Development and Application of Efficient Methods for the Simulation of Charge Transport in Complex Biomolecules

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

Fakultät für Chemie und Biowissenschaften

Karlsruher Institut für Technologie (KIT) – Universitätsbereich

genehmigte

DISSERTATION

von

Dipl. Chem. Paul Benjamin Woiczikowski

aus

Frankfurt an der Oder

Die vorliegende Arbeit wurde von November 2007 bis September 2009 am Institut für Physikalische und Theoretische Chemie der Technischen Universität Carolo-Wilhelmina zu Braunschweig, sowie von Oktober 2009 bis Mai 2011 am Institut für Physikalische Chemie am Karlsruher Institut für Technologie unter Anleitung von Herrn Prof. Dr. Marcus Elstner angefertigt.

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet, sowie die Satzung der Universität Karlsruhe (TH) zur Sicherung guter wissenschaftlicher Praxis beachtet habe.

Paul Benjamin Woiczikowski
Karlsruhe, 31.05.2011

Contents

List of Abbreviations

A	Adenine
Å	Ångström $= 10^{-10}$ meter
a.u.	Atomic units
BO	Born-Oppenheimer
C	Cytosine
CAS-PT2	Complete active space - second order perturbation theory
CG	Coarse-grained
CPD	Cyclobutane pyrimidine dimers
CT	Charge transport
DBA	Donor-bridge-accepter
DFT	Density functional theory
DFTB	Density functional tight-binding
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
DTI	Discrete thermodynamic integration
EFM	Electrostatic force microscopy
EID	Electron-ion dynamics
ESP	Electrostatic potential
Exp.	Experiment/experimental
ET	Electron transfer
FAD	Flavin adenine dinucleotide
FEP	Free energy perturbation
FGTI	Fast-growth thermodynamic integration
G	Guanine
G4	Square-planar unit composed of four guanines
GGA	Generalized gradient approximation
HF	Hartree-Fock
HOMO	Highest occupied molecular orbital
HT	Hole transfer
IP	Ionization potential
KS	Kohn-Sham

KTA	Koopmans' theorem approximation
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation
LYP	Correlation functional from Lee, Yang and Parr
M-DNA	Metalized DNA
MD	Molecular dynamics
MM	Molecular mechanics
MTHF	5,10-Methylenetetrahydrofolate
NDDO	Neglect of diatomic differential overlap
NEGF	Non-equilibrium Green's function
PBC	Periodic boundary conditions
PBE	Density functional from Perdew, Burke and Ernzerhof
PES	Potential energy surface
PL	DNA photolyase, i.e. *E. coli* photolyase
QM	Quantum mechanics
RDF	Radial distribution function
RESP	Restrained electrostatic potential
RMSF	Root mean square fluctuation
RMSD	Root mean square deviation
RNA	Ribonucleic acid
SCC	Self consistent charge
SCF	Self consistent field
SIC	Self interaction correction
T	Thymine
TD	Time-dependent
TI	Thermodynamic integration
TIH	Thermal induced hopping
Trp	Tryptophan side chain
WCP	Watson-Crick pair
xDNA	Size-expanded DNA (insertion of benzene into the nucleobase)

Zusammenfassung

Gegenstand der vorliegenden Dissertation, ist die theoretische Beschreibung des Ladungstransports in komplexen, biomolekularen Systemen mittels Kombination von molekular- und quantenmechanischen Methoden sowie diversen Transportmodellen. Zum einen wird der Transport von Elektronenlöchern durch DNA untersucht, zum anderen der Transfer eines Elektronenloches entlang dreier evolutionär-konservierten Tryptophan-Seitenketten als Teil des Photoaktivierungsprozess im Protein *E. coli* Photolyase.

Für die Untersuchung des Ladungstransports in derartig komplexen Systemen wurde im Rahmen dieser Arbeit eine Methodik zur effizienten Berechnung von Ladungstransfer (LT) Parametern mitentwickelt und intensiv an verschiedenen DNA- und Proteinsystemen evaluiert. Die Methodik basiert auf der genäherten Dichtefunktionalmethode DFTB in Kombination mit einem Fragment-Orbital Ansatz, im folgenden als DFTB-FO Methode bezeichnet. Im Kapitel 3 werden LT-Parameter mit der DFTB-FO Methode für den Elektronenlochtransfer in idealisierten, statischen DNA Strukturen berechnet. Der Vergleich der Ergebnisse mit *ab initio* Methoden ergab eine sehr gute Übereinstimmung, wobei die Rechenzeit der DFTB-FO Methode um bis zu drei Größenordnungen geringer ist. Daher können LT-Parameter sehr effizient und ausreichend genau mit der DFTB-FO Methode berechnet werden. Dennoch sollte für Anwendungen auf neue Systeme stets eine vorherige Validierung an genaueren Methoden oder experimentellen Daten erfolgen.

Der Kern dieser Arbeit besteht darin, den Einfluss von struktureller Dynamik und der Umgebung (Lösungsmittel, Gegenionen, DNA Rückgrat oder Proteinumgebung) in die Simulation des Ladungstransports mit einzubeziehen und zu studieren. Die hervorragende Effizienz der DFTB-FO Methode erlaubt es die LT-Parameter entlang ausgedehnter, klassischer Moleklardynamik (MD) Simulationen in explizitem Lösungsmittel (Wasser, Gegenionen) zu berechnen, wobei auch die elektrostatische Polarisation der Umgebung auf die LT-Parameter berücksichtigt wird. Die DNA Simulationen in Kapitel 4 zeigen wie stark die LT-Parameter fluktuieren, und welchen Einfluss das auf den kohärenten Ladungstransport durch DNA hat. Demnach sind nur wenige DNA Konformationen leitfähig. Interessanterweise führt die dynamische Unordnung in homogenen Sequenzen (keine statische Energiebarriere) zur Unterdrückung des Transports, während es sich bei heterogenen Sequenzen (statische Energiebarrieren G–A) umgekehrt verhält. Der Transport wird durch die Dynamik sogar verstärkt, indem überhaupt erst leitfähige Konformationen entstehen. Demzufolge sollten theoretische LT-Studien nicht an statis-

chen, idealisierten Strukturen durchgeführt werden, vielmehr müssen Konformationsän-
derungen und der Einfluss der Umgebung in realistischer Weise berücksichtigt werden.
Zudem konnten die Simulationen zeigen, dass Fluktuationen benachbarter DNA Basen
nicht gänzlich ungeordnet sind. Diese beachtlichen Korrelationen erhöhen effektiv den
Transport, wurden jedoch in älteren, statistischen Modellen nicht berücksichtigt.

Im Kapitel 5 wurde die Methodik auf die viersträngige DNA Spezies, G4-DNA,
angewendet. Experimentelle und theoretische Arbeiten deuten darauf hin, dass G4-
DNA ein größeres Potential in der molekularen Elektronik hat, da es im Vergleich zu
konventioneller, doppelsträngiger (ds) DNA wesentlich starrer ist und somit eventuell
verbesserte Transporteigenschaften hat. In dieser Arbeit wurde gezeigt, dass G4-DNA
substantiell besser kohärent leitet als dsDNA, was aber nicht an der höheren Rigidität
von G4-DNA liegt, sondern vielmehr an den vielen möglichen LT-Wegen entlang der
Quadruplex. Aufgrund der beachtlichen Kopplungen zwischen den einzelnen Strängen
werden so sehr viele leitfähige Konformationen ermöglicht.

Die Kombination von MD Simulation und Transportrechnungen erlaubt demnach
eine neue, realistischere Beschreibung des kohärenten Ladungstransports unter Ein-
beziehung dynamischer und umgebungsbedingter Effekte. Die Methodik, d.h. Mitteln
des Transports für die vielen Konformationen, setzt allerdings die Annahme voraus, dass
der Ladungstransport schneller ist im Vergleich zur molekularen Bewegung.

Ein anderer, eleganterer Weg dynamischen Transport zu beschreiben, wird im zweiten
Teil dieser Arbeit (Kapitel 7 und 8) am Ladungstransfer entlang dreier Tryptophane in
der Photolyase realisiert. Hier wird die Zeitentwicklung des sich bewegenden Elektro-
nenloches direkt verfolgt, ohne dabei einen bestimmten Transportmechanismus *a priori*
anzunehmen. Dieses nicht-adiabatische Vorgehen ist insbesondere dann geeignet wenn
Elektronen- und Molekularbewegung auf der gleichen Zeitskala ablaufen. Zudem ist das
quantenmechanische (Elektronenloch) direkt an das molekularmechanische (Protein und
Lösungsmittel) System gekoppelt. Diese sogenannte Elektronen-Ionen Dynamik (EID)
ermöglicht neue, mikroskopische Einsichten in den LT-Prozess in Photolyase, welche
mit klassischen LT-Theorien nicht zugänglich sind. Zum einen kann der Grad der De-
lokalisierung des Elektronenloches beobachtet werden, zum anderen werden auch nicht-
Gleichgewichtsprozesse beschrieben. So zeigen die EID Simulationen, dass der LT vom
zweiten zum dritten Tryptophan deshalb so schnell passieren kann, weil die Protein-
und Wasserumgebung nicht genügend Zeit hat, die Ladung auf dem mittleren Trypto-
phan zu stabilisieren. Die errechneten Transferraten stimmen sehr gut mit den experi-
mentellen Daten überein. Auf der anderen Seite konnte klar gezeigt werden, dass klas-
sische Gleichgewichts-LT-Theorien gerade den zweiten Transferschritt nicht beschreiben
kann. Die Analyse der EID Simulationen stellt das umgebende Wasser als die Triebkraft
des Ladungstransfers in Photolyase heraus, welches das Elektronenloch aus dem Inneren
des Proteins „herauszieht". Abschließend lässt sich sagen, dass die EID Methode sich her-
vorragend für die direkte Simulation komplexer Ladungstransferprozesse in biologischen
Systemen eignet.

Summary

This thesis deals with the theoretical description of charge transport in complex, biomolecular systems by means of combined molecular and quantum mechanics methods as well as various transport models. In the first part of this work, the transport of electron holes through DNA is investigated, while the transfer of an electron hole along an evolutionary-conserved chain of three tryptophan residues, as part of the photoactivation process in the protein *E. coli* photolyase, is subject of the second part.

Primarily, a method was developed and tested intensively within the frame of this work, that allows for the efficient computation of charge transfer (CT) parameter in complex systems such as DNA and proteins. The method is based on the approximate density functional method, DFTB, combined with a fragment orbital (FO) approach, in the following referred to as DFTB-FO method. In chapter 3, the DFTB-FO method is used to compute CT parameter for the electron hole transfer (HT) in various idealized, static DNA structures. As a result, a good agreement with higher-level *ab initio* methods is achieved, thereby reducing computational costs immensely by about three orders of magnitude. Therefore, CT parameter can be computed efficiently and sufficiently accurate with the DFTB-FO method. Nevertheless, prior case-specific testing is needed for applications to new systems.

The essence of this work is to include and to study the effect of structural dynamics and the environment (solvent, counterions, DNA backbone or protein remainder) in the simulation of charge transport. The excellent performance of the DFTB-FO method allows for the computation of CT parameter along extended classical molecular dynamics (MD) simulations in explicit solvent, thus also considering the electrostatic polarization on the CT parameter by the environment. DNA simulations presented in chapter 4 exhibit the strong fluctuations of the CT parameter, and the associated influence on the coherent charge transport through DNA. Accordingly, only few DNA conformations are conductive. Interestingly, the dynamical disorder suppresses the transport in homogeneous sequences (no static energy barrier), while the situations is reversed for heterogeneous sequences (static energy barrier G–A). In the latter case, the transport is even enhanced, as conducting conformations can arise only due to the dynamics. As a consequence, theoretical studies on CT should not be based on static, idealized structures, rather conformational changes and the influence of the environment must be accounted for in a realistic way. Furthermore, the simulations revealed that fluctuations

of adjacent nucleobases are not completely disordered. These considerable correlations are shown to effectively enhance the transport, however, they were not considered in previous, statistical models.

In chapter 5, the methodology was applied to the four-stranded DNA species, G4-DNA. Experimental and theoretical studies indicated a larger potential of G4-DNA in the field of molecular electronics, as it was shown to be significantly more rigid than conventional double-stranded (ds) DNA, thus may have improved electronic properties. In this work, G4-DNA was shown to exhibit a substantially higher (coherent) conductivity compared to dsDNA. This, however, is not related to the higher rigidity of G4 as expected beforehand, rather many possible CT pathways can arise along the quadruplex, due to non-vanishing interstrand couplings between the individual G4 strands. Therefore, the number of conductive conformations is increased greatly in G4-DNA.

To sum up the first part of this work, the combination of MD simulation and transport calculations allows for a new and more realistic description of coherent charge transport including dynamical and environmental effects. Notwithstanding, the methodology, that is averaging of transport properties over the vast number of conformations, assumes the charge transport to be faster on the time scale of molecular motion.

Another, more sophisticated way to describe dynamic transport is realized in the second part of this work (chapter 7 and 8) in the context of charge transfer along three tryptophan sidechains in photolyase. Here, the time evolution of the moving electron hole is followed directly without presupposing any transport mechanism *a priori*. This non-adiabatic proceeding is applicable in particular if electron and molecular motion occur on comparable time scales. Moreover, the quantum system (electron hole) is coupled directly to the molecular system (protein and solvent). This so-called electron-ion dynamics (EID) enables new, microscopic insights into the CT process in photolyase, which cannot be deduced from classical CT theories. For instance, the degree of charge delocalization can be observed in real time, and non-equilibrium processes are described as well. Accordingly, the EID simulations revealed that CT from the second to the third tryptophan can occur so fast only because the protein and solvent environment does not have enough time to stabilize the charge on the central tryptophan. The computed CT rates agree very well with the experimental data. On the other hand, it was demonstrated clearly that classical CT theory is not able to describe this particular second CT step. The analysis of the various EID simulations indicates that the surrounding solvent is the driving force of CT in photolyase, that is the solvent "withdraws" the electron hole from the inside of the protein. To conclude, the EID method is suited excellently for the direct simulation of complex CT processes in biological systems.

1

Introduction

Various processes in all fields of science involve chemical reactions that are accompanied by the transfer of electrons or electron holes. Accordingly, electron transfer (ET) reactions are the key steps in biological processes such as photosynthesis and respiration. ET reactions were observed to occur in both most relevant biomolecules: DNA and proteins. Nevertheless, chemical charge transport (CT) is vital in material science as well, in particular in the fast-growing field of molecular electronics, in which the conventional silicon-based materials are supposed to be replaced more and more by organic compounds. Intensive experimental and theoretical research has been focused on an in-depth atomistic understanding of CT in complex systems, especially in proteins and DNA. The latter field, i.e. the CT in DNA, is reviewed in the following, subsequently, an introduction is given in the photoactivation of the DNA repair protein *E. coli* photolyase. Eventually, a short survey over the development of theoretical approaches, regarding the simulation and description of CT, is given. A brief outline of this thesis can be found at the end of the chapter.

1.1   Transfer and Transport of Charges in DNA

After Watson and Crick discovered the double helical structure of DNA in 1953 [1], the research focus was centered predominantly on the fundamental biological relevance of DNA, as carrier of the genetic code and as such, the blueprint of life. The specific binding between the individual DNA strands as well as the ability of self-assembling and recognition are unique structural properties for a molecular system of this dimension [2]. Therefore, DNA has attracted considerable interest in the fields of molecular biology and molecular electronics [3].

On the one hand, CT is expected to play a key role in the self-repair of DNA damage under natural conditions, which is caused, e.g. by oxidative stress [4, 5]. On the other hand, technical advances such as the fully-automatic synthesis of oligonucleotides and modern DNA architecture, i.e. DNA origami [6–8] allow for the construction of any imaginable two or even three dimensional structure, therefore offering a huge potential for applications as nano-scale electronic devices either as templates [9–11] or as active wiring

(a) Extended π orbital in DNA   (b) Hole transfer in DNA
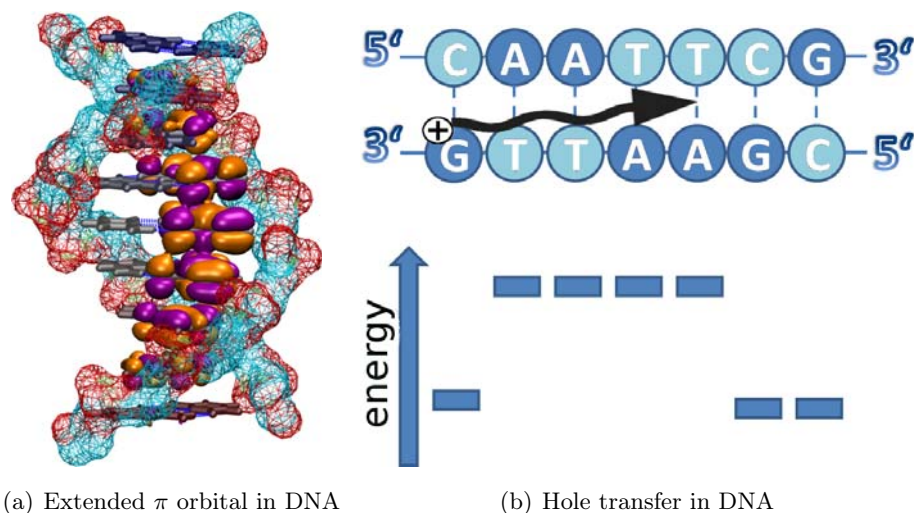
Figure 1.1: a) Illustration of π–π interactions between nucleobases along the DNA double helix. DNA backbone is indicated as blue–red wireframe, MOs are highlighted in orange/purple and nucleobases are shown in gray. b) Schematic representation of hole transfer (HT) in DNA. Only the purine bases A and G (dark blue) are involved in the HT process, while the pyrimidines C and T play a major role in the transfer of excess electrons. The corresponding static energy levels of G and A bases are depicted qualitatively as well.

systems. However, the latter would presuppose well-controlled electronic properties and an in-depth understanding of CT in DNA.

As early as in 1960, the quantum mechanical calculations of Ladik [12] indicated the existence of non-negligible π-electron interaction between the adjacent bases of DNA (Fig. 1.1(a)), which could, therefore, enable electron mobility along the DNA. Two years later, Eley and Spivey were inspired by the semi-conductivity of organic materials and proposed, also, that π-π interactions between the adjacent nucleobases may lead to electronic conduction along the double helical structure [13].

1.1.1 Chemical Experiments

More than 30 years later, this hypothesis was confirmed in the groundbreaking ET experiments by the groups of Barton [14–21], Giese [22–24], Lewis [25–27], Schuster [28–31] and Michel-Beyerle [32–34]. These and similar experiments reported CT to occur over large distances. In fact, a very recent experimental study [35] reported DNA CT over 100 base pairs, which corresponds approximately to a CT distance of 34 nm. In most of these rather chemically motivated experiments, CT is initiated either photoinduced or via chemical reactions [36]. A well-controlled and selective creation of an excess charge (electron or electron hole) can be realized by chemical modification or even replacement of the nucleobases themselves, and also by insertion of intercalators into the DNA duplex. The transferring charge may then be trapped, e.g. due to chemical reaction with water.

In oxidative hole transfer (HT) experiments, guanine (G) is transformed most easily into its radical state $G^{+\bullet}$, as it has the lowest oxidation potential of all the four nucle-

obases. The second lowest oxidation potential is that of adenine (A), which is about 0.4 eV higher in energy. The participation of cytosine (C) and thymine (T) in the HT process is very unlikely, since their cationic radical states are even 1 eV higher in energy. Therefore, the HT process in DNA proceeds over the bases G and A, as illustrated schematically in Fig. 1.1(b). Moreover, a positive charge is even more stabilized on GG and GGG DNA domains, which therefore, function as hole traps within the DNA. The results obtained from HT experiments revealed two mechanistic regimes of CT in DNA: i) strongly exponential distance-dependent CT via single-step superexchange for short distances between two G units separated by not more than three A bases (Sec. 2.1.1), and ii) very weak distance-dependent multi-step hopping for larger distances, i.e. more than three intervening A bases (Sec. 2.1.2). The latter mechanism is also referred to as "G-hopping". Furthermore, the "A-hopping" mechanism was proposed for extended A bridges, that is the hole charge is activated thermally from the G to the A base and may then proceed along the A tract via single-step hopping events [22].

Various theoretical studies could describe both mechanisms, and also the crossover between them [32, 37–42]. Moreover, characteristic CT parameter could be extracted from the applied models, one of them being the electronic coupling between the π-orbitals of stacked nucleobases, as foreseen by Ladik, Eley and Spivey. Therefore, the studies on HT in DNA are rather well advanced at least from the chemical perspective [22], although more recent studies indicate that the clear distinction of both tunneling and hopping regime in dependence on the CT distance may not always be justified. For one thing, hopping transport may be at least partially active even in short DNA fragments, as found in Ref. [43], while, on the other hand, single-step coherent HT was observed in donor/acceptor systems separated by 7–8 intervening A bases [44]. The latter finding was related to a large delocalization of the hole charge over the full turn of the DNA helix. These recent experimental observations emphasize the relevance of sequence-dependent conformational dynamics in the mechanism of CT in DNA [44], that has to be considered for theoretical studies as well. Therefore, some vital issues have not yet been explored adequately, for instance, the extent of charge delocalization, the "true" sequence dependence, and the actual microscopic mechanism of CT with regard to the influence of conformational dynamics and gating effects by solvent and/or counterions. Nevertheless, it can be said that there has been an immense progress in the chemical study of CT in DNA both experimentally and theoretically.

1.1.2 DNA as a Nanowire?

On the contrary, this is not true with regard to the intrinsic conductivity of DNA when coupled to electric leads. In the past 20 years, various DNA conductivity experiments have been conducted leading to very contradicting results. Accordingly, DNA was suggested to be an insulator at room temperature [10, 45–47], a semiconductor with a wide band gap [48, 49], an ohmic conductor [50–52], a metallic conductor [53–55], and eventually even a super conductor at temperatures below 1 K [56]. However, this broad
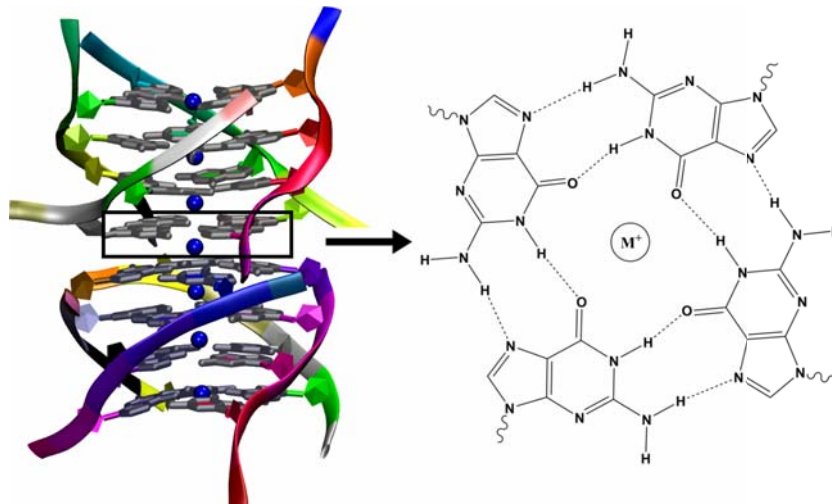
Figure 1.2: Two stacked G4 quadruplexes each composed of four parallel strands (left). Backbones are multicolored, G residues in gray, metal ions within the quadruplex channel are indicated as blue spheres. The chemical structure of a single G4 unit is shown as well (right).

variability is related to the poor control of experimental conditions and also to different setups in general. These may differ with regard to environmental effects, thermal fluctuations and, predominantly, the quality of the electric leads and their coupling to the DNA [2, 48, 57, 58]. Moreover, some experiments revealed a strong dependence of the conduction on the specific DNA sequence [54, 59], which may implicate a large potential in information technology. On the other hand, other experiments [55, 60, 61] reported rather similar electrical currents in the nA regime despite differences in the specific DNA sequence and experimental setups. Therefore, a realistic theoretical description of DNA conduction is impeded by these various contradicting results. There are various attempts to understand at least parts of the observed I–U characteristic from the theoretical point of view [62–67].

1.2 G-Quadruplex DNA: A Better Candidate for Nano-Electronic Applications?

Nucleic acids are known to adopt various conformations, which includes not necessarily solely the double-stranded forms. For instance, C rich DNA strands can form the so-called i-motif structure, a four-stranded quadruplex which is built of two parallel hemi-protonated duplexes intercalated into each other by a head-to-tail orientation [68]. Another quadruplex DNA form is formed by either one, two or four G-rich DNA strands in a parallel or anti-parallel orientation [69]. The latter DNA form is referred to as G4-quadruplex or G4-DNA (G4). A typical G4 structure is illustrated in Fig. 1.2.

G4 occurs naturally in the telomeric regions of DNA and is known to be involved in various biochemical processes such as the inhibition of the enzymes telomerase and HIV integrase [70]. Furthermore, G4 may be vital for the design of anti-cancer drugs, as it

is cytotoxic towards tumor cells [71]. On the other hand, diseases such as Bloom's and Werner's syndromes involve the interaction of G4 with various cell proteins [72].

The structure of G4 is thermally more stable than that of conventional dsDNA [73]. The first crystal structures were resolved in x-ray experiments by Gilbert and Sen [74–76], which proposed the formation of G4 in the presence of monovalent ions. Accordingly, the building block of G4 stacks is a square-planar unit of four guanines, also referred to as G-tetrad or G-quartet, held together by Hoogsteen [77] hydrogen bonds. Moreover, metal ions, mostly alkali ions can be found in the cavity of the quadruplex, coordinated by the carbonyl oxygen atoms of the G bases, as illustrated in Fig. 1.2. However, there is a high structural diversity of G4 attributed to the number of strands and experimental conditions, e.g. different metal coordination. A detailed review on this issue is given in Refs. [69, 78].

In the context of molecular electronics, it was argued that G4 might offer improved electronic properties compared to dsDNA related to its increased structural stability and the increased π–π interactions of two adjacent G4 tetrads. Indeed, electrostatic force microscopy (EFM) measurements exhibited a considerable polarizability of G4-DNA when attached to a solid gold surface, while dsDNA molecules remained electrically silent [79]. Moreover, the progress in chemical synthesis allows for the preparation of up to 300 nm long G4-wires which were observed to be sufficiently stable even in absence of metal ions [80, 81]. In a recent conductivity experiment [82], G4 was shown to maintain a conductive conformation when exerted to strain, while such a behavior is not expected for dsDNA. These experimental findings suggest G4 to be a very promising candidate in the field of nano-electronic devices.

First theoretical studies were focused primarily on structural properties mainly by performing classical MD simulation [83–85]. The main findings of these studies are i) a higher structural rigidity of G4 over dsDNA, and ii) the essential stabilization of short G4 molecules due to the presence of monovalent alkali ions within the quadruplex channel. Nevertheless, the simulations performed in Ref. [85] indicated the stability of long G4 structures with more than 24 G-tetrads in absence of ions, which is in agreement with the experimental findings [80, 81]. Also, the electronic structure of G4 was studied theoretically in-depth by Di Felice and coworkers [86–91], revealing an increased delocalization of states compared to dsDNA. Furthermore, the density of states is extended in G4, also due to the contribution of coordinated metal ions. First transport calculations were performed by Guo and coworkers [92, 93] relying on a coherent transport model. According to these studies, G4 offers considerably larger delocalization lengths at the band centers as compared to dsDNA. Moreover, the delocalization length in G4 could be even increased by inducing statistical disorder trough the backbone. However, these studies assumed static idealized structures, while in chapter 5 of this work, the coherent transport through G4-DNA is studied including dynamical and environmental effects.
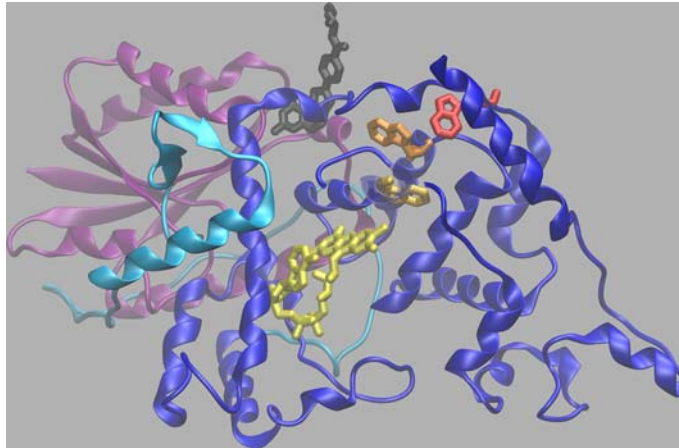
Figure 1.3: Cartoon representation of *E. coli* DNA photolyase from the crystal structure by Park et al. [99]. The FAD cofactor (yellow), MTHF antenna molecule (black) and the three conserved Trp residues 382, 359 and 306 (gold, orange, red) participating in the photoactivation process are displayed. The two-domain structure of the enzyme, with an N-terminal α/β-domain (purple) and an FAD binding α-helical domain (blue) connected by a long inter-domain loop (cyan) is indicated.

1.3   Photoactivation in *E. coli* Photolyase

The photolyases belong to the protein class of photo-reactivating DNA repair enzymes. Moreover, together with the cryptochromes, they are part of the super-family of blue-light driven flavoproteins. Photolyases can bind specifically to DNA domains which contain UV-induced damages, most frequently cyclobutane pyrimidine dimers (CPD). If not repaired, these defects may suppress DNA processing, thus lead to mutation and cell death [94]. The repair process is realized via cyclic ET between the protein's catalytically active cofactor $FADH^{-*}$ in its excited state and the CPD lesion of the DNA [95–97]. First, an electron is transferred from $FADH^{-*}$ to CPD, leading to a splitting into the individual nucleobases of the latter and the oxidized $FADH^{\bullet}$. Subsequently, reverse electron transfer from CPD reforms the protein cofactor to its doubly reduced state $FADH^{-}$, and the DNA substrate is released [98].

An additional second but non-essential cofactor is present in photolyases which functions as a light-harvesting chromophor increasing the repair efficiency due to energy transfer to FAD under limiting light conditions. A prominent representative of photolyase proteins is *E. coli* photolyase (PL). It contains methenyltetrahydrofolate (MTHF) as the second cofactor. A cartoon representation of PL is given in Fig. 1.3.

As early as in 1949, Kelner [100] discovered the fundamental biological function of PL using ground-breaking *in vivo* experiments [101]. Nevertheless, there has been a revival of interest in both homologous protein classes photolyases and cryptochromes during the past decade [102–106]. Photolyases are present predominantly in various microorganisms, whereas cryptochromes can be found in almost all living organism from plants, bacteria to animals, also humans [107]. Many experimental studies were devoted

mainly to the actual repair mechanism, a comprehensive review on this issue is given in Ref. [94].

However, these studies revealed the presence of a second biologically relevant mechanism. Accordingly, an alternative "recycling" of the cofactor FAD is realized within the isolated protein, that is an electron is transferred to the semi-reduced radical FAD$^\bullet$, thus re-generating the fully-reduced, catalytically active FADH$^-$ [104]. This mechanism of action is referred to as photoactivation, as low energy (red light) is absorbed by FAD$^\bullet$ to initiate an intra-protein ET over an evolutionary conserved chain of three tryptophan (Trp) residues (see also Fig. 1.3). This Trp chain is present in other photolyases and cryptochromes as well, thus the process of photoactivation was suggested to be of physiological relevance [108, 109].

In part II of this work, the photoactivation in PL, more precisely the CT within the triple Trp chain, is studied theoretically using two conceptually different strategies: i) classical CT theory with CT parameters obtained from combined classical MD simulations and quantum chemical calculations (chapter 7), and ii) direct and non-adiabatic simulation of CT, that is both quantum and classical systems are propagated simultaneously, directly coupled to each other (chapter 8). The major aim of this part of the work is to emphasize the need of the latter, advanced method (ii), to describe the complex multi-step CT as part of the photoactivation in PL. The time and distance dimensions related to this CT process are both in the nano-scale regime, as found experimentally [110–112]. Moreover, the CT occurs in a very heterogeneous environment, as the FAD cofactor and Trp382 are located inside the protein, while the last residue of the Trp chain (Trp306) is located in the exterior of the protein close to the solvent.

Therefore, a detailed theoretical description of photoactivation in PL is challenging, although various classical and quantum mechanical studies were conducted on PL [113–117], a review is given in Ref. [118]. The combined use of classical MD simulations and quantum chemistry was applied recently to both phenomena, the repair process [119] and the photoactivation [120]. However, the latter work used strategy (i), exhibiting rate constants which are two orders of magnitude too small as compared with recent experimental data [104]. Furthermore, a more realistic theoretical description of CT in PL could address various interesting issues such as the origin of the driving force, the effect of both solvent and protein dynamics, the degree of charge delocalization, the microscopic mechanism of individual CT processes and many more.

One major issue is related to the actual transport mechanism. As for hole transfer in DNA, superexchange and hopping were proposed, yet the very fast CT times of less than 30 ps in PL can be rationalized only with hopping. On the other hand, there is hardly evidence of transient oxidation at Trp359 and Trp382 [104]. The photoactivation in PL represents an excellent example of fast long-range CT in a complex environment, in which the time scales of CT and protein relaxation may overlap. Therefore, the theoretical treatment of such processes has to go beyond classical CT theory by providing a description, which includes dynamical and solvent effects adequately.

(a) Biomolecular CT   (b) Molecular junction: DNA
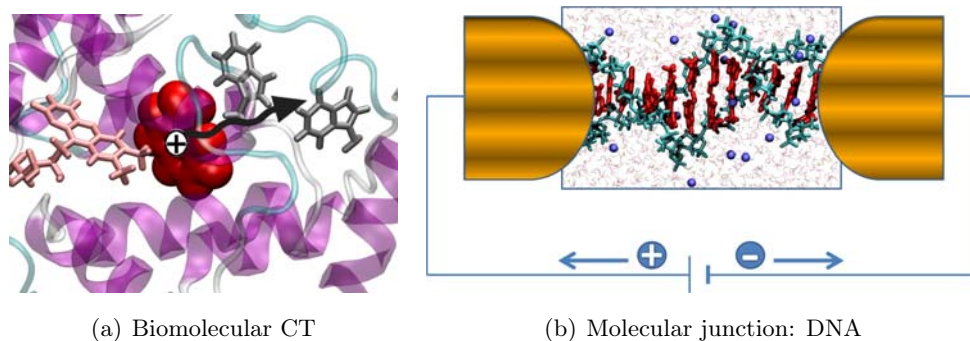
Figure 1.4: Biomolecular CT vs electronic conduction: a) The HT process along the triple Trp chain as part of the photoactivation process in *E. coli* photolyase is depicted (Sec. 1.3). The hole charge is located on the first Trp382, highlighted in red, while the other Trp sidechains are indicated in gray. The flavin part of the FADH cofactor is shown in rose. b) A DNA double helix is coupled to electric leads (gold), as part of an electric circle. DNA backbone and nucleobases are depicted in light blue and red, respectively. Solvent within the molecular box is indicated as thin triangular red lines, while counterions are illustrated as dark blue spheres.

1.4   Towards a Realistic Theoretical Description of CT Phenomena

To begin with, there are two major communities that deal with the theoretical study of complex CT phenomena. On the one hand, there is the (bio)chemical community which is engaged mainly in intra or intermolecular CT occurring within the respective species. An example of biomolecular CT is given in Fig. 1.4(a). The major aim in this research field is to describe and understand biochemical-relevant CT processes and to determine chemical kinetics and corresponding energetic parameters in proteins and other biomolecules [121–125]. For instance, an enormous amount of studies has been devoted to CT processes involved in the photosynthetic reaction cycle [122, 126–138]. The majority of these studies base more or less on the classical theory of ET (Sec. 2.2) developed by the noble laureate Rudolph A. Marcus more than 50 years ago, and its extensions, implemented by many other researchers in the last decades.

On the other hand, there is the rather new field of molecular electronics, in which the community deals with the more complex situation of charge conduction through single molecules, i.e. DNA attached to electric (mostly metallic) leads, as illustrated in Fig. 1.4(b). Such a setup is also referred to as molecular transport junction. Driven by the rapid development of electronic technology to miniaturize electronic building blocks onto the molecular scale, both experimental and theoretical research has been required to explore new physical phenomena, and to understand conductivity on a microscopic level over the past two decades [139, 140]. The coherent conductance of a molecular transport junction can be described phenomenologically by Landauer's formalism (Sec. 2.3), which has a similar meaning in this research field, as Marcus' theory has in the ET field. Indeed, it could be shown that both conductance and ET rate are related directly, as
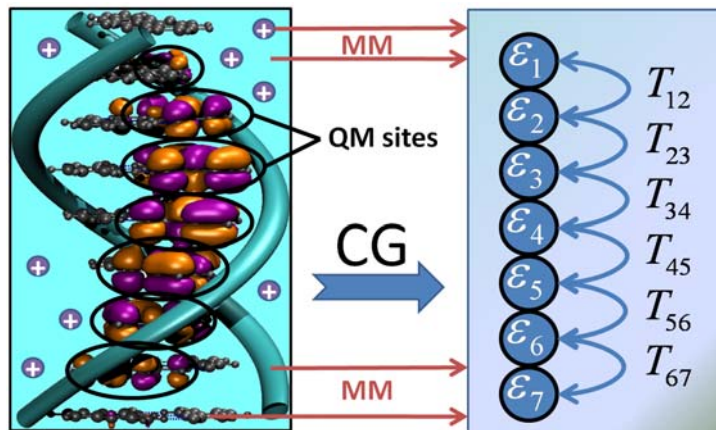
Figure 1.5: Coarse-grained (CG) description of DNA CT: The atomistic structure of the QM part is mapped onto a Hückel Hamiltonian, in which each nucleobase (site) is characterized by an effective site energy ϵ_i. In case of HT, this energy corresponds to the respective highest occupied molecular orbital (HOMO), while the lowest unoccupied molecular orbital (LUMO) is considered for ET. Nearest-neighbor π–π interactions are described by the electronic couplings T_{ij}. The MM part is composed of DNA backbone, "CT-inactive" nucleobases, solvent and counterions, and is considered explicitly in the CT calculation as MM point charges polarizing the QM sites.

both processes are determined by electron tunneling [141]. However, CT and electronic conduction are two fundamentally different processes, since the former is driven predominantly by dynamical relaxation of donor and acceptor sites, while the latter is enabled by an external potential which may change the electronic structure of the molecular junction. Moreover, electrons are injected and absorbed by the continuum of states of the electric leads during the conduction process, thus the molecular bridge is occupied only transiently by the incoming electrons.

1.4.1 Electronic Structure

The theoretical study of both CT and electronic conduction requires the knowledge of the molecules' electronic structure, which can be almost intractable when dealing with large systems containing hundreds or thousands of atoms. Therefore, the computation of CT parameters is realized at different levels of theory depending on the system size. Furthermore, the complexity of the electronic problem is often reduced by using a coarse-grained (CG) model Hamiltonian, instead of a full atomistic Hamiltonian. In case of DNA HT, the atomistic electronic structure can be mapped onto a simple Hückel matrix, which diagonal elements, ϵ_i, represent effective oxidation potentials of the nucleobases i, while the off-diagonal elements comprise the π–π electronic couplings T_{ij} between nearest-neighbor DNA bases i and j (Fig. 1.5). Moreover, the remainder of the system, e.g. DNA sugar-phosphate backbone, solvent and counterions can be described approximately as simple MM point charges polarizing the QM part of the system, thus representing a first-order description of QM/MM interactions. For HT in DNA, the CT parameters, ϵ_i and T_{ij}, have been computed with various quantum chemical methods, such as HF [39],

(a) Static Molecular Chains   (b) Introducing Disorder
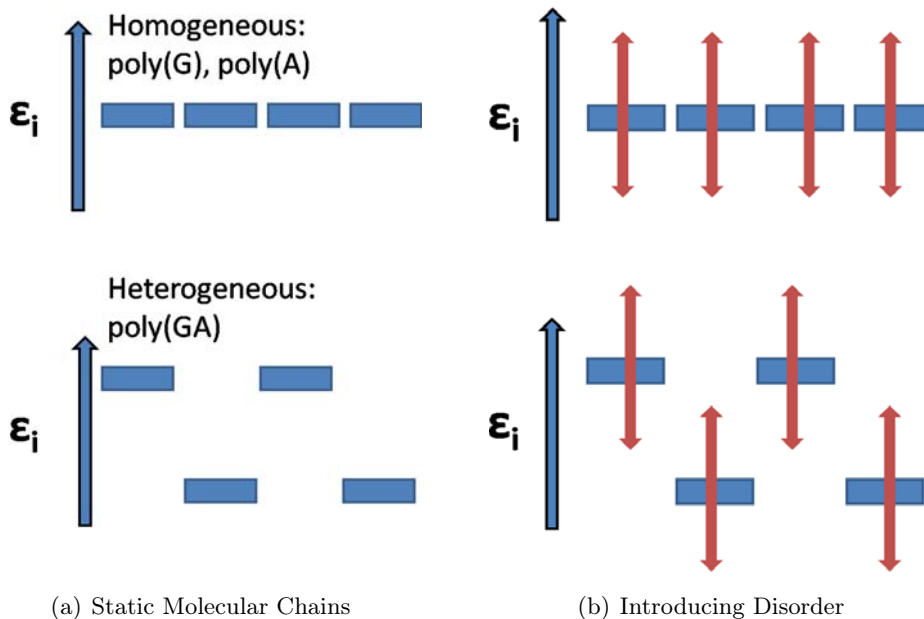
Figure 1.6: Effect of Dynamical Disorder on the energy levels for DNA HT. a) Static, idealized site energies for a homogeneous and a heterogeneous sequence. b) Fluctuation of site energies (indicated by arrows), due to structural and environmental fluctuations.

DFT [3, 142–144], and even with CAS-PT2 [145], over the past decade.

Nevertheless, the computation of CT parameters for a large number of conformations, e.g. a time series of molecular snapshots, is hardly affordable with standard *ab initio* approaches such as DFT or HF, thus vital conformational changes of biomolecules occurring on time scales longer than pico-seconds cannot be accounted for. Therefore, the usage of semi-empirical methods is becoming more popular in this regard [146]. A DFT-based semi-empirical method, referred to as DFTB, is used in this work in combination with a fragment orbital (FO) approach to compute CT parameter in DNA and *E. coli* photolyase. This DFTB-FO method was shown to produce accurate CT parameter for HT in DNA at the level of full DFT methods, at the same time being almost three orders of magnitude faster [147]. As a consequence, the DFTB-FO method allows for the application to extended molecular dynamics (MD) simulations, e.g. up to hundreds of nano-second in case of DNA. The development and testing of the DFTB-FO method represents one part of this work as described in chapter 3.

1.4.2   The Need of Including Dynamical and Environmental Effects

In several theoretical studies, Hückel Hamiltonians were used to compute coherent transport through DNA molecules [92, 148–151]. These computations were based on static, partially idealized structures (Fig. 1.6(a)). However, DNA and other biological molecules are no rigid materials, rather they are structurally flexible species whose conformational dynamics also depend vitally on their environments, i.e. the solvent. Taking into account

the structural dynamics and environmental effects, introduces disorder into the molecular chain (Fig. 1.6(b)). The impact of dynamical disorder on biomolecular CT, due to structural and environmental fluctuations, has been studied in great detail for various systems [130, 137, 152–164]. According to these studies, conformational fluctuations and the polarization by the environment cannot be neglected in a realistic description of biomolecular CT. In case of DNA, the dynamical disorder should not be regarded as a weak perturbation, but rather as a critical component either promoting or hindering charge motion [41, 142, 144, 165–167].

In some previous theoretical studies on DNA conduction, the effect of dynamical disorder was included using statistical models [149, 168–171]. However, the combination of extended molecular dynamics (MD) simulations in explicit solvent and coherent transport calculations, as realized in chapter 4 and 5 in this work, indicate severe shortcomings of the statistical approach, and is shown to include dynamical and environmental effects in a new and more realistic way [167, 172].

1.4.3 Direct Dynamics

Nevertheless, the combination of MD and coherent transport, i.e. configurational averaging, holds only in the adiabatic regime, that is the time scale of electronic motion is presumed to be faster than that of molecular motion. Moreover, the CT is assumed to occur solely via electron tunneling.

Generally, a more promising way to realize actual CT dynamics is to propagate the electronic system, e.g. the Hückel matrix, directly by solving the time-dependent Schrödinger equation without presupposing any mechanistic limits. Such a direct dynamics is computationally more expensive, yet CG and QM–MM separations of the system can reduce effectively the complexity of the electronic problem to be solved. In case of HT, it is sufficient to just follow the time evolution of the hole charge, while the remainder of the system is described by classical molecular dynamics, as explained in Sec. 2.6.

Recently, Kubař and Elstner [173] developed a non-adiabatic method in which both quantum and classical systems, i.e. electrons and ions, are propagated simultaneously. Therefore, such a coupled electron-ion dynamics (EID) allows for a realistic dynamic description of biomolecular CT, especially if the timescales of nuclear and electron motion are comparable, but also under non-equilibrium conditions. The latter situation refers to multistep CT processes, in which the environment is not given enough time to relax fully in the intermediate charge state [128]. Also, the direct approach allows to study the degree of charge delocalization during the CT process, as investigated e.g. for HT in DNA [173]. More importantly, the QM/MM description in the EID approach is improved greatly compared to that for configurational averaging, since the environment now "feels" the excess charge and, therefore, can reorganize itself during the CT process, e.g. rearrangement of water molecules. In chapter 8 of this work, the method from Ref. [173] is used to study the CT in photolyase on a microscopic level.

1.5   Thesis Outline

This dissertation is organized as follows. First, the theoretical background of CT in biomolecular systems is described (chapter 2). More precisely, the phenomenological CT theories are discussed, and the computation of CT parameters via both MM and QM methods is described. Also, the theory behind both MD simulations and the applied QM methods HF, DFT and DFTB are addressed briefly. Subsequently, the work is divided into two major parts I and II. The first one deals with the optimization and testing of the DFTB-FO method, and also the computation of realistic CT parameter for hole transport in DNA (chapter 3). Moreover, the CT parameters are computed along molecular dynamics trajectories (chapter 4). The corresponding time series are analyzed in-depth with regard to dynamical and environmental effects. Eventually, the time series of CT parameters is then used to compute time-dependent electronic transport properties. The same strategy is applied for the DNA derivative G4 (chapter 5). The second part of this work is devoted to the photoactivation process in *E. coli* photolyase, more precisely, the hole transfer between the triple Trp chain as described in Sec. 1.3. The results of classical CT theory (chapter 7) are compared to those of the coupled QM–MM approach (chapter 8). Finally, conclusions for both result parts I and II are provided separately at the end of each part (chapter 6 and 9).

2

Theoretical Framework and Methods

In this chapter, the theoretical background of the applied methods to study biomolecular CT is given. First, two major mechanisms of how CT proceeds in molecular systems, are described briefly. Then, two phenomenological theories are discussed, beginning with Marcus' theory of ET in Sec. 2.2, also including the computation of the involved CT parameter. Subsequently, the Landauer-Büttiker formalism for the study of coherent transport through bridged contact-molecule-contact systems is described in Sec. 2.3, and the linear chain model is introduced.

In this work, the CT parameter are either extracted from classical molecular dynamics (MD) simulations or computed along their trajectories. Therefore, a short introduction is given into the theory of MD simulations in Sec. 2.4. Eventually, the calculation of quantum chemical CT parameter entering the various CT models, is explained in Sec. 2.5. For this purpose, first, the fragment-orbital (FO) method is introduced, followed by a brief description of the applied quantum chemical methods: HF, DFT and DFTB in Sec. 2.5.2, 2.5.3 and 2.5.4, respectively. For the latter DFT-based methods, some issues are discussed in Sec. 2.5.3.1, in particular with regard to the application on CT phenomena. Finally, the coarse-grained non-adiabatic electron-ion dynamics (EID) is discussed in Sec. 2.6.

2.1 Basic Mechanisms of Charge Transport

In the following, a donor-bridge-accepter (DBA) CT system is considered. The most basic expression for the non-adiabatic CT rate in such molecular systems is given by Fermi's golden rule [141]:

$$k_{D \to A} = \frac{2\pi}{\hbar} |V|^2 F. \tag{2.1}$$

Here, $\hbar$ is the reduced Planck constant, V is the effective electronic coupling and F the thermally-averaged nuclear vibrational Franck-Condon factor.

In general, the dependence of the CT-rate on the donor-acceptor distance R_{DA}, can be described by an exponential decay:

$$k_{D \to A} = k_0 \exp\left(-\beta R_{DA}\right), \tag{2.2}$$
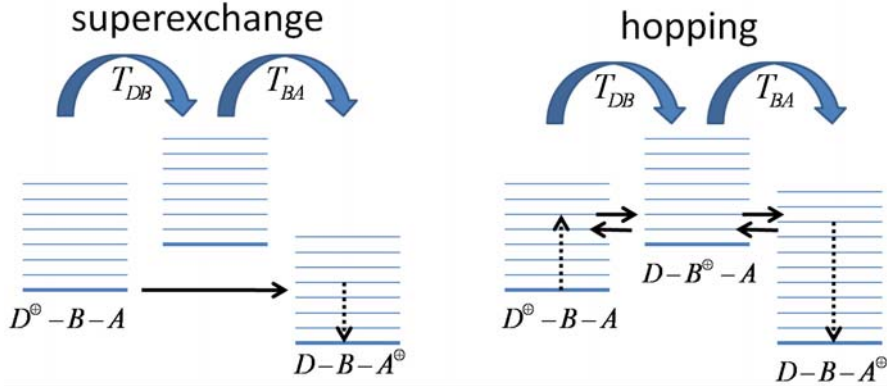
Figure 2.1: Energetic level scheme for hole transfer in a donor-bridge-acceptor (DBA) system. Bold levels indicate the ground state respectively, while vibronic-excited levels are indicate by thin lines. Two distinct CT mechanism are depicted: CT via superexchange (left) and CT via hopping (right). The solid arrows indicate the CT process, while the dashed arrows indicate either thermal activation or vibrational relaxation.

with the pre-exponential factor k_0 and the falloff parameter β. There are two major mechanisms of how CT in molecular DBA systems, such as DNA or proteins, can occur (Fig. 2.1).

2.1.1   CT via Superexchange

For small R_{DA}, the unistep superexchange mechanism was proposed (Fig. 2.1(left)). This process is described as coherent tunneling, as the CT proceeds directly from the donor to the acceptor without any charge occupation on the bridge site, i.e. no chemical intermediates involving oxidized (hole transfer (HT)) or reduced (electron transfer (ET)) bridge sites are observed. The CT rate in the superexchange regime depends strongly on R_{DA}, e.g. for single-step tunneling in DNA, β from Eq. 2.2 is assumed to be larger than 0.6 Å[157]. Though, the bridge states are not populated physically, the off-resonant electronic couplings T_{DB} and T_{DA} as well as the energy differences between the donor and acceptor states to the bridge states ΔE_{DB} and ΔE_{BA} have a considerable impact on the CT rate. All these energetic contributions which involve the molecular bridge states are included in the coupling V from Eq. 2.1. In the superexchange regime, $V = V_{\text{super}}$ can be calculated for donor-acceptor systems with multiple bridges as [37, 41]:

$$V_{\text{super}} = \frac{T_{DB_1} T_{B_n A}}{\Delta E_{DB_1}} \prod_{j=1}^{n-1} \frac{T_{B_{j+1} B_j}}{\Delta E_{DB_{j+1}}}. \tag{2.3}$$

For larger R_{DA} the superexchange mechanism breaks down and the strong distance dependence of the CT rate on R_{DA} vanishes. For instance, single-step tunneling in DNA between two G–C Watson-Crick pairs (WCP) separated by n A–T WCP is said to be the dominant CT mechanism if $n <= 3$ (about 12 to 15 Å). For $n > 4$, the CT in DNA cannot be described by unistep superexchange anymore, instead the R_{DA} dependence

nearly disappears [24]. However, the distance domain for CT via superexchange is said to be considerably larger in proteins [37].

2.1.2 CT via Hopping

The weak R_{DA} dependence of the CT rate at large distances was observed in CT experiments, and the crossover of the mechanisms in DNA had been predicted by theory as well [37, 42]. Interestingly, this different mechanism allows for CT at very large distances. For instance, photochemically or thermally created hole charges in DNA were observed to migrate over 50 Å^{-1} (10-12 WCP) [22, 23]. These findings could be explained only by an incoherent thermally-induced hopping process (TIH). In this multistep process, the charge can be localized on the bridge states as well (see Fig. 2.1), thus the corresponding chemical intermediates may be observed experimentally. The charge on the donor can be activated thermally, which then allows for CT to the bridge site via resonant coupling T_{DA} [39]. By contrast, the CT via superexchange is described by off-resonant electronic couplings T_{DA} [33]. The falloff parameter β from Eq. 2.2 was found to be smaller than 0.1 Å^{-1} if CT occurs via multistep hopping [34].

In the following sections 2.2 and 2.3, two fundamental theoretical concepts are described. These are used in this work to study either i) the transfer of a localized charge in a molecular system (Marcus theory), or ii) the coherent electronic transport through microscopic objects when coupled to electrodes (Landauer-Büttiker formalism).

2.2 Marcus' Theory of Electron Transfer

The theoretical foundation of charge transfer remains Marcus' seminal theory [174–177] and its extensions [178–182]. Within the assumption of weak electronic coupling between the donor and the acceptor, the rate of charge transfer can be obtained in the high-temperature non-adiabatic limit as:

$$k_{D \to A} = \frac{T_{DA}^2}{\hbar} \sqrt{\frac{\pi}{\lambda k_B T}} \exp\left(-\frac{(\Delta G^0 + \lambda)^2}{4\lambda k_B T}\right). \tag{2.4}$$

T_{DA} is the electronic coupling between the donor and the acceptor state, ΔG^0 is the free energy difference between the initial and the final state[1], and λ is the reorganization energy, i.e. the energy required to change the molecular structure of the initial state to that corresponding to the relaxed product state without charge transfer taking place. λ is usually decomposed into two parts $\lambda = \lambda_s + \lambda_i$; the outer-sphere contribution λ_s represents the rearrangement of the surrounding, e.g. the solvent and the protein, upon the redistribution of charge and the internal contribution λ_i describes the change of the molecular structure of the donor and the acceptor themselves.

Comparing the shape of Eqs. 2.4 and 2.1 offers some similarities, in particular concerning the electronic coupling term. However, the Franck-Condon factor is now replaced

---

[1]The superscript ^{0} refers to the standard state of 1 mole reactants at 1 bar and 298.2 K.
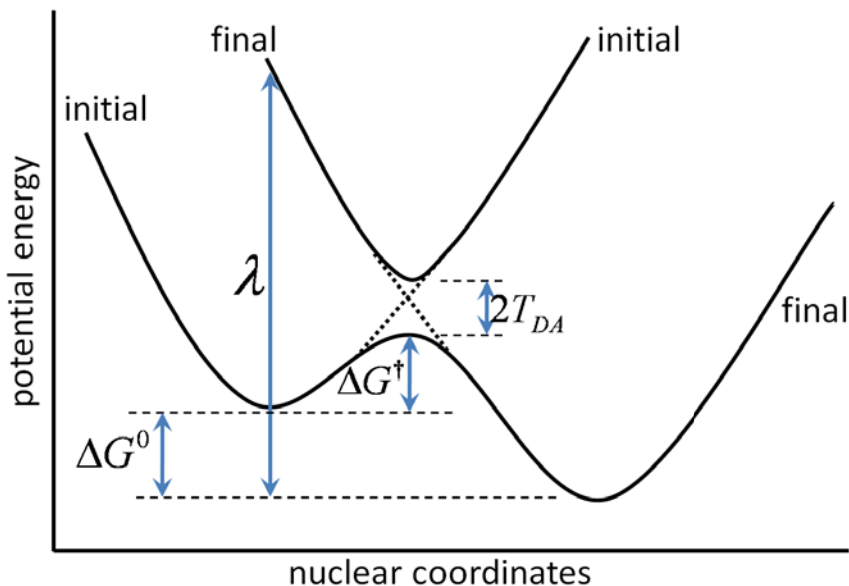
Figure 2.2: Adiabatic (solid lines) and diabatic (dashed lines) potential energy curves for the initial and the final states of a CT reaction: $D^+ + A \rightleftharpoons D + A^+$. The electronic couplings T_{DA} as half of the energy splitting, the reorganization energy λ, the free energy ΔG^0 and the activation barrier $\Delta E^\dagger$ are indicated as well.

by a term which contains well-defined quantities that are more accessible by theoretical calculations. The energetic situation within the Marcus' picture of CT is illustrated in Fig. 2.2.

The potential energy curves of the initial and final diabatic states are usually approximated by harmonic parabolas, and plotted against a generalized reaction coordinate of the whole system. The corresponding adiabatic states due to the energy splitting $2T_{DA}$ are indicated in Fig. 2.2 as well. The CT reaction is said to be non-adiabatic if the energy splitting is weak, i.e. $2T_{DA} < k_B T$ [157]. On the other hand, adiabatic or diabatic CT occurs, respectively, if the energy splitting is either very large or there is no electronic interactions between the diabatic states at all.

The Franck-Condon principle and the law of energy conservation require that CT within Marcus' theory can take place only at the crossing point of the diabatic states [157]. The energy difference between the initial state and that crossing point define the activation barrier $\Delta E^\dagger$. In the classical Marcus' theory, the reorganization energy λ describes the curvature of the parabolas in Fig. 2.2, and as such is assumed to be the same for the forward and backward CT reaction. Within this approximation the activation barrier $\Delta E^\dagger$ may be expressed as:

$$\Delta G^\dagger = \frac{\left(\Delta G^0 + \lambda\right)^2}{4\lambda}. \tag{2.5}$$

Thus, the activation barrier $\Delta G^\dagger$ is $\lambda/4$ for CT reactions between identical species (e.g. $\Delta G^0 = 0$).

The estimation and calculation of the electronic coupling T_{DA}, the reorganization energy λ and the free energy ΔG^0 is the subject of the following subsections.

2.2.1 Donor-Acceptor Coupling T_{DA}

As indicated in Fig. 2.2, the electronic coupling T_{DA} can be calculated as one half of the energy splitting between the adiabatic states.

$$T_{DA} = \frac{1}{2} \left(E_1 - E_2 \right) \tag{2.6}$$

The energy splitting can be calculated easily from the secular equation for the two-state donor-acceptor system:

$$\begin{vmatrix} H_{DD} - E & H_{DA} - ES_{DA} \\ H_{DA} - ES_{DA} & H_{AA} - E \end{vmatrix} = 0, \tag{2.7}$$

with the pure electronic Hamilton and overlap matrix elements $H_{DA} = \left\langle \Psi_D \left| \hat{H} \right| \Psi_A \right\rangle$ and $S_{DA} = \langle \Psi_D | \Psi_A \rangle$, respectively. Here, Ψ_D and Ψ_A are the wave function of the donor and the acceptor.

The solution of Eq. 2.7 leads to the adiabatic energies E_1 and E_2:

$$E_{1/2} = \frac{2 S_{DA} H_{DA} - H_{AA} - H_{DD}}{2 \left(1 - S_{DA}^2 \right)}$$

$$\pm \frac{\sqrt{\frac{1}{4} \left(H_{DD} - H_{AA} \right)^2 + H_{DA}^2 + S_{DA}^2 \left(H_{DD} H_{AA} \right) - S_{DA} H_{DA} \left(H_{DD} + H_{AA} \right)}}{1 - S_{DA}^2}.$$

The first term cancels out when inserting this result into Eq. 2.6. What remains is just the second term. The T_{DA} are defined at the crossing of the diabatic states, where it is assumed that $H_{DD} = H_{AA}$. Using this, leads to:

$$T_{DA} = \frac{H_{DA} - S_{DA} H_{DD}}{1 - S_{DA}^2},$$

with $H_{DD} = \frac{1}{2} \left(H_{DD} + H_{AA} \right)$ one obtains:

$$T_{DA} = \frac{H_{DA} - S_{DA} \left(H_{DD} + H_{AA} \right) / 2}{1 - S_{DA}^2}. \tag{2.8}$$

In the non-adiabatic weak coupling regime, the donor-acceptor coupling is approximated by the Hamilton matrix element, i.e. $T_{DA} \approx H_{DA}$. However, this simplifications turns out to be not practicable for the hole transfer within π-stacked nucleobases in DNA. There, the overlap S_{DA} between the wave functions of adjacent bases is not negligible, thus the second term in the numerator in Eq. 2.8 is of the same order of magnitude than H_{DA} [42, 157].

A different, more economical strategy to compute electronic couplings is the fragment orbital (FO) approach described in Sec. 2.5.

2.2.2 Reorganization Energy λ

As mentioned above, the reorganization energy λ is the free energy change required when moving the system from the initial (reactants) to the final (product) nuclear coordinates
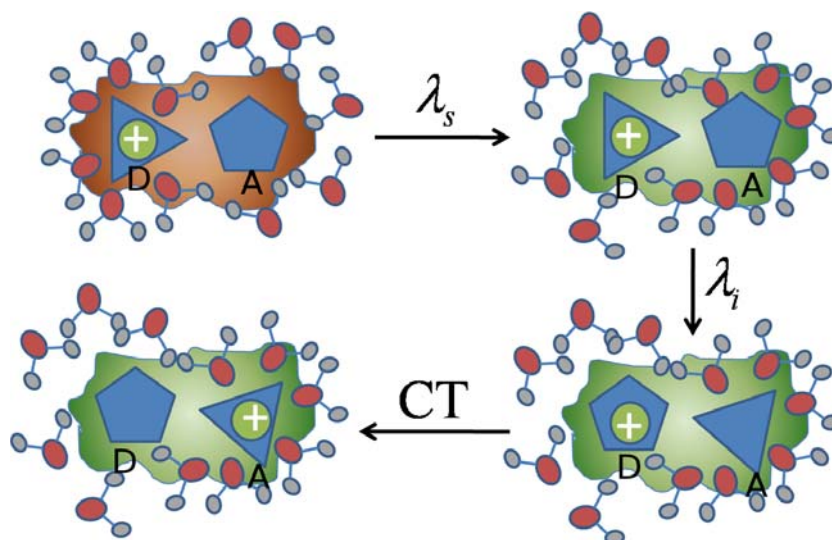
Figure 2.3: Illustration of both the internal λ_i and the outer-sphere λ_s contribution to the reorganization energy λ for a transfer of a positive charge from the donor (D) to the acceptor (A). D and A (blue) are embedded in a larger entity (e.g. CT between amino acids within a protein), which is indicated brown/green. Moreover, D and A are surrounded by small molecules (e.g. solvent). Contributions to λ_s are indicated by an rearrangement of the small molecules around D and A as well as a global structural change of the larger entity (brown→green). By contrast, contributions to λ_i are illustrated as a change of the structure of D and A themselves, which is indicated by the change of shapes (triangle⇌pentagon).

without CT actually taking place. The rearrangement of the surrounding (λ_s) as well as the changes in the donor and acceptor themselves (λ_i) are illustrated in Fig. 2.3. The rearrangement of the system is supposed to take place before the CT happens according to the definition of λ. This definition appears to be arbitrary, since CT can take place only at the crossing point of the diabatic states, as mentioned above.

Whether λ_s or λ_i has a stronger impact on the CT will certainly depend on the type of system one is interested in. It is believed that λ_s will dominate the overall reorganization for CT reactions of biomolecules in polar solvents. In context of hole transfer in DNA, it was pointed out that the driving force of the process is indeed the solvent rearrangement, i.e. a water polaron is formed which effectively slows down the CT [173]. ΔG^0 may be assumed to be approximately zero in systems where CT is taking place between identical species in homogeneous environments (e.g. adjacent guanines in poly(G)). Thus, the CT process is determined critically by λ_s. On the other hand, λ_i might be the critical property for CT in systems with nonpolar solvents or in organic crystals.

In theoretical studies, λ_s is usually estimated classically, whereas λ_i is computed using quantum chemical methods. The first estimations of λ_s were based on electrostatic continuum models [174, 177], while more recent studies rely on atomistic MD simulations.

2.2.2.1 Two-Sphere Model

A very simple way to estimate λ_s was introduced by Marcus [176]. Accordingly, donor and acceptor are treated as conducting spheres with radii r_D and r_A. Then, λ_s for the transfer of a charge ΔQ from the donor to the acceptor can be calculated as:

$$\lambda_s = \frac{\Delta Q^2}{2} \left(\frac{1}{r_D} + \frac{1}{r_A} - \frac{2}{R_{DA}} \right) \left(\frac{1}{\epsilon^{\text{opt}}} - \frac{1}{\epsilon^{\text{st}}} \right), \tag{2.9}$$

with the optical and static dielectric constants ϵ^{opt} and ϵ^{st}. Nevertheless, finding good estimates for ϵ^{opt} and ϵ^{st} in heterogeneous systems, such as proteins in solution, is a rather ambitious task.

The following two approaches to λ_s are based on classical MD simulations. However, the theoretical background of MD simulations is given in Sec. 2.4.

2.2.2.2 λ_s from MD I

Consider a charge shift reaction of a hole charge from the donor to the acceptor:

$$D^{\oplus} + A \rightleftharpoons D + A^{\oplus}.$$
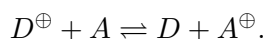
Assuming that both classical MD simulations for the initial system (reactants) and for the final system (products) have been carried out already, λ_s can be obtained by performing energy re-evaluations on the existing trajectories using, respectively, the opposite charge state. For example, the total energies of the trajectory (structural ensemble) of the initial state $D^{\oplus} + A$ are re-calculated using the charge topology of the final state $D + A^{\oplus}$. This proceeding is denoted here as $\left\langle E_{D,A^{\oplus}} \right\rangle_{D^{\oplus},A}$. The index within the brackets denotes the charge state used for the energy re-evaluation, while the outer index represents the structural ensemble (trajectory) on which the re-evaluation is carried out. Following this notation means vertical excitation from the reactants ground state to the product state without any structural relaxations (see Fig. 2.4(a)).

The corresponding energy differences λ_+ and λ_- in Fig. 2.4(a) may be called effective reorganization energies which are obtained as:

$$\begin{aligned} \lambda_- &= \lambda_s - \Delta G^0 = \left\langle E_{D,A^{\oplus}} \right\rangle_{D^{\oplus},A} - \left\langle E_{D^{\oplus},A} \right\rangle_{D^{\oplus},A}, \\ \lambda_+ &= \lambda_s + \Delta G^0 = \left\langle E_{D^{\oplus},A} \right\rangle_{D,A^{\oplus}} - \left\langle E_{D,A^{\oplus}} \right\rangle_{D,A^{\oplus}}. \end{aligned} \tag{2.10}$$

Nevertheless, λ_- and λ_+ are no reorganization energies in the classical sense of Marcus theory as they contain the free energy difference $\pm\Delta G^0$. Eventually, the determination of the desired quantity λ_s is realized by taking into account the assumption of equal reorganization energies for forward and backward reactions:

$$\lambda_s = \left(\lambda_- + \lambda_+ \right)/2. \tag{2.11}$$

Note that once λ_s is known ΔG^0 is determined by Eq. 2.10 as well. The main advantage with this estimation is that only one ensemble of structures is needed, thus reducing

(a) MD I   (b) MD II
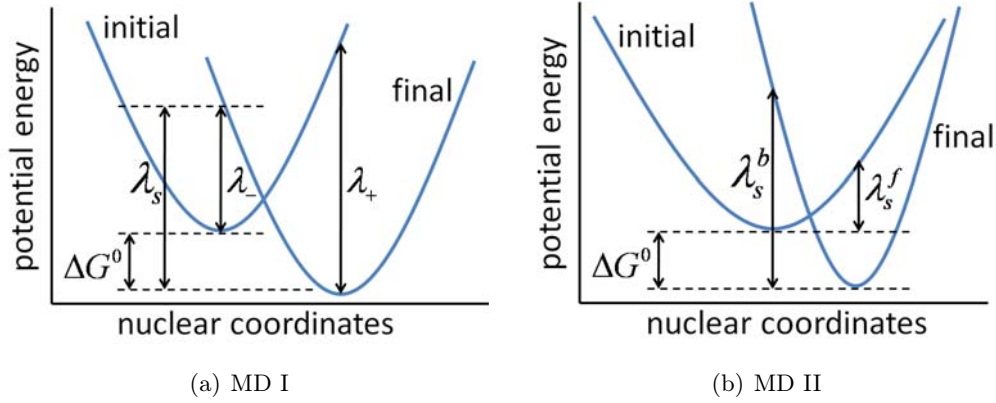
Figure 2.4: MD I: Illustration of the quantities λ_+ and λ_- that are no reorganization energies in the classical sense as they contain $\pm\Delta G^0$. λ_s is assumed to be equal for forward and backward CT reaction. **MD II:** Direct calculation of λ_s^f and λ_s^b, respectively, for the forward and backward reaction following the classical definition of λ. In this example, λ_s^f and λ_s^b, i.e. the curvatures of the parabolas, were chosen to be quite different, thus the approximation $\lambda_s^f = \lambda_s^b$ would be inadequate.

statistical errors. On the other hand, the assumption of equal λ_s for forward and backward CT reactions within Marcus' classical theory may not be appropriate in systems where the charge movement is accompanied by a strongly changing environment, e.g. a CT from the inside to the solvent-exposed exterior of a protein (part II). An example for which the forward (λ_s^f) and backward (λ_s^b) reorganization energies differ significantly is shown in Fig. 2.4(b).

2.2.2.3 λ_s from MD II

Following the notation from the previous paragraph, λ_s for the charge shift reaction above is obtained directly as:

$$\begin{aligned}
\lambda_s^f &= \left\langle E_{D^\oplus,A} \right\rangle_{D,A^\oplus} - \left\langle E_{D^\oplus,A} \right\rangle_{D^\oplus,A}, \\
\lambda_s^b &= \left\langle E_{D,A^\oplus} \right\rangle_{D^\oplus,A} - \left\langle E_{D,A^\oplus} \right\rangle_{D,A^\oplus},
\end{aligned} \tag{2.12}$$

for the forward and backward reaction, respectively. Note that this procedure follows exactly the classical definition of the reorganization energy (Fig. 2.2). The main advantage over the former MD procedure is that λ_s is not assumed to be equal for both the forward and backward reaction (Fig. 2.4(b)), thus this methodology may be regarded as a step beyond the classical Marcus' theory. On the contrary, the chances for statistical errors are increased due to insufficient sampling. The problem is a consequence of comparing total energies of two different independent simulations (trajectories) that can cause large deviations. The protocol explained here was applied successfully to the computation of λ_s for hole transfer in DNA [183].

2.2.2.4 Lack of Polarization

Both classical MD-based methods presented above share a common imperfection as they cannot describe fast electronic polarization effects. It could be shown that reorganization energies computed from classical non-polarizable force field simulations overestimate λ_s significantly [136, 184–186]. Thus, the resulting values of λ_s should be scaled by the inverse optical dielectric constant $1/\epsilon^{\mathrm{opt}}$ of the medium, which is not immediately accessible for heterogeneous environments (e.g. a solvated protein) as mentioned above. Various scaling factors in the range of $1/2 - 1/1.4$ were suggested in the past [127, 134, 135, 187, 188]. Therefore, the MD-based λ_s values for the CT in the protein photolyase (PL) (chapter 7) have to be scaled to account for electronic polarization effects at least empirically. A comparison for two different scaling factors will be provided. Certainly, a change of λ will have a considerable impact on the CT rate, thus results extracted from purely classical MD simulations should be interpreted qualitatively only.

2.2.2.5 Internal Reorganization λ_i

Despite the assumption of $\lambda_s > \lambda_i$ for CT in a polar environment, the internal reorganization energy λ_i can be an important parameter in the CT process, in particular for homogeneous systems such as organic crystals. In this work, λ_i is estimated for a CT reaction between chemically identical species by using quantum chemical calculations [189, 190]. In this particular case, only one, either the donor or the acceptor needs to be computed, and the sum of reorganization energies of deionization and ionization for the charge shift reaction above is given by:

$$\lambda_i = \left(E_{\mathrm{opt.neutral}}^{\oplus} - E_{\mathrm{opt.charged}}^{\oplus} \right) + \left(E_{\mathrm{opt.charged}}^{0} - E_{\mathrm{opt.neutral}}^{0} \right). \tag{2.13}$$

Here, $E^{\oplus}$ is the energy of the ionized molecule, while E^{0} is the energy of the neutral molecule. The subscript specifies the molecular geometry at which the energy is re-evaluated, i.e. either the optimized neutral or ionic structure. This procedure is the same as used in Refs. [191, 192]. Note that Eq. 2.13 accounts already for the relaxation of both the donor and the acceptor. The internal reorganization energy λ_i is usually calculated in vacuo, though recent studies indicated that polarization may have an effect on λ_i as well, using an implicit solvent model [173].

2.2.3 Free Energy Difference ΔG^0

The remaining parameter entering Eq. 2.4 is the driving force ΔG^0. Regarding the exponential term in Eq. 2.4, the CT rate dependence on ΔG^0 can be twofold. On the one hand, the rate increases with ΔG^0 as long as $-\Delta G \leq \lambda$, which is called the "normal" regime. On the other hand, the rate slows down if the driving force is increased further $-\Delta G > \lambda$. The latter behavior is referred to as the "inverted" regime. The maximum CT rate is obtained at the turning point for $-\Delta G = \lambda$. In this case, the CT reaction proceeds without any activation barrier, e.g. $\Delta E^{\dagger} = 0$.

Thermodynamic quantities such as the Gibbs Free energy differences ΔG^0 can be computed from MD simulations. The Helmholtz free energy A^2 of a N-particle system can be expressed as:

$$A = -k_B T \ln Q, \tag{2.14}$$

where Q is the partition function:

$$Q = \iint e^{-\frac{E(\vec{r},\vec{p})}{k_B T}} \, d\vec{r} d\vec{p}. \tag{2.15}$$

Here, $(\vec{r}, \vec{p})$ represents a point in the phase space ($\vec{r} = \{r_1, \cdots, r_{3N}\}$ and $\vec{p} = \{p_1, \cdots, p_{3N}\}$) of the system which is sampled by MD. The canonical probability $\rho(\vec{r}, \vec{p})$ of finding the system in state $(\vec{r}, \vec{p})$ is:

$$\rho(\vec{r}, \vec{p}) = \frac{e^{-\frac{E(\vec{r},\vec{p})}{k_B T}}}{Q}. \tag{2.16}$$

Eq. 2.14 can be re-written using Eq. 2.16 [193]:

$$A = -k_B T \ln \iint e^{\frac{E(\vec{r},\vec{p})}{k_B T}} \rho(\vec{r}, \vec{p}) \, d\vec{r} d\vec{p} \tag{2.17}$$

The fundamental problem is that large energies of the system, i.e. far from the equilibrium state, enter Eq. 2.17 exponentially, thus have a significant impact to the free energy. However, these high-energy regions in the phase space may be reached very rarely within finite simulation time (ergodicity), thus causing considerable errors in the computation of average quantities. Reasonably, the computation of free energy differences ΔG via MD should not rely on absolute energies, i.e. G_A for state A and G_B for state B, as the result can be afflicted significantly with statistical errors. Nevertheless, the driving force ΔG^0 for the donor-acceptor system above may still be estimated approximately this way:

$$\Delta G^0 = \left\langle E_{D,A^\oplus} \right\rangle_{D,A^\oplus} - \left\langle E_{D^\oplus,A} \right\rangle_{D^\oplus,A}. \tag{2.18}$$

Thus, additional errors arise as total energies of two different simulations are compared (same problem as in Eq. 2.12). The estimation of ΔG^0 via Eq. 2.18 will be compared to more advanced methods, described in the following, for the CT in PL (chapter 7). Two major theoretical concepts were developed to overcome the problem described above: i) free energy perturbation (FEP) [194] and ii) thermodynamic integration (TI) [195]. Generally, absolute free energies are not the quantities of interest, rather than relative energy differences, in particular when considering chemical reactions or conformational changes between two states.

---

[2]The Helmholtz free energy A differs from the Gibbs free energies G by the pressure work pV, which is usually negligible when changes in condensed phase systems are studied.

2.2.3.1 FEP

The Gibbs free energy difference ΔG between two states A and B can be written as $(\beta = (1/k_B T))$:

$$\Delta G = G_B - G_A = -k_B T \ln \frac{Q_B}{Q_A} = -k_B T \ln \frac{\iint e^{-\beta E_B(\vec{r}, \vec{p})} d\vec{r} d\vec{p}}{\iint e^{-\beta E_A(\vec{r}, \vec{p})} d\vec{r} d\vec{p}} \tag{2.19}$$

Inserting $1 = e^{-\beta E_A(\vec{r}, \vec{p})} e^{\beta E_A(\vec{r}, \vec{p})}$ into the nominator in Eq. 2.19 leads to:

$$\begin{aligned} \Delta G &= -k_B T \ln \frac{\iint e^{-\beta E_B(\vec{r}, \vec{p})} e^{-\beta E_A(\vec{r}, \vec{p})} e^{\beta E_A(\vec{r}, \vec{p})} d\vec{r} d\vec{p}}{\iint e^{-\beta E_A(\vec{r}, \vec{p})} d\vec{r} d\vec{p}} \\ &= -k_B T \ln \iint e^{-\beta E_B(\vec{r}, \vec{p})} e^{\beta E_A(\vec{r}, \vec{p})} \rho_A(\vec{r}, \vec{p}) d\vec{r} d\vec{p} \\ &= -k_B T \ln \iint e^{-\beta(E_B(\vec{r}, \vec{p}) - E_A(\vec{r}, \vec{p}))} \rho_A(\vec{r}, \vec{p}) d\vec{r} d\vec{p} \end{aligned} \tag{2.20}$$

The integral in the last line of Eq. 2.20 can be interpreted as an expectation value, i.e. quantity times probability distribution [193]:

$$\begin{aligned} \Delta G(A \to B) &= -k_B T \ln \left\langle e^{-\beta(E_B - E_A)} \right\rangle_A \\ \Delta G(A \leftarrow B) &= -k_B T \ln \left\langle e^{-\beta(E_A - E_B)} \right\rangle_B \end{aligned} \tag{2.21}$$

Eq. 2.21 represents the essence of the FEP method developed by Zwanzig [194]. FEP in its standard formulation performs well as long as the energy difference between the states A and B is sufficiently small, i.e. $E_B - E_A < k_B T$ [196]. If the energy difference is larger, the phase space densities of A and B have only a small or no overlap at all. In these cases, intermediate states between A and B can be introduced by making use of a coupling parameter $\Lambda \in \{0, 1\}$[3]. In the simplest case, the coupling between the states can be realized by linear interpolation[4] [196]:

$$E_\Lambda = \Lambda E_B + (1 - \Lambda) E_A \tag{2.22}$$

Note that the transformation path $A_{\Lambda=0} \to B_{\Lambda=1}$ does not necessarily have to be physically realistic as the free energy is a state function which is independent on the way it was reached. In practice, $(N + 1)$ MD simulations have to be carried for N intermediate states Λ.

2.2.3.2 TI

By introducing the same coupling scheme as in the FEP method, the Gibbs free energy becomes dependent on Λ as well:

$$G = G(\Lambda) = -k_B T \ln Q(\Lambda), \tag{2.23}$$

---

[3] Λ is chosen instead of the more common λ to avoid confusion with the reorganization energy

[4] For more complicated systems, higher-order interpolation can be applied as well.

with $G(\Lambda = 0)$ and $G(\Lambda = 1)$, respectively for the states A and B. This allows one to write the free energy difference between the states A and B as a continuous integral with respect to Λ:

$$\Delta G = G_B - G_A = \int_{\Lambda=0}^{\Lambda=1} \frac{\partial G(\Lambda)}{\partial \Lambda} d\Lambda. \tag{2.24}$$

The derivative $(\partial G(\Lambda))/(\partial \Lambda)$ can be rearranged:

$$\begin{aligned}
\frac{\partial G(\Lambda)}{\partial \Lambda} &= -k_B T \frac{1}{Q(\Lambda)} \frac{\partial Q(\Lambda)}{\partial \Lambda} \\
&= -k_B T \frac{1}{Q(\Lambda)} \frac{\partial}{\partial \Lambda} \iint e^{-\beta E_\Lambda} d\vec{r} d\vec{p} \\
&= -k_B T \frac{1}{Q(\Lambda)} \iint \frac{\partial}{\partial \Lambda} e^{-\beta E_\Lambda} d\vec{r} d\vec{p} \\
&= -k_B T \frac{1}{Q(\Lambda)} \iint -\beta \frac{\partial E_\Lambda}{\partial \Lambda} e^{-\beta E_\Lambda} d\vec{r} d\vec{p} \\
&= \iint \frac{\partial E_\Lambda}{\partial \Lambda} \frac{e^{-\beta E_\Lambda}}{Q(\Lambda)} d\vec{r} d\vec{p} \\
&= \iint \frac{\partial E_\Lambda}{\partial \Lambda} \rho_\Lambda(\vec{r}, \vec{p}) \, d\vec{r} d\vec{p} = \left\langle \frac{\partial E_\Lambda}{\partial \Lambda} \right\rangle_\Lambda.
\end{aligned} \tag{2.25}$$

As a result, the derivative of the Gibbs free energy with respect to the coupling parameter Λ is calculated as the ensemble average of the derivative of the total MM energy E_Λ, which is accessible directly in the MD simulation. Eventually, the free energy difference ΔG^0 is computed via integration over $\Lambda \in \{0, 1\}$ [195]:

$$\Delta G^0 = G_B^0 - G_A^0 = \int_0^1 d\Lambda \left\langle \frac{\partial E_\Lambda}{\partial \Lambda} \right\rangle_\Lambda \approx \sum_i \Delta\Lambda \left\langle \frac{\partial E_\Lambda}{\partial \Lambda} \right\rangle_{\Lambda_i}, \tag{2.26}$$

In practice, the integral is approximated numerically by computing $\langle \partial E_\Lambda / \partial \Lambda \rangle$ at various fixed values Λ_i between which linear or higher order interpolation can be applied (Eq. 2.22). The angular brackets in Eq. 2.26 denote Boltzmann-weighted averages, that is averaging is performed by computing $\partial E_\Lambda / \partial \Lambda$ for an ensemble of conformational snapshots generated by molecular dynamics simulations. In this work, the GROMACS software package is used [197, 198] (chapter 7). Accordingly, the TI method consists of various independent MD-simulation windows, in each of which the gradient of the free energy curve is computed at a different Λ-point. The quality of the TI result depends significantly on the simulation length as the ensemble average at each Λ_i requires the corresponding simulation to be in equilibrium, that is the system needs to be adapted to the Λ_i. In context of electron transfer processes both states A and B have to be defined. The charge is assumed to be completely localized on one site at the beginning as well as at the end of the CT process. Therefore, the functions E_A and E_B differ only due to the atomic partial charges of the CT-active sites and E_Λ is constructed by linear mixing. Note that the transition path from $\Lambda = 0$ to $\Lambda = 1$, i.e. moving a charge from site A to site B, does not necessarily have to be a physically realistic one, as mentioned

before for the FEP method. The Gibbs free energy change, e.g. for CT in a solvated protein, is usually decomposed into an inner and an outer contribution. The internal part contains changes in the geometry as well as in the electronic structure of the CT-active species, e.g. amino acid sidechains. On the other hand, the outer part is composed of global conformational changes in the protein as well as in the solvent arrangement, i.e. change of the solvent polarization. The internal part of ΔG^0 cannot be described correctly by molecular mechanics force fields, but in the case of CT between identical species (Trp sidechains in part II), this contribution to ΔG^0 is vanishingly small and may be neglected.

2.2.3.3 Fast-Growth TI

The previous method is also referred to as discrete TI (DTI) as multiple MD simulations are carried out with a different fixed Λ value. However, each of the simulations is supposed to be in equilibrium which may implicate extended simulation times, in particular for large complex system such as CT in solvated proteins. Even simulation lengths of several nano-seconds for each Λ_i may not guarantee converged $\langle \partial E_\Lambda / \partial \Lambda \rangle$ values. A different strategy is to compute the equilibrium property ΔG^0 from non-equilibrium simulations.

Jarzynski has shown that the Helmholtz free energy difference ΔA can be computed from a series of non-equilibrium work simulations [199, 200]:

$$e^{-\beta \Delta A} = \left\langle e^{-\beta W} \right\rangle. \tag{2.27}$$

Here, the average $\langle \rangle$ is taken over an ensemble of short-length trajectories, which were initiated from an equilibrated canonical ensemble. The irreversible work W for each of these non-equilibrium trajectories is given by:

$$W = \int_0^1 \frac{\partial E_\Lambda}{\partial \Lambda} d\Lambda. \tag{2.28}$$

Therefore, the switching process $A_{\Lambda=0} \rightarrow B_{\Lambda=1}$ is performed for each single simulation within a switching time that may be very short (fast growth TI (FGTI)), instead of running simulations at fixed Λ values (DTI). The system stays close to equilibrium for very long switching times, hence Eq. 2.28 describes basically conventional TI and $\Delta G^0 = W$ [201]. The advantage of FGTI over DTI is that the computation time can be spent rather on running many short simulations in a highly parallel fashion, than on running only few extended simulations, in which the potential for parallelization is limited.

There are various variants of the FGTI method. A recent comparative review is given in Ref. [201]. A valuable improvement is to consider the backward reaction $A \leftarrow B$ as well. Accordingly, work distributions $P_f(W)$ and $P_b(-W)$ are obtained for both the forward and backward reaction, respectively. Note that the non-equilibrium work simulation for the forward and backward reaction are started from different canonical ensembles, i.e. equilibrium canonical ensembles of state A for $P_f(W)$ and of state B for $P_b(-W)$. According to Crooks fluctuation theorem (CFT) [202, 203]:
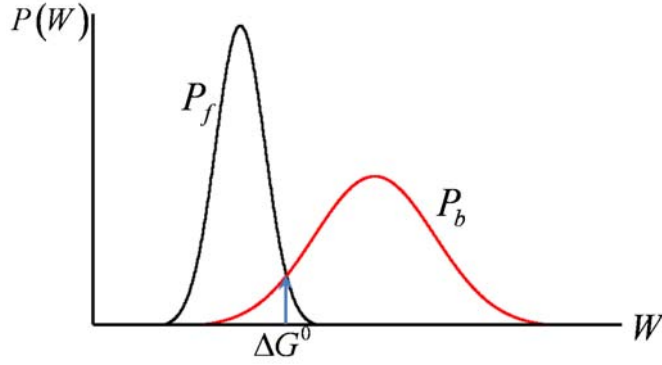
Figure 2.5: Illustration of the Crooks Gaussian intersection (CGI) method. The Gibbs free energy difference ΔG^0 is obtained as the intersection point of the two work distributions P_f and P_b, respectively, for the forward and the backward reaction.

$$\frac{P_f(W)}{P_b(-W)} = e^{\beta(W-\Delta A)}, \tag{2.29}$$

the free energy change is the work W for which the two distributions are equal $P_f(W) = P_b(-W)$ [204]. In other words, that is the point at which both distributions intersect as illustrated in Fig. 2.5.

The combined use of CFT and Gaussian functions for the work distributions is referred to as the Crooks Gaussian intersection (CGI) method. Graphical determination of the intersection point for any distribution shape is usually associated with large statistical errors, in particular if the overlap between $P_f(W)$ and $P_b(-W)$ is small [201]. Therefore, it is common to use Gaussian approximations for the work distributions which exhibit more analytical solutions and more adequate estimations of statistical errors [201, 205].

2.3 Landauer-Büttiker Formalism for Coherent Transport

The conductance G of a macroscopic object is proportional to its cross sectional area A and inversely proportional to its length l:

$$G = \sigma A/l. \tag{2.30}$$

Here, the conductivity σ is a material property which is independent of the object's dimension. The progress in technology made it possible to minimize the dimensions of conducting materials continuously onto the molecular scale. Accordingly, the question arises how small the dimension of a conductor can be before the ohmic behavior in Eq. 2.30 breaks down. Therefore, a deeper insight into the meaning of conductivity on the molecular level needs to be acquired [139]. More importantly, the relation between the electronic structure of a conductor and its transport properties needs to be understood.

For an idealized model, i.e. a small molecule with only one single energy level coupled to two metallic contacts, the conductance is limited by a maximum value G_0. As it turns out, this value is a fundamental entity known as the conductance quantum [140].

$$G_0 \equiv q^2/h = 38.7\mu S = (25.8k\Omega)^{-1} \tag{2.31}$$

Formally, the maximum conductance would be twice as large $2G_0$ considering both degenerate spin states. No larger conductance can be achieved even with perfect contacts.

Based on the formalism of Landauer [206] and Büttiker [207, 208] for coherent electronic transport, the electrical current I in a system comprised of a molecule attached in between a left (L) and a right (R) contact can be calculated as [140]:

$$I = \frac{2e}{h} \int_{-\infty}^{+\infty} T(E) \left[f(E - \mu_L) - f(E - \mu_R) \right] dE. \tag{2.32}$$

Here, $f(E - \mu_L)$ und $f(E - \mu_R)$ are the Fermi functions of both contacts:

$$f(E - \mu_L) = \left(1 + e^{[(E-\mu_L)/k_BT]} \right)^{-1}, \tag{2.33}$$

$$f(E - \mu_R) = \left(1 + e^{[(E-\mu_R)/k_BT]} \right)^{-1}, \tag{2.34}$$

with the chemical potentials μ_L and μ_R, Boltzmann's constant k_B and the temperature T. Fig. 2.6 illustrates a typical model system for which the Landauer-Büttiker formalism can be applied. The whole system is considered to be in thermal equilibrium, that is electrons flow in and out of the contacts, thus the occupation of the molecular energy levels is described via the Fermi functions. The conduction depends on the availability of states around $E = \mu$, thereby, it is irrelevant if these are virtual or occupied states [140]. Current will flow if both potentials μ_L and μ_R differ. The quantity $T(E)$ in Eq. 2.32 is called transmission function as it describes the probability for an electron of a given energy E to pass from one contact to the other by penetrating through the molecule. The quantities $\Sigma_{L/R}$ in Fig. 2.6 are called the self-energy matrices. They account for
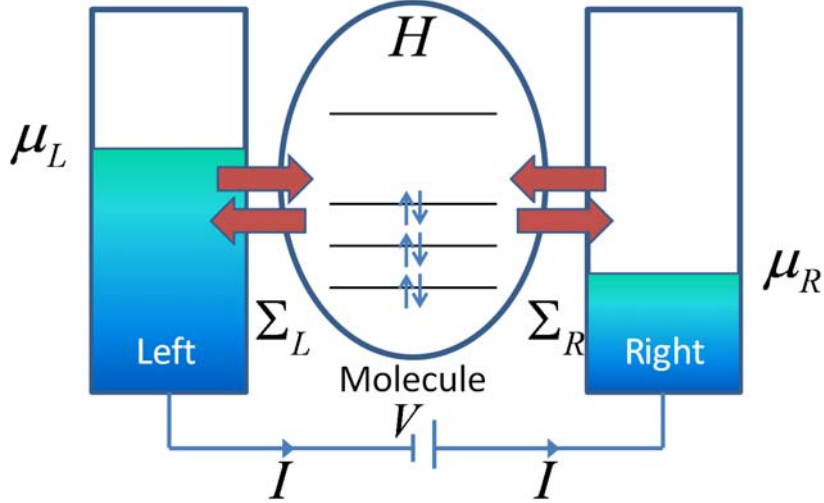
Figure 2.6: A molecule containing occupied and virtual energy levels is coupled to semi-infinite contacts L and R characterized by their chemical potentials μ_L and μ_R. Both the contacts and the molecule are considered to be in equilibrium. The contact-molecule coupling is described by the self-energies Σ_L and Σ_R.

the coupling of the semi-infinite contacts to the molecule [209]. The electronic structure of the isolated molecule is characterized by the Hermitian Hamiltonian H. By contrast, the self-energy matrices Σ_L and Σ_R are not Hermitian. In fact, the anti-Hermitian part of the self-energy is defined as:

$$\Gamma_{L/R} = \mathrm{i} \left[\Sigma_{L/R} - \Sigma_{L/R}^+ \right]. \tag{2.35}$$

2.3.1   Level Broadening

The broadening matrix $\Gamma_{L/R}$ in Eq. 2.35 describes the level broadening of the molecular states once a molecule is attached to a contact. In other words, $\Gamma_{L/R}$ represents the strength of the coupling between the molecule and the contact, thus has a substantial influence on the conductance quality of the molecule [210]. The change of the molecular energy levels due to the molecule-contact coupling is illustrated in Fig. 2.7. The sharp lines for the isolated molecule broaden due to the electronic interaction between the molecule and the contact. This effect can be understood out of the uncertainty principle. An electron injected into the molecule does not reside there forever. The shorter the residence time τ is, the larger the level broadening δE becomes [211]:

$$\delta E \approx \hbar/\tau. \tag{2.36}$$

2.3.2   Non-Equilibrium Green's Functions Formalism

By using the broadening matrix $\Gamma_{L/R}$ and the formalism of non-equilibrium Green's functions (NEGF), the transmission function $T(E)$ can be calculated as the trace over
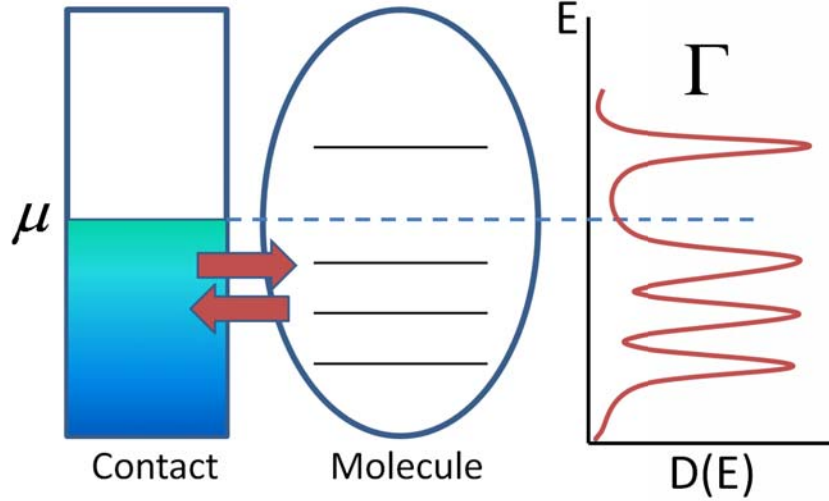
Figure 2.7: Schematic view of the level broadening Γ, induced once a molecule is attached to a contact with the chemical potential μ. $D(E)$ is the density of states.

the matrix product of the Green's functions and the broadening matrices [212–214]:

$$T(E) = \text{Tr}\left[\Gamma_L G^R(E)\Gamma_R G^A(E)\right]. \tag{2.37}$$

The trace in Eq. 2.37 sums over all single-particle levels of the molecule [215]. $G^R(E)$ and $G^A(E)$ are, respectively, the retarded and advanced Green's functions. They describe the dynamics in the conducting media. For isolated molecules represented by the Hamiltonian H these functions can be expressed as [139]:

$$G^R(E) = \left[(E + i\eta)\,I - H\right]^{-1} \qquad (\eta \to 0^+), \tag{2.38}$$

$$G^A(E) = \left[(E - i\eta)\,I - H\right]^{-1} \qquad (\eta \to 0^+). \tag{2.39}$$

The quantity η is an infinitesimal imaginary contribution to the energy E, which is either added or subtracted to incorporate boundary conditions into the differential equations [139]. I is the matrix of identity which is of the same rank as the Hamiltonian matrix. To extend the description to the whole contact-molecule-contact system, e.g. as illustrated in Fig. 2.6, the molecule-contact coupling needs to be included via the self-energy matrices Σ_L und Σ_R [210]:

$$G(E) = [EI - H - \Sigma_L - \Sigma_R]^{-1}. \tag{2.40}$$

The NEGF formalism combined with the Eqs. 2.32 and 2.37 establishes a powerful tool which is widely used to study the transport properties of molecular systems.

2.3.3 Charge Transport in a Linear Chain

The calculation of $T(E)$ can be simplified considerably by reducing the complexity of the molecular electronic structure. A popular approach is to use tight-binding models.
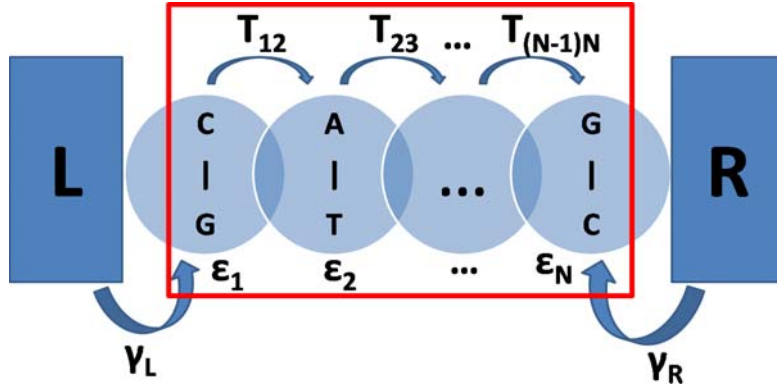
Figure 2.8: Linear chain tight-binding model as applied to study hole transfer in DNA (chapter 4). The model is comprised of N fragments, i.e. the DNA base pairs. The first and the last fragment are coupled to the left and right contact by γ_L and γ_R, respectively. The coarse-grained electronic structure of this model consists of the site energies ϵ_i for each fragment i and the electronic couplings T_{ij} between the two adjacent fragments i and j.

In the simplest case, a linear chain is considered which consists of N fragments where each of which is described by one single energy level only.

$$H = \sum_{i=1}^{N} \epsilon_i c_i^\dagger c_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} T_{ij} \left(c_i^\dagger c_j + \text{h.c.} \right) \tag{2.41}$$

Here, c_i and $c_i^\dagger$ are the operators for the creation and the annihilation of a charge on site i, respectively. The model is characterized by two parameters that turn out to be very useful when studying CT in molecular systems. These parameters are the site energies ϵ_i, i.e. the energy of a charge located on site i, and the electronic coupling T_{ij}, i.e. some kind of probability of a charge going from site i to site j. The calculation of ϵ_i and T_{ij} is described in Sec. 2.5. The electronic couplings are non-zero for two adjacent sites i and j only. Such a model is illustrated in Fig. 2.8 as it will be used to study the transport properties in DNA (chapter 4). Therefore, the model is extended by the coupling to the right and left contact, respectively, by the first and last site of the linear chain. Using the simplification above the linear chain Hückel Hamiltonian has the following matrix form:

$$H = \begin{pmatrix} \epsilon_1 & T_{12} & 0 & 0 & 0 & \cdots & 0 \\ T_{21} & \epsilon_2 & T_{23} & 0 & 0 & \cdots & 0 \\ 0 & T_{32} & \epsilon_3 & T_{34} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & T_{(N-1)N} & \epsilon_N \end{pmatrix} . \tag{2.42}$$

Further approximations are introduced for the contact-molecule coupling using the wide-band approximation. Thus, the ordinarily energy dependent self-energies $\Sigma_{L/R}(E)$ are

replaced by constant values.

$$\Sigma_L = \begin{pmatrix} -i\gamma_L & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \quad \Sigma_R = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & -i\gamma_R \end{pmatrix} \quad (2.43)$$

Note that only one matrix element of the self-energy is non-zero in this simple linear chain model, i.e. $\Sigma_L(1,1) = -i\gamma_L$ and $\Sigma_R(N,N) = -i\gamma_R$ for the left and right contact, respectively. Therefore, the interaction between the molecule and the contacts is described solely by the couplings γ_L and γ_R. Based on the simplifications above and using Eq. 2.35, the general expression from Eq. 2.37 for the transmission function $T(E)$ is simplified to:

$$T(E) = 4\gamma_L\gamma_R \left| G_{1,N}(E) \right|^2 . \quad (2.44)$$

As a consequence, only one matrix elements of the molecular Green's function needs to be computed, which reduces the computational effort immensely.

2.3.3.1 Low Temperature Regime

The Fermi functions $f(E - \mu_L)$ and $f(E - \mu_R)$ in the Eqs. 2.33 and 2.34 can be simplified by assuming the transport to take place at very low temperatures.

$$f(E - \mu_L) = \begin{cases} 1, & \text{if } E > \left(\mu_L - \frac{eU}{2}\right) \\ 0, & \text{otherwise} \end{cases} \qquad f(E - \mu_R) = \begin{cases} 1, & \text{if } E < \left(\mu_R + \frac{eU}{2}\right) \\ 0, & \text{otherwise} \end{cases} \quad (2.45)$$

Here, U is the source-drain voltage. If it is assumed further that $\mu_L = \mu_R = E_F$, where E_F is the Fermi energy, the current for the linear chain model is then obtained as:

$$I(U) = \frac{8e\gamma_L\gamma_R}{h} \int\limits_{E_F-\frac{eU}{2}}^{E_F+\frac{eU}{2}} dE \left| G_{1,N}(E) \right|^2 , \quad (2.46)$$

Note that the I–U characteristics presented in this thesis should be interpreted only qualitatively. The Fermi energy E_F is not computed explicitly, rather E_F is placed arbitrarily as the average of the respective site energies of the system. Therefore, the current could exhibit quite different shapes depending on where E_F is located. However, the width of the current-voltage gap (if existent) is affected by the location of E_F, while the maximum current obtained at high voltages is not altered in this model. The transmission function $T(E)$ and the current $I(U)$ of DNA molecules are computed along molecular dynamics trajectories (chapter 4 and 5). Thus, both the structural fluctuations as well as the electrostatic effects induced by the MM environment (i.e. solvent, counterions and DNA backbone) are taken into account. Eventually, the time series of $T(E)$ and $I(U)$ need to be analyzed statistically as well.

Besides the simplifications introduced above, a central issue in this coherent transport model, is the separability of time scales for the ionic and the electronic motion. Strictly speaking the Landauer formalism of coherent transport, as described here, is valid in the adiabatic regime only. All the calculations presented in chapter 4 and 5 are based on the assumption that the transport of the charge through the molecules is faster than its dominant vibrational modes. This issue will be addressed in more detail in Sec. 4.5.

2.4   Molecular Dynamics Simulations

The aim of molecular modeling is the description of complex matter on a realistic atomistic level. Therefore, accurate predictions of system properties are needed in order to design new materials. The solution of the relativistic time-dependent Schrödinger equation would provide all necessary information of a given system very accurately, though only a few atoms can be handled on this *ab initio* level, according to present state of the art in computer technology. Accordingly, approximations have to be applied in order to describe large systems with many thousands of atoms, such as solid states or biomolecules, e.g. proteins. This also holds for dynamical processes on large timescales as they occur in molecular biology, e.g. electron and proton transfers or global conformational transitions in DNA and proteins. At the end, the larger the system or timescale the more severe the approximations have to be. In classical molecular dynamics (MD) simulations, the *ab initio* approach is replaced by empirical parametrization.

Why molecular dynamics?   On a macroscopic scale, only ensemble-averaged properties, such as binding energies, solubilities or relative stability of molecular conformations are relevant. These are obtained by averaging over the representative statistical ensemble of structures. Thus, a molecular description based on single structures is inadequate. The proper statistical equilibrium ensemble of conformations can be generated either by Monte Carlo [216] or by MD simulations. The latter will be used in this work and is described briefly in the following.

The foundation of MD lies in solving the Newtonian equations of motion [217]. In a system of N particles, the force on each atom i is obtained:

$$\vec{F}_i = m_i \frac{\partial^2 \vec{x}_i}{\partial t^2}. \tag{2.47}$$

Note the vector notation for the forces $\vec{F}$ and the positions $\vec{x}$, i.e. in three dimensions the total system is described by $3N$ coordinates. The forces are obtained as negative derivatives of the potential V, which describes the interaction between the particles.

$$\vec{F}_i = -\frac{\partial V}{\partial \vec{x}_i} \tag{2.48}$$

Both equations have to be solved simultaneously in small time steps Δt. As a result, a trajectory is obtained which contains the particle's coordinates (optionally also the
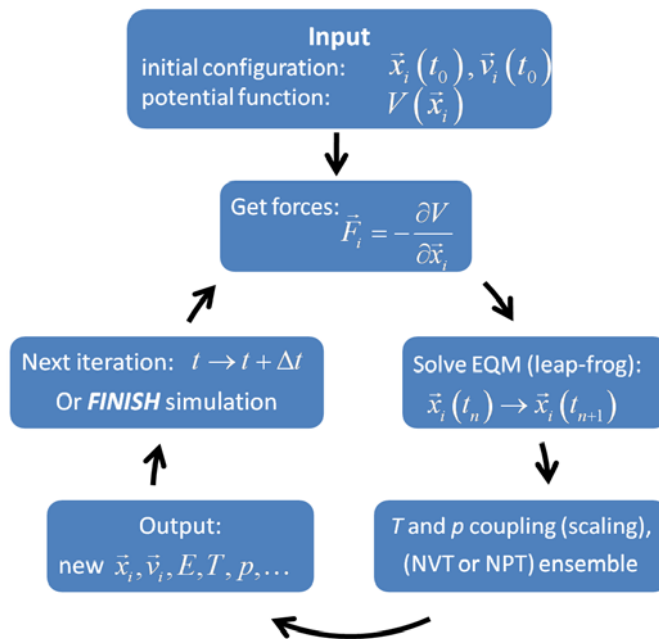
Figure 2.9: Basic MD algorithm

velocities) as a function of time. Macroscopic properties can be computed from the trajectory once the system reaches an equilibrium state. A simplified basic MD algorithm is illustrated in Fig. 2.9, its single steps are described in more detail below.

At this point it is important to be aware of the limitations of MD. In summary, some of those are [218]: i) The simulations are classical, i.e. quantum effects like proton tunneling are not accounted for. ii) Electrons are in the ground state always instantly following the atomic movements according to the Born-Oppenheimer (BO) approximations [219]. In turn, this means electrons transfer, electronic excitations and chemical reaction (e.g. bond breaking) cannot be treated. iii) Forcefields describe the interaction between particles only approximately. Usually, effects like polarization are not accounted for as well. iv) Unnatural boundaries, that is in simulations of bulk systems periodic boundary conditions (PBC) are often applied in order to avoid unwanted boundary of the particle cluster with its environment. However, this is somehow artificial when dealing with liquids or molecules in solution. Nevertheless, the error due to PBC is said to be rather small for large systems [218].

2.4.1 Basic Algorithms

The equations of motion have to be solved numerically. The default procedure in GRO-MACS is the leap-frog algorithm [220], which uses positions $\vec{x}$ at time t and velocities $\vec{v}$ at time $t - \frac{1}{2}\Delta t$:

$$\vec{v}\left(t + \frac{1}{2}\Delta t\right) = \vec{v}\left(t - \frac{1}{2}\Delta t\right) + \frac{F(t)}{m}\Delta t, \tag{2.49}$$

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}\left(t + \frac{1}{2}\Delta t\right)\Delta t. \tag{2.50}$$

Elimination of the velocities leads to [218]:

$$\vec{x}\left(t + \Delta t\right) = 2\vec{x}(t) - \vec{x}(t - \Delta t) + \frac{F(t)}{m}\Delta t^2, \tag{2.51}$$

thus producing trajectories identical to those of the Verlet algorithm [221]. A very similar alternative to leap-frog is the Velocity-Verlet algorithm [222], in which positions $\vec{x}$ and velocities $\vec{v}$ are used at the same time t. Note the equations of motions have to be modified further for the inclusion of temperature and pressure coupling as well as the conservation of structural constraints, as described in more detail below.

MD simulations without couplings to heat and pressure baths generate the so-called NVE ensemble (constant particles, volume and total energy, respectively), also known as the micro-canonical ensemble. Naturally, most physical quantities should be calculated from either the NVT ensemble, i.e. the canonical ensemble, in which the temperature is conserved as well, or the NpT (isothermal-isobaric) ensemble, in which both pressure and temperature are conserved. The latter realization appears to be most closely to laboratory conditions. Therefore, thermostat and barometer have to be introduced representing respectively, temperature and pressure coupling.

2.4.1.1   Temperature Coupling

A weak coupling to an external heat bath can be realized with the Berendsen thermostat [223]. The deviation of the system temperature T from a given reference temperature T_0 is slowly corrected by first order kinetics:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}. \tag{2.52}$$

Thus, the temperature deviation decays exponentially with the time constant τ. The strength of the coupling can be adapted, i.e. for system equilibration τ is taken to be rather short (e.g. 0.01 ps), while for simulations already in equilibrium τ can be much larger (e.g. 0.5 ps) [218]. Technically, the temperature is supposed to change:

$$\Delta T = \frac{\Delta t}{\tau}\left(T_0 - T\right), \tag{2.53}$$

which is realized by scaling the particles' velocities $\vec{v}_i$ by a factor ς. The temperature of a system of N particles is directly related to the velocities of the particles via the kinetic energy E_{kin}:

$$E_{kin} = \frac{3}{2}Nk_BT = \sum_i^N \frac{1}{2}m_i\vec{v}_i^2, \tag{2.54}$$

with k_B being the Boltzmann constant. Accordingly, both temperatures T and T_0, and therefore also $\Delta T = T_0 - T$, can be expressed as:

$$
\begin{aligned}
T &= \frac{1}{3Nk_BT} \sum_i^N m_i \vec{v}_i^2, \\
T_0 &= \frac{\varsigma^2}{3Nk_BT} \sum_i^N m_i \vec{v}_i^2, \\
\Delta T &= \frac{1}{3Nk_BT} \sum_i^N m_i \vec{v}_i^2 \left(\varsigma^2 + 1\right) = T\left(\varsigma^2 + 1\right).
\end{aligned} \tag{2.55}
$$

Combining this result with Eq. 2.53 the scaling factor ς can be calculated:

$$
\varsigma = \sqrt{\frac{\Delta t}{\tau}\left(\frac{T_0}{T} - 1\right) + 1}. \tag{2.56}
$$

Nevertheless, the Berendsen thermostat does not generate a proper canonical ensemble, since it suppresses the fluctuations of the kinetic energy E_{kin} of the system. The error scales with $1/N$, hence most ensemble properties are not significantly affected for large systems[218].

A thermostat which produces the correct canonical ensemble was developed first by Nosé [224] and later modified by Hoover [225], the so-called Nosé-Hoover temperature coupling. Here, a thermal reservoir and friction term are introduced in the equations of motion [218]. Both temperature coupling algorithms, Berendsen and Nosé-Hoover, are implemented in GROMACS [198].

In this work, the weak coupling algorithm of Berendsen is used for the equilibration process, i.e. relaxing the system to the target temperature T_0, while the Nosé-Hoover thermostat is used for the final production run, i.e. the simulation in equilibrium, generating a proper canonical ensemble.

2.4.1.2 Pressure Coupling

Similarly to the temperature coupling, the system can be coupled to a pressure bath by rescaling the coordinates of the particles and the dimensions of the simulation box. Again Berendsen [223] provided an algorithm which relaxes the pressure p of a system to a given reference pressure p_0 by first-order kinetics:

$$
\frac{dp}{dt} = \frac{p_0 - p}{\tau_p}. \tag{2.57}
$$

However, this scheme shares the drawback of the Berendsen Thermostat, i.e. it does not generate the correct NpT ensemble, although it yields simulations with the correct average pressure [218]. The true NpT ensemble can be generated with the Parinello-Rhaman approach [226, 227], which is similar to the Nosé-Hoover thermostat. The Parinello-Rhaman approach is used throughout this work for all constant-pressure simulations.
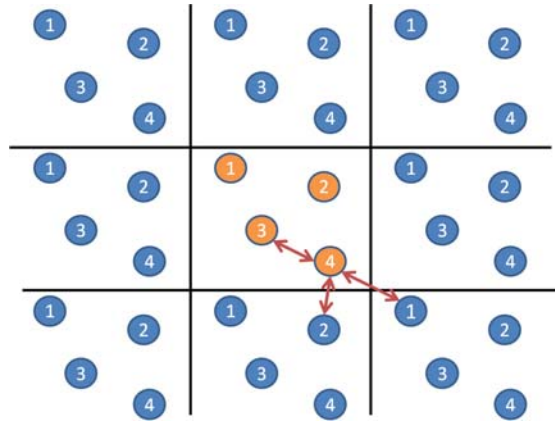
Figure 2.10: Periodic boundary conditions in two dimensions.

2.4.1.3   Constraint Algorithm

For large complex system it is sometimes not affordable, maybe not even necessary to propagate all degrees of freedom in the system. Therefore, constraints can be applied in the simulation, which keep e.g. bonds, angles, etc. at their predefined values. There are two major algorithms which can account for constraints, the SHAKE algorithm [228] and the LINCS algorithm [229], the latter of which is used in this work to keep all hydrogen bonds fixed.

2.4.1.4   Periodic Boundary Conditions

In order to avoid edge effects in the simulations of finite systems usually periodic boundary conditions (PBC) are applied. With PBC the simulations box (unit cell) containing the N atoms is surrounded by an infinite number of copies of itself (images) in all dimensions. In Fig. 2.10 this is illustrated for the 2-dimensional case. Atoms that leave a box enter the adjacent box on the opposite site, thus the number of particles is always conserved. Interactions are treated following the so-called minimum image convention, that is each particle interacts only with the closest images of the remaining $N-1$ atoms. To account for the minimum image convention in a rectangular box the cut-off radius, i.e. maximum distance between interacting particles, has to be less or equal than half the box size. In this work all simulations are carried out using rectangular boxes.

2.4.1.5   Parallelization

Even with the rapid development in computer technology, the systems and length-scales that can currently be studied with MD are limited to about 10^6 atoms and timescales of ns to μs, respectively. Fortunately, the CPU time required for a simulation can be considerably reduced by running simulations in parallel. Nevertheless, linear scaling, i.e. simulations on N CPUs is N times faster, can only be achieved for a small number of processors [218]. The success of parallelization depends on the algorithms that are used and on the system under study itself. Technically, there are two types of parallelization:
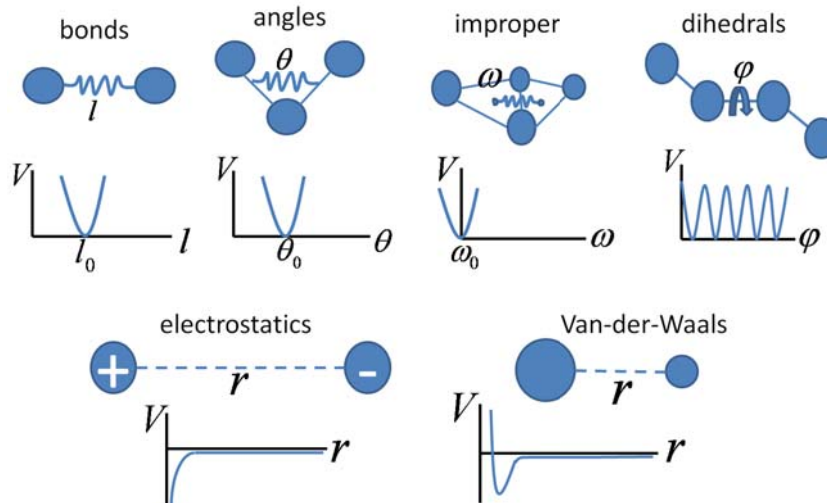
Figure 2.11: Illustration for bonded (top) and non-bonded (bottom) force field terms. The respective potential functions V are depicted qualitatively as well. Bonded interactions are indicated by solid lines (or springs), while dashed lines indicate non-bonded interactions.

particle decomposition and domain decomposition. The latter is the default algorithm in GROMACS, which is faster due to an improved scaling algorithm [198]. In this work, domain decompositions is used for the purely classical simulations, while particle decomposition is used for the QM/MM simulation in chapter 8, due to technical reasons.

2.4.2 Forcefields

As mentioned above, molecular forcefields are purely classical, that is atoms are regarded as spheres carrying partial charges q_i, and bonds between atoms are represented by harmonic springs. Moreover, the forcefield is a function of the nuclei only, i.e. electrons are neglected. Typically, the analytical potential function V, which accounts for all the interaction between the atoms, is decomposed into bonded and non-bonded terms.

$$V\left(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N\right) = V_{\text{bonded}} + V_{\text{non-bonded}} \tag{2.58}$$

The contributions to V_{bonded} are:

$$V_{\text{bonded}} = \sum_{\text{bonds}} K_l \left(l - l_0\right)^2 \;+\; \sum_{\text{angles}} K_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{improper}} K_\omega \left(\omega - \omega_0\right)^2$$
$$+ \sum_{\text{dihedrals}} K_\varphi \left(1 + \cos\left(n\varphi - \varphi_0\right)\right). \tag{2.59}$$

Here, bonding, angle and improper (out-of-plane motion) interactions are described by harmonic potentials based on Hooke's law (e.g. springs between atoms), while for the dihedral potential a periodic function is used. The bonded interactions and their potential functions V are illustrated qualitatively in Fig. 2.11(top).

The non-bonded interactions are usually decomposed into an electrostatic Coulomb part V_{Coulomb} and a part which accounts for both the attractive Van-der-Waals interactions (dispersion) and the Pauli repulsion. The latter part can be combined in either the

widely used Lennard-Jones (LJ) potential or the Buckingham potential. In this work the LJ potential is used, though the Buckingham potential describes the Pauli repulsion more realistically it is much harder to compute [218]. Thus, $V_{\text{non-bonded}}$ can be expressed as:

$$
\begin{aligned}
V_{\text{non-bonded}} &= V_{\text{Coulomb}} + V_{\text{LJ}} & (2.60)\\
V_{\text{Coulomb}} &= \sum_i \sum_{j>i} \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_i q_j}{r_{ij}}\\
V_{\text{LJ}} &= \sum_i \sum_{j>i} 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right)
\end{aligned}
$$

with ϵ_0 and ϵ_r being, respectively, the dielectric constant *in vacuo* and the relative dielectric constant of the medium itself (e.g. 78 in water). The Coulomb and LJ interactions are shown in Fig. 2.11(bottom).

All the parameters given in Eq. 2.59 and 2.60 have to be evaluated using either quantum chemical calculations or, if available, experimental data. The main drawback of such a potential function is the low transferability, that is parameters may have to be adjusted when dealing with different molecular systems.

2.4.2.1 Exclusions and 1-4 Interactions

Atoms that are covalently bound, or linked by one or two atoms are denoted as first, second and third neighbors, respectively. The interactions of atom i with $i+1$ and $i+2$ cannot be modeled by the non-bonded LJ potential, which cannot account for quantum mechanical effects. These are certainly important for atoms that are very close (e.g. covalent bonds). However, these 1-2 and 1-3 interactions are assumed to be already accounted for via the harmonic bonded potentials in Eq. 2.59, i.e. the 1-2 bond term and the 1-3 angle term. Thus, the atoms $i+1$ and $i+2$ are excluded from LJ interaction on atom i.

Even 1-4 (third neighbor) interactions are sometimes too strong, i.e. the repulsive part of the LJ potential may cause deformation or even breaking of the molecule [218]. In practice, 1-4 interactions, sometimes both LJ and Coulomb interactions, are scaled down by an empirical factor.

2.4.2.2 Long Range Electrostatics

As mentioned in the PBC section above, a cut-off can be used when treating non-bonded interactions. Predominantly, this is applied only to LJ interactions, since the LJ potential rapidly vanishes with the distance by $1/r^6$. On the contrary, this is not true for the Coulomb potential which has significant contributions also at large distances, since it falls off with $1/r$. Therefore, a cut-off for long-range Coulomb interactions is not suitable.

With PBC the electrostatic energy of a system with N particles and their periodic

images is [218]:

$$V_{\text{Coulomb}} = \frac{1}{8\pi\epsilon_0\epsilon_r} \sum_{n_x} \sum_{n_y} \sum_{n_z^*} \sum_i^N \sum_j^N \frac{q_i q_j}{r_{ij,\vec{n}}}, \tag{2.61}$$

with the box vector $\vec{n} = (n_x, n_y, n_z)$ and the real distance (not the minimum image) $r_{ij,\vec{n}}$. Note, the star indicates that terms with $i = j$ are omitted if $\vec{n} = \vec{0}$. Unfortunately, the sum in Eq. 2.61 converges very slow. Ewald [230] provided the idea to subdivide the slowly-converging sum into two fast-converging sums plus a constant term [218]:

$$V_{\text{Coulomb}} = V_{\text{Coulomb}}^{\text{direct}} + V_{\text{Coulomb}}^{\text{reciprocal}} + V_{\text{Coulomb}}^0. \tag{2.62}$$

However, the reciprocal sum turns out to be the bottleneck, i.e. its computational cost scales with N^2, thus are hardly affordable for large system [218]. The particle-mesh Ewald (PME) method [231, 232] greatly improves the handling of the reciprocal, hence is widely used for the treatment of long-range electrostatic interaction in MD simulations of large systems.

2.4.2.3 Force Fields for the Simulations of Proteins and DNA

In this work the Amber force-field parm99SB[233] is used for the simulation of Photolyase in part II, where SB denotes the correction for improved protein backbone parameters [234]. The simulations of nucleic acids (chapter 4 and 5) are carried out using the Amber parm99BSC0 forcefield. The BSC0 correction [235] accounts for the refinement of the standard parm99 forcefield for nucleic acids. In either case, the molecules are placed in a rectangular box filled with solvent of the TIP3P [236] water model. Optionally, counterions are added in the simulations of nucleic acids for neutralization only. In part II, simulations are carried out using charged amino acid sidechains and/or co-factors, respectively. The atomic charges of these charged species are derived using the restrained electrostatic potential (RESP) procedure [237]. Technically, charges for both neutral and charged species are computed and the difference is then added to the standard atomic charges in the forcefield topology.

2.4.3 MD Observables

Many quantities can be calculated from the time series of atomic positions and velocities. Besides thermodynamic properties, a systematic structural characterization of the system under study is of considerable interest as well. In the following three structural measures are described.

Root mean square deviation (RMSD) The RMSD measures the deviation of a molecular structure to a given reference structure, e.g. to a x-ray crystal structure, in time. Usually, translational and rotational contributions to the RMSD are avoided by

prior least-square fitting of the molecular structures to the reference structure.

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\vec{x}_i\,(t) - \vec{x}_i^{\,0} \right)^2} \tag{2.63}$$

Besides the energy time series, also the RMSD time series can be used to decide whether a simulation has reached an equilibrium. Moreover, RMSD time series, which show alternating plateaus, may indicate different local or global conformations.

Root mean square fluctuation (RMSF)   The value describes the deviation (fluctuation) of atomic positions to (around) a given reference structure. For the atom i the RMSF is calculated via:

$$\mathrm{RMSF}_i = \sqrt{\frac{1}{t_n} \sum_{j=1}^{n} \left(\vec{x}_i\,(t_j) - \vec{x}_i^{\,0} \right)^2}, \tag{2.64}$$

with n being the number of time steps and t_n the total simulation time. The RMSF is closely related to the Debye-Waller factor [238], B_i, which is well known from x-ray measurements, via the relation [239]:

$$\mathrm{RMSF}_i = \sqrt{\frac{3B_i}{8\pi^2}} \tag{2.65}$$

Using the RMSF value, more or less flexible (thermally active) parts of a system can be assigned, which then can be compared (qualitatively) to x-ray experimental data. Moreover, large RMSF values may be an indicator of conformational changes, e.g. closing and opening of binding pockets in proteins.

Radial pair distribution function (RDF)   The RDF, also known as pair correlation function g_{AB}, between atoms of type A and B, is a measure for finding an atom $i \in A$ at a radial distance r_{ij} of atom $j \in B$.

$$g_{AB}(r) = \frac{1}{N_A N_B} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \langle \delta\left(|r_{ij}| - r\right) \rangle \tag{2.66}$$

In bio-molecular simulations, the RDF is used to analyze the solvent arrangement in the vicinity of the solute, e.g. the protein or the DNA.
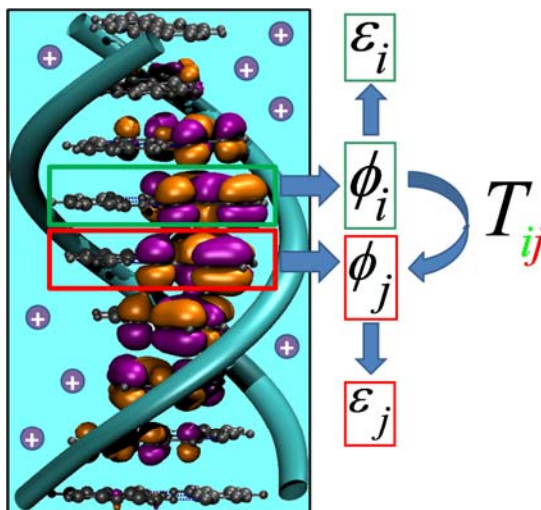
Figure 2.12: Illustration of the fragment orbital (FO) approach as applied to hole transfer in DNA. Color code: nucleobases are indicated in gray, DNA backbone, i.e. sugar and phosphate groups, as well as the solvent (within the box) in light blue, counterions carry a positive charge, and finally the molecular orbitals on the respective nucleobase pairs in orange/purple. Two individual fragments i (green) and j (red) are highlighted. The CT coupling integrals are calculated between the respective FOs i and j using a QM/MM formalism, that is only the nucleobases between which T_{ij} is computed will be treated quantum mechanically. However, the computation of both ϵ_i and T_{ij} fully includes all the MM charges of the remainder of the system (i.e. solvent, backbone, counterions and also the other nucleobases), though only electrostatically via Eq. 2.126.

2.5 Model Parameters from Quantum Chemistry

In this section the calculation of CT parameters, which can only be deduced from quantum chemical methods, is discussed. First, the coarse-grained fragment orbital (FO) approach is introduced. Subsequently, the quantum chemical methods, HF, DFT and DFTB, used in this work are described briefly.

2.5.1 The Fragment-Orbital Approach

The FO approach allows for a direct and well controlled coarse graining of the electronic problem. Since the hole transfer, e.g. in DNA occurs between the HOMO's ϕ_i and ϕ_j of individual DNA bases i and j, the electronic Hamiltonian of the system can be formulated in terms of the site energies of those nucleobases (representing the diagonal elements: $T_{ii} = \epsilon_i$) and the CT (coupling) integrals between the sites i and j (giving the off-diagonal elements T_{ij}) [144]. The orbitals ϕ_i can be obtained by treating only fragments of the whole system, therefore called fragment orbitals, for illustration see Fig. 2.12.

In case of DNA, these fragments can be chosen as isolated bases or Watson-Crick base pairs. Using the Kohn-Sham Hamiltonian $\hat{H}_{\mathrm{KS}}$, the diagonal terms $T_{ii} = \epsilon_i$ can be

calculated as:

$$\epsilon_i = - \left\langle \phi_i \left| \hat{H}_{\mathrm{KS}} \right| \phi_i \right\rangle, \tag{2.67}$$

while the coupling integrals are evaluated as:

$$T_{ij}^0 = \left\langle \phi_i \left| \hat{H}_{\mathrm{KS}} \right| \phi_j \right\rangle. \tag{2.68}$$

Introducing an atomic-basis-set expansion

$$\phi_i = \sum_\mu c_\mu^i \eta_\mu \tag{2.69}$$

both coupling and overlap integrals in the molecular-orbital (MO) basis can be evaluated efficiently as:

$$T_{ij}^0 = \sum_{\mu\nu} c_\mu^i c_\nu^j \left\langle \eta_\mu \left| \hat{H}_{\mathrm{KS}} \right| \eta_\nu \right\rangle = \sum_{\mu\nu} c_\mu^i c_\nu^j H_{\mu\nu}, \tag{2.70}$$

and

$$S_{ij} = \sum_{\mu\nu} c_\mu^i c_\nu^j \left\langle \eta_\mu | \eta_\nu \right\rangle = \sum_{\mu\nu} c_\mu^i c_\nu^j S_{\mu\nu}. \tag{2.71}$$

$H_{\mu\nu}$ and $S_{\mu\nu}$ are the Hamilton and overlap matrices represented in the atomic basis set. These matrices as well as the atomic coefficients c_μ^i can be determined by quantum chemical calculations, e.g. using HF or DFT. In this work, the SCC-DFTB method is used as described in Sec 2.5.4.

The matrix T_{ij}^0 is built from non-orthogonal orbitals ϕ_i and ϕ_j. For many problems, a representation in an orthogonal basis set is more suitable, which can be achieved using the Löwdin transformation:[240]

$$\mathbf{T} = \mathbf{S}^{-1/2} \mathbf{T}^0 \mathbf{S}^{-1/2} \tag{2.72}$$

In the following, T_{ij} will denote the matrix based on the orthonormal basis, while T_{ij}^0 is the matrix in non-orthonormal representation. The T_{ij} can be identified with CT integrals computed e.g. using HF-KTA.[241]

2.5.2 Hartree-Fock

To begin with, consider a system of N non-interacting electrons. In this so-called Hartree picture the electrons do not "see" each other, that is there is neither electron exchange nor correlation. The total wave function of the system can be written as a product of the single particle wave functions ϕ_i [242]

$$\Psi(r_1, r_2, \ldots, r_N) = \phi_1(r_1) \phi_2(r_2) \ldots \phi_N(r_N) = \prod_i^N \phi_i(r_i). \tag{2.73}$$

In the Hartree ansatz, each electron i is affected by a separate potential that is given by the single particle Hamiltonian $\hat{h}_i(r_i)$ operating on the single particle wave function. Therefore, the N electron problem can be decomposed into N single electron problems:

$$\hat{h}_i(r_i) \phi(r_i) = \epsilon_i \phi_i(r_i). \tag{2.74}$$

Applying the Variational principle leads to

$$\hat{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_{\alpha}^{M} \frac{Z_\alpha}{r_{i\alpha}} + \sum_{i\neq j} \int \frac{\rho_j}{r_{ij}} dr \qquad (2.75)$$

with

$$\rho_j(r) = \phi_j^* \phi_j = |\phi_j|^2,$$

being the charge density of electron j. This approximation is known as the Hartree approximation in which the electron spin is completely neglected. Therefore, the Pauli principle is not accounted for, that is two electrons with the same spin may get arbitrarily close, i.e. occupy the same quantum states. To account for the Pauli principle the wave function has to be antisymmetric. This is accomplished in the Hartree-Fock (HF) theory. Introducing spin orbitals $\chi(i)$, which are products of the Hartree single particle wave function $\phi(i)$ and the spin wave functions $\alpha(i), \beta(i)$, the Pauli principle can be ensured elegantly by usage of the Slater determinant:

$$\Psi(r_1, \ldots r_N) = \frac{1}{\sqrt{N!}} \det \begin{pmatrix} \chi_1(1) & \chi_2(1) & \ldots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \ldots & \chi_N(2) \\ \ldots & \ldots & \ldots & \ldots \\ \chi_1(N) & \chi_2(N) & \ldots & \chi_N(N) \end{pmatrix} \qquad (2.76)$$

with $\chi(i) = \phi(i)\alpha(i)$ or $\chi(i) = \phi(i)\beta(i)$.

The total energy in HF can be written as

$$E = \sum_i h_i + \frac{1}{2} \sum_{ij} (J_{ij} - K_{ij}) + V_{nn} \qquad (2.77)$$

with J_{ij} being the Coulomb energy between electrons i and j, K_{ij} the electron exchange energy and V_{nn} the core-core repulsion which is often described classically. The second term in Eq. 2.77 vanishes for $i = j$, thus the well known DFT problem of electron self interaction is not present in HF. The total energy can be minimized via the Lagrange formalism. However, like the Hartree ansatz, HF is a single particle theory as well. The electrons feel an average potential which is built by all other electrons[243].

Interestingly, the sum of the orbital energies ϵ_i is not equal to the HF total energy:

$$\sum_i \epsilon_i = \sum_i \left\langle \chi_i \middle| \hat{F}_j \middle| \chi_i \right\rangle = \sum_i \left\langle \chi_i \middle| \hat{h}_i \middle| \chi_i \right\rangle$$
$$+ \sum_{ij} \left(\langle \chi_i | J_j(i) | \chi_j \rangle - \langle \chi_i | K_j(i) | \chi_j \rangle \right) \qquad (2.78)$$

The difference amounts exactly to $\frac{1}{2}\sum_{ij}(J_{ij} - K_{ij}) + V_{nn}$, since the electron-electron interaction is counted twice by the Fock operator [196]. The Lagrange multipliers ϵ_i have a clear physical meaning as the orbital energies of the system.

In the following, a system of N electrons is considered in which an electron from orbital k is removed. Further, it is assumed that the removal of an electron will not

change the molecular orbitals, i.e. Frozen Orbital Approximation [244]:

$$E_N = \sum_i^N h_i + \frac{1}{2} \sum_i^N \sum_j^N (J_{ij} - K_{ij}) + V_{nn}$$

$$E_{N-1}^k = \sum_i^{N-1} h_i + \frac{1}{2} \sum_i^{N-1} \sum_j^{N-1} (J_{ij} - K_{ij}) + V_{nn}$$

$$E_N - E_{N-1}^k = h_k + \frac{1}{2} \sum_i^N (J_{ik} - K_{ik}) + \frac{1}{2} \sum_j^N (J_{kj} - K_{kj}).$$

The last two terms are identical, thus one can write

$$E_N - E_{N-1}^k = h_k + \sum_i^N (J_{ki} - K_{ki}) = \epsilon_k. \tag{2.79}$$

As a consequence, within the Frozen MO approximation, the Fock eigenvalues ϵ_k can be regarded as ionization potentials or electron affinities, respectively. This result is well known as Koopmans' Theorem [245].

2.5.3 DFT

The HF wave function $\Psi(r_1, \ldots r_N)$ contains $3N$ coordinates, which represents an information overkill, e.g. only 6 coordinates r and r' are needed for the calculation of 2-electron integrals. Moreover, the wave function is no physical observable. On the other hand, the electron density $\rho(r_i)$, which is defined as the probability of finding an electron at place r, can be measured experimentally. To get $\rho(r_1)$ we have to integrate over all other electrons and their positions

$$\rho(r_1) = N \int \Psi^*(r_1, \ldots r_N) \Psi(r_1, \ldots r_N) \, dx_2 \ldots dx_N. \tag{2.80}$$

The total number of electrons N is obtained by integration of the electron density over the whole coordinate space

$$\int \rho(r) \, dV = N. \tag{2.81}$$

In 1964, Hohenberg and Kohn gave the proof that the properties of an electronic system can be determined solely and uniquely by the total number of electrons N and the external potential $v_{ext}(r)$ [246]

$$v_{ext}(r) = \sum_\alpha \int \frac{Z_\alpha}{R_\alpha - r} dr. \tag{2.82}$$

They showed that there are no two external potentials $v_{ext}(r)$ which lead to the same electron density $\rho(r)$, a result which is known as the first Hohenberg-Kohn (HK) theorem. Furthermore, the second HK theorem ensures:

$$E_0 \leq E[\tilde{\rho}]. \tag{2.83}$$

Therefore, any density which is not the true ground state density causes a higher energy than that of the ground state [246].

The functionals of the electron-electron and electron-core interaction are known and correspond to those of the Hartree theory

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(r)\,\rho(r')}{r - r'} dr\,dr' \tag{2.84}$$

$$E_{en}[\rho] = \sum_\alpha \int \frac{Z_\alpha \rho(r)}{R_\alpha - r} dr. \tag{2.85}$$

The challenge within DFT is the description of the so-called Pauli and Coulomb hole via the hole function $h_{xc}(r_1, r_2)$, also known as exchange-correlation (XC) hole. The function represents a negative electron density which diminishes the density of all the other electrons around r_1. This is expressed quantitatively by the sum rule:

$$\int h_{xc}(r_1, r_2) dr_2 = -1. \tag{2.86}$$

Integration over the hole function, thus, eliminates exactly the charge of one electron. Therefore, in addition to the classical part of the electron-electron interaction in Eq. 2.84, the interaction of the electron density with the XC hole has to be included:

$$E_{ee}[\rho] = \frac{1}{2} \iint \frac{\rho(r)\,\rho(r')}{r - r'} dr\,dr' + \frac{1}{2} \iint \frac{\rho(r)\,h_{xc}(r, r')}{r - r'} dr\,dr'. \tag{2.87}$$

The second therm in Eq. 2.87 is known as exchange-correlation energy E_{xc} compensating the overestimation of $J[\rho]$. Now the total DFT energy reads:

$$E[\rho] = T_s[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho]. \tag{2.88}$$

Starting from the Hartree formalism, Kohn and Sham formulated an exact theory [247], which provides both the exact electron density of the ground state ρ_0 as well as the corresponding ground state energy E_0. They applied the variatonal principle to the DFT total energy under the condition $\int \rho(r) dr = N$:

$$\delta E[\rho] - \delta \left(\int \rho(r) dr - N \right) \mu = 0.$$

The results is

$$v_{eff}(r) + \frac{\delta T_s[\rho]}{\delta \rho} = \mu, \tag{2.89}$$

with the effective Potential $v_{eff}(r)$ being the sum over the functional derivatives of the separate energy terms with respect to the density:

$$v_{eff} = \frac{\delta E_{ne}[\rho]}{\delta \rho} + \frac{\delta J[\rho]}{\delta \rho} + \frac{\delta E_{xc}[\rho]}{\delta \rho}$$

or respectively, the potentials

$$v_{eff} = v_{ext}[\rho] + \int \frac{\rho(r')}{r - r'} dr' + v_{xc}[\rho]. \tag{2.90}$$

With the effective potential $v_{eff}(r)$, which now also contains the XC potential $v_{xc}[\rho]$, the well-known Kohn-Sham equation can be constructed: [247]

$$\left[-\frac{1}{2}\nabla_i^2 + v_{eff}\right]\phi_i = \epsilon_i\phi_i. \tag{2.91}$$

It is important to stress that the kinetic energy term $T_s[\rho]$ in Eq. 2.89 is exact only for a system of non-interacting particles. Therefore, in order to obtain the true ground state density for the correlated system v_{eff} has to be chosen accordingly. The solution of the Kohn-Sham equations is carried out self-consistently, that is the procedure is initiated with a guess for the density, from which then Kohn-Sham orbitals are obtained via Eq. 2.91. In turn, these are used to build up an improved density. The cycle is repeated until the change in density between succeeding iterations is marginal.

Like in HF, the sum of Kohn-Sham energy eigenvalues is not identical to the DFT total energy. Again the Coulomb interaction is counted twice:

$$E[\rho] = \sum_i^{occ} \epsilon_i - \frac{1}{2}\iint\frac{\rho(r)\rho(r')}{r-r'}drdr' + E_{xc}[\rho] - \int v_{xc}(r)\rho(r)dr + V_{nn}. \tag{2.92}$$

The big challenge in DFT lies in the choice for the XC functional, i.e. the self interaction error would vanish using an exact functional. Nevertheless, there is no such thing like an exact XC functional, at least until the present day. The approximations used to describe E_{xc} lead to various serious problems within DFT, as is discussed in Sec. 2.5.3.1. The single particle energies ϵ_i in Eq. 2.91 can be associated with molecular energy levels. Moreover, if E_{xc} were exact, the energy of the HOMO would be the molecule's ionization energy as in HF. This is a result stemming from a general concept known as Janak's theorem [248]:

$$\frac{\partial E}{\partial n_i} = \epsilon_i, \tag{2.93}$$

which represents an analogue to Koopmans' theorem in HF. Note, there is no clear physical meaning for the other Kohn-Sham eigenvalues [249].

Besides the condition in Eq. 2.86 there are other mathematically exact conditions which should be fulfilled for the hole function. The local density approximation (LDA) satisfies some of these conditions. LDA is based on the model of a homogeneous electron gas in which the electron density is constant within the volume V:

$$E_{xc}[\rho] = \int \epsilon_{xc}[\rho]\rho dV. \tag{2.94}$$

Here, ϵ_{xc} is the XC energy per particle, which can further be decomposed into exchange and corrlation part:

$$\epsilon_{xc} = \epsilon_x + \epsilon_c. \tag{2.95}$$

The exchange part ϵ_x was already determined by Bloch and Dirac in the late 1920s [250]:

$$\epsilon_x = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3}(r). \tag{2.96}$$

By contrast, it is far more difficult to obtain an analytical expression for ϵ_c. There are solutions for limiting cases, i.e. very large and rather small electron densities, between which one interpolates. Nevertheless, LDA performs quite well in describing molecule geometries and vibrations for systems with approximately low or no density fluctuations at all. On the other hand, it fails in predicting reaction enthalpies as well as activation barriers. Since ϵ_{xc} depends on the density ρ at place r only, it is said to be a local approximation. Obviously E_{xc} depends on the density at r' as well, due to the $h_{xc}(r, r')$ term in Eq. 2.87. At the end, the XC functional is needed to be non-local. In the generalized gradient approximation (GGA) also the gradient of the density is considered via a Taylor series. This results in an improved description of systems with moderate density fluctuations compared to LDA, but GGA functionals are still not non-local. Based on LDA and GGA many functionals have been developed in the past decades. One prominent and widely used exchange functional was proposed by Becke in 1988 [251]:

$$E_x = E_x^{LDA} - \beta \int \rho^{4/3} \frac{s^2}{1 + 6\beta s \sinh^{-1} s} dr, \tag{2.97}$$

with $s = |\nabla \rho| / \rho^{4/3}$. The B88 (or simply B) functional has the correct $1/r$ asymptotic behavior for the energy density, however unfortunately, not for the exchange potential [196]. The functional from Perdew, Burke and Ernzerhof (PBE) is based on the B88 functional, but gets along without any parameters, thus may be called *ab initio* [252].

On the other hand, the corresponding correlation functionals have rather complicated, analytical expressions which can hardly be motivated solely out of physical notions. One of the most used correlation functionals is that from Lee, Yang and Parr (LYP) [253] which contains one empirical parameter and is not based on the homogeneous electron gas. Rather it is derived from the correlation energy of the helium atom which had been evaluated with correlated wave functions by Colle and Salvetti [254]. These functionals can account for dynamical correlation but, on the other hand, perform poorly in cases with static correlation.

Finally exchange and correlation functionals are combined, e.g. the BLYP functional uses the B88 functional for the E_x part and the LYP functional for the E_c part. One of the most used functionals, at least in the chemical community, is a so-called hybrid functional, B3LYP [255],

$$E_{xc}^{B3LYP} = (1 - a)E_x^{LDA} + aE_{xc}^{HF} + bE_x^{B88} + cE_c^{LYP} + (1 - c)E_c^{LSD}. \tag{2.98}$$

Hybrid functionals use the Adiabatic Connection formalism [256], that is combining the exact HF exchange with density functionals. The B3LYP functional contains three parameters, which can be fitted to molecular test sets in order to improve the description of molecular properties such as, e.g. ionization energies and electron affinities. The improvement of GGA over LDA is substantially, and the performance of hybrid functionals such as B3LYP is almost as good (in some cases even better) as with MP2, however, at computational cost similar to HF [196].

2.5.3.1 Shortcomings of DFT

In this section some DFT failures are discussed. These are either attributed to the approximate character of E_{xc} or lie within the fundamentals of Kohn-Sham DFT itself, as a one particle ground state theory. As mentioned earlier the electron self interaction (not present in HF) is inherent in DFT, and as such causes severe problems.

For one thing, DFT tends to overestimate polarization and delocalization, respectively, in conjugated and charged systems [250]. Both effects are related to the so-called delocalization error [257], that is the electron density in such systems is unphysically delocalized over the entire system. In turn, this is a result of the incorrect convex behavior of the DFT energy for fractional numbers of electrons, i.e. a straight-line behavior would be exact. Therefore, a lower energy is artificially achieved for fractional charges in DFT. Already in 1981, Perdew and Zunger [258] suggested a self interaction correction (SIC), in which the self interaction is removed for each orbital. Though, the implementation is rather difficult and the description of molecules can be even worse [250]. Obviously, charge transport applications are considerably affected by the delocalization error. Mantz et al. [259] introduced a simplified SIC scheme for CT in DNA in which only the unpaired electron was affected by the correction. This scheme effectively alleviated the delocalization error.

Moreover, DFT overestimates binding energies ("overbinding"), underestimates reaction barriers and struggles in describing CT complexes and VdW interaction [257]. The latter issue affects weak interaction, which can be of substantial importance in large macromolecules, e.g. the π-stacking interaction between nucleobases in DNA. However, the VdW problem can be overcome by adding an empirical potential to DFT(B), which corrects for the attractive part of the VdW interaction (dispersion) [260, 261].

According to Janak's theorem (Eq. 2.93) the vertical ionization potential is obtained from the Kohn-Sham HOMO energy as:

$$I_v = -\epsilon_{HOMO}. \tag{2.99}$$

However, this is true only in exact Kohn-Sham theory. Because E_{xc} is not exact in pure DFT functionals (LDA and GGA), the Kohn-Sham energies differ significantly (at least 2 eV) from experimental ionization energies [249]. The main reason for this discrepancy lies in the wrong exponential behavior of the exchange potential, which in fact should approach asymptotically with $1/r$. In Hybrid functionals the error is reduced, since 20-30% of the exact (and correct) HF exchange is included. A more promising solution to this problem is to use so-called range-separated functionals [262, 263], in which E_{XC} is splitted. For small distances r the LDA/GGA exchange is used, while for larger r the HF exchange with the correct asymptotic behavior is used.

The Band Gap Problem   It turns out that the fundamental band gap of semiconductors and insulators is severely underestimated within conventional KS-DFT [264]. It could be shown, that this is not related to the form of E_{xc}, i.e. LDA or GGA, rather

it is an intrinsic problem in KS-DFT due to the absence of non-local contributions to E_{xc}. Scham and Schlüter [265, 266] as well as Levy and Perdew [267] showed, that the DFT band gap given by the energy difference of KS HOMO and LUMO is not the true quasi-particle (fundamental) band gap E_g:

$$E_g = \epsilon_{LUMO}^N - \epsilon_{HOMO}^N + \Delta_{xc}. \tag{2.100}$$

The difference term Δ_{xc} is called derivative discontinuity (DD). The DD emerges due to a jump in the true XC potential, when the number of electrons cross the integer N [267]. It turns out that Δ_{xc} is a substantial portion of the fundamental gap [264], which cannot be accounted for with conventional KS-DFT. Therefore, the missing Δ_{xc} term causes about 80% of the well-known DFT error [264]. What is needed, is an energy dependent XC potential, e.g. as provided by the SIC [258]. A different solution operates outside DFT known as the GW approximation [268, 269]. In this approximation, derived from many body theory, the XC potential is replaced by a truly non-local and energy dependent self energy $\Sigma(r, r')$ [264]. The method uses one-electron Green's functions for the interacting electrons (G) and screened Coulomb interaction (W). It could be shown, that band gaps for insulators and semi-conductors obtained with GW are in excellent agreement with experimental data [270].

In the next section an approximate semi-empirical version of DFT is described. The reader should note that the shortcomings of DFT, discussed here, hold for this DFT-derived method as well.

2.5.4 SCC-DFTB

Many molecular properties can be accessed with sufficient accuracy using DFT, yet the computation time increases dramatically with the system size by N^3. As a consequence, the application of standard KS-DFT is limited to about 1000 atoms. The study of large systems such as proteins or DNA, thus, requires well-controlled approximations, which save computation time without losing too much accuracy. A semi-empirical method, which is derived from DFT, make use of the tight-binding formalism, i.e. the density functional-based tight-binding (DFTB) method, which is described in the following.

To begin with, consider the non-self-consistent case, in which the ground state density ρ_0 is already known to sufficient accuracy. Then, the KS eigenstates can be obtained immediately from the KS equations:

$$\left[-\frac{1}{2}\nabla_i^2 + v_{eff}[\rho_0] \right] \phi_i = \epsilon_i \phi_i. \tag{2.101}$$

The KS eigenstates ϕ_i may be expanded in a minimal atomic basis η_μ considering only valence states[5], that is, e.g. 2s,2p$_x$,2p$_y$,2p$_z$ for second row elements and 1s for hydrogen.

---

[5]Omitting the core states is an approximation, though they are assumed to be chemically inactive [271]

Inserting the LCAO ansatz from Eq. 2.69

$$\phi_i = \sum_\mu c_\mu^i \eta_\mu \tag{2.102}$$

into Eq. 2.101 and using $\hat{H}[\rho_0] = \hat{T} + v_{eff}[\rho_0]$ one obtains:

$$\sum_\mu c_\mu^i \hat{H}[\rho_0] \mid \eta_\mu\rangle = \epsilon_i \sum_\mu c_\mu^i \mid \eta_\mu\rangle. \tag{2.103}$$

Multiplication from the left with $\langle \eta_\nu \mid$ leads to the secular equation:

$$\sum_\mu c_\mu^i \left\langle \eta_\nu \left| \hat{H}[\rho_0] \right| \eta_\mu \right\rangle = \epsilon_i \sum_\mu c_\mu^i \langle \eta_\nu | \eta_\mu \rangle \tag{2.104}$$

or in matrix notation:

$$\sum_\mu c_\mu^i \left(H_{\mu\nu}^0 - \epsilon_i S_{\mu\nu} \right) = 0. \tag{2.105}$$

Eq. 2.105 has to be solved just once via diagonalization of the Hamilton matrix $H_{\mu\nu}$. As a result, the single particle energies ϵ_i are obtained which add up to the electronic energy [271]:

$$E_{\text{elec}} = \sum_i^{\text{occ}} \epsilon_i. \tag{2.106}$$

A comparison with the DFT total energy from Eq. 2.92:

$$E_{\text{tot}} = \sum_i^{occ} \epsilon_i - \frac{1}{2} \iint \frac{\rho_0\left(r\right)\rho_0\left(r'\right)}{r - r'} dr dr' + E_{xc}[\rho_0] - \int v_{xc}(r)\rho_0(r)dr + V_{nn} \tag{2.107}$$

reveals that the double-counting terms and the core-core repulsion are missing. However, this time, these terms depend on the start (reference) density ρ_0 only. Further, it is assumed that the XC part decays exponentially, which is also true for the Coulomb part [271]. By a decomposition of the density into atomic contributions $\rho = \sum_\alpha \rho_\alpha$, the total energy can be written as:

$$E_{tot} = \sum_i \epsilon_i + \frac{1}{2} \sum_{\alpha\beta} U_{\alpha\beta}. \tag{2.108}$$

Eq. 2.108 shows the typical form of first empirical tight-binding models. Here, the first electronic term is calculated from Eq. 2.105, while the second term contains the empirical two-body potentials $U_{\alpha\beta}$, which are fitted in order to reproduce molecular properties such as geometries and vibrational frequencies.

2.5.4.1 Non-Self-Consistent Case: DFTB

The atomic orbitals η_μ from above can have rather diffuse shapes, which can be problematic for the description of some molecular properties. Therefore, an additional harmonic potential is introduced in the atomic Kohn-Sham equations [272]:

$$\left[-\frac{1}{2}\nabla_\mu^2 + v_{eff}[\rho_{atom}] + \left(\frac{r}{r_0} \right)^2 \right] \eta_\mu = \epsilon_\mu \eta_\mu \tag{2.109}$$

The confining potential compresses the atomic orbitals η_μ and densities ρ_α. Usually, the parameter r_0 is chosen to be twice the covalent radius of the respective atom, which leads to good results for most molecular properties[273]. On the other hand, quantities which depend more strongly on the long-range overlap of the wave function are severely underestimated. For instance, the electronic couplings T_{ij} from Eq. 2.70 are significantly too low using the standard values for r_0. Thus, the atomic orbitals need to be decompressed, i.e. the confinement radius r_0 needs to be adjusted for CT applications.

The start density ρ_0 in the DFTB Hamiltonian is composed of the single densities of the atoms ρ_α:

$$H_{\mu\nu} = \left\langle \eta_\nu \left| \hat{H}[\rho_0] \right| \eta_\mu \right\rangle = \left\langle \eta_\nu \left| \hat{H} \left[\sum \rho_\alpha \right] \right| \eta_\mu \right\rangle \tag{2.110}$$

Moreover, instead of using the complete start density ρ_0, the 2-center approximation is applied [271]:

$$\left\langle \eta_\nu \left| \hat{H}[\rho_0] \right| \eta_\mu \right\rangle = \left\langle \eta_\nu \left| \hat{H} \left[\rho_\alpha + \rho_\beta \right] \right| \eta_\mu \right\rangle . \tag{2.111}$$

Therefore, the Hamilton matrix contains only the diagonal elements $H_{\mu\mu} = \epsilon_\mu$ from Eq. 2.109 and those 2-center elements $H_{\mu\nu}$, in which the atomic orbital η_μ is located on atom α, while the atomic orbital η_ν is located on atom β. The matrix elements $H_{\mu\nu}$ and $S_{\mu\nu}$ are stored in tables for atom-atom distances up to 10 a.u. [271]. Therefore, no integral calculations have to be performed anymore, rather the matrix elements can be read in for any molecular geometry and then, be orientated in space using the Slater-Koster sin/cos combination rules [274]. Thus, the electronic part of the energy can be calculated with Eq. 2.105. Note, the set of atomic orbitals η_μ is clearly non-orthogonal, i.e. $S_{\mu\nu} \neq \delta_{\mu\nu}$, which makes this scheme more transferable [271].

The missing part of the total energy, i.e. the second term in Eq. 2.108, is referred to as the repulsive potential:

$$E_{\text{rep}}[\rho_0] = \frac{1}{2} \sum_{\alpha\beta} U_{\alpha\beta}. \tag{2.112}$$

To get the two-body potentials $U_{\alpha\beta}$, the energy difference between the DFTB electronic energy and the DFT total energy is evaluated for various distances $R_{\alpha-\beta}$:

$$U_{\alpha\beta} \left(R_{\alpha-\beta} \right) = E_{\text{tot}}^{\text{DFT}} \left(R_{\alpha-\beta} \right) - \sum_i \epsilon_i. \tag{2.113}$$

Technically, various molecular structures containing the atoms α and β are included in the fitting procedure [271].

This is the non-self-consistent variant of DFTB performing well for homo-atomic systems, in which there is no or marginal charge transfer between the atoms [271]. On the other hand, differences in electronegativity between atoms can enhance charge transfer considerably. In this case, the approximation for the density as a superposition of atomic densities is not longer valid. Nevertheless, the scheme provides a good starting point for an extension of DFTB as described in the following section.

2.5.4.2   Self-Consistent Case: SCC-DFTB

To apply DFTB for hetero-atomic systems as well, e.g. biomolecules, the self-consistent charge method SCC-DFTB was developed [272, 273, 275–278]. Starting point is a second order Taylor expansion of the DFT total energy with respect to the charge density fluctuations $\delta\rho$ around a given reference density ρ_0

$$
\begin{aligned}
E = \sum_i^{occ} \left\langle \Psi_i \middle| \hat{H}^0 \middle| \Psi_i \right\rangle + \frac{1}{2} \iint{}' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'}\bigg|_{\rho_0} \right) \delta\rho\delta\rho' \\
- \frac{1}{2} \iint{}' \frac{\rho_0\rho_0'}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]\rho_0 + V_{nn},
\end{aligned}
\tag{2.114}
$$

with $\rho_0' = \rho_0\,(\vec{r}')$ and $\int' = \int d\vec{r}'$. Using again the LCAO ansatz (Eq. 2.69), the first term in Eq. 2.114 can be expressed as:

$$
\left\langle \Psi_i \middle| \hat{H}^0 \middle| \Psi_i \right\rangle = \sum_{\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0,
\tag{2.115}
$$

and thus, can be calculated as in the non-self-consistent scheme above. However, like before, the last four terms in Eq. 2.114 depend solely on the reference density ρ_0, hence are parts of the repulsive energy $E_{\text{rep}}[\rho_0]$ evaluated with Eq. 2.113 as described above.

Eventually, one term remains in Eq. 2.114, E^{2nd}, which is of second order with respect to the density fluctuation $\delta\rho$. As carried out for the density above, the total density fluctuation may be decomposed into atomic contributions as well:

$$
\delta\rho = \sum_\alpha \delta\rho_\alpha.
\tag{2.116}
$$

Moreover, a charge monopole approximation is used for the atomic density fluctuations $\delta\rho_\alpha$, that is the density fluctuation is divided into a radial contribution F_{00}^α and an angular contribution Y_{00}. Making use of Mulliken charges [279], $\delta\rho_\alpha$ can be approximated to:

$$
\delta\rho_\alpha \approx \Delta q_\alpha F_{00}^\alpha Y_{00}.
\tag{2.117}
$$

The basis for this approximation is the assumption of spherical charges and charge density fluctuations as well as the neglect of angular deformation of the charge density change in second order. With these simplifications E^{2nd} reads:

$$
E^{2nd} \approx \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \iint{}' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'}\bigg|_{\rho_0} \right) F_{00}^\alpha F_{00}^\beta Y_{00}^2 dr dr'.
\tag{2.118}
$$

Though Eq. 2.118 appears to be rather complicated, two limiting case can be deduced [271]:

- For large distances $|\vec{r} - \vec{r}'| \to \infty$, the XC term vanishes and the integral describes the regular Coulomb interaction with its $1/R_{\alpha\beta}$ dependence between two partial atomic charges Δq_α and Δq_β:

$$
E^{2nd} \approx \frac{1}{2} \sum_{\alpha\beta} \frac{\Delta q_\alpha \Delta q_\beta}{R_{\alpha\beta}}.
\tag{2.119}
$$

- For very small distances $|\vec{r} - \vec{r}'| \to 0$, the integral in Eq. 2.118 describes the electron-electron interaction on atom α, thus can be approximated to:

$$E^{2nd} \approx \frac{1}{2} \frac{\delta^2 E_\alpha}{\delta q_\alpha^2} = U_\alpha, \tag{2.120}$$

  with U_α being the Hubbard Parameter [280], also known as chemical hardness. It characterizes the energy change of a system upon electron addition or removal.

Technically, both limiting cases have to be combined via an interpolation function γ. This a general concept in semi-empirical methods, such as MNDO, AM1 and PM3 [271], and γ can have a rather simple form, e.g. using the Klopman-Ohne approximation [281, 282]:

$$\gamma_{\alpha\beta} = \frac{1}{\sqrt{R_{\alpha\beta}^2 + 0.25(1/U_\alpha + 1/U_\beta)^2}}. \tag{2.121}$$

Eventually, E^{2nd} reads:

$$E^{2nd} = \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta}. \tag{2.122}$$

Thereby, all terms in Eq. 2.114 can be evaluated and the SCC-DFTB total energy reads:

$$E = \sum_i^{occ} \sum_{\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} + E_{rep}[\rho_0]. \tag{2.123}$$

The partial Mulliken charges Δq_α satisfy the relation $\Delta q_\alpha = q_\alpha - q_\alpha^0$, with q_α^0 corresponding to the number of valence electrons of the charge neutral atom α. On the other hand, the atomic charges q_α are related to the overlap matrix $S_{\mu\nu}$ via:

$$q_\alpha = \sum_i^{occ} \sum_{\mu\nu} c_\mu^i c_\nu^i S_{\mu\nu}. \tag{2.124}$$

Applying the variational principle to Eq. 2.123, leads to a generalized eigenvalue problem, which has to be solved iteratively. As a result, the Hamilton matrix elements $H_{\mu\nu}$ are obtained as [276]:

$$H_{\mu\nu} = H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu} \sum_\xi \Delta q_\xi \left(\gamma_{\alpha\xi} + \gamma_{\beta\xi} \right). \tag{2.125}$$

The Hamilton matrix elements $H_{\mu\nu}$ depend on the Mulliken charges which, in turn, depend on the atomic coefficients c_μ^i. However, these are obtained as solution of the secular equation. Therefore, the energy evaluation has to be performed self-consistently with respect to the charge [283].

In general, the SCC-DFTB method as described above provides an efficient tool for the description of molecular properties, in particular, for organic and biological molecules [273, 284, 285]. SCC-DFTB yields results comparable to DFT for a wide range of molecules, yet it is considerably faster by roughly three orders of magnitude [284, 286]. Nevertheless, large systems, e.g. proteins in solutions, contain many thousands of atoms, thus, can hardly be studied solely with QM methods, even semi-empirical methods like SCC-DFTB would fail.

QM/MM-coupling   For many applications, especially when dealing with large-scale bio-molecular processes, the system is usually separated into a small part described with quantum mechanics (QM), while the remaining part is treated with classical molecular mechanics. Effectively, the charges Q_A from the MM region polarize the QM region. For a comprehensive review see Ref. [287]. Cui et al. [288] provided a straight forward way to incorporate the effect of external MM charges Q_A into the SCC-DFTB Hamiltonian (Eq. 2.125):

$$H_{\mu\nu} = H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu} \left(\sum_{\xi} \Delta q_\xi \left(\gamma_{\alpha\xi} + \gamma_{\beta\xi} \right) + \sum_A Q_A \left(\frac{1}{r_{A\alpha}} + \frac{1}{r_{A\beta}} \right) \right). \qquad (2.126)$$

By making use of Eq. 2.126, the CT parameter evaluation (Sec. 2.5.1) can be performed by taking into account environmental effects, though purely electrostatically. More precisely, the CT parameters between the nucleobases in DNA (chapter 4 and 5) and the tryptophan sidechains in PL (chapter 8) can be calculated in the presence of an external electric field. For the simulations of DNA, this field is built by the solvent, the counterions, the remaining DNA bases and the DNA backbone, i.e. sugar and phosphate groups (see Fig. 2.12). In the simulations of PL, the MM environment is composed of the solvent and the protein remainder.

The Hamilton and overlap matrix elements $H_{\mu\nu}$ and $S_{\mu\nu}$ in Eq. 2.125 and 2.126 are used to calculate the CT matrix elements in Eq. 2.70, thus, producing a very efficient scheme as $H_{\mu\nu}^0$ and $S_{\mu\nu}$ are read in from tables for every geometry and no integral evaluations have to be performed. Moreover, the environmental effects on the CT parameters can be studied by using either Eq. 2.125 or 2.126. The good performance for the evaluation of CT parameters, e.g. in DNA, therefore, allows for the coupling to extended nano-second MD simulations [147].

2.6   Coarse-Grained Electron-Ion Dynamics

The Hamiltonian provided in Eq. 2.126 accounts for the polarization of the MM environment onto the CT-active sites in the QM system. The Landauer transmissions and currents from Sec. 2.3 are obtained using this Hamiltonian. Therefore, all these calculation will include the electrostatic effects of the MM environment, which will lead to a shift and large fluctuation of the site energies, as we will see in chapter 4 and 5. However, these simulation are carried using charge-neutral species, i.e. there is no explicit electron or hole charge in the QM region. Instead, the KTA is used to describe the CT process. As a consequence, the MM environment feels the neutral CT sites only, and the resulting dynamics may differ significantly from that of a simulation in which there is an explicit charge. Introducing a charge, e.g. a hole charge on a nucleobase in DNA, leads to a decrease of the solvent ESP on the respective site by more than 2 V as shown in Ref. [173]. Moreover, the adjacent nucleobases are affected as well, though the stabilization is not as strongly as for the site on which the hole is located. This phenomena had been noticed of as a solvent polaron [289], i.e. water molecules re-orientate themselves stabilizing the hole charge and following its dynamics. This polarization of the excess charge back onto the MM environment is not accounted for in the Landauer simulations. Additionally, it is assumed that the time scales of electronic and ionic motion are separated clearly, which may not be the case for CT in bio-molecules necessarily. At the end, if electron and ion movement occur on a similar time scale, a clearly non-adiabatic dynamics is needed, which accounts for the polarization of the excess charge back to the MM environment as well.

Such an electron-ion dynamics (EID) is computationally very expensive when treated fully atomistic. Therefore, coarse-graining (CG) is necessary in order to reduce the complexity and to allow for extended simulations on the nano-second scale. Besides the approximation due to CG, the reduction of complexity enables the possibility to build up a model which allows for a well-controlled study of the CT problem. The coarse-grained EID method, used in this work to study the photoactivation process in PL (chapter 8), is described in the following.

The methodology for a direct EID simulation of CT in complex systems has been described in detail recently [173]. Here, only the most important concepts are mentioned briefly. Starting point is a charge-neutral closed-shell system with N electrons, for which the DFT ground-state density ρ_0 and the Kohn–Sham (KS) orbitals Ψ_i^0 have been determined. The total DFT energy can be written in terms of the KS matrix elements:[290]

$$E^N[\rho_0] = \sum_i^{N/2} 2 \left\langle \Psi_i^0 | H[\rho_0] | \Psi_i^0 \right\rangle + E_{\mathrm{DC}}[\rho_0], \qquad (2.127)$$

with

$$
E_{\mathrm{DC}}[\rho_0] = -\frac{1}{2} \iint{}' \frac{1}{|r - r'|} \rho_0(r)\, \rho_0(r')\, \mathrm{d}^3 r\, \mathrm{d}^3 r' \tag{2.128}
$$

$$
- \int v_{\mathrm{XC}}[\rho_0]\, \rho_0\, \mathrm{d}^3 r + E_{\mathrm{XC}}[\rho_0] + V_{\mathrm{NN}}.
$$

$E_{\mathrm{DC}}[\rho_0]$ contains the DFT 'double-counting' and nucleus–nucleus repulsion terms V_{NN}. The state of the system containing a hole is described as containing $N-1$ electrons, and its energy can be approximated by an expansion with respect to the density $\rho = \rho_0 + \delta\rho$ around the N-electron reference density ρ_0 up to the second order ($n_{\mathrm{HOMO}} = 1$, $n_i = 2$ otherwise):

$$
E^{N-1}[\rho] \approx \sum_i^{N/2} n_i \langle \Psi_i | H[\rho_0] | \Psi_i \rangle + E_{\mathrm{DC}}[\rho_0] \tag{2.129}
$$

$$
+ \frac{1}{2} \iint{}' \left(\frac{1}{|r - r'|} + f_{\mathrm{XC}}[\rho_0] \right) \delta\rho(r)\, \delta\rho(r')\, \mathrm{d}^3 r\, \mathrm{d}^3 r'.
$$

With $f_{\mathrm{XC}} = \left(\delta^2 E_{XC} \right) / \left(\delta\rho^2 \right)$ being the 2nd derivative of the DFT exchange-correlation energy.

Making use of the frozen-core approximation:

$$
\Psi_i \approx \Psi_i^0 \tag{2.130}
$$

the total energy (using the HOMO Ψ_{h}^0) can be approximated as:

$$
E^{N-1}[\rho] \approx E^N[\rho_0] - \left\langle \Psi_{\mathrm{h}}^0 | H[\rho_0] | \Psi_{\mathrm{h}}^0 \right\rangle + E^{\mathrm{2nd}}. \tag{2.131}
$$

The three contributions to the total energy are then further simplified as follows:

- The total energy of the neutral reference system is computed as the molecular-mechanics total energy

$$
E^N[\rho_0] = E^{MM}. \tag{2.132}
$$

- To compute the matrix elements involving the hole wave-function, Ψ_{h}^0, an expansion in terms of fragment orbitals ϕ_m is used. The FOs are obtained as DFT (Kohn-Sham) orbitals of the molecular fragments. In the case of DNA, these fragments are the individual nucleobases or Watson-Crick base pairs, while in the case of photolyase, these fragments are the sidechains of the involved tryptophan sites [147].

$$
\Psi_i = \sum_{m=1}^{M} a_m^i \phi_m \tag{2.133}
$$

In the simplest case, one orbital per fragment is included only, which is the HOMO ϕ_m of the fragment m for hole transfer simulations[6]. The hole wave function is

---

[6]For more complicated cases, more orbitals per site can be included easily, see e.g. Ref. [172]

then expanded as [147]:

$$\langle \Psi_{\mathrm{h}}^0 | H[\rho_0] | \Psi_{\mathrm{h}}^0 \rangle = \sum_m \sum_n a_m^{h\,*} a_n^h \langle \phi_m^* | H[\rho_0] | \phi_n \rangle = \sum_m \sum_n a_m^{h\,*} a_n^h H_{mn}^0. \quad (2.134)$$

The diagonal matrix elements $\epsilon_m^0 = T_{mm}^0 = H_{mm}^0$ represent effective ionization potentials and will be referred to as site energies, while the off-diagonal elements are the so-called electronic couplings or hopping matrix elements $T_{mn}^0 = H_{mn}^0$. The index '0' refers to the charge-neutral reference density ρ_0.

- In a similar fashion, the differential density $\delta\rho = \rho - \rho_0$ in Eq. 2.130 can be decomposed into contributions located on the individual molecular fragments,

$$\delta\rho = \sum_m \delta\rho_m. \quad (2.135)$$

and the last term in Eq. 2.130, E^{2nd}, can be written as:

$$E^{\text{2nd}} = \frac{1}{2} \sum_{nm} \iint{}' \left(\frac{1}{|r - r'|} + f_{\text{XC}}[\rho_0] \right) \delta\rho_m(r)\, \delta\rho_n(r')\, \mathrm{d}^3r\, \mathrm{d}^3r'. \quad (2.136)$$

Applying a monopole approximation for the integral[7], the expression can be further simplified to:

$$E^{\text{2nd}} = \frac{1}{2} \sum_m U_m \Delta Q_m^2 + \sum_m \sum_{n>m} \frac{\Delta Q_m \Delta Q_n}{R_{mn}}, \quad (2.137)$$

where ΔQ_m denotes the fraction of the hole charge localized on site m. The first term involves the Hubbard parameter U (chemical hardness), which determines the electron-electron repulsion on site m. The interaction between neighboring sites is described as simple Coulomb repulsion of the hole portions on fragments m and n. Note that this notation is used only for simplicity, since it is easily possible to project the fragment charges Q_m to atomic charges q_α, and compute the interaction energy much more accurately. This second order-term contains much of the well known DFT self-interaction error, and a self-interaction correction (SIC) is applied by scaling E^{2nd} by a factor of $\Gamma = 1/6$, see Ref. [173] for more details. U is determined using DFT calculations, leading to a value of $U = 5.33\ eV/e^2$ for the Trp sidechain.

Now, the total energy reads:

$$E^{N-1}[\rho] = E^{MM} \quad - \quad \sum_m a_m^* a_m \epsilon_m^0 - \sum_m \sum_{n>m} a_m^* a_n T_{mn}$$

$$+ \quad \Gamma \left(\frac{1}{2} \sum_m U_m \Delta Q_m^2 + \sum_m \sum_{n>m} \frac{\Delta Q_m \Delta Q_n}{R_{mn}} \right). \quad (2.138)$$

---

[7]Note that this is the same proceeding as in Sec. 2.5.4 for the conventional SCC-DFTB method.

2.6.1   QM/MM Coupling

In the case of hole transfer in PL, only the three Trp sidechains are treated quantum mechanically (QM region), while the remainder of the system is described with classical molecular mechanics (MM-region). Certain parts of the interaction between the QM and MM region, occur in all three energy contributions introduced above. In the CG methodology described above, a part of the QM region is already treated with E^{MM}. Therefore, to describe the interaction of this part with the MM region, the energy expression is simply extended to the entire (QM+MM) system. The CT parameters ϵ_m^0 and T_{mn}^0 are computed in the presence of the solvent and the remainder of the protein using Eq. 2.126. Kubař et al. [164] have shown that the site energies ϵ_m^0 are related directly to the electrostatic potential (ESP) at site m:

$$ESP_m = \sum_{MM charges} q_\alpha^{MM}/R_{m\alpha}. \tag{2.139}$$

Technically, the ESP stemming from the MM environment is included in the calculation of CT parameters [147, 164]. Applying the variational principle to Eq. 2.138 leads to the self-consistent diagonal elements [8]:

$$\epsilon_m = -\epsilon_m^0 + U_m \Delta Q_m + \sum_{n \neq m} \frac{\Delta Q_n}{R_{mn}}. \tag{2.140}$$

Three major factors determine the site energies ϵ_m^0, hence the energetics of CT:

1. *Static differences*: Admittedly, this is no issue in PL as the CT occurs between identical species (i.e. between three Trp sidechains). By contrast, in case of hole transfer in DNA, the site energies of adenine and guanine differ by about 0.4 eV, hence CT is determined critically by the DNA sequence.

2. *QM/MM coupling, i.e. fluctuations of ϵ_m^0*: The electrostatic interaction with the environment, i.e. the electrostatic potential ESP_m affects the site energy substantially. The resulting effects can be twofold: For one thing, the ESP may induce a static shift which may drive the CT, for another, there is a dynamic component [164], that is the dynamics of the MM environment composed of solvent and remainder of the protein leads to considerable fluctuations of the site energies.

3. *Effect of the hole charge*: The positive hole charge polarizes the environment. In case of DNA, this leads to a considerable rearrangement of the environment, which causes a large change in the ESP. In turn, this change leads to a stabilization of the hole charge due to a significant decrease of the respective site energy.

It should be mentioned that the separation of the latter two points is purely phenomenological, that is both effects have the same origin within the methodology.

---

[8]Note that only the CT parameter ϵ_m^0 is referred to as site energy and as such will be used as a measure of an effective IP, while the self-consistent diagonal elements ϵ_m are used for the propagation of the electronic Hamiltonian only.

Note that Eq. 2.140 accounts for the complete interaction with the environment, but the sum over all fragments m does not represent the total energy since the E^{MM} term is missing. When comparing site energy differences, i.e. between the sites m and n, this does not reflect the full total energy difference. Nevertheless, this energy difference gives a first insight into the energetics of the system and is a critical factor that drives the dynamics of the system. As it was shown in detail before, the dynamics is determined by the ESP, since fluctuations induced by solvent and protein will determine the energies of the CT-active sites. Further, the presence of the hole charge at site m polarizes the (protein and water) environment. This leads to a dramatic decrease of the ESP at site m. The site energies were shown to be lowered by 1-2 eV for hole transfer in DNA [173]. This polarization of the environment, conceptualized as a 'polaron' [289], is likewise a dynamic entity following the hole charge.

Internal Reorganization Energy

The inner-sphere reorganization is not included in the equations above, since the dynamics of atoms is driven using classical MD simulations with force field parameters derived for neutral CT sites (except for the atomic charges, which are updated in every step to include the projected hole charge). In this study, the effect of internal reorganization for each site m is approximated as a small correction λ_i^m to the site energies ϵ_m[9]

$$\epsilon'_m = \epsilon_m - \lambda_i^m \Delta Q_m. \tag{2.141}$$

The value of λ_i^m is computed beforehand with quantum chemical methods, as described in Sec. 2.2.2.5. The inclusion of λ_i^m is a minor correction compared to the SIC scaling parameter Γ, thus may be considered as additional localization force.

2.6.2 Calculation of Coarse-Grained Parameters

In previous studies, it was shown in great detail, that CT parameters computed for DNA bases at the SCC-DFTB level agree excellently with those obtained of higher-level DFT, HF and CASPT2 calculations [147]. Recently, Voityuk and coworkers have shown that DFT provides reasonable CT parameter for hole transfer in DNA [291]. Nevertheless, extensive benchmark calculations have to be performed for every new application. For the PL system, benchmark calculations of energies and MOs for the highest occupied levels of a single skatole molecule (used as a model for the tryptophan sidechain) were performed with DFTB, DFT and HF. As a result, an excellent performance of DFTB was confirmed for this application. The data can be found in the appendix (Tab. A.6). Furthermore, the QM/MM coupling at the DFTB level of theory was shown to reproduce the DFT values very well [147]. Therefore, this part of the QM/MM interaction is well treated using SCC-DFTB for the calculation of the CT parameters.

---

[9]Note that in Marcus' theory, λ_i is the overall internal reorganization energy for both donor and acceptor.

2.6.3   Non-Adiabatic Simulations

Applying the Lagrangian formalism, coupled equations of motion for the electronic and ionic degrees of freedom were derived [173, 292]. This leads to classical equations of motion for the nuclei (atoms):

$$M_\alpha \ddot{R}_\alpha = \frac{\partial E^{MM}(q_\alpha^{ion}, q_A^0)}{\partial R_\alpha}, \tag{2.142}$$

with the charges q_α^{ion} of the $(N-1)$-electron QM system ($\alpha \in$ QM) and the charges q_A^0 for the charge-neutral MM system ($A \in$ MM); and effective time-dependent Kohn-Sham (TD-KS) equations (DFT analogue of the time-dependent Schrödinger equation) for the electronic degrees of freedom:

$$\dot{a}_m = i \sum_n a_n T_{mn}, \tag{2.143}$$

with

$$T_{mn} = \begin{cases} \epsilon_m, & \text{if } n = m \\ -T_{mn}^0, & \text{if } n \neq m \end{cases}. \tag{2.144}$$

Thus, only the diagonal elements ϵ_m are affected by the second-order term (Eq. 2.140), while the off-diagonal elements are just the hopping matrix elements T_{mn}^0. Since the two sets of equations are coupled, the electronic degrees of freedom are propagated using the Runge-Kutta integration as implemented in RKSuite [293], with a variable time step, between two time steps of the propagation of the nuclei. The MD time step was chosen to be 1 fs. Effectively, the propagation of the electronic degrees of freedom leads to a new charge distribution, which is projected onto the classical particles (atomic partial charges) in the following MD time step. The propagation of the classical system then proceeds using classical Newtonian mechanics [173]. Effectively, the electronic dynamics leads to an update of MM charges before every MD time step. In this way, the dynamics of the hole charge is driven by the TD-KS equations, while the dynamics of the molecular system is determined by molecular-mechanics using the updated atomic charges obtained from the integration of the TD-KS equation.

Part I

Electronic Transport in DNA and Related Systems

3

CT-Parameter for Hole Transfer in DNA

In this chapter, the computation of CT parameters as described in Sec. 2.5.1 is tested, evaluated and refined for hole transfer in DNA using the SCC-DFTB method (Sec. 2.5.4). Moreover, the CT parameters obtained with SCC-DFTB are compared to those from other quantum chemical methods, such as HF, DFT and CAS-PT2. The CT parameters are computed for the standard nucleobases (G,A,C,T) using various helical parameters as well as for DNA derivatives, i.e. modified nucleobases.

3.1 DFTB-FO Parametrization

The CT parameters are i) the site energies $\epsilon_i = T_{ii}$ (Eq. 2.67), which govern the energetics of charge transfer as effective ionization potentials, and ii) the coupling matrix elements T_{ij} (Eq. 2.68), which were shown to be very sensitive to the choice of the AO basis. Note that the standard SCC-DFTB method uses a confined basis set, with the confinement radius r_0 in Eq. 2.109 set to approximately twice the covalent radius of the respective atom [273, 275]. The relaxation of this constraint leads to more diffuse basis functions and makes the coupling matrix elements more appropriate, as described below.

3.1.1 Ionization Potentials and Site Energies

The energetics of hole transfer between two nucleobases can be estimated approximately by the difference of oxidation potentials (or IP) of these molecules. Therefore, the first important step is to benchmark SCC-DFTB in this respect. Tab. 3.1 compares the vertical IP values calculated with HF-KTA with those determined with B3LYP, PBE (using Gaussian03 [294]) and DFTB using the ΔSCF procedure. The vertical IPs are slightly underestimated by both DFT functionals as well as SCC-DFTB (using the standard

Table 3.1: IP of nucleobases calculated as ΔSCF and ϵ_{HOMO} (ϵ_H) using DFT and HF methods (the 6-31G(d,p) basis set was used). The NDDO-G as well as experimental values for vertical and adiabatic ionization are given for reference. All values in eV.

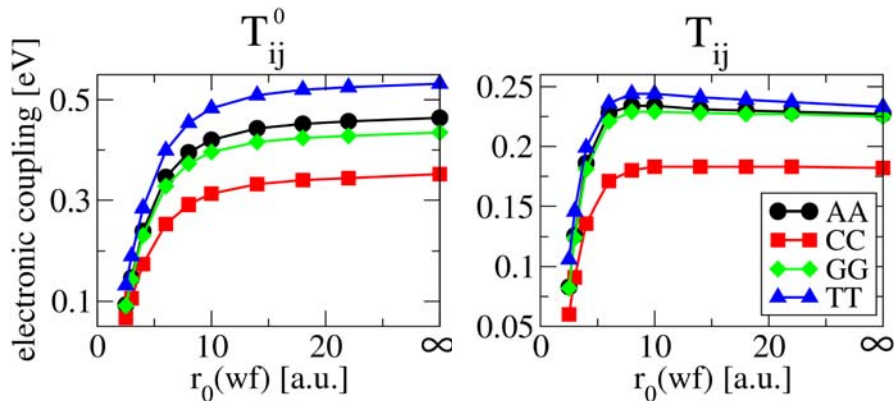| method	guanine		adenine		cytosine		thymine	
	ΔSCF	ϵ_H	ΔSCF	ϵ_H	ΔSCF	ϵ_H	ΔSCF	ϵ_H
DFTB	7.74	4.96	8.12	5.29	8.82	5.17	9.03	5.92
B3LYP	7.53	5.51	7.96	5.89	8.35	6.13	8.72	6.57
PBE	7.36	4.69	7.81	5.06	8.20	5.18	8.48	5.63
HF	—	7.99	—	8.37	—	9.16	—	9.52
NDDO-G [295]	8.10		8.53		9.10		9.15	
Exp.(vert.)[296]	8.24		8.44		8.94		9.14	
Exp.(adiab.)[297]	7.77		8.26		8.68		8.87	

basis), which even gives the best agreement with the experimental values. Further, the SCC-DFTB values agree perfectly with the experimental adiabatic IP. In particular, the relative energies, which are important for the energetics of hole transfer in DNA, are quite well reproduced. Therefore, SCC-DFTB is well suited for the simulation of CT processes in this respect.

The calculation of Kohn-Sham single-particle HOMO energy is a faster way to estimate the site energy than using the ΔSCF procedure. However, several problems arise when using approximative DFT functionals based on the generalized gradient approximation (GGA)[1]. First of all, the HOMO energies are significantly underestimated. While the HOMO energy represents the adiabatic IP value quite well with HF due to Koopmans' theorem [245], it does not match when using GGA or hybrid functionals like PBE or B3LYP, as a consequence of the wrong asymptotic behavior of GGA exchange functionals [285]. Since SCC-DFTB is parameterized to PBE, this holds for SCC-DFTB as well, as can be seen from Tab. 3.1. Actually, it is only the relative HOMO energies of individual bases that affect the CT process, and the HOMO-energy difference provided by SCC-DFTB is correct for guanine and adenine, making this shortcoming irrelevant.

Also, the cytosine HOMO energy is severely underestimated with respect to adenine and guanine by all DFT methods. Nevertheless, the actually high-lying HOMO of cytosine and thymine will prevent a hole from being localized on these bases. Thus, in the context of hole transfer, the C and T bases may be excluded from the quantum-chemical description without loss of accuracy. This way, the DFT failure becomes irrelevant. As can be seen from Tab. 3.1, the relative energies of adenine and guanine are well reproduced by both HOMO energies and ΔSCF values for all the DFT methods. In this case, the SCC-DFTB Kohn-Sham energies can be used as an approximation of the energy of a hole residing on a particular nucleobase.

Another point of imperfection represents the energy difference between HOMO and

---

[1]This issue is addressed in more detail in Sec. 2.5.3.1.

Figure 3.1: T_{ij}^0 (left) and T_{ij} (right) with respect to r_0 (wf) using r_0 (dens) = 7 a.u. for stacked base pairs adenine–adenine (AA), cytosine–cytosine (CC), guanine–guanine (GG) and thymine–thymine (TT) in the configuration with rise = 3.38 Å and twist = 0°.

second-highest occupied MO (HOMO−1) being too small in some cases, as can be seen in Tab. A.1 and A.2 in the appendix. For cytosine, HOMO and HOMO−1 may even swap with DFT, depending on minor changes in the molecular geometry. The HOMO—HOMO−1 energy difference of less than 0.4 eV is problematic, as fluctuations of this magnitude appear typically during MD simulations, as discussed below, and, consequently, the orbitals with such small energy difference may change order in the course of MD simulation. Since HOMO−1 of both guanine and adenine has σ character in DFT, a wrong representation of the electronic structure would result. The hybrid functional B3LYP alleviates this problem slightly due to the HF exchange contribution, yet HOMO—HOMO−1 energy difference of less than 0.4 eV still appears.

Interestingly, the inclusion of the electrostatic environment by means of point charges for DNA, water and counterions reduces this effect considerably: When computing the FO's and the CT integrals in the presence of external point charges, the energy difference becomes much larger and the artificial level crossings disappear, as discussed in more detail below.

3.1.2 A DFTB Basis Set Optimized for Hole Transfer

The standard basis set for DFTB is constructed by using Eq. 2.109 with rather small values of r_0. Usually, r_0 for an atomic orbital is chosen to be twice the covalent radius of the element, yielding typical values for carbon, nitrogen and oxygen between 2 and 3 a.u. This, however, leads to rather small coupling integrals T_{ij} (Eq. 2.68), as the important wave-function tails are suppressed using such a confined basis. Therefore, the tight confinement radius for the basis functions η_μ has to be relaxed. The calculation of DFTB Hamilton matrix elements (Eq. 2.111) consists of two steps: First, the atomic electron density is calculated from Eq. 2.109 using a constant value of $r_0 = 7.0$ a.u. (for a more detailed discussion see Ref. [273]). In the second step, the basis functions η_μ are computed using a different value of the confinement radius r_0. The dependence of
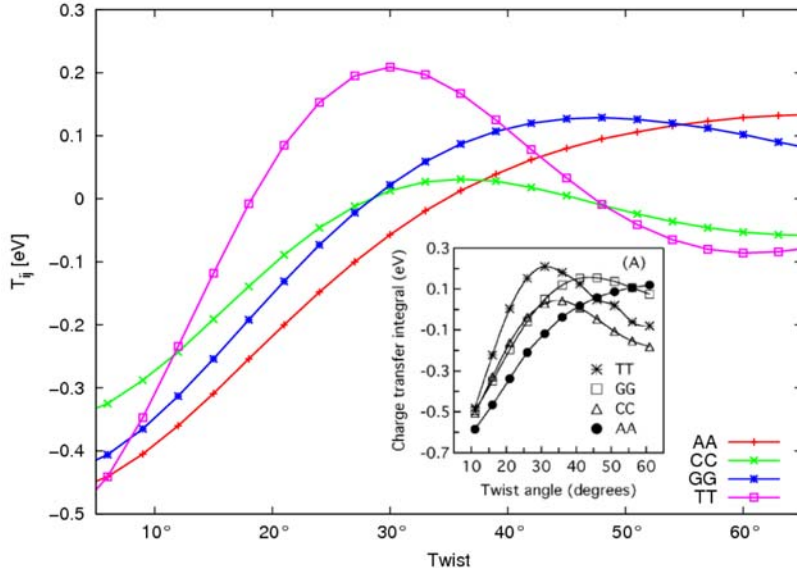
Figure 3.2: T_{ij} with respect to the twist angle (with helical parameter rise 3.38 Å) for stacked pairs adenine–adenine (AA), cytosine–cytosine (CC), guanine–guanine (GG) and thymine–thymine (TT). The inset shows the corresponding DFT results from Ref. [144].

T_{ij}^0 (Eq. 2.68) and T_{ij} (Eq. 2.72) on r_0 for the basis function is shown in Fig. 3.1 (with $r_0 = 7.0$ a.u. for the density).

T_{ij}^0 is constantly increasing with increasing r_0, while T_{ij} assumes a maximum at r_0 around 8 a.u. This is due to the effect of increased overlap S_{ij}, which seems to overrule the r_0-dependence of T_{ij}^0, according to Eq. 2.72, thereby making T_{ij} drop slightly for large r_0. More extensive tests showed that the values r_0 (dens) = 7 a.u. and r_0 (wf) = 8 a.u. lead to quite a good agreement with higher-level calculations. These fixed values (from now on called the "8-7 basis set") will be applied to compute the off-diagonal coupling integrals as used in this work.

3.2   Electronic Couplings with DFTB

Senthilkumar et al. [144] evaluated CT integrals using DFT, based on the FO approach as discussed above. They determined the HOMOs ϕ_i for isolated nucleobases, and calculated the CT integrals using the Kohn-Sham Hamiltonian for the stacked nucleobase dimer: $T_{ij} = \left\langle \phi_i \left| \hat{H}_{\mathrm{KS}} \right| \phi_j \right\rangle$. Exactly the same approach as outlined above is used here, however, using the much faster approximative SCC-DFTB rather than the full DFT-SCF as used in Ref. [144]. Thus, the mentioned work represents an excellent benchmark for our method. Fig. 3.2 shows the dependence of T_{ij} on the twist parameter of a base-pair step at a constant distance ("rise") of 3.38 Å, the bases assuming a parallel configuration.

The DFTB values match very well the DFT data (inset from Ref. [144]), especially the qualitative and quantitative differences between the base pairs are reproduced excellently. Only for small twist angles, the DFTB values are underestimated slightly.
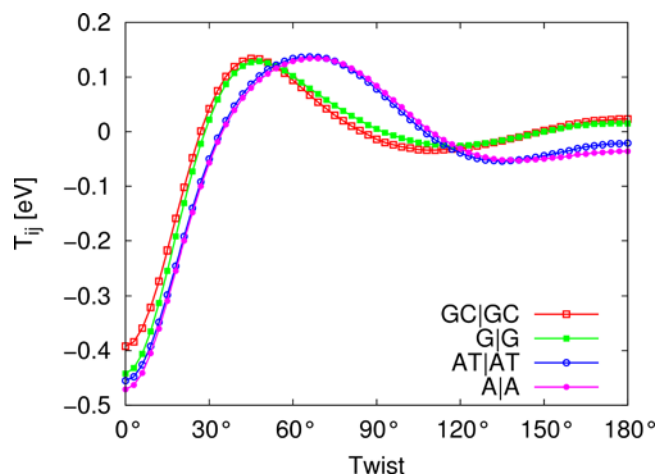
Figure 3.3: Comparison of CT integrals for different FO schemes: the GC|CG and AT|AT curves show the twist dependence of the T_{ij} integrals when the WC pairs GC and AT are treated as fragments, while the G|G and A|A curves were calculated as above, using the single bases A and G as fragments

The CT integrals have been calculated at various levels of theory in recent years, mostly using HF-KTA as reviewed recently [42]. Interestingly, the HF values seem to be overestimated by 40 % when compared with the values calculated on the more accurate CAS-PT2 level [298]. The latter calculations provide an excellent benchmark for other methods and explain the difference between HF and DFT values [144]. For comparison with the ab initio data, the overlap-corrected CT integrals T_{ij} have to be used. As shown in Table 3.2, the T_{ij} calculated by DFTB agree quite well with the CAS-PT2 data,[298] while the larger T_{ij}^0 data resemble the (overestimated) HF values nicely. In general, this holds for the DFT values in Ref. [144] as well – the non-orthogonal CT integrals J are larger than the orthogonalized J'. Anyway, the DFT values do not seem to match better with the CAS-PT2 reference than the SCC-DFTB ones. This illustrates the challenging character of CT integral calculations, in particular for the B-DNA geometries, where these quantities take small values.

In summary, the DFTB-based CT integrals T_{ij} agree well with the accurate CAS-PT2 values, being smaller in magnitude than the widely used HF-based values. Therefore, DFTB can be used for a rapid and reliable evaluation of CT integrals, and several applications follow in the next sections.

3.2.1 Definition of Molecular Fragments

So far, every base has been chosen as a fragment and the CT parameters have been calculated for sets of single bases. This is justified because the HOMO of Watson-Crick hydrogen-bonded base pairs is localized predominantly on the purine base. As shown in Fig. 3.3, the CT integrals do not differ significantly if whole Watson-Crick pairs GC and AT are treated as fragments, in comparison to the choice of single bases as fragments.

The computation of T_{ij} with SCC-DFTB as described in Sec. 2.5.1 is a two-step pro-

Table 3.2: CT integrals T_{ij}^0 and T_{ij} for the intra-strand hole transfer in 5'-XY-3' and 5'-YX-3' in the configuration with twist = 36° and rise = 3.38 Å. Comparison with HF-KTA data from Ref. [39] and CAS-PT2 values from Ref. [298]. Also presented are DFT values J and J' [144]. All values in eV.

X Y	5'-XY-3'						5'-YX-3'					
	T_{ij}^0	T_{ij}	HF	CAS-PT2	J	J'	T_{ij}^0	T_{ij}	HF	CAS-PT2	J	J'
A A	0.013	0.008	0.030	0.004	0.038	0.004						
C C	0.031	0.026	0.041	—	0.042	0.022						
G G	0.087	0.061	0.084	0.051	0.119	0.053						
T T	0.167	0.093	0.158	—	0.180	0.072						
A C	0.044	0.030	0.061	—	0.091	0.042	0.010	0.005	0.029	—	0.008	0.002
A G	0.059	0.037	0.049	0.044	0.013	0.010	0.080	0.052	0.089	0.036	0.186	0.077
A T	0.057	0.028	0.105	—	0.157	0.063	0.116	0.064	0.086	—	0.068	0.031
C G	0.044	0.031	0.042	—	0.026	0.009	0.091	0.057	0.110	—	0.295	0.114
C T	0.044	0.033	0.100	—	0.161	0.055	0.074	0.052	0.076	—	0.066	0.028
G T	0.100	0.049	0.137	0.081	0.334	0.141	0.123	0.070	0.085	0.061	0.044	0.018
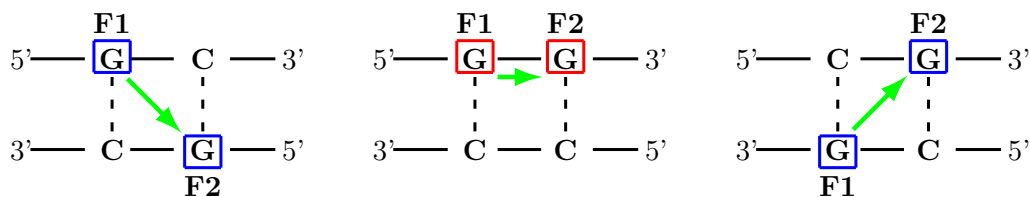
Figure 3.4: G\G denotes the inter-strand 5'−5' hopping (left), G|G the intra-strand hopping (center)[2] and G/G denotes the inter-strand 3'−3' hopping (right).

cedure. First, the isolated fragments, e.g. single DNA bases, are computed to obtain the fragment orbitals (FOs). Then, in a second step, these isolated FOs are used to construct the coarse-grained transfer matrix containing the off-diagnal CT matrix elements T_{ij}. In the following, various possibilities of this proceeding are discussed systematically:

    i First of all, the fragments can be isolated single bases in the gas phase, and the T_{ij} are computed using the Hamiltonian in Eq. 2.125.

    ii Here, the same isolated fragments in the gas phase are chosen, but the T_{ij} are computed using the Hamiltonian in Eq. 2.126, i.e. the remainder of the DNA (other bases and backbone) is included as MM charges.

    iii In the first step, the FOs of each single base are obtained in the presence of MM charges. The second step is the same as in (ii), i.e. using the Hamiltonian in Eq. 2.126.

    iv In this option, each fragment is taken as a whole Watson-Crick base pair. The corresponding FO can be obtained in the first step either in absence (iv-1) or presence (iv-2) of MM charges.

The most exact schemes (iv), which include the major part of the environmental effects already in the initial QM calculation, require the treatment of whole base pairs, leading to the computational complexity increased by a factor of eight, compared to schemes (i)–(iii), due to the doubled fragment size and the $O(N^3)$ scaling of SCC-DFTB. The CT integrals computed for the various options (i)–(iv) are presented in Tab. 3.3.

Interestingly, the effect of hydrogen bonding is captured already by the external point charge calculation (iii), thus yielding very similar results for the couplings. On the other hand, the fragment calculations performed for isolated bases (i) and (ii) cause sizable deviations in some cases. But then, once the fragments were calculated (step 1), the CT integrals were found not to be influenced by the electric field induced by the neighboring bases, as can be seen from the comparison of schemes (i) and (ii) or schemes (iv-1) and (iv-2). The use of Eq. 2.125 or Eq. 2.126 does not make a significant difference for the calculation of the couplings. (Note that this does not hold at all for the site energies

---

[2]Note, for intra-strand hopping between different nucleobases: $X|Y \neq Y|X$.

Table 3.3: CT integrals calculated using various fragment definitions, with and without external charges, for the notation see Fig. 3.4. All values in eV.

| hopping type	fragment definition				
	i	ii	iii	iv-1	iv-2
A|A	0.015	0.013	0.015	0.019	0.015
A/G	0.040	0.039	0.037	0.038	0.044
A|G	0.078	0.078	0.053	0.051	0.045
A/A	0.063	0.060	0.068	0.074	0.080
G\A	0.009	0.010	0.003	0.011	0.015
G\G	0.027	0.023	0.048	0.061	0.066
G|A	0.065	0.058	0.094	0.108	0.108
G/G	0.036	0.037	0.033	0.032	0.037
G|G	0.084	0.077	0.091	0.101	0.097
A\A	0.078	0.077	0.077	0.068	0.070

ϵ_i!) Therefore, single nucleobases can be chosen as fragments without loss of accuracy, as long as the fragments are polarized by the MM charges from the environment.

In summary, a very good agreement was found for the IPs calculated using the ΔSCF approach and the SCC-DFTB Hamiltonian (using the standard basis), compared with experimental data. The calculation of the diagonal terms ϵ_i as the HOMO energy is computationally much more efficient but requires some care. The HOMO energies given by DFT-GGA exhibit a too close energetic spacing, which may lead to orbital swapping due to geometry changes. The calculation of FOs and HOMO energies in the presence of MM point charges Q_A remedies this problem. Therefore, the site energies should be computed using the DFTB standard basis in the presence of external point charges. A slightly less confined basis set is used for the calculation of CT matrix elements. With this basis set, the CT matrix elements match the values yielded by more accurate methods very well. A similar approach is also used to calculate the AO matrix elements of DFTB: the $H_{\mu\nu}$ (Eq. 2.111) are calculated using the compressed, standard DFTB basis, while the atomic site elements ϵ_μ are calculated using the free atomic wavefunctions [272].

3.2.2   Idealized Conformations: A-DNA and B-DNA

In the course of time, the DNA bases assume different relative configurations, described e.g. by the helical parameters twist, rise, roll, slide etc [299]. The dependence of CT integrals on the rise and the twist is studied in detail below. Cytosine and thymine hardly contribute to the hole transfer process due to their large oxidation potentials (see Tab. 3.1), and so the coupling and energetics between adenine and guanine bases are likely to determine the CT rate.

Fig. 3.5 shows the CT integral for the relevant base pair stacks AA, GG, GA and
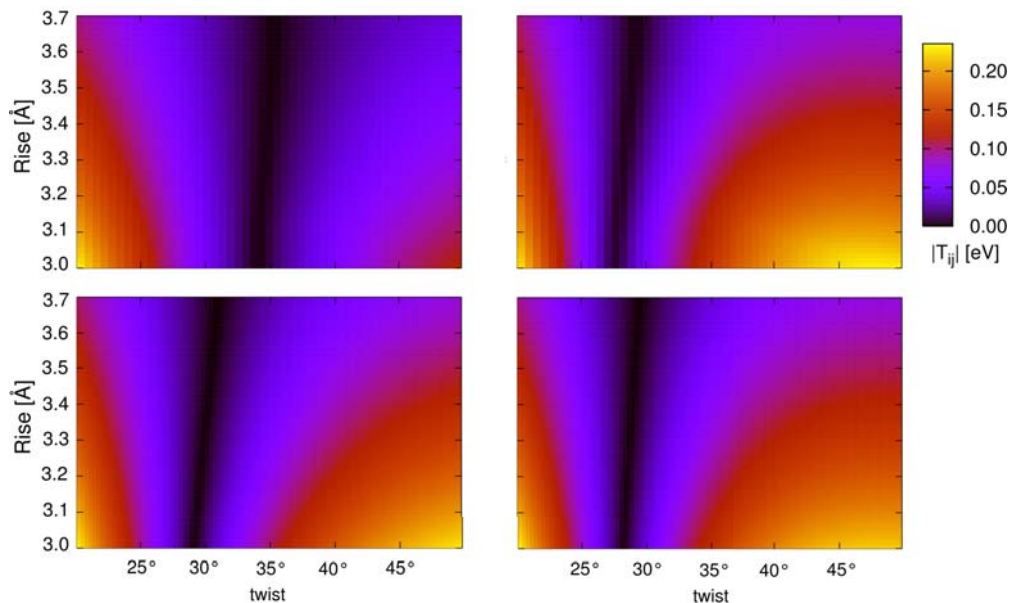
Figure 3.5: CT integral ($|T_{ij}|$) for intra-strand hole transfer in the base-pair steps AA (top left), GG (top right), AG (bottom left) and GA (bottom right); twist $= 20°\dots50°$, rise $= 3.0\dots3.7$ Å.

AG. The AA coupling assumes minimal values around the twist of 34°, a value typical for the B-DNA conformation, while the other combinations assume minima at the twist below 30°, which is close to the A-DNA conformation. In addition, these plots show that there may be non-negligible couplings even for quite large values of the rise, therefore, it is the fluctuation in twist rather than in rise that is responsible for large variance in the CT integrals, in the DNA dynamics.

Table 3.4: T_{ij}^0 and T_{ij} in purine base pair stacks of G and A calculated with DFTB for B-DNA (twist$= 36°$ and rise$= 3.38$ Å) as well as for A-DNA (twist$= 32°$, rise$= 3.20$ Å, roll$= 12°$ and slide$= -1.5$ Å). Parameters were taken from Refs. [300, 301]). Intra-strand coupling prevails if both purine bases are located in one strand, inter-strand coupling otherwise. All values in eV.

| | B-DNA | | | | A-DNA | | | |
| | 5'-XY-3' | | 5'-YX-3' | | 5'-XY-3' | | 5'-YX-3' | |
X Y	T_{ij}^0	T_{ij}	T_{ij}^0	T_{ij}	T_{ij}^0	T_{ij}	T_{ij}^0	T_{ij}
intra-strand								
A A	0.019	0.019	—	—	0.118	0.078	—	—
G G	0.101	0.072	—	—	0.002	0.006	—	—
A G	0.051	0.035	0.108	0.071	0.069	0.046	0.021	0.021
inter-strand								
G C	0.032	0.020	0.084	0.051	0.005	0.001	0.095	0.057
A T	0.074	0.047	0.068	0.032	0.032	0.018	0.211	0.118
A C	0.038	0.025	0.011	0.011	0.015	0.006	0.133	0.080
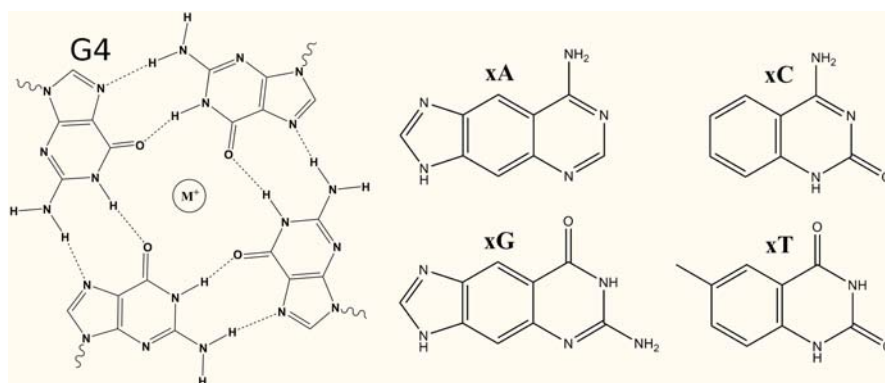
Figure 3.6: Chemical structures of modified nucleobases. Left: Planar guanine quartet, which represents the building block of G4-DNA, containing a coordinated metal ion in its center. Right: Four additional modified nucleobases (xA, xC, xG and xT) for the size-expanded xDNA[3]. The modified xDNA bases are derived from the original four nucleobases by insertion of a benzene ring.

To study the dependence of CT integrals on the overall DNA structure in more detail, ideal A-DNA and B-DNA structures were constructed using the program 3DNA [302]. In accordance with the data presented above, the AA intra-strand couplings for B-DNA in Tab. 3.4 are quite small compared to the GG, AG and GA values (5'-YX-3'). Interestingly, the inter-strand values are still quite large for both A and B forms as shown in Tab. 3.4. On the other hand, the GG coupling vanishes in the A-form, while the AA couplings take rather large values, as can also be inferred from Fig. 3.5.

3.2.3   DNA Derivatives: xDNA and G4

Because of a growing consensus that double-stranded DNA may not be the optimal candidate for molecular electronics applications, several possibilities are explored currently to apply various DNA derivatives in nano-electronic applications. For instance, the G4-DNA wires are formed by stacked guanine quartets (see Fig. 3.6(left)).

These DNA species may be structurally stiffer and more resistant to surface forces than the double helix, while keeping the self-assembling properties of the latter [79]. Evidence of significant polarizability of long G4-DNA molecules, as measured by electrostatic force microscopy, has been recently demonstrated by Porath and coworkers [79]. By contrast, double-stranded DNA molecules co-adsorbed on mica were electrically silent. This firmly suggests that G4-DNA may be a better conductor than double-stranded DNA.

The question is, if a (potentially) enhanced conductance of G4-DNA may be attributed to larger electronic couplings T_{ij}. Therefore, the CT integrals are calculated for an stacked dimer of idealized guanine quartets in dependence on the torsional angle ("twist"). The result is given in Fig. 3.7(left). Although the observed T_{ij} is not distinctly larger than in the case of normal dsDNA (no matter if A or B-like), the larger stability of G4-DNA may still lead to a different dynamical behavior thus make the CT integrals

---

[3]Therefore, xDNA can contain eight different nucleobases A,xA,C,xC,G,xG,T and xT.
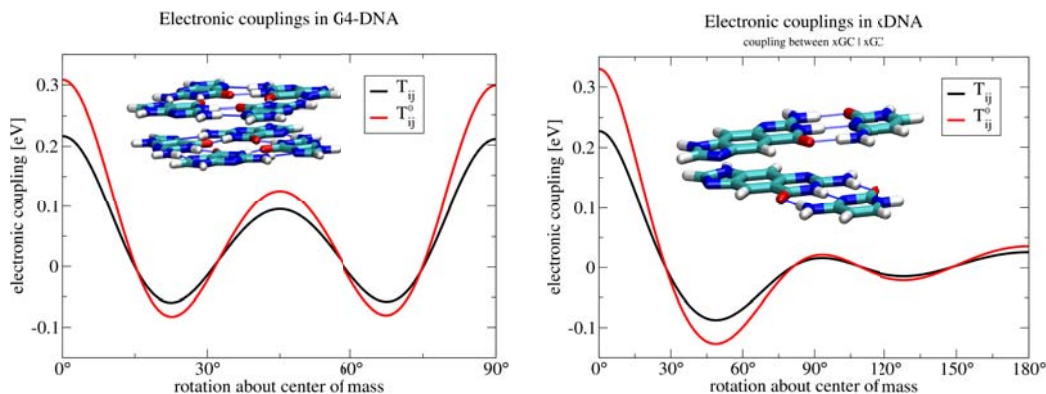
Figure 3.7: CT integrals T_{ij} and T_{ij}^0 in a pair of idealized guanine quartets (left) and in stacked xDNA base pairs xGC|xGC (right); dependence on the torsional angle.

keep larger values on average.

Another interesting modification of natural DNA is the so-called size-expanded xDNA, where an extra benzene ring is inserted into every base molecule, see Fig. 3.6(right). The spatial extension of the bases increases stacking interactions, which in turn leads to higher thermodynamic stability as well as improved fluorescence properties, which may be useful in spectroscopy [303, 304]. Furthermore, xDNA was shown to assemble perfectly with natural DNA [305].

The dependence of electronic couplings T_{ij} on the torsion angle, between two stacked idealized xDNA base pairs (xG–C), is shown in Fig. 3.7. Again, the T_{ij} do not exceed those for natural A-/B-DNA markedly, if compared with Fig. 3.2.

At this point, it is important to be aware that the electronic couplings T_{ij} depend very sensitively on the molecular geometry. Here, idealized model structures for the guanine quartet and the xDNA base pairs (xG-C) were used. The couplings may take quite different values for various structures, e.g. for x-ray structures. This effect was studied in detail recently for electronic couplings between natural and size-expanded DNA [306].

Nevertheless, it is surprising that, at first sight, the electronic π systems in these extended structures do not exhibit a markedly increased overlap, which would result in larger CT couplings. The observed higher polarizability of G4 [79] may then be related only to a higher structural stability, as reported for xDNA[304]. However, these structures may, additionally, assume different average structures, e.g. if sampled in a MD simulation. This would change the average CT couplings significantly. Indeed, first MD simulations indicate slight difference between xDNA and natural B-DNA: the xDNA has slightly smaller values of rise and twist [307]. Whether this leads to a different dynamical behavior and larger average values of CT couplings compared to natural DNA will be discussed in great detail for G4-DNA in chapter 5.

4

Hole Transport in DNA along Classical MD Trajectories

In this chapter, the efficient computation of CT parameter using SCC-DFTB is applied to extended classical MD simulations of DNA molecules in solution. Accordingly, the site energies ϵ_i and the electronic couplings T_{ij} are computed for various snapshots of classical MD simulations accounting for both structural dynamics, i.e. DNA conformation, and environmental effects, i.e. electrostatic interaction of the QM sites with the MM charges of the DNA backbone, the solvent and the counterions.

Eventually, the time series of ϵ_i and T_{ij} are then used to compute the time-dependent transmission function $T(E)$ and the current $I(U)$ based on the Landauer-Büttiker formalism as described in Sec. 2.3.

The CT parameter, hence also the transport properties will be shown to be affected considerably by two major effects: i) the dynamical disorder, i.e. structural and environmental fluctuations, and ii) static disorder, i.e. the DNA sequence. Both effects will be analyzed in great detail below.

4.1 Dynamical and Solvent effects on CT Parameter

4.1.1 Simulation Setup and DNA structures

To study the effect of DNA conformational dynamics as well as the fluctuations of environment on the transport properties of DNA oligomers, classical MD simulations in the ns regime are performed using the AMBER-parm99 force field [233] with the parmBSC0
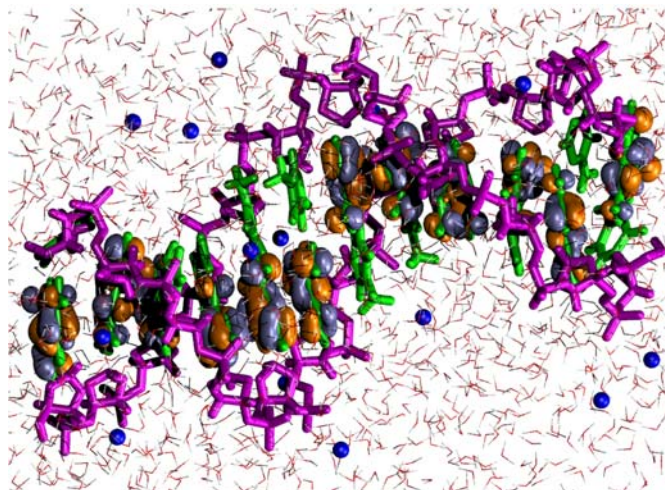
Figure 4.1: Molecular dynamics snapshot of the Dickerson dodecamer DNA: backbone (pink), base pairs (green), solvent molecules (red and white lines), sodium counter ions (blue spheres). Also shown are the corresponding HOMO on each of the base pairs being almost completely localized on the purine bases. DNA backbone, solvent and sodium counterions comprise the electrostatic environment which is described via QM/MM coupling.

extension [235] as implemented in the GROMACS [197] software package. The static idealized A- and B-DNA structures were built with the 3DNA program [302] while the starting structures for the MD simulations were created using the make-na server [308]. MD simulations of six undecamers are carried out using the regular sequences: poly(G), poly(A), poly(AT), poly(GA), poly(GC) and poly(GT) as well as the Dickerson dodecamer[1] [309].

After a standard heating procedure, followed by a 1 ns equilibration phase which was discarded afterwards, 30 ns MD simulations are performed with a time step of 2 fs. The simulations were carried out in a rectangular box using periodic boundary conditions (PBC). The boxes were filled with around 5500 TIP3P [236] water molecules, and 20 (respectively 22 for the Dickerson DNA) sodium counterions for the purpose of neutralization only. Snapshots of the molecular structures were saved every 1 ps, for which the CT parameters were calculated with the SCC-DFTB-FO approach as described in Sec. 2.5.1. To assess the effect of environment, the parameters were computed either with Eq. 2.125 denoted as "vacuo", or with Eq. 2.126 denoted as "QM/MM". The external MM charges render the electrostatic field induced by the DNA backbone, the solvent and the sodium counterions, see also Fig. 4.1.

Additionally, simulations for longer DNA chains of poly(A) and poly(G) are performed, containing up to 31 base pairs, in order to investigate the length dependence of the transport through DNA. To focus on the fluctuations of transport properties in a more time-resolved manner, a 100 ps MD simulation of a poly(G) undecamer is performed with a time step of 1 fs. Here, snapshots were recorded every fs.

---

[1]base sequence: 5'-CGCGAATTCGCG-3'

Table 4.1: Electronic couplings T_{ij} for hole transfer in idealized static A and B-DNA (in vacuo) compared to MD averaged values with standard deviations $\langle T_{ij} \rangle \pm \sigma$ including the QM/MM environment. For helical parameters of the idealized A and B-DNA see ref. [300] and [301]. All values in eV.

	static B-DNA		average MD values		static A-DNA	
	5'-XY-3'	5'-YX-3'	5'-YX-3'	5'-XY-3'	5'-XY-3'	5'-YX-3'
XY	T_{ij}	T_{ij}	$\langle T_{ij} \rangle \pm \sigma$	$\langle T_{ij} \rangle \pm \sigma$	T_{ij}	T_{ij}
intrastrand						
AA	0.013		0.058 ± 0.037		0.070	
GG	0.052		0.029 ± 0.023		0.012	
GA	0.053	0.026	0.034 ± 0.027	0.033 ± 0.028	0.023	0.044
interstrand						
GC	0.017	0.029	0.012 ± 0.012	0.022 ± 0.016	0.006	0.054
AT	0.035	0.031	0.037 ± 0.029	0.045 ± 0.034	0.018	0.107
GT	0.020	0.005	0.016 ± 0.013	0.026 ± 0.023	0.010	0.073

4.1.2 CT Parameter for Static A-DNA and B-DNA

To have an appropriate reference point to estimate the influence of dynamics, the CT parameters are first calculated for ideal A- and B-DNA forms of all the previously listed sequences containing seven Watson-Crick base pairs. These static conformations are characterized by six helical parameters: rise, twist, slide, roll, shift and tilt. Those parameters are taken as 3.2 Å, 32°, −1.5 Å, 12°, 0 Å and 0° for A-DNA; and 3.38 Å, 36°, 0 Å, 0°, 0 Å and 0° for B-DNA, respectively [300, 301]. To calculate the electronic parameters ϵ_i and T_{ij}, only the stacked base pairs are considered without any environment, i.e. the term containing Q_A in Eq. 2.126 is omitted. Small variations occur due to small differences in the base geometries along the DNA chain. The diagonal elements of the Hamiltonian matrix, i.e. the site energies ϵ_i are nearly constant within these idealized heptamers and they take values of −4.53 eV for G-C and −5.21 eV for the A-T base pairs, respectively. The electronic couplings T_{ij}, which are largely determined by the relative orientation of neighboring bases, are shown in Table 4.1 for the ideal A- and B-DNA sequences investigated here.

For B-DNA, the T_{ij} value is much larger in poly(G) than in poly(A). Also, the electronic couplings for poly(GA) and poly(AT) are significantly larger compared with poly(A). In the case of ideal A-DNA, the relation is completely reversed which again confirms that T_{ij} is determined by DNA conformation.

4.1.3 CT Parameter Affected by Dynamical Disorder

The CT parameters are now evaluated along the 30 ns MD trajectories as described above, either omitting the QM/MM term (Q_A) in Eq. 2.126 ("vacuo"), or explicitly

Table 4.2: Average site energies $\langle \epsilon_i \rangle$ with standard deviations for the six regular DNA sequences[2]. The statistics was obtained from 30 ns MD simulations for each DNA sequences, snapshots were saved every ps.

site	AA	AT	GA	GC	GG	GT
1	-5.26±0.38	-5.41±0.38	-4.91±0.40	-4.72±0.40	-4.88±0.39	-4.88±0.39
2	-5.14±0.35	-5.41±0.34	-5.21±0.37	-4.71±0.36	-4.86±0.36	-5.21±0.36
3	-5.21±0.33	-5.37±0.33	-4.77±0.35	-4.72±0.36	-4.84±0.35	-4.86±0.35
4	-5.20±0.33	-5.40±0.32	-5.24±0.32	-4.72±0.34	-4.83±0.36	-5.21±0.33
5	-5.21±0.34	-5.37±0.33	-4.78±0.35	-4.73±0.35	-4.83±0.37	-4.86±0.35
6	-5.23±0.36	-5.39±0.35	-5.25±0.36	-4.66±0.37	-4.84±0.38	-5.19±0.34
7	-5.25±0.38	-5.39±0.38	-4.78±0.40	-4.70±0.38	-4.85±0.40	-4.86±0.38

considering it "QM/MM". The MD is performed for undecamers, however, only seven sites from the core of the helix are considered to avoid end effects. Along these trajectories, the time series $\epsilon_i(t)$ and $T_{ij}(t)$ are computed for the seven sites along the chain.

4.1.3.1 Site Energies ϵ_i

In "vacuo", the site energies ϵ_i fluctuate in the order of 0.15 eV as observed recently in Refs. [147, 164]. The fluctuations originate from the skeletal modes of the DNA bases themselves. The distributions of ϵ_i for a poly(A) heptamer computed in "vacuo" are shown in Fig. 4.2(a). Accordingly, the distributions can be represented almost perfectly by standard Gaussian functions with standard deviations σ of about 0.12 eV for all seven sites.

By contrast, the computation including the full MM environment, i.e. the "QM/MM" case, reveals largely increased ϵ_i fluctuations of about 0.4 eV. Therefore, the corresponding Gaussian functions in Fig. 4.2(c) are broadened considerably compared to the "vacuo" case. These large site energy fluctuations are observed for all the sequences studied here as can be inferred from Tab. 4.2.

Interestingly, the broadening of the ϵ_i distributions in the "QM/MM" case decreases slightly by about 0.06 eV from the outer to the inner Watson-Crick pairs, which is not the case for the data in "vacuo". Thus, the outer base pairs seem to be affected more strongly by the electrostatics of the environment as also the average site energies are shifted slightly to lower energies by about 0.06 eV compared to those for the inner base pairs.

Furthermore, the averages $\langle \epsilon_i \rangle$ in the "QM/MM" case are shifted by 0.2–0.3 eV to lower energies as compared to the "vacuo" values. However, this seems not to be true for poly(A) as no significant site energy shift could be observed.

---

[2]Only the seven central sites are considered.

(a) ϵ_i, "vacuo"

(b) T_{ij}, "vacuo"

(c) ϵ_i, "QM/MM"

(d) T_{ij}, "QM/MM"
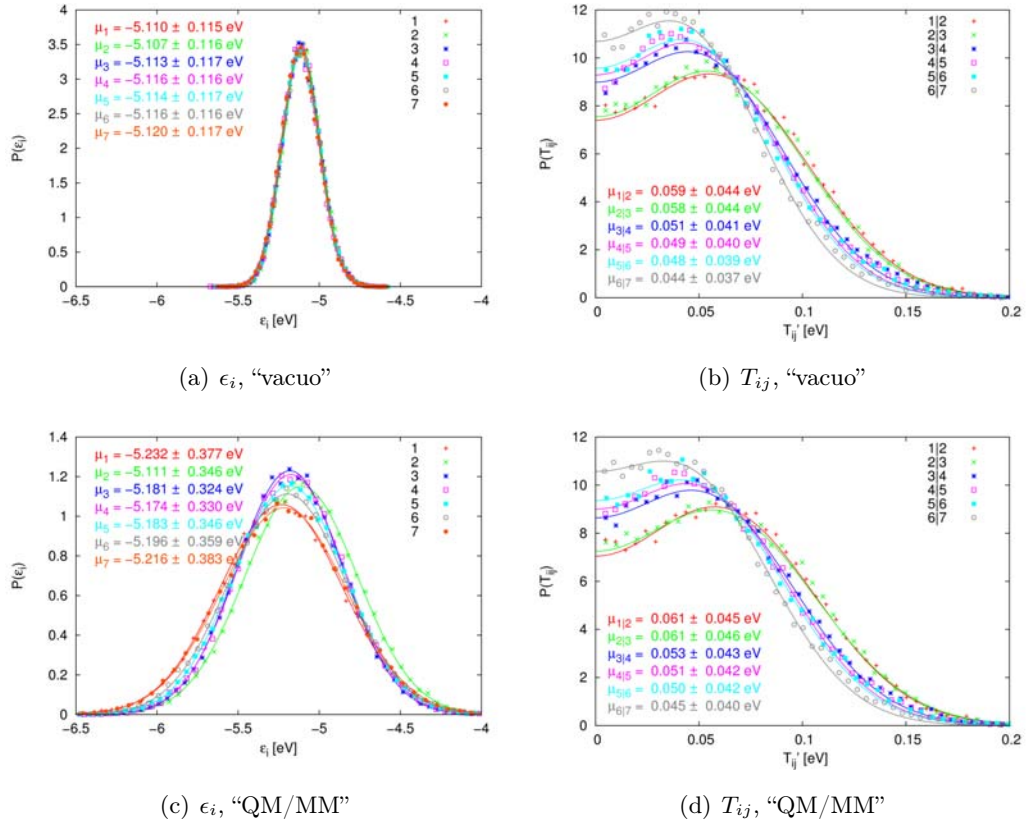
Figure 4.2: ϵ_i (left) and T_{ij} (right) distributions for a poly(A) heptamer obtained from a 30 ns MD trajectory. The electronic parameters were calculated every ps, either with the QM/MM-environment, or in "vacuo". All ϵ_i distributions can be represented by normal Gaussian shapes $p(\epsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\epsilon_i - \mu}{\sigma}\right)^2}$ (solid curves) with expected value μ and standard deviation σ. Interestingly, the broadness of the distributions slightly decreases from the outer to the inner WCPs in the QM/MM case. However, this is not the case for the in vacuo data, so the fluctuations of site energies caused by the surrounding of external charges is bigger for the outer WCPs than for the inner ones. This effect becomes more important with growing length of the DNA. T_{ij} distributions: Since only the absolute values of the couplings are computed, the (in principle normal distributed T_{ij}) can be represented by the fitting function $p(T_{ij}) = \frac{1}{\sigma\sqrt{2\pi}} \left[e^{-\frac{1}{2}\left(\frac{T_{ij} - \mu}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{-T_{ij} - \mu}{\sigma}\right)^2} \right]$.

4.1.3.2   Electronic Couplings T_{ij}

Table 4.1 shows the MD-averaged couplings in comparison to those of the idealized A- and B-DNA structures. In most cases, the MD averages $\langle T_{ij} \rangle$ deviate significantly from those of the ideal sequences. Since the couplings depend sensitively on the DNA conformation, this suggests that the averaged MD structures are significantly different from the ideal ones, as discussed in detail in Ref [164]. The role of fluctuations is further reflected in the variances σ which are of the same order of magnitude as the averages themselves. These results are nearly independent of the interaction with the solvent, indicating that the electronic coupling fluctuations are mainly dominated by the mutual orientation of the base pairs, thus are not sensitive to the electrostatic coupling to the environment [164]. Note however, that the structure of DNA depends sensitively on the solvent pattern, therefore, the coupling parameters may be very different for simulations with varying solvent conditions.

Taking the value of $\langle T_{ij} \rangle$ as a potential indicator for high or low conductance as a first approximation, the results in Tab. 4.1 can be sorted for the various sequences:

$$\mathrm{poly(A)} > \mathrm{poly(AT)} > \mathrm{poly(GA)} > \mathrm{poly(G)} > \mathrm{poly(GT)} > \mathrm{poly(GC)}. \qquad (4.1)$$

This order of T_{ij} will be compared with results for the average transmission below. The comparison may indicate whether the transport on average is dominated by the mean electronic coupling.

4.1.3.3   Analysis of Site Energy Fluctuations

The origin of the site energy fluctuations was analyzed in great detail in Refs. [152, 157, 164]. Some important outcomes in this respect are mentioned briefly in the following. Certainly, these features will affect the Landauer transport calculations in Sec. 4.2. Fig. 4.3 shows a typical time series of site energies for adjacent adenines during a 100 ps simulation of poly(A). Here, the CT parameters were computed each fs. Two major fluctuation periods can be observed. The shorter period of 20 fs can be assigned to the internal vibrational modes of the nucleobases themselves, while the longer period of about 40 fs can be related to solvent modes [164]. Moreover, the site energies of adjacent nucleobases turn out to be correlated, even second-next neighbors are affected. See Ref. [164] for more details.

The ϵ_i fluctuation in time may be transformed into the frequency space, hence exhibit the origin of these fluctuations in more detail. This can be done by using so-called power-spectra, that is the Fourier transform of the auto-correlation function of a certain property in time. Such power-spectra for typical time series of the site energies are given in Fig. 4.4.

There a two major peak regions at about 800 cm^{-1} and 1800 cm^{-1}. Interestingly, the latter one vanishes once the DNA is kept frozen (alternatively kept fixed during the simulations). Accordingly, this fast frequency mode is assigned to the double-bond
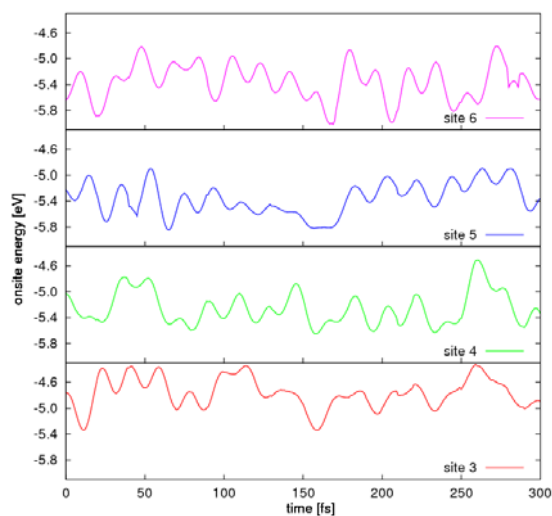
Figure 4.3: Site energies ϵ_i of four central adjacent adenines during a 100 ps MD simulation of a poly(A) heptamer. The parameters were calculated every fs.
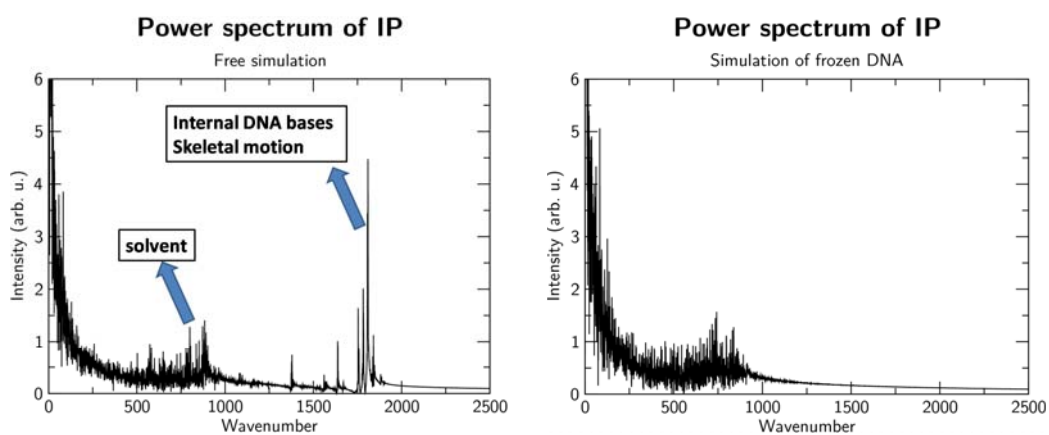
Figure 4.4: Power-spectra, that is the Fourier transform of the auto-correlations function of the site energies ϵ_i, for a free MD simulation (left), and a "frozen-DNA" simulation. Figure has been adapted from Kubař et al., *J. Phys. Chem. B*, **2008**, *112*, 8788-8798.
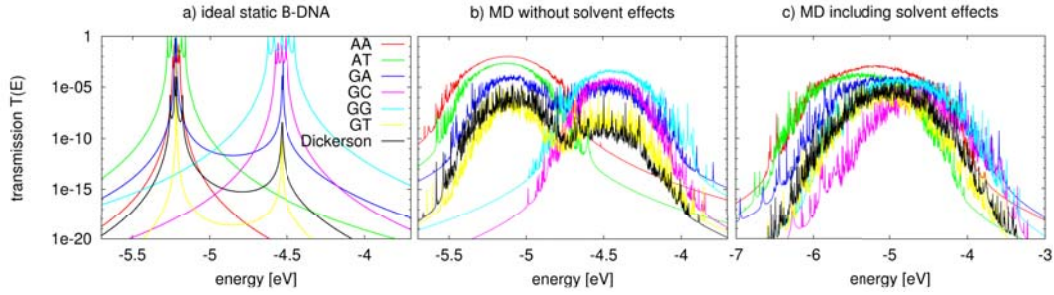
Figure 4.5: Transmission of the ideal chain (a) including dynamical effects (b) and the effect of environment (c) for various DNA sequences. Note the broader energy range in c).

stretch vibrations of the nucleobases. On the other hand, the slower mode at about 800 cm^{-1}, which is also broader, is attributed predominantly to the vibrations of solvent molecules. Nevertheless, a comparison of both spectra in Fig. 4.4 exhibits also some weaker DNA modes in the region between 500 and 1000 cm^{-1}.

4.2 Landauer Transport

For each set of CT-parameters, i.e at each timestep t, the energy dependent transmission function $T(E, t)$ is calculated. Here, the energy E covers the spectral support of the function and it roughly corresponds to the possible charge injection energies.

4.2.1 Idealized B-DNA

Fig. 4.5(a) shows the calculated transmission for various B-DNA sequences. The homogeneous sequences poly(A) and poly(G) show resonances at the eigenvalues of the corresponding Hamiltonian matrices. Due to the very small values of the electronic coupling parameters, these eigenvalues lie very close to the site energies of the respective bases pairs (weak mixing). In heterogeneous sequences like poly(GA) and poly(GT)[3], the transmission is reduced strongly, which is due to the large energy gap between the AT and GC base pair. The irregular Dickerson sequence shows peaks at the energy levels of the AT and GC base pairs as well, though the transmission is reduced considerably due to the large static disorder (different site energies along the chain).

The transmission functions $T(E)$ for the different sequences follows roughly the absolute values of the electronic couplings for the idealized B-DNA in Tab. 4.1. Poly(G) shows clearly the largest transmission followed by poly(AT) and poly(GC). However, this is not true for heterogeneous sequences with a large static disorder, as is the case of poly(GT), for which the transmission is very different, though the electronic couplings (see Tab. 4.1) are rather similar to those of poly(A).

---

[3]Sequences containing both A-T and G-C base pairs will be referred to as heterogeneous, while homogeneous otherwise.

Table 4.3: Maximum current values (at voltage U = 2 V) for seven DNA heptamer sequences. For static B-DNA structures and for the average current $\langle I \rangle$ of the MD structures with and without QM/MM environment. All values in nA.

sequence	static	dyn. vacuo	dyn. QM/MM
AA	202.0	215.7	47.75
AT	1019	39.06	6.404
GA	163.1	1.545	2.787
GC	299.2	0.647	0.086
GG	2511	6.502	1.644
GT	0.093	0.018	0.120
Dickerson[4]	0.023	0.043	0.251

4.2.2 Transport through Fluctuating Bridges: Dynamical Disorder

4.2.2.1 No Solvent Effects

Fig. 4.5(b) shows the transmission for the various sequences. As a result of the broad distribution of the site energies, the transmission spectrum broadens. Further, the dynamical disorder of the site energies increases the transmission of low-conducting (static) structures, while it decreases it for the "high-conducting" ones, as shown in Tab. 4.3.

This is a very interesting result and can be rationalized as follows: The fluctuations of site energies lead to conformations for the "mixed" sequences, such as poly(GA), poly(GT) and Dickerson, where the effective energy gaps become smaller than in the idealized static structures. Therefore, CT-active conformations arise due to the dynamics. On the contrary, the homogeneous sequences become effectively disordered due to the dynamical fluctuations reducing the transmission. This point will be analyzed in more detail below.

4.2.2.2 Influence of DNA Backbone, Water and Counterions

Fig. 4.5(c) shows the transmission of the DNA species in the presence of electrostatic field induced by the environment. Since the environment does not affect the electronic coupling strongly, the main difference from Fig. 4.5(b) arises from the larger fluctuations of site energies. As a result of the wider distribution of site energies as well as the environment-induced energy shifts, the transmission spectra become considerably broader and the clear separation into two transmission subsets found in Fig. 4.5(b) is now blurred out. In principle, the transport properties of the various sequences seem to become more similar to each other than it was the case for the static B-DNA, suggesting that the sequence dependence becomes less important. The current values in Tab. 4.3 also show that the differences in charge transport between the various sequences decrease

---

[4] 8 base pairs instead of 7

when including the dynamical and environmental effects. For static B-DNA structures, there is a difference of $4 - 5$ orders of magnitude in the current between the poly(G) and the Dickerson sequence. However, this sequence dependence decreases dramatically including the full dynamics (dyn. QM/MM). Here, the current values are in a range of $1 - 2$ orders of magnitude. This effect may be related with the experimental results obtained in Refs. [61] and [55]. Accordingly, very similar currents were obtained in both experiments despite the completely different sequences.

Interestingly, poly(A) shows the largest transmission, in contrast to the static case where poly(G) is better conducting. This can be related to the changes of molecular conformation, which make the average coupling increase. Indeed, there is a strong relationship between the average electronic couplings $\langle T_{ij} \rangle$ in Tab. 4.1 and the average transmission $\langle T(E) \rangle$ in Fig. 4.5(c). The order of the $\langle T(E) \rangle$ maxima or, more quantitatively, the $I(U)$ values in Tab. 4.3 for the studied sequences is:

$$\text{poly(A)} \gg \text{poly(AT)} > \text{poly(GA)} > \text{poly(G)} \gg \text{poly(GT)} > \text{poly(GC)}. \tag{4.2}$$

Thus, is in a good agreement with the order of the corresponding average electronic couplings in Eq. 4.1. This shows that the $\langle T_{ij} \rangle$ is a good indicator which reflects the transport properties of DNA species, characterizing the entire ensemble of molecular structures with their very different individual T_{ij} values.

The transmission of the heterogeneous species like poly(GA), poly(GT) and the Dickerson structure are particularly interesting as they contain both types of Watson-Crick Pairs A-T and G-C. Here, the transmission increases substantially compared to the idealized, static case, indicating that the fluctuations of site energies may lead to conformations with smaller effective site energy disorder. The increased conductance of the heterogeneous sequences can also be seen in Tab. 4.3. For all three of them, even for poly(GA), the average current $\langle I \rangle$ becomes larger if the electrostatic environment of the DNA is considered. This confirms the previous statement. Moreover, it suggests that besides the average electronic couplings also the strong fluctuations of site energies seem to play a very important role for charge transport in DNA.

4.2.3 Role of Disorder and Coherent Motion

The width of $\langle T(E) \rangle$ in Fig. 4.5(c) shows that the energy range for transmissive conformations becomes much larger than it was the case in Fig. 4.5(b) and even more so in Fig. 4.5(a). For instance, the energy range for poly(A) where $\langle T(E) \rangle$ is larger than 10^{-5} is 1.7 eV, compared to 0.7 eV without environment and only 0.05 eV in the static case [5]. However, this indicates that the site energy fluctuations are not always completely disordered and therefore destructive, i.e. sometimes they can lead to transmissive conformations for very different energies. The effect can also be followed by computing

---

[5] Although one should notice that the maxima in the transmission decrease from Fig. 4.5(a) to Fig. 4.5(b) and also from Fig. 4.5(b) to Fig. 4.5(c) at least for most of the studied sequences.
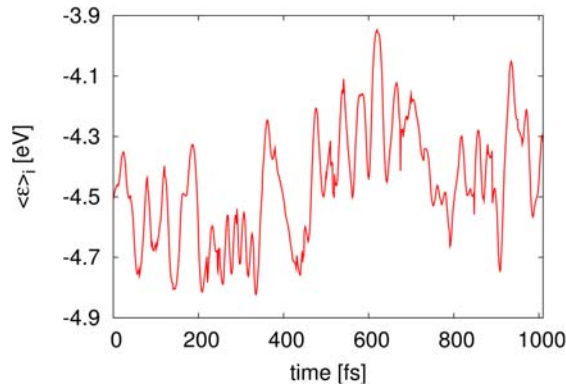
Figure 4.6: Time-dependent average over the site energies $< \epsilon(t) >= (1/N) \sum_j \epsilon_j(t)$ obtained from a 100 ps MD simulation of a poly(G) heptamer ($N = 7$). Snapshots were recorded every fs.

the average site energy $\langle \epsilon(t) \rangle = (1/N) \sum_j \epsilon_j(t)$ along the poly(G) heptamer ($N = 7$) for every snapshot along the MD simulation as shown in Fig. 4.6.

The site energies seem to undergo collective fluctuations in the range of 1 eV, which again explains the large broadening of the transmission function, i.e. conducting conformations at various energies are explored during the time evolution of the system. These calculations demonstrate that it is not meaningful to consider averages for site energies (ionization potentials). Therefore, the charge transport properties of a DNA species cannot be traced back to the averaged site energies in any way.

To analyze the effect of disorder for both ϵ_i and T_{ij} in more detail, additional sets of calculations are performed (Fig. 4.7).

The average transmission $\langle T(E) \rangle$ was computed for all nine possible combinations, where respectively, the ϵ_i or T_{ij} were i) fixed at their idealized B-DNA values, ii) fixed at their MD averages or iii) free to fluctuate. The calculations suggest that the time averages of T_{ij} determine the average transmission but the fluctuations of electronic couplings have only a small impact. On the other hand, taking the averages of the site energies do not seem to be appropriate for such kind of transport calculations. Instead, the results indicate that the fluctuations of ϵ_i have a large impact on the average transmission. They can either hinder or facilitate the transport, which depends on the "initial" static disorder.

In conclusion, the MD time series of T_{ij} may be substituted by the respective mean values for the computation of transmission. Notwithstanding, the transmission at each time step still depends sensitively on the corresponding electronic coupling. Though, the average transmission, i.e. taken over an ensemble of conformations, is dominated by the mean electronic coupling. On the contrary, the fluctuations of site energies cannot be replaced by the average site energies, because the transmission is dominated by few configurations with CT-favorable site energies. This will be analyzed in more detail in Sec. 4.4.
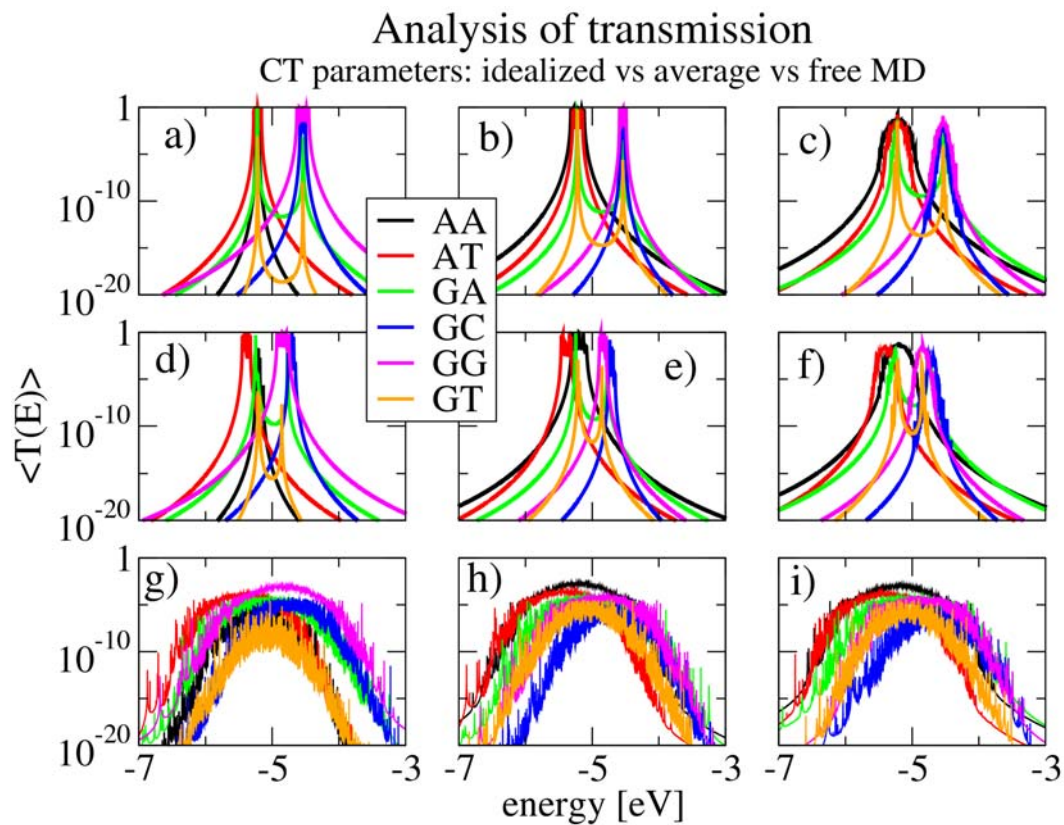
Figure 4.7: Transmission $\langle T(E) \rangle$ for various sequences using all possible combinations of idealized (id), MD-averaged (av) and free MD values (md) for site energies and electronic couplings, respectively. a) $\epsilon_i = \mathrm{id}, T_{ij} = \mathrm{id}$, b) $\epsilon_i = \mathrm{id}, T_{ij} = \mathrm{av}$, c) $\epsilon_i = \mathrm{id}, T_{ij} = \mathrm{md}$, d) $\epsilon_i = \mathrm{av}, T_{ij} = \mathrm{id}$, e) $\epsilon_i = \mathrm{av}, T_{ij} = \mathrm{av}$, f) $\epsilon_i = \mathrm{av}, T_{ij} = \mathrm{md}$, g) $\epsilon_i = \mathrm{md}, T_{ij} = \mathrm{id}$, h) $\epsilon_i = \mathrm{md}, T_{ij} = \mathrm{av}$, i) $\epsilon_i = \mathrm{md}, T_{ij} = \mathrm{md}$.
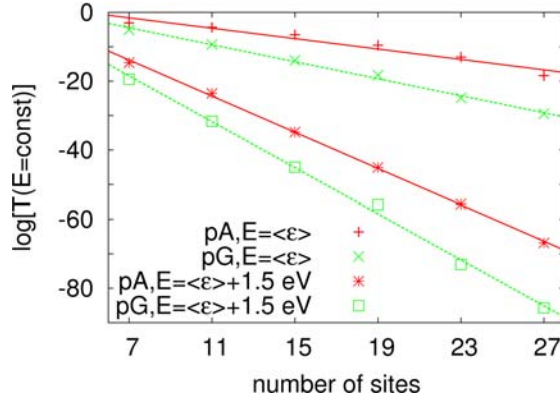
Figure 4.8: Length dependence of $\langle T(E) \rangle$ for poly(G) and poly(A). Shown are logarithmic transmission values for various DNA length, i.e. number of sites N at two constant arbitrary energies which are the average site energies $\langle \epsilon \rangle$ and $\langle \epsilon \rangle + 1.5$ eV, for poly(A) and poly(G), respectively. The data points were fitted by functions of the form $T(E = const) = Ae^{-\beta L}$ where $L = Nd$ (d is the stacking distance ~ 3.4 Å) and β describes the decay rate of transmission. For poly(G) β is 0.36 Å^{-1} and for poly(A) 0.22 Å^{-1} at $E = \langle \epsilon \rangle$. However, if an energy gap of 1.5 eV to $\langle \epsilon \rangle$ is present both β-values increase to 0.77 Å^{-1} and 0.98 Å^{-1} for poly(A) and poly(G), respectively. In both cases the exponential decay of transmission in poly(G) is stronger than in poly(A). The complete $\langle T(E) \rangle$ curves for the different lengths can be found in Fig. C.2 in the appendix.

4.2.4 Length Dependence

In this section, the length dependence on the transport properties of DNA are discussed. For this purpose, 30 ns MD simulations of poly(G) and poly(A) DNA with up to 31 base pairs were carried out. The average transmission values for various DNA length at two constant arbitrary energies is shown in Fig. 4.8.

The first energy is simply the average site energy $\langle \epsilon \rangle$, while the second is separated from the former by a gap of 1.5 eV. The calculations were carried out for both poly(A) and poly(G). Clearly, there is an exponential decay of transmission with increasing number of sites, which can be expected as more sites lead to an increase of the dynamical disorder. Because of that, the probability of conformations with sufficiently large T_{ij}, and at the same time, similar ϵ_i along the chain becomes smaller with increasing number of sites. The exponential decay of transmission has the form $T(E = const) = Ae^{-\beta L}$ where β describes the intensity of decay and the distance $L = Nd$ (N is the number of base pairs and d is the stacking distance ~ 3.4 Å). However, the exponential decay of transmission is considerably stronger for energies which are separated from $\langle \epsilon \rangle$ by a large gap. From Fig. 4.8, decay parameters β for the latter case ($E = \langle \epsilon \rangle + 1.5$ eV) can be extracted, which are 0.77 Å^{-1} for poly(A) and 0.98 Å^{-1} for poly(G), respectively. On the other hand, substantially smaller β-values for the former case ($E = \langle \epsilon \rangle$) are obtained. Here, β takes values 0.22 Å^{-1} for poly(A) and 0.36 Å^{-1} for poly(G) indicating some kind of lower limit for the decay of transmission. Recently, Berlin et al. reported, that there is a minimum value of the decay parameter β in the tunneling regime which is 0.2–0.3 Å^{-1} [153].

The length dependence of transmission is significantly stronger in poly(G) than in poly(A) irrespective of the given energy. This may be due to the different average electronic couplings, which is 0.03 eV for poly(G) and nearly 0.06 eV for poly(A) (see Tab. 4.1). No significant changes in the electronic parameters ϵ_i and T_{ij} with increasing DNA length were found. A statistical analysis is given in Fig. C.1 in the appendix.

Nevertheless, the strong exponential distance dependence of the CT for such long DNA molecules (i.e. up to 31 base pairs) as observed above, must be artificial. More importantly, it reveals the limits of this simple coherent transport model based on the Landauer-Büttiker formalism, as it fails to describe the weak distance dependence of the CT in DNA over large distances. It is well-known from experiment [24] as well as from theoretical calculations [37, 42], that there is a crossover of the superexchange mechanism with its strong exponential distance dependence, and the weak distance dependent hopping mechanism characterized by a falloff parameter $\beta < 0.1$ [34]. The latter mechanism cannot be described adequately with standard Landauer theory of coherent transport, thus results for longer DNA molecules should be interpreted with caution.

4.3   Role of Proper Sampling and Correlation on CT Efficiency

Many theoretical studies of DNA conduction have approximated the dynamical and solvent effects using statistical models [149, 168–171]. Here, the results from the MD simulations of a poly(A) heptamer are compared to those obtained by two statistical models. As has been shown before [164], the fluctuations of site energies exhibit strong correlations between the sites, which may have an impact on conduction. These correlations may be induced by the motion of solvent water molecules, which have been shown to introduce fluctuations of the site energies with a period of 40 fs. The water modes introduce a periodically fluctuating potential at the DNA base sites, which lead to the fluctuation of site energies. A possible explanation for this correlation effect is that neighboring sites are affected by similar electrostatic potentials. The distributions of ϵ_i and T_{ij} obtained from the classical MD simulations have a standard Gaussian shape as shown in Fig. 4.2. To discard the correlations of adjacent sites along the DNA chain, a statistical model is defined, which is referred to as "PDF". In this model, the site energies are taken randomly from normal distributions $P(\epsilon_i)$ for each site i of the chain. These distributions are obtained directly from the corresponding MD simulation of the respective DNA species. This ensures that a realistic site energy distribution for each site is used, while time correlations between the sites are not present.

The second statistical model is based on the Anderson model of disorder [310], where the site energies are randomly drawn from a square-box distribution of width w with uniform probability $P(\epsilon) = 1/2w$. The box width is $w = \sqrt{3}\sigma$, where $\sigma(\epsilon)$ is the standard deviation of site energies resulting from the MD simulations ($\sigma \sim 0.4$ eV).

To make the results comparable, the electronic couplings are kept fixed $T_{ij} = 0.05$ eV,

(a) ϵ_i models

(b) T_{ij} models
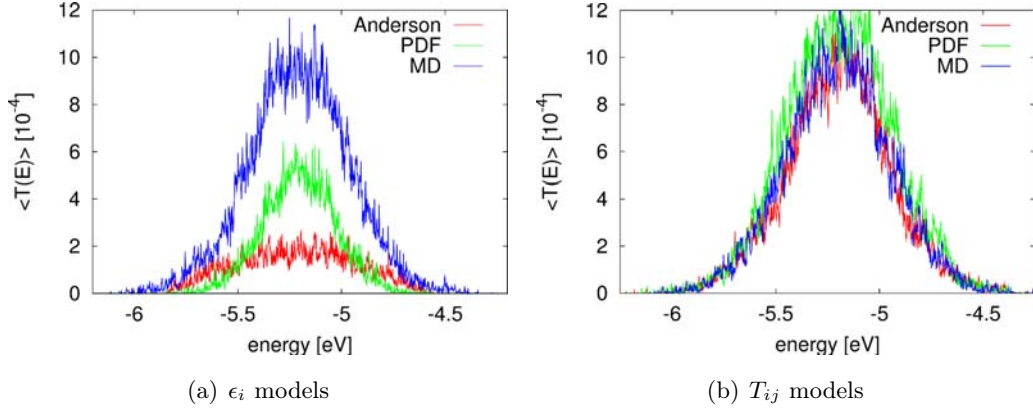
Figure 4.9: Comparison of $\langle T(E) \rangle$ for the MD simulation of a poly(A) heptamer with two statistical models. Left panel: the electronic couplings for the three models are set to 0.05 eV. The average transmission function is calculated for site energies from the MD simulations time series (blue); for site energies drawn from the respective probability distribution functions on each site (green); and the Anderson model (red) where all site energies are drawn randomly from a square-box distribution. Right panel: Now the original MD time series of site energies is used, i.e. the same for the three models, while $\langle T(E) \rangle$ is calculated for electronic couplings T_{ij} from the original MD time series (blue); for T_{ij} drawn from their respective probability distribution functions (green); and the Anderson model (red). The corresponding probability distribution functions for ϵ_i and T_{ij} are shown in Fig. 4.2.

since as shown above, the fluctuations of T_{ij} have only a minor influence on the conduction. The average transmission of the poly(A) heptamer for the two models and the full MD simulation is shown in Fig. 4.9(a).

The three spectra have the same position of maxima, since the transmission functions originates from the distributions of site energies, which are all symmetric with the same mean value. Clearly, the Anderson model largely underestimates the transmission, showing the importance of proper MD sampling of the right distribution of charge transfer parameters. However, also the PDF model based on the MD trajectories but neglecting non-local correlations reproduces only half of the transmission as found for the full MD simulation. The corresponding $I - U$ characteristics is shown in Fig. 4.10. As has been discussed in detail previously [164], the fluctuations of site energies of neighboring sites are highly correlated with a correlation coefficient of about 0.7. Even second neighbors exhibit a significant correlation. Therefore, due to this correlated motion of the site energies a picture emerges, where the site energies in a region of 3-5 bases undergo concerted motion, which is not taken into account in the PDF and Anderson models. This correlated dynamics leads to a distribution of site energies, which has less disorder on average.

The choice of T_{ij} has only a minor effect on the transmission as shown in Fig. 4.9(b). There is no correlation between the couplings which is important for transport. The comparison of both plots shows that the choice of T_{ij} at a reasonable constant value (here 0.05 eV, roughly the MD average of poly(A)) gives nearly the same average trans-
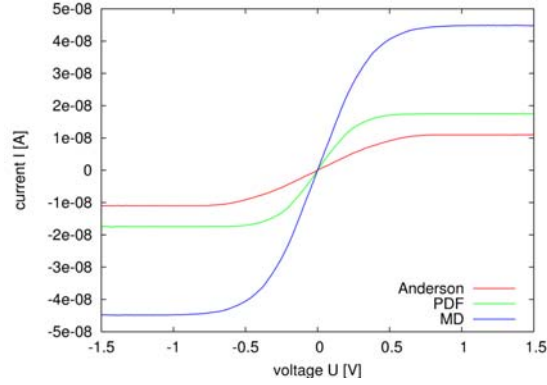
Figure 4.10: Corresponding $I(U)$ characteristics of the site energy models in Fig. 4.9(a).

mission as taking the real time series from MD. In conclusion, the charge transport efficiency depends strongly on the site energies, which are determined predominantly by the environment. Furthermore, our results indicate that DNA charge transport models in which the energetics is based on statistical distributions do not include non-local correlation between adjacent sites, and therefore miss an important contribution to the charge transport.

4.4 Conformation Analysis

In this section, the effect of dynamical disorder on the transmission is analyzed for both electronic parameters ϵ_i and T_{ij} along the DNA chain in more detail. One of the goals is to estimate the amount of conformations which are active for charge transport. The so-called coherence parameter has been analyzed recently in the context of electron transfer in proteins in aqueous solution, in order to understand the effect of fluctuations on charge transfer in more detail [159]. The concept of a coherence parameter is adapted here, and applied to the transmission function and the collective couplings along the DNA chain. The results can be summarized as follows: i) the coherence parameter for the transmission function is much smaller than 1, which indicates that fluctuations dominate the transport, i.e. strong dynamical disorder; ii) the coherence parameter for the transmission is larger in poly(A) than in poly(G) which may be the source of better conductance in poly(A) iii) most of the dynamical disorder is due to the site energies; iv) coherence parameter for the collective electronic couplings along the DNA chain is much closer to 1 indicating once again that the averages of T_{ij} still affect the charge transport in contrast to the ϵ_i.

For a further analysis of the role played by conformational dynamical disorder, two simple measures are proposed for the site energy and hopping disorder along the DNA

(a) maximum transmission

(b) number of conformations
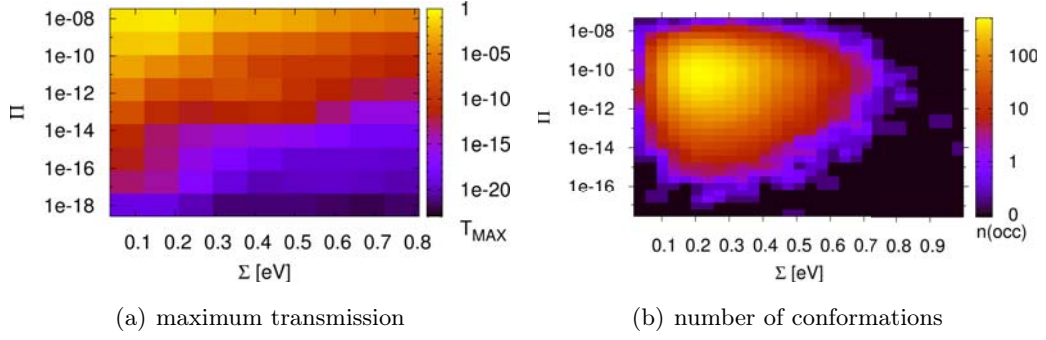
Figure 4.11: Statistical analysis of T_{MAX} in a poly(G) heptamer. Data obtained from a 30 ns simulation, i.e. electronic parameters for every ps (30,000 DNA conformations). T_{MAX} depending on Σ and Π (left); number of conformations found in a given interval of Σ and Π (right).

chain:

$$\Sigma \;=\; \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\epsilon_i - \langle\epsilon\rangle_N)^2} = \sqrt{\langle\epsilon^2\rangle_N - \langle\epsilon\rangle_N^2} \tag{4.3}$$

$$\Pi \;=\; \prod_{i=1}^{N-1} T_{i,i+1} \tag{4.4}$$

The standard deviation Σ is calculated for the ϵ_i along the chain and has an evident meaning. Large values of Σ indicate large differences of neighboring site energies. Note that the index N in $\langle\epsilon\rangle_N$ and $\langle\epsilon^2\rangle_N$ means that averaging is performed for the N sites along the chain. The parameter Π is motivated by the form of the Green's function matrix element $G_{1N}(E)$ required to calculate the transmission function, which scales approximately as the product of electronic couplings. Thus, this quantity determines the transmission efficiency of the system; small values of Π account for conformations with small couplings along the DNA chain. In order to reduce the complexity of further analysis, an additional quantity T_{MAX} is introduced, which is simply the maximum of a given transmission function $T(E)$. Note that the value T_{MAX} can be located anywhere within the respective energy range. All three parameters Σ, Π and T_{MAX} are now calculated for 30,000 snapshots along the 30 ns MD trajectory of a poly(G) heptamer.

The results are shown in Fig. 4.11(a). It appears that none of the measures Σ and Π alone is able to describe the conformations of high conduction, but both seem to contribute nearly linearly to the transmission (note the logarithmic scales for Π and T_{MAX}). However, small Σ and large Π values are required for the transport active conformations.

Moreover, Fig. 4.11(a) reveals that T_{MAX} depends more strongly on Π than on Σ. For instance, if Π is kept fixed at 10^{-8}, then the maximum transmission T_{MAX} is still at least 10^{-7} for all values of Σ. On the other hand, keeping the parameter Σ fixed makes T_{MAX} decrease to almost 10^{-17}. This is true even for very small values of Σ. Fig. 4.11(b) shows the corresponding occupation plot, quantifying how many conformations exhibit a certain combination of Σ and Π. It seems that the number of transport active conformations with

(a) T_{MAX} at fixed Σ (b) conformation analysis
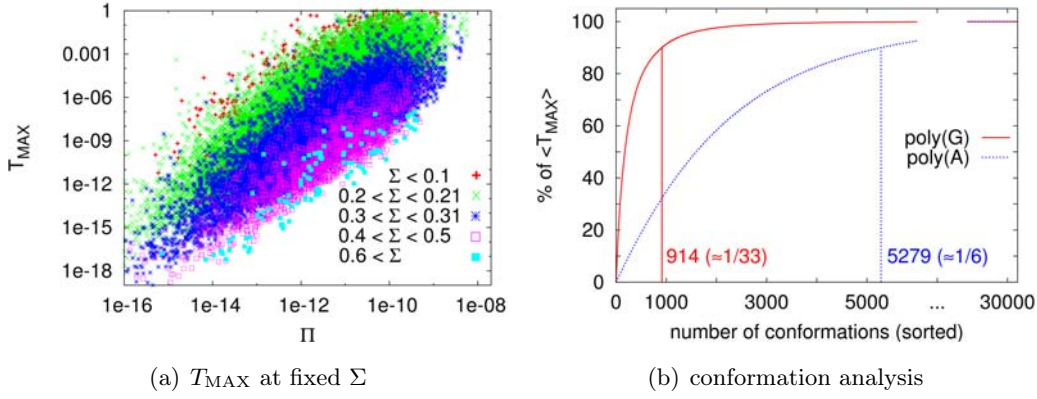
Figure 4.12: Left panel: Plot of T_{MAX} depending on Π for fixed values of Σ, based on the same data as used in Fig. 4.11. Right panel: Conformational analysis, that is to assess the amount of conformations that make up 90% of the average maximum transmission $\langle T_{MAX} \rangle$. Comparison between poly(G) and poly(A) heptamers. The T_{MAX} values are sorted beginning with the largest.

Table 4.4: Correlation coefficient $\rho(X, Y)$ of $\log(T_{MAX})$ with $\log(\Pi)$ and Σ

X	Y	QM/MM	vacuo	fixed Σ	fixed Π
$\log(\Pi)$	$\log(T_{MAX})$	0.76	0.79	0.84	-
Σ	$\log(T_{MAX})$	-0.47	-0.33	-	-0.68

appropriate electronic couplings and site energies is very small. The most conformations have Π values of about 10^{-10} and Σ values of about 0.25 eV, thus may be characterized as "CT-silent".

To study the effects of both parameters Σ and Π in detail, T_{MAX} is plotted against Π for fixed values of Σ. Fig. 4.12(a) again indicates that there is a linear relation between $\log(\Pi)$ and $\log(T_{MAX})$, although the data points are spread largely.

The data sets are nearly parallel to each other for the various Σ, merely shifted along the y-axis due to the different Σ values. A small Σ means less dynamical site energy disorder and leads to larger T_{MAX} values. As can be seen from the values in Tab. 4.4, the correlation coefficient between $\log(\Pi)$ and $\log(T_{MAX})$ for the full MD takes a quite large value of 0.76.

The correlation even increases to 0.84 when Σ is kept fixed. On the other hand, keeping Π fixed leads to a marked anti-correlation of -0.68. This clearly demonstrates that the instantaneous charge transport for a single DNA conformation is indeed affected very strongly by the actual electronic parameters, especially the electronic couplings.

Nevertheless, it is important to stress that these results are by no means in contradiction to those obtained above. The transport on average was shown to depend only on the average electronic couplings.

There are very few charge transport active conformations as indicated in Fig. 4.11(b). The following analysis is carried to put this in a more quantitative way:
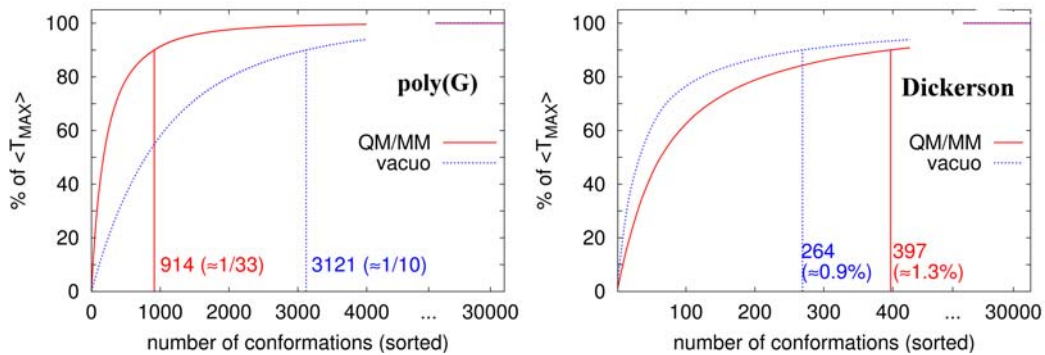
Figure 4.13: Conformational analysis: The amount of conformations that make up 90% of $\langle T_{MAX} \rangle$, based on the calculation of electronic parameters with QM/MM-environment and in vacuo. Comparison between the homogeneous poly(G) heptamer (left) and the heterogeneous Dickerson sequence (right). The T_{MAX} values are sorted beginning with the largest.

1. The T_{MAX} values for the whole MD trajectory (30,000 conformations) are sorted, beginning by the largest.

2. The total average transmission maximum $\langle T_{MAX} \rangle$ is calculated, i.e. over the whole ensemble.

3. It is determined how many of the best-transmissive conformations are necessary to obtain 90% of $\langle T_{MAX} \rangle$.

Fig 4.12(b) shows the conformational analysis for the poly(A) and poly(G) heptamers. In the latter case, only 914 of 30,000 conformations are active for charge transport, i.e. on average only every 33rd conformation. In poly(A), the number of significantly contributing structures increases dramatically to 5279 (on average every sixth). Thus, the difference in conductance between poly(A) and poly(G) may be related to these results, which are in a good agreement with the calculations of respective coherence parameter. The average transport in both sequences is dominated by the dynamical disorder, i.e. only few conformations contribute to transport. However, the dynamical disorder in poly(G) is considerable larger than in poly(A). For the charge transport active conformations, the mean values of Σ and Π are 0.20 eV and 1.34×10^{-7} for poly(A), and 0.19 eV and 3.23×10^{-9} for poly(G), respectively.

Interestingly, dynamical fluctuations can hinder or support charge migration. The latter occurs in cases where the transmission is small in the static structure; structural fluctuations give rise to transmission-active conformations with small Σ values. For instance, the Dickerson sequence, where the average T_{ij} are not small (see Tab. 4.1), profit very much from the fluctuations. As shown in Fig. 4.13, the interaction with the environment suppresses the CT-active conformations and thereby the transmission in poly(G).

On the other hand, the number of CT-active conformations in the Dickerson sequence is increased. This can be seen in the Σ-value, which is 0.35 eV for the static structure

and decreases to 0.25 eV as an average of the CT-active conformations in the QM/MM dynamics. Note, that due to the initial differences in the static disorder (different site energies along the chain) the number of transport active conformations in poly(G) is generally larger than in the heterogeneous Dickerson sequence.

4.5 Time Scales

Up to now, the primary focus was centered on a statistical analysis of the transmission function, which is averaged over snapshots from 30 ns MD simulations. The analysis clearly showed the existence of charge transport active conformations. Notwithstanding, an important issue is the time scale separability of ionic and electronic motion relevant for the transport. Therefore, the question may arise about how parameter averaging should be performed.

Landauer and Büttiker [311] analyzed a fluctuating model potential in order to determine tunneling traversal times, i.e. the time a particle spends in passing a barrier with height V_0 and fluctuation with $V_1 cos(\omega t)$. For low frequencies ω, the particle experiences a static barrier, while high frequencies (compared to the tunneling time) lead to a time-averaging of the potential, i.e. the particle effectively sees the mean potential $V_0 + \langle V_1 \rangle$. A detailed analysis of ionic (τ_1) and electronic (τ_2) time scales for the charge transfer in donor-bridge-acceptor was given in Ref. [153].

For slow dynamic fluctuations (structural fluctuations with characteristic time τ_1), a non-exponential time evolution of the CT process has been reported, caused by the configurational averaging of the electronic parameters. The averaging as performed in this work would address this situation. However, few CT active conformations would enhance/dominate the whole charge transfer. A similar picture emerged e.g. from the analysis of electron tunneling through water [312].

On the other hand, for fast dynamic fluctuations, $\tau_1 \ll \tau_2$, the self-averaging of the electronic parameters as in Ref. [311] leads to a static correction to the time independent rate constant [153].

Therefore, the question arises how to average the transmission in the case of DNA where the fluctuations are large and perturbation theory may not be applicable. Consider e.g. the case of poly(A): There is no barrier for the idealized B-DNA conformation, i.e. there is no V_0 as in the model of Landauer and Büttiker [311]. Averaging increases the transmission continuously with the applied averaging time, as shown in Fig. 4.14. For long averages, the transmission of the idealized structure is retained. This could be expected since the time average of the site energies is nearly equal to those of the idealized structure, thus long time averages lead to a barrier-free situation.

On the contrary, averaging decreases the transmission continuously in poly(GT), as shown in Fig. 4.14. Again, the averages resemble the situation of the idealized structure, thereby deleting the CT active conformations. These two examples show, that fluctuations can act in different directions, increasing or decreasing the transmission or CT rate.
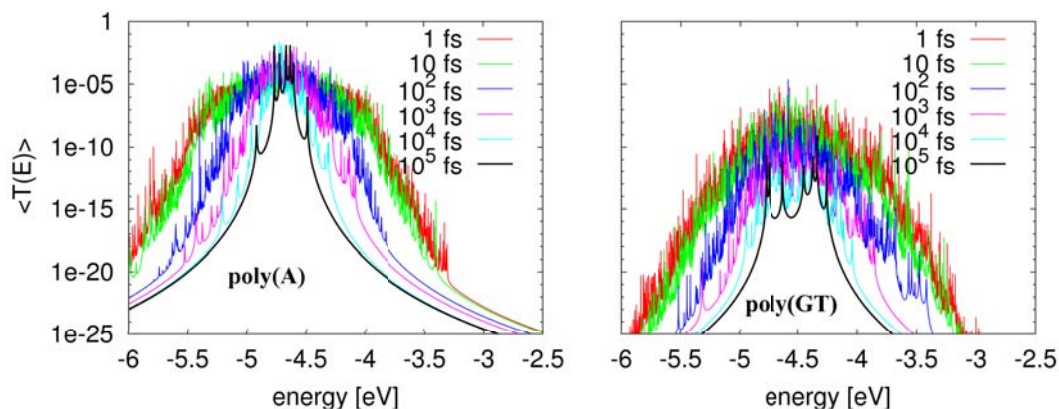
Figure 4.14: Average transmission for various sets of averaged electronic parameters, for poly(A) (left) and poly(GT) (right). Both of them obtained from 100 ps MD data with a timestep of 1 fs.
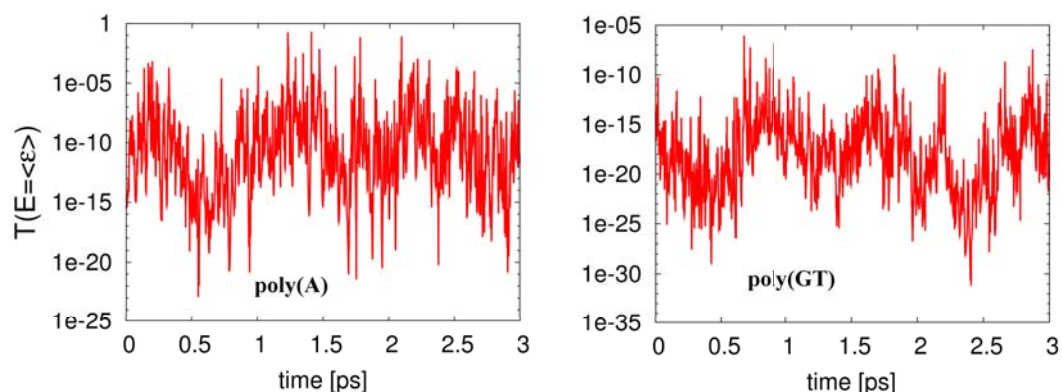

Figure 4.15: Snapshot of a 3 ps time series for the transmission at E=$\langle \epsilon \rangle$ for poly(A) (left) and poly(GT) (right). Data is based on the simulation as used in Fig. 4.14.

Troisi and co-workers [313] derived an expression for the rate constant for CT through fluctuating bridges. The correction to the static limit takes into account the coherence parameter in a unique way. This, however, may not be applicable for DNA as the effect of fluctuations does not affect the CT in the same way.

The averaging problem shows that time scales should be carefully discussed in the context of CT in DNA. The time scales for ionic motion, which are important for CT, are not easy to determine. Fourier transform of the autocorrelation function of time series of CT parameters showed a multitude of modes [164]. The DNA-base skeletal motion with a period of 20 fs, the "water" mode with a period of 40 fs, and the motion of counterions and the DNA backbone, which are in the ps range or even longer. It is difficult however, to asses their relative importance on the transmission. Fig. 4.15 shows a time series of the transmission for poly(A) and poly(GT).

Larger fluctuations in the ps regime are modulated by much smaller fluctuations in the fs regime. The fast fluctuations span a rather large range of transmission. However, it seems that the maxima of the ps time fluctuations indicate the existence of charge transfer active conformations, which are persistent for a short time. In this time, an averaged
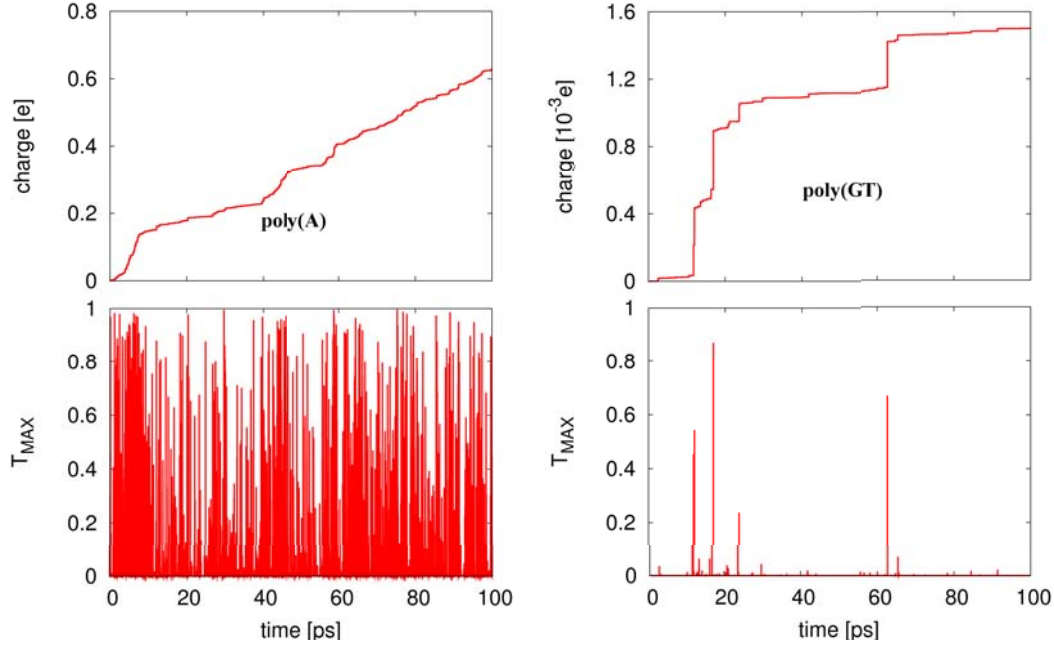
Figure 4.16: Transferred charge in time via $q = \int I \, dt$ as obtained from the current time series (at U = 2 V). Data obtained from a 100 ps MD simulation for poly(A) and poly(GT) with a timestep of 1 fs, respectively. Also given are the corresponding transmission maxima T_{MAX}. Note the different scale for the transferred charge in poly(GT).

potential may be assumed as in the analysis of Landauer and Büttiker, which may allow to compute the amount of charge transferred during the CT active conformations. Additional analyses showing the relation between time domains of large transmission and the transferred charge in time is given in Fig. 4.16.

The analysis of a window of 100 ps shows the appearance of CT active conformations on even longer time scales, i.e. the ps oscillations as shown in Fig. 4.15 are only a modulation on structural modifications on a longer time scale. For poly(A), e.g., during a period of 60 ps three distinct regions appear with high transmission, which leads to an increased charge transfer as competed as the integrated current. A very similar picture arises for poly(GT). According to this analysis, a considerable fraction of charge is transported on a ps time scale. However, it is not clear how to separate the time scales τ_1 and τ_2 in the case of DNA. The analysis above indicates that single conformations determine the CT, i.e. there is a conformational gating situation in that the ionic conformation allows or blocks the CT. The CT "open" conformations are then further modulated by fast fluctuations in the fs regime, in particular due to the water modes, which are important to understand the transmission quantitatively, as analyzed in the preceding sections.

5

Enhanced Conductance in G-Quadruplex DNA


In this chapter, the computational strategy from the previous chapter 4 is applied
to a special DNA form, the G-quadruplex DNA. Only minor changes have to be made
in the computation of the Landauer transmission $T(E)$ and the current $I(U)$ as the CG
Hamiltonian becomes more complex, i.e. no linear chain anymore. Thus, the more general
Eqs. 2.37 and 2.32 have to be used, instead of the simplified Eqs. 2.44 and 2.46. The
major aim of this work is to explore the structural differences between natural dsDNA
and G4-DNA by means of MD simulations. More importantly, to study whether different
structural properties and dynamics lead inevitably to different transport characteristics.
The G-quadruplex is said to be more rigid and exhibits a considerable polarizability
compared to natural DNA as described in Sec. 1.2. However, it has not yet been proven
experimentally whether G4-DNA can support larger currents, hence may be a potentially
better candidate for nano-electronic applications than dsDNA.

5.1 Structural Properties of G4-DNA

5.1.1 Starting Structures and Simulation Setup

The molecules used in this work base on the x-ray crystal structure of a tetrameric
parallel-stranded quadruplex (PDB-code: 244d), formed by the hexanucleotide sequence
d(TG$_4$T) in the presence of sodium ions. The structure had been resolved at 1.2 Å reso-
lution by Laughlan *et al.* [314]. As shown in Fig. 5.1(a), it contains two pairs of quadru-
plexes (TG$_4$T)$_4$, for which each pair is stacked coaxially with opposite polarity at the 5'
ends. Moreover, nine sodium ions are located inside each stack of quadruplexes forming
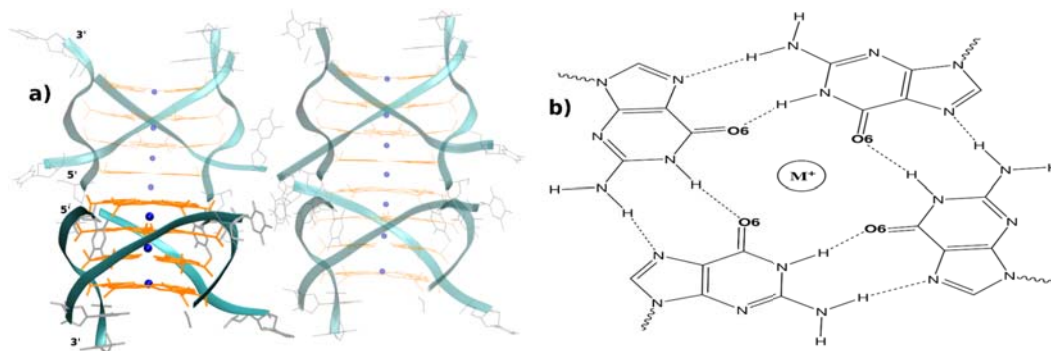
Figure 5.1: a) Tetrameric G4 x-ray crystal structure (244d [314]) with four units $(TG_4T)_4$. Backbones are indicated as light blue ribbons, central coordinated sodium ions as dark blue spheres, terminal thymine and guanine residues are shown in gray and orange, respectively. The highlighted quadruplex is used as starting structure and also to generate longer G4 molecules $(G_{12})_4$ and $(G_{30})_4$. b) Lewis structure of a single G4-tetrad with C_{4h} symmetry containing a monovalent metal ion in its center. Guanines are bound together by 8 hydrogen bonds via Hoogsteen pairing [77]. The metal ion is coordinated either coplanar by 4 or cubic by 8 O6 oxygen atoms of the respective guanines which depends on the ionic radius of M^+.

well ordered G4-DNA constructs. However, the terminal thymine residues, shown in gray, were not completely resolved because of high thermal disorder [314]. As indicated in Fig. 5.1, only one of these four quadruplexes is taken as the basic reference structure, which corresponds to the parallel strands A, B, C and D from the PDB file 244d. This quadruplex, from now on denoted as $(TG_4T)_4$, will be used for simulations and transport calculations. Furthermore, two longer G4 quadruplexes with 12 and 30 tetrads, denoted as $(G_{12})_4$ and $(G_{30})_4$, are generated by omitting the terminal thymine residues and adding subsequently G4 tetrads with a distance of 3.4 Å and twisted by 30°. For comparison, corresponding double-stranded B-DNAs with base sequence 5'-TGGGGT-3', poly(G) denoted as G_{12} and G_{30}, respectively, and a sequence containing 31 base pairs [1] (here referred to as "Scheer-sequence") are built with the make-na server [308].

Depending on the helix length, the G4 and dsDNA molecules are solvated in a rectangular box with 4000 to 11000 water molecules using the TIP3P model [236]; periodic boundary conditions (PBC) are applied. An appropriate number of sodium counterions (Na+) are added, in order to neutralize the total charge of the system, due to the negatively charged backbone. Four simulations are carried out for $(TG_4T)_4$ and $(G_{12})_4$, respectively: First, without any central monovalent alkali ions within the quadruplex, and then in presence of either lithium (Li+), sodium (Na+), or potassium (K+) ions within the respective G4 molecules. Therefore, either 3 $((TG_4T)_4)$ or 11 $((G_{12})_4)$ of these ions are placed subsequently in the center between two G4 tetrads coordinated by O6 oxygen atoms (see also Fig. 5.1(b)). Nevertheless, the longer $(G_{30})_4$ quadruplex is simulated only twice in absence and presence of central sodium ions. All simulations are carried out with the GROMACS software package [197] using the AMBER parm99 forcefield [233] including the parmBSC0 extension [235]. After a standard heating-minimization proto-

---

[1] 5-thiol-dG-GGC GGC GAC CTT CCC GCA GCT GGT ACG GAC
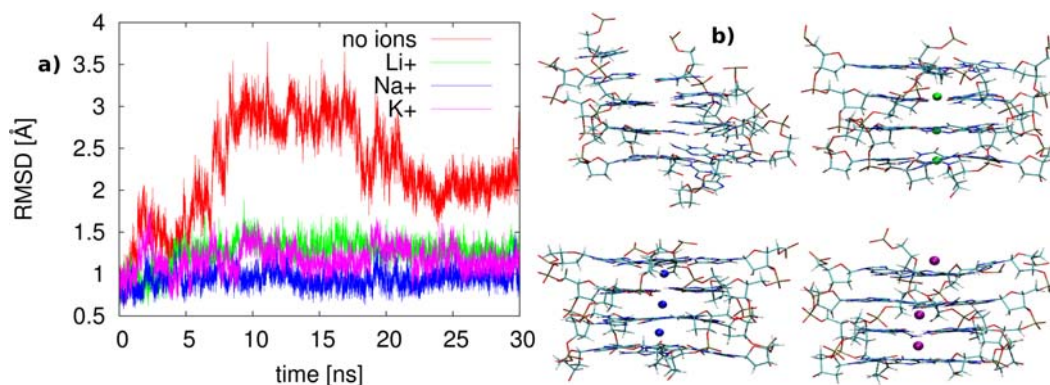
Figure 5.2: MD simulation of parallel stranded quadruplex $(TG_4T)_4$ with: i) no, ii) Li+, iii) Na+ and iv) K+ central ions. a) RMSD calculations, only guanine residues are considered. b) Molecular snapshots after 30 ns, Li+ (green spheres), Na+ (blue spheres) and K+ (purple spheres). Note that the terminal thymine residues are omitted.

col followed by a 1 ns equilibration phase, which is discarded afterwards, 30 ns (50 ns for $(G_{30})_4$) MD simulations are performed with a time step of 2 fs. Snapshots of the molecular structures were saved every 1 ps, for which the charge transfer parameters were calculated with the SCC-DFTB-FO approach as described in Sec. 2.5.

5.1.2 MD Simulations of G-Quadruplex DNA

In this section, the structure as well as the flexibility of G4-DNA, as obtained from MD simulations, is discussed. Results for both G4 structures, the short $(TG_4T)_4$ as well as the $(G_{12})_4$ are presented. Several factors which govern the structure of G4-DNA are analyzed in detail, i.e. the presence of alkali ions within the helix. Eventually, the structure of a longer G4 structure with 30 tetrads is explored.

5.1.2.1 $(TG_4T)_4$

To begin with, 30 ns MD simulations are performed with one of the $(TG_4T)_4$ quadruplexes from the tetramer crystal structure 244D [314] as a starting point (see Fig. 5.1). One of these simulations was carried out without central ions, while the remaining three were carried out in the presence of an ion channel within quadruplex. Therefore, three lithium, sodium and potassium ions, respectively, are placed initially in the center between two G4 tetrads.

Fig. 5.2(a) shows the time evolution of root-mean square deviations (RMSD (Eq. 2.63)) for the simulations considering only the guanine residues.

The reference structure for the RMSD calculation was chosen to be the initial structure of the corresponding production run. The RMSDs for quadruplexes which have central ions are considerably smaller than those where no central ions are present, indicating the structural importance of these ions. In fact, without central ions the quadruplex structure seems to be very flexible suggesting that there is no regular quadruple helix structure anymore. The RMSD values range up to 3 Å with very large irregular

fluctuations emphasizing that the initial structure has been destabilized. This finding is confirmed by various experiments which failed to build short stable guanine quadruplexes in absence of monovalent cations as well as similar theoretical investigations based on MD-simulation of G4-DNA [83, 85, 315]. The quadruplex structures containing sodium ions turns out to be very rigid over the whole simulation time, since the time evolution of RMSD values is rather smooth with only minor fluctuations ranging from 0.7 and 1.2 Å. The quadruplexes containing lithium and potassium ions show slightly larger structural deviations with up to 1.5 Å. Nevertheless, these G4 structures remain stable over the full simulation length. The thymine residues at the 5' and 3' ends are not included in the RMSD calculation, since they show much larger structural fluctuations which is also supported by various x-ray experiments, in which the thymine residues could not be resolved adequately because of their high thermal activity [314].

MD snapshots from Fig. 5.2(b) taken of the last frame for each simulation after 30 ns, underscore the RMSD results. As can be seen, the quadruplex without central ions is destabilized significantly. For one thing, the parallelism of some strands has vanished due to structural changes in one of the backbones. Moreover, the initial guanine pairing is lost almost completely, and replaced by some sort of intermediate base pairing similar to those in dsDNA.

The quadruplexes containing central ions remain regular G4 structures, as expected. The 4-strand parallelism, the tetrad stacking and original base pairing is maintained over the whole simulation time. Obviously, there are structural differences concerning the location of ions within the quadruplex. On the one hand, sodium and potassium ions prefer the center between two tetrads. On the contrary, lithium favors the planar position within the tetrads. One should have in mind that all ion types were initially placed at the same position in the center of two tetrads. Moreover, one of the potassium ions went out of the helix at the 5' end during the simulation. This could be explained by larger repulsive interactions between adjacent potassium ions due to the larger ionic radius compared to sodium. Interestingly, the K^+ ion did not get lost completely, rather it remained coordinated above the 5' tetrad for the remaining simulation time. These different coordination preferences for Li^+, Na^+ and K^+ are in excellent agreement with results obtained previously in Ref. [85].

5.1.2.2 $(G_{12})_4$

The RMSD calculations, given in Fig. 5.3, exhibit a much higher rigidity, i.e. less structural disorder for G-quadruplexes $(G_{12})_4$ in presence of central ions compared to dsDNA G_{12}. However, quadruplexes without central ions appear not to be in equilibrium, as observed before for $(TG_4T)_4$, the RMSD increases for the whole simulation time. Though, the structure is not destroyed entirely, as can be seen from molecular snapshots after 30 ns in Fig. 5.3(c).

The inner tetrads roughly maintain a G-quadruplex form, whereas the outer ones are disordered considerably. On the other hand, the quadruplexes in presence of alkali
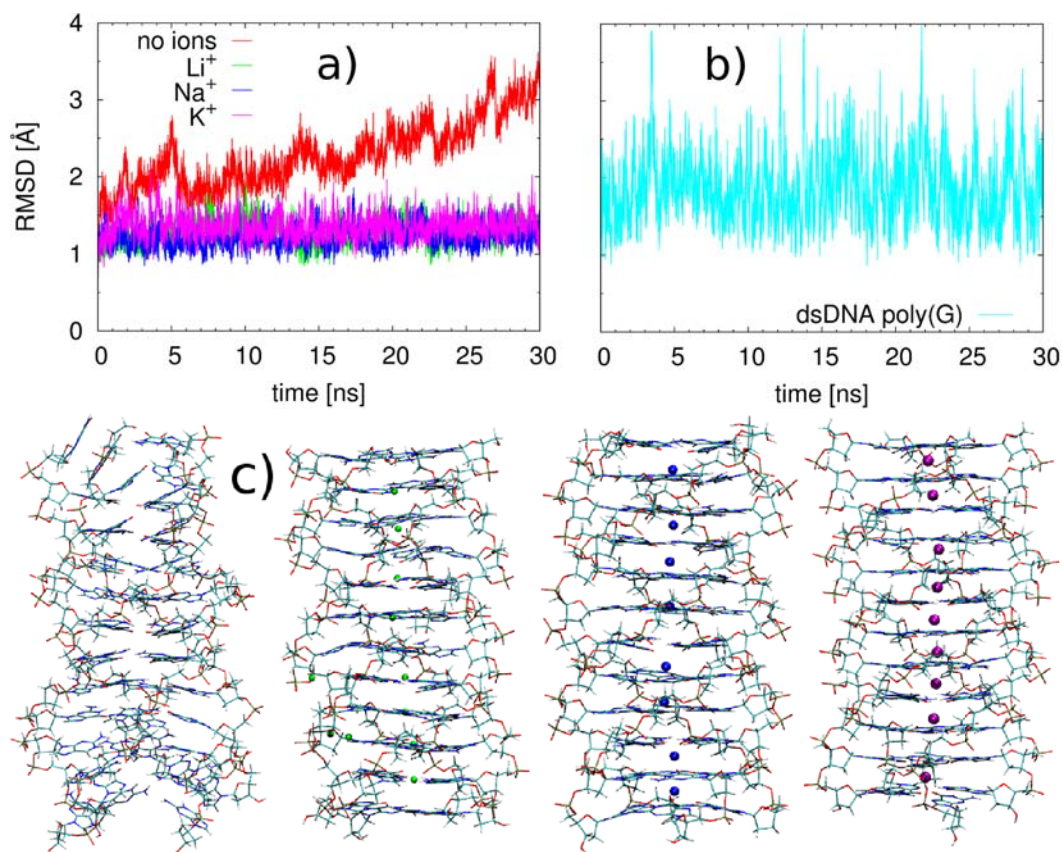
Figure 5.3: RMSD of G4 and poly(G) DNA: a) parallel stranded quadruplex $(G_{12})_4$ in absence and presence of centrals ions Li^+, Na^+ and K^+. b) Comparison with double-stranded DNA G_{12}. c) Molecular snapshot of parallel stranded quadruplex $(G_{12})_4$ after 30 ns MD simulation in absence and presence of central ions Li^+ (green spheres), Na^+ (blue spheres) and K^+ (purple spheres).

ions reveal highly regular 4-stranded structures. The location preference of the ions is the same as observed for $TG_4T)_4$, that is lithium favors the planar position within the tetrads, while sodium and potassium are most likely to be found in the center between adjacent tetrads. Thus, suggesting that this is a general feature, which may not be sensitive to the length of the helix. Nevertheless, lithium and sodium are more flexible than potassium ions, therefore allowing for longitudinal mobility along the quadruplex.

Accordingly, the ions are equally capable of moving out of the quadruplex through the minor groove and the 5' and 3' ends. Respectively, 3 sodium, 3 lithium and 2 potassium ions went out of the respective G4 dodecamers. Sometimes, these ions remain for a certain time in the minor groove. Likewise, solvent molecules are able to infiltrate the quadruplex via the minor groove as well as the 5' and 3' ends, provided that these locations are not occupied already by initially placed ions. The infiltration of solvent molecules into the helix will usually not destabilize the structure significantly, as long as there is at least a minimum amount of central monovalent cations left in the quadruplex. Otherwise, the interior solvent pattern might destroy the regular 4-stranded structure as observed for the structures without central ions in Fig. 5.2(b) and Fig. 5.3(c). These
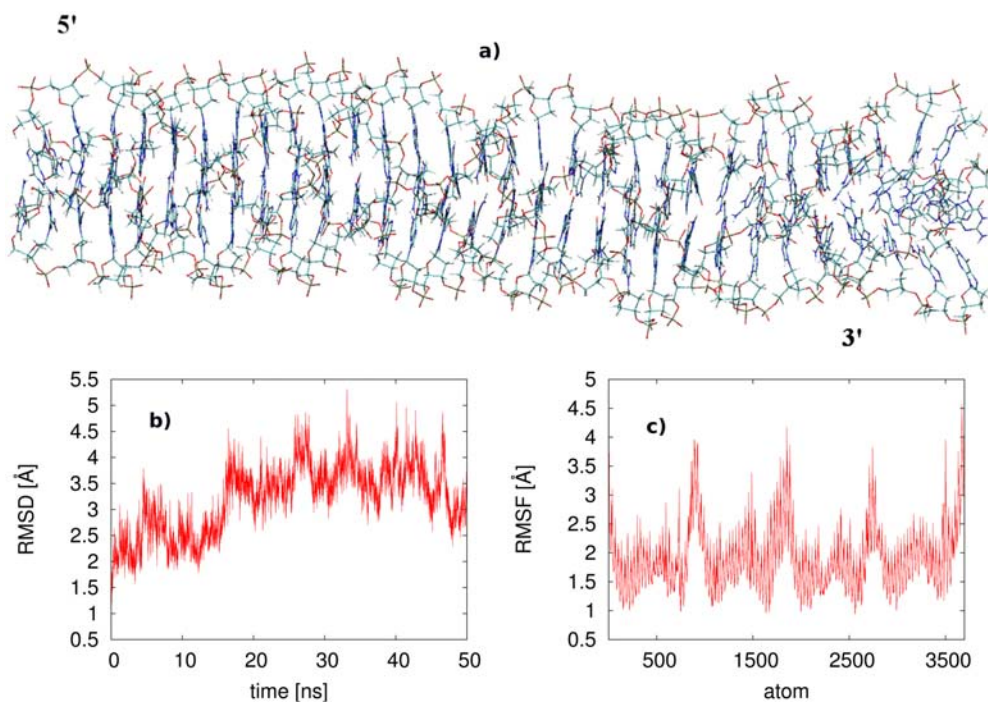
Figure 5.4: MD simulation of $(G_{30})_4$ without central ions: a) Snapshot after 50 ns, b) RMSD, and c) RMSF of atoms (Eq. 2.64)

observations are consistent with results obtained in Refs. [83, 85, 315]. Certainly, it will be interesting to see how these unique structural traits of guanine quadruplexes affect the CT compared to conventional dsDNA.

5.1.2.3 $(G_{30})_4$

Recently, Cavallari et al. have shown in their MD simulations that long G4 quadruplexes can be stable without central ions [85]. This finding is supported by experimental results [80, 81].

Therefore, an additional 50 ns simulation of a quadruplex with 30 tetrads $(G_{30})_4$ is carried out. The last snapshot, shown in Fig. 5.4(a) indicates that the major part of the quadruplex has a quite regular structure with adequate parallel stacking. Thus, it seems to confirm previous results that longer G4 wires become more stable, even in absence of central ions. Reasonably, this is due to increased stacking interaction. However, the quadruplex is deformed towards its ends as observed also for the dodecamer in Fig. 5.3(c). In particular, the 3' terminus is affected. Moreover, RSMD values in Fig. 5.4(b) reveal that the quadruplex structure is not balanced completely. The major part of structural fluctuations originates from the 5' and 3' ends which is indicated by four uprisings in the root mean square fluctuations per atom (RMSF) in Fig. 5.4(c).

Table 5.1: Average diagonal d(O6–O6) distances d_1 and d_2 for simulations of $(TG_4T)_4$. Comparison with x-ray data 244d [314]. All values in Å.

tetrad	diagonal	244d	MD no ions	MD 3Li$^+$	MD 3Na$^+$	MD 3K$^+$
1 (5')	d_1	4.43	7.96±3.46	5.98±0.22	4.37±0.06	4.65±0.07
	d_2	4.45	6.57±1.64	4.66±0.13	4.32±0.06	4.74±0.06
2	d_1	4.49	4.85±0.17	4.09±0.02	4.26±0.17	4.77±0.04
	d_2	4.60	6.79±0.19	4.08±0.03	3.92±0.17	4.85±0.05
3	d_1	4.50	5.48±0.20	4.00±0.02	4.06±0.15	4.26±0.05
	d_2	4.61	6.57±0.26	4.02±0.02	4.11±0.16	4.26±0.05
4 (3')	d_1	4.69	5.59±0.28	4.01±0.04	4.39±0.07	4.64±0.05
	d_2	4.71	6.68±0.64	4.03±0.04	4.48±0.07	4.73±0.06

5.1.2.4 Additional Structural Analysis

For further structural analyses and validation of G-quadruplex simulations for $(TG_4T)_4$ with respect to the crystal structure 244d [314], two average measures were calculated for each tetrad respectively: i) diagonal distances between guanine carbonyl oxygen atoms d(O6–O6) (see Fig. 5.1(b)), and ii) average distance between O6 atoms and the nearest central alkali ion d(O6–M$^+$).

As can be seen from Tab. 5.1, the crystal structure tetrads are highly symmetric, since their two diagonal d(O6-O6) distances d_1 and d_2 are very similar. The symmetry of the quadruplex without central ions appears to be destroyed completely, for the diagonal distances d_1 and d_2 differ by 1–2 Å for all tetrads, respectively. Furthermore, the values are substantially too large compared to the crystal structure pointing out poor binding interactions between the four individual strands. On the other hand, in presence of central ions the difference $|d_1 - d_2|$ is, except of 2 cases, always smaller than 0.1 Å ensuring perfect symmetry independent of the ion type. Since the three Li$^+$ were placed initially in the quadruplex and they most likely are coordinated in-plane, the first tetrad at the 5' terminus is the only one not occupied (see also Fig. 5.2(b)). This results in asymmetry and slight destabilization of the first tetrad of the Li$^+$ quadruplex, which is also supported by the large $|d_1 - d_2|$ difference of 1.3 Å.

More importantly, the diagonal distances for the simulation with central Na$^+$ are in good agreement with those for the crystal data, for which the structure contained sodium ions as well, therefore, confirming the quality of the simulation structures. The diagonal distances d_1 and d_2 for quadruplexes with Li$^+$ are smaller by about 0.4 Å compared with values for Na$^+$. On the contrary, those for K$^+$ are slightly larger by about 0.3 Å showing that the quadruplex cavity size increases for larger ionic radii. However surprisingly, the two inner tetrads of the Na$^+$ structure reveal significantly increased standard deviations, which might be an evidence for larger mobility of Na$^+$ within the quadruplex compared to other alkali ions.

Table 5.2: Average d(O6–M$^+$) distances for simulations of (TG$_4$T)$_4$, comparison with x-ray data 244d [314]. All values in Å.

tetrad	244d	MD 3Li$^+$	MD 3Na$^+$	MD 3K$^+$
1 (5')	2.71	4.273±0.024	2.456±0.005	2.790±0.010
2	2.66	2.082±0.002	2.645±0.023	2.701±0.003
3	2.45	2.036±0.001	2.646±0.056	2.836±0.008
4 (3')	2.35	2.247±0.240	2.459±0.007	2.716±0.004

Similarly, the d(O6–M$^+$) distances in Tab. 5.2 also confirm the structural resemblance of the crystal data with the quadruplex simulation containing central sodium ions. Also consistent is the fact that for both data the d(O6–M$^+$) distance for the 3' end tetrad is substantially smaller than for the remaining planes. However, one should note that in the x-ray structure, the sodium at the 3' terminus is coordinated in-plane. Clearly, d(O6–M$^+$) values are considerable smaller with Li$^+$ coordinated in-plane, i.e. about 2.1 Å, while the simulations with Na$^+$ and K$^+$ (coordination between the tetrads) exhibit larger values of about 2.5-2.8 Å. Although, the values for K$^+$ are slightly larger, in particular for the terminal planes. Again, the standard deviations of d(O6-M$^+$) distances for the inner tetrads are increased for the simulations with central Na$^+$.

In conclusion, the findings from the MD simulation presented in this section, which concern both the higher rigidity of G4-DNA as well as the preferred locations of central ions within the quadruplex, are in perfect agreement with those obtained by Špačková *et al.* [83, 315] as well as by Cavallari *et al.* [85].

5.2   CT Parameter for G4-DNA

In this section, the CT parameter, i.e. the site energies ϵ_i and the electronic couplings T_{ij}, are computed for G-quadruplex DNA. First, DFTB benchmark calculations have to be carried out on single G4 units, with respect to the MOs and its energies involved in the CT.

5.2.1   Molecular States

Similar to the work on double-stranded DNA in chapter 3 and 4, the highest occupied molecular orbitals (HOMOs) are used to describe the hole transfer process in G4-DNA. These orbitals are supposed to have π-symmetry in order to have sufficient MO overlap along the quadruplex. To validate the SCC-DFTB electronic structure of idealized G4 tetrads as well as snapshots from classical MD trajectories, benchmark calculation with HF and DFT methods were carried out.

Table 5.3: Energies of highest occupied molecular orbitals for a idealized G4 tetrad with C_{4h}-symmetry. Comparison between DFTB, DFT and HF, for the latter the 6-31G(d,p) basis set is used. All values in eV.

Method	HOMO	HOMO-1	HOMO-2	HOMO-3	HOMO-4
DFTB	-4.588	-4.593	-4.593	-4.597	-4.870
PBE	-4.356	-4.383	-4.383	-4.408	-5.267
B3LYP	-5.170	-5.196	-5.196	-5.222	-6.637
HF	-7.777	-7.803	-7.803	-7.828	-10.827

5.2.1.1 Delocalized States in Idealized G4

Tab. 5.3 shows energies of the five highest occupied MOs of an idealized G4 unit (tetrad) with C_{4h}-symmetry obtained after geometry optimization with B3LYP/6-31G(d,p). Corresponding visualization of MOs can be found in Tab. A.3 in the appendix. The first four occupied MOs, HOMO to HOMO-3, are very close in energy for all methods respectively, and they have π-symmetry. The energy range is about 0.01 eV for DFTB and 0.05 eV for the other methods. By contrast, HOMO-4 is separated significantly by a gap of about 0.3 eV for DFTB and more than 0.8 eV for the other methods. For DFT based methods this MO has σ-symmetry.

As a result, HOMO to HOMO-3 are the orbitals used for further transport calculations, for the close energy range between them reveals states which are rather delocalized over the four G bases. Thus, supporting the notion of band-like transport, provided large G4 stacks are considered. On the other hand, static disorder in dsDNA, i.e. due to differences in ionization potentials between four different nucleotides (A,C,G and T), leads to large energy gaps for tunneling and therefore to rather localized states. Basically, DFTB MO energies lie between those for PBE and B3LYP.

More importantly, the HOMO to HOMO-3 in Tab. 5.3 appear to be linear combinations of the HOMO for the isolated guanine molecule, which can be shown by comparison with results presented in chapter 3. This offers the possibility to map the electronic structure of G4-DNA onto single guanines rather than onto whole tetrads, hence reducing computational costs as described below.

5.2.1.2 Influence of Central Ions

The structure of G-quadruplexes is affected strongly by the presence of central ions as discussed already in Sec. 5.1. Clearly, the electronic structure is affected by central ions as well, for the distance between the ions and the G bases is very close as shown in Tab. 5.2. Tab. 5.4 shows the highest occupied MOs of a G4 tetrad in the presence of a sodium ion within its center of mass. Two types of calculations were carried out: i) The ion is treated as part of the QM system denoted by "QM", and ii) The ion is represented simply by a MM point charge of +1 denoted as "MM". Results for potassium and lithium

Table 5.4: Highest molecular orbital energies of an idealized G4 tetrad in the presence of Na^+, either as part of the QM system (QM), or as a simple MM point charge (MM). Comparison between $DFTB^2$, DFT and HF, for the latter the 6-31G(d,p) basis set is used. MO energies are given in eV. Results for lithium and potassium ions can be found in Tab. A.4 in the appendix.

MO	DFTB MM	PBE QM	PBE MM	B3LYP QM	B3LYP MM	HF QM	HF MM
HOMO	-7.363	-7.086	-7.041	-7.878	-7.832	-10.212	-10.207
HOMO-1	-7.367	-7.104	-7.067	-7.898	-7.861	-10.238	-10.239
HOMO-2	-7.367	-7.104	-7.067	-7.898	-7.861	-10.238	-10.239
HOMO-3	-7.370	-7.126	-7.091	-7.923	-7.888	-10.268	-10.272
HOMO-4	-8.402	-8.375	-8.341	-9.662	-9.628	-13.317	-13.312

ions can be found in Tab. A.4 in the appendix.

First of all, the pure QM results indicate that HOMO to HOMO-4 remain the same states as those from Tab. 5.3, which means there is no major density located on the central ions. Moreover, there are no new hybridized states due to combination of alkali ion and G4 states in the energy regime of interest, since these ion states are much lower in energy. Therefore, states of alkali ions do not contribute significantly to hole transfer, which is consistent with previous results [86].

However, alkali ions certainly have an effect on the MO energies as revealed from the values in Tab. 5.4. Clearly, all energies are shifted down by at least 2.6 eV due to the presence of a central alkali ion or equivalently a positive charge. Despite the large constant shift, the very small energy range in which HOMO to HOMO-3 are located is not affected by the presence of ions. On the other hand, the gap between these hole transfer states and the HOMO-4 increased significantly from 0.3 to 1.0 eV for DFTB and also to a lesser extent for PBE and B3LYP. This shows that various states, for instance π and σ orbitals, can be affected quite differently by presence of positively charged ions. Results for natural dsDNA from chapter 3 indicated that states being too close in energy may switch during the dynamics, which is somehow a methodical problem for calculation of transport parameters with DFTB. However, in this case the calculations are affected by this problem to a minor degree, which is due to the large energy gap between the hole transfer states (HOMO to HOMO-3) and the HOMO-4 caused by the presence of a positive charge.

Furthermore, the comparison of MM and QM MO energies for the respective methods in Tab. 5.4 (Tab. A.4 for Li^+ and K^+), exhibit a very good agreement with marginal differences of less than 0.03 eV. Accordingly, the central ions within the guanine quadruplexes are treated only electrostatically as MM charges, which is based on the fact that the shape of the hole transfer states is not changed by the presence of Li^+, Na^+ or K^+.

---

[2]Note that DFTB energies are given only for the "MM" option as Slater-Koster parameters for sodium as well as lithium and potassium are not readily accessible.

Table 5.5: Energies of the highest occupied molecular orbitals of a single tetrad of $(TG_4T)_4$ for a MD snapshot containing the full electrostatic environment built of negatively charged backbone, solvent and counterions. Comparison between DFTB, DFT and HF, for the latter the 6-31G(d,p) basis set is used. All values in eV.

Method	HOMO	HOMO-1	HOMO-2	HOMO-3	HOMO-4
DFTB	-5.390	-5.409	-5.467	-5.544	-5.756
PBE	-5.052	-5.156	-5.206	-5.273	-5.986
B3LYP	-5.871	-5.949	-5.994	-6.121	-7.356
HF	-8.322	-8.328	-8.362	-8.711	-11.113

5.2.1.3   Localized States Due to Dynamical Disorder

So far MO energies were obtained for idealized, highly-symmetric G4 tetrads. More importantly, the electronic structure needs to be computed for more realistic conditions, i.e. considering dynamical conformations and the environment composed of solvent, backbone and ions. This can be realized by computing MO energies of a single G4 tetrad for a MD simulation snapshot.

Already single alkali ions placed in the quadruplex cavity possess a substantial impact on the MO energies of hole transfer states, as indicated in the previous section. Therefore, it is vital to include the MM environment in the calculations. The MO energies in Tab. 5.5 are shifted down by about 0.5 to 0.8 eV for all applied methods, compared to those for the idealized tetrad without MM environment in Tab. 5.3.

Furthermore, the energy range in which HOMO to HOMO-3 are located has become considerably larger from 0.01 to about 0.15 eV for DFTB and equally from 0.05 to almost 0.4 eV for DFT and HF. Accordingly, these hole transfer states become rather localized onto the single guanines than onto the whole tetrads as soon as the electrostatic environment is considered, although the energy range is still very small. Snapshots of the corresponding MOs for these states are provided in Tab. A.5 in the appendix.

In conclusion, the effect of the electrostatic environment is of major importance, for it stabilizes and ensures energetically the π states as highest occupied ones used for hole transfer. However, this effect of localization clearly supports the notion to build up a transport model with single guanines as fragments rather than whole tetrads. Such a treatment does not not only have lower computational costs, for only 16 instead of 64 atoms per fragment have to be computed, but it also conceals a larger flexibility, since CT can occur through a multitude of pathways along the quadruplex. Moreover, the effect of central ions will be captured by a classical treatment in the following CT applications.

5.2.2   Site Energies ϵ_i and Electronic Couplings T_{ij}

In this section, the CT parameters, i.e. the site energies ϵ_i and the electronic couplings T_{ij} are computed along the classical MD simulation of $(TG_4T)_4$. The results are compared
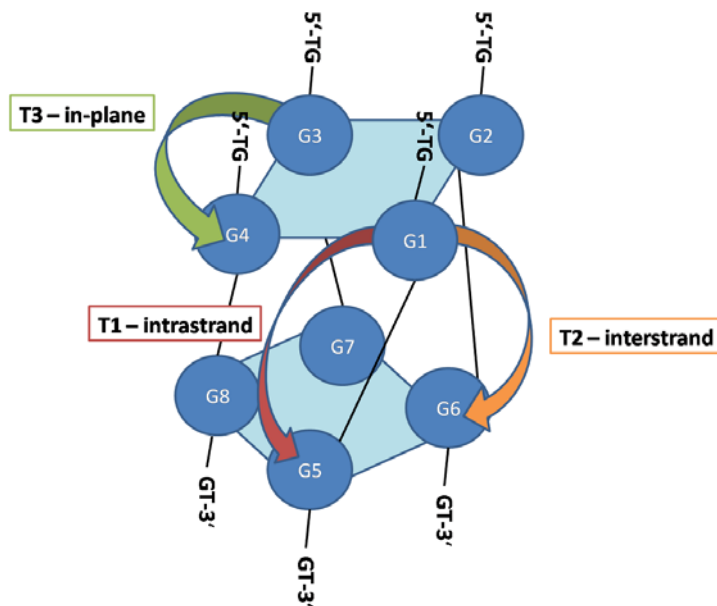
Figure 5.5: Scheme for the illustration of various electronic couplings T_{ij} present in G4-DNA. Calculations are based on the two inner most tetrads in $(TG_4T)_4$. Representatively, T1 indicates intrastrand, T2 interstrand and T3 in-plane couplings.

to those obtained for the crystal structure 244d, as well as to an idealized G4 stack. The parameters ϵ_i and T_{ij} were calculated as described in Sec. 2.5.1. In order to reduce complexity, first, the applied fragment methodology is introduced. The scheme in Fig. 5.5 shows different types of electronic couplings T_{ij} present in G4-DNA, from now on denoted as T1, T2 and T3.

Here, T3 represents electronic couplings within the respective G4-tetrads in-plane. Whereas T1 and T2 denote intra- and interstrand couplings lengthways to the quadruplexes which occur on either one single strand or between two different strands, respectively. In contrast to dsDNA, experiments suggest that also competing horizontal CT can occur in G4-DNA [316]. Clearly, a charge can follow several pathways along the quadruplex strongly dependent on those three couplings. This should be an advantage compared to dsDNA. Even in small G4 stacks there is a large number of possible coupling pathways, although only a minor part of them will be vital for CT in G4-DNA. For instance, T3 couplings are usually quite small, especially the diagonal in-plane couplings in Fig. 5.5 (e.g. G1-G3,G2-G4, ...) are generally negligible. Additionally, most of the T2 couplings are small, except those between adjacent strands (e.g. G1-G6,G2-G7, ...) as these are rather close to each other. The largest couplings are found certainly for T1 (intrastrand) and they are well comparable to those for double-stranded poly(G) DNA (see chapter 3).

A summary of the average electronic couplings and site energies for the simulations of $(TG_4T)_4$ compared to the crystal structure (244d), an idealized G4 dimer stack as well as to the corresponding dsDNA, i.e. the two central guanines in 5'-TGGGGT-3', is given in Tab. 5.6.

Table 5.6: Average site energies $\langle\epsilon\rangle$ and electronic couplings T1, T2 and T3 with standard deviations for the two central guanine tetrads of $(TG_4T)_4$. Energies obtained from MD simulations in absence and presence of central ions are compared to the crystal structure (244d), an idealized G4 dimer stack, and also to the corresponding dsDNA structure, i.e. the two central guanines in 5'-TGGGGT-3'. Note that averaging is carried out not only along the MD time series but also over the eight and two G bases for G4 and dsDNA, respectively. All values in eV.

Type	ϵ	T1	T2	T3
ideal	-4.895	0.028	0.001	0.009
244d	-4.905±0.062	0.051±0.011	0.012±0.002	0.007±0.003
MD no ions	-4.812±0.368	0.039±0.028	0.010±0.013	0.009±0.009
MD Li$^+$	-5.339±0.370	0.031±0.021	0.022±0.014	0.007±0.005
MD Na$^+$	-5.400±0.350	0.031±0.021	0.015±0.012	0.006±0.004
MD K$^+$	-5.201±0.354	0.029±0.020	0.013±0.010	0.007±0.004
dsDNA	-4.790±0.371	0.052±0.034	0.004±0.005[3]	0.012±0.008[4]

To begin with, the guanine site energies ϵ are very similar with about -4.9 to -4.8 eV for all quadruplexes without central alkali ions, hence comparable to those for poly(G) DNA. On the other hand, if central ions are present within the quadruplex, the average site energies are shifted to lower energies by about 0.6 eV for Li$^+$ and Na$^+$ but only 0.4 eV for K$^+$. This difference may be attributed to the larger mobility of sodium and lithium ions due to their smaller ionic radii compared to K$^+$. Therefore, Li$^+$ and especially Na$^+$ may get closer to the guanine bases in the quadruplex, thus increasing the Coulomb interaction. All simulations show the same large fluctuation for ϵ of about 0.35 to 0.37 eV which is the result of dynamical disorder as discussed in chapter 4, and analyzed in detail in Ref. [164]. These large site energy fluctuations are introduced by the solvent, and not by fluctuations of the DNA/G4 structure itself. Thus, they are not affected by the higher structural rigidity. The site energy fluctuations in "vacuo", i.e. neglecting the last term in Eq. 2.126, reduce to 0.1-0.15 eV which then corresponds to structural fluctuations of the DNA bases and is in agreement with findings reported by Hatcher et al. [43] and Řeha et al. [317].

On the other hand, the fluctuation of electronic couplings have the same magnitude as the averages themselves. More importantly, they depend sensitively on DNA conformation, i.e. base stacking and structural fluctuations, and only marginally on the solvent. These findings have been reported by several groups [43, 152, 164, 318].

The three different types of electronic couplings in Tab. 5.6 indicate that T1 is the dominant electronic coupling for all simulations as well as for the static structures (ideal and 244d). Interestingly, T1 is about 0.04 eV for the quadruplex without central ions, hence slightly larger compared to the quadruplexes which contain central ions. The reason for this is that the structure without central ions is more flexible, hence it is

---

[3]Interstrand coupling for G\C, G/C=0.013±0.014 eV
[4]Coupling within the WCP between G and C, note there is an energy gap of 0.4 eV

able to adopt various different conformations, which may sometimes lead to better π–π stacking and therefore to larger couplings. On the other hand, quadruplexes containing central ions are much more rigid as indicated in Sec. 5.1. Therefore, they seem to persist more or less in few conformations with only moderate electronic couplings, i.e. there may be fewer conformations with increased stacking. However, it seems that in this case T1 is not affected significantly by the ion type, since the values are very similar for Li^+, Na^+ and K^+ with about 0.03 eV, respectively, which is found for the idealized static G4 structure as well.

Notably, T1 is considerably larger for the crystal structure 244d with 0.05 eV, thus fits quite well with the G-G intrastrand coupling of double stranded poly(G) DNA. Note that Tab. 5.6 shows only coupling averages taken over all possible steps. For instance, T1 is the average of four similar coupling steps G1–G5, G2–G6, G3–G7 and G4–G8, as illustrated in Fig. 5.5. Large fluctuation of these four coupling steps are obtained for both the crystal structure 244d and the G4 simulation without central ions, ranging from 0.039 to 0.069 eV and 0.027 to 0.049 eV, respectively (see Tab. B.2 and B.3 in the appendix). On the contrary, quadruplexes containing central ions exhibit rather uniform coupling steps for T1 as well as for T2 and T3, indicating once more the very regular quadruplex structure. Nevertheless, one should have in mind that the electronic couplings depend sensitively on the actual DNA conformations, which in turn is determined by the DNA sequence and the number of base pairs. Accordingly, slight structural changes may lead to large differences for the electronic couplings.

As mentioned above, a charge may follow several pathways along the quadruplex. Except the idealized G4 structure, all simulations as well as the crystal structure 244d exhibit substantial interstrand couplings T2, ranging from 0.01 to 0.02 eV. Interestingly, the quadruplex containing Li^+ shows the largest T2 couplings. Presumably, because the structure is more confined compared to quadruplexes containing Na^+ and K^+, as pointed out in Tab. 5.1 and Tab. 5.2. Therefore, the four strands are closer to each other, thus enabling larger interstrand couplings.

On the other hand, only one of the two possible interstrand couplings T2 is comparably large in poly(G), that is C/G, whereas G\C is almost negligible. This reduces possible CT pathways in dsDNA, not mentioning the energy gap between IPs of G and C which has to be overcome. However, there are at least four significant interstrand coupling steps in G4.

Notwithstanding, the T2 coupling for the idealized static quadruplex is close to zero revealing the inadequate structural description by the idealized model and the necessity of proper MD sampling. In the latter case, the four strands appear to be too far away from each other, hence CT between different strands may be unlikely.

The smallest couplings are obtained for the in-plane case, as expected. The T3 couplings for the quadruplex simulations with central ions match the values for the crystal structure with about 0.007 eV considerably well, while the idealized static structure and the MD without central ions reveal slightly larger values of 0.009 eV. However, the in-

plane coupling T3 between G and C within the WCP is significantly larger with 0.012 eV, although there is static energy barrier of about 0.4 eV, as mentioned above.

Basically, the comparison between the simulations and the static structures, i.e. the crystal structure 244d and the G4 simulations containing central ions, exhibit a very good agreement, since the electronic parameters T1–T3 match very well. More importantly, these values characterize the structural resemblance as well, because the electronic couplings somehow mirror the DNA conformation. Nevertheless, the magnitude of couplings in G4 quadruplexes are comparable, i.e. even slightly smaller, to those for double stranded poly(G) DNA. Accordingly, an enhancement of CT in G4 quadruplexes may not be attributed to larger electronic couplings. In addition, the fluctuations of ϵ and T_{ij} are not significantly different as well. This is somehow surprising, for the structural differences between double-stranded and G4 DNA, as pointed out in Sec. 5.1, indicated more regular and rigid structures for the latter. Therefore the fluctuation of electronic couplings should have been decreased.

To sum up, in particular, the T1 couplings turn out to be even lower in the G4 structures with central ions, indicating that these structures conduct even less, when compared to four strands of dsDNA. However, the T2 and T3 values in G4 are rather large, suggesting that interstrand transfer can occur quite frequently. This opens a multitude of pathways for charge transport in G4 (compared to dsDNA), which will be the key to understand G4 conductivity, as discussed in more detail in the following section. As a result, there is no indication that the higher structural stability of G4 in the presence of central ions will lead to a higher conductivity due to reduced dynamical disorder or due to increased electronic couplings, i.e. because of somehow better stacking interactions owing to the more regular structure. Therefore, the potentially higher conductivity must have different reasons. The more stable structure leads to smaller couplings, in contrast to prior expectations [79]. A more detailed and complete set of site energies and electronic couplings for all quadruplex structures studied in this work is given in Tabs. B.1 – B.6 in the appendix.

5.3 Landauer Transport through G4-DNA

The parameters ϵ and T_{ij}, analyzed in the previous section, are now used to compute the CT properties using the Landauer-Büttiker approach as described in Sec. 2.3. It is important to point out that the Fermi energies E_F are not determined explicitly for the calculation of I–U characteristics, rather E_F is placed artificially as average of the site energies for each snapshot. For details see Sec. 2.3.3.1. The reader should be aware that the I–U shape could be quite different depending on E_F.

To begin with, reference calculations of transmission function $T(E)$ and current $I(U)$ are performed on static structures for both idealized G4 and dsDNA models and $(TG_4T)_4$, as based on the x-ray structure 244d. Predominantly, ensemble averages are of major interest, for single snapshots or idealized static structures should not be used
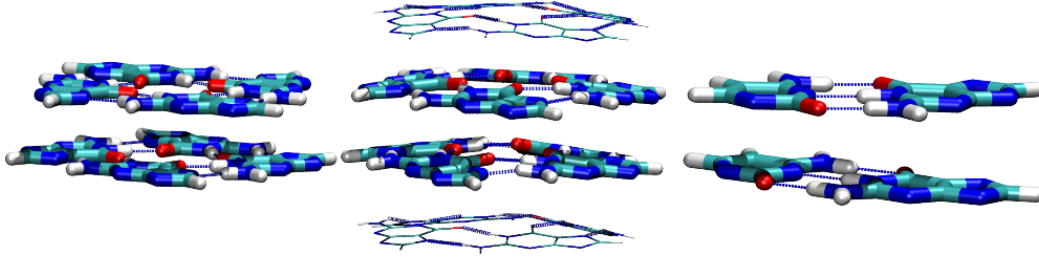
Figure 5.6: Static structures: Idealized G4-dimer (ideal), 30° twist and 3.4 Å rise (left). G4 tetrads of the crystal structure (244d) (center). Double-stranded idealized B-DNA (G$_2$), i.e. two stacked G-C base pairs (36° twist and 3.4 Å rise) (right).

to elucidate the CT process in such complex dynamical systems as DNA. Subsequently, the CT properties of G4-DNA are compared to those of double-stranded DNA, and the differences in CT efficiency is revealed by means of conformational analysis. Finally, the effect of the MM environment, in particular of the central ions, on CT in G4-DNA is analyzed, i.e. similar to the proceeding in the previous chapter 4.
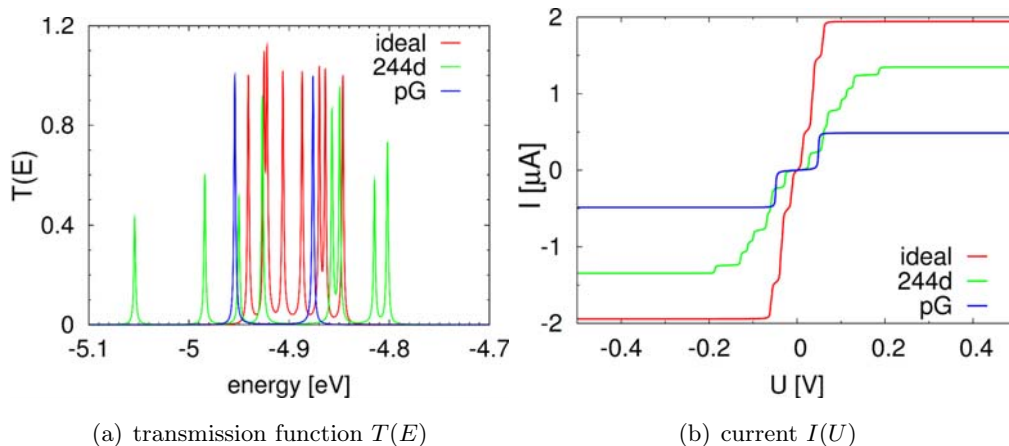
5.3.1 Static Structures

$T(E)$ and $I(U)$ are computed for three static structures: i) The idealized G4-Dimer as used in Tab. 5.6, ii) the two central tetrads of the crystal structure 244d, and iii) double-stranded idealized B-DNA (G$_2$). There are only slight structural differences between the idealized G4 dimer and the central stacked tetrads of the crystal structure 244d as can be seen from Fig. 5.6.

Fig. 5.7(a) exhibits the corresponding transmission functions showing eight resonances for the idealized G4-dimer due to the eight G bases of the transport model. However, these eight peaks lie in a very narrow energy range of about 0.15 eV showing similar transmission maxima of 0.9.

On the other hand, the eight peaks of the two central tetrads of 244d are located in a broader energy range of about 0.26 eV. Moreover, they exhibit lower transmission maxima, ranging from 0.4 to 0.8. Thus, indicating that small structural deviations leads to a broadening and a decrease of the transmission function. Nevertheless, the T1 and T2 couplings are much larger for 244d than for the idealized structure (see Tab. 5.6), thus the reason for the reduced transmission must be related to the site energy disorder. By contrast, the double-stranded poly(G) dimer contains only two G bases, the corresponding resonances are separated by a slightly larger energy gap, thus the transmission in this energy range is reduced strongly compared to the G4 structures.

Fig. 5.7(b) shows the corresponding I–U characteristics. The current for the G4 structures is significantly larger than for poly(G), i.e. in case of the idealized G4-dimer four times as large. This could be expected, because two idealized devices are compared, i.e. poly(G) with one single channel only, and G4 containing four channels. A small voltage gap is observed for the dsDNA as current flows only when a certain threshold

(a) transmission function $T(E)$

(b) current $I(U)$

Figure 5.7: Charge transport in static structures from Fig. 5.6. Comparison of transmission function $T(E)$ and current $I(U)$ for idealized G4-dimer, the two central tetrads of crystal structure 244d and poly(G) (G_2). Note that only the stacked G bases are considered, whereas the cytosine bases and the backbone composed of sugar and phosphate groups are neglected.

voltage is applied. The voltage gap would even increase if, additionally, the cytosine states were considered as well, resulting in a ladder model with both G and C sites separated by a large energy gap of 0.4 eV. Note again that in all current calculations shown in this work, the Fermi energy is artificially placed as the mean value of site energies for each snapshot respectively. Generally, the conductance for double-stranded poly(G) would increase once the cytosine states are included in the calculation as well, though it is well established that the C states have only a minor impact in hole transfer due to their much higher oxidation potential than G.

5.3.2 Dynamical Disorder: G-Quadruplex vs dsDNA

Previous studies revealed that idealized static structures cannot exhibit reasonable CT properties in DNA, since dynamical as well as environmental effects were shown to be crucial [152–156]. Dramatic effects are induced by dynamic disorder, for it suppresses CT in homogeneous sequences on the one hand, but can also enhance CT in heterogeneous (random) sequences on the other. Moreover, only a minority of conformations appears to be CT-active as indicated in chapter 4 (or Ref. [167]). Therefore, neglecting these significant factors or assuming purely random distributions for dynamical disorder may lead to a considerable loss of a vital contribution to the CT in DNA. On this account, the CT properties are evaluated along classical MD trajectories, which then leads to ensemble averaged quantities, that is the average transmission function $\langle T(E) \rangle$ and the current $\langle I(U) \rangle$, i.e. the same proceeding as used in chapter 4.

In Fig. 5.8, $\langle T(E) \rangle$ and $\langle I(U) \rangle$ are shown for both quadruplex molecules containing Na^+, the two central tetrads of $(TG_4T)_4$ (a) and $(G_8)_4$ (b) as well as their corresponding double-stranded poly(G) analogues G_2 and G_8[5].

---

[5] Note that generally, the two last tetrads (base pairs) at the 5' and '3 ends are not considered for
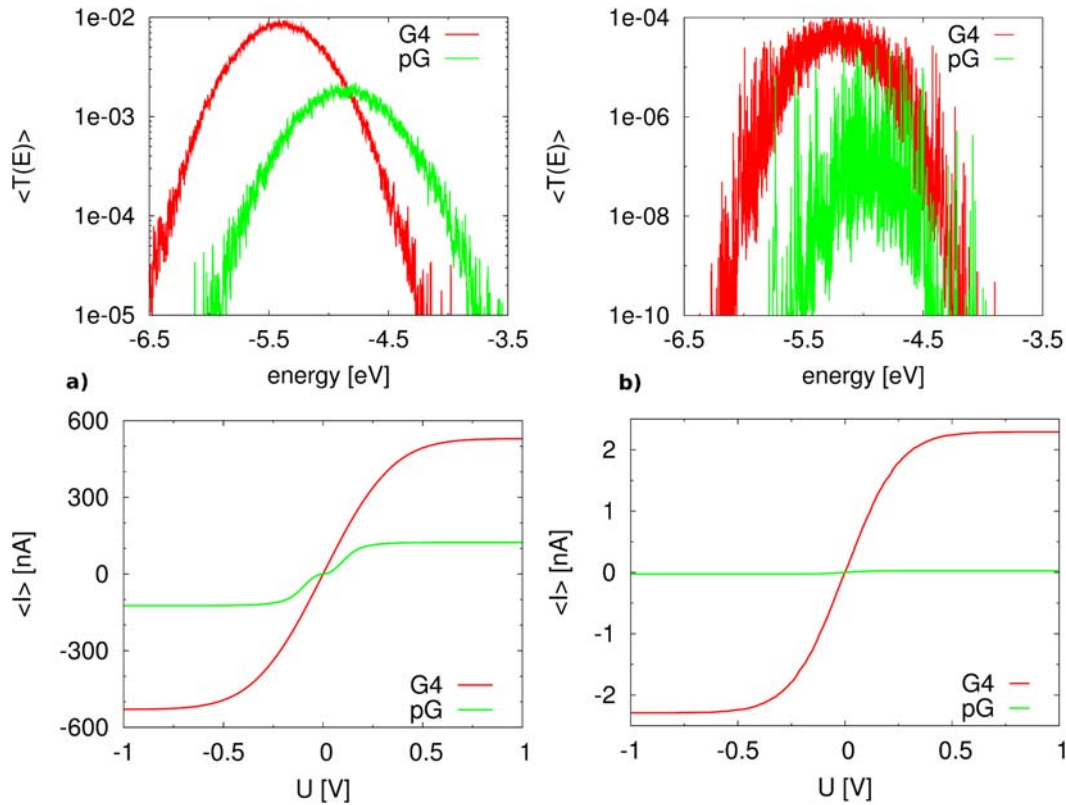
Figure 5.8: Average transmission $\langle T(E) \rangle$ and current $\langle I \rangle$ obtained from MD simulation: Comparison between G-quadruplex and double-stranded poly(G) DNA. a) $(G_2)_4$ (Na^+) and G_2, i.e. the two central tetrads (base pairs) of $(TG_4T)_4$ (5'-TGGGGT-3'), respectively. b) $(G_8)_4$ (Na^+) and G_8.

The transmission function exhibits large broadening for both DNA species due to the substantial site energy fluctuations of about 0.4 eV caused by dynamical disorder, i.e. due to the DNA MM environment and conformational dynamics. The transmission maxima for the quadruplexes are shifted to lower energies by about 0.3 eV due to the presence of the central sodium ions. Basically, the average transmission is reduced strongly compared to the idealized static structures in Fig. 5.7(a). Nevertheless, the maximum of $\langle T(E) \rangle$ for the two central tetrads in $(TG_4T)_4$ is almost five times larger compared to G_2. Accordingly, the maximum current is about 4.4 times larger in the quadruplex.

Regarding the octamers in Fig. 5.8(b), the conductance difference between G4 and poly(G) even increases. Here, the average transmission for the quadruplex in the relevant energy range is to a great extent two orders of magnitude larger than those for the poly(G) sequence. The poly(G) spectra shows much larger spikes at certain energies. These spikes indicate the strong impact of dynamics and may be explained by the existence of few charge transfer active conformations, which dominate the average transmission, i.e. few conformations, which feature a high transmission. This may indicate that even longer

---

CT calculations to avoid end effects, although the simulations were done with 12 tetrads and base pairs for G4 and dsDNA, respectively.
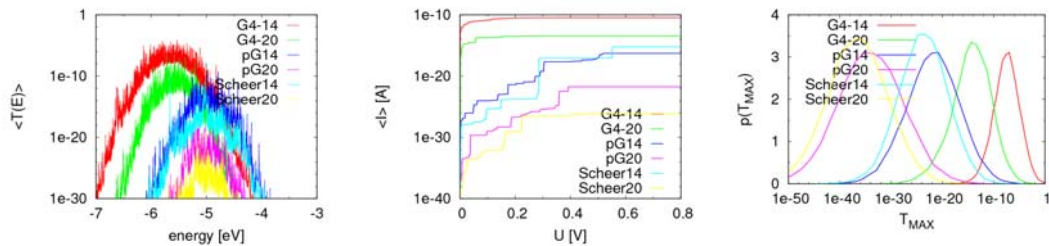
Figure 5.9: Average transmission $\langle T(E) \rangle$ (left) and current $\langle I \rangle$ (center) obtained from MD simulation: Comparison between 14mer and 20mer of G4-DNA with central sodium ions, p(G) and the corresponding central part of the "Scheer-sequence"[6]. Right panel: Probability distribution functions (PDFs) of T_{MAX}.

sampling (more than 30 ns) would be required to converge the spectra. Notwithstanding, it also reveals that in dsDNA the average current is dominated to a much larger degree on few non-equilibrium structures, as concluded in Ref. [167]. As a result, the average current $\langle I(U) \rangle$ for $(G_8)_4$ is almost two orders of magnitude larger than for G_8, hence suggesting that the enhancement of CT in G4 with respect to poly(G) might grow with increasing DNA length.

Length dependence Additional sets of CT calculations were performed for longer DNA species with 14 and 20 tetrads (base pairs), respectively. The corresponding data was obtained from MD simulations of a quadruplex $(G_{30})_4$ (including central Na^+), and double-stranded DNAs G_{30} and the heterogeneous "Scheer-sequence" recently used in a CT-measurement [60]. Note that for CT calculations only the 14 (20) central sites are used. The results are given in Fig. 5.9.

The transmission strongly decreases for both DNA species as expected. However, this effect is not as strong in G4 as in dsDNA. For instance, the current for $(G_{14})_4$ is only about one order of magnitude lower than for $(G_8)_4$, although the current drops significantly additional three orders of magnitude when increasing the length to $(G_{20})_4$.

On the other hand, $\langle I(U) \rangle$ for poly(G) and the "Scheer-sequence" decreases by more than 10 orders of magnitude by increasing the number of base pairs from 14 to 20. This indicates that there is a much stronger distance dependence of CT in dsDNA, hence the notion of coherent CT for longer molecular wires might be considerably more likely in G4 than in dsDNA.

Admittedly, the Landauer formalism used in this work performs well for short DNA species (i.e less than 10 sites), where the transport is assumed to be at least partially coherent. On the other hand, it clearly fails for long DNA sequences, as pointed out in the previous chapter 4. Therefore, the CT results for the longer molecules should be interpreted qualitatively only and with caution. For instance, the currents obtained for the 14mer and 20mer of both dsDNA molecules poly(G) and the "Scheer-sequence" are orders of magnitude smaller than pico-Ampere, which turns out to be far beyond any

---

[6]5-thiol-dG-GGC GGC GAC CTT CCC GCA GCT GGT ACG GAC
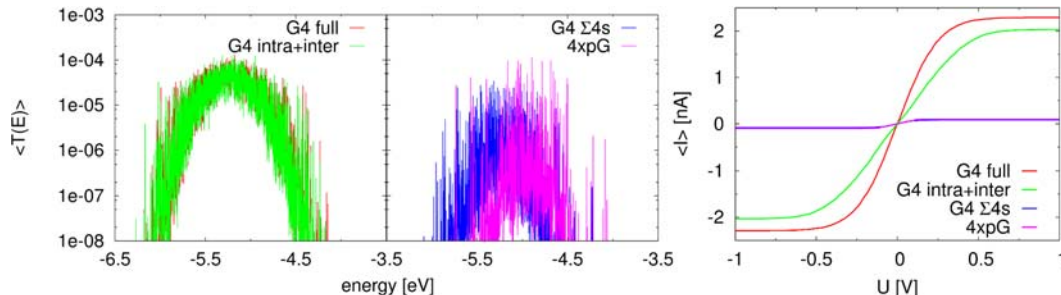
Figure 5.10: Effect of electronic couplings on CT in $(G_8)_4$ (Na+): Average transmission (left) and current (right) calculated i) for the complete electronic Hamiltonian, i.e. including all possible couplings (full), ii) only the intra- and interstrand couplings are non-zero (intra+inter), iii) the quadruplex is separated into its four single strands, for which the intrastrand transport is calculated independently and added up afterwards (Σ4s), and iv) for dsDNA: G_8 multiplied by four (4xpG).

experimentally measurable range.

5.3.3 Analysis of CT Differences in G4 and dsDNA

The significant enhancement of conductance of G4 over dsDNA may not be attributed solely to the fact that G4 is composed of four poly(G) like wires. To analyze this further, the CT in $(G_8)_4$ with central sodium ions is compared to various models, in which i) the complete electronic Hamiltonian is used, i.e. including all possible couplings (full), ii) only intra- and interstrand[7] couplings (i.e. off-diagonal elements of the Hamiltonian) are non-zero (intra+inter), and iii) the Hamiltonian for the quadruplex is separated into its four single strands, for which the intrastrand transport is calculated independently and added up afterwards (Σ4s). Furthermore, the results are analyzed with reference to G_8 (4xpG). Note that the CT quantities of poly(G) are multiplied by four for the purpose of comparison.

As appears from Fig. 5.10, the average transmissions for the two models, full MD and intra+inter, reveal the largest plateaus in the spectral support region. Moreover, they exhibit similar peak structures with moderate fluctuations.

By contrast, the sum of the four single G4 strands (Σ4s) shows much larger fluctuations, comparable with those for G_8 (4xpG). The average transmission for Σ4s is reduced significantly, hence is even slightly lower than 4xpG. Interestingly, there are barely CT-active conformations in Σ4s, i.e. single dominating peaks like in 4xpG, which sometimes even outreach the maximum transmission of the quadruplex (full). This might reflect the structural differences between G4 and dsDNA, as the four strands in the quadruplex are not as flexible as those in poly(G) (see also the RMSD fluctuations in Fig. 5.3). Therefore, the structural phase space is much larger in poly(G), thus several high-transmissive structures can arise.

---

[7]Only T2 interstrand couplings between adjacent strands are considered. See also the scheme in Fig. 5.5.
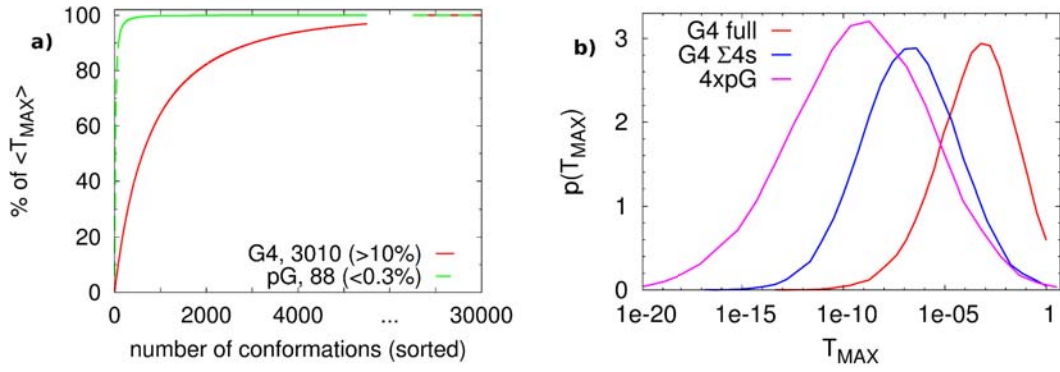
Figure 5.11: Conformation analysis for $(G_8)_4$ (Na^+) and G_8, a): number of conformations that make up 90% of the average transmission maximum $\langle T_{MAX} \rangle$, b): probability distribution function (PDF) of T_{MAX} for $(G_8)_4$ (Na^+) and the two models 4xpG and Σ4s, as used in Fig. 5.10.

On the other hand, the more rigid quadruplex explores a smaller conformational phase space, which indeed contains a large number of structures with moderate CT properties for each single strand, respectively. The results presented in Tab. 5.10 clearly indicate that the most important factors for the enhanced conductance in G4 are the interstrand couplings between the four strands in the quadruplex. Thus, if there are conformations in which the four isolated channels are not transmissive, there is a considerable probability that CT might occur via coupling between the individual strands. Those interstrand couplings are sufficiently large with about 0.01–0.02 eV, as can be extracted from Tab. 5.6. As a consequence, there is a substantial amount of pathways over which the CT might occur through the quadruplex. These findings are supported by the I–V characteristics in Fig. 5.10(right) as well. The current for the intra+inter model almost matches that of the full MD with 2.0 and 2.3 nA, respectively. If the interstrand couplings in the quadruplex are switched off, i.e. Σ4s, the current will drop down to 0.08 nA, which is more than one order of magnitude smaller. A slightly larger current of 0.11 nA is observed for 4xpG.

Further insights into the different CT properties of G4 and dsDNA may be gained by making use of a conformation analysis, as introduced before in the previous chapter 4. Accordingly, the number of conformations which dominate the average CT are determined. Additionally, the distribution of transmissions for the multitude of conformations is explored for both $(G_8)_4$ and G_8. The results are presented in Fig. 5.11.

Panel a) indicates evidently that there are substantially more conformations contributing to the average CT in G4 than in poly(G). Consequently, virtually every 10th G4 conformation is CT-active (about 3000 out of 30,000), while the average CT in poly(G) is characterized by only 128 (again out of 30,000) single non-equilibrium structures.

Furthermore, the transmission probability distribution functions (PDFs) in Fig. 5.11b) reveal that the distribution width for the G4 quadruplex is considerably narrower compared to 4xpG and Σ4s. Note that the x-axis is scaled logarithmic. In addition, the G4 PDF is shifted considerably to higher transmission. This underscores that the majority
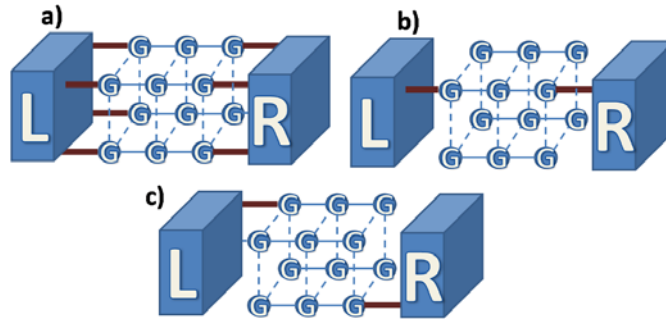
Figure 5.12: Modeling different types of connections to the left and right electrodes: a) all four G strands are connected (1-4), b) only one strand is connected with both termini (1-intra), and c) one strand is connected to the left electrode while one of the remaining is contacted to the right one (12-inter, etc.).

of poly(G) structures is not transmissive, yet a few single conformations are responsible for the average CT. On the other hand, the major part of conformations in $\Sigma 4$s exhibits a higher transmission than in 4xpG, but the single dominating conformations are missing. Thus, these single high-transmissive conformations present in poly(G) are the reason for a better conductivity compared to $\Sigma 4$s. The higher average T1 values in dsDNA are due to few highly conducting conformations. The smaller average of the T1 couplings in G4 resembles a more stable structure, yet not leading to a higher conductance, as could be argued beforehand. The advantage of G4 over dsDNA is attributed to the existence of non-negligible interstrand couplings in G4. The amount of high-transmissive structures is increased remarkably compared to double-stranded poly(G) DNA.
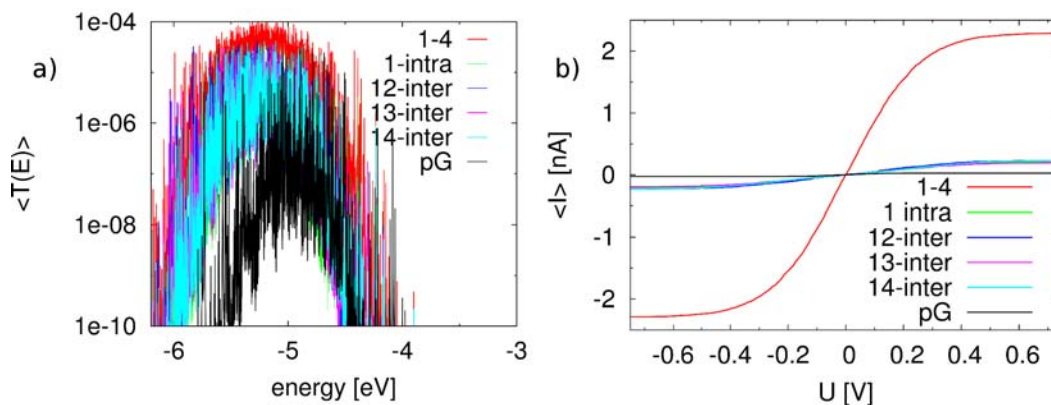
5.3.4   Electrode Connection

In many DNA conductance experiments, the molecules are connected with one strand only to the respective left and right contacts. For instance, in a very recent experiment by Scheer *et al.*, a single stranded nucleotide with sequence 5'-(T*G$_3$[TTAGGG]$_3$T*)-3'[8], which is known to form stable quadruplexes, was attached between two contacts [82].

Thus, the question may arise how CT in G4 depends on various contact connections, i.e. how CT in an all-parallel stranded quadruplex differs if only two or even one strand of the quadruplex is coupled to the left and right contacts, respectively. For that purpose, the CT in $(G_8)_4$ (Na+) is calculated for various connection models, as illustrated in the scheme in Fig. 5.12: i) all 4 strands are connected to the left and right electrodes, respectively (1-4), ii) only one strand is connected to both contacts (1-intra) and iii-v) one strand is attached to the left electrode, while one of the remaining is contacted to the right one (12-inter, etc.). The reader should note that there is no atomistic description of the electrodes in this simple model, rather the wide band limit is applied as described in Sec. 2.3.

Reducing the number of connected strands from four to one decreases $\langle T(E) \rangle$ by

---

[8]Here, T* denotes modified thymine residues.

Figure 5.13: Effect of electrode connections on CT in $(G_8)_4$ (Na^+): Average transmission and current calculated for the various contact models from Fig. 5.12. Comparison with poly(G).

about one order of magnitude for the relevant energy range, as shown in Fig. 5.13a). However, the transmission for the 1-intra and inter models is very similar indicating a minor significance for which strand or strands are connected to the contacts. This finding is supported by the I–V characteristics in Fig. 5.13b) as well, for the current ranges around 0.2 nA for these models, hence is roughly one order of magnitude smaller as though all four strands are attached. Interestingly, there seems to be an increase in the fluctuation of transmission if the quadruplex is contacted by one strand only at each end.

Notwithstanding, the transmission plateau for the one-stranded models (intra and inter) is still about 1.5 orders of magnitude larger compared to that for poly(G), thus leading to an average current which is again one order of magnitude smaller with 0.025 nA. Thus, these purely theoretical results suggest that a higher conductivity should be observed for the quadruplex, independent on the various contact linking schemes of G4 and dsDNA. Certainly, the optimal conductance for all-parallel stranded G4-DNA is ensured if all four strands are coupled to the contacts.

5.3.5   Effect of DNA Environment

The major part of the dynamical disorder is induced by the QM/MM environment, which is built of the MM charges of DNA backbone, solvent and counterions, as analyzed in detail in chapter 4. Previous results have indicated that the disorder due to the DNA environment might not only suppress CT in homogeneous sequences like poly(G), rather it is able to enhance CT in random sequences like the heterogeneous Dickerson dodecamer, as discussed in chapter 4 and Ref. [167].

Recent experiments by Scheer *et al.* confirmed this notion, as the current for the "Scheer-sequence" was found to be two orders of magnitude smaller *in vacuo* than in aqueous solution [60].

As can be seen in Fig. 5.14(a), the transmission maximum for $(G_8)_4$ *in vacuo* is about two orders of magnitude larger than with the QM/MM environment. Besides,
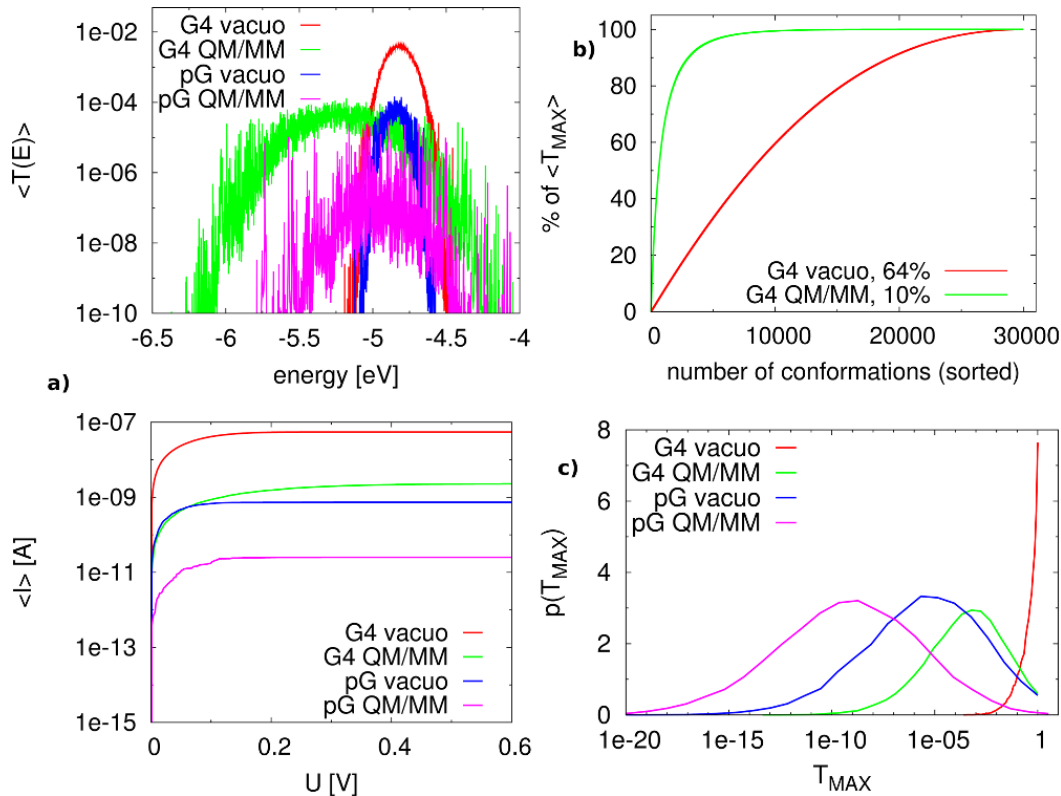
Figure 5.14: Influence of MM environment on CT: Comparison between $(G_8)_4$ (Na^+) and G_8. a) average transmission and current, b) number of conformations that make up 90% of the average transmission maximum $\langle T_{MAX} \rangle$ and c) probability distribution functions (PDFs) of T_{MAX}.

its broadening and as well as the fluctuations are reduced significantly, and the plateau is shifted to higher energies due to the neglect of the electrostatic interaction with the sodium ions within the quadruplex.

The transmission for poly(G) in vacuo shows basically the same features, for the broadening is likewise strongly reduced and the maximum is located in the same energy range as for G4 in vacuo, although it is 2 order of magnitude smaller. In general, the transmission for G4 including the QM/MM environment clearly reveals the largest broadening. This might be an indication that the central sodium ions have a strong additional impact on the dynamic disorder due to longitudinal mobility within the quadruplex, which is not the case in dsDNA. Despite all that, the current at high voltages is larger for G4 with QM/MM than for poly(G) in vacuo.

Basically, a reduced transmission would be observed for heterogeneous sequences in vacuo, e.g. for the "Scheer-sequence" in Fig. 5.9. This is caused by large energy gaps between A and G states. Notwithstanding, for DNA molecules with no static energy gaps like double-stranded poly(G) and G4-DNA (both with uniform DNA bases), the QM/MM environment is most likely to increase the dynamical disorder, thus will suppress CT compared to the vacuo model. As a result, there is no significant difference in the effect of the DNA environment on CT for G4 and poly(G) DNA. This is underscored
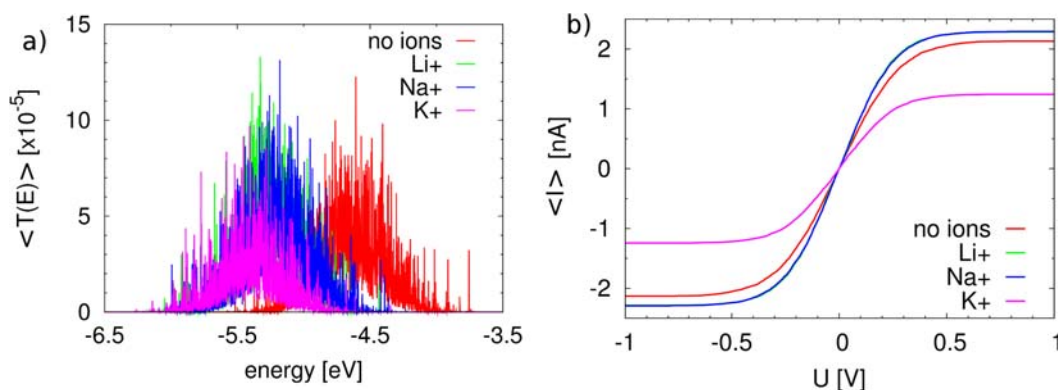
Figure 5.15: Influence of central ions on CT: Calculation of the average transmission and current for $(G_8)_4$ in absence and presence of monovalent central ions Li^+, Na^+ and K^+.

by the conformation analysis given in Panel b) and c) in Fig. 5.14 as well, indicating considerably more CT-active conformations in vacuo than with QM/MM environment for both DNA species G4 and poly(G).

Furthermore, nearly every conformation appears to be high-transmissive for G4 in vacuo, as the average maximum transmission $\langle T_{MAX} \rangle$ increases almost linearly with the number of conformations. Accordingly, the CT in vacuo is affected only marginally by single non-equilibrium conformations, rather the whole ensemble of G4 conformations seems to be CT active, which is indicated in the PDF of transmission (Fig. 5.14c)).

5.3.6   Effect of Central Ions

In this section, the dependence of the CT in G4-DNA on the central ions, located within the quadruplex channel, is analyzed. The structural influence of these ions has been addressed already in detail in Sec. 5.1. Moreover, the central ions were not found to contribute states in the relevant energy range for hole transfer in G4, therefore, as indicated in Tab. 5.4 and Tab. A.4 in the appendix. Therefore, the effect of central ions is investigated purely electrostatically.

In Fig. 5.15, the average transmission and current is given for the quadruplex simulations of $(G_8)_4$ in absence and presence of either lithium, sodium and potassium ions. Obviously, the transmission maximum is affected only marginally by the presence of different types of central ions. Furthermore, the transmission function is shifted to lower energies once central ions are present.

However, the transmission is slightly reduced for potassium, also the broadening is not as large as for the other species. As a consequence, the average current (Fig. 5.15b)) for the simulation with central potassium ions is half as large as for those with lithium and sodium ions, which might be attributed to the different mobilities of Li^+ and Na^+ compared to K^+.

Interestingly, there is no significant difference for CT in absence and presence of ions, although the results in Sec. 5.1 indicated that G4 molecules without central ions exhibit

significant destabilization. Once more, this supports the notion that the enhanced conductance in G4 may not be explained exclusively in terms of higher structural rigidity, i.e. less dynamical disorder, rather it is the multitude of CT pathways via interstrand couplings that lead to an increased number of high-transmissive conformations. Apparently, those interstrand and in-plane couplings T2 and T3 are not altered significantly by the presence of central ions, i.e. by a more rigid quadruplex structure.

6

Conclusion of Part I

6.1 DFTB CT Parameter for Hole Transfer in DNA

It is well-established that π–π interaction between stacked nucleobases play a vital part in the CT in DNA [3]. Therefore, the electronic couplings T_{ij} between those corresponding orbitals are of considerable interest for theoretical studies.

In chapter 3, an efficient method to compute CT parameter for hole transfer in DNA was presented. Site energies ϵ_i and electronic couplings T_{ij} were computed based on the fragment orbital (FO) approach described in Sec. 2.5.1 using atomic Hamilton matrix elements obtained by the semi-empirical method SCC-DFTB (Sec. 2.5.4). The atomic electronic structure is mapped onto a coarse-grained tight-binding model reducing the computational effort immensely, as solely the nucleobases need to be computed quantum-mechanically.

In conclusion, SCC-DFTB values for ϵ_i and T_{ij} compare very well with higher-level methods, such as HF [39], DFT [3, 142–144] and even CAS-PT2 [145], though, the calculations are orders of magnitude faster, i.e. DFTB is usually about thousand times faster than standard GGA-DFT. Moreover, the T_{ij} were shown to depend sensitively on the structure. In fact, slight changes in the molecular geometries have a substantial impact on T_{ij}. Accordingly, various DNA conformations (e.g. A-DNA or B-DNA) exhibit rather different couplings, strongly indicating that theoretical studies of CT in DNA should not be based on static idealized structures.

Furthermore, environmental effects on the CT parameter can be accounted for by making use of the QM/MM formalism, that is the fragments, i.e. the nucleobases, are computed in the presence of the electrostatic surrounding. In case of hole transfer in DNA, this purely electrostatic surrounding may be composed of the MM charges of the solvent, the DNA backbone, the counterions and also the complementary pyrimidine bases. In fact, this framework allows for the detailed study of different effects on the CT parameter due to the various MM components.

Accordingly, the efficient but sufficiently accurate computation of CT parameters using SCC-DFTB can be coupled easily to extended classical MD simulation. This inclusion of dynamical and environmental effects in the study of bio-molecular CT represents the

123

major aim of this work. There are two ways to realize the incorporation of dynamics. First, the CT parameters and related properties can be computed for ensembles of conformations, which then lead to statistical-averaged quantities. On the other hand, CT events may be simulated directly in real time by solving the time-dependent Kohn-Sham equation using the coarse-grained CT Hamiltonian. The latter strategy should give detailed insight into the mechanism of charge flow in DNA.

Generally, the SCC-DFTB-FO approach may be used to study hole and electron transfer in various complex system, such as proteins and organic crystals. Notwithstanding, system-specific testing is required in each case due to the approximate character of DFTB as well as the well-known errors inherent in conventional DFT, i.e. the self-interaction error. The latter affects primarily the CT energetics, that is the site energies (e.g. IPs) and energy differences (e.g. band gaps).

6.2   Transport through Fluctuating Bridges

The SCC-DFTB-FO approach was used to compute CT parameters for hole transfer in DNA along extended classical MD simulation accounting for a realistic representation of dynamical and environmental effects. The corresponding time series of CT parameter were then used to compute the Landauer transmission and current for coherent transport.

To begin with, the DNA conformation is affected dramatically by the dynamics and the surrounding solvent, thus can deviate significantly from the idealized A or B-DNA form. In particular, this is true for the electronic couplings T_{ij} whose average values obtained from MD simulations differ considerably from those of the idealized static DNA structures. More importantly, the time series of T_{ij} reveal rather large fluctuations which are of the same order of magnitude as the averages themselves. As a consequence, a realistic theoretical description of CT in DNA should be based on ensembles of DNA conformations generated from MD, rather than on single static structures. The T_{ij} turn out to be very sensible to changes in the molecular geometry, as indicated in the previous section, while the electrostatic environment, i.e. the solvent and counterions, have a minor influence [164]. Thus, the time evolution of T_{ij} is strongly coupled to the DNA conformation.

However, the statistical analysis of the Landauer transmission in chapter 4 revealed that the CT on average is characterized by the mean T_{ij} only. In fact, the time series of T_{ij} could be replaced by their mean values leading to the same average transport quantities. This observations is contradicting to those obtained in Ref. [317], in which enhanced CT is realized by assuming coherent conformational motion of the DNA. On the other hand, one should note that the CT for a single given snapshot still depends sensitively on the actual values of the T_{ij}.

On the contrary, the transport properties were observed to be affected dramatically by the site energy ϵ_i fluctuations. There are two contributions to these fluctuations, as studied in detail in Ref. [164]. Structural fluctuations of the nucleobases themselves

lead to moderate fluctuations of 0.15 eV in vacuo. By contrast, taking into account the interaction of the nucleobases with their surroundings exhibits considerably larger fluctuations of 0.4 eV, which are caused predominantly by the solvent. Interestingly, the static energy gap (IP difference) between Guanine and Adenine is 0.4 eV as well. Thus, dynamical disorder due to the structural dynamics and the interaction with the electrostatic environment may create situations in which this static G–A barrier vanishes, hence CT via resonant tunneling can occur even in rather heterogeneous sequences with high static disorder.

As a consequence, the average CT is affected differently by the fluctuations of ϵ_i for sequences with low and high static disorder. Accordingly, dynamical disorder suppresses the CT in homogeneous sequences, such as poly(G) or poly(A), very strongly by reducing the number of "CT-active" conformations. On the contrary, the effect is reversed for heterogeneous sequences (e.g. the Dickerson Dodecamer), as "CT-active" conformations arise precisely because of the dynamical disorder leading to an immense increase in the average transmission.

Similarly, the neglect of the solvent effects in such heterogeneous sequences leads to a decrease in the average transmission, for the number of "CT-active" conformation is reduced due to the smaller site energy fluctuations in vacuo. Consequently, these results predict heterogeneous dsDNA to exhibit a higher conduction in solution than in gas phase [167]. This observation has been confirmed by a recent experiment reported in Ref. [60]. Furthermore, the sequence dependence on CT appears to be of minor importance due to the dynamical disorder, which is observed in several experiments as well, in which similar currents were obtained for completely different DNA sequences [55, 61].

However, the analysis of high and low conducting conformations revealed very clearly that only very few conformations contribute to the average transmission, thus the CT described with Landauer theory may be regarded as a highly non-equilibrium process. The "CT-active" conformation are characterized by small site energy difference along the DNA chain and at the same time sufficiently large couplings. During the course of a simulation high conducting structure are explored at various charge injection energies. Therefore, the time series of site energies may not be replaced by their mean values in contrast to the electronic couplings. An adequate description of CT energetics should not rely on mean values of site energies, as used in Refs. [166, 319].

A different strategy to include dynamical effects in the description of CT in DNA, despite performing MD simulations, is usually to assume simple distributions for the CT parameter. Moreover, the fluctuations of CT parameters are supposed to be independent for the individual sites. However, the results reported in Ref. [164] revealed clearly that the fluctuation of ϵ_i is not completely random, rather three to five adjacent nucleobases exhibit considerable site energy correlations, as they feel a similar ESP caused mainly by the solvent. The results presented in chapter 4 indicated the importance of these correlations on the CT. The average transmission and the current are increased by a

factor of two, and four, respectively, once the site energy correlations are considered. A similar observation was reported in Ref. [320] in the context of exciton transport.

Various mechanism of CT in DNA have been discussed in the last decades. The counterion-gating mechanism [321] suggests that CT is accompanied by the motion of counterions. However, the movement of solvent and ions turns out to be strongly related [164], indicating that the motion of both solvent and ions gives rise to "CT-active" conformations. This is supported by the time evolution of the transmission revealing time domains with high transmission in the ps regime.

Barton et al. [15] discussed the mechanism of conformational gating, in which "CT-active" conformations are characterized by increased values of Π (Eq. 4.3). The notion has been adapted in this work, yet indicating that small values for Σ maybe of similar or even more significance. Accordingly, the CT is gated critically by the adjustment of the site energies, whose dynamics is determined largely by the solvent. "CT-active" conformations arise due to the concerted motion of the whole system, i.e. DNA, water and counterions. Thus, in the light of the results presented in chapter 4, the CT in DNA is rather "solvent gated" than conformational gated.

Notwithstanding, there are two major shortcomings associated with the computation of transport properties based on Landauer theory. The first one is related to the fact that the CT is assumed to take place via coherent tunneling. This may be valid for short transfer distances, while the assumption becomes questionable with increasing distance. Nevertheless, recent experimental observations [44] suggest that CT might be at least partially coherent over up to 10 base pairs, although this stands in strong contradiction to most experimental [24] and theoretical [37, 42] results predicting hopping as the dominant mechanism if the number of bridge base pairs exceeds three.

Evidently, the Landauer strategy as applied in this work, predicts far too low currents for longer CT distances as compared with experiments. For instance, the computed current for a sequence used in a recent experiment by Scheer and co-workers [60] turns out to be orders of magnitude too low.

The second issue concerning the limits of the applied Landauer–MD framework is related to the time scales of ionic and electronic motion. The averaging of the transport properties over the ensemble of conformations as carried out in chapter 4 and 5 assumes the CT to be faster than relevant molecular motions of DNA, solvent and counterions. In the opposite case, the transferring charge would encounter an average potential, thus averaging should be performed over various time intervals of the CT parameters. Challenging situations will emerge once both time scales are comparable, then, a non-adiabatic description of the CT is inevitable. However, the extraction of single "CT-relevant" molecular modes and their corresponding time periods is rather difficult for such a large complex system. The time series of transmission revealed "CT-active" plateaus in intervals of several ps, yet large fast-modulating fluctuations in the fs regime were observed as well.

To conclude, several interesting insights concerning the CT in DNA could be gained,

despite the shortcomings of the applied theoretical framework. The notion of promoting solvent modes and the effect of site energy correlations on CT are in agreement with findings obtained in Ref. [173, 322] on the simulation of incoherent hopping in DNA, in which the hole wave-function has been propagated using the TD-KS Hamiltonian. The Landauer calculations as well as the MD time series of CT parameter may serve as an input and reference for improved and more complex model Hamiltonians for the description of CT in complex materials in a more general way [166, 319].

6.3 Enhanced Conductance in G4-DNA

In chapter 5, the transport properties of G-quadruplex DNA was examined and compared to those of natural dsDNA using the same computational strategy as in chapter 4. Therefore, the shortcomings discussed in the previous section hold for the work presented, here, as well.

To begin with, the presence of central ions within the quadruplex leads to a significantly higher structural rigidity of G4-DNA compared to dsDNA. Accordingly, the four-stranded well-ordered G structure is expected to provide considerably increased π–π stacking interaction along the chain [79], therefore, leading to improved transport properties compared to dsDNA.

On the other hand, it is large fluctuations driven by the DNA environment that create conducting conformations, thus dominate the CT in conventional dsDNA, as observed in chapter 4. As a result, double-stranded poly(G) DNA conducts even slightly better than one isolated G4 strand. Therefore, the observed higher conductance of G4 may not be explained in terms of higher structural rigidity. In fact, single dominating "CT-active" conformations are missing regarding the isolated strands of G4 due to the smaller phase space, i.e. the limited structural flexibility compared to dsDNA.

Instead, the number of conducting conformations is increased remarkably in G4 by enabling CT over various pathways along the quadruplex due to the sufficiently large interstrand couplings T2 and T3 (see Fig. 5.5). For instance, if CT within one G4 strand is blocked due to vanishing coupling or large site energy disorder, various other transport channels may be accessible due to vital CT couplings between the individual G4 strands. In conclusion, the greatly increased number of "CT-active" conformations in G4 compared to dsDNA constitutes the basis of the enhanced conductance in G-quadruplex DNA.

Characterizing high-conducting conformations in terms of molecular structure seems to be intractable, for the structural differences compared to low-conducting sequences are marginal. However, "CT-active" G4 conformations exhibit sufficiently large electronic couplings due to π–π stacking interactions for various CT pathways, and at the same time low site energy disorder as discussed in chapter 4 and in Ref. [167].

Furthermore, the improved transport characteristics of G4 over dsDNA are even maintained, if the molecule is contacted to the left and right electrode by only one single G-site, respectively. Though, the conductance of G4 is affected rather sensibly by the

number of connected strands, i.e. all four or only one (either intra or interstrand), which is in agreement with results reported recently in Ref. [151] on dsDNA, yet neglecting dynamical and environmental effects. In this study, the coherent transport for ordered dsDNA sequences was found to be influenced severely by various connection strategies. On the contrary, experiments by Guo et al. [323] on a DNA sequence with high static disorder revealed no significant difference in conductance using different connection strategies, i.e. intra or interstrand. This is in agreement with the results presented in chapter 5 as well, for the average transmission and current in the single-strand-connected case was found not to be influenced by which strand (intra) or strands (inter) are contacted to the electrodes.

Nevertheless, the higher structural rigidity of G4 over dsDNA could be vital in technical applications, although it does not translate directly into higher conductivity, as expected beforehand. If the DNA molecules are exerted to strain due to the contacting procedure, the increased structural stability of G4 may be essential to maintain a conducting conformation, while the CT in dsDNA may be disrupted more easily [82]. Interestingly, a higher variability of I–V characteristics was observed for G4 compared to poly(G) in a recent experiment by Shapir et al. [324] on the electronic structure of G4-DNA. Obviously, this outcome was not expected by the authors as stiffer molecules should give more clear transport characteristics. Thus, it seems to support the previous notion that the CT characteristic of G4 might not be related predominantly to its higher rigidity.

Moreover, G4 reveals a substantially higher conductance in gas phase, as found generally for homogeneous dsDNA sequences such as poly(G). In these cases, dynamical disorder due to dynamics and solvent destroys the good idealized initial conditions, i.e. "CT-highway" with equal site energies, thus reducing the number of "CT-active" conformations immensely. Eventually, the average CT in G4 turns out not to be influenced significantly by the presence of central metal ions within the quadruplex, although these were treated only electrostatically in this work. However, the central ions were shown to be vital for the stability and rigidity of G4-DNA.

Part II

QM/MM Simulation of CT
in *E. coli* Photolyase

7

Classical Description of CT via Marcus' Theory

In the following two chapters, the photoactivation process in *E. coli* DNA photolyase is investigated by means of combined molecular dynamics and quantum chemical approaches. The photoactivation is initiated by electronic excitation of the FAD cofactor inside the protein, presumably via energy transfer from the light-harvesting antenna molecule MTHF, leading to a fast hole transfer to the nearby tryptophan sidechain (Trp382). Subsequently, the FAD cofactor remains negatively charged and the hole is supposed to move further to the exterior of the protein over the evolutionary conserved Trp chain: Trp382, Trp359 and, finally, Trp306. This study focuses solely on the hole transfer between the individual Trp sidechains assuming the hole to be located already on the first residue Trp382, thus electronic excitation of the FAD is not considered. Nevertheless, the effect of a negatively charged FAD cofactor, also with respect to a charge neutral system (i.e. FAD and the three Trp sidechains are charge neutral) on the structure and the energetics of CT is analyzed using classical MD simulations.

In this chapter, classical MD simulations are used to compute structural properties, CT parameter as well as CT rates based on Marcus' theory, while in the following chapter 8, the CT process over the conserved Trp chain is simulated directly using fully-coupled non-adiabatic electron-ion dynamics (EID). Accordingly, the results presented in this chapter are derived from a purely classical description of CT, as the hole charge is always localized completely on one of the involved Trp sidechains.

7.1 Starting Structure and Simulation Setup

The protein structure of *E. coli* DNA Photolyase (PL) used in this study is based on the X-ray crystal structure by Park et al [99], which was used in previous studies as well [118, 120]. A cartoon representation of the structure is given in Fig. 1.3 on page 6. The redox co-factor FAD is decomposed into a riboflavin and an ADP part. For the latter, force field parameters from Ref. [325] are used, whereas the riboflavin and the light harvesting co-factor MTHF (5,10-Methylenetetrahydrofolate) are parameterized according to the gaff forcefield [326]. Atomic charges for both co-factors (including both neutral and negatively charged forms of FAD) are determined from RESP calculations [237]. Additionally, charges for a Trp sidechain radical cation species need to be obtained. To maintain consistency with existing amino acid partial charge parameterization, these were determined by adding atomic charge differences to the respective atoms of a neutral Trp residue. The charge differences are obtained from RESP calculations for both the neutral and positive charged forms of the model molecule 3-methylindole (skatole). Atom labels and corresponding atomic charges of neutral and charged species for the Trp residue can be found in Fig. D.1 and Tab. D.1 in the appendix. Note that bond length, angle and dihedral force constants of the cationic Trp species are not altered compared to the neutral moiety. However, the relaxation of the Trp geometry due to the charge state, i.e. the internal reorganization energy λ_i, is determined beforehand by quantum chemical calculations and will be treated as a constant parameter which is used for the calculation of Marcus rates in Sec. 7.4 as well as for the non-adiabatic simulations as described in Sec. 2.6.1.

The protein is solvated in a rectangular box with 33,753 water molecules using the TIP3P model [236]. Periodic boundary conditions (PBC) are applied. All simulations are carried out with the GROMACS software package [197]. Amber parm99 [233] force field parameters with the extension of improved protein backbone parameters ff99SB [234] are used for the protein parts. The LINCS algorithm [229] is employed to keep bonds involving hydrogen at fixed length. After a standard heating-minimization protocol followed by a 500 ps equilibration phase, which is discarded afterwards, 30 ns MD simulations are performed with a time step of two fs. Snapshots of the molecular structures were saved every ps for which the charge transfer parameters were calculated with the SCC-DFTB-FO approach as described in Sec. 2.5.1.

Five classical MD simulations of PL in solution were carried out with the following setup: A) the co-factor FAD[1] and the three Trp residues carry no excess charge, respectively, B) FAD is negatively charged and C-E) FAD carries a charge of −1 and either one of the Trp residues 382, 359 or 306 carries a charge of +1. In what follows, Trp382, Trp359 and Trp306 are denoted as site 1, 2, and 3, respectively. Tab. 7.1 summarizes the notation which is used throughout the paper.

---

[1]For sake of simplicity the Riboflavin part will be referred to as FAD in the following.

Table 7.1: Simulation labels: Tryptophan residues Trp382, Trp359 and Trp306 are denoted as 1, 2 and 3, respectively. Total fragment charges q for the Trp residues and the riboflavin part of FAD are listed for the various simulations in units of e.

simulation	q(FAD)	q(1)	q(2)	q(3)
A	0	0	0	0
B	-1	0	0	0
C	-1	+1	0	0
D	-1	0	+1	0
E	-1	0	0	+1

7.2 Structural Properties

An important structural parameter controlling the CT process is the distance between the Trp sidechains, i.e. sites 1, 2 and 3, since the rate of CT within the tunneling regime depends exponentially on the distance. The relevant distances for both steps 1→2 and 2→3 are distributed normally, as indicated in Fig. 7.1(a).

Generally, the distances are significantly beyond van der Waals contact distances, exhibiting maxima between 7.2 Å and 8.2 Å. This is more than twice the stacking distance between base pairs in double-stranded DNA, over which fast CT is known to proceed. In comparison, large distances between the Trp residues should result in smaller electronic couplings, rendering a slower CT process in this protein.

The average distance 2→3 is 0.5–1.0 Å smaller than for 1→2 in Sim. A, B, C and D. The distance 2→3 in Sim. E differs from the previous cases in that it exhibits a bimodal distribution with maxima around 7.4 and 8.2 Å. Apparently, rearrangements between two (or more) conformations occur during the MD simulation once site 3 is charged. This observation is analyzed in more detail below.

However, a direct CT from site 1 to site 3 appears unlikely, due to the large distance 1→3 of about 11.5 Å.

Furthermore, the RMSD time series with respect to the crystal structure 1DNP, given in Fig. 7.1(b), reveals a fairly stable system in all simulations. The RMS fluctuation per amino acid (RMSF) for the whole protein (Fig. 7.1(c)) confirms the stability of the structure, although several amino acids exhibit an increased flexibility, i.e. RMSF>2 Å.

However, these thermally more active sidechains are not located in the protein domain which is of interest regarding the CT process, as indicated in the RMSF plot in Fig. 7.1(d) for a more restricted residue range containing the sites 1, 2 and 3 involved in the CT. Accordingly, the protein environment of the three sites appears to be rather stable, though the local domain around site 3 reveals increased thermal activity compared to sites 1 and 2. Moreover, the RMSF values increase slightly for all three sites in Sim. E compared to all the other simulations, thus further indicating the more pronounced dynamics of the protein domain once site 3 carries the hole charge.

(a) distance distribution

(b) RMSD

(c) RMSF complete

(d) RMSF cut-out
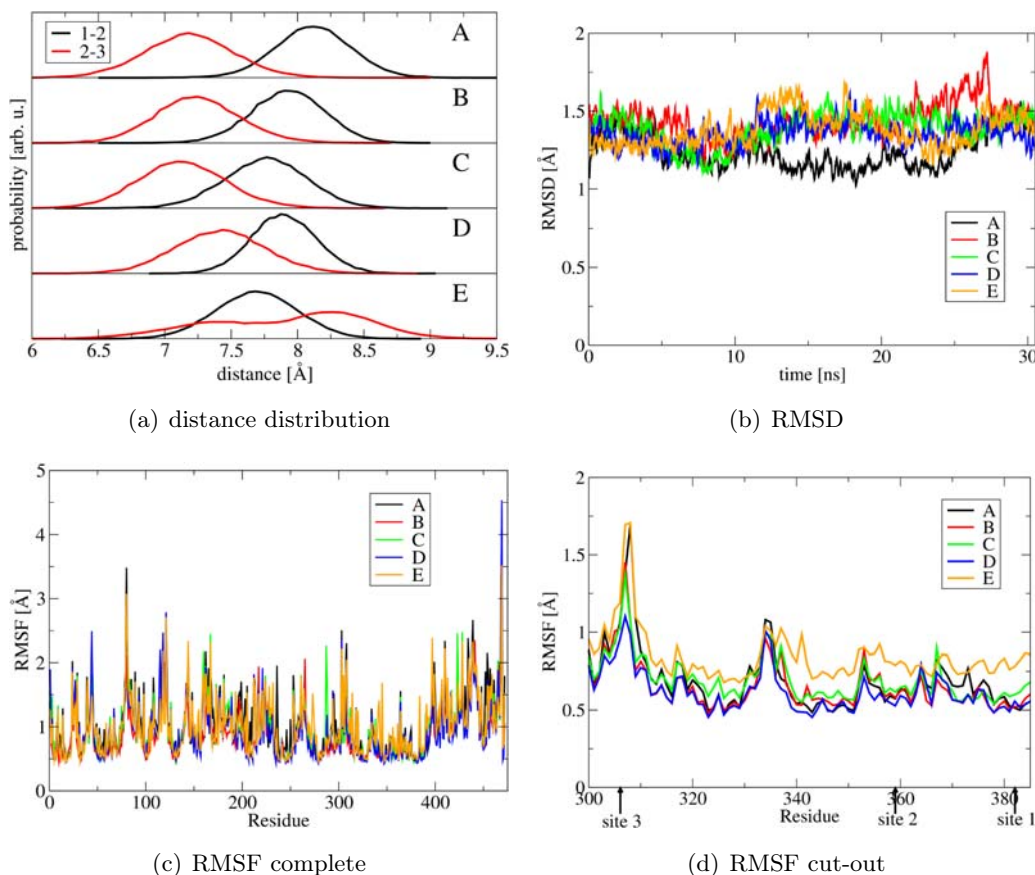
Figure 7.1: Structural analysis of simulation A-E (Notation from Tab. 7.1): a) Distribution of distances between sites 1 and 2 as well as between sites 2 and 3. b) RMSD calculation (Eq. 2.63) of c_α-atoms with respect to the X-ray crystal structure PDB-code: 1DNP [99]. c) and d) RMS fluctuations per residue (RMSF, Eq. 2.64), for the latter, the residue range is restricted to resolve the protein domain around the three Trp residues of interest.

7.2.1   Trp306 a Molecular Switch?

As mentioned above, the distance between sites 2 and 3 in Sim. E is distributed bimodally. The time series of the distance as well as of the corresponding electronic coupling is shown in Fig. 7.2. A strong correlation of these properties is apparent: Small electronic coupling is observed in time intervals where the distance is large, denoted as E2 in the following, while a larger coupling of up to 20 meV occurs in intervals with a shorter distance (E1).

As a consequence, there appear to be two major metastable local conformations exhibiting frequent transitions between them in the course of simulation E. This may be interpreted as some kind of switching behavior. If the CT process is completed, that is the hole charge has reached the final site 3, the corresponding protein domain undergoes local conformational changes, which may attenuate or even prevent back CT. Note that the metastable conformations E1 and E1 persists for several ns, respectively. These two conformations obtained as average structures from the E1 and E2 parts of the trajectory are visualized in Fig. 7.2. Their comparably small RMSD value of 0.44 Å illustrates the
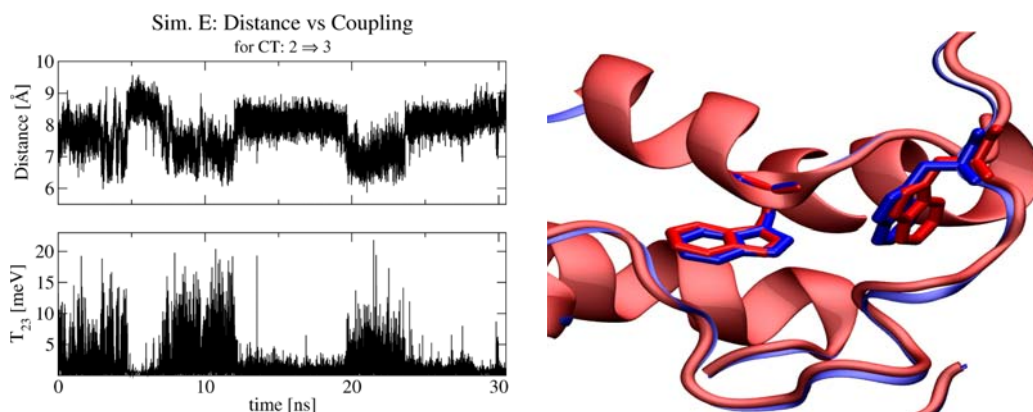
Figure 7.2: Two conformations of the protein are observed in Sim. E. Left: Time series of the distance between sites 2 and 3 as well as the corresponding electronic coupling T_{23}. Right: The averaged structure of the conformations E1 and E2 obtained as averages of the split trajectories.

rather subtle character of structural difference between E1 and E2.

Notwithstanding, the distance between sites 2 and 3 is 1 Å smaller in E1 than in E2, and precisely this difference is sufficient to affect the CT process substantially.

Further analysis is given in Fig. 7.3. The time series of various dihedrals and angles of site 3 are analyzed (Fig. 7.3(b)), to relate the two different conformations in Sim. E to any specific local structural change. As a result, no bimodal distributions are apparent, all curves reveal a smooth Gaussian shape (see Fig. D.1 for naming conventions).

Furthermore, a 2D-RMS plot was produced to determine if more than two sub-conformations are present (Fig. 7.3(a)). However, no clear separation can be made in this regard, only two subregions can be distinguished in the plot. Therefore, the switching between the observed CT domains may not be related necessarily to single local conformational changes in such a complex system as that of a protein in solution.

7.2.2 Water Distribution

The importance of aqueous solvent interacting with localized charges in CT active systems has been noted previously, notably in the context of CT in double-stranded DNA [173]. A hole charge can cause strong orientation polarization of the surrounding solvent, and the resulting induced electrostatic potential (ESP) strongly influences the energetics of CT. Therefore, the effect of polar solvent must not be overlooked in a study of bio-molecular CT. Accordingly, the distribution of solvent is studied in detail in the following. The distribution of solvent around the three CT-active tryptophan sidechains is different due to their different location inside the protein.

Site 1 is located 6–7 Å away from the FAD cofactor in the central region of the protein and there is no water until 8 Å away from it. Therefore, the hole located on site 1 would be expected to be affected marginally by the solvent, while the sites 2 and 3 are substantially closer to the surface of the protein and thus to water. This is reflected by the distribution of water (RDF, Eq. 2.66) setting on at about 5.5 and 4 Å, respectively

(a) 2D-RMSD

(b) Dihedral and angle distribution
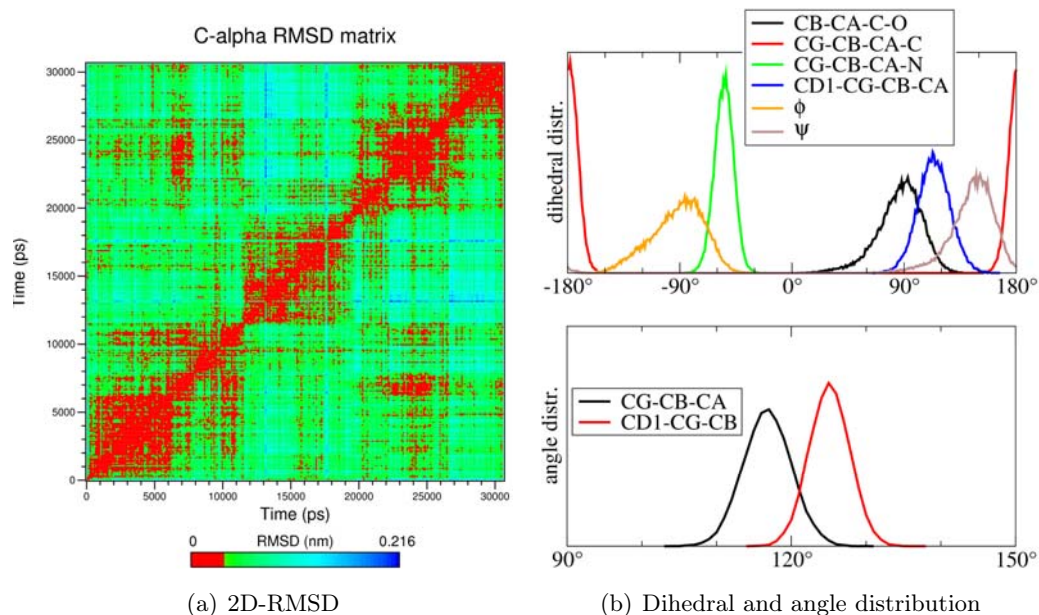
Figure 7.3: Additional structural analysis for Sim. E: a) 2D-RMSD-plot of Sim. E, that is each structure of the trajectory is compared to all the other structures. b) Distribution of various dihedrals (top) and angles (bottom) for site 3 (Trp306). Atom labels were used according to Fig. D.1.

(Fig. 7.4(a)).

On average, six water molecules can be found within 6 Å distance of site 3 once it is charged, while, at that distance, only one water molecule is located in the neighborhood of site 2 once it is charged (Fig. 7.4(b)).

Moreover, a plot of the change in water distribution (Fig. 7.4(c)) exhibits the formation of a solvent polaron around the positively charged sites 2 and 3 after effectively introducing the hole charge, i.e. with reference to Sim. B. For instance, the difference in water distribution for the hole located on the site 3 is negative between 7 and 5 Å while it turns positive between 5 and 3 Å; this can be understood as a rearrangement of water molecules moving closer to the hole charge or as a movement of site 3 towards the solvent. A similar behavior is observed when charging site 2, while no such effect is obtained for site 1 in Sim. C as could be expected beforehand.

As a result, the CT in PL may be affected considerably by the solvent, as revealed by the analysis of water distribution. The directionality of the CT benefits from increasing solvent stabilization of the hole charge along the path 1→2→3, for the distance to the solvent water reduces accordingly. The indication of a small solvent polaron formation may reveal additional stabilization of the hole charge at the final site 3, although this concerns only few water molecules on average. Nevertheless, the energetics of CT will be influenced significantly by the corresponding electrostatic interactions due to the close distance.

Notwithstanding, the protein environment may affect the CT energetics as well, pre-

(a) RDF   (b) Cumulative RDF   (c) ΔRDF with respect to Sim. B
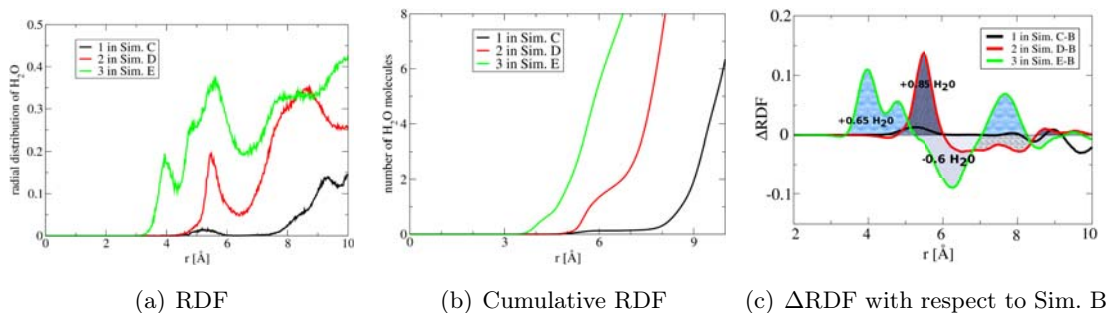
Figure 7.4: Solvent distribution around the center of mass of the CT sites. a) Radial distribution (RDF, Eq. 2.66) of H_2O molecules around the site 1, 2 and 3, respectively, in Sim. C, D and E. b) Cumulative RDF, that is the average number of solvent molecules within a certain distance around the respective CT sites. c) Plot shows the change in radial distribution functions when a site becomes charged, compared to Sim. B. While little change is seen for site 1, H_2O molecules are able to move significantly closer to sites 2 and 3 once one of them carries the hole charge, indicating the formation of a small solvent polaron in either case.

sumably by stabilizing the hole on site 1 inside the protein, thus competing with the solvent. Whether the protein or the solvent dominates the energetics is a major issue in this work and will be analyzed in the following sections in detail.

7.3   CT Parameter

In this section, the site energies ϵ_m (Eq. 2.67) and the electronic couplings T_{ij} (Eq. 2.68) are computed for the hole transfer in PL between the sites 1, 2 and 3. Dynamical and environmental effects, i.e. the separation into protein and solvent part, are analyzed using the data of MD simulations A–E.

Molecular orbitals   Similar to the proceeding in chapter 3 and 5, it has to be ensured that the fragment orbitals (FOs) used to describe the hole transfer reveal π-symmetry, and are described adequately with DFTB compared to higher level methods.

Therefore, the four highest MOs and their corresponding energies are computed with DFTB for an isolated skatole molecule (methyl-capped tryptophan sidechain), and compared to results obtained with DFT methods (B3LYP and PBE) as well as HF. The complete data (energies and MO snapshots) is given in Fig. A.6 in the appendix.

In summary, HOMO, HOMO–1 and HOMO–2 exhibit the same π-MOs with all applied methods. The relative energy differences between these MOs computed with DFTB agree very well with those obtained for the DFT methods. The only exception occurs for the HOMO–3, here a σ-orbital is obtained with DFTB and PBE, whereas B3LYP and HF give a π-orbital.

However, this discrepancy should not affect our calculations, since the energy of the HOMO–3 is separated largely from the energy of the HOMO (2.2 eV in DFTB).

(a) Distribution of ϵ_m (b) Mean ESP at site m
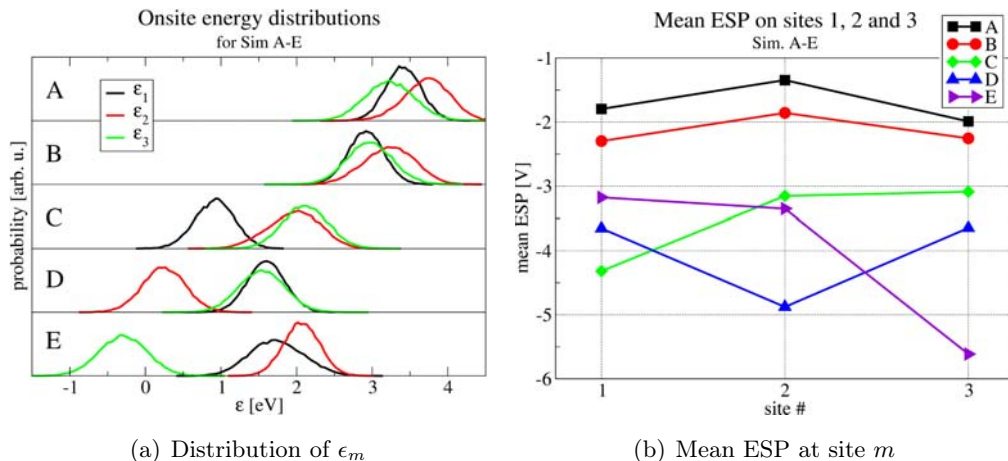
Figure 7.5: Distribution of site energies ϵ_m (left) and mean electrostatic potential (ESP, right) for simulations A-E. Note the direct connection between ϵ_m and ESP_m.

7.3.1   Protein and Water Fluctuations

The site energy ϵ_m is known to be crucial for the energetics of CT, as discussed in detail in the context of hole transfer in DNA in part I of this work.

There are two major contributions which determine the energetics of CT: (i) The fluctuations of the environment lead to fluctuations of the site energies ϵ_m, as analyzed for hole transfer in DNA in Refs. [152, 157, 164]. (ii) The environment is polarized due to the presence of the hole charge. In turn, this leads to a considerable stabilization of the corresponding site which carries the hole charge due to the ESP induced by the environment.

The contribution (i) can be inferred readily from Sim. A, in which all sites including the cofactor FAD are chosen to be charge neutral.

The site energies are distributed normally with standard deviations of 0.25–0.5 eV, as shown in Fig. 7.5(a). Interestingly, the fluctuations of site 1 (0.3 eV) are smaller than those of site 3 (0.5 eV). This is due to the fact that site 3 is more exposed to the solvent, thus the site energy is driven predominantly by solvent fluctuations. On the contrary, the fluctuations of site 1 are dominated by the protein fluctuations, which are slightly smaller but still significant. The magnitude of the fluctuations are not sensitive to the location or even the presence of the hole charge, as the values are rather similar for Sim. A–E.

However, the distributions in Fig. 7.5(a) are shifted very strongly due to charging of the FAD and the CT sites. This will be analyzed in more detail below. A table containing numerical values of the average site energies and corresponding standard deviations, which were obtained from the various simulations, is given in Tab. D.2 in the appendix.

To address the origin of the site energy fluctuations, which is related directly to the ESP (Eq. 2.139), a 10 ps MD simulation with the setup of Sim. B was performed. The

(a) ESP time series

(b) Distribution of ϵ_m

(c) Power spectrum of ϵ_1

(d) Power spectrum of ESP
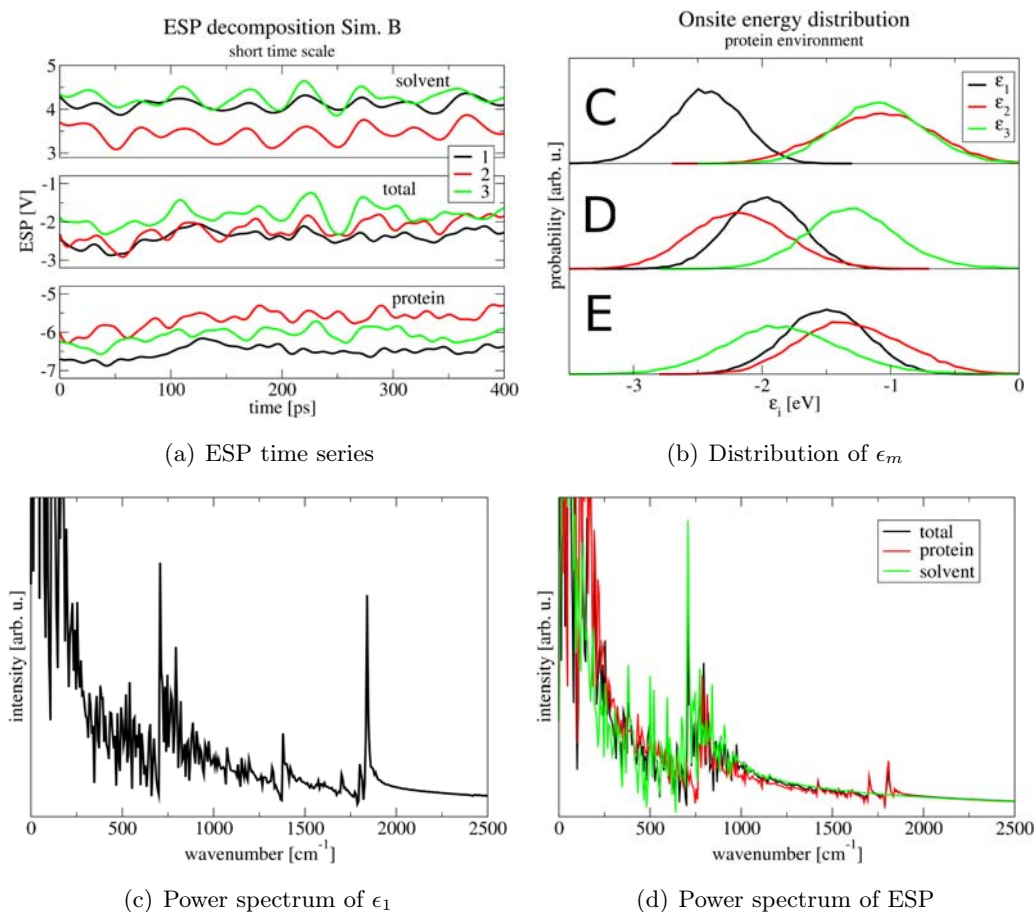
Figure 7.6: Analysis of site energy fluctuations: a) Total ESP vs time in Sim. B, as well as the ESP components due to the protein and solvent environment. b) Distribution of site energies ϵ_m regarding the protein environment only in Sim. C, D and E, neglecting the effect of the water solvent. c) and d) Power spectra (Fourier transform of the autocorrelation function) of, respectively, the site energy time series and the corresponding ESP with its components. Data is given for site 1.

ESP on the respective sites was calculated in intervals of one fs and decomposed into the solvent and protein contributions. The results are presented in Fig. 7.6.

The ESP on site 3 is controlled largely by the solvent component which is illustrated by the correlation coefficient of almost 0.8 given in Tab. 7.2 as well. By contrast, it is the surrounding parts of the protein that affect the electronic structure of sites 1 and 2 more markedly. Moreover, quite regular fluctuations of the ESP due to the solvent with a period of 40 to 50 fs are revealed, as shown in Fig. 7.6(a).

The corresponding power spectra of the site energy as well as the ESP time series are given in Figs. 7.6(c) and 7.6(d).

Two distinct bands are observed in the power spectrum of the site energy. The first band around 800 cm^{-1} can be assigned to the fluctuations of the environment, whereas the second band around 1800 cm^{-1} corresponds to the internal stretching modes of the tryptophan sidechains.

Table 7.2: Correlation of the total ESP and its components: protein and solvent environment obtained from a 10 ps trajectory of Sim. B.

$\rho(x, y)$	1	2	3
total, solvent	0.56	0.53	0.77
total, protein	0.70	0.69	0.33
solvent, protein	-0.20	-0.24	-0.34

Moreover, the decomposition of the ESP reveals bands around 750 cm^{-1} in the spectrum of solvent as well as around 800 cm^{-1} in the spectrum of the protein environment. Both movements manifest themselves in a combined band in the power spectrum of the site energy around 800 cm^{-1}. This corresponds to a oscillation period of about 40 fs. The less distinct band at 1700 cm^{-1} in the power spectrum of ESP originates from Amide I modes of the protein backbone.

7.3.2 Polarization of the Environment by the Hole Charge

As discussed above, the polarization of the environment due to the hole charge will lead to a drastic change in the ESP at site m, where the hole is localized. As a result, the site energy ϵ_m is lowered drastically by more than 1 eV, which leads to a (self-) trapping of the hole charge. Such a phenomenon may be regarded as a 'polaron' referring to similar effects in solid states physics. Conwell and coworkers [289] reported a similar observation in context of hole transfer in DNA, which was was discussed later on in great detail in Ref. [173]. The methodology of the latter work is adapted in this study for a protein indicating a similar finding, as revealed in Fig. 7.5(b).

The mean ESP at sites 1–3 are shown for Sims. A–E. The values appear to be considerably different in Sims. C–E, while they are quite similar for Sims. A and B, shifted only slightly due to the negative charge on FAD in Sim. B.

The ESP decreases by 1-2 eV at the site where the charge is located. Similarly, the shift of the occupied site in Sims. C–E is observed in the ϵ_m distributions in Fig. 7.5(a) as well, since ESP$_m$ and ϵ_m are related directly. It is important to note that this polarization effect leads to a self-stabilization or self-localization of the hole charge. A detailed discussion on this issue is given in Ref. [173].

As shown in Fig. 7.5(a), the site energies are ca. 0.5 eV smaller in Sim. B than in Sim. A, which is only due to the negative charge of the FAD co-factor. Clearly, the decrease is larger for the sites 1 and 2 (0.5 eV) than for the site 3 (0.3 eV), because the latter is located further away from the FAD moiety.

The introduction of a hole charge on site 1 (Sim. C) leads to a strong decrease of ϵ_1 by 2 eV (compared to Sim. B). Similarly, the adjacent sites 2 and 3 experience a drop of 1.3 and 0.9 eV, respectively. This is a consequence of the orientation polarization of the environment (solvent and protein), undergoing conformational reorganization due to the

Table 7.3: Mean electrostatic potentials (ESP) on CT sites 1–3 for Sims. A–E decomposed into protein and solvent environment. All values in V.

Sim.	ESP$_1$		ESP$_2$		ESP$_3$	
	protein	solvent	protein	solvent	protein	solvent
A	-5.3	3.5	-4.8	3.4	-5.5	3.5
B	-6.6	4.3	-5.9	4.0	-6.2	3.9
C	-7.7	3.3	-6.3	3.1	-6.4	3.3
D	-7.2	3.5	-7.4	2.5	-6.6	3.0
E	-6.7	3.5	-6.4	3.1	-7.3	1.7

hole charge and thereby, inducing a strong electric field back on the site 1 as well as the adjacent sites.

In either case, the strongest site energy drop is observed for the site that carries the hole charge. Note that this drop in ϵ_m results from a strong decrease in the induced ESP$_m$, as shown in Fig. 7.5(b).

Furthermore, the site energies for the sites carrying the hole charge in Sims. C, D and E decrease from site 1 to site 2 by 0.6 eV and further from site 2 to site 3 by another 0.5 eV. Therefore, this indicates a down-hill type CT reaction, in which the hole charge is stabilized most strongly on site 3.

Accordingly, the driving forces of the first and second CT step can be approximated to $\Delta\epsilon_{1\to2} = -0.7$ and $\Delta\epsilon_{2\to3} = -0.5$ eV, respectively, by taking the differences of the respective ϵ_m for Sims. C–E. Note however that the differences of ϵ_m neglect a part of the energy, as discussed in Sec. 2.6, as well as entropic effects. Therefore, these values should be used as a first estimate only. A more rigorous calculation of $\Delta G_{m\to n}$ is presented in Sec. 7.4.

In a similar fashion, the ϵ_m differences within one simulation can be used to estimate the sum of $\Delta G_{m\to n} + \lambda$, as discussed in Sec. 7.4 as well.

Protein and Solvent Contributions

As indicated in Tab. 7.3, the negative charge located on the FAD (included as part of the "protein" here) induces a decreased negative ESP, stabilizing the (positive) hole charge on the specific tryptophan sidechain. By contrast, the solvent induces an increased positive ESP in Sim. B, therefore destabilizing the hole by 0.8, 0.6 and 0.4 eV for sites 1, 2 and 3, respectively. However, the effect of the protein part (FAD) prevails. It should be noted that ESP values should be discussed only with respect to an arbitrary zero, therefore any interpretation should focus on ESP differences only.

Introducing the hole charge in Sims. C–E, leads to a decrease of the ESP induced by the protein and the solvent. The hole stabilization of the protein part is more pronounced for the sites 1 and 2. On the contrary, the solvent part of ESP on site 1 is hardly affected

by the location of the hole, since the values in Sims. C–E are rather similar.

On the other hand, the solvent potential on site 3 drops considerably by 1.3-1.6 V in Sim. E compared to Sims. C and D. Apparently, a hole charge on site 1 is stabilized predominantly by the protein, whereas the solvent plays the key role on site 3. However, both solvent and protein environment seem to play an important role for site 2.

The fluctuations of ESP in both solvent and protein environment reveal the same magnitude of 0.3 to almost 0.5 eV, with the exception of the ESP fluctuations on site 1, which are reduced considerably compared to sites 2 and 3. This indicates a less flexible environment around site 1, as suggested earlier.

Accordingly, the simulations exhibit an important functional role of water. On the one hand, it introduces fluctuations of the site energies, thereby enhancing the dynamics of CT. On the other hand, the solvent stabilizes predominantly site 3 and, to a lesser degree, site 2, thus enabling a strong of driving force of the CT.

Fig. 7.6(b) shows the distribution of site energies ϵ_m when the effect of the solvent environment is neglected. For Sim. C, a considerable stabilization of site 1 is observed, while for Sim. D and E this stabilization is absent. This is due to the fact that the stabilization of a hole charge on the sites 2 and 3 is mostly due to water polarization. As a result, sites 2 and 3 are higher in energy than site 1, i.e. the CT $1\rightarrow2\rightarrow3$ would be endothermic.

In conclusion, the results indicate that the CT in PL is driven predominantly by the surrounding water.

7.3.3 Charge Transfer Couplings T_{ij}

The previous sections revealed the specific energetic characteristics of the CT in PL by means of site energies and electrostatic potentials. However, the CT is enabled only if the electronic couplings are of sufficient magnitude, even though the reaction is strongly exothermic.

Therefore, a detailed analysis of the CT couplings T_{ij} for the various charge states along the CT $1\rightarrow2\rightarrow3$ is of major significance as well. Fig. 7.7 displays the distributions of electronic couplings T_{12} and T_{23} for Sims. A–E.

While the T_{12} exhibits a Gaussian-like distribution around the maximum at 5–10 meV, the coupling T_{23} is much smaller on average and seems to be distributed around zero. Note that the couplings are significantly smaller than those observed for the hole transfer in DNA in part I, where the distance between neighboring sites is only 3–4 Å.

Numerical values of the average T_{ij} with standard deviations for Sims. A–E are given in Tab. D.2 in the appendix. Interestingly, the mean values for T_{12} increase when introducing charges in Sim. B and C, while the opposite is true for T_{23}, as the mean value decreases from 4.3 meV in Sim. A to only 1.5 meV in Sim. E. The latter observation could be interesting regarding the probability of back CT once the hole charge reaches the site 3. This could be related with the "switching-like" behavior discussed in Sec. 7.2.1, thus indicating that back CT $1\leftarrow2\leftarrow3$ may be suppressed.
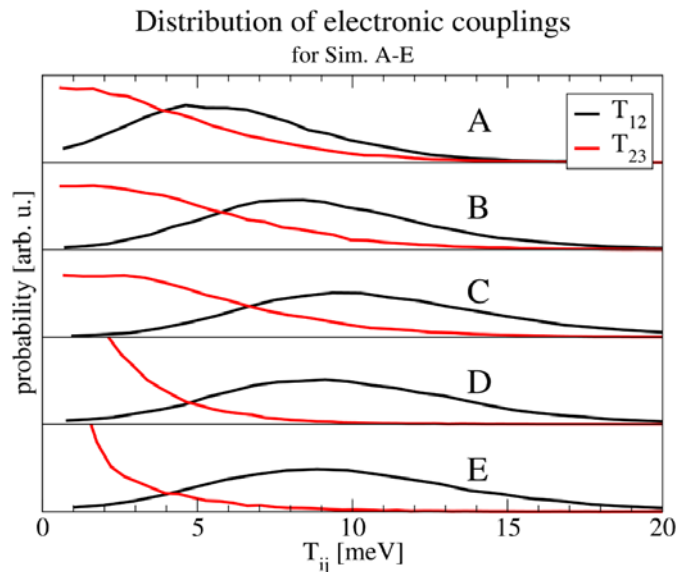
Figure 7.7: Distribution of electronic couplings T_{ij} for simulations A–E.

Furthermore, the mean values of T_{ij} for Sim. C–E in Tab. D.2 agree quite well with those obtained by Krapf et al. [120], although a different computational strategy was used in this work.

As a consequence, the distribution of the couplings in Fig. 7.7 suggests already that the first step of the CT may be described adequately using average couplings, while the transfer from site 2 to 3 may be dominated by the fluctuations of the couplings, giving rare conformations with high T_{23}-values an outsize influence on the CT.

The effect of fluctuations on the electronic couplings in biomolecules (especially proteins) was investigated in great detail by Beratan, Skourtis and coworkers, who studied a large number of CT systems in proteins and DNA [159, 160]. As described in Ref. [159], the coherence parameter C

$$C = \frac{\langle T_{ij} \rangle^2}{\langle T_{ij}^2 \rangle} = \frac{1}{1 + \frac{\sigma^2}{\langle T_{ij} \rangle^2}}, \tag{7.1}$$

introduces a good measure for the importance of fluctuations.

Based on Eq. 7.1 $C \approx 1$ indicates that CT is controlled mainly by the average coupling, while $C \ll 1$ indicates that the transfer is dominated by coupling fluctuations due to strong dynamical disorder. The coherence parameter for electronic couplings T_{12} and T_{23} of both CT steps 1→2 and 2→3 is given in Tab. 7.4.

Table 7.4: Coherence parameter for both electronic couplings T_{12} and T_{23}, for Sim. A–E.

	A	B	C	D	E
1→2	0.77	0.84	0.85	0.84	0.82
2→3	0.62	0.60	0.62	0.50	0.35

While C for both couplings is larger than typical values in DNA [166, 167], the transfer 1→2 appears to be considerably more "coherent" than 2→3. The coherence of the second step drops markedly in Sim. E. Therefore, the fluctuations of T_{ij} might be less important than one would think beforehand. This issue will be addressed further in Sec. 8.4.

7.4   Thermodynamic Quantities and CT Rates

In this section, the thermodynamic quantities entering the Marcus equation (Eq. 2.4), that is the free energy difference ΔG and the reorganization energy λ, are computed from the classical MD simulations. Various computational strategies are presented in either case. Finally, these parameters are used to estimate CT rates based on Marcus theory of ET. The theoretical concepts of the various computations of λ and ΔG are described in detail in Sec. 2.2.2 and Sec. 2.2.3, respectively.

7.4.1   Outer-Sphere Reorganization λ_s

The reorganization of the environment is expected to be the major contribution to the overall λ for inter-molecular CT in aqueous solution. Nevertheless, the internal reorganization λ_i might be the dominant contribution for CT in systems with non-polar solvents or in organic crystals.

In this work, λ_s for the various CT steps is obtained from the classical trajectories of Sims. A–E using three different computational strategies:

i The energies λ_{s_f} and λ_{s_b} are obtained directly via Eq. 2.12 for the forward and backward CT, respectively. The approach is described in detail in Sec. 2.2.2.3. Note that this is already a step beyond Marcus' theory, as λ_s is not assumed to be the same for the forward and backward CT.

ii λ_s is computed using Eq. 2.10 and Eq. 2.11 as described in Sec. 2.2.2.2. By contrast to strategy (i), this calculation is based on the assumption of equal reorganization energies for forward and backward CT.

iii λ_s is determined as the difference of ESP (ΔESP) between the two sites involved in the CT step, e.g. the ESP difference between the sites 1 and 2 in Sim. C. This is possible, since the complete interaction energy of the CT sites with the MM environment can be expressed as the product of the charge e and the ESP on the respective site. Again, it is assumed that $\lambda_{s_f} = \lambda_{s_b}$.

Recently, protocol (i) has been applied successfully to determine λ_s for the hole transfer in DNA [183].

The values of λ_s obtained with the various strategies (i), (ii) and (iii) are presented in Tab. 7.5. Moreover, the results are compared to those obtained in Ref. [120].

Table 7.5: Outer-sphere reorganization energies λ_s obtained from Sims. C–E, calculated using the computational strategies (i)–(iii). Comparison to results obtained in Ref [120]. All values in eV.

CT step	$\lambda_{s_f}^{(i)}$	$\lambda_{s_b}^{(i)}$	$\lambda_s^{(ii)}$	$\lambda_s^{(iii)}$	$\lambda_s^{\text{Ref. [120]}}$
1-2	0.9	1.6	1.3	1.2	1.0
2-3	1.5	1.9	1.7	1.7	1.0
1-3	1.4	2.3	1.8	1.8	1.2

To begin with, protocols (ii) and (iii) exhibit very similar λ_s, which is basically the average of forward and backward λ_{s_f} and λ_{s_b} in (i). The value of λ_s for the CT step $1\rightarrow2$ is considerably smaller than for the step $2\rightarrow3$. This is surprising as the average distance between the sites 1 and 2 was observed to be larger than between the sites 2 and 3 (see Fig. 7.1(a)). Note that previous studies showed λ_s to increase asymptotically with the distance [183, 191].

However, the smaller values for $1\rightarrow2$ may be explained by assuming the reorganization of water molecules to be the major contribution to λ_s. Accordingly, the site 1 and to lesser extent site 2 are located inside the protein, which reduces the influence of solvent on the first CT step.

Moreover, λ_s for $2\rightarrow3$ and $1\rightarrow3$ are of the same magnitude due to the asymptotic behavior of λ_s with respect to the distance[2].

More importantly, the values obtained with (i) demonstrate clearly that λ_{s_f} and λ_{s_b} can be substantially different. In this case, the system reorganization for the backward CT requires considerably more energy than for the forward CT. As a consequence, the results suggest to use protocol (i) for further applications. The assumption of $\lambda_{s_f} = \lambda_{s_b}$ in (ii) and (iii) appears to be inadequate in such heterogeneous CT systems.

By contrast to the good agreement of procedures (i), (ii) and (iii), the reorganization energies reported in Ref. [120] are considerably smaller, especially for $2\rightarrow3$ and $1\rightarrow3$.

Notwithstanding, various theoretical studies revealed that reorganization energies computed from classical non-polarizable force field simulations overestimate λ_s, as discussed in more detail in Sec. 7.4.4.

7.4.2   Internal Reorganization λ_i

In this work, λ_i is estimated by performing quantum chemical calculations on a single 3-methylindole (skatole) molecule. The internal reorganization is obtained via Eq. 2.13 as described in Sec. 2.2.2.5. Values of λ_i were computed using various quantum chemical methods. The data is presented in Tab. 7.6.

As a result, the DFT methods provide smaller values of λ_i, while a considerably larger value is obtained with HF. Therefore, B3LYP reveals an intermediate value which

---

[2]Note the average distance between the sites 1 and 3 of ca. 11.5 Å.

Table 7.6: Internal reorganization energies λ_i for the hole transfer between two tryptophan sidechains, estimated with Eq. 2.13 for a single 3-methylindole molecule. Comparison of quantum chemical methods: DFTB, PBE, HF and B3LYP using a 6-31G(d,p) basis set. For the latter also the 6-311G(2df,2p) basis set was used. All values in eV.

method	λ_i
DFTB	0.25
PBE	0.26
HF	0.45
B3LYP	0.34
B3LYP[3]	0.36

is expected to be more adequate than the pure DFT or HF values. Additionally, the B3LYP value is affected only marginally on the basis set.

Note that these QM calculations are carried out in vacuo, and the values may change significantly if the molecular environment is considered, e.g. using an implicit solvent model. Nevertheless, a value of 0.36 eV is used for λ_i as a parameter in the following applications.

7.4.3   Free Energy Difference ΔG^0

The driving force of the CT $1{\to}2{\to}3$ and its single steps was estimated in Sec. 7.3 using the site energy difference $\Delta\epsilon$, yet, neglecting parts of the MM energy contribution as well as entropic effects. However, an adequate calculation of the driving force ΔG is realized by performing so-called free energy simulations. There are various computational concepts which are based either on free energy perturbation (FEP) or on thermodynamic integration (TI). All applied methods are described in detail in Sec. 2.2.3.

In this work, the free energy difference ΔG is computed using the following theoretical approaches:

- **ΔE**: The free energy difference is estimated by the difference of the total MM energies obtained from the respective MD simulations via Eq. 2.18. For instance, ΔG for the CT $1{\to}2$ is computed as the difference of total energies obtained in Sims. C and D. This protocol may be afflicted with large statistical errors, as described in Sec. 2.2.3.

- **DTI**: This is the standard version of thermodynamic integration (Eq. 2.26), as the derivative of the MM energy with respect to the coupling parameter Λ is computed for fixed values of Λ. The theory behind DTI is described in Sec. 2.2.3.2. Accordingly, 11 equilibrium TI simulations with a length of 40 ns were performed for each CT step, respectively, using fixed values of Λ between 0 and 1 in steps of

---

[3]Basis set: 6-311g(2df,2p)
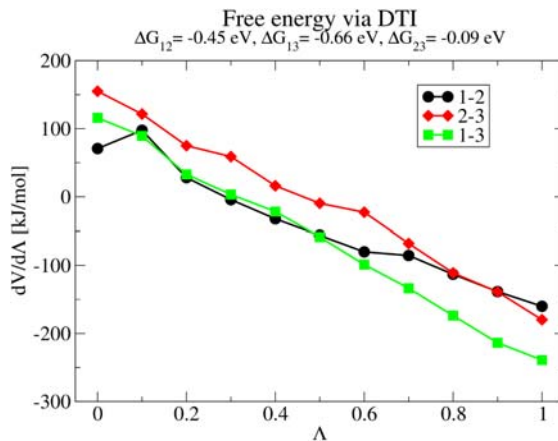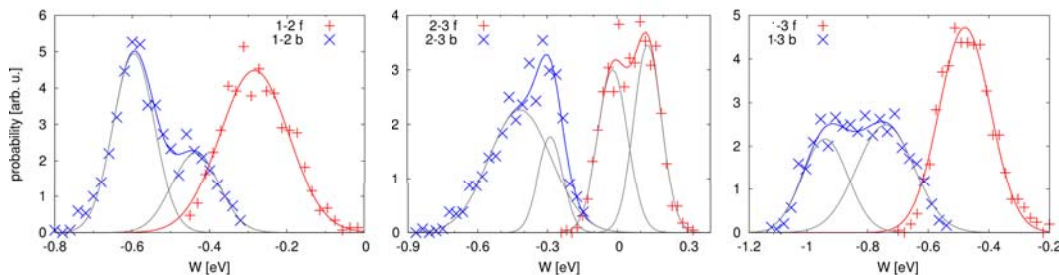
Figure 7.8: DTI calculations of ΔG: For each CT step, the values of $\left\langle \frac{\partial E_\Lambda}{\partial \Lambda} \right\rangle_{\Lambda_i}$ are shown for the various fixed Λ. Eventually, the corresponding ΔG values are computed via numerical integration (Eq. 2.26). For each CT step and each Λ, a 40 ns MD simulations was performed at a fixed value of Λ.

0.1^4. A summary of the DTI results, i.e. the $\left\langle \frac{\partial E_\Lambda}{\partial \Lambda} \right\rangle_{\Lambda_i}$ values for the various Λ and the three single CT steps is given in Fig. 7.8.

- **FGTI-JAR**: By contrast to the DTI method, ΔG is computed from a series of non-equilibrium work simulations using the Jarzynski equality (Eq. 2.27). The concept of FGTI is described in-depth in Sec. 2.2.3.3. Technically, 1000 free energy simulations with a length of 80 ps were carried out for each CT step, respectively. In each of these simulations an irreversible work W is obtained by quickly changing Λ from 0 to 1 within the short simulation time of 80 ps. The simulations were started from different configurations of the corresponding equilibrated ensemble, e.g. the Sim. C for the CT step $1\rightarrow2$.

- **FGTI-CFT**: The former FGTI method is extended, as now two series of W are obtained for the forward and backward CT, respectively. The free energy can then be computed as the crossing point of the corresponding work distributions according to Eq. 2.29. The method is described in Sec. 2.2.3.3 as well. Therefore, in addition to the simulations from the previous method for the forward CT steps, the same amount of work simulations is performed for the backward CT as well. For instance, the calculation of ΔG for the first CT step $1\rightarrow2$ requires additional 1000 work simulations starting from different configurations of Sim. D. Note that the definition of states corresponding to $\Lambda = 0$ and $\Lambda = 1$ is reversed for the back CT. The results of the FGTI-CFT method is illustrated in Fig. 7.9 by means of the work distributions computed, respectively, for the forward and backward CT.

The values of ΔG obtained with the various methods and those obtained in Ref. [120] are given in Tab. 7.7. The ΔE approach exhibits considerably larger values, in particular

---

[4]Here, $\Lambda = 0$ refers to the initial state, e.g. hole on site 1, while $\Lambda = 1$ refers to the product state, e.g. hole on site 2.

Figure 7.9: FGTI-CFT calculations of ΔG: The crossing of the non-equilibrium irreversible work distributions for the forward and backward CT is designated as the driving force ΔG (Eq. 2.29). For each CT step and each direction, the data is obtained from 1000 free energy simulations with a length of 80 ps, quickly shifting Λ from the initial to the final state.

for the CT steps 2→3 and 1→3, as compared to the more advanced approaches and also compared to Ref. [120] in which the DTI method was used as well. This indicates that the computation of the free energy based on total MM energies is inadequate and may lead to severe overestimations of the driving force ΔG for CT in PL.

Furthermore, the applied TI methods show a very good agreement, though the ΔG value for the second CT step 2→3 is slightly increased with FGTI-CFT (-0.15 eV) compared to DTI and FGTI-JAR (-0.09 eV). On the other hand, the values differ significantly from those reported in Ref. [120]. All applied methods exhibit a stronger driving force for the first CT step 1→2, while the situations is reversed for the values from Ref. [120]. On the contrary, the values for the long-distance CT step 1→3 are rather similar.

In both studies, the same starting structure as well as the same force field is used, only the employed solvent model is different. In this work, the TIP3P solvent model is used, while the TIP4P-EW model was used in Ref. [120]. However, it is not clear if the qualitative difference of the ΔG values may be related to the different solvent models.

Usually, the work distributions in the FGTI-CFT method are expected to reveal Gaussian-shaped functions. Here, this holds only for the distributions which were obtained from Sim. C, i.e. in which the hole charged is located on site 1. Interestingly, the distributions for the remaining cases can be represented by the sum of two Gaussian functions, as indicated in Fig. 7.9. This implicates the presence of two local system conformations once either site 2 or 3 carries the hole charge. For the latter case, this may be related to the structural changes observed in Sec. 7.2.1. Nevertheless, the presence of

Table 7.7: Free energy differences ΔG obtained from Sims. C, D and E. All values in eV.

CT	ΔE	DTI	FGTI-JAR	FGTI-CFT	Ref. [120]
1-2	-0.6	-0.45	-0.37	-0.40	-0.1
2-3	-0.5	-0.09	-0.09	-0.15	-0.3
1-3	-1.1	-0.66	-0.58	-0.62	-0.5
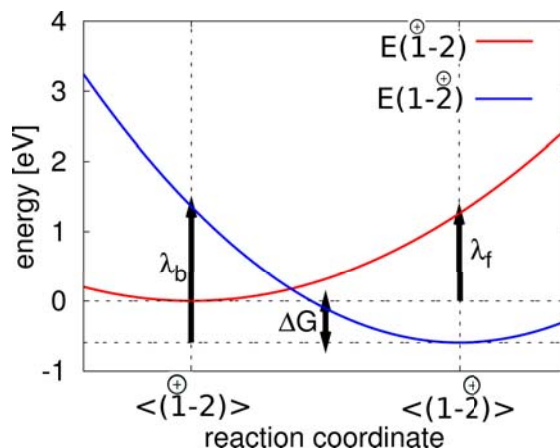
Figure 7.10: Energetics of forward (k_f) and backward (k_b) hole transfer between the sites 1 and 2 based on reorganization energy $\lambda_{f/b} = \lambda_{s_{f/b}} + \lambda_i$ and ΔG from Tabs. 7.5 and 7.7. The equilibrium energy of the initial state ($1^\oplus - 2$) is set to zero.

non-Gaussian work distributions may also indicate that the FGTI-CFT results are not yet converged completely.

Generally, the hole transfer $1{\Rightarrow}2{\Rightarrow}3$ is strongly exothermic, yet λ_s turns out to be larger than ΔG for all CT steps, thus placing the observed CT firmly into the classical realm of the Marcus parabola.

7.4.4 CT Rates

In this section, the thermodynamic parameters computed in the preceding section are employed to estimate rates of CT in the PL system via the Marcus equation (Eq. 2.4). To begin with, the mean squared electronic coupling $\left\langle T_{ij}^2 \right\rangle$ is used for the donor-acceptor coupling in Eq. 2.4 [5]. The values of λ_s are taken from option (i) in Tab. 7.5, while the ΔG values are taken from the FGTI-CFT method in Tab. 7.7. Moreover, an internal reorganization energy of $\lambda_i = 0.36$ eV is used.

The scheme depicted in Fig. 7.10 illustrates the energy profile of the first CT step $1{\rightleftarrows}2$. Note that this represents already a step beyond the framework of Marcus' original theory as the curvatures of the PES corresponding to the initial and the final states are different due to the different λ for the forward and backward reactions.

Evidently, the backward CT $1{\leftarrow}2$ is disfavored because of the large reaction free energy of 0.4 eV and the reorganization energy amounting to almost 2 eV in total ($\lambda_s + \lambda_i$). This makes the exponential term in Eq. 2.4 quite small, hence the rate will be extremely slow. Combining the parameters λ and ΔG with the mean electronic couplings T_{ij} taken from Tab. D.2 in the appendix, the corresponding Marcus' rates can be calculated for the separate CT steps. The results are presented in Tab. 7.8.

The values should be discussed only in a qualitative manner because of the expo-

---

[5] As indicated in Tab. 7.4, the coherence parameter is smaller than 1, i.e. the transfer is dominated mainly by coupling fluctuations, thus $\left\langle T_{ij}^2 \right\rangle$ is used in Eq. 2.4 rather than $\left\langle T_{ij} \right\rangle^2$

Table 7.8: Reaction rates calculated with Eq. 2.4 according to Marcus' theory of electron transfer. T_{ij} taken from Tab. D.2 in the appendix. Note that the internal reorganization energy λ_i is included in the rate evaluation as $\lambda_{f/b} = \lambda_{s_{f/b}} + \lambda_i$. λ and ΔG in eV, k in s^{-1}

step	λ_{s_f}	λ_{s_b}	ΔG	no scaling		scaled $\lambda_s{}^a$		scaled $\lambda_s{}^b$	
				k_f	k_b	k_f	k_b	k_f	k_b
1→2	0.9	1.6	-0.40	5×10^9	1×10^0	5×10^{10}	9×10^1	3×10^{11}	2×10^3
2→3	1.5	1.9	-0.15	2×10^4	6×10^{-1}	2×10^6	2×10^2	4×10^7	1×10^4
1→3	1.4	2.3	-0.62	1×10^3	9×10^{-13}	3×10^4	4×10^{-10}	3×10^5	3×10^{-8}

^{a} λ_s scaled by $1/1.4$; ^{b} λ_s scaled by $1/2$.

nential dependence of the rate of CT on the energies used, which makes modest changes of the parameters result in outsize effects. As indicated in Tab. 7.8, λ_{s_b} is significantly larger than λ_{s_f}, thus resulting in dramatically slow back CT rates. Interestingly, the rate for 1→2 is five orders of magnitude larger than for 2→3, since $\lambda + \Delta G$ is about 1 eV smaller, plus the coupling T_{12} is five times larger than T_{23}.

It is well known that reorganization energies computed from non-polarizable force field simulations overestimate λ_s significantly [136, 184–186]. Accordingly, the values of λ_s should be scaled by the inverse optical dielectric constant of the medium, $1/\epsilon^{\mathrm{opt}}$. However, the quantity ϵ^{opt} is not accessible directly for heterogeneous environments such as a solvated protein. Various scaling factors in the range of $1/2 - 1/1.4$ were suggested in the past [127, 134, 135, 187, 188]. Therefore, results for two different scaling factors $1/2$ and $1/1.4$ are presented in Tab. 7.8. The large effect that a change of λ has on the rate k is readily apparent, indicating again that the results should be taken qualitatively only. Furthermore, the scaling decreases the rate for the first CT step 1→2 into the ps range, while $k_{2\to3}$ remains much smaller.

According to the results presented in Tab. 7.8, the CT from site 1 to site 2 takes about 3 to 20 ps depending on the scaling factor used. This agrees very well with experimental findings that report the first CT step 1→2 to occur within 9 ps [98, 104, 111, 112].

On the other hand, these experiments revealed only slightly larger transfer times for the second CT step 2→3 of less than 30 ps. Therefore, the computed rate for the second step is orders of magnitude too small with transfer times ranging from 25 to 500 ns.

Notwithstanding, the surprisingly small rates for the second CT step 2→3 can be explained. In the Marcus' picture, the whole CT process 1→2→3 is decomposed into two separate, independent reactions, for which it is assumed that each initial state (i.e. hole on site 1 or 2, respectively) is fully equilibrated, that is the surrounding was given enough time to stabilize the charge due to reorganization. However, this does not seem to be the case here. While the solvent reorganization is rather fast, namely within several hundreds of fs, there may be additional relaxations occurring on a substantially longer time scale of up to hundreds of ps involving reorganization in the protein environment. This seems to be the case for the second step 2→3, as will be discussed in Sec. 8.3.

8

Electron-Ion Dynamics

The results in the previous section indicate that the charge transfer in PL requires a theoretical description beyond Marcus' equilibrium theory. The charge motion and the structural relaxation seem to occur on comparable time scales, hence a clearly non-adiabatic framework is needed, in which the charge motion is directly coupled to the motion of the environment. However, reduction of complexity is needed by means of molecular coarse-graining (CG) in order to get such a computational framework to work for such an intricate system. Surprisingly, this approximate reduction of complexity may, on the other side, lead to a more comprehensive understanding of the CT in PL, as will be discussed throughout this chapter.

Therefore, non-adiabatic coupled electron-ion dynamics simulations (EID) were performed to study the charge redistribution in time directly. The methodology is described in Sec. 2.6, and in-depth in Ref. [173], in which it had been applied in the study of hole transfer in DNA. The setup assumes the photoactivation process to proceed according to the following mechanism:

To begin with, CT is initiated by electronic excitation of the FAD co-factor (possibly via excitation transfer from MTHF) followed by a fast hole transfer to the adjacent tryptophan residue site 1 (Trp382). At this point, further dynamics occur solely in the electronic ground state. The cofactor remains negatively charged, while the hole charge is free to move between site 1, the adjacent site 2 (Trp359) and site 3 (Trp306) located on the exterior of the protein.

In the following, the focus is centered on this three site hopping process, i.e. the direct charge transfer $1 \rightleftharpoons 2 \rightleftharpoons 3$, starting with the hole charge completely located on site 1 and treating all three tryptophan residues with the QM/MM methodology described
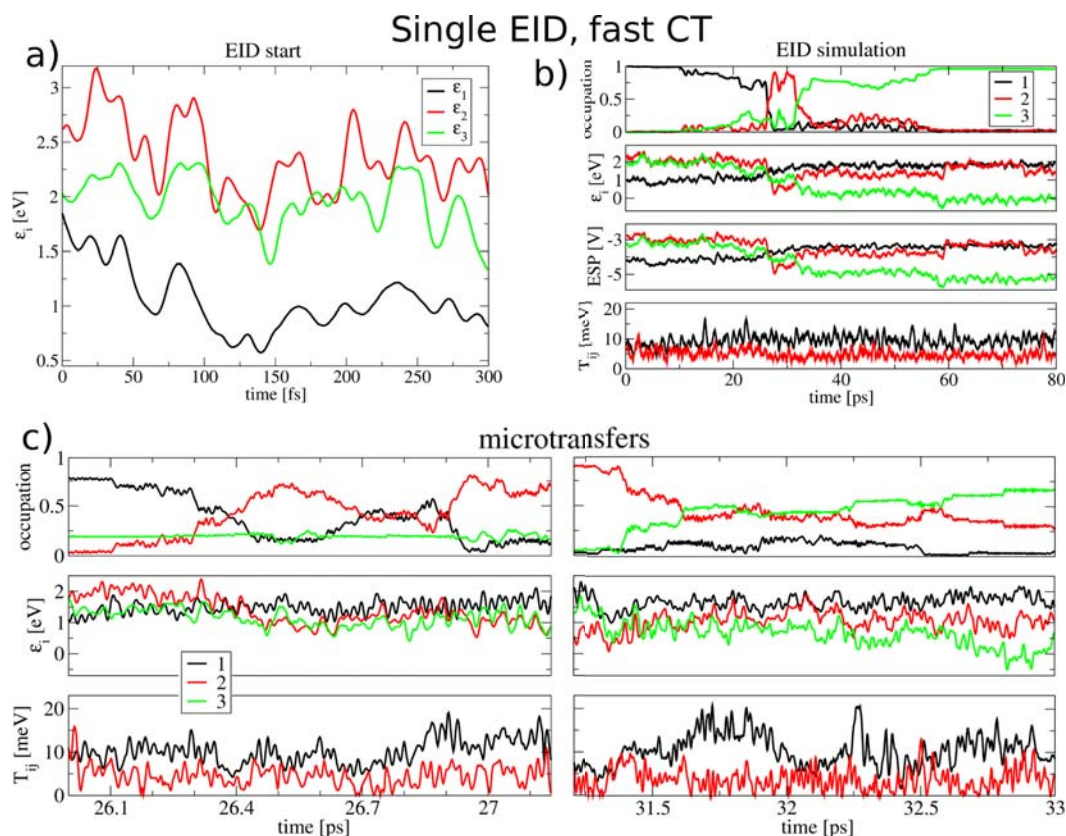
Figure 8.1: Non-adiabatic Ehrenfest dynamics (EID) in PL, example of a fast CT: a) Site energies for the first 300 fs, b) CT parameters, occupations and ESP for all three sites along an 80 ps trajectory (running averages over 50 fs are used to smoothen the graphs), and c) a more detailed cut-out for the two CT events at 27 and 32 ps (bottom).

above. Further reactions like deprotonation of the Trp radical cation are not considered. No prior protein relaxation is considered, since the initial charge separation is assumed to be faster.

Moreover, the simulations were started from various configurations taken from the trajectory of Sim. A. In total, 42 EID simulations were conducted, respectively, with a length of 1.05 ns. Note that the applied method does not presuppose a sequential or direct transfer from site 1 to site 3, hence the simulations are unbiased with regard to the detailed CT mechanism.

8.1   Individual CT Events

First, two individual examples of typical EID simulations are discussed, beginning with one in which CT occurs rapidly (Fig. 8.1). The hole is stabilized on site 1 during the first 100–200 fs, the energy of site 1 drops by more than one eV, establishing an energy gap between site 1 and the sites 2 and 3 (Fig. 8.1a)). The gap can be monitored as an effective reaction coordinate [327–329], indicating the reorganization of protein and solvent environment due to the relocation of the positive charge, an effect discussed in

detail in Sec. 7.3.

To enable CT from site 1 to site 2, the energy gap has to become small, $\epsilon_1 \approx \epsilon_2$, while the electronic couplings T_{12} have to assume sufficiently large values. This situation occurs for the first time after about 10 ps, when site 2 becomes gradually occupied until about 25 ps of simulation time (Fig. 8.1b)). Interestingly, fractions of the hole charge on site 2 are able to pass on directly to site 3. By accumulating more and more positive charge on site 2 and site 3, ϵ_2 decreases due to the effects described above. The partial hole charge on site 2 changes the polarization of the environment, hence leading to a partial reorganization which in turn lowers the energy of site 2. As a consequence, the energy of site 1 increases and a crossing of ϵ_1 and ϵ_2 occurs at about 27 ps, where ϵ_2 becomes the lowest energy state and the hole becomes almost completely localized on site 2 with minor fluctuations between the sites 2 and 3. Having nearly the complete charge on site 2, a substantial relaxation of site 2 should be expected, as seen for site 1 after charge injection. On the contrary, this is not observed. Instead, a significant amount of charge is transferred quickly to site 3 as well. Therefore, the gap between the sites 2 and 3 becomes never as large as that between 1 and 2 in the beginning of the simulation. This explains the rapid transfer 2→3 and the corresponding residence time on site 2 of only roughly 5 ps, which is significantly shorter than on site 1.

The period from 25–35 ps is particularly interesting, showing very similar site energies (Fig. 8.1c)). There is no constant delocalization of charge between the sites but rather an oscillation with short intervals of delocalization between sites 2 and 3, which occur for a fraction of a pico-second, similarly to the situation in DNA [173]. About 20% of the hole charge remains on the site 1 and 2 until a simulation time of 60 ps, oscillating back and forth until the whole amount of charge relaxes to site 3. Note the significant additional drop in the energy ϵ_3 after 60 ps, establishing a gap of about 2 eV to the other sites, thus preventing any back-transfer. As analyzed in Sec. 8.3, this last step is coupled to a further relaxation of the protein.

Moreover, the time dependence of site energies and ESP is nearly identical, as revealed in Fig. 8.1b). They only differ by fast and small amplitude oscillations resulting from internal vibrations of the CT sites themselves, which are negligible compared to the oscillations introduced by the ESP due to the environment. The reader is referred to Ref. [164] for an in-depth analysis of this issue. Accordingly, a drop in the site energy is always correlated closely to a drop in the ESP indicating a structural reorganization in the environment, as discussed previously in Ref. [173].

On the other hand, a 'slow' CT trajectory exhibits rather different characteristics (Fig. 8.2). Here, the hole moves from site 1 to site 2 after about 0.25 ns, yet does not pass on further to site 3. Instead, a shift in the site energies due to relaxation of the environment occurs. The site energies become clearly separated by a gap of about 1 eV[1] until a second CT event occurs at about 0.55 ns where the hole moves eventually to site

---

[1]Note this value is in agreement with the corresponding site energy difference obtained from the statistical analysis in Sec. 7.3.
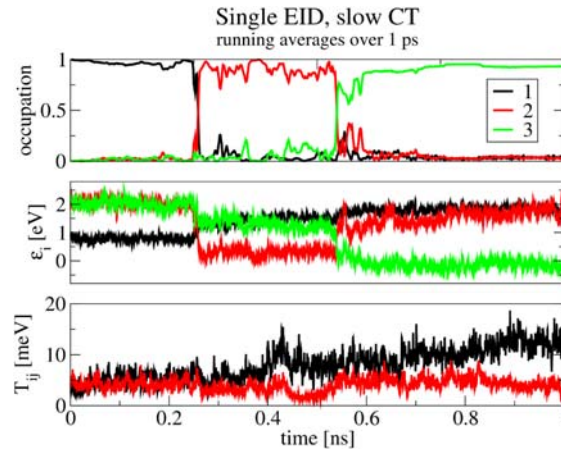
Figure 8.2: Non-adiabatic Ehrenfest dynamics (EID) in PL, example of a slow CT: Hole probability, site energies and electronic couplings between sites 1, 2 and 3 during 1 ns of a 'slow' CT trajectory. To get a clearer picture, running averages over 1 ps are used to smoothen the graphs.

3. In this case, the hole has a similar residence time on site 2 as it had on site 1.

In conclusion, different CT mechanisms seem to be possible, depending on whether the hole stays at site 2 or passes directly on to site 3. This depends on the fact that at the same time the gap is vanishing and the couplings are non-zero. A similar situation was observed in case of hole transfer in DNA [322]. When the hole is hindered to move on to site 3, the environment relaxes and stabilizes the hole at site 2. This situation is assumed implicitly in Marcus' theory, that is each charge state is fully equilibrated to estimate the reorganization energy.

However, the fact that ϵ_2 is not lowered in simulations with very short residence times at site 2, points at an incomplete reorganization. As a consequence, Marcus theory should not be applied in the way it has been in the previous chapter. Therefore, the localization/delocalization of charge plays an important role for the charge transfer mechanism.

8.2   Statistics

The two individual EID simulations revealed rather contradicting CT characteristics implicating that the analysis of CT should not rely on single trajectories, rather various initial configurations have to be realized in order to explore a wide range of possible CT pathways. Eventually, statistical quantities can be extracted from the ensemble of CT trajectories.

Fig. 8.3a) shows the time evolutions of occupation numbers for the three sites for all 42 EID simulations that have been conducted. In most of the simulations, the first CT step 1→2 occurs within 120 ps, although in a few cases it takes up to 600 ps. Even though the majority of simulations reveal the final CT 2→3 to be completed within 300 ps, several effective back transfers 3→2 are observed. A small fraction of charge
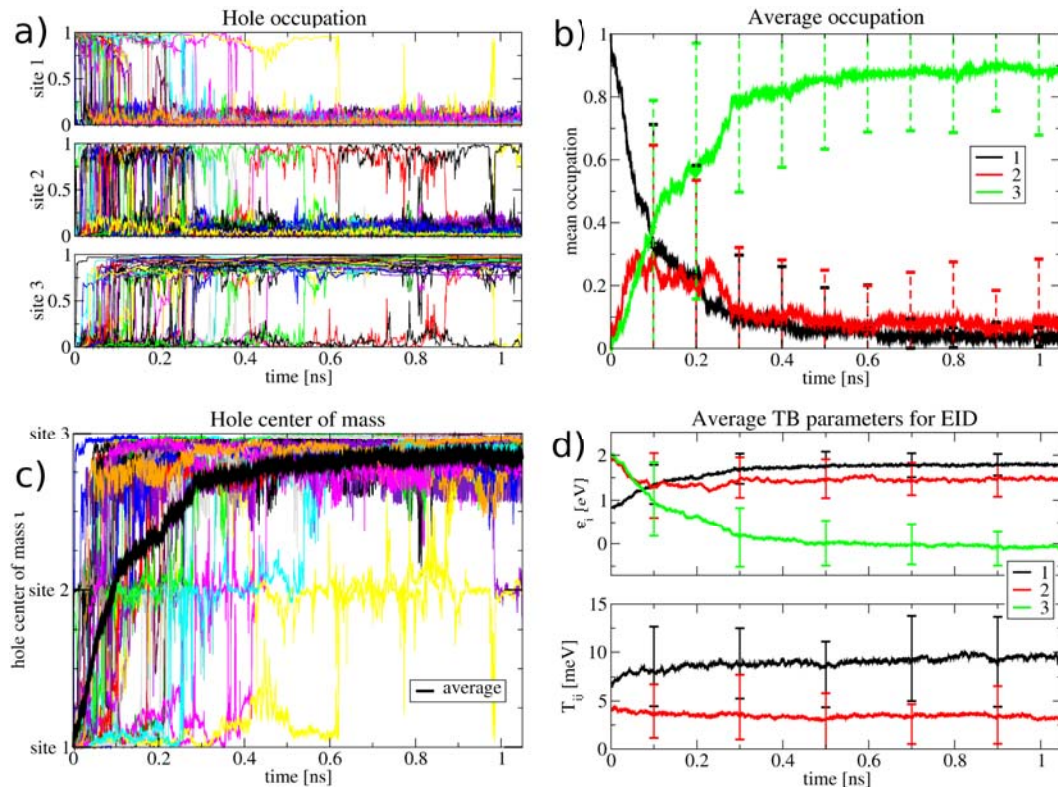
Figure 8.3: EID statistics over 42 trajectories, each 1.05 ns long: a) Site occupations for the three Trp side-chains involved in the CT for each single EID. Note, the different trajectories are realized by starting from different conformations obtained from simulation A. b) Averaged occupations with standard deviations over all 42 single EID trajectory. c) Movement of the hole center of mass $\iota = \sum_{i=1}^{3} i \Delta Q_i$ for each single EID as well as the average over all EIDs. And d), mean TB parameters, i.e. averages with standard deviations over all 42 EIDs (shown are running averages over 5 ps).

remains on sites 1 and 2 until a final relaxation step of the protein, as discussed above. Nevertheless, the various individual simulations exhibit very diverse CT pathways.

Furthermore, the corresponding averaged occupation numbers are presented in Fig. 8.3b), indicating that the site 2 is populated only transiently on average, up to an occupation of about 0.2, within the first 300 ps. Interestingly, the population of site 3 rises simultaneously with that of the site 2 within the first 100 ps. This shows that during these 100 ps more than 60% of the charge is transferred to sites 2 and 3 from which the major part is passed on directly to the final site 3.

Moreover, the system dynamics allow the site energies to be sufficiently close for short periods, so that charge oscillations can occur, although the charge is not delocalized substantially between the sites. The residence times on the respective site during the CT process are not sufficient on average for the system to equilibrate in the corresponding charge state. In particular, site 2 seems to be occupied transiently (on average) so that its site energy is not lowered significantly due to environmental relaxation. This enables the rapid follow-up transfer to site 3. Note the large standard deviation bars for the average occupations given in Fig. 8.3b), indicating again the diversity of individual CT

reactions, in particular in the interval from 100–300 ps.

The combined information of the individual site occupations may be given in a compressed form by the hole center of mass $\iota = \sum_{i=1}^{3} i \Delta Q_i$. This quantity is adapted from Ref. [173], in which it was applied to the analysis of localization/delocalization of a hole charge in DNA. The corresponding time course for the 42 simulations as well as the average $\langle \iota \rangle$ is presented in Fig. 8.3c). The hole charge is localized typically on the respective sites before and after the CT steps $1 \rightarrow 2$ and $2 \rightarrow 3$. However, the hole charge is delocalized during the transfer process itself, which usually occurs within 1–2 ps. Therefore, the hole seems to be localized on one site most of the time but delocalization, which is a result of the oscillation between two sites, occurs during the CT events. Notwithstanding, the average hole center of mass did not reach its final value of 3 even after 1 ns which implicates the slow progress of CT in some of the trajectories.

Moreover, the time series of the corresponding CT parameters, i.e. the average site energies and the couplings, are given in Fig. 8.3d). It turns out that the site energy for the central site is never the lowest one on average, though there is a large overlap for the distribution of site energies until about 200 ps. The average site energy gap between the central sites 2 and 3 amounts to 1.5 eV already after about 450 ps. Furthermore, the average couplings for both CT steps are rather comparable at the beginning of the simulation, i.e. they match the average values for Sim. A (see Tab. D.2 in the appendix). Nevertheless, T_{12} increases from 6 to 10 meV, while T_{23} decreases as the simulation proceeds. Both the large site energy gap between sites 2 and 3 as well as the rather small electronic coupling T_{23} lowers effectively the chances for successful back-transfers $2 \leftarrow 3$.

8.3 Relaxation

The analysis so far implicates that some trajectories allow for a fast transfer from site 1 to site 3, with site 2 being occupied only temporarily. For a more in-depth insight, the residence times on site 2 are computed (Fig. 8.4a). For the purpose of this analysis, the hole was defined as residing on site 2 if the occupation exceeded a value of 0.8.

A few very fast transfers can be seen, but the bulk of CT events exhibits a residence time on site 2 in the order of 10–100 ps. This has to be compared with the time scale of protein relaxation. The dynamics of the ESP at the sites 1–3 is monitored in the following by employing classical MD simulations.

The time course of the ESP after moving the hole charge from site 1 to site 2[2] is shown in Fig. 8.4c). It takes about 200–300 fs for the ESP to converge closely to their new values. Therefore, the site energies of the sites 2 and 3 are sufficiently close to allow for a further transfer from site 2 to site 3 only during the first hundreds of fs after the transfer from $1 \rightarrow 2$.

On the contrary, the complete relaxation for the second step $2 \rightarrow 3$ seems to be much

---

[2]In this case, the charge topology from Sim. D is used for equilibrated conformations from Sim. C.
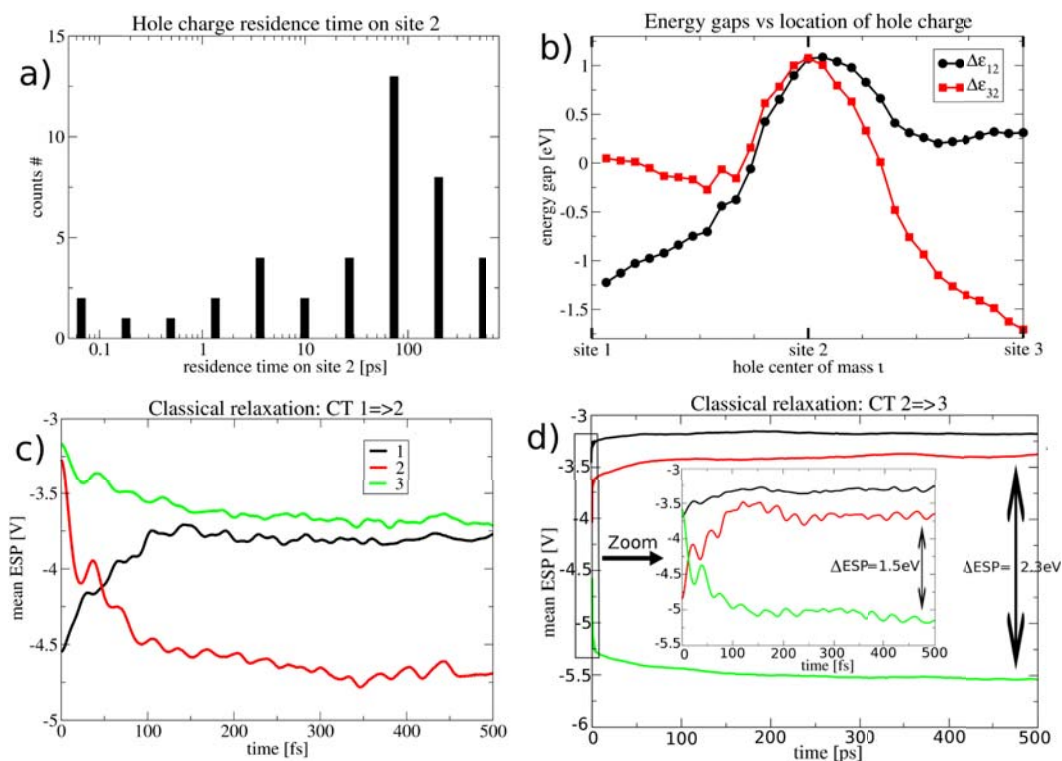
Figure 8.4: Analysis of CT in PL: a) Residence times at the central site 2, b) site energy gaps of the central site to sites 1 and 3 with respect to the hole center of mass ι, and c),d) ESP relaxations obtained as averages from, respectively, 50 classical MD simulations for the separate CT steps $1 \rightarrow 2$ and $2 \rightarrow 3$. For the latter, the relaxation is followed additionally on a longer ps time scale.

slower, as the ESP converges not before several hundreds of ps. This may explain why back-transfer events for the last step are readily possible, as found e.g. for the simulation with the longest residence time on site 2 (Fig. 8.5). The classical relaxation of ESP is analyzed further by decomposing the ESP into protein and solvent components for both CT steps $1 \rightarrow 2$ and $2 \rightarrow 3$. As as result, both solvent and protein response for the first CT $1 \rightarrow 2$ are rather fast compared to those for the CT $2 \rightarrow 3$. For the latter, the ESP relaxation stemming predominantly from the protein takes place on a very long ns time scale. In general, the solvent response occurs on a much shorter time scale (several 100 fs) than the protein response (up to ns may be even more). The data is given in Fig. D.2 in the appendix.

The explanation for the fast transfer can be given using the average site energy gaps evaluated in dependence of the hole center of mass ι, as shown in Fig. 8.4b). Accordingly, if the site 2 is occupied only partially, the site energy gap between ϵ_2 and ϵ_3 decreases, thus allowing for a faster transfer to the next site. This indicates that the transfer cannot be modeled as consisting of two separate hopping steps, in which the protein environment is allowed to relax fully after each transfer, as assumed above when computing the CT parameters for the Marcus model.

Furthermore, two additional sets of simulations $1 \rightleftharpoons 2 \rightleftharpoons 3$ emphasize this point further:
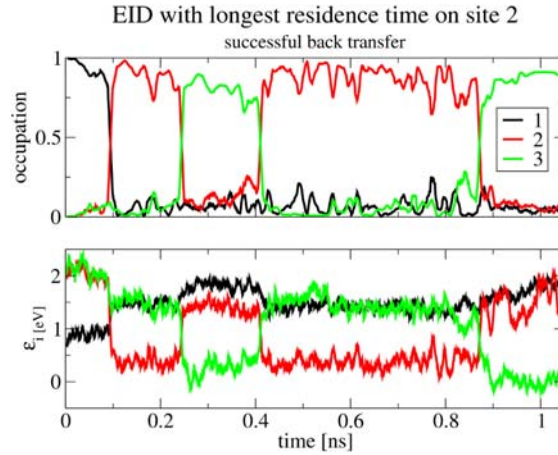
Figure 8.5: Site occupation and energy levels for the simulation with the longest charge res- idence on site 2 (running averages over 5 ps). Moreover, one successful back-transfer $2 \leftarrow 3$ is observed, after which the hole charge remains on the central site for almost half a nano-second. Finally at 0.9 ns, the hole charge hops back again to site 3, followed by a stronger relaxation (compared to the first CT $2 \rightarrow 3$ at 0.25 ns), thus effectively widening the site energy gap $\Delta\epsilon_{23}$.

In the first set, the simulations were started with the hole charge placed on site 1 and using starting conformations from Sim. C, where the environment has been equilibrated with a hole charge at site 1. The second set of simulations was started with the charge on site 2 and conformations from Sim. D, where the environment has been equilibrated with a hole charge at site 2. Thus, both simulations begin with the charges in full equilibrium to their surroundings (Fig. 8.6).

Moreover, the data is compared with results of the unbiased simulations from Fig. 8.3b). Interestingly, the respective transfer rates are found to be even slightly faster for the first set of simulations starting with the hole on site 1 (left panel), while significantly slower rates are observed for the second set starting with the hole on site 2 (right panel). Appar-
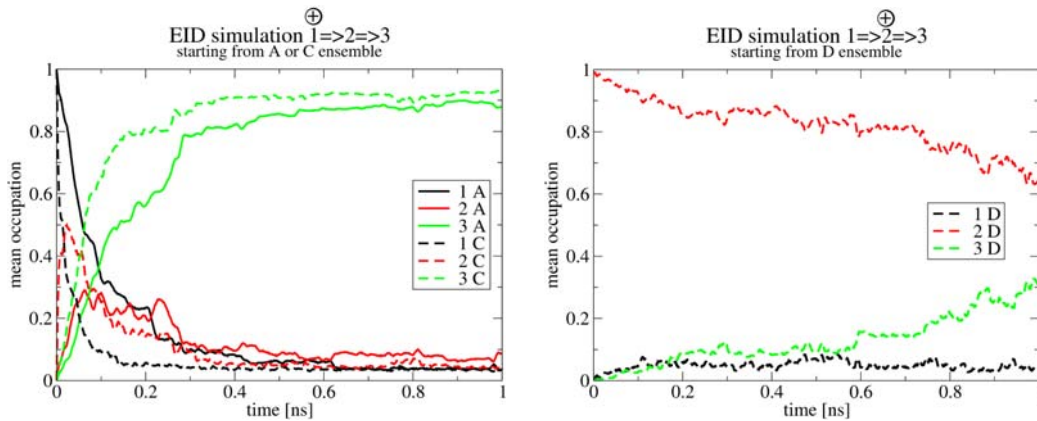
Figure 8.6: EID simulations $1 \rightleftharpoons 2 \rightleftharpoons 3$ starting from equilibrated systems. Left: CT initiated with the hole charge on site 1 and starting conformations from Sim. C and A, respectively (the latter for comparison, data taken from Fig. 8.3b). Right: CT initiated with the hole charge on site 2 and starting conformations taken from Sim. D.

ently, a charge on site 1 is capable of rapidly transferring along the Trp chain regardless of the starting conformation, while full equilibration of a charge on site 2 suppresses fast CT.

In conclusion, the results implicate that the CT times of the order of 10–30 ps, as found experimentally, cannot involve full relaxation of the environment at intermediate steps. These findings are corroborated by additional simulations with a restricted QM system presented in Fig. D.5. Moreover, a complete section with supplementary informations and analysis is given in Sec. D.2 in the appendix.

8.4   Constant Couplings

The distance between the CT active sites in PL is substantially larger compared to that of the hole transfer in DNA. The distance between stacked nucleobases in DNA is about 3.4 Å, whereas the Trp residues in PL are separated by at least 7 Å (see the distance distributions in Fig. 7.1(a)). As a result, the electronic couplings, which vary exponentially with the distance, are about one order of magnitude smaller in the protein on average. Thus, values of about 2–10 meV are computed in PL, compared to T_{ij} of ca. 50 meV for two stacked guanines in DNA [147]. Previous studies on hole transfer in DNA emphasized that CT is affected only slightly by fluctuations of T_{ij} and therefore couplings may be replaced by their averages [167]. The relation between CT rate and electronic coupling may be similar in PL. Even though, the dynamics of the CT process is assumed to be dominated by the down-hill energetics caused by nuclear coordinate fluctuations, still one can evaluate the minimum coupling T_{ij} necessary to facilitate the transfer.

Various constant couplings for single EID simulations are tested, all starting from the same snapshot of Sim. A. As a result, a constant coupling $T_{12} = T_{23}$ of only 1 meV does not result in successful CT over 500 ps. Couplings of 10 meV enable CT 1→2→3 within tens of ps. Further increasing T_{ij} to 100 meV accelerates the transfer by about one order of magnitude. However, if MD average values are used, i.e. $T_{12} = 9.78$ meV and $T_{23} = 2.62$ meV (from Tab. D.2 in the appendix), the overall CT takes more than 300 ps, as depicted in Fig. 8.7a).

The small constant coupling T_{23} leads to a considerable increase of the residence time on site 2. Interestingly, about 20% of the hole charge, which has not yet been transferred to site 3 until 300 ps, can hop back and forth between the sites 1 and 2 facilitated by the rather large coupling T_{12}. The MD average electronic coupling T_{23} appears to be barely sufficient for CT 2→3, and single non-equilibrium fluctuations seem to play a crucial role here. By contrast, the CT 1→2 seems to be hardly affected by single fluctuations, thus the coupling T_{12} is sufficiently large to achieve the transfer most of the time, as indicated in the corresponding distributions in Fig. 7.7. Therefore, the time series may be replaced by its average in this case.

For the purpose of statistics, 22 EID simulations with constant couplings taken as
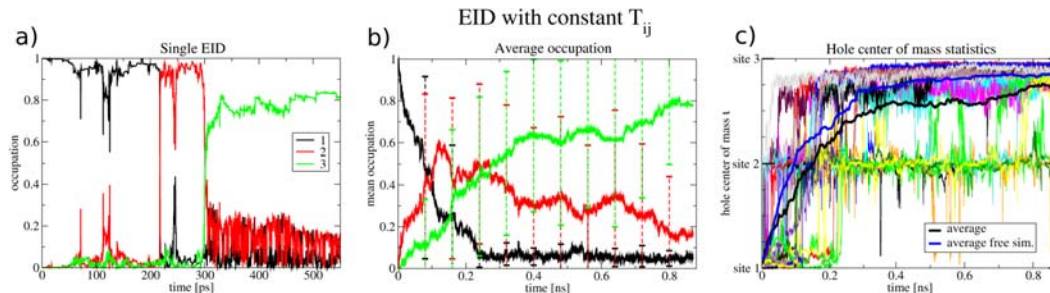
Figure 8.7: EID with constant couplings taken as MD averages $T_{12} = 9.78$ meV and $T_{23} = 2.62$ meV from Tab. D.2 in the appendix). a) Occupations during a single simulation, b) averaged occupation with standard deviation over 22 EID simulations, and c) hole center of mass including the average $\langle \iota \rangle$ over all 22 simulations, also compared to that of the free simulations, i.e. from Fig. 8.3c).

MD averages have been carried out (Fig. 8.7b) and c)). Taking MD average values as constant couplings slows down the overall CT rate $1{\rightarrow}2{\rightarrow}3$ compared to the "free T_{ij}" simulation, which is indicated by the average hole center of mass in Fig. 8.7c) as well. Within 850 ps simulation time, only 80% of the hole charge has reached the final site 3, while the same amount of charge was transfered within 300 ps in the free simulation.

However, the effect of constant couplings is different for the two CT steps. The rate $1{\rightarrow}2$ turns out to be rather similar to that of the free simulation, i.e. taking the MD average for T_{12} results in about the same rate and the first CT step might be called "coherent" based on the definition given in Ref. [159]. On the other hand, the use of constant couplings makes the CT $2{\rightarrow}3$ considerably slower, since major fluctuations of T_{23} are necessary to speed up the transfer in the free simulation. As a result, site 2 becomes significantly more populated, and the major fraction of the hole is located there for almost 200 ps. Moreover, the occupations on sites 2 and 3 undergo larger fluctuations. A broad range of different rates for $2{\rightarrow}3$ is observed, whereas $1{\rightarrow}2$ is completed within 200 ps (Fig. 8.7b). Generally, the findings are in a good agreement with the corresponding coherence parameters obtained for T_{12} and T_{23} from Tab. 7.4.

Summarily, the impact of T_{ij} fluctuations in this particular protein is not as strong as in other biomolecular systems. For instance, in DNA CT, the maximum of T_{ij} distributions was obtained around 0 meV, although the mean values were rather high with about 50 meV for poly(G), so unlike in PL, hole transfer events in DNA are driven predominantly by non-equilibrium conformations [167].

8.5   Protein vs Solvent

In Sec. 7.3, the ESP acting on the respective CT sites has already been decomposed into its two components of the protein and solvent environment. Here, the impact of both components is analyzed during the direct dynamics of the hole charge separately. To begin with, a CT simulation has been carried out in which neither protein nor solvent environment has been taken into account. This can be realized by using Eq. 2.125
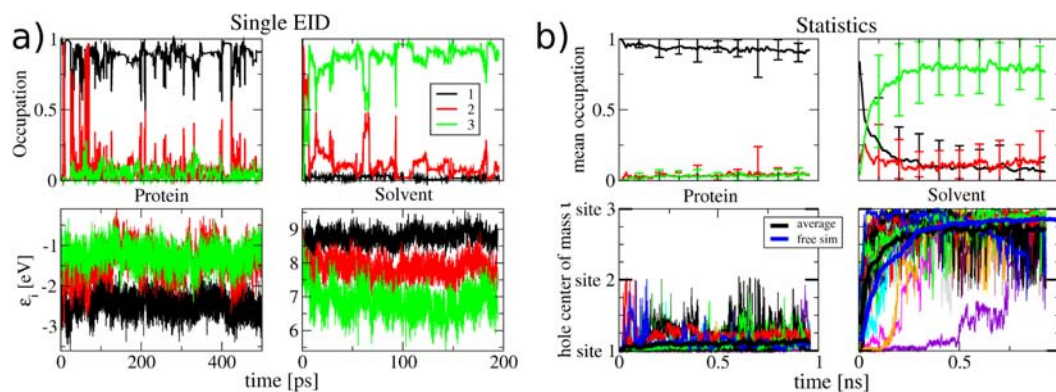
Figure 8.8: EID in either solely protein or solvent environment: a) Occupations and ϵ_i during a single EID. b) Statistics over 22 simulations: Average occupations and standard deviations as well as hole center of mass ι. The latter quantity is compared also to that of the free simulation using the full QM/MM interaction (Fig. 8.3c).

instead of Eq. 2.126 for the computation of CT parameters, thus neglecting the QM/MM interaction. As a result, the hole charge is likely to hop back and forth between the three sites for there is almost no energy barrier between the corresponding CT states. The data is depicted in Fig. D.7 in the appendix.

In the following, the MM charges of only well-defined parts of the environment, either the protein or the solvent, are included in the QM/MM interaction of the CT sites via Eq. 2.125. To begin with, the hole occupation and site energy time series are depicted for two single simulations (Fig. 8.8a)), in which either solely the protein or the solvent environment is used to calculate the ESP at the QM sites. No significant CT is observed within 500 ps if only the protein environment is considered. Clearly, site 1 inside the protein is stabilized most strongly by the protein surrounding, thus ϵ_1 is shifted down by 1–1.5 eV compared to ϵ_2 and ϵ_3. Therefore, the barrier $\Delta\epsilon_{12}$ can be overcome only very rarely via large site energy fluctuations, leading to transient hole charge accumulation. Moreover, the site energies for the sites 2 and 3 exhibit larger fluctuations than obtained for site 1 (see also Tab. 7.3), suggesting that the protein environment around sites 2 and 3 is more flexible.

By contrast, the pure solvent environment enables the CT by lowering substantially the energy of site 3, which is exposed most strongly to the solvent, hence stabilizing the hole charge on this site. The overall CT takes only 7 ps in this particular simulation. There is no initial stabilization of the hole on site 1, since the protein influence is missing and the solvent is too far away to affect it markedly. Temporary back transfers 3→2 appear to be possible at first glance, for the energy gap is only 1 eV. However, the ESP stemming from the solvent is correlated strongly for all three sites, i.e. if site 2 is lowered site 3 is affected in rather the same way and the chances of $\epsilon_2 \approx \epsilon_3$ are small. On the contrary, the ESP stemming from the protein environment appears to be more heterogeneous, as discussed in Sec. 7.3.

For the aim of statistics, again, 22 CT simulations were carried out for both cases,

respectively (Fig. 8.8b)). The average occupation in solely the protein surrounding indicates clearly that there is no effective CT, as the hole charge remains on site 1 where it is stabilized most strongly. By contrast, the CT process is enabled if the solvent environment is active. In fact, the average rate is even slightly faster within the first 200 ps than that for the simulation using the full MM environment.

Furthermore, a comparison between the free simulations (full MM environment) and the solvent-only simulations exhibits a stronger charge delocalization in the latter. For instance, in one simulation, ι takes on a value of 1.5 over about 200 ps, which is a strong indicator of delocalization. Moreover, the CT process is completed to only 80% on average, which is 10% less than for the free simulation.

Notwithstanding, the decomposition of the environment into solvent and protein parts is purely artificial, especially considering only the solvent environment (effectively placing Trp residues into a cavity of the protein's exclusion volume, with solvent at least 5 Å away). However, this proceeding is yet believed to be useful for the purpose of this analysis, as it shows that it is predominantly the solvent which drives the CT process, while the protein surrounding plays a minor role.

Nevertheless, the interplay between both appears to be substantial for the localization of the hole charge during as well as at the end of the overall CT $1\rightarrow2\rightarrow3$. In conclusion, the protein aides the movement of charges, but it does not require any functional role (i.e. conformational changes) for CT to take place. On the other hand, the reorganization and fluctuation of the solvent alone would be sufficient to drive CT to completion.

8.6   Adiabatic vs Non-Adiabatic Dynamics

In this section, the results obtained from the non-adiabatic EID simulations in Sec. 8.2 are compared to those from adiabatic Born-Oppenheimer (BO) simulations. The movement of electrons and ions is not coupled directly in BO simulations, rather the electronic energy of the hole charge is minimized in each step of the classical MM trajectory. Therefore, the hole charge remains always in the electronic ground state. As a consequence, there is no direct quantum mechanical propagation of the hole charge in time, as compared to the EID simulations. The BO approximation holds as long as the energy splitting between the adiabatic states, i.e. the electronic coupling between charge donor and acceptor, is large.

On the other hand, the mean field approach of the EID leads to a system state which is a mixture of the adiabatic eigenstates. There may be cases in which EID fails dramatically as it overestimates delocalization, e.g. in organic crystals. In case of CT in PL, the site with the lowest energy is likely to be populated in BO, as the small electronic coupling represents only a minor alteration when diagonalizing the electronic Hamiltonian. Therefore, direct transfers $1\rightarrow3$ are conceivable in BO, which is rather doubtful considering the large distance between sites 1 and 3 of 11.5 Å on average.

As depicted in Fig. 8.9, the central site becomes hardly populated on average, for
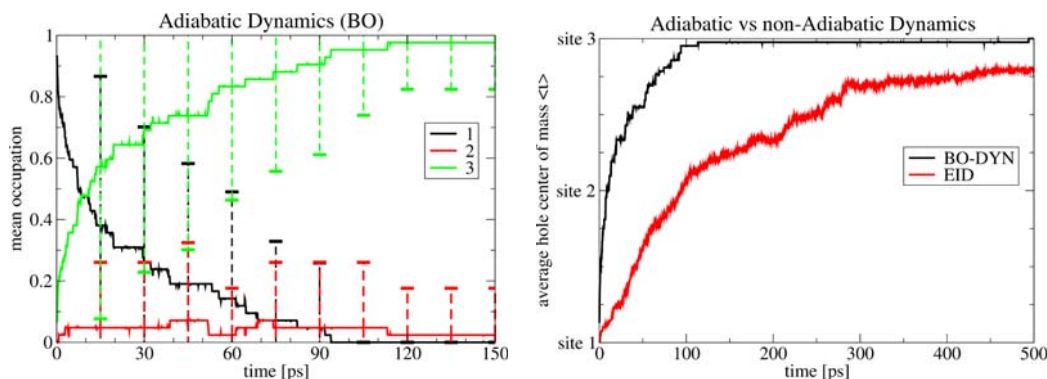
Figure 8.9: BO vs EID simulations: Mean occupation with standard deviation for an ensemble of 42 adiabatic (BO) simulations (left). Hole center of mass, comparison with non-adiabatic (EID) simulations from Fig. 8.3c) (right).

the lowest adiabatic state is that of a hole on site 3 in most cases. Apparently, the direct CT 1→3 is observed for the majority of simulations, strongly in contradiction to the EID results as well as experimental observations [104]. Accordingly, the overall transfer time is significantly shorter by nearly one order of magnitude compared to the non-adiabatic EID simulations. This is due to the vigorous lowering of ϵ_3, hence the strong exothermic character of the reaction. Moreover, the transfer appears to be almost complete, i.e. the hole charge is localized to 100% on the final site after about 100 ps. In fact, no delocalization is observed in the BO simulations, that is the hole charge is localized on one single site all the time. A comparison of BO and EID simulations was carried out previously in the study of hole transfer in DNA. There, the situation was reversed, that is BO exhibited considerably smaller rates than the non-adiabatic EID simulations [173]. This can be attributed to the missing driving force in that case.

In conclusion, the adiabatic (BO) approach is believed not to be suitable for this particular application on the Trp chain in PL. The average couplings of both CT steps are smaller than $k_B T$, thus the CT is clearly outside the adiabatic regime. More importantly, the CT pathway described with BO is unphysical, although the obtained transfer times are closer to the experimental values of about 30 ps [104]. However, this turns out to be the result of error cancellation, as the lack of polarization described for the Marcus rates in chapter 7 holds for the EID and BO simulations as well. Once this electronic polarization is included, the expected transfer times would decrease. Thus, BO would overestimates the experimental transfer rates significantly. The inclusion of polarization may be realized e.g. by scaling the MM charges during the direct dynamics, which might be a subject for future applications.

8.7   Kinetic Models

The data obtained in Sec. 8.2 is used to derive chemical rate parameters from analytical and numerical expressions for various kinetic models of the CT reaction in PL. In the following, the hole occupation of the sites 1, 2 and 3 is denoted, respectively, with $[A]$, $[B]$ and $[C]$. Furthermore, the initial hole occupations are $[A]_0 = 1, [B]_0 = 0, [C]_0 = 0$ at $t = 0$ according to the EID simulation setup.

8.7.1   Analytical Model I

$$1 \xrightarrow{k_1} 2 \xrightarrow{k_2} 3$$
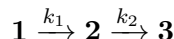
Exclusively consecutive forward reactions 1→2 and 2→3 are considered. The differential equations for the time evolution of occupations read:

$$\frac{d\,[A]}{dt} = -k_1\,[A]$$
$$\frac{d\,[B]}{dt} = k_1\,[A] - k_2\,[B]$$
$$\frac{d\,[C]}{dt} = k_2\,[B]$$

which lead to the following integrated rate equations

$$
\begin{aligned}
[A] &= e^{-k_1 t} \\
[B] &= \frac{k_1}{k_2 - k_1}\left(e^{-k_1 t} - e^{-k_2 t}\right) \\
[C] &= 1 + \frac{k_1 e^{-k_2 t} - k_2 e^{-k_1 t}}{k_2 - k_1}.
\end{aligned}
\tag{8.1}
$$

The functions given in Eq. 8.1 are fitted to the mean occupations obtained for the non-adiabatic simulations (Fig. 8.3b)). The data is illustrated in the left panel of Fig. 8.10. This simple procedure, without allowing for any back transfers, gives rate constants of k_1 and k_2 of the order of ns^{-1} (see Tab. 8.1). Notwithstanding, the fit should be interpreted only qualitatively, as it appears to be rather crude to assume both transfers 1→2 and 2→3 to be totally irreversible, i.e. after finite time $[A] = [B] \approx 0$ and $[C] \approx 1$.

First of all, the rate obtained for the first transfer 1→2 ($8.8\,\mathrm{ns}^{-1}$) agrees well with the corresponding Marcus' rate ($5.0\,\mathrm{ns}^{-1}$). On the other hand, the kinetic model suggests the second step 2→3 to be slightly faster than the first one 1→2. This reveals quantitatively the large disagreement to the Marcus' rate obtained for 2→3 being almost five orders of magnitude slower than for 1→2. This emphasizes the previous conclusion on the applicability of Marcus' theory which is reasonable for the first CT step, in which the relaxation of the environment is fast, while certainly not for the second step due to the presence of long-term protein relaxations, as discussed in Sec. 8.3.

Notwithstanding, the kinetic model exhibits rather similar rates for both CT steps in agreement with the experimental rates given in Tab. 8.1, although the absolute values
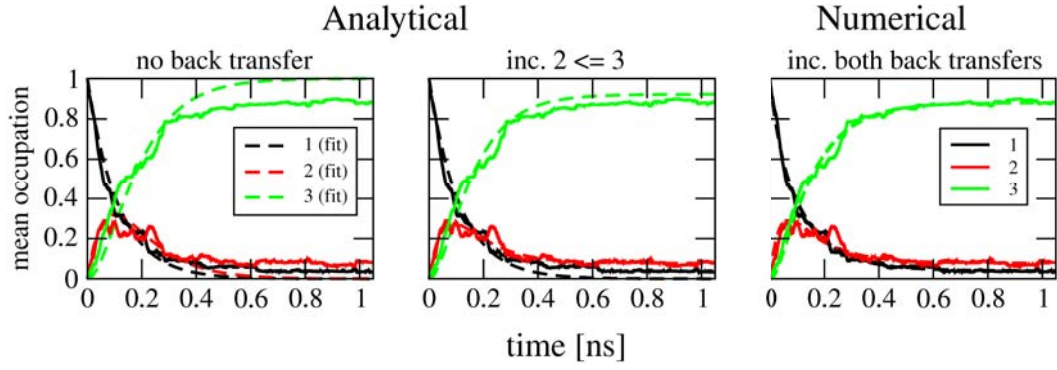
Figure 8.10: Analytical and numerical fits for mean occupations of the EID simulations from Fig. 8.3b).

for the latter are about ten and three times larger for the first and second CT step, respectively.

8.7.2 Analytical Model II

$$1 \xrightarrow{k_1} 2 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} 3$$
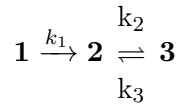
Here, the previous model is extended to include the back transfer 2←3, as observed in several EID simulations, see e.g. Fig. 8.5. This still allows for an analytical solution. In the following, **1** and **2** are considered to be in equilibrium $K = \frac{k_2}{k_3} = \frac{[C]_{equil}}{[B]_{equil}}$. Further, it is assumed that $k_1 \neq k_2$, thus the time dependent occupation of C can be written as $[C] = [A]_0 - [A] - [B]$. Then the differential equations read

$$\frac{d[A]}{dt} = -k_1[A]$$
$$\frac{d[B]}{dt} = k_1[A] - k_2[B] + k_3[C]$$
$$\frac{d[C]}{dt} = k_2[B] - k_3[C]$$

leading to integrated rate expressions

$$[A] = e^{-k_1 t}$$
$$[B] = \frac{k_1 - k_3}{k_2 + k_3 - k_1}\left(e^{-k_1 t} - e^{-(k_2+k_3)t}\right) + \frac{k_3}{k_2 + k_3}\left(1 - e^{-(k_2+k_3)t}\right) \tag{8.2}$$
$$[C] = 1 - e^{-k_1 t} - \frac{k_1 - k_3}{k_2 + k_3 - k_1}\left(e^{-k_1 t} - e^{-(k_2+k_3)t}\right) - \frac{k_3}{k_2 + k_3}\left(1 - e^{-(k_2+k_3)t}\right).$$

The agreement between fit and data for the sites 2 and 3 is improved strongly, that is the correct asymptotic behavior is obtained, whereas for site 1 the situation is still unchanged due to the neglect of back transfer 1←2 (central panel of Fig. 8.10).

However, the inclusion of the back transfer 2←3 increases slightly k_2 to 15.9 ns^{-1}, yet it also exhibits a considerable back rate constant k_3 of 1.3 ns^{-1}, as shown in Tab. 8.1.

Table 8.1: Chemical transfer rates k_{1-4} for the three kinetic models. Comparison with the Marcus' rates (with unscaled λ) obtained from Tab. 7.8, and with experimental data[3] from Ref. [104]. All values in ns^{-1}.

rate	Analyt. Model I	Analyt. Model II	Num.Model	"Marcus"	Exp. [104]
k_1	8.8	8.8	12.3	5.0	110
k_2	11.6	15.9	15.4	2×10^{-5}	33
k_3	—	1.3	1.4	—	—
k_4	—	—	6.3	—	—

Nevertheless, the ratio $K = \frac{k_2}{k_3}$ is rather large with about 12 indicating the major part of the hole charge is located on the final site 3 most of the time at steady state. The ratio K may be regarded as an equilibrium constant from which an energy difference between both states can be computed via $\Delta G = -RT\ln(K)$, with T being the temperature (300 K) and $R = 8.314$ J/(K·mol) the universal gas constant. This results in an energy difference of ΔG = -6.18 kJ·mol^{-1} (-0.064 eV) which is quite comparable with the corresponding ΔG values obtained in Tab. 7.7.

8.7.3 Numerical Model

$$1 \underset{k_4}{\overset{k_1}{\rightleftharpoons}} 2 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} 3$$
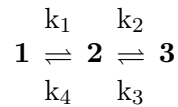
In this model, the back transfer 1←2 is included as well, leading to equations without analytical solutions. Instead, the equations were solved numerically by varying randomly the four rate constants until optimal agreement to the mean occupations is found. A set of 100 optimizations of 10,000 steps each converged to the result shown here, exhibiting excellent agreement to the simulation data as depicted in the right panel of Fig. 8.10.

Interestingly, the rate constants k_2 and k_3 are very similar to those from the analytical model II. However, k_1 is increased in the numerical model, hence rather comparable to k_2, though a considerable rate constant k_4 of 6.3 ns^{-1} is obtained for the back transfer 1←2, which is half as large as for the forward direction. Nevertheless, the back transfer 1←2 does not play a major role in the overall CT process, since the hole occupation of site 2 will always be small, thus the result is not very sensitive to k_4. The reader should note that the obtained rate constants, may not be interpreted quantitatively, since the standard error of such numerical fits can be quite substantial.

Nevertheless, the numerical model exhibits a rate constant k_1 of the first CT step that is within one order of magnitude compared to the experimental value, while the rate constant k_2 for the second step is even closer. However, the electronic polarizability missing in the EID simulations as well, would enhance the transfer rates, as pointed out

---

[3] The rates were determined from the upper limit of experimental transfer times 9 ps and 30 ps for the first and and second CT step, respectively.

in the previous section. The inclusion of electronic polarization for the Marcus' rates in Sec. 7.4.4, by means of empirically scaled λ values, led to rather predictable results. On the other hand, it is not clear how strong the EID simulations would be affected, and whether both CT steps would be influenced in the same way, as the implementation of electronic polarization is different.

In general, the interpretation of the analytical and numerical fits, in particular regarding the back-transfer rates, is partially misleading, as they are derived from EID occupation time series. As stated in the previous section, the EID mean field approach suffers from the trend of over-delocalization. This might be a reason that most of the individual EID simulations do not converge completely, that is 100 % of the hole charge is located finally at site 3, rather the converged values range between 0.8–0.95 (see Fig. 8.3a)). Subsequently, this missing minor portion of the hole is then misinterpreted as back-transfer of the hole charge in the kinetic models. Nevertheless, a few back-transfers were observed for the second CT step, while no successful back-transfer was observed for the first one. This is corroborated by the large reaction free energy of -0.4 eV for the first CT step as well.

In conclusion, the CT in PL exhibits a complex kinetic which can be described approximately with the applied simple chemical models, yet care has to be taken when interpreting the corresponding back-transfer rate constants.

9

Conclusion of Part II

The hole transfer along an evolutionary conserved triple Trp chain as part of the photoactivation process in *E. coli* photolyase was studied in-depth using two different theoretical strategies: Marcus' theory of electron transfer in chapter 7 and non-adiabatic electron-ion dynamics (EID) in chapter 8. Moreover, structural and electronic properties accompanying the CT process were determined as well.

9.1 Structure and CT Parameter

Extended classical MD simulations revealed a stable protein structure for the various steps of the overall CT process indicating no global conformational transition within tens of ns. However, local conformational changes in the CT region of the protein were observed once the final site 3 (Trp306) carried the hole charge. The two corresponding sub-conformations exhibited only minor structural differences, yet leading to significantly different electronic couplings. Accordingly, the coupling in one of the sub-conformations is reduced strongly, indicating a possible mechanism to prevent back CT to site 2 (Trp359). Interestingly, the presence of two system configurations is suggested for a hole charge on site 2 as well, which is a result of the computation of free energies via FGTI.

Furthermore, the hole charge experiences strong stabilization due to the solvent environment. This effective stabilization increases considerably along the CT pathway from site 1 (Trp382) located inside the protein over site 2 to site 3 located on the protein surface close to the solvent. More importantly, a rearrangement of solvent water is observed once the sites 2 and 3 become charged, effectively enhancing the solvent stabilization, and indicating a small "water-polaron" formation which accompanies the CT.

On the other hand, the hole charge on site 1 is stabilized predominantly by the protein environment at the beginning of the CT process, in particular by the negative charge on the adjacent FAD co-factor. However, the protein stabilization is exceeded by that of the solvent as the CT proceeds. Eventually, the product of the CT reaction is a rather stable radical Trp306$^{+\bullet}$, allowing for further reactions such as proton transfer to the solvent, which is expected to be significantly slower compared to the CT [104]. As a

result, the whole CT reaction including its single steps is found to be strongly exothermic exhibiting an overall driving force ΔG of -0.62 eV. These findings were obtained with advanced free energy simulations and are comparable to results reported in Ref. [330] using electrostatic calculations.

The impact of static and dynamical disorder on biomolecular CT due to environmental or structural fluctuations was studied for various systems [130, 137, 159–163]. Similarly, this issue was investigated in great detail for electronic transport in DNA in part I of this work. However, the first CT step in PL may be regarded as rather "coherent", as the transfer is not influenced significantly by the fluctuation of electronic couplings. Moreover, the coupling for this first step appears to be sufficiently large to enable CT at any time. On the other hand, the majority of conformations reveal rather small couplings for the second CT step, thus fluctuations leading to conformations with increased couplings have a stronger influence in that case. This was confirmed by EID simulations using fixed electronic couplings in which the corresponding average values of the MD were taken. No changes in the CT dynamics were observed for the first CT step, while the hole occupation on the central site 2 was considerably increased, as the CT to site 3 was suppressed due to the absence of fluctuations. Nevertheless, only the kinetics was slightly altered, while the transfer times of the CT process were affected marginally, indicating a minor importance of changes in the electronic couplings in general.

On the contrary, the results pointed out that the CT dynamics in PL is determined critically by the energetics due to protein and solvent environment. This is confirmed by large values for the outer-sphere reorganization energy λ_s as well. In addition to the strong driving force of the CT reaction, these factors are responsible for the unidirectionality of the CT exhibiting vanishingly small rates for back CT.

9.2   Timescales of CT and Relaxation

To understand the CT dynamics in PL, it is vital to consider the relaxation times of protein and solvent environment due to the various charge states along the CT reaction. Most of the individual EID simulations exhibited residence times of the hole charge on site 2 of less than 100 ps. This residence time has to be compared with the time scale related to the relaxation of solvent and protein environment. A complete relaxation of the latter was shown to take place on a substantially longer time scale of at least several ns.

Furthermore, a complete relaxation would require a fully localized hole on site 2, as realized in the classical simulation D. However, this happens only very rarely, as the hole charge is transfered in several smaller portions in most of the EID simulations due to delocalization. Therefore, both effects, the delocalization of charge and the long relaxation time of predominantly the protein environment enable rapid CT from site 2 to site 3. Thus, the central site 2 is occupied by the hole charge only transiently on average.

More importantly, the two CT steps cannot be considered separately, as the majority of CT $2 \rightarrow 3$ proceeds out of a non-equilibrium conformational ensemble. Therefore, the Marcus' rates computed for the second step are dramatically underestimated, for the applied equilibrium theory assumes a fully localized hole charge on site 2 in a completely relaxed environment. On the contrary, the results revealed that the time for stabilization of a hole charge on site 2 by the environment is not sufficient in most cases, as compared to the residence time.

As a consequence, the energetic term $\lambda + \Delta G$ in the exponent of the Marcus' equation is overestimated largely within the classic regime of two separate CT reactions, thus leading to a dramatically slower rate for the second CT step. Eventually, delocalization, which would reduce λ even further, and the slow relaxation increase the rate $2 \rightarrow 3$ by orders of magnitude, as observed with the EID simulations.

Twenty years ago, the first experiments reported merely an upper limit of 1 μs for the time of CT between the FAD cofactor and site 3 [331]. However, recent more advanced experimental studies provided accurate information for the various sub-processes of the photoactivation in PL [98, 104, 111, 112]. Accordingly, the first CT step 1→2 proceeds within 9 ps, while the second CT step 2→3 is slightly slower with an upper limit of 30 ps. The corresponding rate constants were compared to those of the Marcus' model in Tab. 8.1, exhibiting an excellent agreement for the first CT step 1→2, provided an adequate scaling factor is used for the reorganization energy λ. This is because the computation of λ from classical MD trajectories misses electronic polarization. On the contrary, the rate constant for the second step is largely underestimated even with scaled λ values, as discussed above. Therefore, equilibrium Marcus' theory is able to describe the first CT step, while it fails to describe the fast CT 2→3 out of non-equilibrium conformational ensembles.

In this work, the Marcus' rates were computed using an empirical scaling of λ and including the internal reorganization energy λ_i. Both effects were not considered in a previous study [120], in which Marcus' parameter were computed from classical MD simulations as well, leading to considerably smaller rate constants of 0.11 ns^{-1} and 0.45 ns^{-1} for the first and second CT step, respectively. However, the inclusion of λ_i would even lower the observed rates. Moreover, qualitatively different ΔG and λ values for the separate CT steps were observed in Ref. [120], as compared to this work. Nevertheless, both studies come to the conclusion that the use of separate λ values is vital when describing CT in a heterogeneous environment, such as a CT from the inside to the exterior of a protein close to the solvent, as it is the case in PL. The energy required for the rearrangement of the system turned out to be considerably larger for the backward CT, thus the approximation of equal λ within conventional Marcus' theory appears to be inadequate. Therefore, the use of separate λ for forward and backward CT may be regarded as a step beyond classical Marcus' theory.

Notwithstanding, unlike the Marcus' results, the rate constants computed from the EID simulations are in very good agreement, within one order of magnitude, with the

experimental data for both CT steps, as shown in Tab. 8.1. However, the EID rate constants are slightly larger, in particular for the first CT step, indicating presumably the lack of electronic polarization. Therefore, the EID results on the CT in PL, and in bio-molecules in general, may be improved be extending the applied framework for the inclusion of electronic polarization. This may be realized, e.g. by scaling the MM charges during the direct dynamics. Nevertheless, the fine and complex interplay between the hole charge dynamics and the dynamics of the environment can be resolved in a unique way with the applied CG-EID framework. Thus, the individual CT steps are described realistically, and the EID approach offers a vital contribution to the theoretical understanding of the CT in PL, that is not accessible with an adiabatic description or using equilibrium Marcus' theory.

To conclude, the CG-EID approach, as applied in this work, comprises a powerful tool for the direct simulation of CT processes in complex systems, such as biomolecules in polar solvents representing rather heterogeneous environments, in particular if the charge motion and the dynamics of environment cannot be separated. In the light of the results presented in this work, the CT over the triple Trp chain as part of the photoactivation process in PL is driven by the solvent which withdraws the electron hole from the inside of the protein.
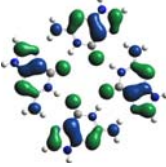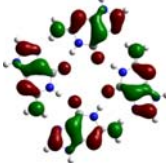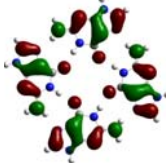
Appendices

A

Molecular Orbitals and Energies

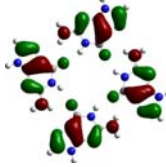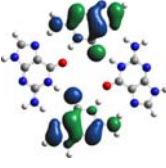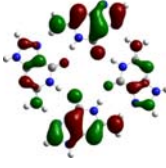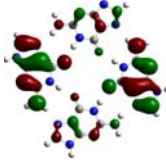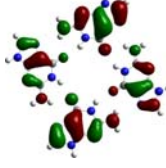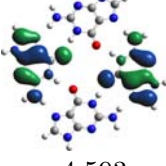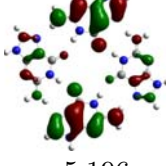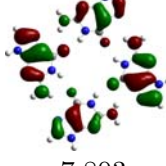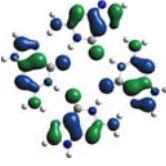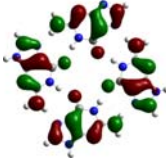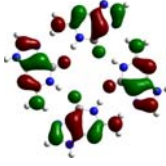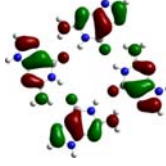**Table A.1:** Character and energy of highest occupied MOs of DNA nucleobases using B3LYP, BLYP, PBE and HF (basis set: 6-31G(d,p)), as well as DFTB (structures optimized with the respective method). Energies in eV.

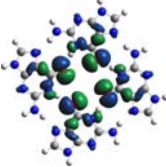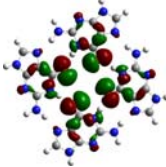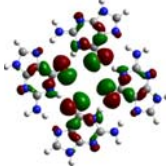| nucleobase	method	HOMO MO	HOMO energy	HOMO-1 MO	HOMO-1 energy	HOMO-2 MO	HOMO-2 energy
adenine	HF	π_{A1}	-8.373	π_{A2}	-9.988	σ_{A1}	-10.955
	B3LYP	π_{A1}	-5.891	σ_{A1}	-6.692	π_{A2}	-7.138
	BLYP	π_{A1}	-4.853	σ_{A1}	-5.253	π_{A2}	-5.991
	PBE	π_{A1}	-5.064	σ_{A1}	-5.426	π_{A2}	-6.240
	DFTB	π_{A1}	-5.296	σ_{A1}	-5.627	π_{A2}	-6.303
cytosine	HF	π_{C1}	-9.158	π_{C2}	-10.236	σ_{C1}	-11.131
	B3LYP	π_{C1}	-6.133	σ_{C1}	-6.618	π_{C2}	-6.777
	BLYP	π_{C1}	-4.981	σ_{C1}	-5.060	π_{C2}	-5.523
	PBE	π_{C1}	-5.185	σ_{C1}	-5.215	π_{C2}	-5.722
	DFTB	σ_{C1}	-5.174	π_{C1}	-5.544	π_{C2}	-6.399
guanine	HF	π_{G1}	-7.993	π_{G2}	-10.886	σ_{G1}	-11.423
	B3LYP	π_{G1}	-5.512	σ_{G1}	-6.827	σ_{G2}	-7.161
	BLYP	π_{G1}	-4.499	σ_{G1}	-5.282	σ_{G2}	-5.657
	PBE	π_{G1}	-4.695	σ_{G1}	-5.426	σ_{G2}	-5.833
	DFTB	π_{G1}	-5.004	σ_{G1}	-5.363	σ_{G2}	-6.156
thymine	HF	π_{T1}	-9.522	π_{T2}	-11.650	σ_{T1}	-11.948
	B3LYP	π_{T1}	-6.568	σ_{T1}	-7.168	π_{T2}	-7.724
	BLYP	π_{T1}	-5.449	σ_{T1}	-5.572	σ_{T2}	-6.338
	PBE	π_{T1}	-5.629	σ_{T1}	-5.723	σ_{T2}	-6.506
	DFTB	σ_{T1}	-5.954	π_{T1}	-6.176	σ_{T2}	-6.392

Table A.2: Character and energy of highest occupied MOs of DNA nucleobases using B3LYP, BLYP, PBE and HF (6-31G*) (geometries from 3DNA). Energies in eV.

nucleobase	method	HOMO		HOMO-1		HOMO-2	
		MO	energy	MO	energy	MO	energy
adenine	HF	π_{A1}	-8.264	π_{A2}	-9.940	σ_{A1}	-10.983
	B3LYP	π_{A1}	-5.894	σ_{A1}	-6.671	π_{A2}	-7.124
	BLYP	π_{A1}	-4.855	σ_{A1}	-5.175	π_{A2}	-5.988
	PBE	π_{A1}	-5.083	σ_{A1}	-5.353	σ_{A2}	-6.193
	DFTB	π_{A1}	-5.277	σ_{A1}	-5.720	π_{A2}	-6.315
cytosine	HF	π_{C1}	-9.094	π_{C2}	-10.257	σ_{C1}	-11.021
	B3LYP	π_{C1}	-5.988	σ_{C1}	-6.233	π_{C2}	-6.882
	BLYP	σ_{C1}	-4.626	π_{C1}	-4.812	σ_{C2}	-5.583
	PBE	σ_{C1}	-4.805	π_{C1}	-5.026	σ_{C2}	-5.771
	DFTB	σ_{C1}	-4.833	π_{C1}	-5.344	σ_{C2}	-6.331
guanine	HF	π_{G1}	-7.874	π_{G2}	-10.830	σ_{G1}	-11.319
	B3LYP	π_{G1}	-5.476	σ_{G1}	-6.539	σ_{G2}	-7.063
	BLYP	π_{G1}	-4.447	σ_{G1}	-4.912	σ_{G2}	-5.517
	PBE	π_{G1}	-4.670	σ_{G1}	-5.093	σ_{G2}	-5.703
	DFTB	π_{G1}	-4.916	σ_{G1}	-5.054	σ_{G2}	-6.233
thymine	HF	π_{T1}	-9.473	π_{T2}	-11.499	σ_{T1}	-11.886
	B3LYP	π_{T1}	-6.547	σ_{T1}	-7.051	π_{T2}	-7.625
	BLYP	π_{T1}	-5.410	σ_{T1}	-5.428	σ_{T2}	-6.162
	PBE	σ_{T1}	-5.592	π_{T1}	-5.611	σ_{T2}	-6.343
	DFTB	σ_{T1}	-5.846	π_{T1}	-6.239	σ_{T2}	-6.350

Table A.3: Highest molecular orbitals of an idealized G4 tetrad with C$_{4h}$-symmetry, comparison between DFTB, DFT and HF, for the latter the 6-31G(d,p) basis set is used, MO energies in eV.
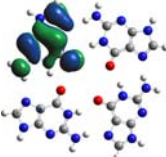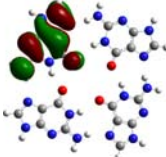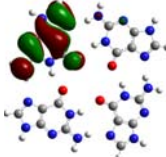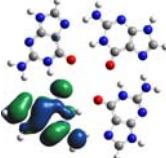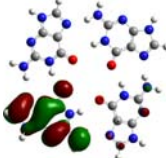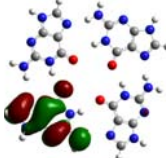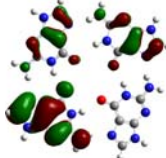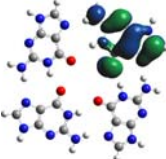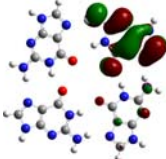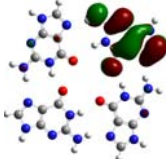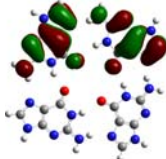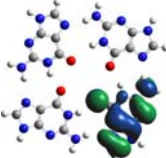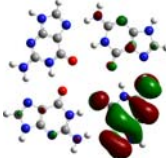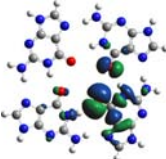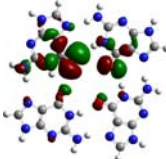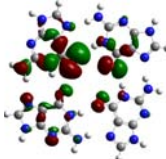
MO	DFTB	PBE	B3LYP	HF
HOMO	-4.588	-4.356	-5.170	-7.777
HOMO-1	-4.593	-4.383	-5.196	-7.803
HOMO-2	-4.593	-4.383	-5.196	-7.803
HOMO-3	-4.597	-4.408	-5.222	-7.828
HOMO-4	-4.870	-5.267	-6.637	-10.827

Table A.4: Highest molecular orbital energies of an idealized G4 tetrad in the presence of a single central monovalent ion: i) sodium, ii) lithium or iii) potassium, either as part of the QM system or as MM point charge. Comparison between DFTB, DFT and HF, for the latter the 6-31G(d,p) basis set is used, MO energies in eV.

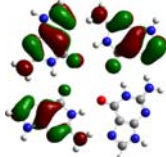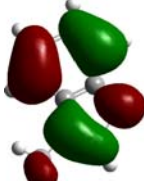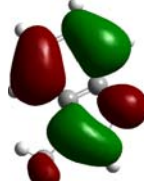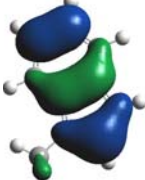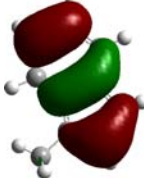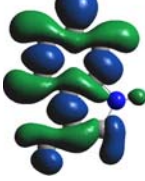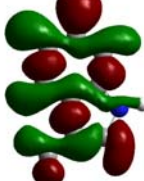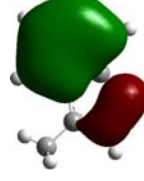|	MO	DFTB MM	PBE QM	PBE MM	B3LYP QM	B3LYP MM	HF QM	HF MM
	HOMO	-7.363	-7.086	-7.041	-7.878	-7.832	-10.212	-10.207
	HOMO-1	-7.367	-7.104	-7.067	-7.898	-7.861	-10.238	-10.239
Li	HOMO-2	-7.367	-7.104	-7.067	-7.898	-7.861	-10.238	-10.239
	HOMO-3	-7.370	-7.126	-7.091	-7.923	-7.888	-10.268	-10.272
	HOMO-4	-8.402	-8.375	-8.341	-9.662	-9.628	-13.317	-13.312
	HOMO	-7.363	-7.096	-7.041	-7.888	-7.832	-10.223	-10.207
	HOMO-1	-7.367	-7.111	-7.067	-7.905	-7.861	-10.246	-10.239
Na	HOMO-2	-7.367	-7.111	-7.067	-7.905	-7.861	-10.246	-10.239
	HOMO-3	-7.370	-7.134	-7.091	-7.931	-7.888	-10.277	-10.272
	HOMO-4	-8.402	-8.381	-8.341	-9.668	-9.628	-13.323	-13.312
	HOMO	-7.363	-7.072	-7.041	-7.856	-7.832	-10.191	-10.207
	HOMO-1	-7.367	-7.095	-7.067	-7.882	-7.861	-10.222	-10.239
K	HOMO-2	-7.367	-7.095	-7.067	-7.882	-7.861	-10.222	-10.239
	HOMO-3	-7.370	-7.117	-7.091	-7.906	-7.888	-10.251	-10.272
	HOMO-4	-8.402	-8.371	-8.341	-9.649	-9.628	-13.306	-13.312

Table A.5: Four highest occupied molecular orbitals of a G4 tetrad as extracted from a MD snapshot. Comparison between DFTB, DFT and HF, for the latter the 6-31G(d,p) basis set is used, MO energies in eV. Note that these orbitals were obtained in presence of the electrostatic field of the environment.

MO	DFTB	PBE	B3LYP	HF
HOMO	-5.390	-5.052	-5.871	-8.322
HOMO-1	-5.409	-5.156	-5.949	-8.328
HOMO-2	-5.467	-5.206	-5.994	-8.362
HOMO-3	-5.544	-5.273	-6.121	-8.711
HOMO-4	-5.756	-5.986	-7.356	-11.113

Table A.6: Highest molecular orbitals of a tryptophan side-chain capped by a methyl group (3-methylindole), comparison between DFTB, DFT and HF, for the latter the 6-31G(d,p) basis set is used, MO energies in eV.

MO	DFTB	PBE	B3LYP	HF
HOMO	-5.331	-4.486	-5.224	-7.397
HOMO-1	-5.884	-5.101	-5.827	-7.952
HOMO-2	-6.886	-6.276	-7.196	-10.107
HOMO-3	-7.574	-7.619	-8.796	-12.392

B

Electronic Parameters

Table B.1: Site energies (diagonal) and electronic couplings (off-diagonal) of an idealized, static G-quadruplex dimer, stacking distance 3.4 Å and rotation angle of 32°, all values in eV.

	G1	G2	G3	G4	G5	G6	G7	G8
G1	-4.894	0.009	0.000	0.010	0.030	0.000	0.000	0.000
G2		-4.896	0.010	0.000	0.000	0.026	0.002	0.000
G3			-4.894	0.009	0.000	0.000	0.030	0.000
G4				-4.893	0.002	0.000	0.000	0.026
G5					-4.894	0.009	0.000	0.009
G6						-4.898	0.009	0.000
G7							-4.896	0.009
G8								-4.896

Table B.2: Site energies (diagonal) and electronic couplings (off-diagonal) of the eight central guanines of the parallel stranded 5'-TG**GG**GT-3' tetramer, with no channel ions, for static crystal structure 244d without solvent and backbone, all values in eV.

	G1	G2	G3	G4	G5	G6	G7	G8
G1	-4.829	0.008	0.000	0.007	0.048	0.010	0.000	0.000
G2		-4.965	0.005	0.000	0.000	0.048	0.016	0.000
G3			-4.843	0.006	0.000	0.000	0.069	0.012
G4				-4.897	0.010	0.000	0.000	0.039
G5					-4.930	0.010	0.000	0.009
G6						-4.871	0.001	0.000
G7							-5.029	0.007
G8								-4.874

Table B.3: Average site energies (diagonal) and electronic couplings (off-diagonal) of the eight central guanines of the parallel stranded 5'-T**GGGG**GT-3' tetramer from MD simulations (including standard deviation), with no channel ions, all values in eV.

	G1	G2	G3	G4	G5	G6	G7	G8
G1	-4.922±0.375	0.021±0.017	0.000±0.000	0.007±0.004	0.049±0.028	0.000±0.000	0.000±0.000	0.000±0.000
G2		-4.853±0.351	0.007±0.005	0.000±0.000	0.008±0.006	0.039±0.030	0.007±0.005	0.000±0.000
G3			-4.688±0.386	0.006±0.003	0.000±0.000	0.038±0.028	0.025±0.015	0.000±0.000
G4				-4.856±0.359	0.008±0.008	0.000±0.000	0.000±0.000	0.027±0.020
G5					-4.762±0.357	0.009±0.008	0.000±0.000	0.006±0.003
G6						-4.857±0.351	0.008±0.004	0.000±0.000
G7							-4.746±0.377	0.008±0.003
G8								-4.815±0.335

Table B.4: Average site energies (diagonal) and electronic couplings (off-diagonal) of the eight central guanines of the parallel stranded 5'-T**GGGG**GT-3' tetramer from MD simulations (including standard deviation), with 3 lithium channel ions, all values in eV.

	G1	G2	G3	G4	G5	G6	G7	G8
G1	-5.382±0.350	0.007±0.005	0.000±0.000	0.006±0.004	0.028±0.020	0.022±0.012	0.000±0.000	0.001±0.001
G2		-5.145±0.350	0.008±0.005	0.000±0.000	0.000±0.000	0.001±0.001	0.033±0.022	0.023±0.015
G3			-5.269±0.327	0.008±0.004	0.000±0.000	0.000±0.000	0.031±0.022	0.019±0.014
G4				-5.079±0.335	0.021±0.013	0.000±0.000	0.000±0.000	0.031±0.020
G5					-5.563±0.348	0.008±0.005	0.000±0.000	0.000±0.000
G6						-5.340±0.331	0.008±0.005	0.001±0.000
G7							-5.451±0.329	0.007±0.005
G8								-5.480±0.312

Table B.5: Average site energies (diagonal) and electronic couplings (off-diagonal) of the eight central guanines of the parallel stranded 5'-TG**GGG**GT-3' tetramer from MD simulations (including standard deviation), with 3 sodium channel ions, all values in eV.

	G1	G2	G3	G4	G5	G6	G7	G8
G1	-5.416±0.349	0.006±0.004	0.001±0.001	0.007±0.005	0.028±0.020	0.013±0.010	0.000±0.000	0.000±0.000
G2		-5.360±0.355	0.006±0.004	0.001±0.001	0.001±0.000	0.032±0.022	0.017±0.012	0.000±0.000
G3			-5.355±0.357	0.006±0.004	0.000±0.000	0.000±0.000	0.033±0.022	0.016±0.012
G4				-5.404±0.347	0.014±0.011	0.000±0.000	0.000±0.000	0.031±0.022
G5					-5.476±0.343	0.006±0.004	0.001±0.001	0.006±0.004
G6						-5.411±0.337	0.006±0.004	0.001±0.001
G7							-5.371±0.346	0.006±0.004
G8								-5.408±0.350

Table B.6: Average site energies (diagonal) and electronic couplings (off-diagonal) of the eight central guanines of the parallel stranded 5'-TG**GGG**GT-3' tetramer from MD simulations (including standard deviation), with 3 potassium channel ions, all values in eV.

	G1	G2	G3	G4	G5	G6	G7	G8
G1	-5.100±0.327	0.008±0.003	0.000±0.000	0.007±0.003	0.027±0.019	0.012±0.010	0.000±0.000	0.000±0.000
G2		-5.081±0.341	0.006±0.003	0.000±0.000	0.000±0.000	0.027±0.019	0.011±0.009	0.000±0.000
G3			-5.092±0.344	0.006±0.003	0.000±0.000	0.000±0.000	0.031±0.021	0.013±0.010
G4				-5.098±0.340	0.014±0.011	0.000±0.000	0.000±0.000	0.032±0.022
G5					-5.318±0.338	0.006±0.004	0.000±0.000	0.006±0.004
G6						-5.336±0.325	0.007±0.004	0.000±0.000
G7							-5.308±0.328	0.006±0.004
G8								-5.277±0.346

C

Length Dependence of CT in DNA

Figure C.1: Length dependence of CT parameter: mean values and standard deviations for ϵ_i (top) and T_{ij} (bottom) for poly(A) (left) and poly(G) (right)
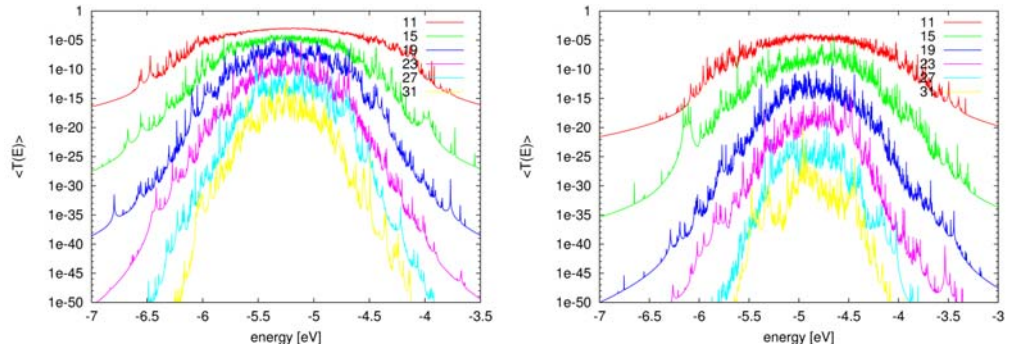
Figure C.2: Length dependence of time-averaged transmission function $\langle T(E) \rangle$ for poly(A) (left) and poly(G) (right), for DNA molecules with 11 to 31 base pairs. Data was obtained from 30 ns MD simulation, respectively.

D

Supplementary Material for part II:
CT in *E. coli* Photolyase
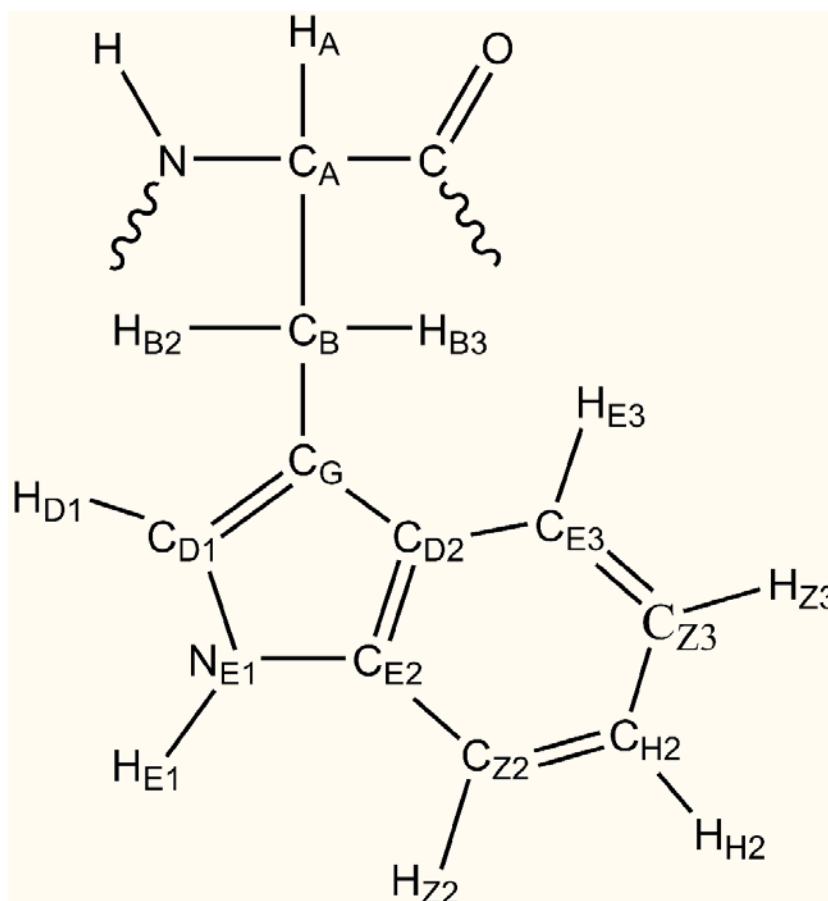
Figure D.1: Atom labels for the tryptophan (Trp) amino acid side chain.

Table D.1: Partial atomic charges for neutral and cationic Trp side chains used in the simulations in part II. Atom labels are assigned according to Fig. D.1. Charges given in e.

label	neutral species	cationic species
N	-0.41573	-0.415700
H	0.27192	0.271900
CA	-0.02750	-0.027500
HA	0.11231	0.112300
CB	-0.00500	-0.094300
HB2	0.03390	0.109100
HB3	0.03390	0.109100
CG	-0.14151	0.078300
CD1	-0.16381	-0.138100
HD1	0.20622	0.228700
NE1	-0.34183	-0.014300
HE1	0.34123	0.346300
CE2	0.13801	0.027400
CZ2	-0.26012	-0.105500
HZ2	0.15721	0.176700
CH2	-0.11341	-0.148700
HH2	0.14171	0.185800
CZ3	-0.19722	-0.037600
HZ3	0.14471	0.163100
CE3	-0.23872	-0.159900
HE3	0.17001	0.193000
CD2	0.12431	0.110500
C	0.59735	0.597300
O	-0.56794	-0.567900

Table D.2: Site energies and electronic couplings for Sim. A through E as mean $\pm$ standard deviation. ϵ_i in eV, T_{ij} in meV (T_{13} in μeV). Electronic couplings are compared to values obtained by Krapf et al., Ref. [120].

Sim.	site energies			electronic couplings		
	ϵ_1	ϵ_2	ϵ_3	T_{12}	T_{23}	T_{13} [μeV]
A	3.4±0.2	3.7±0.3	3.2±0.3	5.9±3.3	4.3±3.3	4±4
B	2.9±0.2	3.2±0.3	2.9±0.3	8.7±3.8	3.9±3.2	5±5
C	0.9±0.3	1.9±0.3	2.1±0.3	10.5±4.4	4.3±3.4	7±7
D	1.6±0.3	0.2±0.3	1.5±0.3	9.4±4.1	2.0±2.0	3±3
E	2.0±0.2	1.7±0.4	-0.3±0.3	9.5±4.5	1.5±2.0	2±3
Ref. [120]	—	—	—	9.4	2.5	5

D.1 Relaxation

Fig. D.2 shows the ESP relaxation decomposed into separated protein and solvent MM environments for both CTs 1→2 and 2→3. For the latter, the relaxations are provided on the short as well as on the long timescale.
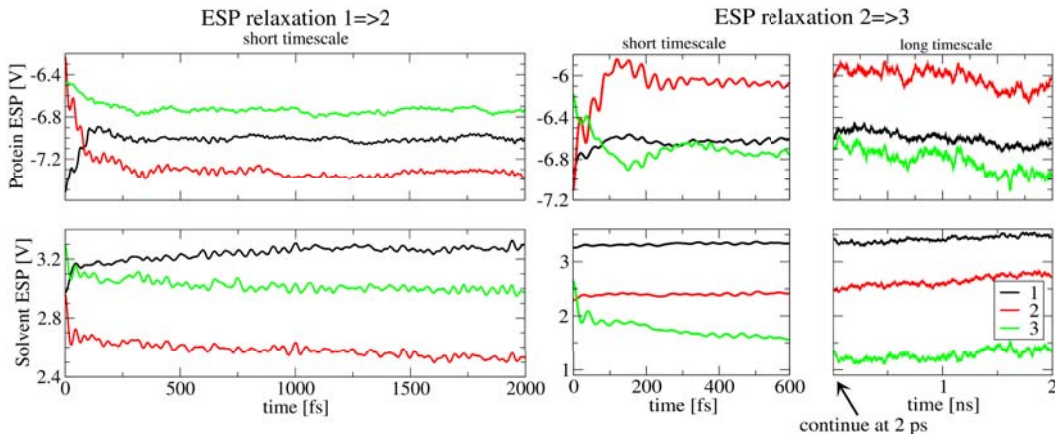
Figure D.2: ESP relaxation of separated protein (top) and solvent (bottom) MM environments for both CT steps 1→2 (left) and 2→3 (right). Data obtained as averages, respectively, from 50 classical MD simulations for both CT steps.

D.2 EID in Restricted QM Systems

In this section both single CT steps 1→2 and 2→3 are simulated and analyzed independently, i.e. treating only two sites quantum mechanically in the simulations. Typical simulations are depicted in Fig. D.3. Both transfers occur on a very short time scale of about 30 and 20 ps for 1→2 and 2→3 respectively. Moreover, there is no successful back transfer within 3 ns. Occasionally, the average site energy gap $\Delta\epsilon_{12}$ of about 1 eV can be overcome via fluctuations due to solvent and protein environment. If at the same time T_{12} is sufficiently large, fractions of the hole charge can be transfered back temporarily at that very instant. However, these rare incidents are transient, visible only as short-lived peaks in the occupation spectrum of 1→2. On the other hand, the gap for the second step $\Delta\epsilon_{23}$ is about 0.5 eV larger, plus T_{23} is significantly smaller on average, hence no such peaks are observed for 2→3. Again, more detailed information can be extracted by analyzing ensembles of simulations.

D.2.1 Statistics of Single CT: 1→2 and 2→3

As depicted in Fig. D.4, more transient fluctuations of the occupation are observed for 1→2. In fact, these fluctuations may lead to a few successful back transfers with residence times of several ps in the initial state. Nevertheless, despite few back transfers and fluctuations, the hole charge arrives at site 2 in less than 850 ps and for the majority of simulations the rate is much faster (about 200 ps).
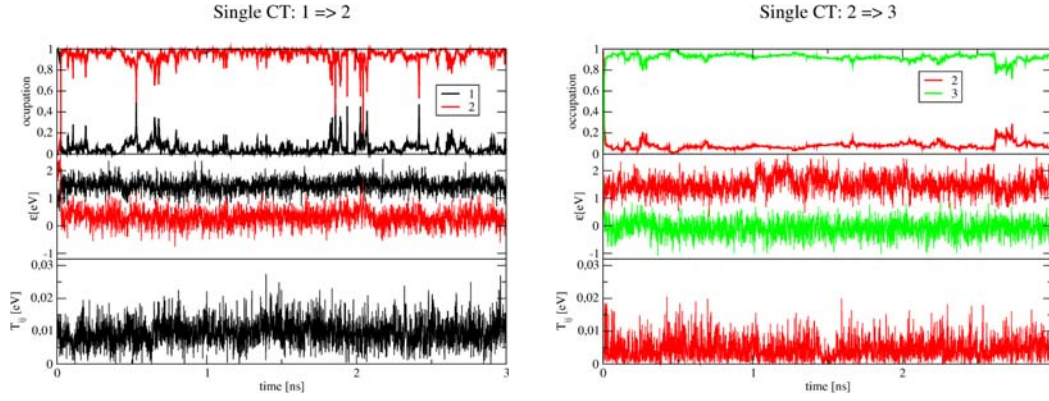
Figure D.3: Single steps of the CT process: 1→2 (left) and 2→3 (right). Top panels: hole probability, center: site energies and bottom: electronic couplings of site 1, 2 and 3.

On the other hand, the transfer 2→3 occurs within 700 ps, while exhibiting significantly less fluctuation of occupations. Several back transfers are observed for 2→3 as well, i.e. in two of them the hole charge persists almost 200 ps in the initial state. However, after 700 ps, the hole charge is localized almost completely on the final site 3 showing practically no fluctuations. Basically, both reaction rates have the same order of magnitude, although 1→2 is slightly faster, i.e. the average occupation of the final states is already 0.8 at about 200 ps, while twice as much simulation time is needed for 2→3. The small difference for the reaction rates is indicated by the average hole center of mass as well, which is given in Fig. D.4. More diversity (various rates) are obtained for 2→3 which is indicated by slightly larger standard deviations as shown in Fig. D.4. On the contrary, simulations for 1→2 seem to be more homogeneous, i.e. disregarding occupation fluctuations and back transfers for the moment, the rates are more similar, thus the transfer process itself may be seen as more coherent. The higher activity for 1→2 can be explained in terms of both $\Delta\epsilon$ and T_{ij}, since $T_{12} > T_{23}$ and $\Delta\epsilon_{12} < \Delta\epsilon_{23}$, as can be extracted from Tab. D.2.

D.2.2   Starting from Different Ensembles

In Fig. D.5, single step EID simulations are compared by means of different starting configurations, that is the EID simulations were either started from Sim. A or from Sim. C and D, respectively, for 1→2 and 2→3.
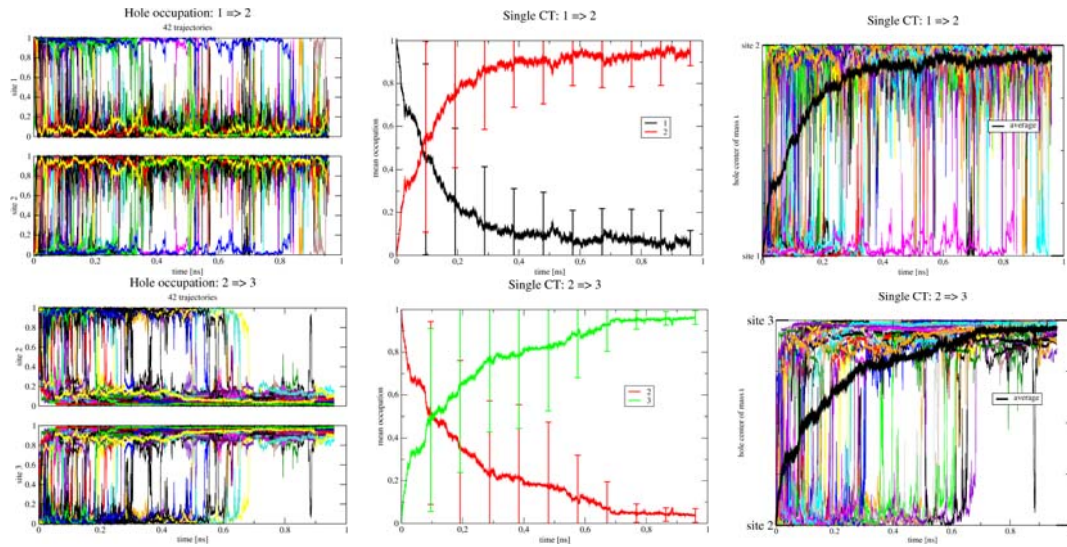
Figure D.4: Statistics of single CT 1→2 (top) and 2→3 (bottom): Occupation time series (left), average occupations with standard deviations (center), and hole center of mass (right) obtained from 42 EID simulations, respectively.
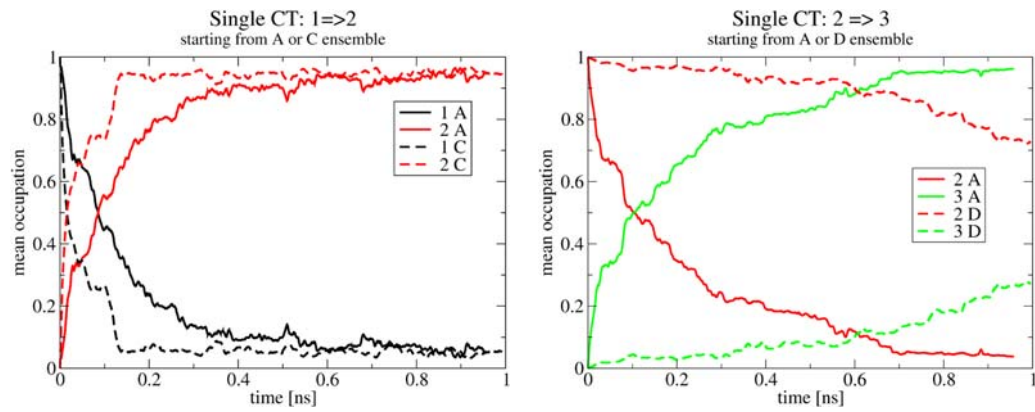
Figure D.5: Average occupations for single CT steps 1→2 (left) and 2→3 (right). Comparison between different starting configurations, i.e. starting the EID simulations from Sim. A ensemble or rather on structures taken from Sim. C or D, respectively. For the latter option, the system is supposed to be already relaxed according to the hole charge on the respective site.
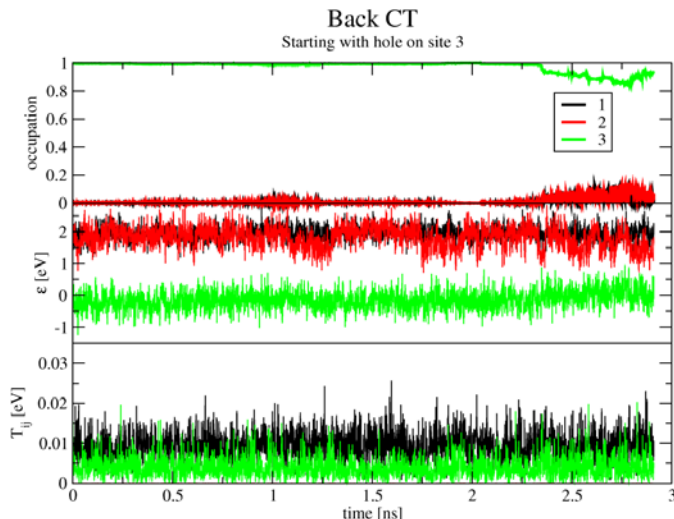
Figure D.6: (No) back-transfer of the hole charge from $3 \to 2 \to 1$. Top: hole probability, center: site energies, and bottom: electronic couplings of sites 1, 2 and 3 along the direct CT simulation.

D.2.3 Likelihood of Back Transfers

Evidently, a strong exothermic character of the CT process $1 \to 2 \to 3$ is exhibited, which is achieved predominantly by energetic phenomena due to the interplay of protein and solvent environment. Nevertheless, the question arises whether the CT process is reversible, since several simulations indicated that back transfers are indeed possible. Note that in these simulations the hole charge has never been localized completely on the final site. In order to verify the (ir)reversibility of the CT process a direct simulation is carried out in which the hole charge has been placed initially on site 3.

In Fig. D.6, occupation, site energy and electronic coupling time series are given for the CT simulation $3 \to 2 \to 1$. To begin with, there is no back transfer of the hole charge within 3 ns simulation time. The hole appears to be stabilized significantly on site 3 which is indicated by $\Delta\epsilon_{23}$ taking values of about 1.5 to 2 eV. Such a high energy barrier is hard to overcome even via considerable fluctuations of ϵ_i of about 0.4 eV due to protein and solvent environment. Additionally, the electronic coupling T_{23} is fairly small with about 2 meV (see Tab. D.2). The major part of the stabilization of the hole charge on site 3 comes from the polarization of the surrounding water molecules, i.e. the reorganization of the solvent molecules. As already discussed, the stabilization is pronounced most strongly on site 3 which is closest to the solvent. A hole charge located on site 2 may apparently benefit slightly from solvent reorganization as well, even though not as strongly. Site 1 is affected only marginally by the solvent, since it is located close to the center of the protein. The overall charge stabilization by the solvent appears to outweigh that by the protein environment.

Nevertheless, starting at about 2.3 ns very small portions of the hole charge are transferred from site 3 to site 2, since the energy gap $\Delta\epsilon_{23}$ is reduced and CT is enabled via fluctuations due to the environment. Moreover, it appears that two or more local,
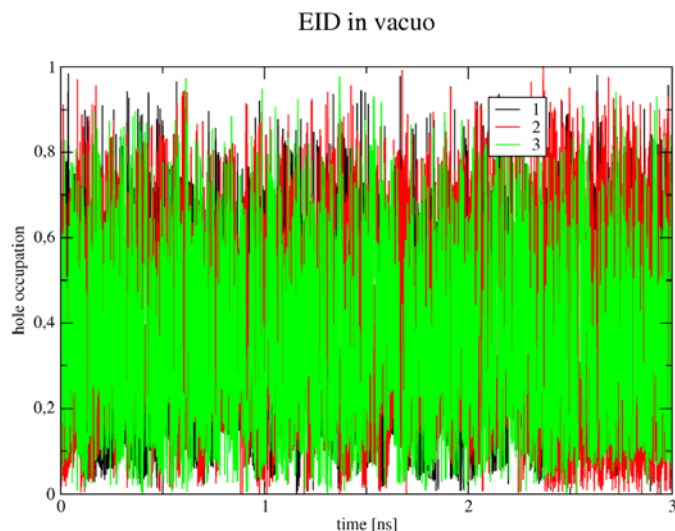
Figure D.7: EID occupations in vacuo, i.e. the QM/MM term in the Hamiltonian (Eq. 2.126) is switched off.

metastable states might be present. For instance, at about 1.3 ns there is a sudden jump of the site energy of the central site only, leading to an increased gap $\Delta\epsilon_{23}$ of more than 2 eV which is maintained over several hundred ps, after which ϵ_2 drops back again to its former value.

D.3 EID in Vacuo

The ESP acting on the respective CT sites has been decomposed already into both its components of protein and solvent environment. Here, a CT simulation has been carried out in which neither protein nor solvent has been taken into account (Fig. D.7), i.e. the QM/MM term in the electronic Hamiltonian (Eq. 2.126) has been switched off. As a result, the hole charge is likely to hop back and forth between the three sites, for there is almost no energy barrier between the corresponding states.

Bibliography

[1] Watson, J. D.; Crick, F. H. A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.

[2] Hihath, J.; Chen, F.; Zhang, P.; Tao, N. Thermal and electrochemical gate effects on DNA conductance. *J. Phys. Cond. Matter* **2007**, *19*, 215202.

[3] Endres, R. G.; Cox, D. L.; Singh, R. R. P. Colloquium: The quest for high-conductance DNA. *Rev. Mod. Phys.* **2004**, *76*, 195–214.

[4] Boon, E. M.; Livingston, A. L.; Chmiel, N. H.; David, S. S.; Barton, J. K. DNA-mediated charge transport for DNA repair. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12543–12547.

[5] Holman, M. R.; Ito, T.; Rokita, S. E. Self-Repair of Thymine Dimer in Duplex DNA. *J. Am. Chem. Soc.* **2007**, *129*, 6–7.

[6] Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **2006**, *440*, 297–302.

[7] Shih, W.; Quispe, J.; Joyce, G. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **2004**, *427*, 618–621.

[8] Douglas, S. M.; Dietz, H.; Liedl, T.; Hoegberg, B.; Graf, F.; Shih, W. M. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **2009**, *459*, 414–418.

[9] Keren, K.; Berman, R. S.; Buchstab, E.; Sivan, U.; Braun, E. DNA-Templated Carbon Nanotube Field-Effect Transistor. *Science* **2003**, *302*, 1380.

[10] Braun, E.; Eichen, Y.; Sivan, U.; Ben-Yoseph, G. DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature* **1998**, *391*, 775–778.

[11] Mertig, M.; Kirsch, R.; Pompe, W.; Engelhardt, H. Fabrication of Highly Oriented Nanocluster Arrays by Biomolecular Templating. *Eur. Phys. J. D* **1999**, *9*, 45.

[12] Ladik, J. Investigation of the electronic structure of desoxyribonucleic acid. *Acta physica Academiae Scientiarum Hungaricae* **1960**, *11*, 239–258.

[13] Eley, D. D.; Spivey, D. I. Semiconductivity of organic substances. Part 9.—Nucleic acid in the dry state. *Trans. Faraday Soc.* **1962**, *58*, 411–415.

[14] Kelley, S. O.; Barton, J. K. Electron transfer between bases in double helical DNA. *Science* **1999**, *283*, 375–381.

[15] O'Neil, M. A.; Barton, J. K. DNA Charge Transport: Conformationally Gated Hopping through Stacked Domains. *J. Am. Chem. Soc.* **2004**, *126*, 11471–11483.

[16] Treadway, C. R.; Hill, M. G.; Barton, J. K. Charge Transport Through a Molecular π-Stack: Double Helical DNA. *Chem. Phys.* **2002**, *281*, 409.

[17] Murphy, C. J.; Arkin, M. R.; Jenkins, Y.; Ghatlia, N. D.; Bossmann, S. H.; Turro, N. J.; Barton, J. K. Long-range photoinduced electron transfer through a DNA helix. *Science* **1993**, *262*, 1025.

[18] Turro, N. J.; Barton, J. K. Paradigms, Supermolecules, Electron Transfer and Chemistry at a Distance. What's the Problem? The Science or the Paradigm? *J. Biol. Inorg. Chem.* **1998**, *3*, 201.

[19] Boon, E. M.; Barton, J. K. Charge Transport in DNA. *Curr. Op. Struct. Biol.* **2002**, *12*, 320.

[20] Wan, C.; Fiebig, T.; Kelley, S. O.; Treadway, C. R.; Barton, J. K. Femtosecond dynamics of DNA-mediated lectron transfer. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6014.

[21] Kelley, S. O.; Jackson, N. M.; Hill, M. G.; Barton, J. K. Long Range Electron Transfer Through DNA Films. *Angew. Chem. Int. Ed.* **1999**, *38*, 941.

[22] Giese, B. Long-distance electron transfer through DNA. *Annual Review of Biochemistry* **2002**, *71*, 51–70.

[23] Meggers, E.; Michel-Beyerle, M.; Giese, B. Sequence Dependent Long Range Hole Transport in DNA. *J. Am. Chem. Soc.* **1998**, *120*, 12950–12955.

[24] Giese, B.; Amaudrut, J.; Köhler, A.; Spormann, M.; Wessely, S. Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling. *Nature* **2001**, *412*, 318–320.

[25] Lewis, F. D.; Wu, T.; Zhang, Y.; Letsinger, R. L.; Greenfield, S. R.; Wasielewski, M. R. Distance-dependent electron transfer in DNA hairpins. *Science* **1997**, *277*, 673–676.

[26] Lewis, F. D.; Liu, X.; Liu, J.; Miller, S. E.; Hayes, R. T.; Wasielewski, M. R. Direct measurement of hole transport dynamics in DNA. *Nature* **2000**, *406*, 51–53.

[27] Lewis, F. D.; Letsinger, R. L.; Wasielewski, M. R. Dynamics of photoinduced charge transfer and hole transport in synthetic DNA hairpins. *Acc. Chem. Res.* **2001**, *34*, 159–170.

[28] Ly, D.; Sanii, L.; Schuster, G. B. Mechanism of Charge Transport in DNA: Internally-Linked Anthraquinone Conjugates Support Phonon-Assisted Polaron Hopping. *J. Am. Chem. Soc.* **1999**, *121*, 9400.

[29] Henderson, P. T.; Hampikian, G.; Jones, D.; Kan, Y.; Schuster, G. B. Long-distance charge transport in duplex DNA: the phonon-assisted polaron-like hopping mechanism. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8353.

[30] Schuster, G. B. Long-Range Charge Transfer in DNA: Transient Structural Distortions Control the Distance Dependence. *Acc. Chem. Res.* **2000**, *33*, 253–260.

[31] *Long-Range Charge Transfer in DNA I and II*; Schuster, G. B., Ed.; Springer: Berlin, 2004; Vol. 237.

[32] Giese, B.; Wessely, S.; Spormann, M.; Lindemann, U.; Meggers, E.; Michel-Beyerle, M. E. On the mechanism of long-range electron transfer through DNA. *Angew. Chem. Int. Ed.* **1999**, *38*, 996–998.

[33] Jortner, J.; Bixon, M.; Langenbacher, T.; Michel-Beyerle, M. E. Charge transfer and transport in DNA. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12759.

[34] Bixon, M.; Giese, B.; Wessely, S.; Langenbacher, T.; Michel-Beyerle, M. E.; Jortner, J. Long-range charge hopping in DNA. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11713.

[35] Slinker, J. D.; Muren, N. B.; Renfrew, S. E.; Barton, J. K. DNA charge transport over 34 nm. *Nature Chemistry* **2011**, *3*, 228–233.

[36] *Charge Transfer in DNA: From Mechanism to Application*; Wagenknecht, H.-A., Ed.; Wiley-VCH, Weinheim, 2005.

[37] Jortner, J.; Bixon, M.; Voityuk, A. A.; Rösch, N. Superexchange mediated charge hopping in DNA. *J. Phys. Chem. A* **2002**, *106*, 7599–7606.

[38] Voityuk, A.; Rösch, N. Quantum Chemical Modeling of Electron Hole Transfer through π-Stacks of Normal and Modified Pairs of Nucleobases. *J. Phys. Chem. B* **2002**, *106*, 3013–3018.

[39] Voityuk, A. A.; Rösch, N.; Bixon, M.; Jortner, J. Electronic coupling for charge transfer and transport in DNA. *J. Phys. Chem. B* **2000**, *104*, 9740–9745.

[40] Troisi, A.; Orlandi, G. The hole transfer in DNA: Calculation of electron coupling between close bases. *Chem. Phys. Lett.* **2001**, *344*, 509–518.

[41] Berlin, Y. A.; Kurnikov, I. V.; Beratan, D.; Ratner, M. A.; Burin, A. L. DNA Electron Transfer Processes: Some Theoretical Notions. *Topics in Current Chemistry* **2004**, *237*, 1–36.

[42] Rösch, N.; Voityuk, A. A. Quantum Chemical Calculation of Donor-Acceptor Coupling for Charge Transfer in DNA. *Topics in Current Chemistry* **2004**, *237*, 37–72.

[43] Hatcher, E.; Balaeff, A.; Keinan, S.; Venkatramani, R.; Beratan, D. N. PNA versus DNA: Effects of Structural Fluctuations on Electronic Structure and Hole-Transport Mechanisms. *J. Am. Chem. Soc.* **2008**, *130*, 11752–11761.

[44] Genereux, J. C.; Wuerth, S. M.; Barton, J. K. Single-Step Charge Transport through DNA over Long Distances. *J. Am. Chem. Soc.* **2011**, *133*, 3863–3868.

[45] DePablo, P. J.; Moreno-Herrero, F.; Colchero, J.; Gómez-Herrero, J.; Herrero, P.; Baró, A. M.; Ordejón, P.; Soler, J. M.; Artacho, E. Absence of dc-conductivity in λ-DNA. *Phys. Rev. Lett.* **2000**, *85*, 4992–4995.

[46] Storm, A. J.; VanNoort, J.; DeVries, S.; Dekker, C. Insulating behavior for DNA molecules between nanoelectrodes at the 100 nm length scale. *Appl. Phys. Lett.* **2001**, *79*, 3881–3883.

[47] Zhang, Y.; Austin, R. H.; Kraeft, J.; Cox, E. C.; Ong, N. P. Insulating behavior of λ-DNA on the micron scale. *Phys. Rev. Lett.* **2002**, *89*, 198102/1–198102/4.

[48] Porath, D.; Bezryadin, A.; DeVries, S.; Dekker, C. Direct measurement of electrical transport through DNA molecules. *Nature* **2000**, *403*, 635–638.

[49] Rakitin, A.; Aich, P.; Papadopoulos, C.; Kobzar, Y.; Vedeneev, A. S.; Lee, J. S.; Xu, J. M. Metallic conduction through engineered DNA: DNA nanoelectronic building blocks. *Phys. Rev. Lett.* **2001**, *86*, 3670–3673.

[50] Fink, H.; Schönenberger, C. Electrical conduction through DNA molecules. *Nature* **1999**, *398*, 407–410.

[51] Cai, L.; Tabata, H.; Kawai, T. Self-assembled DNA networks and their electrical conductivity. *Appl. Phys. Lett.* **2000**, *77*, 3105–3106.

[52] Tran, P.; Alavi, B.; Gruner, G. Charge transport along the λ-DNA double helix. *Phys. Rev. Lett.* **2000**, *85*, 1564–1567.

[53] Yoo, K.-H.; Ha, D. H.; Lee, J.-O.; Park, J. W.; Kim, J.; Kim, J. J.; Lee, H.-Y.; Kawai, T.; Choi, H. Y. Electrical Conduction through Poly(dA)-Poly(dT) and Poly(dG)-Poly(dC) DNA Molecules. *Phys. Rev. Lett.* **2001**, *87*, 198102.

[54] Xu, B.; Zhang, P.; Li, X.; Tao, N. Direct conductance measurement of single DNA molecules in aqueous solution. *Nano Letters* **2004**, *4*, 1105–1108.

[55] Cohen, H.; Nogues, C.; Naaman, R.; Porath, D. Direct measurement of electrical transport through single DNA molecules of complex sequence. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11589–11593.

[56] Kasumov, A. Y.; Kociak, M.; Guéron, S.; Reulet, B.; Volkov, V. T.; Klinov, D. V.; Bouchiat, H. Proximity-induced superconductivity in DNA. *Science* **2001**, *291*, 280–282.

[57] Van Zalinge, H.; Schiffrin, D. J.; Bates, A. D.; Starikov, E. B.; Wenzel, W.; Nichols, R. J. Variable-temperature measurements of the single-molecule conductance of double-stranded DNA. *Angew. Chem. Int. Ed.* **2006**, *45*, 5499–5502.

[58] Hihath, J.; Xu, B.; Zhang, P.; Tao, N. Study of single-nucleotide polymorphisms by means of electrical conductance measurements. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16979–16983.

[59] Mahapatro, A. K.; Jeong, K. J.; Lee, G. U.; Janes, D. B. Sequence specific electronic conduction through polyion-stabilized double-stranded DNA in nanoscale break junctions. *Nanotechnology* **2007**, *18*, 195202.

[60] Kang, N.; Erbe, A.; Scheer, E. Observation of negative differential resistance in DNA molecular junctions. *Appl. Phys. Lett.* **2010**, *96*, 023701.

[61] Xiao, X.; Xu, B. Q.; Tao, N. J. Measurement of Single Molecule Conductance: Benzenedithiol and Benzenedimethanethiol. *Nano letters* **2004**, *4*, 267.

[62] Gutièrrez, R.; Mandai, S.; Cuniberti, G. Quantum transport through a DNA wire in a dissipative environment. *Nano Letters* **2005**, *5*, 1093–1097.

[63] Gutièrrez, R.; Mohapatra, S.; Cohen, H.; Porath, D.; Cuniberti, G. Inelastic quantum transport in a ladder model: Implications for DNA conduction and comparison to experiments on suspended DNA oligomers. *Phys. Rev. B* **2006**, *74*, 235105.

[64] Malyshev, A. V. DNA double helices for single molecule electronics. *Phys. Rev. Lett.* **2007**, *98*, 096801.

[65] Guo, A.; Xu, H. Effects of interbase electronic coupling and electrode on charge transport through short DNA molecules: A numerical study. *Physics Letters, Section A: General, Atomic and Solid State Physics* **2007**, *364*, 48–53.

[66] Wei, J. H.; Chan, K. S. Charge transport in polyguanine-polycytosine DNA molecules. *J. Phys. Cond. Matter* **2007**, *19*, 286101.

[67] Tada, T.; Kondo, M.; Yoshizawa, K. Theoretical Measurements of Conductance in an (AT)12 DNA Molecule. *ChemPhysChem* **2003**, *4*, 1256–1260.

[68] Phan, A.; Mergny, J. Human telomeric DNA: G-quadruplex, i-motif and watson-crick double helix. *Nucl. Acids Res.* **2002**, *30*, 4618–4625.

[69] Davis, J. G-Quartets 40 Years Later: From 5-GMP to Molecular Biology and Supramolecular Chemistry. *Angew. Chem. Int. Ed.* **2004**, *43*, 668–698.

[70] Williamson, J. G-Quartet Structures in Telomeric DNA. *Annual Review of Biophysics and Biomolecular Structure* **1994**, *23*, 703–750.

[71] Mergny, J. L.; Mailliet, P.; Lavelle, F.; Riou, J. F.; Laoui, A.; Helene, C. The development of telomerase inhibitors: the G-quartet approach. *Anti-Cancer Drug Design* **1999**, *14*, 327–339.

[72] Arthanari, H.; Bolton, P. Functional and dysfunctional roles of quadruplex DNA in cells. *Chemistry & Biology* **2001**, *8*, 221–230.

[73] Maizels, N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature Structural & Molecular Biology* **2006**, *13*, 1055–1059.

[74] Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988**, *334*, 364–366.

[75] Sen, D.; Gilbert, W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **1990**, *344*, 410–414.

[76] Sen, D.; Gilbert, W. Guanine quartet structures. *Methods in Enzymology* **1992**, *211*, 191–199.

[77] Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica* **1963**, *16*, 907–916.

[78] Simonsson, T. G-quadruplex DNA structures - Variations on a theme. *Biological Chemistry* **2001**, *382*, 621–628.

[79] Cohen, H.; Sapir, T.; Borovok, N.; Molotsky, T.; Di Felice, R.; Kotlyar, A. B.; Porath, D. Polarizability of G4-DNA observed by electrostatic force microscopy measurements. *Nano Letters* **2007**, *7*, 981–986.

[80] Kotlyar, A. B.; Borovok, N.; Molotsky, T.; Cohen, H.; Shapir, E.; Porath, D. Long, monomolecular guanine-based nanowires. *Advanced Materials* **2005**, *17*, 1901–1905.

[81] Borovok, N.; Molotsky, T.; Ghabboun, J.; Porath, D.; Kotlyar, A. Efficient procedure of preparation and properties of long uniform G4-DNA nanowires. *Anal. Biochem.* **2008**, *374*, 71–78.

[82] Liu, S.-P.; Weisbrod, S. H.; Tang, Z.; Marx, A.; Scheer, E.; Erbe, A. Direct Measurement of Electrical Transport Through G-Quadruplex DNA with Mechanically Controllable Break Junction Electrodes. *Angew. Chem. Int. Ed.* **2010**, *49*, 3313–3316.

[83] Špačková, N.; Berger, I.; Šponer, J. Nanosecond molecular dynamics simulations of parallel and antiparallel guanine quadruplex DNA molecules. *J. Am. Chem. Soc.* **1999**, *121*, 5519–5534.

[84] Špačková, N.; Berger, I.; Šponer, J. Structural dynamics and cation interactions of DNA quadruplex molecules containing mixed guanine/cytosine quartets: Revealed by large-scale MD simulations. *J. Am. Chem. Soc.* **2001**, *123*, 3295–3307.

[85] Cavallari, M.; Calzolari, A.; Garbesi, A.; Di Felice, R. Stability and migration of metal ions in G4-wires by molecular dynamics simulations. *J. Phys. Chem. B* **2006**, *110*, 26337–26348.

[86] Felice, R. D.; Calzolari, A. In *Electronic structure of DNA derivatives and mimics by Density Functional Theory*; Modern Methods for Theoretical Physical Chemistry of Biopolymers; Elsevier, 2006; Chapter 20, pp 485–508.

[87] Calzolari, A.; Di Felice, R.; Molinari, E.; Garbesi, A. Electron channels in biomolecular nanowires. *J. Phys. Chem. B* **2004**, *108*, 2509–2515.

[88] Calzolari, A.; Di Felice, R.; Molinari, E.; Garbesi, A. G-quartet biomolecular nanowires. *Appl. Phys. Lett.* **2002**, *80*, 3331–3333.

[89] Calzolari, A.; Di Felice, R.; Molinari, E. Electronic properties of guanine-based nanowires. *Solid State Commun.* **2004**, *131*, 557–564.

[90] Di Felice, R.; Calzolari, A.; Zhang, H. Towards metalated DNA-based structures. *Nanotechnology* **2004**, *15*, 1256–1263.

[91] Di Felice, R.; Calzolari, A.; Garbesi, A.; Alexandre, S. S.; Soler, J. M. Strain-dependence of the electronic properties in periodic quadruple helical G4-wires. *J. Phys. Chem. B* **2005**, *109*, 22301–22307.

[92] Guo, A.-M.; Xiong, S.-J. Effects of contact and efficient charge transport in G4-DNA molecules. *Phys. Rev. B* **2009**, *80*, 035115.

[93] Guo, A.-M.; Yang, Z.; Zhu, H.-J.; Xiong, S.-J. Influence of backbone on the charge transport properties of G4-DNA molecules: a model-based calculation. *J. Phys. Cond. Matter* **2010**, *22*, 065102.

[94] Weber, S. Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Biochim. Biophys. Acta Bioenerg.* **2005**, *1707*, 1–23.

[95] Sancar, A. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem. Rev.* **2003**, *103*, 2203–2237.

[96] Byrdin, M.; Sartor, V.; Eker, A. P. M.; Vos, M. H.; Aubert, C.; Brettel, K.; Mathis, P. Intraprotein electron transfer and proton dynamics during photoactivation of DNA photolyase from E. coli: review and new insights from an "inverse" deuterium isotope effect. *Biochim. Biophys. Acta Bioenerg.* **2004**, *1655*, 64–70.

[97] Carell, T.; Burgdorf, L. T.; Kundu, L. M.; Cichon, M. The mechanism of action of DNA photolyases. *Curr. Op. Chem. Biol.* **2001**, *5*, 491 – 498.

[98] Byrdin, M.; Lukacs, A.; Thiagarajan, V.; Eker, A. P. M.; Brettel, K.; Vos, M. H. Quantum Yield Measurements of Short-Lived Photoactivation Intermediates in DNA Photolyase: Toward a Detailed Understanding of the Triple Tryptophan Electron Transfer Chain. *J. Phys. Chem. A* **2010**, *114*, 3207–3214.

[99] Park, H.; Kim, S.; Sancar, A.; Deisenhofer, J. Crystal structure of DNA photolyase from Escherichia coli. *Science* **1995**, *268*, 1866–1872.

[100] Kelner, A. Effect of Visible Light on the Recovery of Streptomyces Griseus Conidia from Ultraviolet Irradiation Injury. *Proc. Natl. Acad. Sci. USA* **1949**, *35*, 73–79.

[101] Sancar, A. Structure and Function of Photolyase and in Vivo Enzymology: 50th Anniversary. *J. Biol. Chem.* **2008**, *283*, 32153–32157.

[102] Moldt, J.; Pokorny, R.; Orth, C.; Linne, U.; Geisselbrecht, Y.; Marahiel, M.; Essen, L.; Batschauer, A. Photoreduction of the folate cofactor in members of the photolyase family. *J. Biol. Chem.* **2009**, *284*, 21670–21683.

[103] Essen, L.; Klar, T. Light-driven DNA repair by photolyases. *Cell. Mol. Life Sci.* **2006**, *63*, 1266–1277.

[104] Brettel, K.; Byrdin, M. Reaction mechanisms of DNA photolyase. *Curr. Op. Struct. Biol.* **2010**, *20*, 693–701.

[105] Kodali, G.; Siddiqui, S.; Stanley, R. Charge Redistribution in Oxidized and Semiquinone E. coli DNA Photolyase upon Photoexcitation: Stark Spectroscopy Reveals a Rationale for the Position of Trp382. *J. Am. Chem. Soc.* **2009**, *131*, 4795–4807.

[106] Beukers, R.; Eker, A.; Lohman, P. 50 years thymine dimer. *DNA Repair* **2008**, *7*, 530–543.

[107] Lin, C.; Todo, T. The cryptochromes. *Genome Biology* **2005**, *6*, 220.

[108] Giovani, B.; Byrdin, M.; Ahmad, M.; Brettel, K. Light-induced electron transfer in a cryptochrome blue-light photoreceptor. *Nature Struct. Biol.* **2003**, *10*, 489–490.

[109] Biskup, T.; Schleicher, E.; Okafuji, A.; Link, G.; Hitomi, K.; Getzoff, E.; Weber, S. Direct Observation of a Photoinduced Radical Pair in a Cryptochrome Blue-Light Photoreceptor. *Angew. Chem. Int. Ed.* **2009**, *48*, 404–407.

[110] Byrdin, M.; Eker, A.; Vos, M.; Brettel, K. Dissection of the triple tryptophan electron transfer chain in Escherichial coli DNA photolyase: Trp382 is the primary donor in photoactivation. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8676–8681.

[111] Aubert, C.; Vos, M.; Mathis, P.; Eker, A.; Brettel, K. Intraprotein radical transfer during photoactivation of DNA photolyase. *Nature* **2000**, *405*, 586–590.

[112] Lukacs, A.; Eker, A.; Byrdin, M.; Brettel, K.; Vos, M. Electron hopping through the 15 A Triple tryptophan molecular wire in DNA photolyase occurs within 30 ps. *J. Am. Chem. Soc.* **2008**, *130*, 14394–14395.

[113] Chatgilialoglu, C.; Guerra, M.; Kaloudis, P.; Houee-Levin, C.; Marignier, J.; Swaminathan, V.; Carell, T. Ring opening of the cyclobutane in a thymine dimer radical anion. *Chemistry* **2007**, *13*, 8979–8984.

[114] Tachikawa, H.; Kawabata, H. Interaction between thymine dimer and flavin-adenine dinucleotide: a DFT and direct ab initio molecular dynamics study. *J. Phys. Chem. B* **2008**, *112*, 7315–7319.

[115] Prytkova, T. R.; Beratan, D. N.; Skourtis, S. S. Photoselected electron transfer pathways in DNA photolyase. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 802–807.

[116] Acocella, A.; Jones, G.; Zerbetto, F. What is adenine doing in photolyase? *J. Phys. Chem. B* **2010**, *114*, 4101–4106.

[117] Sadeghian, K.; Bocola, M.; Merz, T.; Schütz, M. Theoretical Study on the Repair Mechanism of the (6–4) Photolesion by the (6–4) Photolyase. *J. Am. Chem. Soc.* **2010**, *132*, 16285–16295.

[118] Harrison, C. B.; O'Neil, L. L.; Wiest, O. Computational Studies of DNA Photolyase. *J. Phys. Chem. A* **2005**, *109*, 7001–7012.

[119] Masson, F.; Laino, T.; Röthlisberger, U.; Hutter, J. A QM/MM Investigation of Thymine Dimer Radical Anion Splitting Catalyzed by DNA Photolyase. *ChemPhysChem* **2009**, *10*, 400–410.

[120] Krapf, S.; Koslowski, T.; Steinbrecher, T. The thermodynamics of charge transfer in DNA photolyase: using thermodynamic integration calculations to analyse the kinetics of electron transfer reactions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9516–9525.

[121] Warshel, A. Molecular dynamics simulations of biological reactions. *Acc. Chem. Res.* **2002**, *35*, 385–395.

[122] Hervas, M.; Navarro, J.; Rosa, M. D. L. Electron transfer between membrane complexes and soluble proteins in photosynthesis. *Acc. Chem. Res.* **2003**, *36*, 798–805.

[123] Schultz, B.; Chan, S. Structures and proton-pumping strategies of mitochondrial respiratory enzymes. *Ann. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 23–65.

[124] Gray, H. B.; Winkler, J. R. Electron Transfer in Proteins. *Annu. Rev. Biochem.* **1996**, *65*, 537–561.

[125] Marcus, R. A. Electron-Transfer Reactions in Chemistry - Theory and Experiment. *Rev. Mod. Phys.* **1993**, *65*, 599–610.

[126] Treutlein, H.; Schulten, K.; Brunger, A. T.; Karplus, M.; Deisenhofer, J.; Michel, H. Chromophore Protein Interactions and the Function of the Photosynthetic Reaction Center – A Molecular-Dynamics Study. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 75–79.

[127] Alden, R. G.; Parson, W. W.; Chu, Z. T.; Warshel, A. Calculations of Electrostatic Energies in Photosynthetic Reaction Centers. *J. Am. Chem. Soc.* **1995**, *117*, 12284–12298.

[128] Parson, W. W.; Chu, Z. T.; Warshel, A. Reorganization energy of the initial electron-transfer step in photosynthetic bacterial reaction centers. *Biophys. J.* **1998**, *74*, 182–191.

[129] Nonella, M.; Schulten, K. Molecular-Dynamics Simulation of Electron-Transfer in Proteins – Theory and Application to $Q_A \rightarrow Q_B$ Transfer in the Photosynthetic Reaction Center. *J. Phys. Chem.* **1991**, *95*, 2059–2067.

[130] Sterpone, F.; Ceccarelli, M.; Marchi, M. Linear response and electron transfer in complex biomolecular systems and a reaction center protein. *J. Phys. Chem. B* **2003**, *107*, 11208–11215.

[131] Wang, H.; Lin, S.; Allen, J. P.; Williams, J. C.; Blankert, S.; Laser, C.; Woodbury, N. W. Protein dynamics control the kinetics of initial electron transfer in photosynthesis. *Science* **2007**, *316*, 747–750.

[132] Warshel, A.; Parson, W. W. Dynamics of biochemical and biophysical reactions: insight from computer simulations. *Q. Rev. Biophys.* **2001**, *34*, 563–679.

[133] Parson, W. W.; Warshel, A. Dependence of Photosynthetic Electron-Transfer Kinetics on Temperature and Energy in a Density-Matrix Model. *J. Phys. Chem. B* **2004**, *108*, 10474–10483.

[134] Ceccarelli, M.; Marchi, M. Simulation and modeling of the Rhodobacter sphaeroides bacterial reaction center II: Primary charge separation. *J. Phys. Chem. B* **2003**, *107*, 5630–5641.

[135] Schulten, K.; Tesch, M. Coupling of Protein Motion to Electron-Transfer – Molecular Dynamics and Stochastic Quantum-Mechanics Study of Photosynthetic Reaction Centers. *Chem. Phys.* **1991**, *158*, 421–446.

[136] Marchi, M.; Gehlen, J.; Chandler, D.; Newton, M. Diabatic Surfaces and the Pathway for Primary Electron-Transfer in a Photosynthetic Reaction Center. *J. Am. Chem. Soc.* **1993**, *115*, 4178–4190.

[137] Balabin, I.; Onuchic, J. Dynamically Controlled Protein Tunneling Paths in Photosynthetics Reaction Centers. *Science* **2000**, *290*, 114–117.

[138] Skourtis, S. S.; Beratan, D. N. Photosynthesis from the protein's perspective. *Science* **2007**, *316*, 703–704.

[139] Datta, S. *Electronic Transport in Mesoscopic Systems*; Cambridge University Press, 1997.

[140] Datta, S. Electrical resistance: An atomistic view. *Nanotechnology* **2004**, *15*, 433–451.

[141] Jortner, J.; Nitzan, A.; Ratner, M. A. In *Foundations of Molecular Electronics – Charge Transport in Molecular Conduction Junctions*; Cuniberti, G., Fagas, G., Richter, K., Eds.; Lecture Notes in Physics; Springer: Berlin, 2005; Chapter 1, pp 13–54.

[142] Grozema, F. C.; Siebbeles, L. D. A.; Berlin, Y. A.; Ratner, M. A. Hole mobility in DNA: Effects of static and dynamic structural fluctuations. *ChemPhysChem* **2002**, *3*, 536–539.

[143] Senthilkumar, K.; Grozema, F. C.; Guerra, C. F.; Bickelhaupt, F. M.; Siebbeles, L. D. A. Mapping the Sites for Selective Oxidation of Guanines in DNA. *J. Am. Chem. Soc.* **2003**, *125*, 13658–13659.

[144] Senthilkumar, K.; Grozema, F. C.; Guerra, C. F.; Bickelhaupt, F. M.; Lewis, F. D.; Berlin, Y. A.; Ratner, M. A.; Siebbeles, L. D. A. Absolute rates of hole transfer in DNA. *J. Am. Chem. Soc.* **2005**, *127*, 14894–14903.

[145] Blancafort, L.; Voityuk, A. A. CASSCF/CAS-PT2 study of hole transfer in stacked DNA nucleobases. *J. Phys. Chem. A* **2006**, *110*, 6426–6432.

[146] Voityuk, A. A. Assessment of semiempirical methods for the computation of charge transfer in DNA π-stacks. *Chem. Phys. Lett.* **2006**, *427*, 177–180.

[147] Kubař, T.; Woiczikowski, P. B.; Cuniberti, G.; Elstner, M. Efficient calculation of charge-transfer matrix elements for hole transfer in DNA. *J. Phys. Chem. B* **2008**, *112*, 7937–7947.

[148] Cuniberti, G.; Craco, L.; Porath, D.; Dekker, C. Backbone-induced semiconducting behavior in short DNA wires. *Phys. Rev. B* **2002**, *65*, 241314.

[149] Roche, S.; Bicout, D.; Macia, E.; Kats, E. Long range correlations in DNA: Scaling properties and charge transfer efficiency. *Phys. Rev. Lett.* **2003**, *91*, 228101.

[150] Shih, C.-T.; Roche, S.; Römer, R. A. Point-Mutation Effects on Charge-Transport Properties of the Tumor-Suppressor Gene *p*53. *Phys. Rev. Lett.* **2008**, *100*, 018105.

[151] Brisker-Klaiman, D.; Peskin, U. Coherent Elastic Transport Contribution to Currents through Ordered DNA Molecular Junctions. *J. Phys. Chem. C* **2010**, *114*, 19077–19082.

[152] Voityuk, A. A.; Siriwong, K.; Rósch, N. Environmental Fluctuations Facilitate Electron-Hole Transfer from Guanine to Adenine in DNA π Stacks. *Angew. Chem. Int. Ed.* **2004**, *43*, 624–627.

[153] Berlin, Y. A.; Grozema, F. C.; Siebbeles, L. D. A.; Ratner, M. A. Charge transfer in donor-bridge-acceptor systems: Static disorder, dynamic fluctuations, and complex kinetics. *J. Phys. Chem. C* **2008**, *112*, 10988–11000.

[154] Sadowska-Aleksiejew, A.; Rak, J.; Voityuk, A. A. Effects of intra base-pairs flexibility on hole transfer coupling in DNA. *Chem. Phys. Lett.* **2006**, *429*, 546–550.

[155] Voityuk, A. A. Fluctuation of the electronic coupling in DNA: Multistate versus two-state model. *Chem. Phys. Lett.* **2007**, *439*, 162–165.

[156] Grozema, F. C.; Tonzani, S.; Berlin, Y. A.; Schatz, G. C.; Siebbeles, L. D. A.; Ratner, M. A. Effect of structural dynamics on charge transfer in DNA hairpins. *J. Am. Chem. Soc.* **2008**, *130*, 5157–5166.

[157] Siriwong, K. Ph.D. thesis, Technische Universität München, 2004.

[158] Keinan, S.; Venkatramani, R.; Balaeff, A.; Beratan, D. N. Is MD Geometry Sampling Sufficient for Nucleobase Electronic Structure Analysis of ET Reactions? Comparing Classical MD and QM/MM Methods. *J. Phys. Chem. C* **2010**, *114*, 20496–20502.

[159] Balabin, I. A.; Beratan, D. N.; Skourtis, S. S. Persistence of structure over fluctuations in biological electron-transfer reactions. *Phys. Rev. Lett.* **2008**, *101*, 158102.

[160] Skourtis, S. S.; Waldeck, D. H.; Beratan, D. N. Fluctuations in biological and bioinspired electron-transfer reactions. *Annu. Rev. Phys. Chem.* **2010**, *61*, 461–485.

[161] Beratan, D. N.; Skourtis, S. S.; Balabin, I. A.; Balaeff, A.; Keinan, S.; Venkatramani, R.; Xiao, D. Steering Electrons on Moving Pathways. *Acc. Chem. Res.* **2009**, *42*, 1669–1678.

[162] Blumberger, J.; Klein, M. L. Reorganization free energies for long-range electron transfer in a porphyrin-binding four-helix bundle protein. *J. Am. Chem. Soc.* **2006**, *128*, 13854–13867.

[163] Cascella, M.; Magistrato, A.; Tavernelli, I.; Carloni, P.; Röthlisberger, U. Role of protein frame and solvent for the redox properties of azurin from Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19641–19646.

[164] Kubař, T.; Elstner, M. What governs the charge transfer in DNA? The role of DNA conformation and environment. *J. Phys. Chem. B* **2008**, *112*, 8788–8798.

[165] Cramer, T.; Krapf, S.; Koslowski, T. DNA Charge Transfer in an External Field: An Atomistic Approach. *J. Phys. Chem. C* **2007**, *111*, 8105–8110.

[166] Gutiérrez, R.; Caetano, R. A.; Woiczikowski, B. P.; Kubař, T.; Elstner, M.; Cuniberti, G. Charge Transport through Biomolecular Wires in a Solvent: Bridging Molecular Dynamics and Model Hamiltonian Approaches. *Phys. Rev. Lett.* **2009**, *102*, 208102.

[167] Woiczikowski, P. B.; Kubař, T.; Gutiérrez, R.; Caetano, R. A.; Cuniberti, G.; Elstner, M. Combined density functional theory and Landauer approach for hole transfer in DNA along classical molecular dynamics trajectories. *J. Chem. Phys.* **2009**, *130*, 215104.

[168] Roche, S. Sequence Dependent DNA-Mediated Conduction. *Phys. Rev. Lett.* **2003**, *91*, 108101.

[169] Klotsa, D.; Römer, R. A.; Turner, M. S. Electronic Transport in DNA. *Biophys. J.* **2005**, *89*, 2187.

[170] Rodríguez, A.; Römer, R. A.; Turner, M. S. Sequence dependence of electronic transport in DNA. *Physica Status Solidi (B) Basic Research* **2006**, *243*, 373–377.

[171] Komineas, S.; Kalosakas, G.; Bishop, A. R. Effects of intrinsic base-pair fluctuations on charge transport in DNA. *Phys. Rev. E* **2002**, *65*, 061905.

[172] Woiczikowski, P. B.; Kubař, T.; Gutiérrez, R.; Cuniberti, G.; Elstner, M. Structural stability versus conformational sampling in biomolecular systems: Why is the charge transfer efficiency in G4-DNA better than in double-stranded DNA? *J. Chem. Phys.* **2010**, *133*, 035103.

[173] Kubař, T.; Elstner, M. Coarse-Grained Time-Dependent Density Functional Simulation of Charge Transfer in Complex Systems: Application to Hole Transfer in DNA. *J. Phys. Chem. B* **2010**, *114*, 11221–11240.

[174] Marcus, R. Theory of Oxidation-Reduction Reactions Involving Electron Transfer. *J. Chem. Phys.* **1956**, *24*, 966–978.

[175] Marcus, R. Electrostatic Free Energy and Other Properties of States Having Nonequilibrium Polarisation. *J. Chem. Phys.* **1956**, *24*, 979–989.

[176] Marcus, R. Chemical + Electrochemical Electron-Transfer Theory. *Ann. Rev. Phys. Chem.* **1964**, *15*, 155–196.

[177] Marcus, R. A.; Sutin, N. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta Bioenerg.* **1985**, *811*, 265–322.

[178] Levich, V.; Dogonadze, R. Theory of Non-Radiation Electron Transitions from Ion to Ion in Solutions. *Doklady Akademii Nauk SSSR* **1959**, *124*, 123–126.

[179] Hush, N. S. Adiabatic theory of outer sphere electron-transfer reactions in solution. *Trans. Faraday Soc.* **1961**, *57*, 557–580.

[180] Hopfield, J. Electron-Transfer between Biological Molecules by thermally Activated Tunneling. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 3640–3644.

[181] Jortner, J. Temperature-Dependent Activation-Energy for Electron-Transfer between Biological Molecules. *J. Chem. Phys.* **1976**, *64*, 4860–4867.

[182] Moser, C.; Keske, J.; Warncke, K.; Farid, R.; Dutton, P. Nature of biological electron transfer. *Nature* **1992**, *355*, 796–802.

[183] Kubař, T.; Elstner, M. Solvent Reorganization Energy of Hole Transfer in DNA. *J. Phys. Chem. B* **2009**, *113*, 5653–5656.

[184] Blumberger, J.; Lamoureux, G. Reorganization free energies and quantum corrections for a model electron self-exchange reaction: comparison of polarizable and non-polarizable solvent models. *Mol. Phys.* **2008**, *106*, 1597–1611.

[185] Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P.; Sprik, M. Calculation of redox properties: Understanding short- and long-range effects in rubredoxin. *J. Phys. Chem. B* **2007**, *111*, 3969–3976.

[186] Kuharski, R.; Bader, J.; Chandler, D.; Sprik, M.; Klein, M.; Impey, R. Molecular-Model for Aqueous Ferrous Ferric Electron-Transfer. *J. Chem. Phys.* **1988**, *89*, 3248–3257.

[187] Ando, K. Solvent nuclear quantum effects in electron transfer reactions. III. Metal ions in water. Solute size and ligand effects. *J. Chem. Phys.* **2001**, *114*, 9470–9477.

[188] King, G.; Warshel, A. Investigation of the Free-Energy Functions for Electron-Transfer Reactions. *J. Chem. Phys.* **1990**, *93*, 8682–8692.

[189] Rauhut, G.; Clark, T. Electron-transfer reactions: AM1 and ab initio studies on self-exchange in p-diaminobenzene systems. *J. Am. Chem. Soc.* **1993**, *115*, 9127–9135.

[190] Nelsen, S. F.; Blackstock, S. C.; Kim, Y. Estimation of inner shell Marcus terms for amino nitrogen compounds by molecular orbital calculations. *J. Am. Chem. Soc.* **1987**, *109*, 677–682.

[191] Siriwong, K.; Voityuk, A. A.; Newton, M. D.; Rösch, N. Estimate of the Reorganization Energy for Charge Transfer in DNA. *J. Phys. Chem. B* **2003**, *107*, 2595–2601.

[192] Berashevich, J. A.; Chakraborty, T. Energetics of the hole transfer in DNA duplex oligomers. *Chem. Phys. Lett.* **2007**, *446*, 159 – 164.

[193] Kubař, T. *Theoretische Biophysikalische Chemie: Modellierung biomolekularer Strukturen mit Hilfe von empirischen Kraftfeldern und Moleklardynamik-Simulationen*, University Lecture, KIT, 2010.

[194] Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

[195] Kirkwood, J. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.

[196] Jensen, F. *Introduction to Computational Chemistry*; Wiley, 1999.

[197] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701.

[198] Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

[199] Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.

[200] Jarzynski, C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E* **1997**, *56*, 5018–5035.

[201] Goette, M.; Grubmüller, H. Accuracy and convergence of free energy differences calculated from nonequilibrium switching processes. *J. Comput. Chem.* **2009**, *30*, 447–456.

[202] Crooks, G. E. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **1998**, *90*, 1481–1487.

[203] Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **1999**, *60*, 2721–2726.

[204] Jarzynski, C. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E* **2006**, *73*, 10.

[205] Hummer, G. Fast-growth thermodynamic integration: Error and efficiency analysis. *J. Chem. Phys.* **2001**, *114*, 7330–7337.

[206] Landauer, R. *IBM Philos. Mag.* **1970**, *21*, 863.

[207] Büttiker, M.; Imry, Y.; Landauer, R.; Pinhas, S. Generalized Many-Channel Conductance Formula with Application to Small Rings. *Phys. Rev. B* **1985**, *31*, 6207–6215.

[208] Büttiker, M. Four-Terminal Phase-Coherent Conductance. *Phys. Rev. Lett.* **1986**, *57*, 1761–1764.

[209] Di Felice, R.; Calzolari, A.; Varsano, D.; Rubio, A. In *Electronic Structure Calculations for -olecular Systems*; Cuniberti, G., Fagas, G., Richter, K., Eds.; Lecture Notes in Physics; Springer: Berlin, 2005; Chapter 3, pp 77–116.

[210] Zahid, F.; Paulsson, M.; Datta, S. *Electrical Conduction through Molecules*; Academic Press, 2003.

[211] Atkins, P. W. *Physikalische Chemie*; Wiley-VCH, 1990.

[212] Caroli, C.; Combescot, R.; Nozieres, P.; Saint-James, D. Direct calculation of the tunneling current. *J. Phys. C* **1971**, *4*, 916.

[213] Fisher, D. S.; Lee, P. A. Relation between conductivity and transmission matrix. *Phys. Rev. B* **1981**, *23*, 6851–6854.

[214] Meir, Y.; Wingreen, N. S. Landauer formula for the current through an interacting electron region. *Phys. Rev. Lett.* **1992**, *68*, 2512–2515.

[215] Hänggi, P.; Kohler, S.; Lehmann, J.; Strass, M. In *AC-driven transport through molecular wires*; Cuniberti, G., Fagas, G., Richter, K., Eds.; Lecture Notes in Physics; Springer: Berlin, 2005; Chapter 2, pp 55–76.

[216] Metropolis, N.; Ulam, S. The Monte Carlo Method. *Journal of the American Statistical Association* **1949**, *44*, 335–341.

[217] Newton, I. *Philosophiæ naturalis principia mathematica*; London, 1687.

[218] van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual version 4.5.4*; 2010.

[219] Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Annalen der Physik* **1927**, *389*, 457–484.

[220] Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **1974**, *14*, 148–158.

[221] Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98.

[222] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637–649.

[223] Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[224] Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

[225] Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

[226] Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

[227] Nosé, S.; Klein, M. L. A study of solid and liquid carbon tetrafluoride using the constant pressure molecular dynamics technique. *J. Chem. Phys.* **1983**, *78*, 6928–6939.

[228] Ryckaert, J.; Ciccotti, G.; Berendsen, H. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

[229] Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

[230] Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **1921**, *369*, 253–287.

[231] Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[232] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

[233] Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049.

[234] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function and Genetics* **2006**, *65*, 712–725.

[235] Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* **2007**, *92*, 3817.

[236] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.

[237] Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

[238] Waller, I. Zur Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen. *Zeitschrift für Physik A Hadrons and Nuclei* **1923**, *17*, 398–408.

[239] Willis, B. T. M.; Pryor, A. W. *Thermal vibrations in crystallography / B. T. M. Willis, A. W. Pryor*; Cambridge University Press, London ; New York :, 1975; pp xvi, 280 p. ;.

[240] Löwdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365–375.

[241] Newton, M. D. Quantum Chemical Probes of Electron-Transfer Kinetics: The Nature of Donor-Acceptor Interactions. *Chem. Rev.* **1991**, *91*, 767–792.

[242] Reinhold, J. *Quantentheorie der Moleküle*; Teubner Studienbücher Chemie, 2004.

[243] Levine, I. N. *Quantum Chemistry*; Prentice Hall, 1999.

[244] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover Publications, 1996.

[245] Koopmans, T. C. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* **1933**, *1*, 104.

[246] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

[247] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

[248] Janak, J. F. Proof that $\frac{\partial E}{\partial n_i} = \epsilon_i$ in density-functional theory. *Phys. Rev. B* **1978**, *18*, 7165–7168.

[249] Politzer, P.; Abu-Awwad, F. A comparative analysis of Hartree-Fock and Kohn-Sham orbital energies. *Theoret. Chem. Acc.* **1998**, *99*, 83–87.

[250] Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH, 2001.

[251] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

[252] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[253] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

[254] Colle, R.; Salvetti, O. Approximate calculation of the correlation energy for the closed shells. *Theoretica Chimica Acta* **1975**, *37*, 329–334.

[255] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[256] Harris, J. Adiabatic-connection approach to Kohn-Sham theory. *Phys. Rev. A* **1984**, *29*, 1648–1659.

[257] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into current limitations of density functional theory. *Science* **2008**, *321*, 792–794.

[258] Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **1981**, *23*, 5048–5079.

[259] Mantz, Y. A.; Gervasio, F. L.; Laino, T.; Parrinello, M. Charge localization in stacked radical cation DNA base pairs and the benzene dimer studied by self-interaction corrected density-functional theory. *J. Phys. Chem. A* **2007**, *111*, 105–112.

[260] Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density functional-theory based treatment. *J. Chem. Phys.* **2001**, *114*, 5149–5155.

[261] Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

[262] Gerber, I. C.; Ángyán, J. G. Hybrid functional with separated range. *Chem. Phys. Lett.* **2005**, *415*, 100–105.

[263] Toulouse, J.; Colonna, F.; Savin, A. Long-range - Short-range separation of the electron-electron interaction in density-functional theory. *Phys. Rev. A* **2004**, *70*, 062505–1–062505–16.

[264] Godby, R. W.; Schlüter, M.; Sham, L. J. Self-energy operators and exchange-correlation potentials in semiconductors. *Phys. Rev. B* **1988**, *37*, 10159–10175.

[265] Sham, L. J.; Schlüter, M. Density-Functional Theory of the Energy Gap. *Phys. Rev. Lett.* **1983**, *51*, 1888–1891.

[266] Sham, L. J.; Schlüter, M. Density-functional theory of the band gap. *Phys. Rev. B* **1985**, *32*, 3883–3889.

[267] Perdew, J. P.; Levy, M. Physical content of the exact kohn-sham orbital energies: Band gaps and derivative discontinuities. *Phys. Rev. Lett.* **1983**, *51*, 1884–1887.

[268] Hedin, L.; Lundqvist, B. I. Explicit local exchange-correlation potentials. *J. Phys. C* **1971**, *4*, 2064–2083.

[269] Von Barth, U.; Hedin, L. A local exchange-correlation potential for the spin polarized case. I. *J. Phys. C* **1972**, *5*, 1629–1642.

[270] Hybertsen, M. S.; Louie, S. G. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys. Rev. B* **1986**, *34*, 5390–5413.

[271] Elstner, M. *Semi-empirical DFT: SCC-DFTB*, University Lecture, KIT, 2010.

[272] Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B* **1995**, *51*, 12947–12957.

[273] Elstner, M. The SCC-DFTB method and its application to biological systems. *Theoret. Chem. Acc.* **2006**, *116*, 316–325.

[274] Slater, J. C.; Koster, G. F. Simplified LCAO Method for the Periodic Potential Problem. *Phys. Rev.* **1954**, *94*, 1498–1524.

[275] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.

[276] Elstner, M. Ph.D. thesis, Universität-Gesamthochschule Paderborn, 1998.

[277] Seifert, G. Tight-Binding Density Functional Theory: An Approximate Kohn-Sham DFT Scheme. *J. Phys. Chem. A* **2007**, *111*, 5609–5613.

[278] Elstner, M. SCC-DFTB: What is the proper degree of self-consistency. *J. Phys. Chem. A* **2007**, *111*, 5614–5621.

[279] Mulliken, R. S. Electronic Population Analysis on LCAO[Single Bond]MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833–1840.

[280] Hubbard, J. Electron Correlations in Narrow Energy Bands. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1963**, *276*, 238–257.

[281] Ohno, K. Some remarks on the Pariser-Parr-Pople method. *Theoret. Chem. Acc.* **1964**, *2*, 219–227.

[282] Klopman, G. A Semiempirical Treatment of Molecular Structures. III. Equipotential Orbitals for Polyatomic Systems. *J. Am. Chem. Soc.* **1965**, *87*, 3300.

[283] Hoffmann, M. Ph.D. thesis, Department Physik der Fakultät fur Naturwissenschaften Universität Paderborn, 2006.

[284] Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules. *physica status solidi (b)* **2000**, *217*, 357–376.

[285] Otte, N.; Scholten, M.; Thiel, W. Looking at self-consistent-charge density functional tight binding from a semiempirical perspective. *J. Phys. Chem. A* **2007**, *111*, 5751–5755.

[286] Krüger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. Validation of the density-functional based tight-binding approximation method for the calculation of reaction energies and other data. *J. Chem. Phys.* **2005**, *122*, 114110.

[287] Senn, H. M.; Thiel, W. QM/MM studies of enzymes. *Curr. Op. Chem. Biol.* **2007**, *11*, 182 – 187.

[288] Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* **2001**, *105*, 569–585.

[289] Conwell, E. Charge transport in DNA in solution: The role of polarons. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8795–8799.

[290] Parr, R. G.; Yang, W. *Density-functional theory of atoms and molecules*; Oxford University Press: New York and Oxford England, 1989.

[291] Félix, M.; Voityuk, A. A. DFT performance for the hole transfer parameters in DNA π-stacks. *International Journal of Quantum Chemistry* **2011**, *111*, 191–201.

[292] Niehaus, T. A.; Suhai, S.; Della Sala, F.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, T. Tight-binding approach to time-dependent density-functional response theory. *Phys. Rev. B* **2001**, *63*, 851081–851089.

[293] Brankin, R. W.; Gladwell, I.; Shampine, L. F. *RKSUITE: a suite of Runge-Kutta codes for the initial value problem for ODEs.*; Technical Report, 1992.

[294] Frisch, M. J. et al. *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford, CT, 2004.

[295] Voityuk, A. A.; Jortner, J.; Bixon, M.; Rösch, N. Energetics of hole transfer in DNA. *Chem. Phys. Lett.* **2000**, *324*, 430–434.

[296] Hush, N. S.; Cheung, A. S. Ionization potentials and donor properties of nucleic acid bases and related compounds. *Chem. Phys. Lett.* **1975**, *34*, 11–13.

[297] Orlov, V.; Smirnov, A.; Varshavsky, Y. Ionization potentials and electron-donor ability of nucleic acid babes and their analogues. *Tetrahedron Letters* **1976**, *17*, 4377–4378.

[298] Voityuk, A. A. Estimation of electronic coupling in π-stacked donor-bridge-acceptor systems: Correction of the two-state model. *J. Chem. Phys.* **2006**, *124*, 064505.

[299] Olson, W. et al. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **2001**, *313*, 229–237.

[300] Lu, X. .; El Hassan, M. A.; Hunter, C. A. Structure and conformation of helical nucleic acids: Rebuilding program (SCHNArP). *J. Mol. Biol.* **1997**, *273*, 681–691.

[301] Calladine, C. R.; Drew, H. R. *Understanding DNA; The Molecule & How it Works*; Academic Press, London, 1992.

[302] Lu, X. .; Olson, W. K. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucl. Acids Res.* **2003**, *31*, 5108–5121.

[303] Liu, H.; Gao, J.; Maynard, L.; Saito, Y.; Kool, E. Toward a New Genetic System with Expanded Dimensions: Size-Expanded Analogues of Deoxyadenosine and Thymidine. *J. Am. Chem. Soc.* **2004**, *126*, 1102–1109.

[304] Liu, H.; Lynch, S.; Kool, E. Solution structure of xDNA: A paired genetic helix with increased diameter. *J. Am. Chem. Soc.* **2004**, *126*, 6900–6905.

[305] Gao, J.; Liu, H.; Kool, E. T. Assembly of the complete eight-base artificial genetic helix, xDNA, and its interaction with the natural genetic system. *Angew. Chem. Int. Ed.* **2005**, *44*, 3118–3122.

[306] Migliore, A.; Corni, S.; Varsano, D.; Klein, M.; DiFelice, R. First principles effective electronic couplings for hole transfer in natural and size-expanded DNA. *J. Phys. Chem. B* **2009**, *113*, 9402–9415.

[307] Fuentes-Cabrera, Z., M. abd Xiongce; Kent, P.; Sumpter, B. Electronic structure of xDNA. *J. Phys. Chem. B* **2007**, *111*, 9057–9061.

[308] http://structure.usc.edu/make-na/server.html.

[309] Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2179–2183.

[310] Anderson, P. W. Absence of diffusion in certain random lattices. *Phys. Rev.* **1958**, *109*, 1492–1505.

[311] Büttiker, M.; Landauer, R. Traversal Time for Tunneling. *Phys. Rev. Lett.* **1982**, *49*, 1739–1742.

[312] Peskin, U.; Edlund, A.; Bar-On, I.; Galperin, M.; Nitzan, A. Transient resonance structures in electron tunneling through water. *J. Chem. Phys.* **1999**, *111*, 7558–7566.

[313] Troisi, A.; Nitzan, A.; Ratner, M. A. A rate constant expression for charge transfer through fluctuating bridges. *J. Chem. Phys.* **2003**, *119*, 5782–5788.

[314] Laughlan, G.; Murchie, A.; Norman, D.; Moore, M.; Moody, P.; Lilley, D.; Luisi, B. The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* **1994**, *265*, 520–524.

[315] Fadrná, E.; Špačková, N.; Štefl, R.; Koča, J.; Cheatham III, T.; Šponer, J. Molecular dynamics simulations of guanine quadruplex loops: Advances and force field limitations. *Biophys. J.* **2004**, *87*, 227–242.

[316] Ndlebe, T.; Schuster, G. Long-distance radical cation transport in DNA: Horizontal charge hopping in a dimeric quadruplex. *Org. Biomol. Chem.* **2006**, *4*, 4015–4021.

[317] Reha, D.; Barford, W.; Harris, S. A multi-scale method for the calculation of charge transfer rates through the Π-stack of DNA: application to DNA dynamics. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5436–5444.

[318] Steinbrecher, T.; Koslowski, T.; Case, D. Direct simulation of electron transfer reactions in DNA radical cations. *J. Phys. Chem. B* **2008**, *112*, 16935–16944.

[319] Gutiérrez, R.; Caetano, R.; Woiczikowski, P. B.; Kubař, T.; Elstner, M.; Cuniberti, G. Structural fluctuations and quantum transport through DNA molecular wires: a combined molecular dynamics and model Hamiltonian approach. *New J. Phys.* **2010**, *12*, 023022.

[320] Strümpfer, J.; Schulten, K. The effect of correlated bath fluctuations on exciton transfer. *J. Chem. Phys.* **2011**, *134*, 095102.

[321] Barnett, R. N.; Cleveland, C. L.; Joy, A.; Landman, U.; Schuster, G. B. Charge Migration in DNA: Ion-Gated Transport. *Science* **2001**, *294*, 567.

[322] Kubař, T.; Kleinekathofer, U.; Elstner, M. Solvent Fluctuations Drive the Hole Transfer in DNA: A Mixed Quantum-Classical Study. *J. Phys. Chem. B* **2009**, *113*, 13107–13117.

[323] Guo, X.; Gorodetsky, A. A.; Hone, J.; Barton, J. K.; Nuckolls, C. Conductivity of a single DNA duplex bridging a carbon nanotube gap. *Nature Nanotechnology* **2008**, *3*, 163–167.

[324] Shapir, E.; Sagiv, L.; Molotsky, T.; Kotlyar, A. B.; Felice, R. D.; Porath, D. Electronic Structure of G4-DNA by Scanning Tunneling Spectroscopy. *J. Phys. Chem. C* **2010**, *114*, 22079–22084.

[325] Meagher, K.; Redman, L.; Carlson, H. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **2003**, *24*, 1016–1025.

[326] Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[327] Hwang, J. K.; Warshel, A. Microscopic Examination of Free-Energy Relationships for Electron-Transfer in Polar Solvents. *J. Am. Chem. Soc.* **1987**, *109*, 715–720.

[328] Tachiya, M. Generalization of the Marcus Equation for the Electron-Transfer Rate. *J. Phys. Chem.* **1993**, *97*, 5911–5916.

[329] Tachiya, M. Reaction coordinate in electron transfer: What physical quantity should we use for it? *J. Chem. Phys.* **2008**, *129*, 066102.

[330] Popovic, D.; Zmiric, A.; Zaric, S.; Knapp, E. Energetics of Radical Transfer in DNA Photolyase. *J. Am. Chem. Soc.* **2002**, *124*, 3775–3782.

[331] Li, Y.; Heelis, P.; Sancar, A. Active-site of DNA photolyase - tryptophan-306 is the intrinsic hydrogen-atom donor essential for flavin radical photoreduction and DNA-repair in vitro. *Biochemistry* **1991**, *30*, 6322–6329.

List of Publications

1. T. Kubař, P. B. Woiczikowski, G. Cuniberti, and M. Elstner; "Efficient calculation of charge-transfer matrix elements for hole transfer in DNA"; *Journal of Physical Chemistry B*; **112**(26); pp. 7937-7947 (2008).

2. R. Gutiérrez, R. A. Caetano, P. B. Woiczikowski, T. Kubar, M. Elstner, and G. Cuniberti; "Charge transport through biomolecular wires in a solvent: Bridging molecular dynamics and model hamiltonian approaches"; *Physical Review Letters*; **102**(20); pp. 208102 (2009).

3. P. B. Woiczikowski, T. Kubař, R. Gutiérrez, R. A. Caetano, G. Cuniberti, and M. Elstner; "Combined density functional theory and Landauer approach for hole transfer in DNA along classical molecular dynamics trajectories"; *Journal of Chemical Physics*; **130**(21); pp. 215104 (2009).

4. R. Gutiérrez, R. A. Caetano, P. B. Woiczikowski, T. Kubař, M. Elstner, and G. Cuniberti; "Structural fluctuations and quantum transport through DNA molecular wires: a combined molecular dynamics and model Hamiltonian approach"; *New Journal of Physics*; **12**(2); pp. 023022 (2010).

5. P. B. Woiczikowski, T. Kubař, R. Gutiérrez, G. Cuniberti, and M. Elstner; "Structural stability versus conformational sampling in biomolecular systems: Why is the charge transfer efficiency in G4-DNA better than in double-stranded DNA?"; *Journal of Chemical Physics*; **133**(3); pp. 035103 (2010).

6. P. B. Woiczikowski, T. Steinbrecher, T. Kubař, and M. Elstner; "Nonadiabatic QM/MM Simulations of Fast Charge Transfer in Escherichia coli DNA Photolyase"; *Journal of Physical Chemistry B*; http://dx.doi.org/10.1021/jp204696t (2011).

7. M. Wolter, P. B. Woiczikowski, M. Elstner, and T. Kubař; "How the conductivity of DNA responds to stretching"; *submitted for publication* (2011).

8. A. Heck, P. B. Woiczikowski, T. Kubař, M. Elstner, and T. Steinbrecher; "Charge Transfer in Model Peptides – Obtaining Marcus' Parameters from Molecular Simulation"; *manuscript in preparation* (2011).

Danksagung

An dieser Stelle möchte ich all denen danken, die direkt oder indirekt zum Gelingen dieser Arbeit beigetragen haben:

- Prof. Dr. Marcus Elstner, für die Betreuung meiner Promotion und die wissenschaftliche Ausbildung, insbesondere der Teilnahmen an internationalen Workshops. Aber vor allem auch, für die fortwährende Motivation und das scheinbar unerschöpfliche Maß an kreativen Ideen und Denkansätzen, die nicht nur die meinige Forschungsarbeit immer weiter voran getrieben haben.

- Dr. Tomáš Kubař, ohne dessen ambitionierte wissenschaftliche Arbeit und Engagement die hier vorgestellten Projekte und damit diese Arbeit nicht möglich gewesen wäre. Er und Prof. Dr. Marcus Elstner haben unter Anderem die spannende Methodik der selbst-konsistenten, nicht-adiabatischen Ladungstransfer-Dynamik entwickelt, die im Kapitel 8 dieser Arbeit angewendet wurde.

- Dr. Thomas Steinbrecher, der das Photolyase Projekt nicht nur initiiert, sondern auch entscheidend daran mitgewirkt hat.

- Der Arbeitsgruppe um Prof. Dr. Gianaurelio Cuniberti an der TU Dresden, insbesondere Dr. Rafael Gutiérrez und Dr. Rodrigo Caetano (jetzt Universidade Federal de Alagoas, Brasilien), für die sehr angenehme Zusammenarbeit, die unermüdliche und kompetente Unterstützung bei den physikalischen Transportrechnungen, sowie für den stets sehr unterhaltsamen Dialog zwischen Chemie und Physik.

- Der gesamten Arbeitsgruppe, für das stets sehr entspannte und freundschaftliche Arbeitsklima, sowohl in Karlsruhe (Thomas, Tomáš, Kai, Tino, Mario, Alex, Steve und Hiroshi) als auch in Braunschweig (Jan, Prasad, Albrecht und Michi).

- Den Korrekturlesern Kai Welke und Dr. Thomas Steinbrecher, für die mühevolle und kritische Durchsicht des Manuskripts.

Besonderer Dank gebührt meiner Familie, insbesondere meinen Eltern, die mich in jedweder Lebenslage unterstützt und an mich geglaubt haben.