Langevin thermostat is used here for temperature control by rescaling velocity. Additional terms that account for friction and random forces are applied to the momenta of the particles within the equations of motion to induce the probability of collision. That way, a canonical distribution for many-body simulations is provided. The major advantage of this thermostat is that it allows the use of larger time steps during the equilibration of the system. After heating the system, the temperature is kept constant (at 300 K) with the Langevin thermostat. The actual production run was performed with the *NPT* ensemble under isothermic-isobaric conditions. As the volume was kept under constraint in the previous heating step, it is now variable and, thus, the system will expand in order to fully relax. Therefore, the pressure P is constant and maintained at a value of 1.0 bar by the Berendsen barostat[70] which acts as a piston by rescaling the volume with an external reference pressure. Thus, barostats rely on the same principles as thermostats but rescale the position of atoms rather than their velocity. For our purposes, 100 geometries from the MD simulation are selected representing a thermal ensemble of the DNA double strand. Those structures were extracted only from the second half of the production run (see Chapter 3) in order to avoid initial artifacts caused by the equilibration steps.

2.2 Quantum Mechanics / Molecular Mechanics

The Quantum Mechanics / Molecular Mechanics (QM/MM) scheme is a method to describe large systems with accurate precision. This has been an impossible task for many years until the development of this hybrid technique in 1976[71]. It combines QM with MM methods. The simple idea is to partition the system at hand into two regions that are calculated differently. A small QM region will be treated with a quantum theory of choice, whereas the surrounding environment is described using a (biomolecular) FF employing classical equations but lacking chemical accuracy[72]. Thus, the result for the QM region will be more accurate but due to the large computational costs the subsystem can comprise only a small number of atoms. QM/MM was primarily developed for modeling biomolecules, e.g. in explicit aqueous solution. A small part of the molecule, preferably atoms involved in a reorganization of their electronic structure, are set as the QM region, for example active residues in an enzymatic reaction[56] or - like in this work - nucleobases involved in photoabsorption. The MM region then contains all other atoms and molecules that are not believed to be actively involved in any kind of reaction like the QM part but rather serve as a scaffold (in case of a protein), hydration shell or bulk water (in case of an aqueous solution).

The QM/MM Energy

In order to obtain the total energy using a QM/MM calculation two possible ways are presented: the subtractive and the additive schemes. The former requires a MM calculation of the total system S_{MM} as well as of the inner region I_{MM} (QM atoms) and a QM calculations of the inner region I_{QM}[72][56]. The total energy $E(S_{QM/MM})$ is then obtained via

$$E(S_{QM/MM}) = E(S_{MM}) - E(I_{MM}) + E(I_{QM}). \tag{8}$$

The main advandtage of the subtractive scheme is the simplicity provided since it basically cuts the inner MM expression and replaces it with the more accurate QM result; therefore the calculation does not need additional coupling terms.

$$E(S_{QM/MM}) = E(O_{MM}) + E(I_{QM}) + E(C_{QM/MM}) \tag{9}$$

The additive scheme (equation 9),which is employed in this work, requires the calculation of the outer part on MM level O_{MM}, a QM calculation of the inner subsystem I_{QM} and an explicit coupling term $C_{QM/MM}$ that accounts for the interactions between QM and MM region.

The coupling term includes bonded and non-bonded contributions. The latter comprises van-der-Waals and electrostatic interactions. Since they form the most important part of

the coupling the electrostatics are explained first. The interaction between the QM charge density and the MM charge model can be classified according to the level of increasing polarizability as mechanical, electrostatic or polarizable embedding. The former applies the same charge model of the MM force field to the atoms in the QM subsystem and describes the interaction between QM and MM similar as MM-MM electrostatics. Hence, the charges of the outer region do not influence the QM charge density. Additionally, the charges of the inner system cannot be updated when redistribution of QM charge density occurs, e.g. during a reaction, since this would lead to a discontinuity in the potential energy surface.

To go one step further, electrostatic embedding incorporates the MM point charges as one-electron terms in the QM Hamiltonian representing a QM calculation in the presence of MM charges.

$$\hat{H}^{el}_{QM/MM} = -\sum_{i}^{N}\sum_{J\in O}^{L}\frac{Q_J}{|x_i - R_J|} + \sum_{a\in I}^{M}\sum_{J\in O}^{L}\frac{Q_J Q_a}{|R_a - R_J|} \tag{10}$$

In equation 10 (in a.u.) the first term describes the electrostatic interaction of all L outer point charges Q_J at R_J with N electrons of the QM part at position x_i, whereas the second term considers the repulsion of all outer MM point charges with M nuclei charges Q_a at R_a.

In this way the QM density is polarized by changes in the charge distribution of the outer part.

Polarizable embedding[73] expands the previous idea by introducing flexibility to the MM charges in a way that they can be polarized by the QM electric field. Either the MM charges are restricted to not be allowed to act back on the QM region or the polarizable MM model is incorporated fully self-consistently in the QM Hamiltonian which enables mutual polarization. This requires a polarizable MM force-field[74]. This work makes use of electrostatic embedding.

In contrast to the electrostatic interaction, the van-der-Waals contribution and the bonded coupling terms are described fully on MM level, thus it is not important if the subtractive or additive scheme is applied. The van-der-Waals interactions among the two subsystems is simply computed via a Lennard-Jones potential. Due to the short-range nature of these interactions, only the atoms close to the boundary region play a pivotal role.

When it comes to bonded interactions, one has to be careful when the boundaries between the two regions cut covalent bonds. If this is the case, the most common approach is the link-atom scheme[75] which is also employed in this work. It saturates the free valency that results from the cut with a (hydrogen) link atom. The QM calculation then comprises the inner region plus all link atoms.

2.3 Density Functional Theory

Density functional theory (DFT) tries to simplify the electronic Schrödinger equation by using the electronic density as a variable.

$$\hat{H}_{el}\Psi_{el} = E_{el}\Psi_{el} \tag{11}$$

$$\hat{H}_{el} = \hat{T}_e + \hat{V}_{ext} + \hat{V}_{ee} \tag{12}$$

Under Born-Oppenheimer restrictions the nuclei are fixed, the electronic wavefunction Ψ_{el} only depends on the information about electrons (and only parametrically on coordinates of nuclei) and, thus, the electronic Hamiltonian $\hat{H}_{el}$ comprises the kinetic energy of electrons $\hat{T}_{el}$, the interaction between electrons and nuclei, which is termed external potential $\hat{V}_{ext}$ in DFT, and the repulsion between electrons $\hat{V}_{ee}$.

First, the electron density will be defined and some characteristics are outlined. After that, the Hohenberg-Kohn theorems are introduced representing the first landmarks in DFT and showing why density functionals are a good replacement for wavefunctions. Subsequently, the Kohn-Sham method is outlined as the procedure of how to appoximate the right density functional in ground state DFT. From Subsection 2.3.2 onwards, time-dependent density functional theory (TDDFT) as a possible method of how to extract information about excited states from the ground state electron density is explained.

The probability density, more commonly termed electronic density $\rho(\vec{r})$ is an expression for the probability of finding any of the N electrons - regardless of the spin σ- within a volume element $d\vec{r_1}$ while the other N-1 electrons are arbitrarily distributed over the space independent of their spin[76]. This quantity is the central variable in density functional theory. Note that $\vec{x}$ comprises electron coordinates $\vec{r}$ and spin information σ collectively.

$$\rho(\vec{r_1}) = N \int ... \int |\Psi(\vec{x_1}, \vec{x_2}, ..., \vec{x_N})|^2 d\sigma_1 d\vec{x_2} ... d\vec{x_n} \tag{13}$$

Equation 13 shows that the function of the electron density only depends on the three spatial variables of the volume element of one particular electron ($\vec{r_1}$). The electron density vanishes for infinite separation between nuclei ($\vec{R} \rightarrow \infty \Rightarrow \rho(\vec{r}) \rightarrow 0$) and it integrates to the total number of electrons ($\int \rho(\vec{r_1}) d\vec{r_1} = N$); furthermore, the electron density $\rho(\vec{r})$ gives information about the position of the nuclei since, here, the density shows cusps, and last but not least the density at the position of the nuclei contains information about the nuclear charges Z_i.

2.3.1 The Hohenberg-Kohn theorems and Kohn-Sham methodology

The theorems derived by Hohenberg and Kohn (HK) in 1964 were the first real proof that the density is a plausible replacement for wavefunctions[76]. The first theorem states that there cannot be two external potentials V_{ext} that lead to the same ground state electron density. This means that the ground state electron density uniquely specifies the external potential. Thus, it is clear that the electronic density is a meaningful variable that contains all the information we need to solve the (approximated) Schrödinger equation.

Following these findings, the second HK theorem states the way to find the exact ground state electron density ρ_0. Here, they simply apply the variational principle and get (for clarity $[\rho(\vec{r})] = [\rho], [\tilde{\rho}(\vec{r})] = [\tilde{\rho}], [\rho_0(\vec{r})] = [\rho_0]$)

$$E_0 \leq E[\tilde{\rho}] = T[\tilde{\rho}] + E_{ee}[\tilde{\rho}] + E_{ext}[\tilde{\rho}], \tag{14}$$

which means that every trial electron density $\tilde{\rho}$ yields an energy equal to or above the exact ground state energy. When the result is exactly the ground state energy, then the trial electron density is the exact ground state electron density. Thus, minimizing the energy in equation 14 will guide us to the exact ground state electron density and energy, respectively.

In DFT the ground state energy as a functional of the ground state electron density is

$$E_0[\rho] = T[\rho_0] + E_{ee}[\rho_0] + E_{ext}[\rho_0]. \tag{15}$$

This can be separated into system-dependent and system-independent parts. The third term of the summation is the only system dependent part. The other two summands are universally valid but unknown. They can be combined into the *Hohenberg-Kohn functional* $F_{HK}[\rho_0]$. By inserting the expression of the external energy given due to the electron-nuclei attraction, it can be rewritten as

$$E_0[\rho] = F_{HK}[\rho_0] + \int \rho_0(\vec{r}) V_{ext} d\vec{r}. \tag{16}$$

The HK functional contains the kinetic energy of the electrons and the electron-electron repulsion term of the ground state electron density. One of the main goals of modern DFT is to find the right functional for $F_{HK}[\rho_0]$.

One year after the famous HK theorems, Kohn and Sham (KS) developed their methodology in order to find the right functional[76]. The overall achievement of this method is to extract as many things as possible from the electronic kinetic energy term that can be solved and describe the rest in an approximate manner. First, the classical Coulomb part $J[\rho]$ can be extracted from $E_{ee}[\rho]$. The remaining non-classical part $E_{ncl}[\rho]$ accounts for the self-interaction correction, exchange and Coulomb correlation. The unkown functional

$F_{HK}[\rho]$ is approximated with a reference system (subscript S) of non-interacting electrons. The kinetic energy of the non-interacting reference system T_S will be different than the one of the real system when using the same electron density. Aware of these differences, we can rewrite:

$$F_{HK}[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho]. \tag{17}$$
$$E_{XC}[\rho] = (T[\rho] - T_S[\rho]) + E_{ncl}[\rho]$$

In this way, all unknown parts are combined in the exchange-correlation term $E_{XC}[\rho]$. Approximating the unknown potential with a function of the electron density is part of the *local density approximation* (LDA)[77]. In LDA the potential is local, meaning one approximates the exchange correlation potential V_{XC} of $E_{XC}[\rho]$ as a function of the electron density at a certain point in space $V_{XC}(\vec{r})$. Urging towards a more accurate description, the *generalized gradient approximation* (GGA)[78] includes the derivative of the electron density. By mixing a specific fraction of the exact (Hartree Fock) exchange energy to GGA exchange and correlation energy, the so-called *hybrid* functionals[79] are developed, one of which is used in this work.

2.3.2 Time dependent DFT and linear response

To introduce the time dependence in DFT the methodology is basically very similar to the previously described time independent part. Runge and Gross first described in 1984 the time dependent HK theorem and found a one-to-one correspondence to the time independent result, particularly that the time dependent potential and the wavefunction are unique functionals of the density[80].

To extract information about excited states, time-dependent density functional theory makes use of the linear response (LR) method[81]. In LR theory a perturbation, e.g. an electric field, at time $t = 0$ is introduced to the ground state electron density $\rho_{GS}(r,t)$.

$$\rho(r,t) = \rho_{GS}(r,t) + \delta\rho(r,t) \tag{18}$$

Assuming that this perturbation is weak, representing common spectroscopic concepts, the electron density is simply the ground state density with an additional small deviation $\delta\rho(r,t)$.

In a last step the susceptibility χ has to be introduced. This value connects the response of the ground state to the small change in the time-dependent external potential δv_{ext}

$$\delta\rho(r,t) = \int dt' \int d^3r' \chi[\rho_{GS}](r,r',t-t')\delta v_{ext}(r',t'). \tag{19}$$

Equation 19 indicates that the susceptibility χ describes the change in the density at a different time t and position r when a change to the external potential is introduced before (at t', r'). Similar to the previous procedure of the KS methodology, we generate the susceptibility for the reference system of non-interacting electrons, χ_{KS}, as well. Thus, we now incorporate the change in the KS potential $\delta v_{KS}(r', t')$ that includes the classical Coulomb interactions and additionally the differences to the real system instead of only the external potential δv_{ext} in equation 19.

$$\delta\rho(r,t) = \int dt' \int d^3r' \chi_{KS}[\rho_{GS}](r,r',t-t')\delta v_{KS}(r',t') \tag{20}$$

Equation 20 elucidates the response of the non-interacting KS density to a slight change in the KS potential δv_{KS}, just like equation 19 denotes the response of the real system to a change in the external potential. At this point, let us recall that the density of the reference system equals the one of the real system where electron interaction is not neglected. Transfering this to the last two equations it becomes clear that the change in the external potential of the real system and the change in the KS potential of the reference system will also reveal the exact change in electron density, hence, the expressions on the left side of equations 19 and 20 are the same. Because of that, we can simply equate them taking into account the exchange correlation kernel $f_{XC}[\rho_{GS}](r,r',t-t')$ of equation **??** and the explicit definition of the KS potential and get the central equation of TDDFT linear response in frequency space:

$$\chi(r,r',\omega) = \chi_{KS}(r,r',\omega) + \tag{21}$$
$$\int d^3r_1 \int d^3r_2 \chi_{KS}(r,r_1,\omega)\, \frac{1}{|r_1 - r_2|} + f_{XC}(r_1,r_2,\omega)\, \chi(r_2,r',\omega)$$

All components of equation 21 are functionals of the ground state electron density. Concluding, when ω is exactly the frequency of a real transition of the system, the response function χ has a pole as a function of ω. Summarizing, to extract information about excited states, the system will be perturbed with a weak electric field in the beginning and time-dependent KS equations will be propagated. After Fourier transformation of this function of time, the optical absorption spectrum can be obtained.

Different ways of extracting information about excited states in TDDFT, one of which is LR, yield the same eigenvalue equation in matrix notation that is solved by most quantum chemistry programs today[82][83]:

$$\begin{bmatrix} A & B \\ A^* & B^* \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}. \tag{22}$$
$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + (ia|jb) + (ia|f_{XC}|jb)$$

$$B_{ia,jb} = (ia|bj) + (ia|f_{XC}|bj)$$

i and j are occupied (molecule) orbitals, a and b virtual (molecule) orbitals. ϵ_a and ϵ_i are the one-electron orbital energies of orbitals a and i, respectively, and δ corresponds to the Kronecker delta. Matrix elements A include the energy difference of the orbitals from and to which an electron is excited as well as the antisymmetrized two-electron integrals concerning the exchange interaction and the response of the exchange correlation potential, respectively. The last two terms are also included in B. Finally, X and Y are the excitation- and de-excitation vectors, respectively.

2.4 Wavefunction Analysis

In this work the analysis of excited states was performed with the TheoDORE[58][54][59] code. It concerns the one-electron transition density matrix $\tilde{X}$ in the basis of non-orthogonal atomic orbitals (AOs) μ and ν identified with the excitation vector X_{ia} of equation 22 between two (molecular) orbitals i and a (MOs).

$$\tilde{X}_{\mu\nu} = \sum_{i,a} C_{\mu i} X_{ia} C_{\nu a} \tag{23}$$

According to equation 23 the transition density matrix in AOs $\tilde{X}$ can be derived from the excitation vector in MOs X by a transformation taking into account the elements of the respective MO coefficients (C).

Further analytic steps require an intuitive division of the system into appropriate fragments. An excitation is then termed local when both orbitals, e.g. μ and ν of the abovementioned matrix element of the transition density matrix $\tilde{X}$ are located on the same fragment M; the respective matrix element of the CT number matrix Ω_{MM} is the weight of local excitations on M. Consequently, when the orbitals belong to different fragments the excitation has CT character. The respective CT number Ω_{MN} is a weight for the CT character between the two fragments M and N. To obtain the relevant elements of the CT number matrix, the sum over all contributions on the fragments of interest has to be calculated according to equation 24 using non-orthogonal atomic orbitals and the overlap matrix S.

$$\Omega_{MN} = \frac{1}{2} \sum_{\mu \in M, \nu \in N} [(\tilde{X}S)_{\mu\nu}(S\tilde{X})_{\mu\nu} + \tilde{X}_{\mu\nu}(S\tilde{X}S)_{\mu\nu}] \tag{24}$$

Meaningful physical information can be extracted from this CT number matrix in form of more compact descriptors. This work is concerned about the delocalization of the excitation as well as the character of the states (Frenkel exciton vs CT) and utilizes the relevant descriptors - *DL* for the delocalization length and *CT* for charge transfer character, respectively - derived from the TheoDORE analysis of the transition density matrix. Note, that

here the italic abbreviations *DL* and *CT* are used for the descriptors of TheoDORE.

$$CT = \frac{1}{\Omega} \sum_{M,N \neq M} \Omega_{MN} \tag{25}$$

According to equation 25 the CT character of a state α is simply the summation over all off-diagonal elements of the CT number matrix with the normalization factor $\Omega = \sum_{M,N} \Omega_{MN}$. The values for *CT* range from 0 for local exctations and Frenkel excitons to 1 for full CT states.

$$DL = \frac{\Omega^2}{\sum_M (\sum_N \frac{\Omega_{MN} + \Omega_{NM}}{2})^2} \tag{26}$$

The *DL* is calculated as an inverse participation ratio (equation 26) that obtains the number of fragments involved in the excitation of a specific state. Thereby, *DL* includes the delocalization of the initial orbitals (hole) and the final orbitals where the electron is distributed. Logically, *DL* yields 1 for local states, 2 for a delocalized state involving two fragments and so on. Thus, *DL* tells how many fragments, but not which ones exactly are contributing to the excited state.

To give an example, Figure 3 shows three possible excited states. These can be distinguished according to the descriptors previously introduced using a single nucleobase as one fragment. For a local exciton on an adenine (3a) *DL* will be 1 and since the electron and the hole are both on the same fragment, *CT* is 0. If an exciton state is delocalized over four nucleobases (3c), *DL* will yield 4 while *CT* remains 0. An interstrand CT state that involves two nucleobases will reveals a *DL* of 2 and a *CT* of 1 (3b).

With the previous descriptors it is not yet possible to distinguish between inter- and intrastrand states, since, e.g., the *DL* and *CT* values of the state depicted in Figure 3c would remain the same if all transitions would be distributed solely over four adenines. To separate this, the exact position of the excited state in the QM stack has to be known. A possible quantity to analyse this is the position descriptor of the hole orbital POS_I and the orbital where the electron will be located POS_F and consequently the mean position *POS* of the excitation. For the separation between inter- and intrastrand states the QM region is divided into two fragments, each containing all nucleobases of one strand. If the adenine strand is fragment one and the thymine chain is defined as fragment two, the excited state is labeled intra-adenine and intra-thymine if *POS* is 1 and 2, respectively, and interstrand for values in between.

$$POS = \frac{(POS_I + POS_F)}{2} \tag{27}$$

$$POS_I = \frac{\sum_M M(\sum_N \Omega_{MN})}{\Omega} \qquad POS_F = \frac{\sum_N N(\sum_M \Omega_{MN})}{\Omega}$$

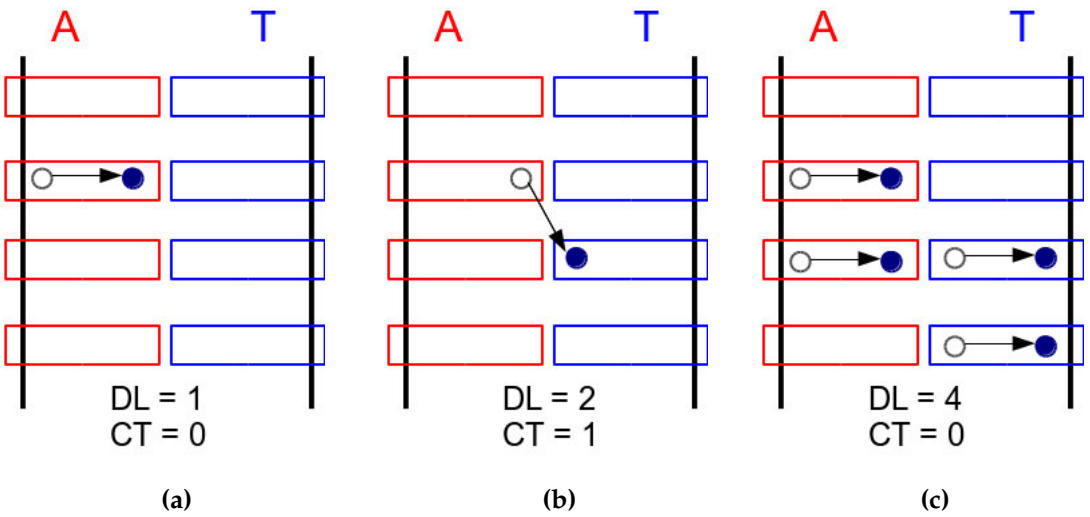


Figure 3: Exemplary excited states in a $(dA)_{20}\cdot(dT)_{20}$ double strand and belonging values for the delocalization length (DL) and charge transfer contribution (CT) resulting from the TheoDORE analysis. An empty circle depicts the hole orbital, a filled circle the final orbital. (a) Local intra-adenine excited state, (b) Diagonal interstrand CT state, (c) Delocalized interstrand exciton.

3 Computational details

The system was built with the nucleic acid builder of AmberTools16[66], a code that enables building and manipulating macromolecules, and comprised a double stranded B-DNA consisting of 20 Adenine-Thymine nucleobase pairs $(dA)_{20}\cdot(dT)_{20}$ as well as a 12 Å solvation shell of explicit water molecules and counterions to obtain a neutral system, the topology and parameters of which were generated with the *leap* module of AmberTools16[66]. The unit cell is a octahedral box around the nucleotides in the center. After minimization for 10,000 steps according to the steepest descent method followed by another 10,000 steps of conjugate gradient minimization, a preparation run of two heating steps each for 100 ps (50,000 steps with a step size of 2 fs) was performed. The first one heated the system to 300 K applying constraints with a force constant of 10 $kcal/(mol\text{Å}^2)$ on the entire DNA strand in order to equilibrate the solvation shell. During the second heating step only bonds involving hydrogen atoms in the strand were set as constraints. All steps so far were performed with the NVT ensemble to provide constant volume and a Langevin thermostat for temperature control with a collision frequency of 1 ps^{-1}. The following production run comprised $25{*}10^6$ steps representing 50 ns in 5000 frames. The NPT ensemble guaranteed isobaric conditions in order to fully equilibrate the volume of the system. Therefore, the Berendsen barostat[70] was used with a pressure relaxation time of 5.0 ps to maintain a pressure of 1 bar. Again, the Langevin thermostat was used. Throughout

all previously described simulation steps the non-bonded interactions were cut after 10 Å, the Particle Mesh Ewald (PME) method[84] was used to calculate the electrostatic energy bond lengths where involved hydrogen atoms were constrained according to the SHAKE algorithm[85] [86] and all atoms were described using force fields (ff14SB[87] for DNA, TIP3P[88] for water molecules) obtaining fully classical molecular dynamics with periodic boundary conditions conducted by the *pmem* module of Amber14[89].

From the 5000 snapshots 100 equidistant geometries were extracted beginning from the middle of the simulation. Additionally, one geometry was minimized acting as an optimized geometry. All following computational details apply for the calculations and results presented in sections 4.2 and following. Deviant settings that were used to find the appropriate QM size in Section 4.1 are mentioned there. The snapshots were used as starting coordinates for following QM/MM calculations of the first 60 vertical excitations (Franck-Condon excited states) per geometry (resulting in a total of 6000 states for the thermal ensemble) employing electrostatic embedding with TeraChem[90][91]. If not further noticed, the QM region comprises five nucleobase pairs (without the sugar phosphate backbone) in the middle of the strands under the conditions of the link atom scheme. Those ten individual nucleobases were calculated with TD-DFT[92][93] on CAM-B3LYP[94] level of theory with the Ahlrichs SV(P) basis set[95] (double zeta valence and a single set of polarization functions on non-hydrogen atoms) and dispersion correction. The MM charges to generate the surrounding electric field were taken from the respective FF. The total amount of electronic excited states were analysed with TheoDORE[58][54][59] - if not further noticed - neglecting the states that are located on the edge nucleobases, i.e. have a high contribution from these chromophores. Nucleobases on the edge of the stack face a different (MM-) environment and are likely to generate artifacts as shown in Ref. [50] where they elucidated that the DL is underestimated since neighbouring nucleobases are contained in the MM region. These states are identified by adding up the electron and hole contribution from edge nucleobases, respectively, obtained from the electron-hole population analysis in TheoDORE. If the resulting value for the electron or hole is larger than 0.5, meaning either the hole or the electron is located to more than 50% on edge nucleobases, the state is not considered for further analysis. Additionally, all states above 6 eV are not considered either since they are already part of the second absorption band. This work concentrates on the features of the first band only. All absorption- and decomposed spectra in the results section are generated convulting the remaining excited states with Gaussian functions with a 0.20 eV full width at half maximum. An additional red shift of 0.90 eV was applied to all spectra to match the experimental values of Ref. [40] because the small SV(P) basis set overestimates excited state energies. As Ref. [96] explains, a larger basis set will induce a red shift in the spectrum resulting in better agreement with the experiment. Additionally, Ref. [50] could show that an extended basis set (6-311+G(2d,p)) red shifts the computed

spectrum of an oligonucleotide consisting of four stacked adenine nucleobases by 0.23 eV compared to SV(P). Moreover, the shape of the band and a similar decomposition analysis remained unaltered. They concluded that an even larger basis set would induce further shifting towards the experimental values. For the $(dA)_{20}\cdot(dT)_{20}$ strand the size of the QM region and the conformational sampling with 100 geometries prevents the application of an extensive basis set since the computational costs would be infeasible. Furthermore, the plots are normalized with respect to the graph with the highest intensity in each plot.

4 Results and discussion

The electronic characterization of UV absorption bands of DNA strands is quite challenging. Indeed, different researchers, theoreticians and experimentalists, disagree on the details of the UV absorption behavior of double-stranded DNA (dsDNA). Topics of the discussion comprise whether excited states are localized on single nucleobases or if the electronic excitation can be delocalized along the DNA fragment generating excitons. Furthermore, different statements have been published in the literature regarding the character of the excitation and the position in energy of Frenkel excitons and CT states. It is not yet fully understood which role is played by interbase interactions like hydrogen bonding and π stacking. These issues are addressed in the present investigation. To this aim, the excited-state electronic structure of $(dA)_{20}\cdot(dT)_{20}$ is analysed. Therefore, a total of 6000 electronic excited states calculated with QM/MM from a thermal ensemble of 100 geometries (60 states per structure) were analysed with TheoDORE. With this wave-function analysis program the DL and the CT contribution for each state can be extracted, which helps to reveal the spatial extent and the character of the transition, respectively. The first section (4.1) is devoted to find the appropriate size of the QM subsystem in order to investigate on the DL of excited states. After deciding the size of the QM region, a general description of the first band of the absorption spectrum is provided in Section 4.2 while a decomposition according to the *DL* (equation 26) for single nucleobase fragments is given. In Section 4.3, the excited states are further separated in intra- and interstrand excitations using the *POS* descriptor (equation 27). With this classification and the discussion in terms of their DL and CT contribution, the influence of hydrogen bonding and π stacking interactions on excited state formation in dsDNA can be evaluated.

4.1 Size of the QM region

To calculate the electronic excited states a QM/MM scheme is used. This section deals with different QM regions that increase from 4 to 24 nucleobases. The goal is to find a size that is suitable for describing the electronic excited states of a $(dA)_{20}(dT)_{20}$ DNA stack. To

achieve this goal, the DL, defined as the number of nucleobases involved in an excitation, was computed for different QM region sizes. In particular, five equidistant geometries of the classical MD simulation were selected. For each, the first 60 excited states were calculated employing different QM regions increasing by one nucleobase pair at a time. For the analysis all states under 6 eV were taken into account and each nucleobase defines a single fragment. The average DL ($<DL>$) and the oscillator weighted average DL ($<DL>_f$) were calculated according to the following equations where all sums run over all considered states n with f_i being the oscillator strength of state i:

$$< DL >= \frac{\sum_{i=1}^{n} DL_i}{n} \qquad < DL >_f = \frac{\sum_{i=1}^{n} DL_i * f_i}{\sum_{i=1}^{n} f_i}.$$

Considering five MD geometries (Figure 4), the values for $<DL>$ and $<DL>_f$ keep increasing until a QM size of 14 nucleobases. Then, they oscillate around a value of three nucleobases and do not expand significantly when including more chromophores in the QM subsystem. Further analysis excluded the excited states where the hole or electron contribution of edge nucleobases of the stack is larger than 50%. Nucleobases on the edges of the QM stack face a different MM environment and are likely to be affected by artifacts since the delocalization cannot be elongated across MM nucleobases[50]. If these states are omitted, the average DL reaches the plateau close to three nucleobases already for ten QM chromophors (dashed lines). These results suggest that a QM size of ten nucleobases is sufficient for the analysis of the excited states when an ensemble of MD geometries is considered and the excited states that are constituted mainly by contributions from edge nucleobases are neglected.

4.2 Delocalization Length

The spectrum of the first absorption band of the $(dA)_{20}\cdot(dT)_{20}$ DNA stack is represented in Figure 5a (black line). The spectrum ranges from 3.8 - 5.2 eV and develops a continously broad peak with a maximum around 4.6 eV and a shoulder around 4.3 eV. This spectrum is red shifted (by 0.9 eV) to match the experimental results of Ref. [45], which is justified since only the energy values of excited states are overestimated due to the small basis set. However, spectral features are correctly reproduced compared to experimental[40, 45] and more accurate computational[50, 96] results, as explained in Section 3. Based on experimental measurements, Markovitsi et al have claimed that the shoulder can be attributed to low-lying CT and/or $n\pi^*$ states[45]. They rely on experiments performed by Krawczyk et al[97]. However, Krawczyk and colleagues investigate on single-stranded polynucleotides in A-DNA conformation. They observe low-energy transitions with increased polarizability and transition dipole moments and attribute them to mixed $\pi\pi^*$ / $n\pi^*$ states with

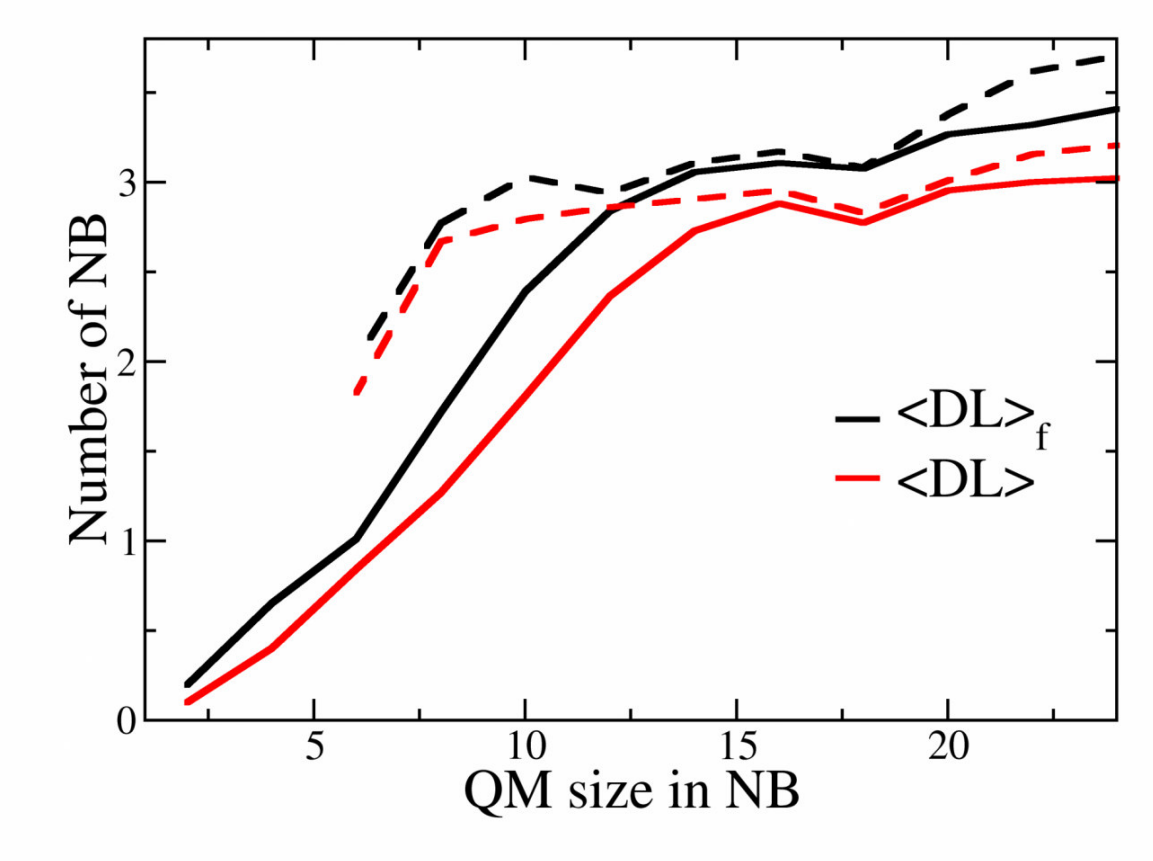


Figure 4: Average delocalization length ($<DL>$) and oscillator weighted average DL ($<DL>_f$) of 300 excited states calculated employing different quantum mechanical (QM) regions of 4 to 24 nucleobases (NB) for 5 MD geometries. Dashed lines represent an analysis omitting excited states located mainly on the edges of the QM nucleobase stack.

charge resonance contribution. This is supported by a model with which they try to explain that coupling of $\pi\pi^*$ with $n\pi^*$ transitions results in red-shifting and an increased intensity compared to monomeric transitions[97]. This is significant for single strands consisting only of thymine, cytosine or guanine. The weakness of this model is that it is based entirely on Coulombic interactions between neighboring nucleobases, which are calculated via transition dipole moments under the point dipole approximation. This approximation breaks down if the separation between neighboring nucleobases is smaller than the spatial extent of the transition dipole moment and is, thus, unable to properly account for dipolar coupling[46]. Moreover, various studies claimed that atomic transition charges derived from transition densities are more suitable to calculate Coulombic interactions[98, 99]. A different computational approach predicts that $n\pi^*$ states are destabilized upon double helix formation, which makes them insignificant for the first absorption band[100]. Additionally, taking only Coulombic interactions into account neglects important effects arising from nucleobase pairing and -stacking, e.g. exchange interactions and orbital overlap. Moreover, they emphasize that the increased absorption intensity for states with significant $n\pi^*$ transitions is dependent on the conformation (A- vs. B-DNA) and that, once downshifted, they become hard to distinguish experimentally from pure $\pi\pi^*$ states[97]. Thus, no definite evidence is provided that the shoulder in the absorption spectrum of the AT double helix really arises due to the contribution of CT or $n\pi^*$ transitions. This issue will be addressed later when a decomposition according to CT contribution is performed.

Analysis concerning the *DL* was carried out in order to investigate the spatial delocalization of the excited states with the TheoDORE program[54, 58, 59] and a fragmentation scheme according to Figure 5b. Here, every nucleobase represents a single fragment. A decomposition of the total absorption spectrum into different *DL* contribution was performed. For example, all states with a *DL* between 2.5 and 3.5 contribute to *DL* = 3. Looking at Figure 5a, an electronic excitation is most probably delocalized over two (23.9%) and three nucleobases (21.7%) affecting almost half of the states. This is consistent with a previous finding of Voityuk and colleagues[44]. They employed an even larger QM region (at semiempirical level) and concluded a delocalization over two to three nucleobases for the AT double strand. Local excitations (12.2%), which dominate the red tail of the spectrum and delocalization over four (17.4%) as well as five (8.9%) and six (8.4%) nucleobases still represent an important fraction. The rest of the states (<8%) are delocalized over more than six nucleobases. The average *DL* is 3.4 nucleobases. Additionally, this analysis reveals the presence of interstrand excitations since *DL*s of six and higher occur while the QM region spans only five nucleobases per strand. Interstrand excited states will be analysed later (see section 4.3.2).

Comparing these findings to a previous study of a polyadenine single strand, where the electronic excitation involves at most six nucleobases (employing a QM region of eight

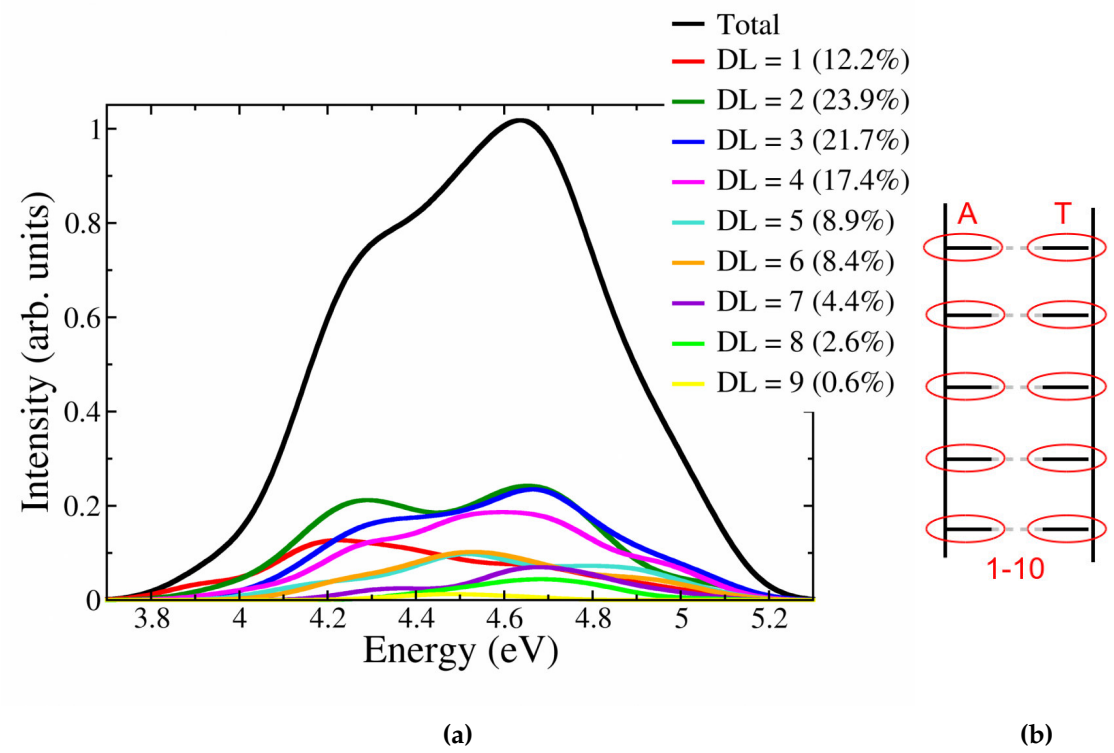


Figure 5: First band of the absorption spectrum of $(dA)_{20}\cdot(dT)_{20}$ DNA (black) and decomposed spectra according to the delocalization length (DL) using the *DL* descriptor of TheoDORE (a). Fragmentation scheme employed for the analysis of DL of the elecronic excitation (b).

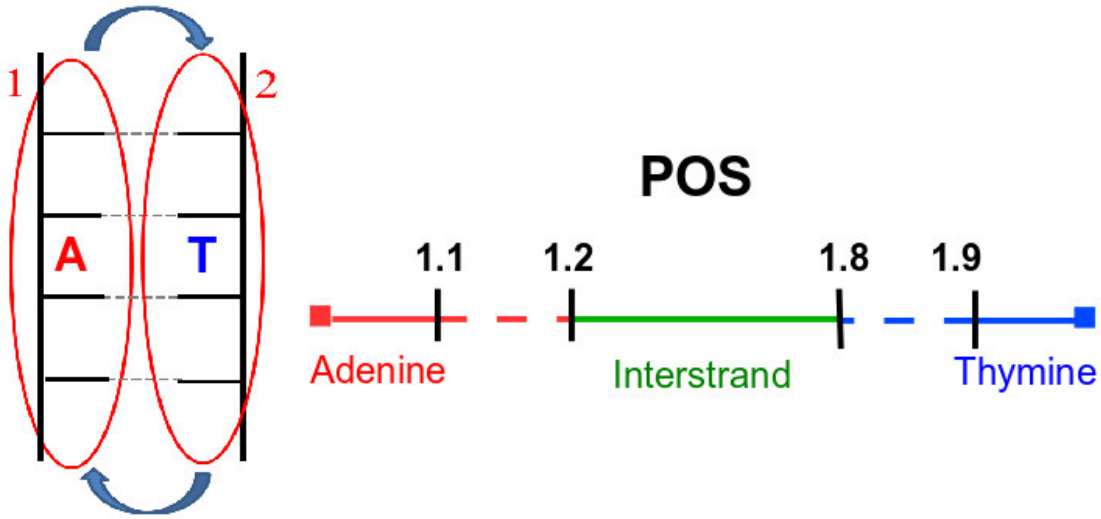
nucleobases) and the average *DL* was 2.2[50], it can be concluded that delocalization is enhanced in the double strand. Physical techniques performed on single-molecule level could prove that single-stranded polynucleotides are more flexible while double helices show more rigid structures[101]. Thus, enhanced delocalization of electronic excitation in the double strand can be attributed to the increased stability of dsDNA compared to a single strand. Although stacking interaction has been seen as the more important effect for the stability of double helices in the literature[102], hydrogen bonds introduce rigidity and, thus, facilitate stronger nucleobase stacking which accounts for more pronounced delocalization.

This analysis is very general and with the fragmentation scheme of Figure 5b it cannot yet be said, if excitations with a *DL* smaller than six remain in one strand or spread across hydrogen bridges. To further investigate on this, the next section provides a sorting of the states in intra- and interstrand excited states.

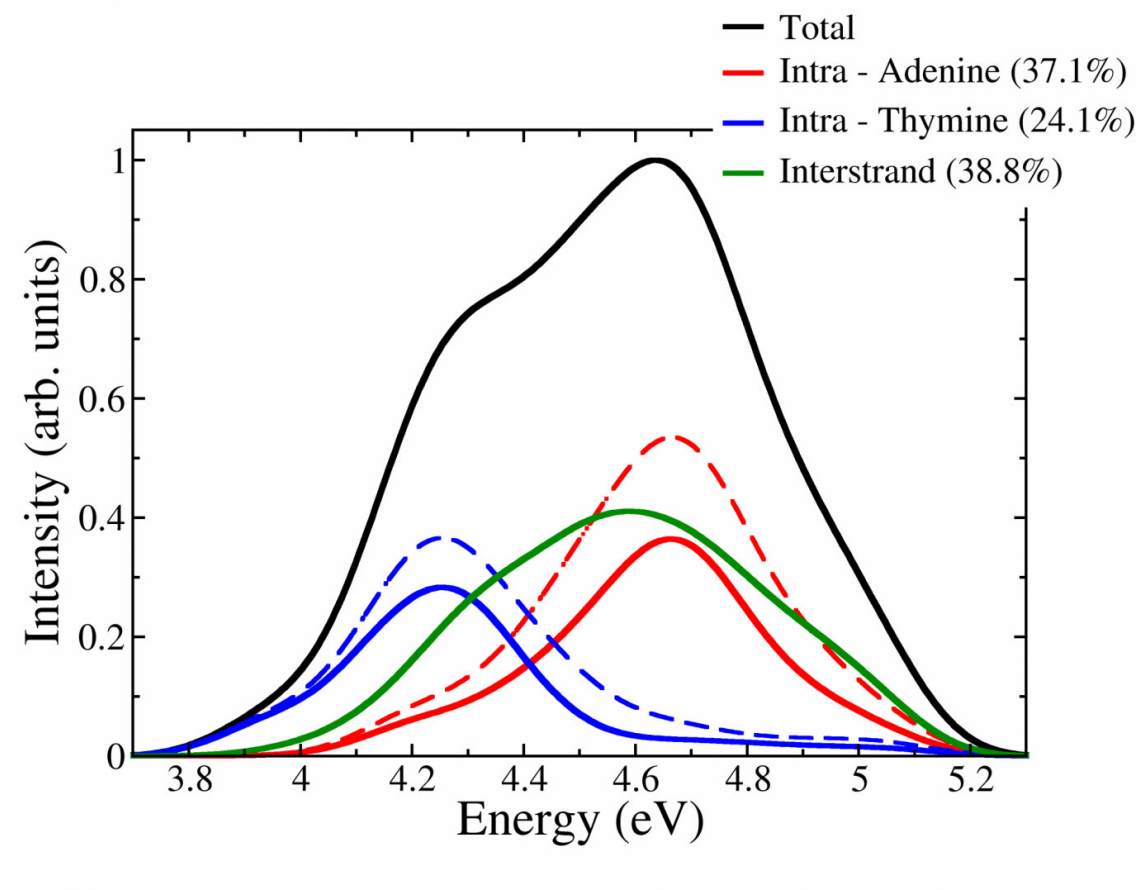
4.3 Separation between intra- and interstrand states

With a decomposition analysis using the information obtained from different fragmentation schemes and descriptors in TheoDORE, more specific information about the absorption behavior can be extracted. The average position of an excitation - the *POS* value explained in Section 2.4 - enables a separation between intra- and interstrand states. The system is divided into two fragments, each containing all nucleobases of one strand (Figure 6a). Ideally, if an excitation reveals a *POS* value of 1 this state is located entirely on the adenine stack. Similarily, in the case of *POS* = 2 the electron and the hole will both be located on the thymine strand. Both possibilities define an intrastrand excitation whereas a value in between involves both strands and is labeled as an interstrand state. However, *POS* will never be exactly 1 or 2 and, thus, we introduce arbitrary thresholds that define ranges (shown in Figure 6a) for classifying the states. *POS* values up to 1.1 and larger than 1.9 indicate intra-adenine and intra-thymine states, respectively. Interstrand states range from *POS* values 1.2 to 1.8. States with a *POS* value between 1.1-1.2 and 1.8-1.9 define buffer regions (dashed lines) and are not considered in further analysis (except in percentage values shown in Figure 6b) since the preference is to investigate ”pure” intra- and interstrand excitations.

Figure 6b shows that excitations located on the thymine strand occur in the red part of the spectrum (lower in energy) and intra-adenine states in the blue part. Interstrand states are in the middle of the spectrum. Additionally, the dashed lines reveal that the separation looks qualitatively similar when changing the thresholds, which strengthens the validity of this arbitrary classification. States involving only adenine nucleobases make a larger fraction (37.1%) of the total amount of states than pure thymine states (24.1%), and al-



(a)



(b)

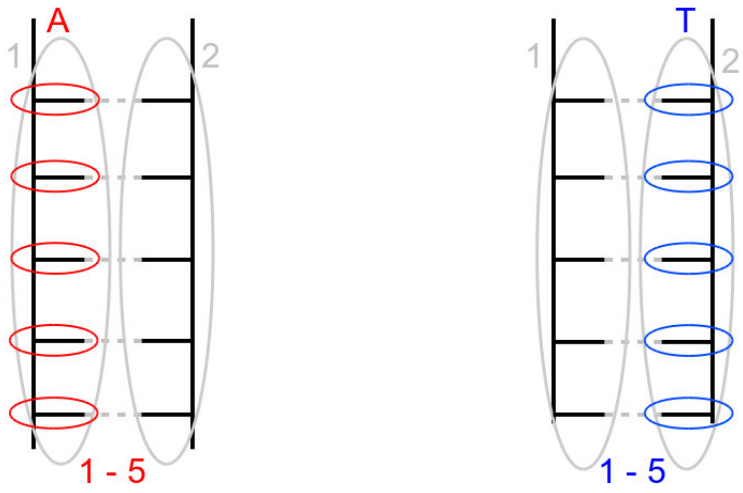
Figure 6: Fragmentation scheme (left) and thresholds for the classification of inter- and intrastrand states (a). First band of the absorption spectrum of $(dA)_{20}\cdot(dT)_{20}$ DNA and decomposition according to the *POS* descriptor into intra- and interstrand excitations (b). Dashed lines in (a) and (b) define buffer regions, see text.

most the same as interstrand excitations (38.8%). Furthermore, the maxima of the red and green lines overlap with the total maximum around 4.6 eV indicating that adenine states contribute more to the interstrand states than thymine ones. The thymine peak seems to account for the shoulder between 4.2 and 4.3 eV. The calculated absorption spectrum of a $(dA)_2 \cdot (dT)_2$ tetramer confirms that the shoulder in the absorption band is due to a delocalized intra-thymine state[100]. For single nucleobases[103] and mono nucleotides[40] the low lying absorption behavior of thymine compared to adenine has been reported and can be explained by the smaller HOMO-LUMO gap for the thymine molecule[104].

Summarizing, the analysis according to the position of the excitations in the DNA stack revealed intra-thymine excitations in the red part, intra-adenine in the blue part of the spectrum and interstrand states distributed in the middle of the absorption band. It is still not clear how many nucleobases are really involved in each type of state. To reveal this information, further investigations on the DL and CT contribution for intra-adenine, -thymine and interstrand excitations, respectively, are performed in following subsections.

4.3.1 Intrastrand excitations

The intrastrand states will be decomposed regarding their DL and CT contribution. To do so, both decompositions of the previous sections will be combined. First, the states are selected according to the position (intra-adenine vs. -thymine). In a second decomposition analysis those states are accounted for their *DL* and *CT* when every single nucleobase in the respective strand defines a single fragment similar to the fragmentation scheme of Figure 5b. Figure 7 summarizes the decomposition analysis and shows the division into single nucleobase fragments (red and blue) when the states were previously selected according to their *POS* value (grey). Since only the individual strands are considered, *DL* can reach values from 1 to 5 nucleobases. For the CT analysis the range between *CT* values of 0 and 1 is devided into ten bins. The distribution of *CT* values will allow a detailed examination of the character of these excited states (Frenkel excitons vs. CT states) and provides information about the location of CT states in the spectrum.



(a) (b)

Figure 7: Fragmentation scheme for the decomposition analysis according to delocalization length (*DL*, 1 - 5) and charge transfer (*CT*) contribution of intra-adenine (a) and intra-thymine (b) excited states.

Intra-Adenine excited states

While discussing the results of the DL and excitation character of intra-adenine states, these findings will be subsequently compared to the theoretical investigations on a poly-adenine single strand of Ref. [50] to investigate the effect of hydrogen bonding. Looking at the plot in Figure 8a, most states (37.3%) are delocalized over two nucleobases. 25% of the states spread over three nucleobases and local states build the third largest fraction (20.2%) that dominate the red tail of the spectrum. A significant amount of states (13.3%) is even spread over four nucleobases while the remaining 3.6% reaches five nucleobases. The maxima of the decomposed graphs overlap close to 4.7 eV except the one for the highest *DL* which is red shifted. A similar study on a polyadenine single strand (Figure 8b, adapted from Ref. [50]) revealed a *DL* of 2 for almost 50% of the states, approximately 23% for local states as well as for a delocalization over three nucleobases. Note that for the single strand the QM region contained eight nucleobases which is the reason why some excited states spread over six chromophores[50]. However, a delocalization involving four nucleobases or more is negligible. This is not the case for intra-adenine states (Figure 8a); here, four or more nucleobases contribute to an excited state to almost 17%. This increased DL leads to the conclusion that the spatial extent of electronic excitation is enlarged upon formation of the double helix. Hydrogen bridging between the strands causes a reduction of dynamic disorder. That is why stacking interactions to neighbouring nucleobases are stronger in the adenine strand and facilitates the delocalization of electronic excitation.

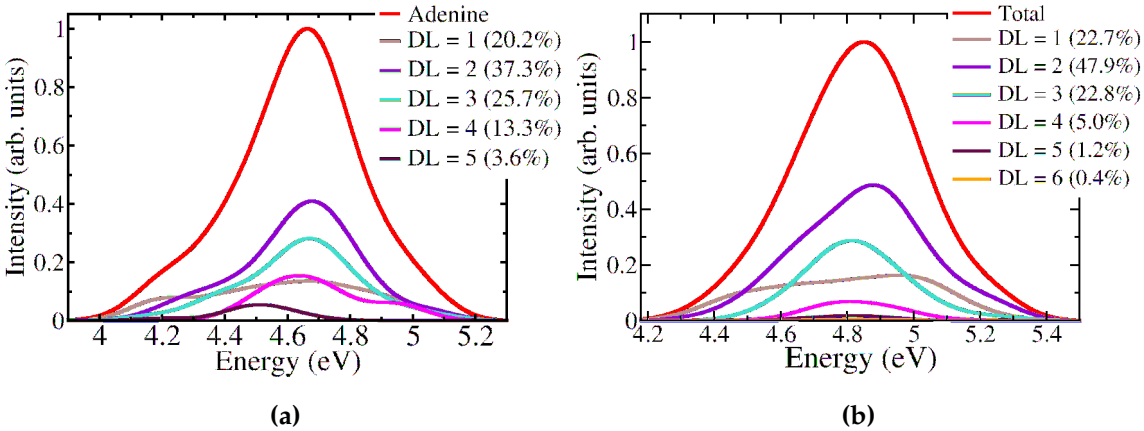


Figure 8: Decomposed absorption spectra regarding the delocalization lengths (DL) in the intra-adenine band of a $(dA)_{20}\cdot(dT)_{20}$ double strand (a) and a $(dA)_{20}$ single strad (b) adapted from Ref. [50].

Figure 9a shows the analysis according to the CT contribution of intra-adenine states. All states with a *CT* value of 0.3 or higher are collected in one curve. 25% of all states have a negligible amount of CT contribution (less than 10%) and are defined ”pure” Frenkel excitons. The majority (approximately 42%) of all intra-adenine excitations have 10-20% CT character whereas 15.3% of the states have a *CT* value between 0.2 and 0.3. Everything above 0.3 is defined as a state with significant amount of CT contribution which applies for 17.8% of all intra-adenine states. The latter states dominate the blue part of the spectrum. An example of a pure intra-adenine CT state (with a $CT > 0.9$) is shown in Figure 10. In the first transition the electron is transferred from two adenines on one end of the stack to two nucleobases on the other end. The second transtion involves a more localized hole in the middle of the adenine strand and an electron orbital with contributions from all adenine nucleobases. The overall DL for this state spreads over five nucleobases.

A computational study with a similar decomposition analysis on the polyadenine single strand (Figure 9b, adapted from Ref. [50]) shows a clear domination of Frenkel excitonic states (51.0%) followed by a CT contribution between 0.1 and 0.2 of 30.3%. Only 7.1% of excited states were characterized by a *CT* of 0.3 and 11.6% of all states present higher CT contribution. Excitations with a non-negligible amount of CT ($CT > 0.3$) are blue shifted with respect to the remaining states[50]. Thus, for a polyadenine oligonucleotide excitons outweigh CT states. However, the CT contribtution of adenine excitations increases when the strand is part of a double helix and more states with a high CT character can be found at even lower energies (e.g. between 4.6 and 4.8 eV) compared to the single strand. A reason for that is the increased rigidity of the double strand due to hydrogen bridges be-

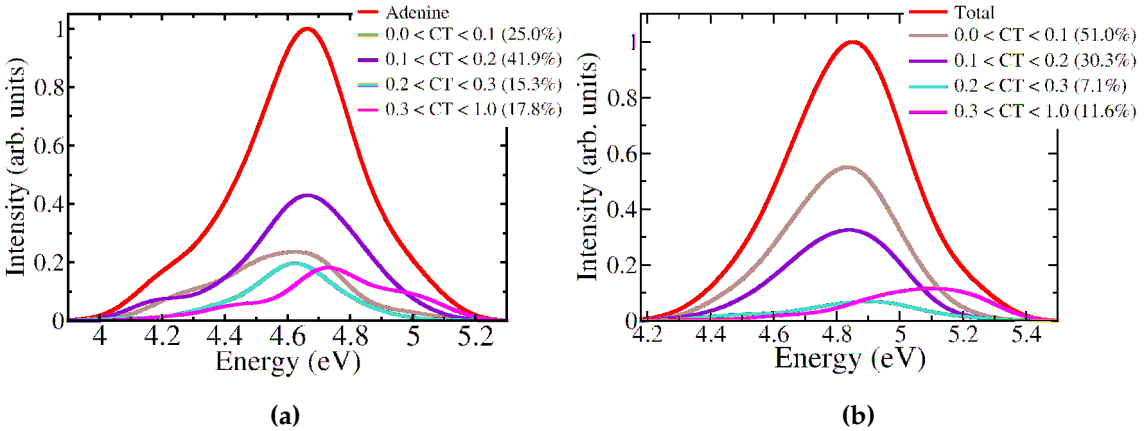


Figure 9: Decomposed absorption spectra regarding the charge transfer (CT) contribution in the intra-adenine band of a $(dA)_{20}\cdot(dT)_{20}$ double strand (a) and a $(dA)_{20}$ single strand (b) adapted from Ref. [50].

tween nucleobase pairs. Consequently, the stacking interaction between step nucleobases is enhanced and causes small interbase distances enabling orbital overlap. Orbital mixing of stacked nucleobases is the predominant requirement for charge separation[105].

To sum up the evaluation of intra-adenine excitations, the states present most likely exciton character with a DL over two nucleobases. CT character delocalization are promoted in a duplex structure in comparison with the single-stranded polyadenine oligonucleotide. Due to the rigidity of the double strand arising from hydrogen bridges, improved alignment and spatial proximity of stacked nucleobases provide orbital overlap and delocalization of the electronic excitation. Local states dominate the red tail and high CT contribution is found towards the blue end of the spectrum, which is consistent with similar analysis in a single adenine strand[50]. After examining the major features of the intra-adenine band, the next section evaluates the same descriptors for intra-thymine states. A comparison to the intra-adenine excitations will be drawn.

Intra-Thymine excited states

Excitations involving only thymine nucleobases are evaluated similar to the previous analysis of intra-adenine states. Figure 11a shows the intra-thymine states of Figure 6b and their decomposition according to the *DL* descriptor. Local and delocalized states over two nucleobases have an equal contribution (37.6%) to the spectrum. A DL over three chromophores make up the third largest fraction (17.9%) and everything above affects the minority of the states (7.0%). The red tail of the spectrum consists of local excitations while the extension towards higher energies is dominated by states involving two thymine nu-

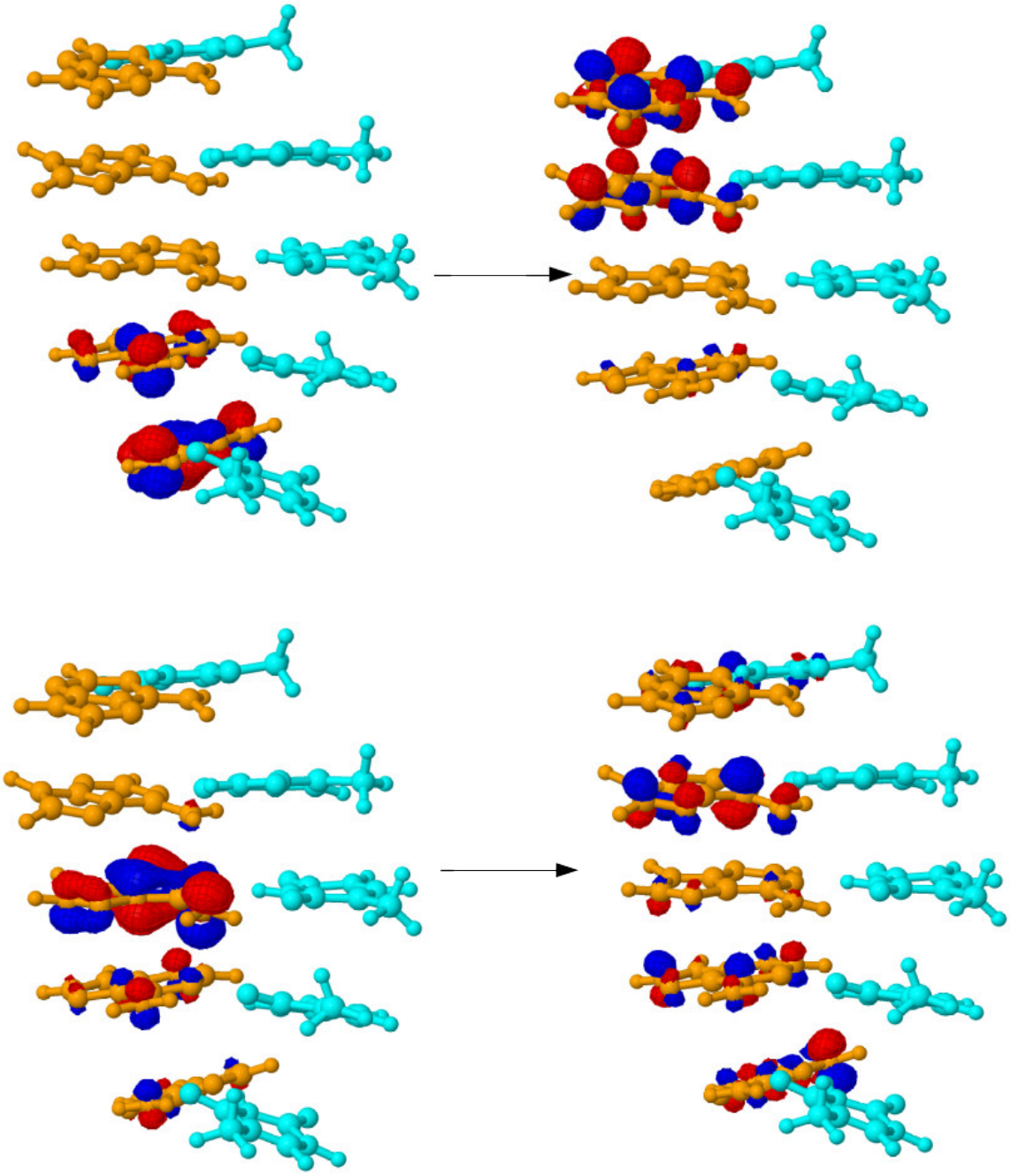


Figure 10: Nucleobases incorporated in the quantum mechanical (QM) region; color code: adenines in orange, thymines in cyan, orbitals in red and blue. Panels on the left and right show hole and electron orbitals, respectively. Two transitions (top and bottom) contribute to this collective intra-adenine charge transfer state ($CT > 0.9$, DL=5 for single nucleobase fragments).

cleobases. All decomposed graphs share their maximum between 4.2 and 4.3 eV except for a *DL* of 5 which is slightly red shifted, similar to hihgly delocalized intra-adenine states. Therefore, high delocalization (over five nucleobases) of intrastrand states is rare but the maxima of their distribution are red shifted compared to more localized states. Generally speaking, the small probability hints to the fact that five stacked nucleobases of a thermal ensemble of geometries are unlikely to align at the same time in order to facilitate delocalization but when they do so, their interactions between each other seems to lower the absorption energy. Nonetheless excitations that require the smallest amount of energy are localized on one nucleobase. Compared to the intra-adenine DL (Figure 8a), thymine excitations are less delocalized. A possible explanation is that thymine, a member of the pyrimidine family, consists of one aromatic ring whereas adenine has two. Hence, the conjugated π electron system of thymine is smaller and stacking interactions are less pronounced compared to purines, such as adenine. Additionally, the experimental absorption spectrum of a poly(dT) single strand resembles the spectrum of its single constituents, which states that stacking interaction between thymines in a single strand is not very strong due to a high structural distortion of the stack[106]. Furthermore, internal conversion in a thymine single strand occurs on a subpicosecond time scale, which is a characteristic feature of monomer thymine nucleobases and reveals that stacking interactions between thymines are insignificant[36]. In the double strand the overall conformational fluctuation is reduced due to hydrogen bonding but the stacking interactions of neighbouring thymine nucleobases seem to remain weaker than those of adenine and, thus, intra-thymine excitation tend to be more localized than intra-adenine states. Experimental results of transient absorption spectroscopy could show that the preferred relaxed excited state pathway for poly(dA)·poly(dT) and poly(dAdT)·poly(dAdT) double helices involve formation of intrastrand excimers[36]. This means that excited states decay to dimeric excitons (excitons delocalized over two adjacent nucleobases) on either of the strands. The calculated vertical excitations describe the absorption behavior in the Franck-Condon region and excited-state dynamics are neccessary to prove this finding computationally. But even in the Franck-Condon region (for the ground state geometry) the formation of excitons delocalized over two nucleobases are the predominant states in the adenine strand while those and local states are equally probable in the thymine strand. Our results show, together with the experimental findings, that electronic delocalization is enhanced in intra-thymine excitations in a double helix compared to the thymine single strand. This highlights the importance of hydrogen bonds for more ideal stacking interactions between neighboring nucleobases that enable excitonic coupling and, thus, delocalization.

The intra-thymine CT decomposition depicted in Figure 11b reveals that 67% of all states are Frenkel excitons (with a CT contribution of less than 10%). The second largest fraction of all thymine states (17.9%), has a *CT* value between 0.2 and 0.3. The remaining

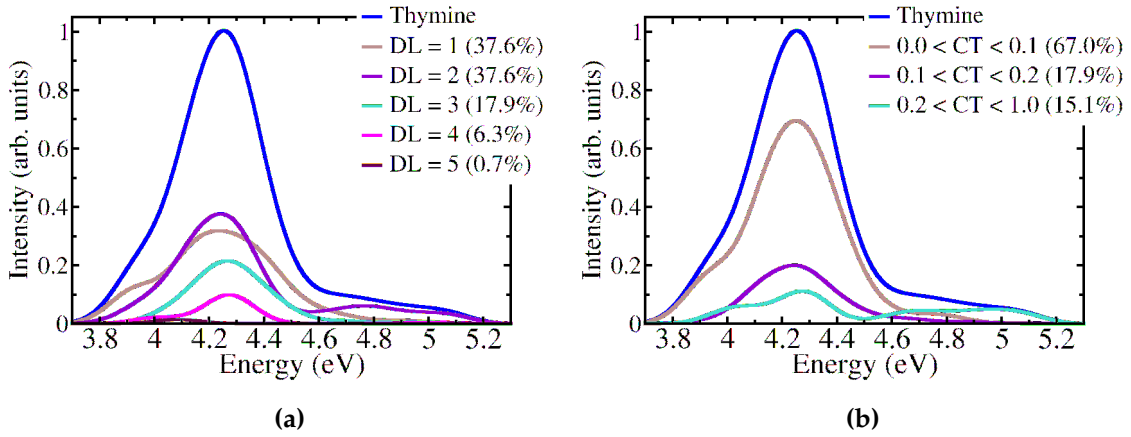


Figure 11: Decomposition of the intra-thymine states according to their delocalization length DL (a) and charge transfer character CT (b).

15.1% show higher CT contributions and occur in the blue extension of the spectrum. In contrast, the red tail is dominated by excitonic states. Compared to intra-adenine states, intra-thymine states favor pure excitons even more (25.0% vs. 67.0%) since stacking interactions and consequently orbital overlap between thymines are reduced, which accounts for less CT. Concomitantly, thymine nucleobases are less likely to present CT states due to their higher ionization potential compared to adenine[107]. Thus, it takes more energy to extract an electron from thymine than from adenine. To give an example of one of the most common thymine transitions, Figure 12 represents an exciton delocalized over two neighbouring thymines.

A more detailed picture of the distribution of *CT* values of intrastrand adenine- and thymine-states can be seen in Figure 13. This way it is even more clear that intrastrand excitations are dominated by excitonic states since small *CT* values have a high probability. This effect is more important for intra-thymine excitons than for intra-adenine states. Moreover, mixed intrastrand states with more or less equal amounts of exciton and CT character and pure CT states have only little contribution to the spectrum for both types of nucleobases.

In summary, thymine is less likely to present delocalized electronic excitation and states with significant CT contribution than adenine. The more local character arises mainly due to the decreased stacking interactions between thymines while the smaller CT rate comes from higher ionization potential[107] as well as from smaller stacking interactions.

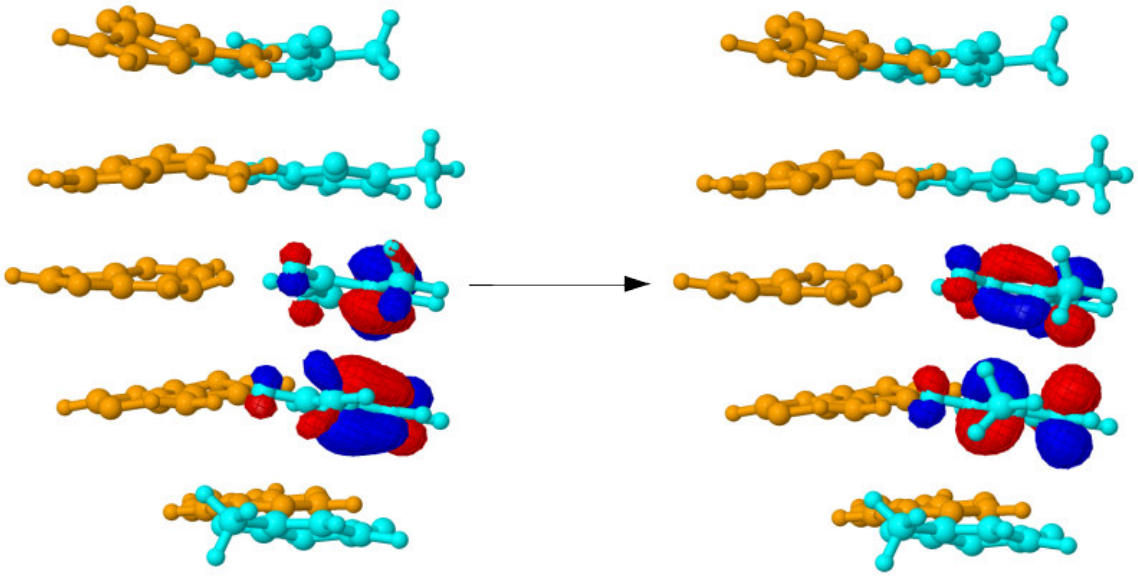


Figure 12: Quantum mechanical (QM) region of 10 individual nucleobases; color code: adenines in orange, thymines in cyan, orbitals in red and blue. Exemplary intra-thymine exciton (CT=0, DL=2 for single nucleobase fragments). Hole (left) and electron (right) orbitals spread over two nucleobases.

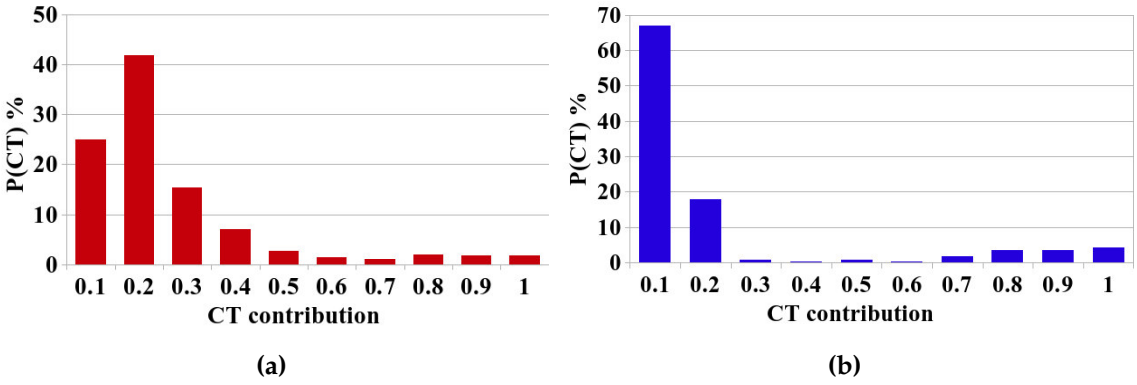


Figure 13: Charge transfer contributions (CT) of intra-adenine (a) and intra-thymine (b) excitations similar to the graphs in Figures 9a and 11b, respectively, in more detailed representations showing all ten bins in histograms.

4.3.2 Interstrand excitations

Excited states that are not limited to one strand are selected according to a *POS* value between 1.2 and 1.8, as explained in section 4.3. This means that each strand is contributing to the state at least 20%. With a subsequent division into single nucleobase fragments (see Figure 14a), the extent of the delocalization across strands can be investigated. Figure 14b shows the normalized absorption for this class of states. The decomposed graphs were multiplied by two in order to obtain a clearer picture since multiple bins have similar contributions to the total spectrum. This is the reason why the pink line in Figure 14b overlays the green interstrand band at the very red end of the spectrum. This plot demonstrates that interstrand excitations are broadly distributed across multiple nucleobases. Interestingly, states delocalized over four nucleobases dominate the red tail of the decomposed absorption band. Five out of eight *DL* bins show a contribution of more than 10%, namely 4 (22.3%), 6 (18.2%), 3 (16.3%), 5 (16.0%) and 7 (11.4%), illustrating the strong interaction between the strands. Although the excitation seems to be never delocalized over all ten nucleobases, the states with *DL* values of 8 and 9 (8.1%) make together a fraction larger than that of the states that involve only two nucleobases (7.7%). Figure 14c collects the information in form of a histogram which highlights the various contributions of different *DL*s to the total amount of interstrand states. It has to be pointed out that excitations over eight nucleobases are still important. This large delocalization across the strands is a consequence of hydrogen bonds between the helices that stabilize the duplex and, furthermore, enable electronic coupling, which facilitates delocalization of electronic excitation. Thus, the importance of hydrogen bonding is evident.

Interstrand CT states are identified as a transfer of an electron from an orbital located on one strand to an orbital of the other strand. Therefore, the system is divided in two fragments, each containing all nucleobases of one type as schematically shown in Figure 15a. Figure 15b illustrates that 21.1% of interstrand states are pure excitons ($CT < 0.1$). About the same amount of states have CT values between 0.1 and 0.2 or 0.2 and 0.3. Most states (57.4%), however, are identified with a significant (>30%) CT contribution. Although these states are most likely concentrated in the middle and towards the blue end of the interstrand peak, they also dominate the very red tail of the band. However, at the location of the shoulder in the total spectrum (4.3 eV) the interstrand mixed states show only partial contribution to the spectrum and interstrand excitons are dominating. Comparing interstrand excitations with intrastrand states, the former have much larger CT contribution. Thus, hydrogen bonding favors CT character of electronic excited states. To further analyse the CT contribution, Figure 15c shows the probability distribution of CT fractions. It can be seen that, in contrast to intrastrand excitations, interstrand states have most probably mixed character with a significant amount of CT character ($0.2 < CT < 0.9$). Investigations

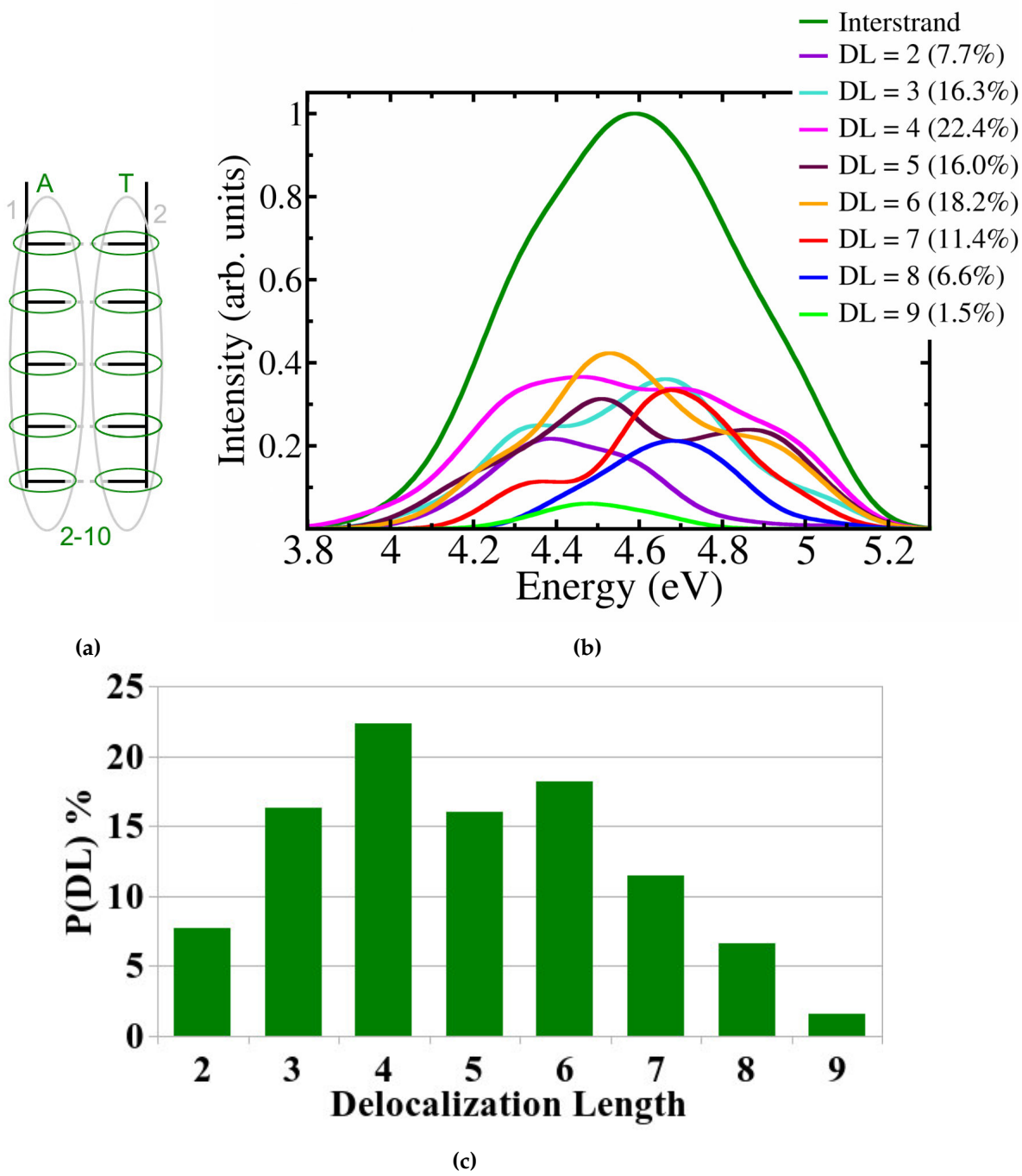


Figure 14: Fragmentation scheme employed analysing the delocalization length (DL) of interstrand excitations (a). Decomposition of interstrand states according to their DL using the *DL* descriptor in TheoDORE shown as graphs (b), decomposed graphs (*DL*=2 to *DL*=9) multiplied by 2 for clarity, and as a histogram (c).

on "real" interstrand CT states ($CT > 0.8$) revealed that electron transfer occurs solely from adenine (electron donor) to thymine (electron acceptor). An example can be seen in Figure 16. This can be justified with the lower ionization potential of adenine[104].

Comparing the CT contribution of intra- (Figure 13) and interstrand (Figure 15b) states, the latter are more likely of mixed character with stronger CT contribution, which leads to the conclusion that hydrogen bonding favors CT more than interbase stacking interactions. This is in agreement with quantum chemical calculations that revealed orbital interactions between hydrogen bond donor and acceptor atoms extending the pure electrostatic picture of hydrogen bridges[41]. Recalling, orbital interactions enable the formation of CT states[40]. A comparable computational study of $(dA)_{15}\cdot(dT)_{15}$ in aqueous solution employing conformational sampling with a QM region of six nucleobase pairs found that only intrastrand CT occurs while interstrand electron transfer was not observed[44]. However, the results of the previous sections elucidated that intrastrand states are mainly of exciton character while interstrand excitations involve high CT contribution. This contrast may arise from different quantum mechanical methods (semiempricial in Ref. [44] vs. TDDFT), time windows spanned by the sampling analysis (100 ps in Ref. [44] vs. 25 ns) and subsequent analysis procedures (difference densities in Ref. [44] vs. transition densities). Due to the more sophisticated computations performed here, it is assumed that interstrand states with high CT contribution play an important role in the electronic delocalization of nucleic acid duplexes upon photoexcitation. In a review by Markovitsi it is stated that mixed Frenkel exciton and CT character appears as a fingerprint in the red tail of the main absorption band[105]. Our analysis on the CT character of excited states revealed that mixed states play an important role only for interstrand excitations, which are more likely to be located in the middle of the total absorption band. Within the interstrand peak, mixed states are concentrated towards higher energies, as seen in Figure 15b. After examining all decomposed spectra, it can be assumed that the shoulder in the total absorption spectrum (e.g. in Figure 5a) is unlikely to arise only from low-lying CT states as Ref. [45] suggested. Mixed states with high CT contribution appear more probably towards the blue extension of the respective spectra albeit more relevant for excitation across hydrogen bridges. With our analysis we found a combination of intra-thymine local states and excitons with mainly interstrand excitons at the position of the shoulder in the total absorption band (4.3 eV). Interstrand mixed states show only partial contribution to the spectrum at this energy.

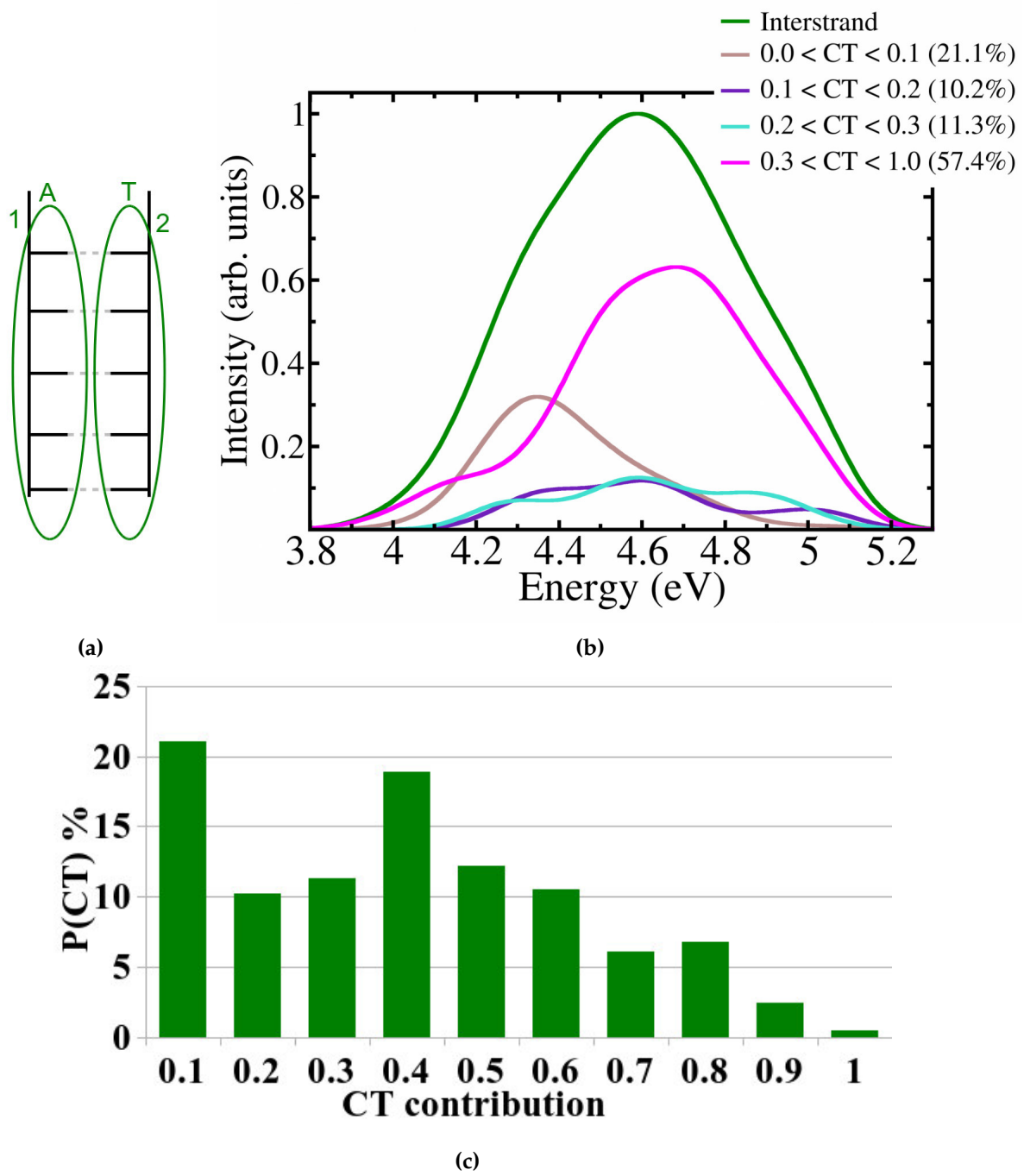


Figure 15: Fragmentation scheme employed analysing interstrand charge transfer (CT) states (a). Decomposition of interstrand states according to their CT character using the *CT* descriptor in TheoDORE shown as graphs (b) and as a histogram (c).

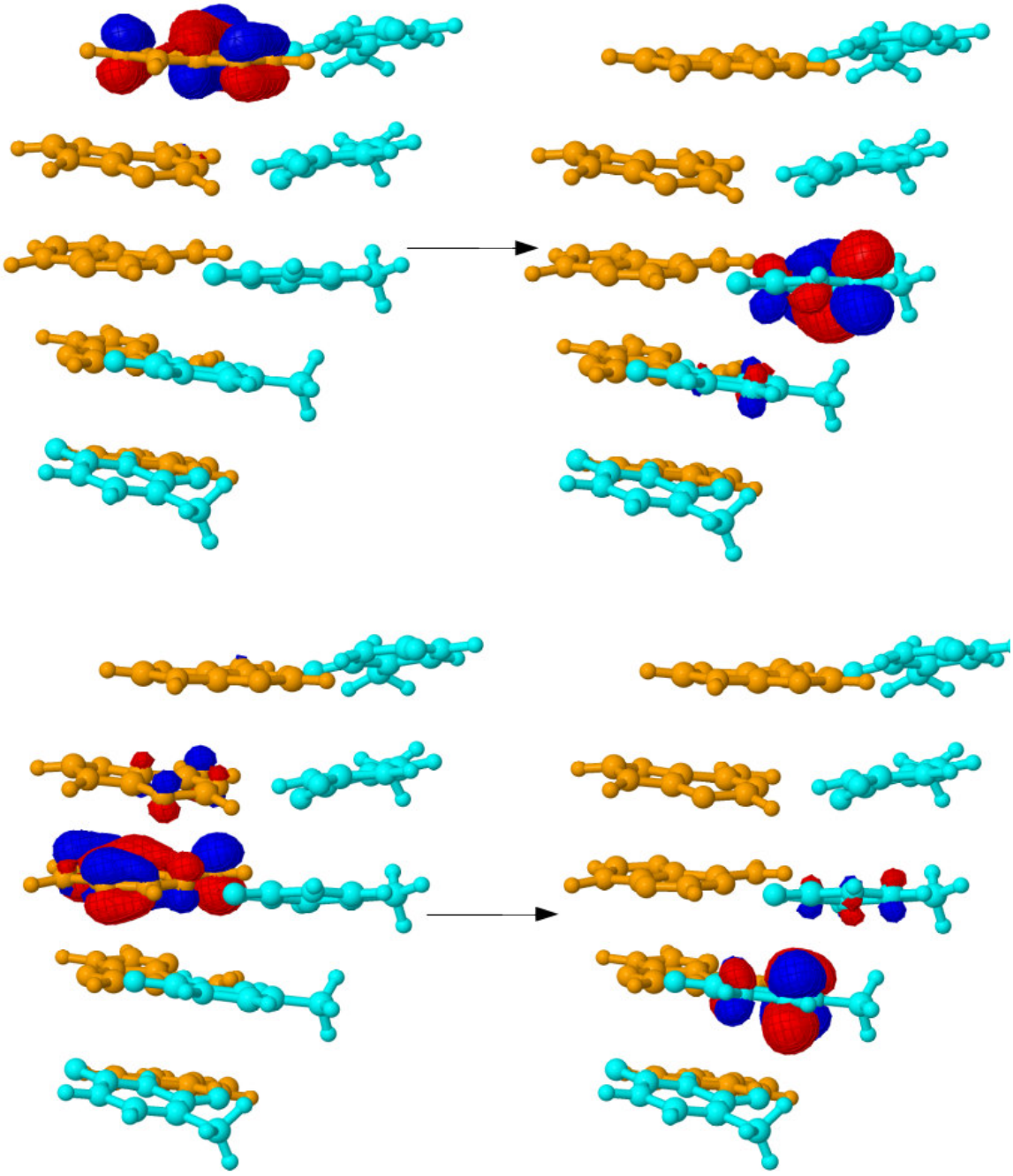


Figure 16: Interstrand charge transfer state (CT>0.9, DL=2 for single strand fragments) consisting of two diagonal transitions, color code: adenines in orange, thymines in cyan, orbitals in red and blue. Hole orbitals are located on the adenine strand (electron donor) while the electron is transferred to thymine orbitals (electron acceptor).

5 Conclusion and outlook

In the following the major findings of this work are summarized. After that, a quick outlook on possible future projects in this field is provided.

This study has shown that for a theoretical evaluation of electronic excited states in a $(dA)_{20}\cdot(dT)_{20}$ DNA it is neccessary to include at least five base pairs in the QM region when performing QM/MM calculations and investigating on the DL since otherwise the spatial extent of the delocalization is underestimated. Furthermore, nucleobases on the edges of the QM stack face a different MM environment and are, thus, likely to introduce artifacts.

The convoluted UV absorption spectrum, generated using the first 60 electronic excited states per geometry considering a total of 100 geometries, peaks at 4.6 eV and shows a shoulder energetically below the absorption maximum at 4.3 eV. According to the separation between inter- and intrastrand excitations, this shoulder is mainly attributed to intra-thymine local states and excitons as well as interstrand excitons. The assumption by Markovitsi et al[45], that the shoulder arises due to low-lying CT states can only partially be confirmed, as interstrand mixed states with significant CT contribution play a minor role at 4.3 eV. The most abundant delocalization of the electronic excitation in the thermal ensemble occurs over two and three nucleobases, respectively. Evidence is provided that delocalization of electronic excitation is enhanced in the double strand due to hydrogen bonds between the strands. These weak interactions are responsible for a more rigid structure of the DNA stack. That way, neighboring nucleobases present higher alignment inducing stronger stacking interactions, which in turn produce larger electronic coupling between nucleobases and result in enhanced delocalization.

A separation between intra- and interstrand excited states showed intra-thymine states in the red, intra-adenine states in the blue part of the spectrum and interstrand excitations in the middle of the absorption peak. Excited states on adenine nucleobases are mainly of exciton character but the overall CT contribution and DL are enhanced in contrast to a single-stranded adenine oligonucleotide[50]. Intra-thymine states are in general less delocalized compared to intra-adenine excitations due to weaker stacking interactions between thymines compared to adenines. This is explained with the larger conjugated π electron system of purines. However, compared to experimental results on a polythymine single strand[106], intra-thymine states of a double strand are more delocalized. This highlights the importance of hydrogen bridges through which the rigidity of the double strand is increased and stacking interactions are enhanced. Thymine excitations favor Frenkel excitons even more than adenine states since it is easier to detach an electron from adenine[107] and, additionally, orbital overlap, that accounts for CT, is less distinct in the thymine stack compared to its purine counterpart.

Interstrand excitations are highly delocalized and present significant CT contribution. Again, hydrogen bonds account for the increased interbase interactions. These conclusions give rise to the assumptions that hydrogen bonds favor CT states while stacking interactions give preference to exciton states.

This study has shed light on the complex nature of electronic excited states in a double-stranded DNA fragments and stressed the importance of hydrogen bonds for the rigidity of the double helix and in excited state formation. The reduced flexibility enables enhanced couplings among nucleobases in the stack that result in electronic delocalization. Furthermore, interbase separation is reduced due to a more compact structure, which accounts for the formation of orbital overlap and, thus, CT contribution.

After summarizing the main results of this study, the following paragraphs give an idea of what is still missing in the investigation of the absorption behavior of polynucleotides and provide a suggestion for future work in this field.

The example for interstrand CT states (Figure 16) shows a diagonal electron transfer. It would be interesting to investigate whether parallel interstrand CT states, which involve only one nucleobase pair in the transfer, are present as well. Therefore, one could think of a proper way to distinguish between those states in the analysis and determine the ratio between diagonal and parallel transfer to find out which transition is favored and why.

To be consistent, a similar theoretical study of a poly(dT) single strand would have to be performed in order to prove the aforementioned assumptions regarding intra-thymine states and confirm experimental results[36]. This way, intra-thymine excitations occurring in the double strand could be better compared to excited states of an all thymine strand.

One could as well investigate the origin of the hypochromism as one phenomenon in UV absorption of polynucleotides, as outlined in the introductory chapter. Research in this area is still ongoing. While experimentalists claim that high energy mixed states with important CT character are responsible for the decrease in absorption intensity[49], a more recent theoretical study suggests the influence of long-range effects on the perturbation of monomer like excitations as a possible reason for hypochromism[50].

Optical spectroscopy is not able to detect non-radiative relaxation pathways after photoexcitation in a direct manner[51]. Here, vertical excitations were calculated that help to understand the primary absorption behavior in the Franck-Condon region. To better link these findings to experimental results and to reveal the relaxation mechanisms outside the Franck-Condon region, nonadiabatic excited state dynamics have to be performed[108]. Since this requires extensive computational costs, this procedure is only feasible for single nucleobases or small model systems[109]. Further methodological and technical development could extend the capacity of quantum mechanical calculations and enable to predict and simulate excited state dynamics of bigger systems, e.g. a double stranded polynucleotide. These calculations could unravel the relaxation mechanisms in double strands

that explain the evolution from vertical excitation to the experimentally observed long-living excited states and reveal the differences to excited state relaxation in monomers[110].

A decomposition analysis like in this study would be interesting for guanine-cytosine nucleobase pairs of a DNA duplex. As mentioned in the introductory chapter and in Figure 1b, GC pairs possess one hydrogen bond more than AT. Thus, a similar analysis on GC DNA could help to outline the role of hydrogen bonds for excited state formation in more detail. According to our results, more CT contribution and delocalization is expected. Furthermore, experimental investigations using fluorescence upconversion spectroscopy revealed a reduced lifetime for relaxed excited states in a GC DNA compared to its single nucleobases[111]. For DNA double strands consisting only of adenine and thymine nucleobases this is not the case. Excited states of the latter show increased lifetime in the experiment compared to single adenine and thymine monomers[36]. Thus, a computational analysis on GC excited states in a duplex structure could help to understand the differences observed experimentally. In addition to that, testing the sequence dependence of excited state formation could be performed by using alternating nucleobase pairs as in Ref. [54] or a random combination of successive nucleobase pairs. With that, the consequences of UVR for DNA could be understood not only for a model system[112].

Last but not least, studies on DNA molecules include the investigation on charge delocalization. It is known that the three dimensional structure of the DNA double helix enables charge migration along the helical axis which makes it likely to become a potential biomaterial for nanoelectronic sensors[113]. Guanine, having the lowest ionization potential, is carrier of the cationic species after charge separation[107]. Adenine can be a charge carrier as well, if GC pairs are separated from each other by AT pairs inbetween. The charges are conserved and transfer is occurring over a long distance, e.g. 200 Å. However, the environment plays a pivotal role as for example water molecules are able to trap the charge and prohibit further migration[114] or ions in the proximity of the stack mediate the transport[115]. Research in this field yields towards conductivity of DNA using modified nucleobases that increase the rigidity of the stack and subsequently enhance charge delocalization. While this is already done experimentally, further computational studies could help to improve the understanding of electron- and hole-transfer kinetics[113].

6 Appendix

Zusammenfassung

UV-Strahlung schädigt das Erbgut und fördert dadurch die Mutagenese und die Entwicklung von Hautkrebs. Sowohl die photophysikalischen Prozesse, die zu diesen Schäden führen, als auch deren biologische Auswirkungen werden intensiv erforscht. Trotzdem ist die Photophysik von Nukleinsäuren bislang nur wenig verstanden. Benachbarte Nukleinsäuren (Basen) eines DNA-Fragments sind elektronisch gekoppelt und weisen deshalb ein kooperatives Absorptionsverhalten auf. Somit entstehen elektronisch angeregte Zustände, die sich auf mehrere Basen verteilen, man spricht von einer Delokalisation der Anregung. Zu diesen delokalisierten Zuständen zählen Excitons und charge-transfer-Anregungen ("Ladungsaustausch"). Das räumliche Ausmaß der Delokalisation und die Rolle der charge-transfer-Anregungen für das Absorptionsspektrum werden in der Literatur diskutiert. Außerdem ist nicht bekannt, wie Wasserstoffbrückenbindungen und π-π-Wechselwirkungen zwischen den Basen die Formation angeregter Zustände beeinflussen. Diese Arbeit soll zur computergestützten Analyse elektronisch angeregter Zustände eines doppelsträngigen DNA-Fragments, die durch UV-Absorption gebildet werden, beitragen. Ziel ist es einerseits, das räumliche Ausmaß der Delokalisation zu bestimmen. Weiterhin soll die Rolle von charge-transfer-Zuständen und die Auswirkung der Wechselwirkungen zwischen den Basen auf die Formation angeregter Zustände beurteilt werden. Dafür werden elektronisch angeregte Zustände eines $(dA)_{20}\cdot(dT)_{20}$ Ensembles, das einer klassischen Molekulardynamik-Simulation entnommen wurde, mittels Quantenmechanik-Molekularmechanik berechnet. Detaillierte Wellenfunktionsanalyse dieser Zustände konnte zeigen, dass π-π-Wechselwirkungen zwischen den Basen Excitons begünstigen. Wasserstoffbrückenbindungen hingegen unterstützen die Ausbildung von Zuständen mit signifikantem charge-tranfer-Charakter. Daraus resultiert, dass innerhalb eines Strangs fast ausschließlich Excitons vorherrschen. Charge-transfer-Zustände spielen für Thymine noch weniger eine Rolle als für Adenine. Dies wird durch die geringe Orbitalüberlappung zwischen den Basen eines Strangs verursacht, welche für benachbarte Thymine kleiner ist als für Adenine. Außerdem breiten sich Exciton Anregungen im Adenin-Strang durchschnittlich über mehr Basen aus als im Thymin-Strang. Zustände, die beide Stränge involvieren, sind noch delokalisierter und weisen ausgeprägteren charge-transfer-Charakter auf. Allgemein kann gesagt werden, dass die Anregung im Doppelstrang delokalisierter ist als in separaten Einzelsträngen, was auf die rigide Struktur des Doppelstrangs zurückzuführen ist, welche durch Wasserstoffbrückenbindungen zwischen den Strängen hervorgerufen wird. Zustände mit signifikantem charge-transfer-Charakter sind nur zwischen den Strängen zu beobachten, wobei ein Elektron stets von Adenin an Thymin übertragen wird. Sie sind eher im Bereich höherer Energien zu finden. Die hier präsentierten Ergebnisse fungieren als wichtiger Bestandteil der Erforschung elektronisch angeregter Zustände in doppelsträngigen Nukleinsäurepolymeren.

Abstract

UV radiation induces photodamage to DNA that promotes mutagenesis and the development of skin cancer. The photophysics behind the formation of photolesions and their consequences have been intensively studied. However, the former remains poorly understood. Electronic coupling between neighboring nucleobases in DNA fragments accounts for cooperative absorption behavior and, thus, delocalized electronic excited states (excitons and charge-transfer states), which are absent in nucleobase monomers, are relevant in the polymer. Thereby, the spatial extent of delocalization and the importance of charge-transfer states for the spectrum are highly discussed in the literature. Furthermore, the effects of hydrogen bonding and π stacking between nucleobases for excited-state formation are unknown. This thesis contributes to the computational evaluation of the excited states of the first UV absorption band of a double-stranded DNA fragment and aims to unravel the spatial extent of delocalization, the role of charge-transfer states and the contribution of interbase interactions. Therefore, excited states of a thermal ensemble of $(dA)_{20}\cdot(dT)_{20}$ from a classical molecular dynamics simulation in explicit water are calculated with a quantum mechanics/molecular mechanics scheme. A detailed wavefunction analysis of excited states showed that π stacking favors exciton formation, whereas hydrogen bonding accounts for charge-transfer contribution. As a consequence of that, intrastrand excited states are mainly exciton like, while highly delocalized interstrand excitations possess mixed character with significant charge-transfer contribution. The spatial extent of intra-adenine excitons is larger than that of intra-thymine states. Compared to single-stranded fragments, delocalization is enhanced upon formation of the double helix, which can be attributed to the increased rigidity of the duplex due to hydrogen bonds between strands. Additionally, intra-thymine states show less charge-transfer character than intra-adenine excited states since orbital overlap, which facilitates charge-transfer contribution, is reduced in stacked pyrimidine nucleobases. States with significant charge-transfer character are pivotal only for interstrand states (electron transfer from adenine to thymine) and are concentrated towards the blue end of the spectrum. Thus, the results of this work are important for understanding excited-state formation in double-stranded nucleic acid polymers.

7 References

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Molecular biology of the cell", *Classic textbook now in its 5th Edition*, New York: Garland Science; 2002.

[2] O. T. Avery, C. M. MacLeod, and M. McCarty, "Studies on the chemical nature of the substance inducing transformation of pneumococcal types", *Journal of experimental medicine*, vol. 79, no. 2, pp. 137–158, 1944.

[3] J. D. Watson, F. H. Crick, *et al.*, "Molecular structure of nucleic acids", *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[4] S. Zamenhof, G. Brawerman, and E. Chargaff, "On the desoxypentose nucleic acids from several microorganisms", *Biochimica et Biophysica Acta*, vol. 9, pp. 402–405, 1952.

[5] M. Nirenberg and P. Leder, "RNA codewords and protein synthesis", *Science*, vol. 145, no. 3639, pp. 1399–1407, 1964.

[6] M. Nirenberg, P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'neal, "RNA codewords and protein synthesis, VII. on the general nature of the RNA code", *Proceedings of the National Academy of Sciences*, vol. 53, no. 5, pp. 1161–1168, 1965.

[7] A. Sancar, "Structure and function of DNA photolyase", *Biochemistry*, vol. 33, no. 1, pp. 2–9, 1994.

[8] F. L. Gates, "A study of the bactericidal action of Ultra Violet Light", *The Journal of General Physiology*, vol. 14, no. 1, pp. 31–42, 1930.

[9] J. Cleaver, "Defective repair replication of DNA in Xeroderma Pigmentosum", *Nature*, vol. 218, no. 5142, pp. 652–656, 1968.

[10] M.-D. Chou and K.-T. Lee, "Parameterizations for the absorption of solar radiation by water vapor and ozone", *Journal of the Atmospheric Sciences*, vol. 53, no. 8, pp. 1203–1208, 1996.

[11] International Agency for Research on Cancer, IARC monographs on the evaluation of carcinogenic risks to humans: solar and ultraviolet radiation. Vol 55. Lyon, France: World Health Organization", *International Agency for Research on Cancer*, 1992.

[12] M. Peak and J. Peak, "Effects of solar ultraviolet photons on mammalian cell DNA.[UVA (320-400 nm): a2]", Argonne National Lab., IL (United States), 1991.

[13] M. J. Peak, J. G. Peak, and C. A. Jones, "Different (direct and indirect) mechanisms for the induction of DNA-protein crosslinks in human cells by far-and near-ultraviolet radiations (290 and 405 nm)", *Photochemistry and Photobiology*, vol. 42, no. 2, pp. 141–146, 1985.

[14] M. J. Peak, J. G. Peak, and B. A. Carnes, "Induction of direct and indirect single-strand breaks in human cell DNA by far-and near-ultraviolet radiations: action spectrum and mechanisms", *Photochemistry and Photobiology*, vol. 45, no. 3, pp. 381–387, 1987.

[15] J. G. Peak and M. J. Peak, "Ultraviolet light induces double-strand breaks in DNA of cultured human p3 cells as measured by neutral filter elution", *Photochemistry and Photobiology*, vol. 52, no. 2, pp. 387–393, 1990.

[16] S. Mouret, C. Baudouin, M. Charveron, A. Favier, J. Cadet, and T. Douki, "Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation", *Proceedings of the National Academy of Sciences*, vol. 103, no. 37, pp. 13765–13770, 2006.

[17] R. Davies, "Ultraviolet radiation damage in DNA", *Biochemical Society Transactions*, vol. 23, no. 2, pp. 407–418, 1995.

[18] R. Setlow and W. Carrier, "Pyrimidine dimers in ultraviolet-irradiated DNA's", *Journal of Molecular Biology*, vol. 17, no. 1, pp. 237–254, 1966.

[19] G. P. Pfeifer, "Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment", *Photochemistry and Photobiology*, vol. 65, no. 2, pp. 270–283, 1997.

[20] J. R. Pehrson and L. H. Cohen, "Effects of DNA looping on pyrimidine dimer formation", *Nucleic Acids Research*, vol. 20, no. 6, pp. 1321–1324, 1992.

[21] D. E. Brash and W. A. Haseltine, "UV-induced mutation hotspots occur at DNA damage hotspots", *Nature*, vol. 298, no. 5870, pp. 189–192, 1982.

[22] J.-S. Taylor, H.-F. Lu, and J. J. Kotyk, "Quantitative conversion of the (6–4) photoproduct of TpdC to its Dewar valence isomer upon exposure to simulated sunlight", *Photochemistry and Photobiology*, vol. 51, no. 2, pp. 161–167, 1990.

[23] P. H. Clingen, C. F. Arlett, L. Roza, T. Mori, O. Nikaido, and M. H. Green, "Induction of cyclobutane pyrimidine dimers, pyrimidine (6-4) pyrimidone photoproducts, and Dewar valence isomers by natural sunlight in normal human mononuclear cells", *Cancer Research*, vol. 55, no. 11, pp. 2245–2248, 1995.

[24] http://www.parp4.com/antibodies_against_dna_damage.htm.
accessed 15th August 2017.

[25] X. Qin, S. Zhang, M. Zarkovic, Y. Nakatsuru, S. Shimizu, Y. Yamazaki, H. Oda, O. Nikaido, and T. Ishikawa, "Detection of ultraviolet photoproducts in mouse skin exposed to natural sunlight", *Cancer Science*, vol. 87, no. 7, pp. 685–690, 1996.

[26] J. A. Lippke, L. K. Gordon, D. E. Brash, and W. A. Haseltine, "Distribution of UV light-induced damage in a defined sequence of human DNA: detection of alkaline-sensitive lesions at pyrimidine nucleoside-cytidine sequences", *Proceedings of the National Academy of Sciences*, vol. 78, no. 6, pp. 3388–3392, 1981.

[27] E. C. Friedberg, G. C. Walker, W. Siede, and R. D. Wood, *DNA Repair and Mutagenesis*. American Society for Microbiology Press, 2005.

[28] K. Smith, *The Science of Photobiology*. Springer Science & Business Media, 2013.

[29] M. Barbatti, A. J. Aquino, J. J. Szymczak, D. Nachtigallová, P. Hobza, and H. Lischka, "Relaxation mechanisms of UV-photoexcited DNA and RNA nucleobases", *Proceedings of the National Academy of Sciences*, vol. 107, no. 50, pp. 21453–21458, 2010.

[30] M. Daniels and W. Hauswirth, "Fluorescence of the purine and pyrimidine bases of the nucleic acids in neutral aqueous solution at 300 K", *Science*, vol. 171, no. 3972, pp. 675–677, 1971.

[31] J. Eisinger and R. Shulman, "Excited electronic states of DNA", *Science*, vol. 161, no. 3848, pp. 1311–1319, 1968.

[32] E. Sage, "Distribution and repair of photolesions in DNA: genetic consequences and the role of sequence context", *Photochemistry and Photobiology*, vol. 57, no. 1, pp. 163–174, 1993.

[33] E. R. Bittner, "Frenkel exciton model of ultrafast excited state dynamics in AT DNA double helices", *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 190, no. 2, pp. 328–334, 2007.

[34] P.-O. Löwdin, "Proton tunneling in DNA and its biological implications", *Reviews of Modern Physics*, vol. 35, no. 3, p. 724, 1963.

[35] T. Schultz, E. Samoylova, W. Radloff, I. V. Hertel, A. L. Sobolewski, and W. Domcke, "Efficient deactivation of a model base pair via excited-state hydrogen transfer", *Science*, vol. 306, no. 5702, pp. 1765–1768, 2004.

[36] C. E. Crespo-Hernández, B. Cohen, and B. Kohler, "Base stacking controls excited-state dynamics in AT DNA", *Nature*, vol. 436, no. 7054, p. 1141, 2005.

[37] J. Singh, "Theory of excitons", in *Excitation Energy Transfer Processes in Condensed Matter*, pp. 1–45, Springer, 1994.

[38] L. Blancafort, A. A. Voityuk, "Exciton delocalization, charge transfer, and electronic coupling for singlet excitation energy transfer between stacked nucleobases in DNA: An MS-CASPT2 study", *The Journal of Chemical Physics*, vol. 140, no. 9, p. 03B602_1, 2014.

[39] B. Bouvier, J.-P. Dognon, R. Lavery, D. Markovitsi, P. Millié, D. Onidas, and K. Zakrzewska, "Influence of conformational dynamics on the exciton states of DNA oligomers", *The Journal of Physical Chemistry B*, vol. 107, no. 48, pp. 13512–13522, 2003.

[40] D. Markovitsi, D. Onidas, T. Gustavsson, F. Talbot, and E. Lazzarotto, "Collective behavior of Franck-Condon excited states and energy transfer in DNA double helices", *Journal of the American Chemical Society*, vol. 127, no. 49, pp. 17130–17131, 2005.

[41] C. F. Guerra, F. M. Bickelhaupt, J. G. Snijders, and E. J. Baerends, "The nature of the hydrogen bond in DNA base pairs: the role of charge transfer and resonance assistance", *Chemistry-A European Journal*, vol. 5, no. 12, pp. 3581–3594, 1999.

[42] F. Santoro, V. Barone, and R. Improta, "Influence of base stacking on excited-state behavior of polyadenine in water, based on time-dependent density functional calculations", *Proceedings of the National Academy of Sciences*, vol. 104, no. 24, pp. 9931–9936, 2007.

[43] A. W. Lange and J. M. Herbert, "Both intra- and interstrand charge-transfer excited states in aqueous B-DNA are present at energies comparable to, or just above, the $\pi\pi^*$ excitonic bright states", *Journal of the American Chemical Society*, vol. 131, no. 11, pp. 3913–3922, 2009.

[44] A. A. Voityuk, "Effects of dynamic disorder on exciton delocalization and photoinduced charge separation in DNA", *Photochemical & Photobiological Sciences*, vol. 12, no. 8, pp. 1303–1309, 2013.

[45] D. Markovitsi, T. Gustavsson, and F. Talbot, "Excited states and energy transfer among DNA bases in double helices", *Photochemical & Photobiological Sciences*, vol. 6, no. 7, pp. 717–724, 2007.

[46] B. Bouvier, T. Gustavsson, D. Markovitsi, and P. Millié, "Dipolar coupling between electronic transitions of the DNA bases and its relevance to exciton states in double helices", *Chemical Physics*, vol. 275, no. 1, pp. 75–92, 2002.

[47] I. Tinoco Jr, "Hypochromism in polynucleotides 1", *Journal of the American Chemical Society*, vol. 82, no. 18, pp. 4785–4790, 1960.

[48] W. Rhodes, "Hypochromism and other spectral properties of helical polynucleotides", *Journal of the American Chemical Society*, vol. 83, no. 17, pp. 3609–3617, 1961.

[49] I. Vayá, J. Brazard, M. Huix-Rotllant, A. K. Thazhathveetil, F. D. Lewis, T. Gustavsson, I. Burghardt, R. Improta, and D. Markovitsi, "High-energy long-lived mixed Frenkel–charge-transfer excitons: From double stranded (AT) n to natural DNA", *Chemistry-A European Journal*, vol. 22, no. 14, pp. 4904–4914, 2016.

[50] J. J. Nogueira, F. Plasser, and L. Gonzalez, "Electronic delocalization, charge transfer and hypochromism in the UV absorption spectrum of polyadenine unravelled by multiscale computations and quantitative wavefunction analysis", *Chemical Science*, 2017.

[51] P. Marquetand, J. J. Nogueira, S. Mai, F. Plasser, and L. González, "Challenges in simulating light-induced processes in DNA", *Molecules*, vol. 22, no. 1, p. 49, 2016.

[52] A. N. Alexandrova, J. C. Tully, and G. Granucci, "Photochemistry of DNA fragments via semiclassical nonadiabatic dynamics", *The Journal of Physical Chemistry B*, vol. 114, no. 37, pp. 12116–12128, 2010.

[53] R. Improta, "Photophysics and photochemistry of thymine deoxy-dinucleotide in water: A PCM/TD-DFT quantum mechanical study", *The Journal of Physical Chemistry B*, vol. 116, no. 49, pp. 14261–14274, 2012.

[54] F. Plasser, A. J. Aquino, W. L. Hase, and H. Lischka, "UV absorption spectrum of alternating DNA duplexes. Analysis of excitonic and charge transfer interactions", *The Journal of Physical Chemistry A*, vol. 116, no. 46, pp. 11151–11160, 2012.

[55] S. Mai, M. Richter, P. Marquetand, and L. González, "The DNA nucleobase thymine in motion–intersystem crossing simulated with surface hopping", *Chemical Physics*, vol. 482, pp. 9–15, 2017.

[56] H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems", *Angewandte Chemie International Edition*, vol. 48, no. 7, pp. 1198–1229, 2009.

[57] Anonymous Author, "Interaction of UV radiation with DNA,", *Royal Society of Chemistry*, 2013.

[58] F. Plasser and H. Lischka, "Analysis of excitonic and charge transfer interactions from quantum chemical calculations", *Journal of Chemical Theory and Computation*, vol. 8, no. 8, pp. 2777–2789, 2012.

[59] F. Plasser, M. Wormit, and A. Dreuw, "New tools for the systematic analysis and visualization of electronic excitations. I. formalism", *The Journal of Chemical Physics*, vol. 141, no. 2, p. 024106, 2014.

[60] S. A. Adcock and J. A. McCammon, "Molecular dynamics: survey of methods for simulating the activity of proteins", *Chemical Reviews*, vol. 106, no. 5, pp. 1589–1615, 2006.

[61] R. Battiti, "First- and second-order methods for learning: between steepest descent and Newton's method", *Neural Computation*, vol. 4, no. 2, pp. 141–166, 1992.

[62] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, vol. 49. NBS, 1952.

[63] J. M. Redondo, "Turbulence, entropy and dynamics", *UPC Lecture Notes*, 2014.

[64] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters", *The Journal of Chemical Physics*, vol. 76, no. 1, pp. 637–649, 1982.

[65] J. A. McCammon and S. C. Harvey, *Dynamics of proteins and nucleic acids*. Cambridge University Press, 1988.

[66] R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, *et al.*, "Ambertools 16", 2016.

[67] D. C. Rapaport, R. L. Blumberg, S. R. McKay, W. Christian, *et al.*, "The art of molecular dynamics simulation", *Computers in Physics*, vol. 10, no. 5, pp. 456–456, 1996.

[68] O. N. de Souza and R. L. Ornstein, "Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with Particle-mesh Ewald method", *Biophysical Journal*, vol. 72, no. 6, pp. 2395–2397, 1997.

[69] W. F. van Gunsteren and H. J. Berendsen, "Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry", *Angewandte Chemie International Edition*, vol. 29, no. 9, pp. 992–1023, 1990.

[70] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak, "Molecular dynamics with coupling to an external bath", *The Journal of Cemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[71] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme", *Journal of Molecular Biology*, vol. 103, no. 2, pp. 227–249, 1976.

[72] G. Groenhof, "Introduction to QM/MM simulations", *Biomolecular Simulations: Methods and Protocols*, pp. 43–66, 2013.

[73] A. Warshel and S. T. Russell, "Calculations of electrostatic interactions in biological systems and in solutions", *Quarterly Reviews of Biophysics*, vol. 17, no. 3, pp. 283–422, 1984.

[74] T. A. Halgren and W. Damm, "Polarizable force fields", *Current Opinion in Structural Biology*, vol. 11, no. 2, pp. 236–242, 2001.

[75] M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations", *Journal of Computational Chemistry*, vol. 11, no. 6, pp. 700–733, 1990.

[76] W. Koch and M. C. Holthausen, *A chemist's guide to density functional theory*. John Wiley & Sons, 2015.

[77] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects", *Physical Review*, vol. 140, no. 4A, p. A1133, 1965.

[78] W. Kohn, A. D. Becke, and R. G. Parr, "Density functional theory of electronic structure", *The Journal of Physical Chemistry*, vol. 100, no. 31, pp. 12974–12980, 1996.

[79] A. D. Becke, "Becke's three parameter hybrid method using the LYP correlation functional", *J. Chem. Phys*, vol. 98, pp. 5648–5652, 1993.

[80] E. Runge and E. K. Gross, "Density-functional theory for time-dependent systems", *Physical Review Letters*, vol. 52, no. 12, p. 997, 1984.

[81] M. Marques, *Time-dependent density functional theory*, vol. 706. Springer Science & Business Media, 2006.

[82] M. E. Casida and M. Huix-Rotllant, "Progress in time-dependent density-functional theory", *Annual review of physical chemistry*, vol. 63, pp. 287–323, 2012.

[83] A. Dreuw and M. Head-Gordon, "Single-reference *ab initio* methods for the calculation of excited states of large molecules", *Chemical Reviews*, vol. 105, no. 11, pp. 4009–4037, 2005.

[84] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An $N\cdot\log(N)$ method for ewald sums in large systems", *The Journal of chemical physics*, vol. 98, no. 12, pp. 10089–10092, 1993.

[85] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes", *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.

[86] S. Miyamoto and P. A. Kollman, "SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models", *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 952–962, 1992.

[87] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB improving the accuracy of protein side chain and backbone parameters from ff99SB", *Journal of Chemical Theory and Computation*, vol. 11, no. 8, pp. 3696–3713, 2015.

[88] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.

[89] D. A. Case, V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. Cheatham Iii, T. Darden, R. Duke, H. Gohlke, *et al.*, "Amber 14", 2014.

[90] I. S. Ufimtsev and T. J. Martinez, "Quantum chemistry on graphical processing units. 3. analytical energy gradients, geometry optimization, and first principles molecular dynamics", *Journal of Chemical Theory and Computation*, vol. 5, no. 10, pp. 2619–2628, 2009.

[91] A. V. Titov, I. S. Ufimtsev, N. Luehr, and T. J. Martinez, "Generating efficient quantum chemistry codes for novel architectures", *Journal of Chemical Theory and Computation*, vol. 9, no. 1, pp. 213–221, 2012.

[92] C. M. Isborn, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, "Excited-state electronic structure with configuration interaction singles and Tamm–Dancoff time-dependent density functional theory on graphical processing units", *Journal of Chemical Theory and Computation*, vol. 7, no. 6, pp. 1814–1823, 2011.

[93] E. Gross and W. Kohn, "Time-dependent density-functional theory", *Advances in Quantum Chemistry*, vol. 21, pp. 255–291, 1990.

[94] T. Yanai, D. P. Tew, and N. C. Handy, "A new hybrid exchange–correlation functional using the coulomb-attenuating method (CAM-B3LYP)", *Chemical Physics Letters*, vol. 393, no. 1, pp. 51–57, 2004.

[95] A. Schäfer, H. Horn, and R. Ahlrichs, "Fully optimized contracted gaussian basis sets for atoms Li to Kr", *The Journal of Chemical Physics*, vol. 97, no. 4, pp. 2571–2577, 1992.

[96] P. Caruso, M. Causà, P. Cimino, O. Crescenzi, M. D'Amore, R. Improta, M. Pavone, and N. Rega, "Effects of molecular dynamics and solvation on the electronic structure of molecular probes", *Theoretical Chemistry Accounts*, vol. 131, no. 4, p. 1211, 2012.

[97] S. Krawczyk and R. Luchowski, "Electronic excited states of polynucleotides: A study by electroabsorption spectroscopy", *The Journal of Physical Chemistry B*, vol. 111, no. 5, pp. 1213–1221, 2007.

[98] T. Miyata and S. Yomosa, "π-electronic structures of the base pairs of DNA", *Journal of the Physical Society of Japan*, vol. 27, no. 3, pp. 720–727, 1969.

[99] E. Emanuele, D. Markovitsi, P. Millié, and K. Zakrzewska, "UV spectra and excitation delocalization in DNA: influence of the spectral width", *ChemPhysChem*, vol. 6, no. 7, pp. 1387–1392, 2005.

[100] F. Santoro, V. Barone, and R. Improta, "Absorption spectrum of A–T DNA unraveled by quantum mechanical calculations in solution on the $(dA)_2 \cdot (dT)_2$ tetramer", *ChemPhysChem*, vol. 9, no. 17, pp. 2531–2537, 2008.

[101] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, "Single-molecule studies of DNA mechanics", *Current Opinion in Structural Biology*, vol. 10, no. 3, pp. 279–285, 2000.

[102] M. Petersheim and D. H. Turner, "Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp", *Biochemistry*, vol. 22, no. 2, pp. 256–263, 1983.

[103] C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández, and B. Kohler, "DNA excited-state dynamics: from single bases to the double helix", *Annual Review of Physical Chemistry*, vol. 60, 2009.

[104] M. Preuss, W. G. Schmidt, K. Seino, J. Furthmüller, and F. Bechstedt, "Ground- and excited-state properties of DNA base molecules from plane-wave calculations using ultrasoft pseudopotentials", *Journal of Computational Chemistry*, vol. 25, no. 1, pp. 112–122, 2004.

[105] D. Markovitsi, "UV-induced DNA damage: The role of electronic excited states", *Photochemistry and Photobiology*, vol. 92, no. 1, pp. 45–51, 2016.

[106] W. J. Schreier, T. E. Schrader, F. O. Koller, P. Gilch, C. E. Crespo-Hernández, V. N. Swaminathan, T. Carell, W. Zinth, and B. Kohler, "Thymine dimerization in DNA is an ultrafast photoreaction", *Science*, vol. 315, no. 5812, pp. 625–629, 2007.

[107] D. Roca-Sanjuán, M. Rubio, M. Merchán, and L. Serrano-Andrés, "*Ab initio* determination of the ionization potentials of DNA and RNA nucleobases", *The Journal of Chemical Physics*, vol. 125, no. 8, p. 084302, 2006.

[108] J. C. Tully, "Perspective: Nonadiabatic dynamics theory", *The Journal of Chemical Physics*, vol. 137, no. 22, p. 22A301, 2012.

[109] S. Mai, P. Marquetand, M. Richter, J. González-Vázquez, and L. González, "Singlet and triplet excited-state dynamics study of the keto and enol tautomers of cytosine", *ChemPhysChem*, vol. 14, no. 13, pp. 2920–2931, 2013.

[110] C. E. Crespo-Hernández, B. Cohen, P. M. Hare, and B. Kohler, "Ultrafast excited-state dynamics in nucleic acids", *Chemical Reviews*, vol. 104, no. 4, pp. 1977–2020, 2004.

[111] F.-A. Miannay, Á. Bányász, T. Gustavsson, and D. Markovitsi, "Ultrafast excited-state deactivation and energy transfer in guanine- cytosine DNA double helices", *Journal of the American Chemical Society*, vol. 129, no. 47, pp. 14574–14575, 2007.

[112] M. Rooman and R. Wintjens, "Sequence and conformation effects on ionization potential and charge distribution of homo-nucleobase stacks using M06-2X hybrid density functional theory calculations", *Journal of Biomolecular Structure and Dynamics*, vol. 32, no. 4, pp. 532–545, 2014.

[113] K. Kawai and T. Majima, "Hole transfer kinetics of DNA", *Accounts of Chemical Research*, vol. 46, no. 11, pp. 2616–2625, 2013.

[114] B. Giese, "Long-distance electron transfer through DNA", *Annual Review of Biochemistry*, vol. 71, no. 1, pp. 51–70, 2002.

[115] R. N. Barnett, C. L. Cleveland, A. Joy, U. Landman, and G. B. Schuster, "Charge migration in DNA: ion-gated transport", *Science*, vol. 294, no. 5542, pp. 567–571, 2001.