Institut für Physikalische und Theoretische Chemie
der Technischen Universität München

A Combined Quantum Mechanics and Molecular Dynamics Study of Charge Transfer in DNA

Khatcharin Siriwong

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende: Univ.-Prof. Dr. Klaus Köhler

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Notker Rösch
2. Univ.-Prof. Dr. Sevil Weinkauf

Die Dissertation wurde am 27.4.2004 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 16.6.2004 angenommen.

dedicated to my parents and Chomsri

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to Prof. Dr. Notker Rösch for giving me the opportunity to join his group and do research in an exciting field, as well as, for advising, understanding, encouraging, and especially for teaching me not only how to be a good researcher but also how to be a good teacher.

My very special thanks go to Dr. Alexander Voityuk who helped me whenever I had doubts about my work. This thesis was significantly enriched by his valuable discussion and advice.

I would like to thank Prof. Marshall D. Newton, Brookhaven National Laboratory, as collaborator in some part of this thesis; Dr. Sven Krüger and Dr. Konstantin Neyman for making my time in Germany more convenient; Dr. Alexei Matveev for helping with all kinds of computer problems; Alexander Genest and Florian Schlosser for translating all German documents.

I thank other Rösch's group members for the great working atmosphere and for being not only colleagues but also friends: C. Inntam, P. Chuichay, Dr. C. Bussai, D. Gayushin, Dr. M. Fuchs-Rohr, Dr. A. Woiterski, M. Suzen, M. Girju, D. Dogaru, Dr. W. Alsheimer, Dr. M. Garcia-Hernandez, Dr. L. Moskaleva, K.-H. Lim, Dr. Z. Chen, Dr. S. Majumder, Dr. R. Deka, A. Deka, Dr. V. Nasluzov and S. Bosko.

I also would like to thank those who agreed to be the referees of this thesis and allocated their valuable time in order to evaluate the quality of the work. The financial support of Deutsche Forschungsgemeinschaft, Volkswagen Foundation, and Fonds der Chemischen Industrie are gratefully acknowledged.

My absolute acknowledgement is dedicated to my parents, my wife Chomsri, and my brothers and sisters for their inspiration and encouragement throughout the entire study. They will most certainly be glad to know that I am finally finishing my education and starting life in the wide world.

Contents

List of Abbreviations and Symbols

8-oxoG	7,8-dihydro-8-oxoguanine
Å	Ångstrom
A	adenine
a	acceptor
arb.	arbitrary
avg.	average
B	base
B3LYP	hybrid Becke-3-parameter exchange and Lee-Yang-Parr correlation approximation
C	cytosine
cm	centimeter
CNDO	complete neglect of differential overlap
CT	charge transfer
d	donor
DFT	density functional theory
DNA	deoxyribonucleic acid
E	energy
e.g.	for example
Eq.	equation
ET	electron transfer
et al.	and others
etc.	and so forth
eV	electron Volt
FCD	fragment charge difference
FDPB	finite difference Poisson–Boltzmann
fs	femtosecond
G	guanine
GB	generalized Born
GMH	generalized Mulliken–Hush
HF	Hartree–Fock

HOMO	highest occupied molecular orbital
HT	hole transfer
i.e.	that is
IP	ionization potential
IR	infrared
K	Kelvin
k_B	Boltzmann constant
kcal/mol	kilocalories per mole
k_{CT}	rate constant of charge transfer
M	molar
MD	molecular dynamics
NDDO	neglect of diatomic differential overlap
NDDO-G	special parameterization of the NDDO method
NMR	nuclear magnetic resonance
ns	nanosecond
OxP	oxidation potential
P	product
PB	Poisson–Boltzmann
PME	particle mesh Ewald
ps	picosecond
QM	quantum mechanics
R	reactant
r	distance
R_{da}	donor–acceptor distance
RDF	radial distribution function
Ref.	reference
RMSD	root mean square deviation
s	second
S. D.	standard deviation
SCF	self-consistent filed
SFCD	simplified fragment charge difference
T	thymine
T	temperature
t	time

V_{da}	electronic coupling between donor and acceptor states
ΔG^{o}	driving force
$\Delta G^{\dagger}$	activation free energy change
β	decay parameter of the rate constant
β_{el}	decay parameter due to electronic contributions
β_{s}	decay parameter due to solvent contributions
ε	dielectric constant
λ	reorganization energy
λ_{i}	intramolecular reorganization energy
λ_{s}	solvent reorganization energy

Chapter 1

Introduction

1.1 Why is charge transfer in DNA important?

Deoxyribonucleic acid (DNA) is well known to play a pivotal role in biology as the carrier of genetic information. However, during recent years scientists have been strongly interested in the electronic properties of this "molecule of life". The DNA structure is suitable for electron transfer because of the overlap of orbitals belonging to the nucleobases along the DNA axis. Soon after Watson and Crick discovered the double-helix structure of DNA in 1953,[1] Eley and Spivey were the first to suggest that DNA could serve as an electronic conductor.[2] Over 40 years, both experimentalists[3–6] and theoreticians[7–14] have become increasingly interested in the electron transfer in DNA for a number of reasons, including in particular: (1) understanding the mechanisms of oxidative damage and photorepair of DNA,[15,16] and (2) the potential role of DNA in the design of nanoelectronic devices.[17–19]

Oxidative damage of DNA is caused mainly by radiation and reactive oxidizing agents, such as superoxide ($O_2^{\bullet -}$) and hydroxyl ($HO^{\bullet}$), formed constantly in the human body.[20,21] The chemical events associated with this DNA damage start with the oxidation of guanine base (G) which has the lowest oxidation potentials of all the nucleobases; the others are adenine (A), cytosine (C), and thymine (T).[16,22–25] The loss of an electron from a guanine unit can be the start of a serious problem, because a guanine radical cation $G^{\bullet +}$ reacts with water yielding oxidized guanine (8-oxoG).[20] During the continuous process of DNA replication as it occurs within the cells, 8-oxoG has an increased tendency to bind A

instead of C, leading to mutations. Although cells have developed various mechanisms to detect and repair oxidative damage, these surveillance systems cannot pick up everything, and mutations caused by oxidizing agents are among the leading causes of cancer, genetic disease and cell death. In 1996, Barton and co-workers confirmed that oxidative damage in DNA really can occur at a distance.[26] By using a specially designed oxidizing agent to pluck electron from a specified guanine, the researchers showed that damage could occur up to 11 base pairs from the original attack. There was only one explanation, namely electron holes were migrating along the DNA strand.[26]

As DNA had been proposed to be a conductor,[2] it could be used to design electronic devices. One expects that molecular devices would be able to perform all the operations currently done by silicon transistors. The basic idea of molecular electronics is to use individual molecules as wires, switches, rectifiers, and memories.[27–29] The great advantage of molecular electronics in the frame of the continued device miniaturization is the intrinsic nanoscale size of molecular building blocks. However, so far experiments of different groups have provided results on DNA conductivity which are in part contradictory.[30] Some experimentalists reported that there is a very weak distance dependence for charge transfer through DNA.[31] This supports the idea that DNA acts as a molecular wire by conducting electrons via the π-stack. In contrast, other results indicated a strong distance dependence of charge transfer, which would suggest that DNA is effectively an insulator.[32]

The ambiguity of these results was partly due to the variety of experimental conditions and DNA sequences. In particular, Giese and co-workers showed that long-range charge transport through DNA is sequence dependent[5] (more details will be discussed in Chapter 2). To understand why the DNA sequence makes a difference, we need to compare the relative energies of DNA bases. These energies can be deduced from computational models and experiments.[16,22–25] It was found that G base has a lower oxidation energy than the other three bases. Additionally, GG units have an even lower oxidation energy than single G, and GGG units have a still lower oxidation energy. The important feature is that a hole (i.e. positive charge) is more stable on a G·C pair than on an A·T pair. Thus, an electron hole will localize on G·C base pairs. A hole can transfer to another G·C site, and can then either hop back to the first G·C site or move on to the next one. Because holes on A·T base pairs have a higher energy, such A·T units act as a barrier

to hole transfer. A hole will migrate along a DNA strand until it hits a GG or GGG site. Therefore, for DNA containing G-rich sequences, transport hole from one end of the nanomolecule to another should be feasible.

Of course, without understanding of different other factors which control DNA-mediated electron transfer, it would be impossible to create DNA based electronic devices, as well as to develop novel diagnostic tools for screening and analyzing nucleic acids.

1.2 Quantum mechanics/molecular dynamics study

Computational modeling of charge transfer within DNA at the atomic level is a complicate task. One has to determine the structure of the system and its thermal fluctuations in solution, as well as the relative energies of various states of localized electrons or holes.

Quantum mechanics (QM) methods have been adopted to elucidate the different factors that control charge transfer in DNA using models of DNA fragments with fixed geometries.[10,11,33–36] These studies highlight the strong dependence of the charge transfer parameters on the structure of a DNA duplex and therefore suggest that the motion of the bases may be crucial for a proper description of the mechanism of charge transfer. In fact, several experimental studies on oligonucleotides[37–39] indicated that structural distortion may play a major role in the charge transfer. Furthermore, both experimental[40,41] and theoretical[42] results demonstrated that the solvent can substantially affect the rate of charge transfer, and thus, it is necessary to include explicitly the solvent medium surrounding a DNA oligomer. Thus, for a computational investigation on the charge transport in DNA that accounts for the effects of structural fluctuations of DNA and for the motion of the solvent, a method is indispensable which properly describes the dynamics of the stacks of nucleobases within a duplex.

Molecular dynamics (MD) simulations based on empirical force field are well established tools for studying thermal fluctuations of the structure of DNA and its environment.[43,44] In addition, experimental results showed that an electron hole in a π-stack of nucleobases has a life time of tens to hundreds of picoseconds.[45] This is long enough for

the local environment of the charged state in a duplex to undergo DNA structural relaxation.

In the investigations on the charge transfer in DNA reported in this dissertation, we employed QM methods to calculate the factors controlling the charge transfer and MD simulations to quantify the effects of thermal fluctuation on these factors. Thus, this strategy is called a hybrid *QM/MD* approach.

Chapter 2

Charge Transfer Theory in DNA Double Helix

2.1 Principle of DNA structure

The three dimensional structure of the DNA double helix was correctly elucidated by James Watson and Francis Crick half a century ago.[1] DNA is a polymer, whose units are nucleotides, and thus the polymer is known as a "polynucleotide". Each nucleotide consists of a deoxyribose sugar, a phosphate group, and a nucleobase. Four kinds of nucleobases are found in DNA: two purines, adenine (A) and guanine (G), and two pyrimidines, thymine (T) and cytosine (C).

The central feature of DNA structure is[46,47] a sugar-phosphate backbone (a chain of deoxyribose linked by phosphodiester bridges, Figure 2.1), which winds around the outside, whereas the purine and pyrimidine bases are on the inside of the helix (Figure 2.1); the strands run in opposite directions. Each nucleobase is attached to the C1′ atom of a sugar unit, Figure 2.1; sugar atoms are numbered with a prime (′) to distinguish from the numbering of the nucleobase atoms. The two chains are held together by hydrogen bonds formed between pairs of bases. Pairing is highly specific, namely A pairs to T with two H-bonds and G pairs to C with three H-bonds (Figure 2.2). Base pairs are roughly perpendicular to the helical axis, and are stacked on top of each other by van der Waals contact. The sequence of bases along the polynucleotide chain is not restricted and the precise sequence of bases carries the genetic information. The twist of the backbone defines two grooves in the helix (major and minor grooves), see Figure 2.1.

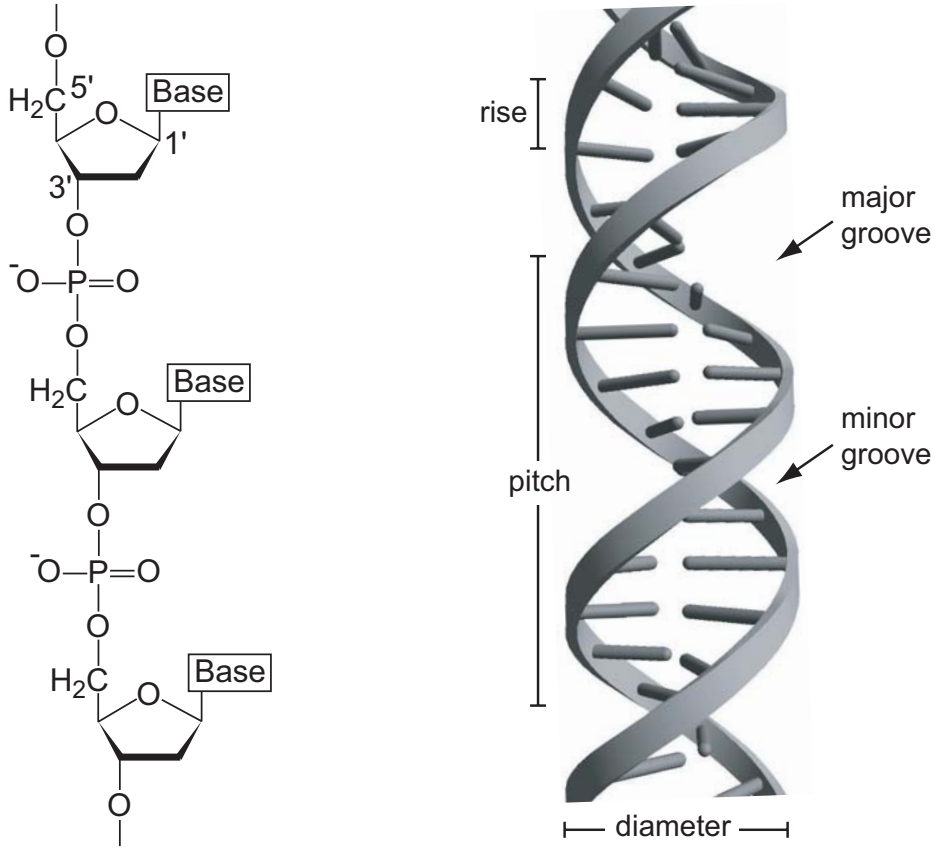

Figure 2.1 Structure of part of a DNA strand (left), and DNA double helix (right).

Table 2.1 Average structural parameters for various DNA forms.[46,47]

Parameter	A-DNA	B-DNA	Z-DNA
Orientation	Right-handed	Right-handed	Left-handed
Helix diameter (Å)	26	20	18
Rise (Å)	2.56	3.38	3.70
Pitch (Å)	28.2	33.8	44.5
Base pairs/turn	11	10	12
Helix twist (°)	32.7	36.0	–30.0
Major groove width[a] (Å)	2.7	11.7	2.0
Minor groove width[a] (Å)	11.0	5.7	8.8

[a] Groove width is the perpendicular separation of helix strands drawn through phosphate groups, added by 5.8 Å to account for van der Waals radii of phosphate groups.

Generally, there are three forms of DNA (A, B and Z), which are classified by DNA geometric parameters, such as orientation of the double helix (right- or left-handed), inter-strand phosphate distance (groove width), diameter of the helix, distance between two adjacent base pairs (rise), rotation angle per base pair (twist), length of the helix per turn, and number of base pairs per turn (pitch). These parameters are summarized in Table 2.1.

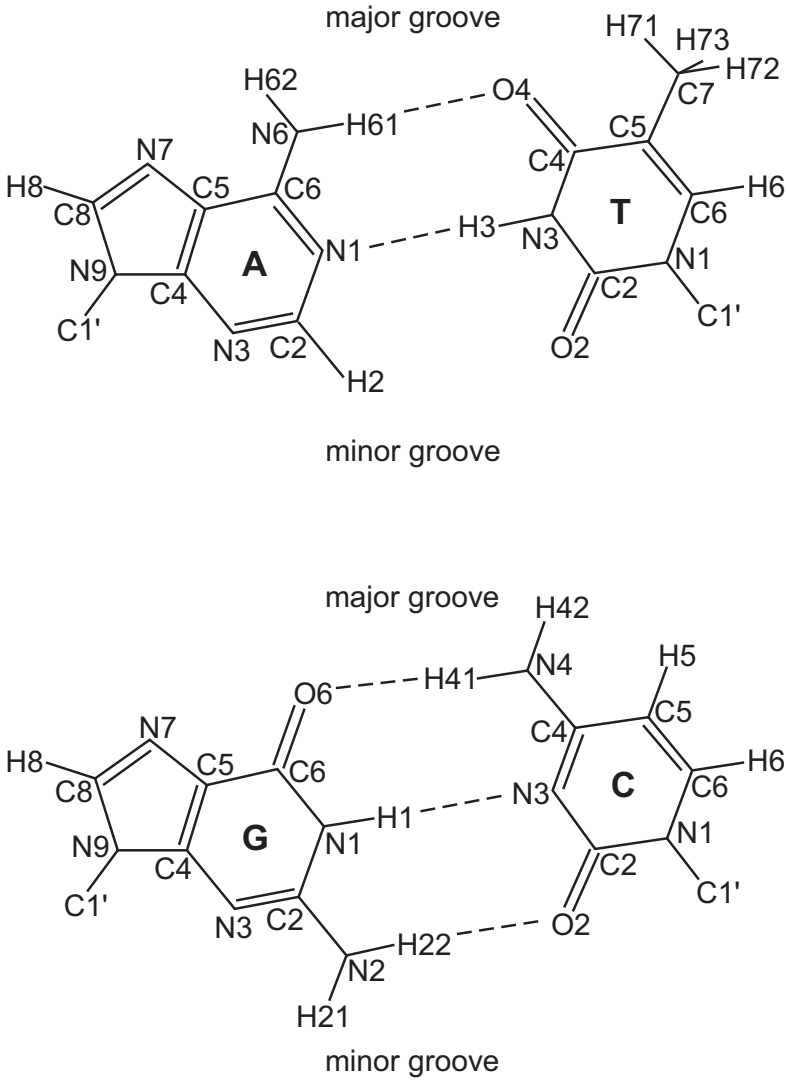

Figure 2.2 Watson–Crick base pairs (A·T and G·C) found in double stranded DNA. The DNA grooves are indicated; the minor groove is on the side where the sugars are attached.

2.2 Basic charge transfer theory

In this section, the concepts and theoretical fundamentals of electron transfer (ET) are introduced. Electron transfer comprises the transition of a single electron from one state, the electron donor, to another state, the electron acceptor. Such a process is usually referred to as *hole transfer* (HT) when one wants to focus on the inverse process, i.e. the migration of a missing electron (a hole or positive charge) from a hole donor (electron acceptor) to a hole acceptor (electron donor). In the following, we will refer to both electron donor/acceptor and hole donor/acceptor as *charge transfer* (CT).

When considering an electron transfer reaction, we have to deal with (at least) two different electronic states: one characterizing the reactants R (before electron transfer), and one characterizing the products P (after electron transfer). In a complex system, the electronic states have their potential energy surface in many-dimensional atomic configuration space. These two surfaces intersect each other. The intersection surface can be reached by any suitable fluctuation of atomic coordinates. The electronic interaction between reactant and product leads to the usual splitting of states (surfaces); such a cross section of potential energy curves is sketched in Figure 2.3. One can classify electron transfer reactions according to the magnitude of this splitting. A reaction is called *diabatic* if the two electronic states do not interact, hence no electron transfer reaction occurs (dashed lines in Figure 2.3). For sufficient electronic interaction resulting in a large spitting, the reaction is called *adiabatic.* In this case, a system passes through the intersection during a fluctuation of atom configuration and will always stay on the lowest energy hypersurface (Figure 2.3). When the electronic interaction is weak (splitting $< k_B T$), the reaction is neither diabatic nor adiabatic. This intermediate type is called *non-adiabatic.* In this case, the system tends to retain its original electronic configuration when passing across the intersection.

In 1950s and 1960s, Rudolph A. Marcus, who received the Nobel Prize in Chemistry in 1992, derived a theory for outer-sphere electron transfer.[48–51] The classical Marcus theory provides a quantitative way to describe a certain type of electron transfer reactions; it has been applied to several biological charge transfer systems.[52,53] In this theory, the rate constant of charge transfer, k_{CT}, from a donor d to an acceptor a for the non-adiabatic limit is expressed as[54]

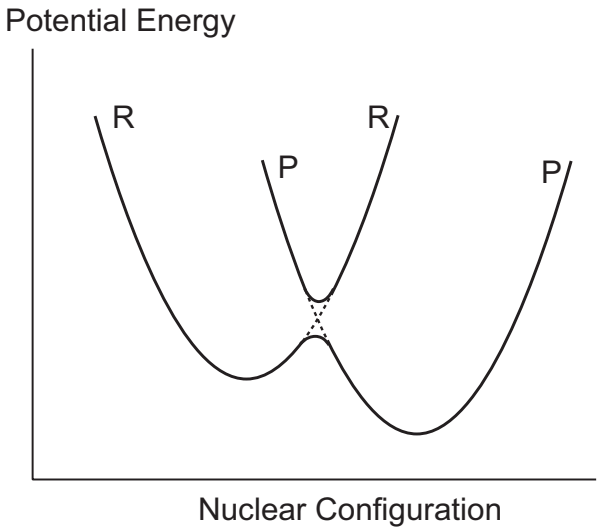

Figure 2.3 Potential energy surfaces of an electron transfer reaction plotted against a generalized reaction coordinate of the entire system. Curves R and P denote reactants and products, respectively. Dashed lines show the intersection of the surfaces (diabatic case) and solid lines indicate the splitting in the adiabatic limit.

$$k_{\mathrm{CT}} = \frac{2\pi}{\hbar\sqrt{4\pi\lambda k_{\mathrm{B}}T}} V_{da}^{2} \exp\left[-(\Delta G^{\mathrm{o}} + \lambda)^{2} / 4\lambda k_{\mathrm{B}}T\right] \tag{2.1}$$

where V_{da} is the electronic coupling matrix element between redox centers and describes the degree to which wave functions of donor and acceptor sites overlap. k_{B} is the Boltzmann constant, T is the temperature, ΔG^{o} is reaction free energy or driving force, and λ is the so-called reorganization energy. According to Figure 2.4, the electronic coupling is one-half of the splitting at the intersection, the reorganization energy is defined as the energy of the reactants at the equilibrium nuclear configuration of the products, and the driving force is the energy difference of the energy minima of the two states.

The activation free energy change $\Delta G^{\dagger}$ is related to the parameters ΔG^{o} and λ as

$$\Delta G^{\dagger} = \frac{(\Delta G^{\mathrm{o}} + \lambda)^{2}}{4\lambda} \tag{2.2}$$

$\Delta G^{\dagger}$ equals $\lambda/4$ when donor and acceptor are equivalent, i.e. $\Delta G^{\mathrm{o}} = 0$. More details on these three factors (V_{da}, ΔG^{o} and λ) that control the rate of charge transfer will be described in subsequent sections.

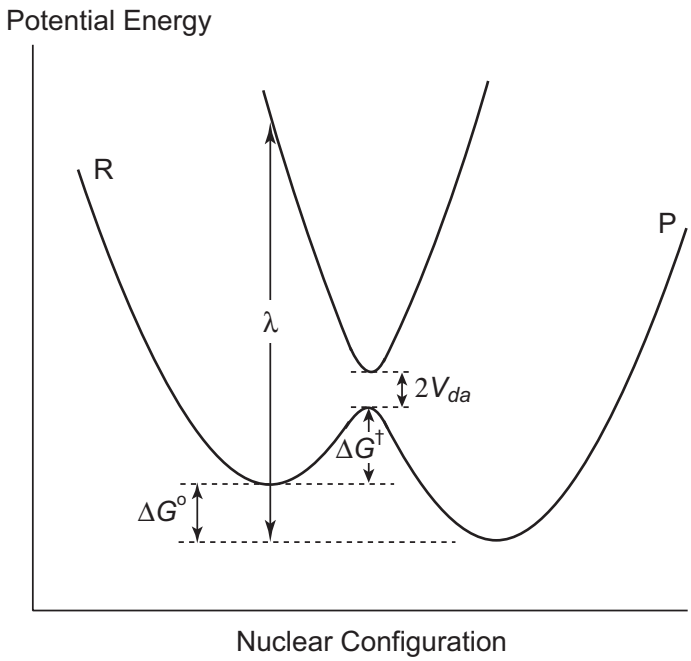

Figure 2.4 Potential energy diagram for an electron transfer reaction in the non-adiabatic regime. The definitions of the driving force ΔG^{o}, the activation free energy $\Delta G^{\dagger}$ and reorganization energy λ are indicated. Also shown is the splitting at the surface intersection which is twice the electronic coupling V_{da}.

It is an essential feature of Marcus theory that electron transfer takes place only at the cross-over point. According to the Frank–Condon principle, electron transfer occurs so rapidly that no change in nuclear configuration can occur during the transfer. This requires that the transfer is a vertical transition in the diagram. Conservation of energy requires that the transition is a horizontal in the diagram. The only place where both conditions are fulfilled is the crossing point of the potential energy surfaces. The crossing point determines activation barrier $\Delta G^{\dagger}$.

2.2.1 Electronic coupling matrix element

Non-adiabatic charge transfer reactions are characterized by weak electronic interaction between reactants and products at the transition-state nuclear configuration. This coupling is directly related to the strength of the electronic interaction between donor and acceptor.[55]

A number of methods based on quantum chemical calculations have been proposed and applied to obtain estimates of electronic coupling V_{da};[56–61] for a review, see Ref. 62.

Let us consider a system comprising a donor *d*, with wave function ψ_d, and an acceptor *a*, with wave function ψ_a. Within the framework of a two-state model, the splitting between the two surfaces can be obtained by solving the 2×2 secular equation[56]

$$\begin{vmatrix} H_{dd}-E & H_{da}-ES_{da} \\ H_{da}-ES_{da} & H_{aa}-E \end{vmatrix} = 0 \tag{2.3}$$

Here, $H_{da} = \langle \Psi_d | H | \Psi_a \rangle$ and $S_{da} = \langle \Psi_d | \Psi_a \rangle$ are Hamiltonian matrix element and overlap matrix, respectively, and *H* is the total electronic Hamiltonian, excluding nuclear kinetic energy and nuclear repulsion terms. There are two roots E_+ (higher energy) and E_- (lower energy) for the energy eigenvalue *E*:

$$E_{\pm} = \frac{(H_{dd}+H_{aa}-2H_{da}S_{da}) \pm 2\sqrt{\frac{1}{4}(H_{dd}-H_{aa})^2 - (H_{dd}+H_{aa})H_{da}S_{da} + H_{dd}H_{aa}S_{da}^2 + H_{da}^2}}{2(1-S_{da}^2)}$$

Conventionally, one-half of the separation evaluated at the seam of the crossing, where $H_{dd} = H_{aa}$, is designated as electronic coupling V_{da}. With the latter simplification one obtains

$$\begin{aligned} V_{da} &= \frac{1}{2}(E_+ - E_-) \\ &= \frac{H_{da} - S_{da}(H_{dd}+H_{aa})/2}{1-S_{da}^2} \end{aligned} \tag{2.4}$$

In the non-adiabatic limit in which the overlap S_{da} is very small, the electronic coupling is estimated to be equal the Hamiltonian matrix element between donor and acceptor states, i.e. $V_{da} \simeq H_{da}$. However, for the coupling of with π-stacks of DNA, such a simplification is not acceptable.[62]

Alternatively, in case donor and acceptor are equivalent by symmetry, the magnitude of the splitting can be easily approximated as one-half of the energy difference between the two adiabatic states, E_1 and E_2. This approach is referred to as *minimum splitting method*[57]

$$V_{da} = \frac{1}{2}(E_2 - E_1) \tag{2.5}$$

Invoking Koopmans' theorem in the Hartree–Fock method, E_1 and E_2 for hole transfer can be estimated as the Hartree–Fock orbital energies of HOMO (highest occupied molecular orbital) and the subsequently lower-lying orbital HOMO–1, respectively, of the system.[10,58] For systems where donor and acceptor are not equivalent (or off resonance), one has to apply an external perturbation to bring donor and acceptor electronic levels into resonance. However, a disadvantage of this method is that, applying an external electric field for the large system is very time consuming.[62]

Another straightforward method to compute the electronic coupling is the *generalized Mulliken–Hush* (GMH) formalism developed by Cave and Newton.[59,60] These authors showed that the coupling can be estimated from the transition dipole moment between the states of interests and the diabatic states can be defined by diagonalizing the dipole moment matrix. According to the GMH method, the coupling is calculated as

$$V_{da} = \frac{(E_2 - E_1)\left|\mu_{12}\right|}{\sqrt{(\mu_1 - \mu_2)^2 + 4\mu_{12}^2}} \tag{2.6}$$

where μ_{12} is the transition dipole moment between electronic states 1 and 2; μ_1 and μ_2 are the dipole moment of the ground and excited states, respectively. Examples of this approach for calculating the electronic coupling are provided by the studies of Elliott et al.[63] and Rust et al.[64]

The *fragment charge difference* (FCD) procedure to estimate the donor–acceptor coupling has been recently introduced by Voityuk and Rösch and used to study the hole transfer in DNA.[61] In this method, the coupling is expressed as

$$V_{da} = \frac{(E_2 - E_1)\left|\Delta q_{12}\right|}{\sqrt{(\Delta q_1 - \Delta q_2)^2 + 4\Delta q_{12}^2}} \tag{2.7}$$

Here, Δq_1 and Δq_2 are the *d–a* charge difference in the adiabatic states 1 and 2, respectively. $\Delta q_{12} = q_{12}(d) - q_{12}(a)$, where $q_{12}(d)$ and $q_{12}(a)$ are calculated using the transition density $\langle \Psi_1 | \Psi_2 \rangle$. Additionally, they have derived a more straightforward formula for estimating the electronic coupling. This procedure is called *simplified FCD* (SFCD) method and expressed as[61]

$$V_{da} = \frac{1}{2}(E_2 - E_1)\sqrt{1-\Delta q^2} \tag{2.8}$$

where Δq is the difference of the charges on donor and acceptor in the ground state. Notice, when donor and acceptor are in resonance, i.e. their energies are equal ($H_{dd} = H_{aa}$), and $\Delta q_1 = \Delta q_2$ (in Eq. (2.7)) or $\Delta q = 0$ (in Eq. (2.8)), then both FCD and SFCD models reduce to the minimum splitting method, Eq. (2.5).

An advantage of GMH and FCD expressions is that they can be used in general case, where donor and acceptor are in resonance or off-resonance, without applying an external electric field. Both methods allow on to go beyond the two-state model. Nevertheless, when one takes more states into account, the FCD model is more robust than the GMH approach.[61]

2.2.2 Reorganization energy

The reorganization energy λ is the free energy required to deform or repolarize the nuclear modes of donor and acceptor as well as of their environment from the equilibrium configuration of a given electronic state, say the reactant state, to the equilibrium configuration of the product state whilst the initial electronic state is maintained.[65] In other words, this is the change in free energy needed to move the reactants to the product configuration without actually transferring the electron.[66] The reorganization energy includes two components (see Figure 2.5). The first term, inner-sphere or intramolecular reorganization energy, λ_i, is caused by changes in geometries of the donor and acceptor. The second contribution, outer-sphere or solvent reorganization energy, λ_s, results from reorientation of solvent molecules in response to the charge redistribution in the donor and acceptor. For electron transfer reactions in polar solvents, the dominant contribution to λ arises from the solvent. Experimentally, λ_i can be obtained from resonance Raman spectroscopy.[67,68] In turn, λ_s can be obtained from measurements of electron transfer rates as a function of free energy.[45,69] Theoretically, the term λ_s is treated classically[48,54] (older methods rely on a continuum model; modern approaches are based on MD results), whereas the term λ_i can be calculated with a quantum chemical approach.[70,71]

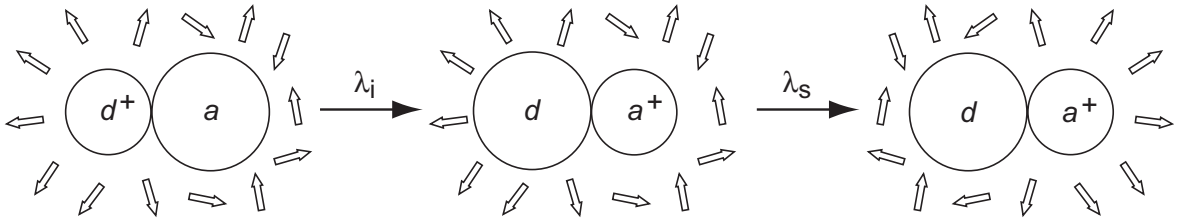

Figure 2.5 Changes in the inner-sphere are indicated by different sizes of the donor *d* and acceptor *a*. Changes in the outer-sphere are indicated by reorientation of the solvent molecules (arrows around *d* and *a*).

In the simple two-sphere model introduced by Marcus for estimating λ_s, the donor *d* and acceptor *a* are treated as conducting spheres of radii r_d and r_a, separated by the distance R_{da}. λ_s is given as[72]

$$\lambda_s = \frac{\Delta q^2}{2}\left(\frac{1}{r_d}+\frac{1}{r_a}-\frac{2}{R_{da}}\right)\left(\frac{1}{\varepsilon^{op}}-\frac{1}{\varepsilon^{st}}\right) \tag{2.9}$$

where Δq is the charge transferred from *d* to *a*, ε^{op} and ε^{st} are optical and static dielectric constants of solvent, respectively.

λ_s can be calculated as a difference of solvation energies, E_{solv}:[73,74]

$$\lambda_s = E_{solv}(\Delta\rho, \varepsilon^{op}) - E_{solv}(\Delta\rho, \varepsilon^{st}) \tag{2.10}$$

where $\Delta\rho$ is the electron density change of the solute upon ET. The electrostatic continuum model has been very successful in estimating the solvation energy.[75] This method is based on the finite difference solution of the Poisson–Boltzmann equation (FDPB)[75,76] for the electrostatic potential in the medium divided into regions with different dielectric constants. The boundary of each region is defined by the solvent accessible surface,[77] which is generated by rolling a sphere of water molecule (typically with a radius of 1.4 Å) along the van der Waals surface of such region. In the FDPB method, the dielectric regions are mapped onto a cubic lattice. Values are assigned at each point for the charge density, dielectric constant, and ionic strength parameters in the Poisson–Boltzmann equation. The electrostatic potential at grid points is computed to estimate the solvation energy.

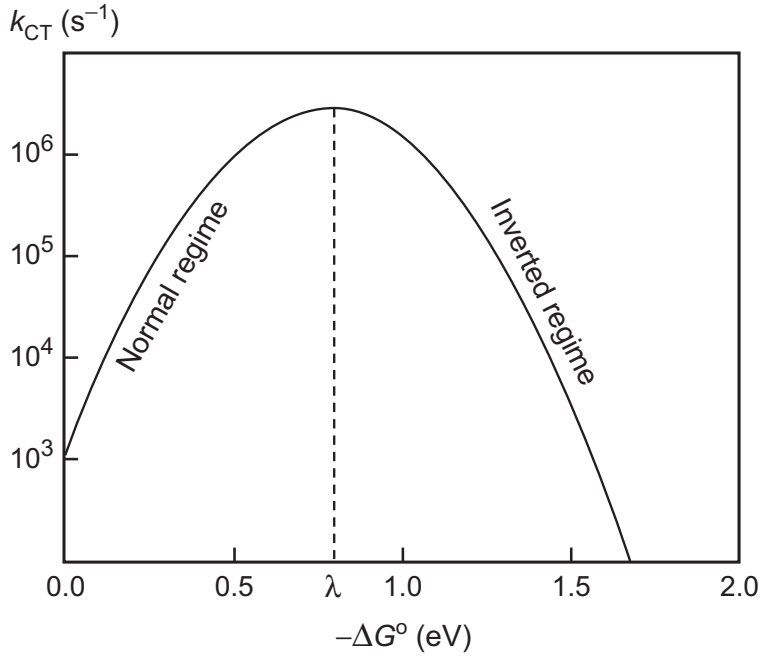

Figure 2.6 Schematic drawing of driving force dependence of charge transfer rate. Figure adapted from Gray, H. B.; Winkler, J. R. *Ann. Rev. Biochem.* 1996, *65*, 537.

2.2.3 Driving force

According to the Marcus formula, Eq. (2.1), the last parameter controlling the rate of charge transfer is the driving force ΔG^o. As shown in Figure 2.4, the driving force can be estimated by using redox potentials:[6]

$$\Delta G^o = E_{ox} - E_{red} \tag{2.11}$$

Here, E_{ox} is the oxidation potential of the ground-state donor and E_{red} is the reduction potential of the ground-state acceptor. According to Eq. (2.1), values of k_{CT} can be found experimentally by measuring the rate constant for the reaction under different conditions, giving different values for ΔG^o.[53,78] A characteristic plot of k_{CT} against ΔG^o is bell-shaped and two free energy regimes can be distinguished, depending on the relative magnitudes of ΔG^o and λ. This dependence of k_{CT} on ΔG^o at constant λ is illustrated in Figure 2.6.

In the "normal" regime where $-\Delta G^o < \lambda$, the rate constant increases as the driving force for the charge transfer becomes more favorable. For $-\Delta G^o = \lambda$, the product parabola

intersects the reactant parabola at the minimum, the activation free energy is zero, and the reaction proceeds without activation barrier. If the driving force increases even further, $-\Delta G^{o} > \lambda$, then the rate constant will decrease with increasing driving force. This situation is referred to as "inverted" regime.

2.3 Charge transfer mechanisms

As described in Chapter 1, insight into the mechanism of CT in DNA is important for understanding the mechanisms of DNA damage and repair, and for electronic applications. In general, there are two possibilities for CT in DNA,

(*i*) DNA is oxidized, an electron is subtracted from the DNA, and a radical cation is generated in the DNA strand.

(*ii*) DNA is reduced, an excess electron is injected into the DNA, and a radical anion is generated in the DNA strand.

In the following we will focus on the migration of the holes in donor-bridge-acceptor systems, where the hole can be generated by radiation or oxidizing agents,[20,21] as well as by hole injection.[79]

2.3.1 Unistep superexchange and multistep hopping mechanisms

On the basis of a theoretical analysis,[80] two distinct mechanisms are considered for hole transfer and transport in DNA. The first is a *unistep superexchange* mechanism. This hole transfer process is described as a "coherent tunneling process" because the hole can transfer from donor to acceptor without stopping in mid-journey. In other words, no observable radicals of nucleobases in the bridge between the donor and acceptor exist; only donor and acceptor radicals can be observed. The second mechanism is referred to as *multistep hopping*. In this incoherent process, charge can localize on bases of the bridge, and the corresponding radical cations can be observed as intermediates.

The CT mechanism can be characterized by an exponential decay of the CT rate (k_{CT}) with the distance R_{da} between donor and acceptor:[54,79,81]

$$k_{CT} = k_0 \exp(-\beta R_{da}) \tag{2.12}$$

where k_0 is a pre-exponential factor and β is the falloff parameter. Note that the smaller the value of β, the weaker the distance dependence of the CT rate. β values determined for CT reactions in DNA can be found in a wide range from less than 0.1 $Å^{-1}$ to 1.5 $Å^{-1}$.[30,82] For single-step tunneling processes, the characteristic values of β are larger than 0.6 $Å^{-1}$; this is considered as strong distance dependence. In contrast, the hopping mechanism, characterized by $\beta < 0.1$ $Å^{-1}$, exhibits a weak distance dependence of the rate.[83] Additionally, the CT mechanism is also determined by the relative energies of the charge donor and the bridge. This energetic control is based on the distinction between off-resonance and resonance donor-bridge coupling.[80] Superexchange occurs for off-resonance coupling between the lowest electronic state of the hole donor d^+ and the vibronic manifold of the bridge, i.e. $\delta E > 0$, see Figure 2.7a. The hole hopping mechanism, however, occurs when the lowest vibronic state of the electronic origin of d^+ is in resonance with a highly degenerate bridge vibronic manifold, i.e. $\delta E < 0$ (Figure 2.7b). Resonant coupling results in the injection of a hole into the appropriate base (B_1) of the DNA bridge; hole hopping to the next base (B_2) etc. will follow until an acceptor is reached which traps the hole.

It is useful to differentiate between charge transfer and charge transport in DNA. A transfer process involves a single-step tunneling, whereas a multistep hopping reaction can be associated with a transport process.[79]

2.3.2 G-hopping and A-hopping

In the multistep hopping mechanism, the hole can migrate from a donor (e.g. $G^{\bullet+}$) to an acceptor (e.g. GGG unit) by hopping on nucleobases of the bridge. For a bridge consisting of A·T and G·C pairs (e.g. the system $G^{\bullet+}$TTGTTGTTGGG), the intermediate G bases are oxidized and behave like relay stations for charge transport, called "G-hopping". To investigate whether and how charge migrates from $G^{\bullet+}$ to a cluster GGG, Giese and co-workers[5,79] have developed an assay by selective charge injection into a G unit. The charge transport to GGG has been monitored by competing trapping of the hole by a reaction with surrounding water molecules; thereby, the oxidation of a guanine is detected by the trapping reaction of the radical cations $G^{\bullet+}$ and $G^{\bullet+}$GG with water, which results in different fragmentation patterns P_G and P_{GGG}, respectively, of the DNA duplex. Then the

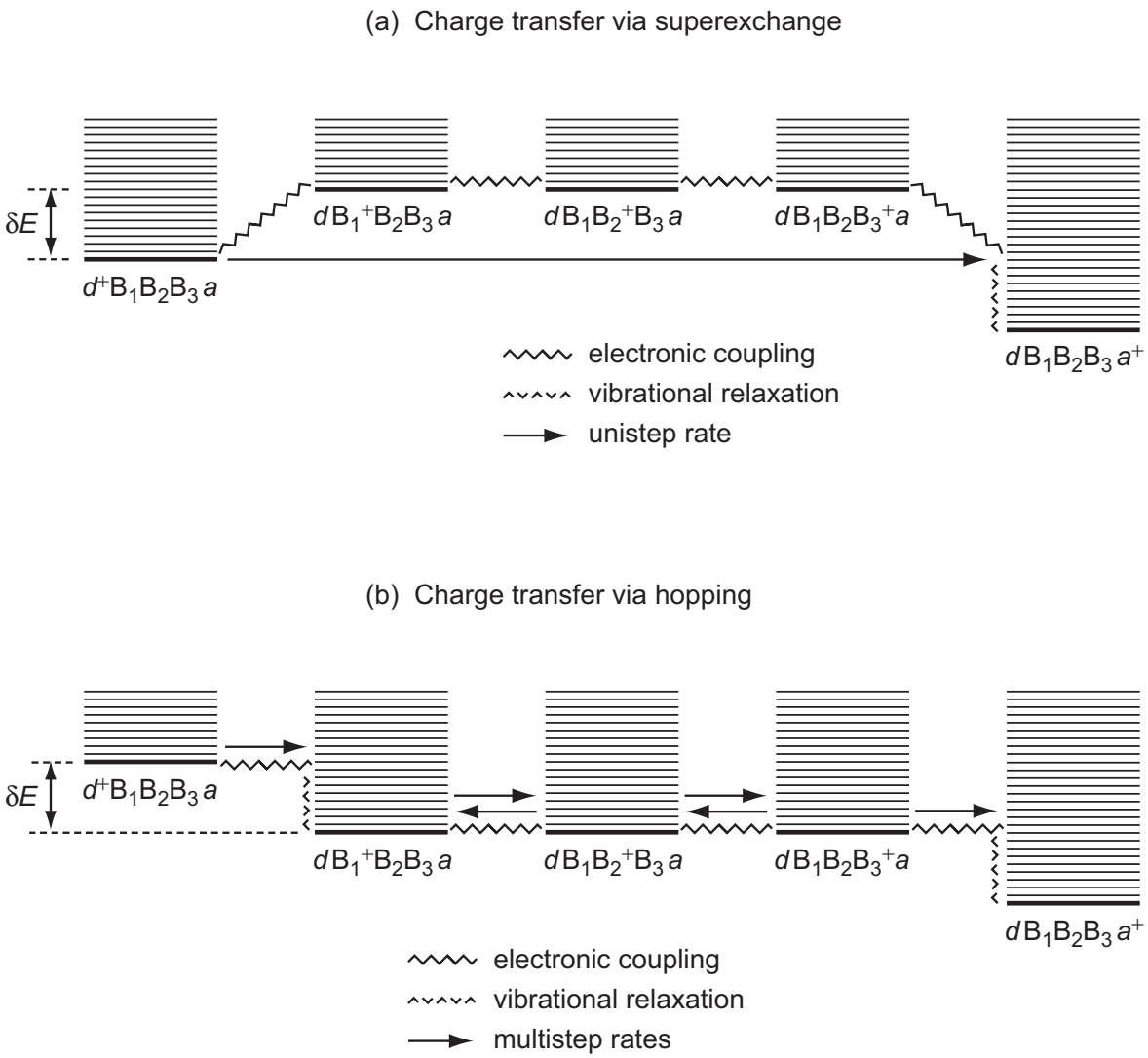

Figure 2.7 Vibronic level scheme for two distinct charge migration mechanisms in DNA, (a) unistep charge transfer via superexchange and (b) multistep change transfer via hopping. Figure adapted from Jortner, J.; Bixon, M.; Langenbacher, T.; Michel-Beyerle, M. E. *Proc. Natl. Acad. Sci. USA* 1998, *95*, 12759.

efficiency of charge transfer is measured by the P_{GGG}/P_G ratio.[84] It was found that in DNA strands where G units are located between a $G^{\bullet+}$ donor and a GGG acceptor sequence, long-distance charge transport occurs by a multistep hopping mechanism. Its efficiency depends on the number and the lengths of the hopping steps as well as on the rate of the trapping reaction by water. In addition, the slowest (or the longest) of the hopping steps determines the efficiency of the multistep charge transport process.[5,79] It was found that the maximum distance for hole hopping in a duplex with guanine bases separated by single A·T pairs is about 300 Å.[83]

Furthermore, Giese and co-workers measured the efficiency of charge transfer from $G^{\bullet+}$ to GGG unit, separated by $(A{\cdot}T)_n$ bridges of various lengths, i.e. n = 1–16, in double strands $G^{\bullet+}(T)_nGGG$.[84] The results showed that in short $(A{\cdot}T)_n$ bridges (n = 1–3) the efficiency of the transfer decreased by about a factor of 10 per each intervening A·T base pair. This yields a strong distance dependence of single-step tunneling process with β = 0.6 $Å^{-1}$, which is in very good agreement with experiments of Meggers et al.[32] In longer sequences ($n \geq 4$), the distance dependence nearly disappears. This behavior (switching from coherent superexchange charge transfer at short distances to a different process at longer distances) has been predicted by theory,[85–88] and can be interpreted with thermally induced hopping of charges between adenine bases (A-hopping). The thermodynamically favored tunneling between $G^{\bullet+}$ and G is so slow at long $(A{\cdot}T)_n$ bridges that an endothermic oxidation of the adjacent A occurs, which has the second lowest ionization potential of the four nucleobases. Once A is oxidized to the adenine radical cation $A^{\bullet+}$ in a thermally activated reaction step, the positive charge can migrate very quickly between the neighboring adenines. Thus, the rate determining step is the oxidation of A by $G^{\bullet+}$, and the number of A·T base pairs influences only slightly the overall rate of the hole transport.[79]

Chapter 3

Molecular Dynamics Simulations of Nucleic Acids

3.1 Methodological aspects of molecular dynamics simulations

Computational methods are playing an increasingly more important role in the investigation of molecular structures and the dynamic behavior of molecular systems. Quantum mechanical methods deal with the electronic structure of molecules, and such computational results can reach the accuracy of experimental data. These calculations are, however, time-consuming and only small molecular systems can be treated in a regular fashion. On the other hand, for biological macromolecules such as nucleic acids and proteins, molecular mechanical or force field methods are used to treat the systems as a function of the nuclear positions only, bypassing the use of wave functions or the total electron density. In this section we will discuss general concepts of the force field approach, followed by molecular dynamics simulations, which employ the force field method to determine molecular structures and properties. Then some techniques used for simulations, e.g. periodic boundary conditions, are described. Finally, we will describe molecular dynamics simulations of DNA.

3.1.1 Force field

At the root of so-called molecular mechanics is a force field which describes the potential energy surface of the system under study. It consists of various contributions, such as bonded or valence terms (bond stretching, angle bending and torsion) and non-bonded

terms (mainly represented by van der Waals and Coulomb forces). Each of these terms contains empirical parameters. These parameters are optimized to obtain the best fit of experimental values such as geometries, energies and spectroscopic properties. If some parameters are not experimentally available, quantum chemical calculation of representative fragments can be used to obtain the desired values. Furthermore, the parameters are fitted to specific groups or types of molecular systems and therefore the force fields are generally designed to treat specific classes of molecules. For instance, AMBER,[89] CHARMM[90] and GROMOS[91] force fields are designed to treat proteins and nucleic acids, while the MMX[92] force field has been designed for treating organic compounds.

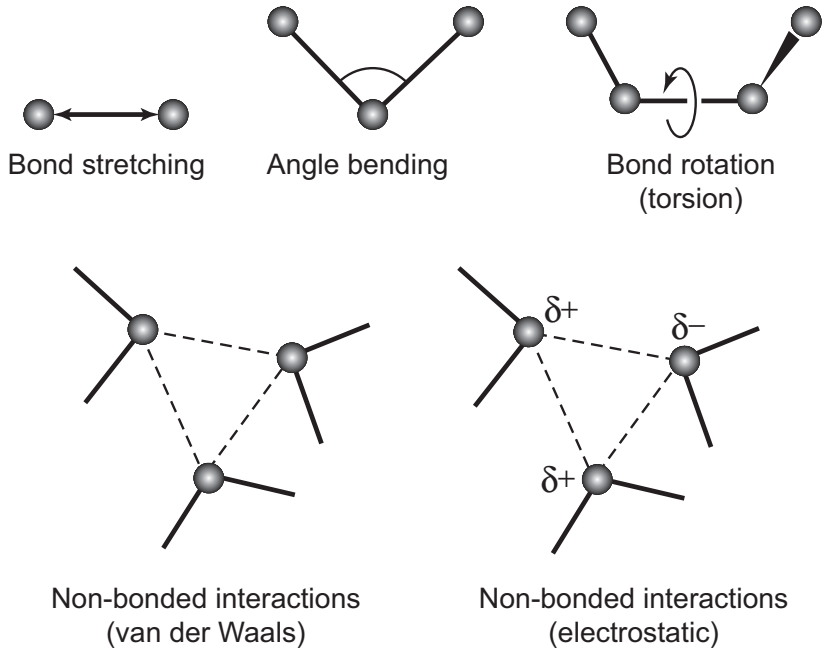

Figure 3.1 Schematic representation of the four contributions to a force field: bond stretching, angle bending, torsion and non-bonded interactions. Figure adapted from Leach, A. R. *Molecular modelling: Principles and applications*, Longman, 1996.

The potential energy surface $V(\mathbf{R})$ of a molecular systems is a function of the atomic positions, **R**, of all atoms in the system under study. A molecular mechanics force field consists of four components, as mentioned above (Figure 3.1):

$$V(\mathbf{R}) = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{non-bonded}} \tag{3.1}$$

The first term in Eq. (3.1) represents the energy between pairs of bonded atoms. The individual terms are often assumed as harmonic potentials according to Hooke's law:

$$E_{\text{bond}} = \sum_{\text{bonds}} \frac{k_b}{2}\left(l - l_0\right)^2 \tag{3.2}$$

The force constants k_b are specific for each pair of bound atoms. The energy increases when the bond is compressed or stretched, i.e. when the bond length l deviates from the reference value l_0. The second term in the equation above is associated with the deviation of a bond angle θ from the corresponding reference value θ_0. This energy is also represented by a harmonic potential

$$E_{\text{angle}} = \sum_{\text{angles}} \frac{k_\theta}{2}\left(\theta - \theta_0\right)^2 \tag{3.3}$$

The contribution of each angle is characterized by the force constant k_θ and the reference value θ_0. The torsion angle potential, the third term in Eq. (3.1), models steric barriers between atoms separated by three covalent bonds. The motion associated with this term is a rotation around the middle bond (between atoms 2 and 3). This torsion potential is assumed to be periodic and can be expressed as a cosine series expansion

$$E_{\text{torsion}} = \sum_{\text{torsions}} \frac{V_n}{2}\left(1 + \cos(n\omega - \gamma)\right) \tag{3.4}$$

The barrier V_n controls the amplitude of this periodic function, n is the multiplicity describing the number of minima on the potential energy surface, ω is the torsion angle, and the so-called phase factor γ is the value when the torsional angle passes the minimum. The last term in Eq. (3.1) represents non-bonded interactions. The non-bonded term usually contains two potentials describing van der Waals and Coulomb interactions

$$E_{\text{non-bonded}} = \underbrace{\sum_i \sum_{j>i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)}_{\text{van der Waals}} + \underbrace{\sum_i \sum_{j>i} \frac{q_i q_j}{\varepsilon r_{ij}}}_{\text{Coulomb}} \tag{3.5}$$

The van der Waals interaction between two atoms i and j separated by distance r_{ij} arises from a balance between attractive dispersion and repulsive forces and is often described by a Lennard-Jones potential (the first term in Eq. (3.5)). A_{ij} and B_{ij} are parameters depending on a pair of atoms i and j. The Coulomb potential represents the electrostatic interaction

within a pair of atoms; q_i and q_j are charges on atoms i and j, respectively, and ε is the dielectric constant of the medium.

Sometimes force fields include additional terms for specific interactions, such as hydrogen bonding or dipole-dipole interaction. Furthermore, the Morse potential for bonds stretching, Taylor expansions with higher terms and cross-terms between potentials can be used to improve the agreement of calculated experimental data. More details can be found in text books on molecular modeling, e.g. Ref. 93.

3.1.2 Basic theory of molecular dynamics

Molecular dynamics (MD) is a computer simulation technique which provides information on the time-dependent behavior of a molecular system.[93–95] The molecular dynamics method was first introduced by Alder and Wainwright[96] in 1957 using a hard sphere model. Many important insights concerning the behavior of simple liquids emerged from their study. The first biomolecular simulation was carried out in 1977 by McCammon et al.[97] Today one routinely finds in the literature hundreds papers on molecular dynamics simulations of proteins and nucleic acids; e.g., see the review by Cheatham and Kollman[98] and references therein. In the following section, basic aspects of a MD simulation will be described.

In a molecular dynamics simulation, one describes the motion of the nuclei of a system by classical mechanics. Following Newton's second law, the equation of motion for each atom i is

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2} \tag{3.6}$$

where $\mathbf{F}_i$ is the force acting on atom i, m_i is its mass and $\mathbf{a}_i$ is its acceleration (the second derivative of the location vector $\mathbf{r}_i$ with respect to time t). The force acting on atom i can be computed directly from the negative derivative of the potential energy surface, Eq. (3.1), with respect to the coordinate $\mathbf{r}_i$:

$$\mathbf{F}_i = -\frac{dV(\mathbf{R})}{d\mathbf{r}_i} \tag{3.7}$$

The acceleration $\mathbf{a}_i$ is thus expressed as

$$\mathbf{a}_i = -\frac{1}{m_i}\frac{dV(\mathbf{R})}{d\mathbf{r}_i} \tag{3.8}$$

Integration of these equations of motion results in a trajectory that describes the positions and velocities of the atoms as they vary with time. Due to the complicated nature of the potential energy function of a many-atom system, there is no analytical solution to the equations of motion. They must be solved numerically using a *finite difference method*. The equations are solved step-by-step at discrete time intervals δt, the so-called *time steps*. The accelerations of the atoms are computed from the force, Eq. (3.8), which is assumed to be constant during a time step. The accelerations are then combined with the positions and velocities at time t to calculate new positions and velocities at time $t+\delta t$. The determined forces on the atoms are used in turn to estimate new positions and velocities at time $t+2\delta t$. This procedure is repeated for each step of the simulation, yielding the MD trajectory. There are various numerical algorithms developed for integrating the equations of motion. Below we describe several algorithms which are widely used in molecular dynamics simulations.

3.1.3 Integration algorithms

All integration algorithms assume that the positions, velocities and accelerations can be approximated as Taylor series expansions

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 + \frac{1}{6}\mathbf{b}(t)\delta t^3 + \frac{1}{24}\mathbf{c}(t)\delta t^4 + \cdots \tag{3.9}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \mathbf{a}(t)\delta t + \frac{1}{2}\mathbf{b}(t)\delta t^2 + \frac{1}{6}\mathbf{c}(t)\delta t^3 + \cdots \tag{3.10}$$

$$\mathbf{a}(t+\delta t) = \mathbf{a}(t) + \mathbf{b}(t)\delta t + \frac{1}{2}\mathbf{c}(t)\delta t^2 + \cdots \tag{3.11}$$

$$\mathbf{b}(t+\delta t) = \mathbf{b}(t) + \mathbf{c}(t)\delta t + \cdots \tag{3.12}$$

where $\mathbf{v}$ is the velocity (the first derivative of the position $\mathbf{r}$ with respect to time), $\mathbf{a}$ is the acceleration (the second derivative), $\mathbf{b}$ is the third derivative, and so on.

Verlet algorithm

The Verlet method[99] is a direct solution of the second order differential equations. To calculate the new positions at time $t+\delta t$, i.e. $\mathbf{r}(t+\delta t)$, the Verlet algorithm uses the positions at time *t*, $\mathbf{r}(t)$, the accelerations at time *t*, $\mathbf{a}(t)$, and the positions of the previous step, $\mathbf{r}(t-\delta t)$. The relationships between these quantities and the velocities can be written down as Taylor series expansion

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 + \cdots \tag{3.13}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 - \cdots \tag{3.14}$$

From these two equations on derives

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \mathbf{a}(t)\delta t^2 \tag{3.15}$$

One can see from the last equation that the Verlet method uses no explicit velocities. The velocities, used to estimate the kinetic energy, can be calculated by the difference in positions at times $t-\delta t$ and $t+\delta t$ divided by the time difference,

$$\mathbf{v}(t) = [\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)]/2\delta t \tag{3.16}$$

The advantage of the Verlet algorithm is that it is straightforward and the storage requirements are modest. However, this method is of moderate precision.

Leap-frog algorithm

In the leap-frog algorithm,[100] the velocities at time $t+1/2\,\delta t$ (namely half-step velocity) are first calculated from the velocities at time $t-1/2\,\delta t$ and the accelerations at time *t*,

$$\mathbf{v}\left(t+\frac{1}{2}\delta t\right) = \mathbf{v}\left(t-\frac{1}{2}\delta t\right) + \mathbf{a}(t)\delta t \tag{3.17}$$

Then the positions at time $t+\delta t$ are calculated from these velocities together with the positions at time *t*

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}\left(t+\frac{1}{2}\delta t\right)\delta t \tag{3.18}$$

In this way, the velocities leap over the positions to give their values at $t+1/2\,\delta t$, then the positions leap over the velocities to give their values at $t+\delta t$. The advantage of the leap-frog method is that the velocities are explicitly calculated. The velocities at time *t* can be expressed as

$$\mathbf{v}(t)=\frac{1}{2}\left[\mathbf{v}\left(t+\frac{1}{2}\delta t\right)+\mathbf{v}\left(t-\frac{1}{2}\delta t\right)\right] \tag{3.19}$$

The disadvantage, however, is that the positions and velocities are not synchronized, and, therefore, it is not possible to calculate the kinetic energy at the same time as the positions are defined.

Velocity Verlet algorithm

The handling of the kinetic energy is not ideal in either of the above algorithms. The velocity Verlet algorithm[101] gives positions, velocities and accelerations at the same time and does not compromise precision. These quantities can be computed in the following way:

$$\mathbf{r}(t+\delta t)=\mathbf{r}(t)+\mathbf{v}(t)\delta t+\frac{1}{2}\mathbf{a}(t)\delta t^2 \tag{3.20}$$

$$\mathbf{v}(t+\delta t)=\mathbf{v}(t)+\frac{1}{2}\left[\mathbf{a}(t)+\mathbf{a}(t+\delta t)\right]\delta t \tag{3.21}$$

3.1.4 Time step and SHAKE algorithm

As mentioned previously, in a molecular dynamics simulation, the equations of motion are solved at each time step. Choosing an appropriate time step δt is essential for a successful simulation. A time step that is too small will make it necessary to run more iterations, and thus the simulation takes longer to run. In contrast, too large a time step will cause instabilities in the simulation due to fact that atoms move too far. Generally, the time step should be one order of magnitude less than the time scale of the shortest motion (e.g. bond vibrations, especially bonds involving hydrogen atoms have much higher vibrational frequencies). Nevertheless, high frequency motions are usually of less interest than lower frequency motions which correspond to major conformational changes. A typical time step

for all-atom force fields of biological molecules is 1 femtosecond (fs) or even less. It is possible to enlarge the time step, through constraining bonds and angles (involving hydrogen atoms, for example) to specific values. The most commonly used method for applying constraints is the SHAKE algorithm.[102] A time step of 1–2 fs is commonly used with SHAKE on the hydrogen atoms.

In the SHAKE method, N_c constraints are formulated as[93,103]

$$\sigma_k(\mathbf{r}_i) = r_{ij}^2 - d_{ij}^2 = 0, \qquad k = 1, \ldots, N_c \tag{3.22}$$

where σ_k is the deviation of bond distance for constraint k, $\mathbf{r}_i$ is the position vector of atom i, r_{ij} is the distance between atoms i and j involved in constraint k, and d_{ij} is the reference bond length. Using Lagrange multipliers $l_k(t)$, the constraints are added to the potential energy function $V(\mathbf{R})$, to yield the equations of motion as[103]

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\frac{\partial}{d\mathbf{r}_i}\left[V(\mathbf{R}) + \sum_{k=1}^{N_c} l_k(t)\sigma_k(\mathbf{r}_i) \right] \tag{3.23}$$

Recall that the equations of motion without constraints are given by Eqs. (3.6) and (3.7). The first term on the right-hand side of Eq. (3.23) represents the unconstrained force acting on atom i, whereas the second term represents the overall constrain force that affect atom i. By solving the equations of motion in Eq. (3.23), see Section 3.1.3, the new coordinates $\mathbf{r}(t+\delta t)$ should satisfy the constraint equations, Eq. (3.22), and thus lead to a system of N_c quadratic equations for the undetermined multipliers that must be solved simultaneously. Two approximations are then made: (*i*) the system of equations is linearized by neglecting any quadratic term in the multipliers, and (*ii*) the multipliers are determined independently in sequence by omitting the coupling between distance constraints involving a common atom. As a consequence of these approximations, the procedure must be performed iteratively until all constraints are satisfied within a specific tolerance, e.g. 10^{-5} Å^2. In other words, simultaneously solving the equations of motions, Eq. (3.23), gives the new positions of all atoms in the system, in which the new positions keep the bond at the required distances d_{ij}, satisfying Eq. (3.22).

3.1.5 Periodic boundary conditions

The idea of a computer simulation is to enhance the ability of small systems to represent properties of large systems. An *infinite* system is usually simulated by using of *periodic boundary conditions*, which are required to keep the constant density. This condition leads to an infinite array of identical boxes in three dimensions. The original box is surrounded by an infinite number of image boxes which have the same number of molecules and the same configuration, as illustrated in a two-dimensional picture in Figure 3.2.

In a simulation, as a molecule moves in the original box, its periodic image in each of the neighboring boxes moves in exactly the synchronous way. Thus, as a molecule leaves the central box, one of its images will enter through the opposite face. As shown in Figure 3.2, the duplicate boxes are labeled A, B, C and so on. When particle 1 moves through a boundary, its images, 1_A, 1_B, 1_C and so on (where the subscript specifies in which box the image lies) move across the corresponding boundaries. Consequently, the number density in the central box (and hence in the entire system) is conserved.

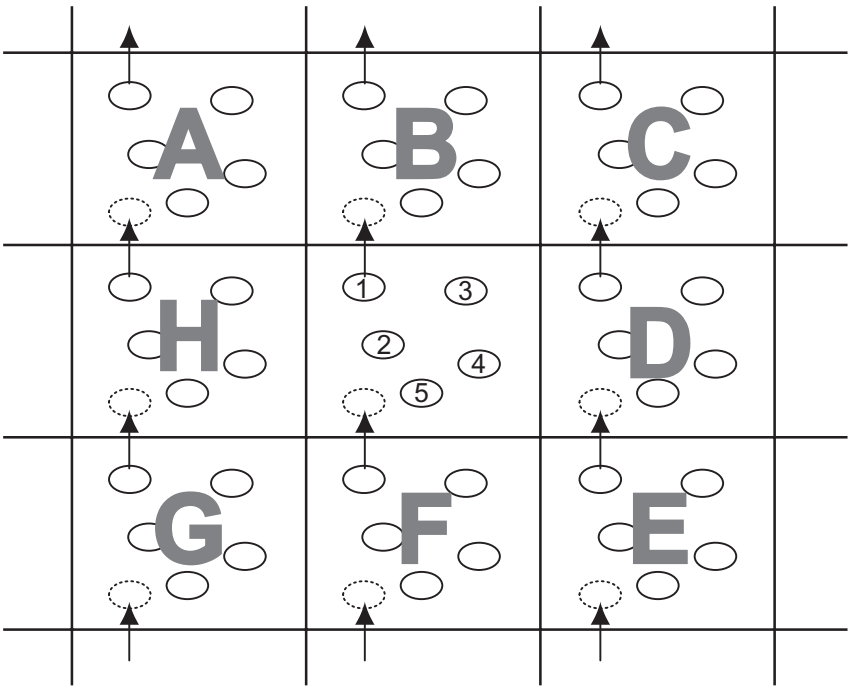

Figure 3.2 A two-dimensional periodic array.

3.1.6 Treatment of long-range electrostatic interactions

The most time consuming part of any molecular dynamics simulation is the calculation of the electrostatic interactions. These interactions fall off as $1/r$, where r is the separation between charges, and have consequently to be considered as long-range. A great number of approximations have been developed to make this problem tractable with available computational means. In the following, these methods will be briefly described, focusing on their advantages and drawbacks.

Truncation methods were developed to limit the computational effort for evaluating the long-range forces. These methods neglect electrostatic interactions between two atoms beyond a certain distance called the "truncation" or " cutoff" distance, r_{off}.[104–106] The interactions can be truncated suddenly at the cutoff distance, namely straight truncation method, i.e. the electrostatic interactions are zero at $r > r_{\text{off}}$. Alternatively, the interactions are smoothed by applying some smoothing schemes, either *shift* or *switch* scheme. Both schemes were developed by Brooks et al.[107] The shifting function increases the magnitude of the interactions before it is smoothed to zero and thus the short-range interactions are disturbed. Such distortion and overestimation of short-range interactions are a drawback of the shifting function scheme. For other smoothing scheme, the switching function switches off the electrostatic interactions in the range $r_{\text{on}} \le r \le r_{\text{off}}$. With a proper value of r_{on}, the short-range interactions are not distorted, giving continuous forces or a continuous potential energy. Although the truncation methods can significantly reduce the amount of computational time for evaluating the electrostatic interactions, these methods are inaccurate because of the finite cutoff distance which severely restricts the infinite character of the system. This may result an unstable geometry for a long simulation.

Ewald summation method which was first devised by Ewald[108] has been widely used to handle the problem of the long-range interactions. It allows one to calculate all electrostatic interactions in an infinity array of periodic boundary condition. In this method, the electrostatic potential is expressed in two terms. The first is a short-range term which is summed in real space, or direct space, up to a cutoff distance and the second is long-range term that is calculated in reciprocal space (taking advantage of the periodicity). The Ewald trick makes both sums converge with a Gaussian fall-off. The high accuracy of the Ewald summation method comes with a large computational expense that scales in

order of $N^{3/2}$ where N is a number of atoms in the system. The *particle mesh Ewald* (PME) method based on interpolation of the reciprocal space Ewald sums has been developed by Darden et al.[109] The main difference from the standard Ewald method is in treatment of the reciprocal space summation. A fast Fourier transformation is used to evaluate the convolutions of the interaction energy and force. The efficiency of these procedures greatly reduces the cost of the reciprocal space sum. The amount of time for PME method is in order of $N \cdot \log(N)$.

In molecular dynamics simulations of biomolecules, an accurate treatment of the long-range interactions is crucial for achieving stable nanosecond trajectory. The importance of such a treatment and also a comparison of the methods will be described below.

3.2 Status of MD simulations of DNA: An overview of methodology and previous results

MD simulations providing information on the time-dependent behavior of a complex molecular system are a powerful computational approach to studying structure and motions of macromolecules. A realistic goal of a simulation is to obtain the molecular structure and properties in agreement with the corresponding experimentally observed quantities. Nowadays, advances in computer technology, methodology and empirical force fields have led to significantly improvements in MD simulations of nucleic acids in terms of both system size and length of the simulation. MD simulations of DNA with explicit water molecules and counterions using method such as particle mesh Ewald for handling the long-range electrostatic interactions[109] and empirical force fields, such as AMBER,[89] CHARMM[90] and GROMOS,[91] yield stable trajectories in nanosecond-length MD simulations and have had excellent success in predicting nucleic acid structures and interactions that agree with experimental results.[98,106,110–113] However, run lengths of even nanoseconds are still not long enough to access many nucleic acid phenomena of interest and thus MD modeling is a work in progress.[114] In the following, we will summarize the results of MD simulation on DNA conformations, as affected by the empirical force fields, simulation protocols, and system preparations used.

3.2.1 Force field dependence of DNA conformation

An accurate description of the interatomic forces at play in chemical systems is necessary for a correct description of their dynamical behaviour.[98,114,115] Refined versions of common force fields used for MD simulations of nucleic acids, such as AMBER,[89] CHARMM[90] and GROMOS,[91] are regularly included in associated MD packages. MD simulations conducted independently by several groups using different methodologies and parameter sets have shown that a given force field may favour one DNA form over another. These studies therefore raised the issue of a possible "force field dependent" polymorphism. Cheatham and Kollman[110] performed four simulations of one nanosecond of the duplex d(CCAACGTTGG) (≡ double helix 5′-CCAACGTTGG-3′) surrounded by 18 Na^+ ions, with the AMBER force field. Two simulations were started from canonical A-DNA and canonical B-DNA structures. As judged from root mean square deviation (RMSD) criteria, the authors reported an A-to-B transition for the simulations starting from the A-form; the simulations starting from the B-DNA structures remained close to the B geometry. This study suggested that simulations starting from different DNA forms converged toward an identical average structure, in reasonable agreement with available data from NMR spectroscopy and X-ray crystallography.[110] At the same time, however, Yang and Pettitt[116] performed a 3.5 ns simulation starting from the B-form crystallographic structure of duplex d(CGCGAATTCGCG) dodecamer in a 0.45 M NaCl aqueous solution, using the CHARMM23 parameter set. They observed the reverse B-to-A transition. Some explanations were suggested for the different behaviors noted by these groups.[110,116] The observed transitions may be force field driven, but could also result from the different salt conditions used in these simulations, or from a combination of both factors.

In order to address these points, MD simulations on DNA molecules in various environments were conducted. Young et al.[117] as well as Duan et al.,[44] produced nanosecond simulations of the B-form d(CGCGAATTCGCG) duplex surrounded by 22 Na^+ counterions. These simulations, which used the AMBER force field, showed a stabilization of the B-form structures. An extension of the 1.5 ns MD trajectories produced by Young et al.[117] to 5 ns led to a remarkably stable model of B-form DNA.[111] For the same DNA sequence and under the same ionic conditions, Cieplak et al.[118] observed a convergence of the canonical A- and B-DNA forms toward the structure determined by X-ray experiments. They also noted that the A-to-B transitions simulated with AMBER

parameters are probably not sequence dependent since similar results were obtained for the d(CCAACGTTGG) duplex.[110]

A careful comparison of two 4 ns simulations using the AMBER and the CHARMM force fields, and starting from the same initial A-like model-built structure of the duplex d(CCCCCTTTTT) in an approximately 0.5 M NaCl aqueous solution, was performed by Feig and Pettitt.[119] MD average structures generated with the CHARMM force field displayed an A-form base geometry, whereas the AMBER results showed an intermediate between A and B. The backbone assumes a B-form on both strands with the AMBER force field, whereas the CHARMM23 MD results produce heterogeneous structures with the purine strand in A form and the pyridine strand in dynamical equilibrium between A-form and B-form conformations. The results failed to reproduce an overall B conformation in the A·T region, particularly with CHARMM23; experimentally, A·T base pairs have a preference for a B form structure, whereas C·G pairs have a preference for an A form.[120] A full description of the new CHARMM27 force field has just appeared.[121] Applications of CHARMM27 have been reported for DNA, including details of the dynamical properties of A- and B-DNA sequences under a variety of conditions.[122] The results show the new force field to provide an improved account of the distributions of conformational properties which yields satisfactory agreement with a variety of experimental observables.

The GROMOS simulation program was recently discussed by van Gunsteren and colleagues[91] and several recent papers reported results for MD simulations on DNA using the GROMOS force field. Tapia et al.[123] applied GROMOS to a DNA decamer that remained stable in the B-form, but required a modification of certain Lennard-Jones terms to make the aliphatic carbons more hydrophobic and to constrain the counterions around the protein–DNA complex to keep them from collapsing on the DNA. A fuller discussion of the parameterization of the Lennard-Jones parameters for aliphatic united-atom carbons has been provided.[124] At this point, the GROMOS force field has not received as extensive critical validation for studies on nucleic acid systems as AMBER and CHARMM, but work in this direction is in progress. Moreover, it is not possible to establish a classification of the different force fields available. Additionally, it is both useful and insightful to develop simultaneously several force fields because this allows comparisons between theoretical and experimental results from different perspectives. It should be noted

that the choice of a force field is often conditioned by the use of the associated MD package.

3.2.2 Long-range electrostatic interaction treatment

The selection of an accurate force field is not the only factor determining the quality of MD simulations. As will be shown below, the use of an accurate method for calculating the long-range electrostatic interactions is equally decisive because long-range interactions are known to play an important role in highly charge biomolecules like DNA.[125] In particular, an accurate treatment of the long-range interactions is crucial for achieving stable trajectories for several nanoseconds.[106,125,126]

As mentioned above, truncation methods affect the stability of long-time simulations and their effect has recently been evaluated for MD simulations of DNA in aqueous solution.[106,125] Application of switching functions, which bring the potential energy to zero in the interval from 7.5 Å to 8.5 Å and thereafter in a simulation of the duplex d(CGCGAATTCGCG) showed that the distribution of interphosphorus distances exhibited a peak close to the 8.5 Å boundary which corresponds to a stretch of the DNA backbone.[106] At this distance, the interaction energies between adjacent phosphate groups is close to zero. A comparison with the experimental inter-phosphorus distance distribution, calculated from a set of 61 B-DNA structures extracted from the nucleic acid data base,[127] clearly demonstrates that this peak is an artifact. Although the selection of larger cut-off distances should attenuate the truncation problems. However, as shown in ion−ion radial distribution functions calculated from an MD simulation of a 1.0 M aqueous NaCl solution MD using a 16 Å truncation distance[128] (which is larger about 2 times than the previous study[106]), large artifact peaks appear clearly at the cut-off boundary. Moreover, this truncation technique resulted in a rapid drift in RMSD, with respect to the starting structure of DNA. A comparative study using the particle mesh Ewald (PME) method[109] showed, nevertheless, a very high dynamical stability. Further, the above described artifacts have not been detected in simulations using Ewald summation techniques.[129]

Alternatively, force-shifting functions, which scale the force (instead of potential energy) to zero at a specific distance, have been applied for MD simulations of DNA

hexamer in aqueous solution with a 12 Å cutoff.[125] This method produces stable nanosecond MD simulations of a DNA molecule very similar with the simulations performed by PME method. Thus, the force-shifting approach can be also employed. However, the PME method is recommended to be used for long-range electrostatic interactions[125] and, recently, it has become very popular in MD simulation studies.[130]

3.2.3 Continuum and explicit solvent models

In addition to the improvement of force fields, one of the main computational challenges remains the simulation of large systems over long times. One approach to this goal is the replacement of explicit solvent with hybrid explicit/implicit[131,132] or entirely implicit solvent model.[133,134] In the explicit solvent model, the condensed phase environment is treated as a cluster containing a large number of solvent molecules, whereas in the implicit model the effect of the solvent is introduced indirectly via a distance (r)-dependent dielectric screening constant (ε), e.g. $\varepsilon = r$.[135] Alternatively, the semi-explicit solvent model involves inclusion of only the first hydration layer of explicit water molecules whose motion is governed by standard MD forces, and a distance-dependent dielectric function.[131,132]

The most reliable implicit method is Poisson–Boltzmann (PB) method. This approach has proved to be capable of reliably predicting both solvation energies and solvent-dependent conformational changes, but this computation is very expensive, and therefore there has been limited use of PB method in MD simulation.[136] More recent efforts have focused on generalized Born (GB) model which is much faster and can be formulated to include both salt and cavity contributions.[133,137] Such method can be parametrized to yield reasonable solvation energies[137,138] and pK_a shifts,[137,139] and it has already been used for MD simulations on nucleic acids.[140–142] Srinivasan and co-workers[141] examined the stability of A- and B-forms of DNA, RNA and phosphoramidate DNA helices using both PB and GB methods. Further studies of the conformational stability comparing the sequence-dependent preferences of $dA_{10}\cdot dT_{10}$ and $dG_{10}\cdot dC_{10}$ show that the G·C base pairs have a greater preference for A-form DNA than A·T base pairs, as is seen in the van der Waals interaction energies.[143] Tsui and Case[140] showed that the GB model results in good agreement with simulations using explicit water solvent in term of both

structure and energetics of DNA decamer. Srinivasan et al.[144] provided a detailed comparison of the difference between PB and GB simulations of four nucleic acid structures. They showed that the GB approach does an excellent job of reproducing the PB solvation energies of the system at zero ionic strength. At low salt concentrations (0.1 M), however, they noted a large differences of the solvation energy between PB and GB models. Hence, for many purposes, the GB method offers an alternative to the more expensive PB method and explicit water simulations for studies of nucleic acid energies and structure.[140,144]

3.2.4 DNA conformational stability and transitions

DNA structures can adopt a variety of canonical forms; they are strongly influenced by solvent and ionic environment as well as the DNA base sequence.[46,98,114,145] Early crystallographic evidence from DNA fibers demonstrated a preference for the A-form of DNA in mixed organic-water solutions and the B-form in pure water.[146] The role of the base pair composition is better understood because, as mentioned already, C·G base pairs generally tend less toward B-like conformations than A·T base pairs.[120] Solvent conditions which lower the activity of water by co-solvents or salt tend to cause a sequence-dependent B-to-A transition.[147]

Cheatham and Kollman[110] carried out two simulations of A and B forms of d(CCAACGTTGG) duplexes in pure water with 18 Na^+ ions. Their MD results showed that the starting B-DNA canonical is stabilized throughout the course of simulation, whereas the starting A-form is spontaneous A to B transitions seen in 500 ps. In contrast, B-DNA to A-DNA transitions have been observed in solutions with counterions Li^+ and Mg^{2+}.[148] B to A transitions have also been reported in mixed ethanol-water solutions in the presence of Na^+, K^+, and Cs^+.[149–151]

As already discussed, a force field bias may cause a conformational transition of DNA. To overcome some of these force field deficiencies, Langley[152] modified a set of CHARMM-type parameters coupled with AMBER charges, known as the Bristol-Myers-Squibb (BMS) force field, for nucleic acids. Using the BMS force field, a stable B-form is reproduced at low salt concentrations. B-to-A DNA transitions were observed for the duplex d(GGGCCC) in 75% ethanol/water, whereas an A-to-B DNA transition was

obtained for the duplex d(AAATTT) in a 75% ethanol/water mixed solvent. These results confirm the effect of the base sequence on a conformational transition of DNA.[120]

3.2.5 Water and ion distributions

Water molecules and counterions are known to be essential structural elements of nucleic acid systems.[98,114,115,153] A large array of experimental results, coming mainly from NMR spectroscopy and X-ray crystallography, confirmed this view.[154–156] From the discussion in the previous section, it is also clear that modern simulation protocols and force fields are sufficiently robust even to represent subtle environmental dependencies of a DNA duplex structure. Given that MD simulations provide a time history for all atomic motions in the system, in simulations that include explicit water and counterions, precise details for all specific ion and water interactions with the nucleic acid can be monitored. An average picture of the hydration can be obtained by calculating the radial distribution functions.[157,158] A detailed picture of the hydration or ion association can be obtained by fixing the DNA to a reference structure and accumulating relevant populations or atomic densities on a grid for visualizing the hydration.[43,44,111] These results can be compared directly with specific high-resolution crystal structures or to average analyses obtained by looking at many different crystal structures.[154–156] In addition, populations and lifetimes for various DNA hydrogen bond donor or acceptor interactions to water or ions can be tabulated,[159,160] and then compared directly with NMR measurements.

IR spectroscopic experiments on DNA fibers at various water activities demonstrated that roughly 18–23 water molecules per nucleotide are in the primary hydration shell.[47] The binding duration of water molecules was estimated from MD simulations; such lifetimes were suggested to fall in the range from 10 to 600 ps, which is, in general, consistent with experimental data and suggests lifetimes of < 1 ns.[159–161] Simulation studies of the hydration of the duplex d(CGCGAATTCGCG) were able to reproduce the experimentally observed spine of water molecules located in the DNA minor groove.[44,111] Variations in the calculated hydration pattern suggest a sequence dependence. Similarly, a spine of water molecules was observed in the minor groove of the duplex d(CCAACGTTGG).[43] The local dielectric environment in MD simulations of DNA was analyzed by Young and co-workers;[162] they computed the dielectric profile near DNA and

found that the effective dielectric constant increases rather rapidly with distance and displayed bulk behavior beyond 5 Å. An analysis of the dielectric function revealed that the relative permittivity in the first shell of DNA obeys the following trend: phosphate backbone > major groove > minor grove. Estimates of the local dielectric constants in the major groove are consistent with interpretations based on fluorescence measurements;[163] these finding indicated that MD models of the solvent close to a DNA strand account in reasonably accurate fashion for the local solvent environment of a complicated polyelectrolyte. The calculated dielectric profile was fit to a sigmoidal function, which could be used to estimate the strength of charge–charge interactions near DNA.[162]

MD simulations provided information on the sequence-specific localization of monovalent cations in the DNA grooves. Hamelberg et al.[113] analyzed the major structure and the correlation between ion populations and the width of the minor groove based on a 10 ns MD trajectory of the duplex d(CGCGAATTCGCG) in the presence of Na^+ ions and water molecules. The results clearly showed a time-dependent influence of ion positions on the structure of the minor groove. When no ions interact with the groove, the groove is wide, whereas it narrows, especially in AATT region, when counterions reside in it. Similar correlations between the entrance of Na^+ into the minor groove and the groove narrowing were also observed in other studies.[112,117] Ion–water interactions narrow the groove through two distinct interactions: (*i*) ions interact directly with DNA bases in the minor groove, such as cross-strand thymine oxygens (O2) in the sequence above to give an internal ion-spine of hydration, or (*ii*) ions interact with phosphate groups in the AT sequence while water molecules in the minor groove interact directly with the bases.[113] In addition, recent contributions[113,164,165] have helped to clarify the localization of counterions in the minor groove. The localization is sequence-dependent, with AT-rich sequences, and particularly A-tracts, being favored over GC-rich or mixed sequences. An MD simulation of a pure GC sequence showed essentially no localization of mono cations in the minor groove.[165] This trend is consistent with the minor groove surface being more electronegative near AT-rich sequences than close to GC-rich sequences.[117] Alternatively, for the ion association in the major groove, McConnell and Beveridge[164] reported Na^+ near the G·C base pairs of the duplex d(CGCGAATTCGCG) and binding near a particular CG step was found to cause a slight narrowing of the major groove. Furthermore, in the study of Feig and Pettitt with 12 ns MD simulations of the duplex d(C_5T_5) in explicit water with

0.8 M additional NaCl,[112] the residence times of Na^+ around the duplex were found to fall in the range of tens to hundreds of picoseconds.

3.3 Objective

As the main goal of our study concerns charge transfer in DNA, in the subsequent sections we present the MD simulations of DNA duplexes in explicit water. The result of each simulation is an MD trajectory as time series of configurations. These trajectories, which can be used for calculating modeling charge transfer parameters, were routinely analyzed (e.g. root mean square deviations of structural parameters, distribution of water molecules and counterions, and canonical form of the duplexes) and compared with results of previous studies to characterizing the quality of our simulation protocol.

3.4 Setup and running MD simulations

In general, a molecular dynamics simulation comprises four steps: minimization, heating, equilibration, and production. In the first phase, the initial configuration of the system (which may be obtained from crystallographic data or a graphically built model) is subjected to energy minimization to relieve any major stress. Velocities are then assigned to each atom to increase slowly the system's kinetic energy until the target temperature is achieved. Equilibration follows, in which velocities corresponding to the target temperature are reassigned constantly, and finally one enters the production stage of the simulation.

Let us consider several details of a MD simulation. We carried out two types of MD simulations on the canonical B-form of DNA fragments. Calculations of the first type were a so-called "constrained simulations", where a DNA duplex was completely kept fixed with positional restraints during the simulation, i.e. it was treated as a rigid duplex. The structure of the DNA duplex was assumed to be ideal; such a structure is characterized all step parameters being zero except for a rise of 3.38 Å and a twist of 36°; see Section 4.2 for a definition of the step parameters. The ideal structure was generated by the program 3DNA[166] using geometries of nucleobases optimized in a DFT calculation at the B3LYP

level.[167] In calculations of the second type, a so-called "unconstrained simulations" were carried out where a DNA oligomer was allowed to move. Starting structures were generated by module NUCGEN of the program AMBER 6.[168] All studied duplexes are listed in Table 3.1. In all cases, missing hydrogen atoms were added to the starting structures using module LEaP of AMBER 6. To relax large steric overlaps and electrostatic inconsistencies, the minimization of 1000 steps was performed for the added hydrogen atoms using SANDER module. Then additional 1500 minimization steps on the whole DNA molecule were done. Next, each fragment was embedded into a rectangular box of TIP3P water[169] using the LEaP module. The water box was extended by ~10 Å in each direction from the DNA duplex; the resulting box sizes and numbers of water molecules as collected in Table 3.2. The systems were then neutralized by adding sodium counterions (Na^+). Note that the number of sodium cations is $2m-2$, where m is the number of Watson–Crick base pairs (see Table 3.2).

The force field described by Cornell et al.[89] was employed in all simulations. These simulations were run using module SANDER with the SHAKE procedure on all hydrogen atoms[102] and a time step of 1 fs. A cutoff of 9 Å was applied to the non-bonding Lennard-Jones interactions. The simulations were performed at constant volume V and total energy E, i.e. for a microcanonical (*NVE*) ensemble. Each simulation consists of three consecutive steps: heating, equilibration, and production run. The heating step for both constrained and unconstrained simulations was done as follows. First, we performed 1000 minimization steps varying the positions of ions and water molecules and keeping fixed the DNA structure. Then molecular dynamics for 25 ps was performed while the temperature was raised from 100 K to 300 K. After this heating step, the equilibration (see below) and production steps were carried out using the particle mesh Ewald method[109] to account for long-range interactions.

For a constrained simulation, equilibration was done by performing molecular dynamics for 100 ps dynamics. The total energy and volume of the system were monitored to make sure that they achieved constant values. After that, the production run was initiated and the coordinates were stored every 1 ps.

For unconstrained simulation, equilibration was performed in two steps. First, the constraints of the duplex were released by minimizations. This allowed also water molecules and counterions to relax around the duplex. In the second step, 100 ps molecular

dynamics was performed. Like in a constrained simulation, the total energy and volume of the system were monitored. After achieving equilibration, the production run was initiated.

For a statistically meaningful representation, long simulation times (of the order of nanoseconds) are essential. Our simulations were carried out up to 12 ns.

In addition, we carried out a molecular dynamics simulation of a neutral duplex 5′- TTGTTTTTTTGTT -3′ of modified DNA, namely with a backbone of methylphosphonate moieties (see Figure 3.3), where one phosphoryl oxygen atom of each phosphate group was replaced with a methyl group. A chirally pure R_P-stereoisomer (a phosphorus center) was obtained. Most studies carried out so far indicate that the R_P-isomer forms more stable complexes than the S_P-isomer.[170–174] Steric, electrostatic and solvent interaction effects are believed to contribute to the increased stability of the R_P-isomer. Since the total charge of the modified duplex was zero, the simulation box contained water molecules, but no counterions. We used the force field of Cornell et al.[89] for the normal part of DNA and the force field of Hamelberg et al.[175] for the modified part of DNA, i.e. the methylphosphonate groups. The MD simulation was done as described above.

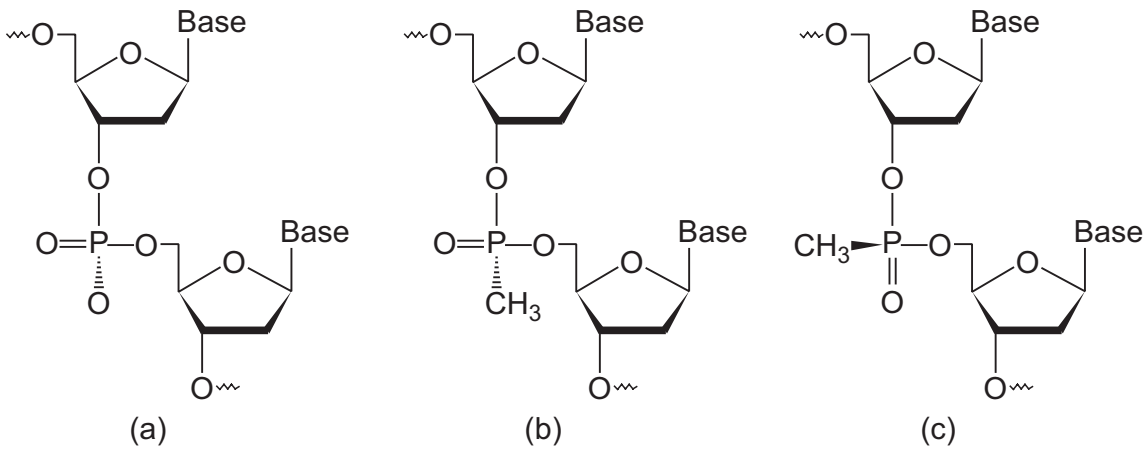

Figure 3.3 Chemical structure of (a) normal DNA, and methylphosphonate modified DNA of (b) R_P-stereoisomer and (c) S_P-stereoisomer.

Table 3.1 List of DNA duplexes for which MD simulations have been carried out.

	Duplex
1	5′- C C A A C G T T G G -3′ 3′- G G T T G C A A C C -5′
2	5′- G G G G G G -3′ 3′- C C C C C C -5′
3	5′- G G G T G G G -3′ 3′- C C C A C C C -5′
4	5′- G G G T T G G G -3′ 3′- C C C A A C C C -5′
5	5′- G G G T T T G G G -3′ 3′- C C C A A A C C C -5′
6	5′- G G G T T T T G G G -3′ 3′- C C C A A A A C C C -5′
7	5′- G G G T T T T T G G G -3′ 3′- C C C A A A A A C C C -5′
8	5′- G G G T T T T T T G G G -3′ 3′- C C C A A A A A A C C C -5′
9	5′- T T T T T G T G T T T T T -3′ 3′- A A A A A C A C A A A A A -5′
10	5′- T T T T T G T T G T T T T T -3′ 3′- A A A A A C A A C A A A A A -5′
11	5′- T T T T T G T T T G T T T T T -3′ 3′- A A A A A C A A A C A A A A A -5′
12	5′- T T T T T G T T T T G T T T T T -3′ 3′- A A A A A C A A A A C A A A A A -5′
13	5′- T T T T G T T T T G T T T T -3′ 3′- A A A A C A A A A C A A A A -5′
14	5′- T T T G T T T T T T G T T T -3′ 3′- A A A C T A A A A A C A A A -5′
15	5′- T T G T T T T T T T T G T T -3′ 3′- A A C A T A A A A A A C A A -5′
16	5′- T T G T T T T T T T T G T T -3′ 3′- A A C A T A A A A A A C A A -5′ (methylphosphonate modified duplex)

Table 3.2 Characteristics of the MD models for the systems listed in Table 3.1.

Duplex	Number of base pairs	Number of counterions	Number of water molecules[a]	Box size[a] ($Å^3$)
1	10	18	3256	46×48×64
2	6	10	1983	44×44×46
3	7	12	2218	44×45×50
4	8	14	2347	44×45×53
5	9	16	2508	44×45×56
6	10	18	2676	44×46×60
7	11	20	2825	44×46×63
8	12	22	2945	44×46×67
9	13	24	3793	46×48×74
10	14	26	3936	46×48×77
11	15	28	4132	46×48×81
12	16	30	4276	46×48×84
13	14	26	3945	46×48×77
14	14	26	3942	46×48×77
15	14	26	3938	46×48×77
16	14	–	4286	48×49×77

[a] For unconstrained systems.

3.5 Analysis of MD results and discussion

3.5.1 MD structure and stability

The root mean square deviation (RMSD) of a MD trajectory as a function of time shows variations of the DNA structure with respect to a reference structure, e.g. the starting structure or an experimental structure. RMSD is computed from the mass-weighted mean square difference of all DNA atoms of each snapshot. Using the starting structure as reference, the averaged RMSD for each system was about 3.1–3.6 Å. These results indicated that the structures were stable along the MD trajectories throughout the full

simulations of several nanoseconds. The oscillations of these RMSDs were ~0.4–0.6 Å. A selected RMSD profile of duplex 5′-CCAACGTTGG-3′ is shown in Figure 3.4. Additionally, RMSDs of MD trajectories were also analyzed with respect to their corresponding averaged structure. The given results were ~12% smaller than those with respect to the initial structure. For instance, a value of 2.92 Å was obtained for the duplex 5′-CCAACGTTGG-3′. This value agrees with that of the study by Cheatham and Kollman[110] (2.90 Å) for the corresponding duplex. Furthermore, analysis of an average structure (of 1000 snapshots of a 1 ns MD trajectory) of this fragment with respect to the crystal structure[176] gave an RMSD of 3.15 Å. This result is consistent with that reported by Cheatham and Kollman (RMSD = 3.14 Å), and agrees rather well with the corresponding X-ray crystal structure.[110]

3.5.2 Analysis of DNA canonical form

The major differences between canonical A-DNA and canonical B-DNA relate to differences in helicoidal parameters[177] of the duplex. The "x-displacement (*dx*)" is the

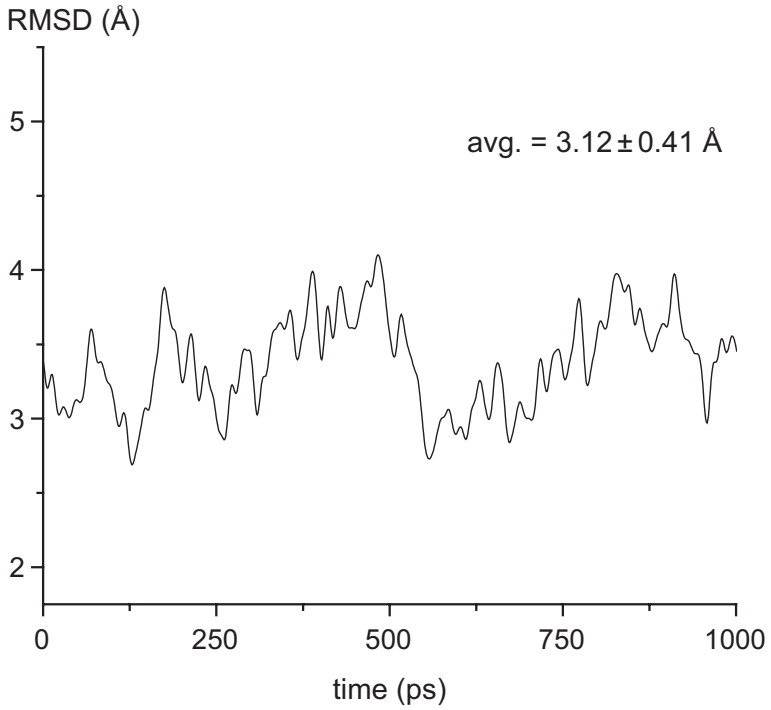

Figure 3.4 Root mean square deviation (RMSD) of duplex 5′-CCAACGTTGG-3′ with respect to the starting structure as a function of time for a 1 ns MD trajectory.

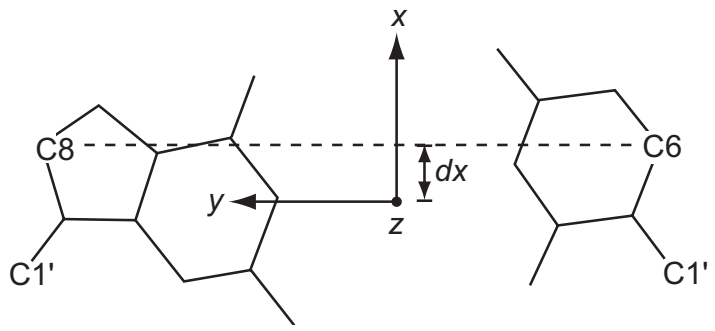

Figure 3.5 Definition of the x-displacement (*dx*) as the distance between the base pair axis (drawn from purine C8 to pyrimidine C6; dashed line) and the helix axis *z* (pointing out of the plane of the paper).

parameter which shows the largest difference for both forms. The x-displacement is defined as shown in Figure 3.5. It is positive if the axis passes by the major groove side of the base pair and negative if it passes by the minor groove (see Section 2.1 for DNA structure). The x-displacement is in the range of –2.0 to 0.0 Å in B-DNA and in the range of –6.0 to –5.0 Å in A-DNA. Another distinguishing characteristic is the minor groove width, measured by close inter-strand phosphate distances. The minor groove width is much larger in A-DNA (~11 Å) than in B-DNA (~6 Å), see Table 2.1.

To analyze the DNA canonical form of our simulations, we extracted a number of snapshots (one for each 500 ps time interval). The structure analysis was done with the program 3DNA.[166] We found that all extracted snapshots of all systems expressed the characteristic of canonical B-form. For instance, an x-displacement of about –0.18 Å and a minor groove width of 6.4 Å were obtained for the duplex 5′-CCAACGTTGG-3′. Since we employed B-DNA as initial structure, one can conclude that we did not observe a transition from B-DNA to A-DNA in the course of our simulations. This is not surprising because the simulations were performed in aqueous solution in which a B-to-A transition is not expected to occur (see Section 3.2.4).

3.5.3 Water and counterion distributions

The most common way to analyze the structure of a solution is to employ the radial distribution function (RDF), denoted by $g_{\alpha\beta}(r)$.[178] The RDF is a dimensionless quantity

that presents the probability of finding an atom of type β at a radial distance r from an atom of type α:

$$g_{\alpha\beta}(r) = \frac{N(r)}{\rho 4\pi r^2 dr} \tag{3.24}$$

Here, $N(r)$ is the average number of type β atoms in a spherical shell of thickness dr at a radial distance r from atom α. ρ is the number density of species β. The average number of β atoms within a sphere of radius r around atom α is obtained from the corresponding running integration number $n_{\alpha\beta}(r)$ defined as

$$n_{\alpha\beta}(r) = \rho \int_0^r g_{\alpha\beta}(r) 4\pi r^2 dr \tag{3.25}$$

The corresponding radius is normally assigned by the solvation shell, at the first minimum of RDF. In general, it is more convenient to understand the solvation pattern around DNA if one analyzes the radial distribution functions of an individual DNA site, e.g. a phosphate group, a sugar group, minor groove and major groove.[112,179,180] In the following, however, only the distributions of water molecules and counterions around DNA are analyzed.

We consider first the distribution of counterions. To eliminate end effects, the radial distribution functions were calculated by ignoring the two terminal base pairs. The results show that, radial distribution functions $g(r)$ referring to the surface of DNA surrounded by ions show a sharp peak of the first solvation shell (the first minimum) at about 2.9 Å. This corresponds to the running integration number $n(r)$ of about 2 ions. Additionally, a second solvation shell at about 5.4 Å with $n(r)$ of 10 Na^+ ions has been observed. These results are consistent with earlier calculations on a dodecamer B-DNA reported by Young et al. using Monte Carlo and MD simulations.[111] Within 10 Å from the closest heavy DNA atom, 70% of all sodium ions are found, compared to a value of 76% found by Feig and Pettitt.[112] The latter carried out an MD simulation of the decamer $d(C_5T_5)\cdot d(A_5G_5)$ by adding 32 additional Na/Cl ion pairs, yielding 1.2 M Na^+ and 0.8 M Cl^-. This may be responsible for the difference of the number of sodium ions found within that distance. In Figure 3.6, we present a plot of the calculated DNA–Na^+ $g(r)$ and $n(r)$ for a MD trajectory of the duplex 5′- TTGTTTTTTTTGTT -3′. Moreover, we found that the counterions localize in the vicinity of negatively charged phosphates; some of them reside in the grooves of DNA

helix.[117] The locations of Na^+ ions, taken from 500 snapshots of a MD trajectory, are superimposed in Figure 3.7.

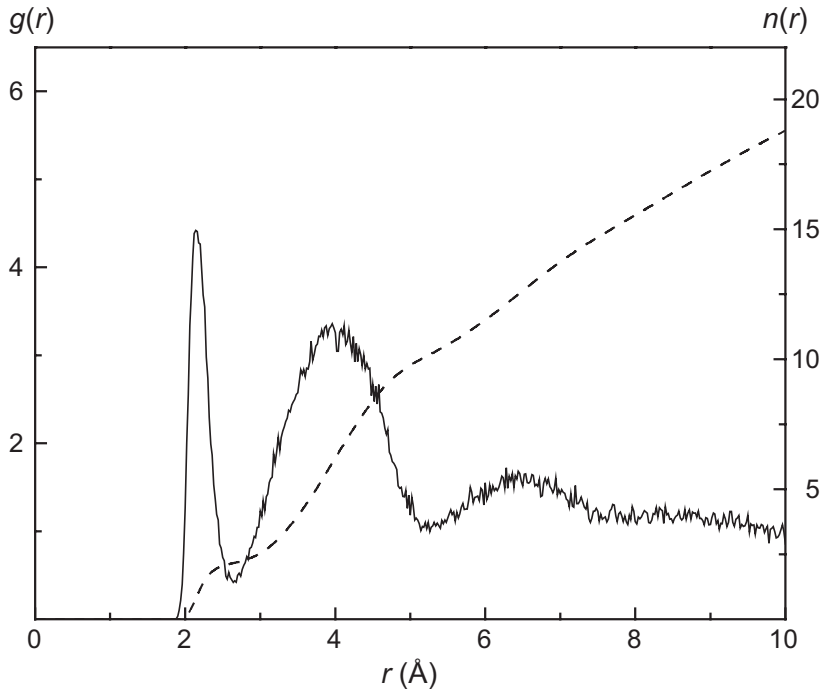

Figure 3.6 Calculated distribution function $g(r)$ (solid line) and integration number $n(r)$ (dashed line) of Na^+ ions around DNA in the duplex 5′- TTGTTTTTTTTGTT -3′.

Next we consider the result concerning the hydration of DNA. Selected results for the calculated RDF of the entire DNA duplex 5′- TTTTTGTTGTTTTT -3′ are shown in Figure 3.8 for the oxygen atoms of water. The calculation was again done without accounting for the two terminal pairs. The plot shows only a first solvation shell, as indicated by a well-defined peak at about 3 Å, but there is no second shell. In addition, $g(r)$ at the first minimum is shallow, see Figure 3.8. This indicates a weak DNA–water interaction (compared to the DNA–Na^+ interaction) which leads to an exchange between water molecules in the solvation shell. The running integration number, integrated up to the corresponding first minimum of 3.4 Å is 410 water molecules; this translated to an average of 17 water molecules per nucleotide. This number is consistent with the value of 18–23 water molecules per nucleotide in the primary hydration shell detected by IR spectroscopy.[47] Note that the hydration pattern and counterion distribution given by

constrained MD trajectories are similar to those of unconstrained simulations, in terms of both the radial distribution function and the integration number.

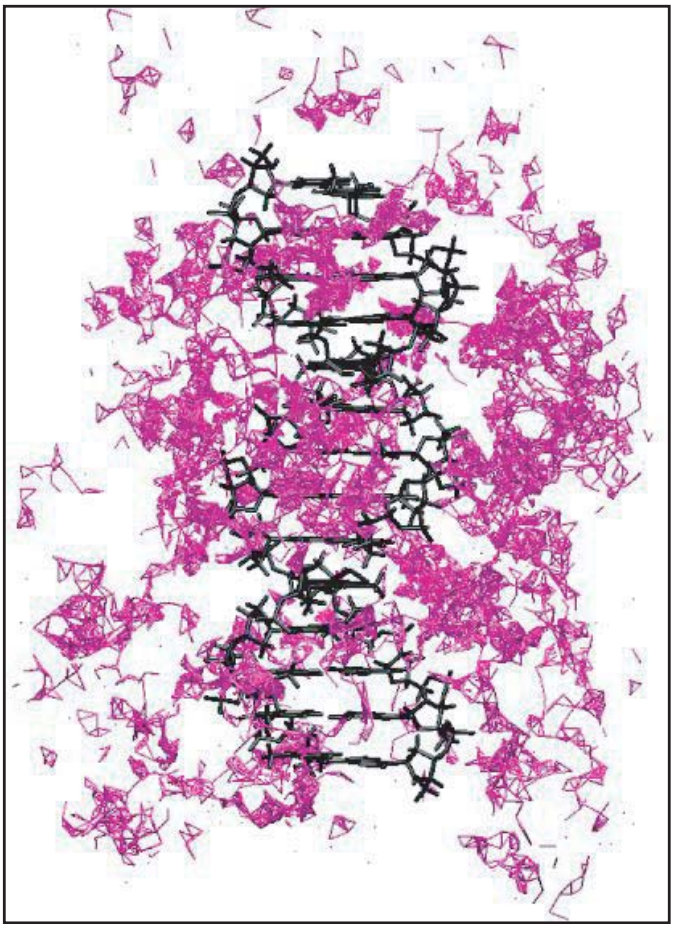

Figure 3.7 Distribution of Na^+ ions over 500 snapshots superimposed on the average DNA structure of the duplex 5′- TTTTTGTTGTTTTT -3′ .

3.6 Conclusion

We carried out molecular dynamics simulations of 16 DNA duplexes in explicit aqueous solution, neutralized by sodium counterions. The simulations were performed for constrained and unconstrained duplexes. Employing canonical B-DNA as initial structure and the AMBER force field, a well-documented methodology and an approved simulation protocol, we obtained stable 2–12 ns trajectories. The results were found consistent with earlier MD simulations. Analysis of the hydration and counterion distribution around DNA

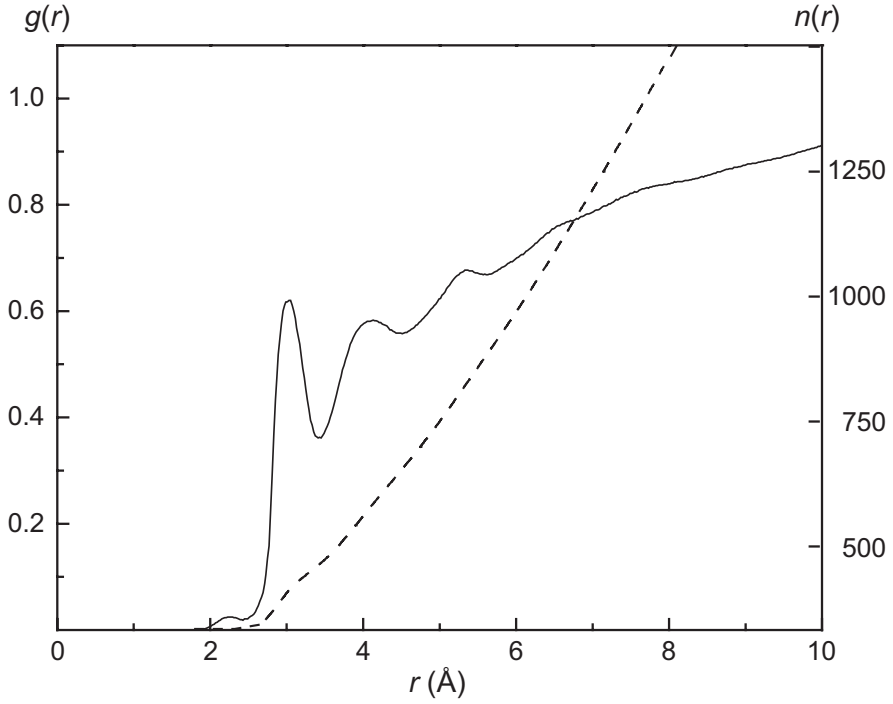

Figure 3.8 Calculated distribution function $g(r)$ (solid line) and the integration number (dashed line) $n(r)$ of water molecules around DNA in the duplex 5′-TTTTTGTT-GTTTTT-3′.

showed good agreement between theoretical and experimental data for both the solvation pattern and the numbers of water molecules and Na^+ ions in the solvation shell.

Furthermore, it is worth recalling important aspects of molecular dynamics simulations of nucleic acids. For accurate simulations, long-range electrostatic interactions have to be calculated accurately and a reliable evaluated force field has to be used. Variations of the simulation conditions can lead to differences in the dynamical behavior of the system, e.g. A- to B-DNA conformational transitions. However, due to the growing interest in MD methodology, improved force fields and more efficient methods for treating long-range interactions are now available.

Chapter 4

Sensitivity of Electronic Coupling on Conformational Changes

4.1 Introduction

As described in Chapter 2, the electronic coupling between nucleobases and relative energetics of their cation radical states are key parameters to describe the hole transfer in DNA. These quantities have recently been calculated using quantum chemical methods.[10,11,33,35,181] While standard semi-empirical methods provide reasonable estimates of the electronic coupling via σ-bridges, these methods considerably underestimate the electronic coupling between π-stacked donor and acceptor. Therefore, the characteristic decay parameter for hole transfer through DNA calculated with a CNDO/2-based scheme[33] was obtained at 1.6 $Å^{-1}$, a value twice as large as those extracted from experiments.[4–6] Previously, electronic coupling matrix elements in model systems containing two nucleobases have been considered.[35] Because of their large dipole moments, nucleobases polarize each other to a considerable extent when they form Watson–Crick pairs. As a result, their electronic properties change. In particular, the formation of such hydrogen-bonded complexes further stabilizes guanine and adenine radical cations.[33,182] To take these effects of nucleobase pairing into account, the electronic coupling between Watson–Crick pairs should be considered.[10] In our previous calculations, we assumed for simplicity that the model systems exhibit ideal structures as determined by averaging many X-ray results of DNA.[47,183] However, it is well recognized that the positions of base pairs in DNA undergo significant fluctuations at room temperature.[47] Molecular dynamics simulations

provide a detailed microscopic description of molecular structure and motion of DNA at finite temperature and in aqueous solution.[98] In particular, molecular dynamics consistently suggest that the motion of nucleobases and sugar-phosphate backbone fragments occurs on a time scale of ~10–100 picosecond.[98] In the following we will consider effects of such conformational fluctuations on the electronic couplings of A·T base pairs as a first attempt to answering the question how sensitive hole transfer is to structural changes. These results will provide deeper insight into the role of orientational coherence in DNA mediated charge transfer.

4.2 Electronic coupling calculation

4.2.1 Model

As described above, bridges of A·T pairs mediate hole transfer between G bases via superexchange. Therefore, estimates of the electronic coupling matrix element between A·T pairs are pertinent to a quantitative description of how such bridges affect the hole transfer. Four arrangements of two A·T pairs may occur in a DNA fragment: two equivalent configurations

$$\begin{array}{l} 5'-A-A-3' \\ \quad\ \ | \quad | \\ 3'-T-T-5' \end{array} \quad \text{and} \quad \begin{array}{l} 5'-T-T-3' \\ \quad\ \ | \quad | \\ 3'-A-A-5' \end{array}$$

where both adenines are in the same strand – we denote them as [(AT),(AT)] and [(TA),(TA)], respectively, – and two different configurations with inter-strand arrangements of the adenine moieties,

$$\begin{array}{l} 5'-A-T-3' \\ \quad\ \ | \quad | \\ 3'-T-A-5' \end{array} \quad \text{and} \quad \begin{array}{l} 5'-T-A-3' \\ \quad\ \ | \quad | \\ 3'-A-T-5' \end{array}$$

represented by [(AT),(TA)] and [(TA),(AT)], respectively.

The mutual positions of base pairs in DNA can be characterized by three translations: rise, shift and slide, and three rotations: twist, tilt, and roll;[47,184] see Figure 4.1. The ideal (or regular) structure corresponds to a configuration of the base pairs where rise

and twist are 3.38 Å and 36°, respectively, while all other step parameters are zero. Previous investigations[10,35] had been carried out at this configuration and, in the following, we shall take it as reference structure. To estimate how changes of each step parameter can affect the electronic coupling matrix element, we generated a series of alternative configurations where only one parameter was changed relative to the reference structure. We used the following increments of the step parameters: rise ±0.5 Å, shift ±0.5 Å, slide ±1.0 Å, twist ±5°, tilt ±2°, and roll ±5°; see Figure 4.1 for the definition of the signs of these parameters. These increments are in line with the standard deviations of the parameters determined from 400 nucleotide steps in X-ray crystal structures of DNA oligomers[183] as well as with the results of our MD simulation of the double-stranded decamer 5′-CCAACGTTGG-3′ (see below). The structures of the models were constructed with the program SCHNArP.[184]

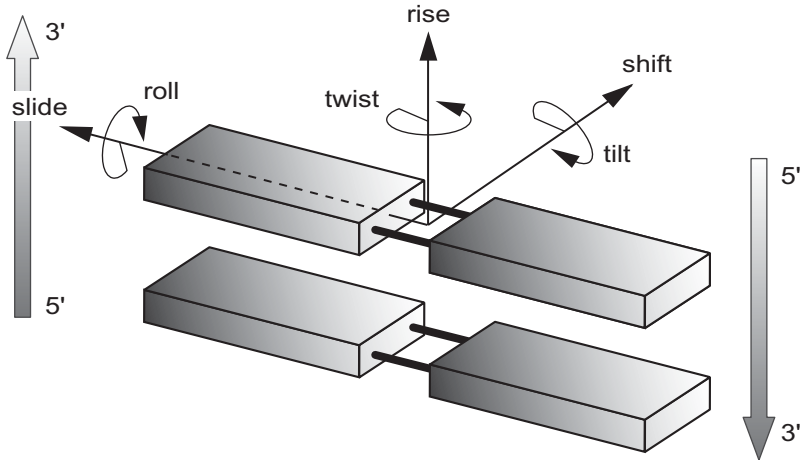

Figure 4.1 The six parameters used to define the conformation of a Watson–Crick base-pair step in DNA.

Different from previous investigations[10,35] where the experimental averaged geometries of individual nucleobases were used,[185] in the present study we employed geometries of nucleobases as optimized by the hybrid density functional method B3LYP.[186,187] We found that electronic couplings calculated for average "experimental" geometries and the corresponding structures optimized at the B3LYP/6-31G* level of theory do not differ in a significant fashion; differences are at most 5%. The present

strategy based on calculated geometries has the advantage that this approach can subsequently be extended to chemically modified nucleobases and other related species.

4.2.2 Electronic coupling

To calculate the electronic matrix elements for hole transfer between Watson–Crick pairs, we applied the two-state model of electron transfer[61,62,188] at the Hartree–Fock SCF level; this theoretical approach had been successfully used in previous investigation on the regular DNA structure.[10] In this procedure, the electronic coupling matrix element is estimated as half of the minimum splitting Δ between two adiabatic states. To find that minimum splitting, where charge transfer is allowed since electron donor and acceptor levels are in resonance, we employed an external electric field oriented along the axis of the DNA duplex.[10,35] If one invokes Koopmans' approximation, then one is able to estimate the energy splitting $\Delta = E_2 - E_1$ between the cation radical states as the difference $\Delta \approx \varepsilon_{\mathrm{HOMO-1}} - \varepsilon_{\mathrm{HOMO}}$ of the one-electron energies of the HOMO (highest occupied molecular orbital) and the adjacent lower lying orbital HOMO−1 of the corresponding neutral closed-shell system[10] (see Section 2.2.1).

The quantum chemical calculations have been carried out with the program Gaussian98 using the standard basis set 6-31G*.[167]

4.3 Results and discussion

4.3.1 Electronic coupling in the reference systems

The calculated electronic coupling matrix elements for hole transfer between two A·T pairs are presented in Table 4.1. The duplexes [(AT),(AT)] ≡ [(TA),(TA)], [(AT),(TA)] and [(TA),(AT)] of regular structure (rise = 3.38 Å, twist = 36°) serve as reference systems. According to the calculations, HOMO and HOMO−1 of these dimer duplexes are formed almost exclusively by adenine orbitals. Thus, hole transfer occurs between these purine nucleobases. Therefore, the intra-strand A–A interaction is responsible for the coupling in

the duplex [(AT),(AT)] while the inter-strand A–A interaction determines the coupling in the duplexes [(AT),(TA)] and [(TA),(AT)]. Comparison of the matrix elements of the reference duplexes shows that the intra-strand A–A interaction, 0.027 eV, is smaller than the inter-strand interactions, 0.056 eV ([(AT),(TA)]) and 0.051 eV ([(TA),(AT)]).

Table 4.1 Dependence of the electronic coupling matrix element (in eV) between adjacent A·T Watson–Crick pairs on the step parameters.[a]

		[(AT),(AT)]	[(AT),(TA)]	[(TA),(AT)]
Ideal structure[b]		0.027	0.056	0.051
Rise	2.88	0.074	0.124	0.098
	3.88	0.010	0.022	0.024
Shift	–0.50	0.124	0.047	0.042
	0.50	0.090	0.048	0.042
Slide	–1.00	0.003	0.027	0.119
	1.00	0.004	0.013	0.016
Twist	31.00	0.029	0.053	0.044
	41.00	0.067	0.058	0.059
Tilt	–2.00	0.039	0.056	0.051
	2.00	0.017	0.056	0.051
Roll	–5.00	0.046	0.042	0.069
	5.00	0.010	0.072	0.037

[a] Translations rise, shift, and slide in Å; rotations roll, tilt, and twist in degree.

[b] Rise = 3.38 Å, twist = 36°, all other parameters are zero.

4.3.2 Structure sensitivity of the intra-strand electronic coupling

Next we consider how conformational changes affect the electronic coupling of A·T pairs in duplexes. We performed calculations of twelve structures constructed with positive and negative increments for each of the six step parameters (Table 4.1). The increments were taken in accord with average deviations of the step parameters as described above.

As expected, the matrix element decreases very rapidly with increasing distance R between two nucleobases. The values range from 0.074 eV (rise 2.88 Å) to 0.010 eV (rise 3.88 Å). The coupling V_{da} may be well approximated by the function $V_{da}(R) = 0.027 \exp[-\alpha(R - 3.38)]$ with α = 2.0 Å^{-1} (V_{da} in eV, rise R in Å). As to other parameters, their effect on the electronic matrix element can hardly be predicted. In fact, a "shift" of a Watson–Crick pair by ±0.5 Å causes a considerable increase of the coupling whereas a "slide" by ±1.0 Å leads to a remarkable decrease.

The helical twist parameter determines the extent of DNA winding. A structure formed by totally unwound fragments with a twist of 0° exhibits a maximum overlap between the molecular orbitals of the two A·T pairs and, as a consequence, it is expected to feature a very strong coupling between the pairs. In fact, the matrix element calculated at this symmetric structure, 0.448 eV, is considerably larger than that for the regular structure, 0.027 eV. However, a partial unwinding, with the twist decreased from 36° to 31°, has only a rather slight influence on the matrix element while an increase of the twist to 41° causes the coupling to increase notably to 0.067 eV.

Variations of the parameters roll and tilt produce similar effects. The coupling rises when the increment of the angles is negative and becomes smaller for rotations in the opposite directions.

4.3.3 Structure sensitivity of the inter-strand electronic coupling

Inspection of Table 4.1 reveals the following results for the coupling matrix elements of the duplexes [(AT),(TA)] and [(TA),(AT)]. Like the intra-strand interaction, the coupling matrix elements are very sensitive to the intermolecular distance. The electronic couplings decrease by a factor of 5 if the rise parameter changes from 2.88 Å to 3.88 Å. The dependence of the inter-strand coupling may approximately be described by $V_{da}(R) = 0.050 \exp[-\alpha(R - 3.38)]$ (V_{da} in eV, R in Å) with α = 1.7 Å^{-1} for [(AT),(TA)] and with α = 1.4 Å^{-1} for [(TA),(AT)].

The matrix elements for both configurations decrease slightly when the shift changes by ±0.5 Å. Therefore, the ideal structure (shift = 0 Å) corresponds to a local

maximum of the electronic coupling with respect to this parameter for both [(AT),(TA)] and [(TA),(AT)]. As can be seen from Table 4.1, the matrix elements are very sensitive to a variation of the slide. Positive increments result in a decreased coupling; for instance, at a slide of 1.0 Å, the matrix elements are reduced to 0.013 eV and 0.016 eV for [(AT),(TA)] and [(TA),(AT)], respectively.

At variance with the intra-strand coupling, inter-strand interactions depend relatively little on the rotational structure parameters. By partial unwinding the duplex as the twist changes from 41° to 31°, the A–A coupling in [(TA),(AT)] decreases from 0.059 eV to 0.044 eV. Increase of the "roll" parameter is accompanied by a notably stronger

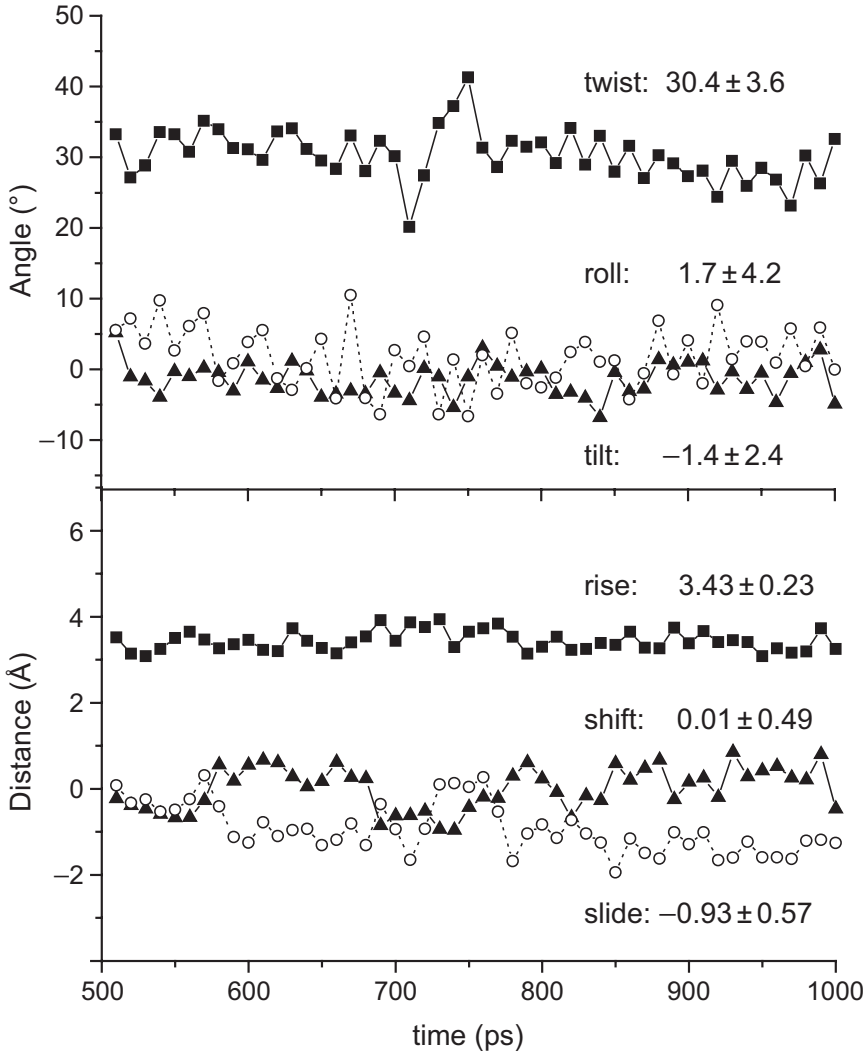

Figure 4.2 Fluctuations of the step parameters of the base steps A_3A_4 in the duplex 5′- CCAACGTTGG -3′ extracted from a MD trajectory of 1 ns.

coupling within [(AT),(TA)] and a smaller coupling within [(TA),(AT)]. On the other hand, the interaction between these nucleobase pairs remains almost unchanged when the tilt angles is varied (Table 4.1).

These results suggest that electronic coupling matrix elements may change considerably depending on the configuration of duplexes. We found that the intra-strand A–A interaction within [(AT),(AT)] is especially sensitive to geometry variations; for the investigated structure changes, it value ranges from 0.003 eV to 0.124 eV. In particular, while in duplexes of regular structure the intra-strand interaction is smaller than the inter-strand coupling, mutual shifting of A·T pairs can reverse this ordering (Table 4.1). If one takes into account that the rate constant for hole transfer depends on the square of the coupling matrix element, such structural changes can alter the transfer rates by factors of several hundred or even thousand for different reasonable conformations of [(AT),(AT)]. Similarly, for inter-strand hole transfer within [(AT),(TA)] and [(TA),(AT)], rate constants of various configurations are anticipated to vary by factors of 100.

These results suggest that the structural dynamics of DNA has to be considered when the couplings between nucleobase pairs in duplexes are to be described in an accurate fashion. In particular, molecular dynamics provides a more quantitative sampling of the various configurations where all step parameters of nucleobase pairs change simultaneously with time.

4.3.4 Molecular dynamics simulated structures

In aqueous solution, the MD model of duplex 5′-CCAACGTTGG-3′ has a stable B-form structure.[110] The RMSD between the average structure of the fragment and the corresponding canonical B-DNA structure amounts to 2.92 Å which agrees very closely with a RMSD value of 2.90 Å found previously.[110] The averaging was performed for all DNA atoms over structure snapshots taken at 1 ps intervals. Detailed inspection of the computational results reveals that the decamer is reasonably equilibrated after about 300 ps.

In the duplex, there are two relevant fragments A_3A_4 and T_7T_8 for which the intra-strand A–A couplings may be calculated. Figure 4.2 demonstrates the fluctuations of the

A_3A_4 step parameters in the course of time. The conformation of this stack exhibits remarkable changes over a period of 100 ps. Similar data were also obtained for the T_7T_8 step. Average base-step parameters of the A_3A_4 step, their standard deviations, and their ranges are collected in Table 4.2. These data represent 500 snapshots of the structures along the MD trajectory from 500 to 1000 ps. For comparison, we present also experimental maximum and minimum values of the step parameters in Table 4.2.[183] As can be seen, the step parameters as obtained from the MD simulation fluctuate considerably; however, the parameters remain inside of the experimental ranges. Extreme values of the parameters are far outside their standard deviations; therefore, the corresponding conformations are very rare and may be neglected.

Table 4.2 Average A_3A_4 base-step parameters, their standard deviations as well as maximum and minimum values obtained from a MD trajectory of the duplex 5′-CCAACGTTGG-3′.

	Rotations (degree)			Translations (Å)		
	Twist	Tilt	Roll	Rise	Shift	Slide
Average	30.4	–1.4	1.7	3.43	0.01	–0.93
Standard deviation	3.6	2.4	4.2	0.23	0.49	0.57
Maximum	41.3	5.2	9.8	3.94	0.85	0.31
Experiment[a]	53.9	6.6	24.8		1.58	2.82
Minimum	20.2	–6.8	–6.6	3.08	–0.96	–0.94
Experiment[a]	19.8	–7.8	–8.6		–0.99	–2.40

[a] Derived from X-ray structures containing 400 base pairs.[183]

Table 4.3 lists the step parameters of the conformations A_3A_4 and T_7T_8 at selected snapshots and the electronic coupling matrix elements calculated for these structures. Given the sensitivity of the electronic coupling, rather large fluctuations of the electronic matrix element V_{da} are expected. In fact, the intra-strand coupling of the A·T pairs in the fragments A_3A_4 and T_7T_8 vary by a factor of thirty and more (Table 4.3). When one follows the step parameters along the duplex over time, one notes their oscillations around

the average values. For instance, a twist above average in one step is compensated by value below average in the following step. Thus, one may expect that a decrease of one transfer matrix element is counteracted by an increase of another coupling. Nevertheless, based on the presented results we conclude that the hole transfer rate (proportional to the effective coupling squared) mediated by a short bridge can change with time by factors of ~100–1000. Unfortunately, the expense of Hartree–Fock SCF calculations prevents a quantitative statistical analysis of this issue at the present computational level.

Table 4.3 Step parameters of the pairs A_3A_4 and T_7T_8 generated by snapshots from an MD simulation of the decamer duplex 5′-CCAACGTTGG-3′ and the corresponding intra-strand A–A electronic coupling matrix element V_{da} of these configurations.

time (ps)		Step parameters[a]						V_{da} (eV)
		Rise	Shift	Slide	Twist	Tilt	Roll	
500	A_3A_4	3.52	–0.23	0.08	33.23	5.18	5.54	0.010
	T_7T_8	3.47	0.49	–0.10	37.75	–2.85	–5.34	0.102
600	A_3A_4	3.46	0.55	–1.25	31.08	1.06	3.87	0.024
	T_7T_8	3.63	0.13	0.21	36.10	–1.36	–4.09	0.041
700	A_3A_4	3.44	–0.63	–0.94	30.10	–3.34	2.72	0.065
	T_7T_8	3.26	–0.07	–0.95	33.24	–1.98	–3.44	0.006
800	A_3A_4	3.30	0.23	–0.83	32.08	0.07	–2.58	0.009
	T_7T_8	3.25	0.19	–0.24	41.83	2.15	–3.59	0.135
900	A_3A_4	3.38	0.16	–1.29	27.34	1.04	4.08	0.010
	T_7T_8	3.19	1.23	0.58	37.58	1.90	–3.44	0.310
1000	A_3A_4	3.25	–0.47	–1.26	32.56	–4.93	–0.05	0.042
	T_7T_8	3.61	0.75	–0.10	37.21	–1.19	–3.82	0.105

[a] Translations in Å, rotations in degree.

4.4 Conclusion

Electronic matrix elements for hole transfer between adjacent Watson–Crick pairs in DNA have been calculated at the Hartree–Fock SCF level for various conformations of the duplexes [(AT),(AT)], [(AT),(TA)], and [(TA),(AT)]. Configurations of the dimer [(AT),(AT)] were also extracted from molecular dynamics simulations of the decamer duplex 5′-CCAACGTTGG-3′. The electronic coupling matrix elements were found to be very sensitive to variations of the mutual position of the Watson–Crick pairs. For instance, the electronic coupling increases by factors between 2 and 3 when the Watson–Crick pairs move toward each other by 0.5 Å, i.e. when the rise parameter decreases from 3.38 Å to 2.88 Å (see Table 4.1). Considerable changes in the electronic coupling have also been found for the parallel displacements shift and slide of two pairs as well as for changes in the winding (twist) of the DNA helix. The intra-strand A–A interaction was calculated to be more sensitive to conformational changes than the corresponding inter-strand coupling.

Taking into account thermal fluctuations of the structures of DNA oligomers (Figure 4.2), the electron transfer coupling between adjacent Watson–Crick nucleobase pairs varies considerably with time. Charge transfer properties of thermally accessible conformers of DNA fragments may differ significantly from those predicted for duplexes of regular structure. The results obtained suggest that structural fluctuations in DNA may have a twofold effect on the rate of charge migration. First, as generally recognized, hole donors and acceptors have to be brought into resonance. Second, a bridge has to exhibit a conformation that is favorable for the electronic coupling. The last assumption is a consequence of the high sensitivity of the electron coupling to conformational changes in DNA as demonstrated by the present study.

The computational approach employed here relies on *ab initio* quantum chemical calculations for estimating the electronic coupling between nucleobase pairs in DNA as well as on a structure sampling based on force field generated trajectories. A straightforward attempt to expand this strategy to a reasonably quantitative description of the variations of the electronic coupling over time is limited by two factors. First, the computational effort of calculating the coupling matrix elements of Watson–Crick pairs by the Hartree–Fock SCF approach is relatively costly. Available semi-empirical methods do not provide a reliable alternative since the underlying wave functions are not accurate

enough as demonstrated by test calculations. As an alternative, one may construct a semi-empirical method that aims at more accurate estimates of electronic matrix elements in order to achieve a more quantitative approach to the time variation of the electronic coupling between nucleobase pairs in DNA. Second, force fields such as AMBER 95 used here are parameterized to reproduce the structure of biophysical systems under standard conditions. They may provide only an approximate description of the structural consequences of localized nucleobase cations on Watson–Crick pairs as well as on the local structure of DNA when a localized electron hole interacts with phosphate groups of the DNA backbone. Such effects likely require a combined quantum mechanics/molecular mechanics approach.[189]

The study described here[190] demonstrated for the first time that structural fluctuations have to be taken into account in a quantitative description of the charge migration in DNA and it provides guidelines for constructing a more sophisticated theoretical approach.

Chapter 5

Estimate of the Reorganization Energy

5.1 Introduction

A recent experimental investigation of the apparently anomalous distance dependence of charge transfer rates in DNA duplexes containing a covalently appended acridinium chromophore revealed that, in addition to the electronic coupling, the solvent reorganization energy can influence the distance dependence of the charge transfer rate in an essential fashion.[191] On the basis of measured activation energies, the reorganization energy λ for different donor–acceptor distances was estimated.[45] In the following we will focus mainly on the calculation of the solvent contribution λ_s for hole transfer in a number of DNA duplexes. Evaluation of the solvent contribution λ_s for electron transfer was considered in detail previously,[48,191] and particular applications to DNA have been reported recently.[13,162]

Tavernier and Fayer[66] studied the effects of dielectric heterogeneity (both in DNA and its environment) on λ_s for charge transfer in DNA, employing model calculations based on a total of four different zones (three distinct dielectric zones surrounding the donor *d* and acceptor *a* sites), and with the electric displacement field approximated by the vacuum field. Note that the term "zone" refers to a region of fixed dielectric constant, which in general may involve spatially disjoint components, e.g. as in the case of zone 1, displayed in Figure 5.1. Subject to this approximation, λ_s may be formulated additively in terms of contributions obtained for individual dielectric zones. The authors compared their results with experimental data of Lewis et al.[192] and concluded that the distance

dependence of λ_s must be taken into account when interpreting the experimental data on DNA-mediated charge transfer in the range R_{da} < 15 Å. These calculations[66] may be expected to overestimate the reorganization energy somewhat because a rather large value of the static dielectric constant, $\varepsilon = 12.4$, was assigned to the base stack zone, in contrast to other estimates in the range of 2 to 4.[193] Furthermore, the geometry of the DNA fragment and its environment were kept fixed,[66] and the possible role of structural fluctuations was not considered. Such fluctuations may be significant for the interaction of DNA with surrounding counterions and water molecules.[193] Thus, for understanding of the charge-transfer kinetics, it is of interest to estimate the significance of these effects.

The physical framework for a quantitative treatment of charge transfer in DNA has also been considered recently by Beratan and coworkers.[12,13] They evaluated λ_s for several DNA-based systems, using a model comprised of two dielectric zones; a DNA zone with $\varepsilon = 4$ ($\varepsilon = 2$ was also considered) and a solvent zone with $\varepsilon = 81$.[13] Once again, a significant increase of λ_s with R_{da} was obtained, roughly linear for the case of one to three A·T base pairs between holes localized on single guanine *d* and *a* sites (1.1–1.5 eV).[13] Solvation energetics was shown to favor localization of the hole, although delocalization over two or three guanines may be energetically feasible, depending on the details of the DNA and its environment.[12]

In the vicinity of DNA, namely in the "bound water" zone, the local dielectric constant deviates from the bulk value as a consequence of the much lower mobility of water molecules which interact with charged and polar groups of DNA. Beveridge et al.[111,162] studied the dielectric behavior of solvent water and found in particular that the average dielectric constant in the zone within 3 Å from DNA is considerably lower than that of the bulk water. The structure of the "bound water" zone was also recently explored in terms of X-ray diffraction data,[194] yielding a picture in which the chain of phosphates is surrounded by two solvation shells of regular geometry. The first layer is partially occupied by counterions, and because of site-specific binding of cations, the structure of this "spine" depends on the DNA sequence.[194]

Here we report calculations of λ_s for hole transport in DNA duplexes, using a heterogeneous dielectric model which exploits recent estimates of zone-specific dielectric constants (both within the DNA itself and within its aqueous environment) obtained from the analysis of molecular dynamics (MD) simulations.[111,162,195] Our calculations employ

the new Delphi Poisson equation solver.[196,197] The relevant structures for DNA and aqueous solvent were taken from MD simulations, which also permitted us to evaluate the effect of structural fluctuations on λ_s. In this study, we used the duplex fragments $5'\text{-}GGGT_nGGG\text{-}3'$ (n = 0–6).

5.2 Details of calculation

5.2.1 Solvent reorganization energy

The calculation of λ_s for hole transfer between donor and acceptor was based on a scheme described previously[73,74] (Section 2.2.2) in which the dielectric function (generally a spatially varying quantity, $\varepsilon(\mathbf{r})$) is treated in a piecewise fashion as constant within each of k zones: ε_j (j = 1–k). λ_s for a charge transfer process is the difference of solvation energies E_{solv} calculated using static and optical dielectric constants, ε_k^{st} and ε_k^{op} (see Section 2.2),

$$\lambda_s = E_{solv}\left(\varepsilon_1^{op}, \varepsilon_2^{op}, ..., \varepsilon_k^{op}, \Delta\mathbf{q}_{if}\right) - E_{solv}\left(\varepsilon_1^{st}, \varepsilon_2^{st}, ..., \varepsilon_k^{st}, \Delta\mathbf{q}_{if}\right) \tag{5.1}$$

where $\Delta\mathbf{q}_{if} = \mathbf{q}_f - \mathbf{q}_i$ represents the difference of the charge distributions for the initial and final states of the *d* and *a* sites.[73,74] The quantities $\mathbf{q}_i$ and $\mathbf{q}_f$ are vectors corresponding to a point charge representation of the charge distributions (one atom-centered charge for each heavy atom, to which the charge of any bonded H atoms has been added). For each hole state, the set of $\mathbf{q}$'s ($\mathbf{q}_i$ or $\mathbf{q}_f$) sums to +1e, where e is the magnitude of the electronic charge. In the results reported below, the $\mathbf{q}$'s are based on a Mulliken population analysis; alternative charges, such as those based on electrostatic potentials, yielded very similar results for λ_s, i.e. larger ~1% compared to that given by Mulliken population analysis. In most cases, the *d* and *a* hole states were confined to single guanine (G) sites, indicated by a designator in italics; e.g., $5'\text{-}GG^{+}GT_nGGG\text{-}3'$ and $5'\text{-}GGGT_nGG^{+}G\text{-}3'$. We also considered some cases where the hole is delocalized over two guanine units.

The solvation energies were calculated using the program Delphi II,[196,197] which employs a finite difference solver of the Poisson equation for systems comprised of multiple zones with different dielectric constants. The sodium counterions were included explicitly in the Delphi II calculations.[76,198]

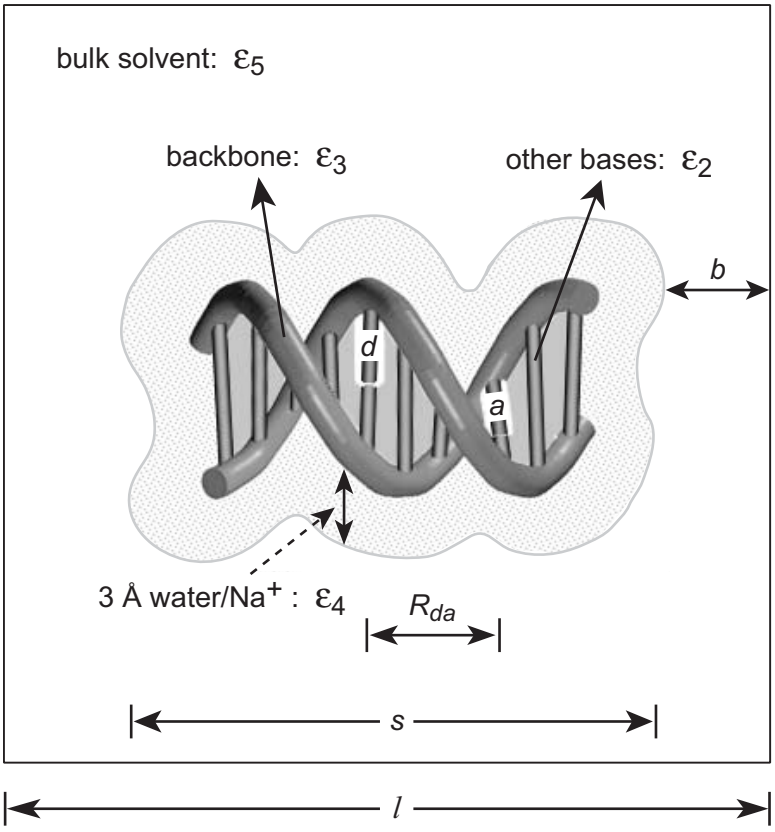

Figure 5.1 Sketch of five dielectric zones: (1) donor (*d*) and acceptor (*a*), (2) other nucleobases of the DNA π-stack, (3) sugar-phosphate backbone, (4) "bound water" within 3 Å from the surface of DNA, (5) bulk water. The following parameters of the dielectric model are shown: box size *l*, length *s* of the system (DNA oligomer and bound water), minimum distance *b* between the system and the edge of the box, and distance R_{da} between donor and acceptor.

Each system under consideration was divided into five dielectric zones as shown in Figure 5.1. For the *d* and *a* zones, the static and optical dielectric constants were taken as $\varepsilon_1^{st} = \varepsilon_1^{op} = 1$. The next two zones corresponded to bases and sugar-phosphate backbone, respectively. Based on previous results,[195] the static dielectric constants of these zones were assigned the values 3.4 and 20.6, respectively. A value of 2 was used for the corresponding optical dielectric constants, $\varepsilon_2^{op} = \varepsilon_3^{op} = 2$. Two types of water zone were considered, to take into account the fact that water molecules near DNA exhibit a reduced mobility compared to bulk water. A zone of "bound water" was constructed as a shell of thickness 3 Å, adjacent to the surface of the DNA fragment (corresponding roughly to the

first hydration shell.[111,194] Because the static dielectric constant of such bound water is not precisely determined, we carried out calculations using five different values of ε_4^{st}: 2, 4, 8, 32, and 80. For the bulk water zone, the value ε_5^{st} = 80 was used. The optical dielectric constant of water was taken as $\varepsilon_4^{op} = \varepsilon_5^{op} = 1.8$.

In all calculations with Delphi II, we used a cubic lattice with 201 grid points in each dimension and a scale parameter of 2 grid points/Å (distance p = 0.5 Å between grid points), yielding a box size of 100 Å. This box size is large enough to accommodate all DNA fragments considered here. The dielectric boundary between DNA and the solvent was taken as the molecular surface, defined by a 1.4 Å probe sphere rolling on atomic spheres with radii R_H = 1.0, R_C = 1.7, R_N = 1.5, R_O = 1.4, R_{Na+} = 1.8, and R_P = 2.0 Å.[75] The calculated λ_s results were based on snapshot structures extracted from MD trajectories, from which averaged λ_s values were then obtained.

5.2.2 Internal reorganization energy

The internal reorganization energy λ_i was calculated at the B3LYP/6-31G(d) level.[167] For radical cation states, the unrestricted Kohn-Sham method was applied. To estimate λ_i for the charge transfer process,[70,71] the following terms were computed for donor and acceptor: (1) energies of neutral species at optimized geometries, $E_0(d)$ and $E_0(a)$, (2) energies of the corresponding radical cations at optimized geometries, $E_+(d^+)$ and $E_+(a^+)$, (3) energies of neutral d and a, calculated at the geometries of the corresponding cation radicals, $E_+(d)$ and $E_+(a)$, and (4) the energies $E_0(d^+)$ and $E_0(a^+)$ of the oxidized states d^+ and a^+ at the geometries of corresponding neutral molecules. Then λ_i for the charge transfer reaction $d^+ + a \rightarrow d + a^+$ is a sum of the reorganization energies of donor and acceptor, $\lambda_i = \lambda_i(d) + \lambda_i(a)$, where d and a were taken as single G·C pair *in vacuo*. The reorganization energy can be defined both from the side of reactants $\lambda_i'(\mathrm{X}) = E_0(\mathrm{X}^+) - E_0(\mathrm{X})$ and from the side of products $\lambda_i''(\mathrm{X}) = E_+(\mathrm{X}) - E_+(\mathrm{X}^+)$, where X = d, a. The averaged energy λ_i(X) is

$$\lambda_i(\mathrm{X}) = \frac{1}{2}\left[E_+(\mathrm{X}) - E_+(\mathrm{X}^+) + E_0(\mathrm{X}^+) - E_0(\mathrm{X}) \right]. \qquad (5.2)$$

5.3 Results and discussion

5.3.1 Sensitivity to parameters of the dielectric model

We performed several tests to check the sensitivity of the λ_s values to parameters of the dielectric model used in the Delphi II calculations: the box length l, related to the shortest distance, b, between the edge of the system and the box boundary (Figure 5.1), and the distance p between grid points. Three B-DNA duplexes of different overall length s and donor–acceptor distance R_{da} were considered:

(*i*) 5′-G*G*GTG*G*G-3′, with R_{da} = 13.5 Å and s = 35 Å

(*ii*) 5′-G*G*GTTTTTTG*G*G-3′, with R_{da} = 30.4 Å and s = 53 Å

(*iii*) 5′-AGTGTG*G*GTTGTTG*G*GTTATT-3′, with R_{da} = 27.0 Å and s = 83 Å.

Each system was placed at the middle of a cube of length $l = 2b + s$, where b ranged from 3 Å to 25 Å. Calculations were performed with grid parameter p ranging from 0.2 Å to 1.6 Å.

Calculated data for the duplex 5′-G*G*GTG*G*G-3′ are given in the Table 5.1. When varying the parameters of the model, we found the results to be very robust, with λ_s remaining almost unchanged when the grid parameter p increased from 0.2 Å to 1.0 Å. As expected, the magnitudes obtained were also insensitive to the distance b over the range studied. Furthermore, the calculations on larger duplexes, namely 5′-G*G*GTTTTTT-G*G*G-3′ and 5′-AGTGTG*G*GTTGTTG*G*GTTATT-3′ (see Table 5.1), also support the conclusion that the calculated values of λ_s are practically independent of the parameters of the model. Changes over a large range in parameter space were at most 1 kcal/mol or below 3% (Table 5.1). However, we do not recommend to use a grid parameter $p > 1.0$ Å.

5.3.2 Solvent reorganization energy

Rigid duplexes

Let us first consider the λ_s results obtained for duplexes of ideal structure (Table 5.2). Note that these values λ_s and their standard deviations reported in tables below were obtained by

Table 5.1 Solvent reorganization energy λ_s (in kcal/mol) calculated with different box lengths *l* and grid parameters *p* for rigid duplexes. Donor and acceptor sites are indicated in italics.

- Duplex 5′- G*G*GTG*G*G -3′

l (Å)	*b* (Å)	λ_s					
		p (Å) = 0.2	0.3	0.5	0.7	1.0	1.6
41	3	43.8	43.9	44.0	44.1	44.2	44.8
45	5	43.8	43.9	44.0	44.0	44.2	44.8
55	10	43.8	43.8	43.9	44.0	44.1	44.8
65	15	43.8	43.8	43.9	44.0	44.1	44.8
75	20	43.8	43.8	43.9	44.0	44.1	44.8
85	25	43.8	43.8	43.9	44.0	44.1	44.8

- Duplex 5′- G*G*GTTTTTTG*G*G -3′

l (Å)	*b* (Å)	λ_s				
		p (Å) = 0.3	0.5	0.7	1.0	1.6
59	3	50.7	50.8	50.9	51.0	51.7
63	5	50.7	50.8	50.9	51.0	51.7
73	10	50.7	50.7	50.9	50.9	51.7
83	15	50.6	50.7	50.8	50.9	51.6
93	20	50.6	50.7	50.8	50.9	51.6
103	25	50.6	50.7	50.8	50.9	51.6

- Duplex 5′- AGTGTG*G*GTTGTTG*G*GTTATT -3′

l (Å)	*b* (Å)	λ_s				
		p (Å) = 0.3	0.5	0.7	1.0	1.6
89	3	49.3	49.4	49.5	49.5	50.1
93	5	49.3	49.4	49.5	49.5	50.1
103	10	49.3	49.3	49.4	49.5	50.1
113	15		49.3	49.4	49.5	50.1
123	20		49.3	49.4	49.4	50.0
133	25		49.3	49.4	49.4	50.0

Table 5.2 Solvent reorganization energy λ_s (in kcal/mol) calculated for hole transfer between guanine units (*G*) within the rigid duplexes 5′-G*G*GT$_n$G*G*G-3′ (n = 0–6), for different values of the dielectric constant ε_4^{st} of the "bound water" zone.

n	R_{da} (Å)	λ_s				
		ε_4^{st} = 2	4	8	32	80
0	3.4[a]	17.4 ± 0.8	21.4 ± 0.4	24.3 ± 0.9	27.9 ± 1.3	29.0 ± 1.5
0	6.8[a]	27.9 ± 0.7	32.4 ± 0.5	35.8 ± 1.0	40.3 ± 1.4	41.8 ± 1.6
0	10.1[a]	33.6 ± 0.8	37.8 ± 0.5	41.2 ± 1.0	45.6 ± 1.4	47.1 ± 1.6
1	13.5	36.2 ± 0.7	40.6 ± 0.4	44.0 ± 0.6	48.2 ± 0.7	50.1 ± 0.8
2	16.9	38.4 ± 0.6	42.9 ± 0.3	46.5 ± 0.5	51.1 ± 0.7	52.7 ± 0.8
3	20.3	39.7 ± 0.7	44.3 ± 0.4	48.0 ± 0.6	52.7 ± 0.9	54.4 ± 1.0
4	23.7	40.8 ± 0.7	45.4 ± 0.4	48.9 ± 0.5	53.5 ± 0.7	55.2 ± 0.7
5	27.0	42.1 ± 0.8	46.7 ± 0.5	50.2 ± 0.7	54.7 ± 1.0	56.3 ± 1.1
6	30.4	42.9 ± 0.6	47.5 ± 0.3	51.0 ± 0.5	55.4 ± 0.9	57.2 ± 1.0

[a] Different donor–acceptor separations were considered for the 6-mer 5′-G$_1$G$_2$G$_3$G$_4$-G$_5$G$_6$-3′ duplex: 3.4 Å (G$_3$G$_4$), 6.8 Å (G$_2$G$_4$) and 10.1 Å (G$_2$G$_5$).

averaging over 100 snapshots selected for every 10 ps of a 1 ns MD trajectory. These include the 6-mer duplex 5′-G$_1$G$_2$G$_3$G$_4$G$_5$G$_6$-3′ with several assumed *d* and *a* sites and associated R_{da} values: 3.4 Å (G$_3$G$_4$), 6.8 Å (G$_2$G$_4$) and 10.1 Å (G$_2$G$_5$). A broad range of assumed values for the dielectric constant of the bound water zone were also considered (ε_4^{st} = 2, 4, 8, 32 and 80). The quantitative influence of this parameter is displayed in Table 5.2. For example, in the 6-mer, λ_s for hole transfer between *d* and *a* sites separated by two base pairs (R_{da} = 10.1 Å) was calculated at 33.6, 37.8, 41.2, 45.6, and 47.1 kcal/mol when ε_4^{st} was 2, 4, 8, 32 and 80, respectively, thus increasing by roughly a factor of 1.4 over this interval of ε_4^{st}. The dependence of λ_s on R_{da} calculated with different ε_4^{st} is shown in Figure 5.2. As expected, a sharp dependence of λ_s on R_{da} is found, most pronounced at short separations (R_{da} < 15 Å), and leveling off beyond 20 Å; e.g., for ε_4^{st} = 32, λ_s was calculated at 27.9 kcal/mol and 48.2 kcal/mol when R_{da} was 3.4 Å and 13.5 Å, respectively. In the following discussion we will focus on the results computed with ε_4^{st} = 32, a value

comparable to the mean of the range of values within the first hydration shell obtained by Beveridge et al.[162] In all cases, the first hydration shell (zone 4) generated by the MD simulations was found to contain about ~10% of the Na^+ counterions.

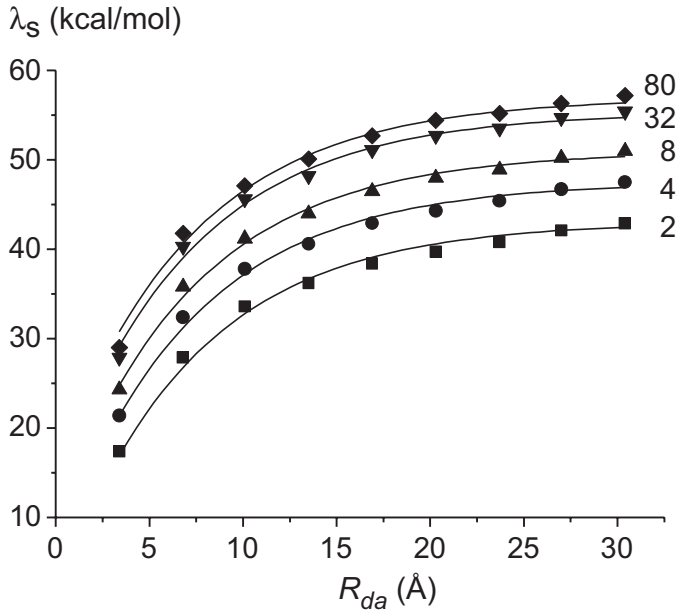

Figure 5.2 Dependence of the solvent reorganization energy λ_s on the distance R_{da} between donor and acceptor sites. Calculated results for different values of ε_4^{st}. The data were fitted with the function $A(\varepsilon_4^{st}) - B\exp(-CR_{da})$, with B = 42 kcal/mol, and C = 0.14 $Å^{-1}$; for values of A, see Table 5.6. For alternative fits in R_{da}^{-1}, see Figures 5.4 and 5.5.

We now consider the degree of localization expected for guanine-based hole states, particularly with respect to the guanine triads dealt with in the present study. There is, of course, a close relationship between localization, solvation, and reorganization effects. Because the redox potentials of the guanine moieties in the GGG triads are similar,[33] the charge can delocalize over more than one guanine. On the other hand, since the polar environment tends to localize the hole,[12] an increase in medium polarity will increase λ_s due to both the larger solvent dielectric constant and the greater degree of localization of the hole. This tendency has been observed in recent theoretical studies which found hole states over one to three adjacent guanines to be energetically accessible at room temperature in the presence of a polar medium.[12] We have estimated the effect of charge delocalization on λ_s for several systems, as illustrated in Table 5.3. In both initial and final

states of the *d* and *a* sites, the hole was assumed to be equally distributed over the first two guanines of the triad (the 5′ and central guanines). An example of results supporting this assignment of charge is given by *in vacuo* HF/6-31G* calculations within the Koopmans' approximation, which yielded a hole distribution in the duplex $TG_1G_2G_3T$ localized on G_1 (39%) and G_2 (59%). For the case of such delocalized holes, the effective donor–acceptor distance can be estimated as the mean value: $R_{da} = (R_{d'a'} + R_{d''a'} + R_{d'a''} + R_{d''a''})/4$, where d' and d'' are two neighboring guanines of the donor site and a' and a'' are neighboring G units of the acceptor site. Comparing λ_s values obtained for localized (Table 5.2) and delocalized (Table 5.3) hole states with common R_{da} values and with $\varepsilon_4^{st} = 32$, one notes that charge delocalization leads to a consistent decrease of the reorganization energy by ~12 kcal/mol, almost independent of the donor–acceptor separation. For instance, for R_{da} = 16.9 Å we calculated λ_s at 51.1 and 39.6 kcal/mol for localized and delocalized holes, respectively.

Table 5.3 Effect of hole delocalization on the solvent reorganization energy λ_s (in kcal/mol), calculated with $\varepsilon_4^{st} = 32$ for the rigid duplexes 5′-*GG*G(T)$_n$*GG*G-3′, n = 0–6.[a]

n	R_{da}[b] (Å)	λ_s
0	6.8[c]	27.1 ± 0.8
0	10.1[c]	33.5 ± 0.8
1	13.5	37.0 ± 0.5
2	16.9	39.6 ± 0.5
3	20.3	41.1 ± 0.5
4	23.7	42.0 ± 0.5
5	27.0	42.8 ± 0.6
6	30.4	43.8 ± 0.6

[a] The hole is distributed equally over adjacent guanines in the *GG* donor and acceptor sites.

[b] R_{da} is the mean distance between the *GG* donor and acceptor sites.

[c] In the 6-mer 5′-$G_1G_2G_3G_4G_5G_6$-3′ duplex, two donor–acceptor separations were considered: R_{da} = 6.8 Å between (G_2G_3) and (G_4G_5) and R_{da} = 10.1 Å between (G_1G_2) and (G_4G_5).

In other calculations we have examined the sensitivity of λ_s for a given R_{da} value to the chemical nature of the *d* and *a* bases, finding that for adenine bases, λ_s is consistently larger, by ~1–2 kcal/mol, than the λ_s values obtained for guanines; e.g., at R_{da} = 10.1 and ε_4^{st} = 32, the results are 46.9 kcal/mol for adenine (Table 5.4) and 45.6 kcal/mol for guanine (Table 5.2). However, hole states localized on adenine may be appreciably higher in energy than those localized on guanine, and in such a case the nature of their role in thermal hole transport would require a detailed kinetic analysis.[199]

Table 5.4 Solvent reorganization energy λ_s (in kcal/mol) for varying ε_4^{st} of the rigid duplex $3'\text{-}CCCA_4A_5A_6A_7A_8A_9CCC\text{-}5'$, calculated for different values ε_4^{st} of the dielectric constant of the "bound water" zone (see text). Various adenine bases act as donor *d* and acceptor *a*.

d–*a*	R_{da} (Å)	λ_s				
		ε_4^{st} = 2	4	8	32	80
A_6–A_7	3.4	18.2 ± 0.7	22.1 ± 0.5	24.8 ± 0.8	28.3 ± 1.3	29.5 ± 1.6
A_5–A_7	6.8	28.8 ± 0.8	33.2 ± 0.5	36.5 ± 0.9	40.9 ± 1.5	42.2 ± 1.9
A_5–A_8	10.1	34.8 ± 0.8	39.2 ± 0.6	42.3 ± 1.0	46.9 ± 1.5	48.5 ± 2.0
A_4–A_8	13.5	37.9 ± 0.9	42.1 ± 0.6	45.5 ± 1.0	49.7 ± 1.6	51.3 ± 2.1
A_4–A_9	16.9	40.3 ± 0.9	44.7 ± 0.7	48.2 ± 1.1	52.6 ± 1.7	53.9 ± 2.1

The calculations discussed so far involved fixed structures for the DNA fragments, and the small fluctuations in λ_s magnitudes (Table 5.2) are due to fluctuations of the DNA environment. More specifically, the number of water molecules found within the cutoff distance of 3 Å (the "bound water" zone) were 170 ± 5, 195 ± 5, 223 ± 5, 251 ± 7, 278 ± 7, 302 ± 8, and 328 ± 6 for duplexes $5'\text{-}GGGT_nGGG\text{-}3'$, n = 0, 1, ..., 6, respectively. These fluctuations lead to small changes of the zone boundaries of our model (Figure 5.1) and thus, to variations of λ_s.

Flexible duplexes

Unconstrained MD simulations, in which structural fluctuations of the DNA fragments were allowed, yielded λ_s values within ~1 kcal/mol (Table 5.5) of the corresponding values based on rigid ideal DNA structures. Taking into account structural changes of the duplexes and further averaging of results allows to minimize any numerical inaccuracy of the Delphi solver because of different placement of the DNA duplex relative to the grid.[76] Furthermore, the standard deviations due to fluctuations in the unconstrained cases are quite similar in magnitude (≤ 1 kcal/mol) to those for the rigid systems (Table 5.2), where only the solvent fluctuations contribute. The following discussion, therefore, deals only with the results obtained for rigid duplexes.

Table 5.5 Solvent reorganization energy λ_s (in kcal/mol) calculated for hole transfer between two guanine units *G* within the flexible duplexes 5′-G*G*GT$_n$G*G*G-3′ (n = 0, 2, 4, 6), for different values of the dielectric constant ε_4^{st} of the "bound water" zone.

n	R_{da} (Å)	λ_s				
		ε_4^{st} = 2	4	8	32	80
0	3.4 ± 0.3[a]	18.7 ± 1.2	22.7 ± 1.0	25.7 ± 1.1	29.3 ± 1.4	30.4 ± 1.5
0	6.8 ± 0.5[a]	28.8 ± 1.2	33.4 ± 1.0	36.9 ± 1.0	41.3 ± 1.3	42.9 ± 1.4
0	10.2 ± 0.5[a]	34.2 ± 1.2	38.7 ± 0.9	42.3 ± 1.0	46.8 ± 1.3	48.4 ± 1.4
2	16.9 ± 0.5	38.7 ± 1.1	43.3 ± 0.7	46.9 ± 0.8	51.4 ± 1.2	53.0 ± 1.3
4	23.6 ± 0.6	40.7 ± 1.1	45.4 ± 0.7	48.9 ± 0.8	53.6 ± 1.2	55.3 ± 1.3
6	30.2 ± 0.6	42.7 ± 1.0	47.3 ± 0.6	50.9 ± 0.7	55.3 ± 1.0	57.1 ± 1.1

[a] Various donor–acceptor separations were considered for the 6-mer 5′-$G_1G_2G_3G_4G_5G_6$-3′ duplex: 3.4 Å (G_3G_4), 6.8 Å (G_2G_4) and 10.1 Å (G_2G_5).

Contribution of different dielectric zones

The λ_s calculations reported here, based on a multi-zone dielectric model implemented with Delphi II,[196,197] do not permit a rigorous additive partitioning into contributions from the individual zones, in contrast to previous models in which the displacement field was

approximated by the vacuum field.[66] That work treated a multi-zone case additively by combining results obtained from two-zone calculations based on the original Delphi code.[200] Contributions for isolated zones are not well-defined, since the displacement field in one zone depends on the details of the full multi-zone dielectric system. These "image effects" are implicitly included in calculations carried out with the Delphi II code, which uses a scalar charge density/electrostatic potential formulation, equivalent to one based on displacement fields.

Bearing in mind the above considerations, we have nevertheless explored approximate additivity schemes which might be useful in the present context. Consider the following expression,

$$\lambda_s = \sum_{j=1}^{4} \lambda_s(j) \tag{5.3}$$

where

$$\lambda_s(j) = E_{\text{solv}}\left(\varepsilon_j^{\text{op}}, \varepsilon_{j+1}^{\text{op}}, \Delta\mathbf{q}_{\text{if}}\right) - E_{\text{solv}}\left(\varepsilon_j^{\text{st}}, \varepsilon_{j+1}^{\text{st}}, \Delta\mathbf{q}_{\text{if}}\right) \tag{5.4}$$

and where the parameters in Eq. (5.4) are the same as defined for Eq. (5.1). In the simple case of a set of nested spherical dielectric zones, in which the scalar $\Delta\mathbf{q}_{\text{if}}$ represents the shift in magnitude of a single point charge at the center, Eqs. (5.3) and (5.4) give an exact partitioning of the total λ_s into contributions from each $j/j+1$ zone interface (j = 1–4). Applied, for example, to the rigid duplex 5′- G*G*GTTG*G*G -3′ included in the present study (with $\varepsilon_4^{\text{st}}$ = 32), we obtained $\lambda_s(j)$ values, 22.1, 18.5, 3.6, and 0.4 kcal/mol, respectively, for j = 1–4. The sum of these contributions (44.6 kcal/mol) is within 15% of the value 51.3 kcal/mol given by the full calculation, Eq. (5.1). Note that the result of 51.3 kcal/mol, based on a single "snapshot" structure, is very close to the mean value given in Table 5.2 (51.1 kcal/mol). It warrants further study to explore how useful Eqs. (5.3) and (5.4) or other partitioning schemes may be in the present context.

Decay parameter

Experimental studies of the distance dependence of hole transfer through DNA indicate[4–6] that the corresponding rate constants in several cases exhibit an overall exponential decay with donor–acceptor distance, R_{da}:

Table 5.6 Non-linear fitted parameters and their standard deviations (S. D.) of λ_s as function of R_{da} (Figure 5.2), $\lambda_s = A_i - B\exp(-CR_{da})$, i = 1–5, where B = 42 kcal/mol, C = 0.14 Å^{-1}, and A_i (in kcal/mol).

ε_4^{st}	A_i	S. D.
2	43.0	0.5
4	47.5	0.4
8	51.0	0.4
32	55.3	0.6
80	57.0	0.7

$$k(R_{da}) = k_0 \exp(-\beta R_{da}) \tag{5.5}$$

where the decay parameter β depends on the nature of the *d* and *a* sites and the intervening bridge. Since the electronic coupling and the reorganization energy participate in separate factors in Eq. (2.1), the parameter β in Eq. (5.5) can be expressed approximately as a sum of two terms $\beta = \beta_{el} + \beta_s$. Note that the Boltzmann factor in Eq. (2.1) contains only a reorganization contribution, because $\Delta G^o = 0$ for the thermoneutral charge shift processes dealt with here. It is of interest to know the extent to which β parameters may be viewed as global constants over a broad range of R_{da}, an issue which may be addressed on the basis of theoretical calculations. The parameter β_{el} can be estimated on the basis of quantum chemical calculations of the electronic coupling matrix elements.[11,201] Here we consider the second parameter, β_s, which reflects the distance-dependence of λ_s. For the purpose of numerical fitting, this dependence on R_{da} is accurately expressed as (see Figure 5.2) $\lambda_s(\text{kcal/mol}) \approx A_i - B\exp(-CR_{da})$, where R_{da} is in Å. The parameters A_i (i = 1–5 for ε_4^{st} = 2, 4, 8, 32 and 80), B, and C were obtained simultaneously using a non-linear least-squares fitting procedure. The resulting values are A_i = 43.0, 47.5, 51.0, 55.3, and 57.0 kcal/mol for ε_4^{st} = 2, 4, 8, 32 and 80, respectively, B = 42 kcal/mol, and C = 0.14 Å^{-1} (see Table 5.6). We may now express β_s as: $\beta_s = -d\ln k'_{da}/dR_{da}$, where k'_{da} denotes the Boltzmann factor on the right-hand side of Eq. (2.1). Since $\Delta G^o = 0$ and λ_i is taken as independent of R_{da} (see below), the resulting expression for β_s is $\beta_s(\text{Å}^{-1}) \approx 2.5\exp(-0.14R_{da})$. For systems with one intervening base pair between *d* and *a* sites (R_{da} ~6.8 Å), β_s is quite large (~1.0 Å^{-1}),

but it is nearly an order of magnitude smaller (0.15 $Å^{-1}$) for systems containing five intervening pairs between *d* and *a* sites (R_{da} ~20 Å).

5.3.3 Internal reorganization energy

The internal reorganization energy was calculated at the B3LYP/6-31G* level for A·T and G·C units. For hole transfer between two G·C base pairs, we obtained λ_i = 16.6 kcal/mol; hole transfer between two A·T pairs yields a notably smaller value, λ_i = 10.2 kcal/mol.

Since $\Delta G^o = 0$, the activation barrier for the charge shift process can be taken as $(\lambda_s + \lambda_i)/4$; see Eq. (2.2). Combining λ_i = 16.6 kcal/mol with the λ_s values from Table 5.2 (for ε_4^{st} = 32) yields activation energies for charge transfer in the duplex $5'\text{-}GGGT_nGGG\text{-}3'$ ranging from 16.2 kcal/mol for *n* = 1 to 18.0 kcal/mol for *n* = 6.

5.3.4 Comparison with other λ_s estimates

Figure 5.3 compares λ_s values for hole transfer between GGG moieties in DNA, calculated with different dielectric models, all with hole states confined to single guanine moieties. The estimated values vary considerably among the different schemes. For systems with a *d*–*a* separation of ~10 Å (two intervening nucleobase pairs), λ_s ranges from ~31 kcal/mol, calculated by Tong et al.,[13] to ~69 kcal/mol, as calculated by Tavernier and Fayer,[66] whereas the present calculations yield the value 45.6 kcal/mol (with ε_4^{st} = 32). As already discussed, delocalization of the hole over two bases can decrease the reorganization parameter by ~12 kcal/mol, yielding results that happen to be rather close to those of Tong et al. (see Figure 5.3), which involved a notably different (two-zone) model with holes localized on single guanine bases.[13] For systems with at least one intervening nucleobase pair between *d* and *a* sites, and with hole states localized on single guanines ($R_{da} \geq 7$ Å), the λ_s values calculated in the present work based on estimates of dielectric constants obtained from analysis of MD simulations[162,195] are systematically larger by ~15 kcal/mol than the corresponding values obtained by Tong et al.[13] while at the same time being systematically smaller by ~25 kcal/mol than the results of Tavernier and Fayer.[66] These differences may be attributed to differing definitions of the dielectric zones (both the

structure and the dielectric constants) used to describe the response of the DNA and its environment to the charge transfer process.

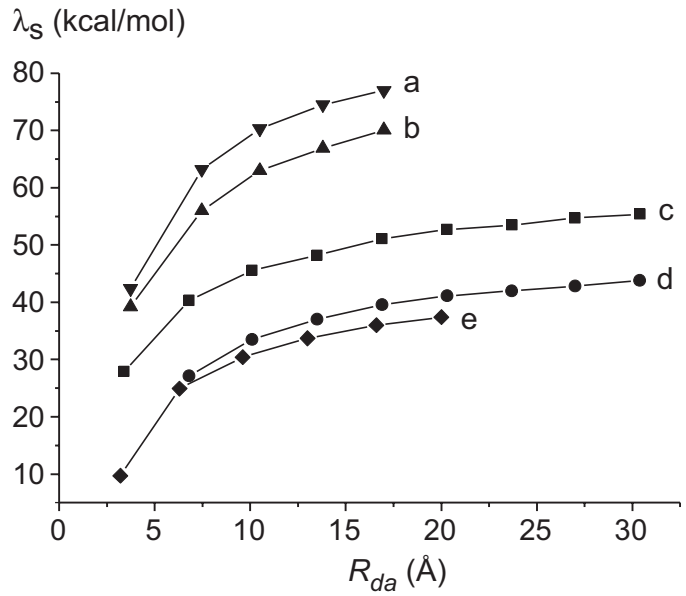

Figure 5.3 Solvent reorganization energies calculated with different dielectric models for charge transfer between nucleobases in DNA: a and b – results of Tavernier and Fayer[66] calculated with ε_2^{st} = 12 and 4 for the nucleobase zone, respectively, c and d – data obtained in the present study with ε_4^{st} = 32 for localized and delocalized holes, respectively, and e – data obtained by Tong et al.[13] using the ε^{st} = 4 for all DNA zones.

We have emphasized the sensitivity of the calculated values of λ_s to the choice of the dielectric constants, in particular for the "bound water" zone (see Figure 5.2). Although computational studies of DNA provide estimates of dielectric constants for different zones of solvated DNA,[162,195] significant uncertainties remain concerning the construction of a multi-zone dielectric model, including the degree to which distinct zones may be meaningfully distinguished. In the present five-zone model, the sugar and phosphate fragments in DNA are treated as one zone with mean dielectric constant set to 20.6, whereas distinct zones (with dielectric constants of 2 and 33, respectively) were identified in Ref. 195. Tavernier and Fayer[66] employed a four-zone model in which all water was treated as bulk, and Tong et al.[13] assigned a mean dielectric constant to the entire DNA duplex. Of course, ultimately one desires a molecular-level treatment beyond the dielectric

continuum framework. Similar to our calculations, Ratner and co-workers[202] have recently studied λ_s of synthetic DNA hairpins with a stilbene chromophore. However, in that paper, neither the various zones nor the dielectric constants were specified. Also, the van der Waals atomic radii used[203] (2.9, 3.3, 3.2, 3.0 and 3.8 Å for atoms H, C, N, O and P, respectively) are considerably larger than the corresponding parameters used by us (1.0, 1.7, 1.5, 1.4 and 2.0 Å, respectively). As a result, Ratner et al. calculated smaller reorganization energies than in our work: their values increase from 5 kcal/mol to 25 kcal/mol when R_{da} increases from 3.6 Å to 16.8 Å. Although these values agree better with experimental findings,[6,69] there is no independent justification for such a choice of radii.

Another issue concerns the spatial extent of the different dielectric zones, especially the *d* and *a* sites. It is thus of interest to consider simple models which help to gauge the effective size of these latter sites. We first attempted to model the λ_s data in terms of the simple two-sphere model of Marcus, where λ_s is given as (see Eq. (2.9))[72]

$$\lambda_s = \frac{\Delta q^2}{2}\left(\frac{1}{r_d}+\frac{1}{r_a}-\frac{2}{R_{da}}\right)\left(\frac{1}{\varepsilon^{op}}-\frac{1}{\varepsilon^{st}}\right) \tag{5.6}$$

Here r_d and r_a are the effective radii of donor and acceptor and Δq is the magnitude of the point charge which is transferred from the center of one sphere to the other. The formula is applicable to non-intersecting spheres, $R_{da} > (r_d + r_a)$. Aside from the structural simplification entailed in the two-sphere model for present purposes, we note that corrections of the form $C_4 / R_{da}^4 + C_6 / R_{da}^6 + ...$ should be added to Eq. (5.6) when the spheres come into contact or begin to overlap.[204] Inclusion of the first correction term allows all of the present λ_s data (including the contact case at R_{da} = 3.4 Å) to be fitted accurately (Table 5.7 and Figure 5.4). As noted above, the exponential form is quite accurate for describing the calculated dependence of λ_s on R_{da} (Table 5.6) and convenient from a practical point of view, although no physical justification can be given. According to Eq. (5.6), λ_s should decrease linearly with $1/R_{da}$. Previous computational studies involving *d*–*a* systems departing strongly from a simple two-sphere geometry[73,74,205] reveal, nevertheless, that the form of Eq. (5.6) may account quite well for the R_{da} dependence of λ_s while at the same time providing definitions of effective *d* and *a* radii. Figure 5.5 shows that, indeed, the λ_s data calculated with the present five-zone dielectric model is linear in $1/R_{da}$ for $R_{da} \geq 6.8$ Å (see also Table 5.8). Moreover, the Marcus two-

sphere model can be employed to rationalize why delocalization of the hole over two guanine units reduces λ_s in a manner almost independent of the R_{da}. From this point of view delocalization of the transferring charge over neighboring bases would imply an increase of the effective radii of *d* and *a* sites (r_d and r_a , respectively).

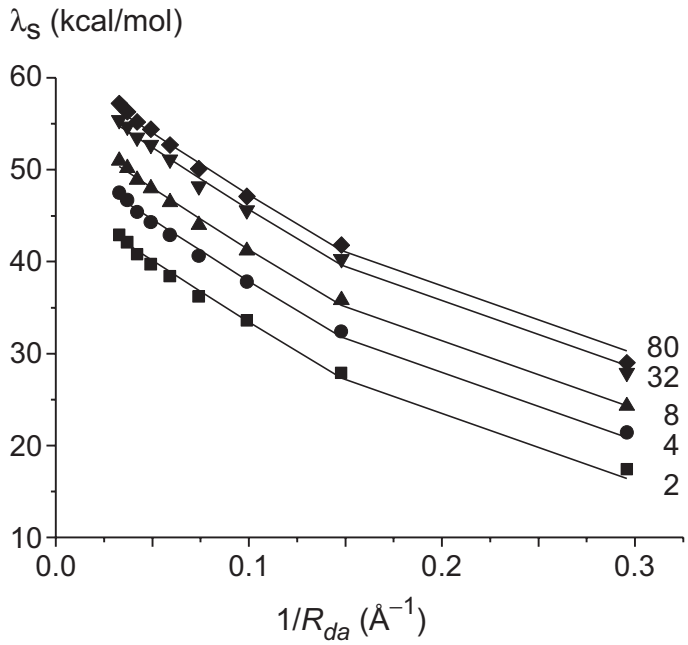

Figure 5.4 Non-linear fit of $\lambda_s = F_i - \frac{G}{R_{da}} + \frac{H}{R_{da}^4}$ as function of R_{da}^{-1}; the distance R_{da} = 3.4 Å was taken into account. Fitting parameters are listed in Table 5.7.

Table 5.7 Results of the non-linear fit of λ_s as function of R_{da}^{-1} (see also Figure 5.4); $\lambda_s = F_i - \frac{G}{R_{da}} + \frac{H}{R_{da}^4}$, i = 1–5, where G = 137 kcal/mol·Å, H = 1299 kcal/mol·Å^4, and F_i (in kcal/mol).

ε_4^{st}	F_i	S. D.
2	47.0	0.3
4	51.4	0.3
8	54.9	0.2
32	59.3	0.3
80	60.9	0.5

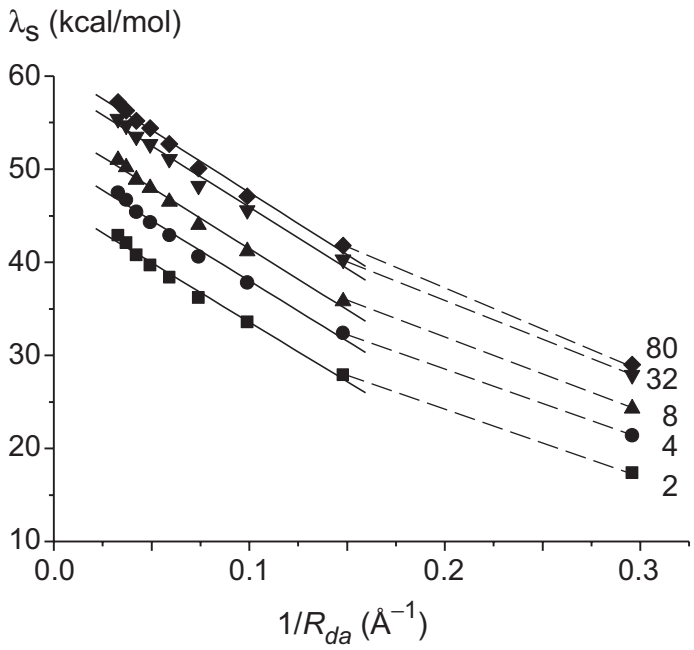

Figure 5.5 Calculated solvent reorganization energy λ_s as a function of R_{da}^{-1} and the corresponding linear fit of function $\lambda_s = D_i - E_i R_{da}^{-1}$ (based on data for $n > 0$). Fitting parameters are listed in Table 5.8.

Table 5.8 A linear fit of λ_s as function of R_{da}^{-1} including all cases except contact (R_{da} = 3.4 Å), see also Figure 5.5; $\lambda_s = D_i - E_i R_{da}^{-1}$, i = 1–5, where D_i (in kcal/mol) and E_i (in kcal/mol·Å).

ε_4^{st}	D_i	E_i	S. D.
2	46.3	127.6	0.5
4	51.0	129.6	0.6
8	54.6	131.1	0.6
32	59.1	131.9	0.7
80	60.9	133.9	0.6

Returning to the case of localized guanine holes, we find that the fits displayed in Figure 5.5 in conjunction with Eq. (5.6) yield (to within 10%) $r_d = r_a$ = ~3 Å, where the dielectric factor (the last factor in Eq. (5.6)) has been assigned a value of 0.5; this is an approximate effective value representing the heterogeneous environment of the *d* and *a*

groups. Similar estimates are obtained from model two-zone Delphi calculations of either the solvation energy for a single guanine cation or λ_s for hole transfer between a pair of guanines. In each of these model calculations, the guanines were given the same molecular structure as employed in the five-zone calculations described above, and were immersed in a homogeneous aqueous solvent (ε^{st} = 80 and ε^{op} = 1.8). Fitting the Delphi results for solvation energy and λ_s, respectively, to one-sphere (Born) or two-sphere (Eq. (5.6)) models yield once again an effective guanine radius of ~3 Å. A value of this magnitude is appreciably larger than the radii (1.87 Å) assigned by Tavernier and Fayer to the spheres which constituted the *d* and *a* sites in their calculations;[66] this difference may help to explain why their λ_s values are appreciably larger than the present values.

Finally, we note a previous analysis[205] of effective radii for atoms in solvated species. That study suggested to employ distinct radii for high-frequency and low-frequency medium response. The resulting effective radii for low-frequency response were appreciably greater than the scaled van der Waals radii used for the high frequency part of the reaction field. This resulted in the corresponding reduction of λ_s values.

5.4 Conclusion

We carried out a computational study of the solvent reorganization energy λ_s for hole transfer through several DNA fragments with different donor–acceptor distances. We estimated λ_s as the difference of the solvation free energies, calculated with static and optical dielectric constants, by solving the Poisson equation for models comprised of five different dielectric zones. We showed λ_s to be rather sensitive to the parameters of the model, noting in particular, the influence of the dielectric constant (ε_4^{st}) used for the "bound water" zone in the immediate vicinity of the DNA. We found λ_s for hole transfer between guanine units to increase rapidly at short donor–acceptor distances R_{da} < 15 Å. The R_{da} dependence of λ_s (for ε_4^{st} = 32 and with hole states confined to single guanines) can be accurately fitted by $\lambda_s(\text{kcal/mol}) \approx 55 - 42\exp(-0.14R_{da})$, with R_{da} in Å. The corresponding falloff parameter then becomes $\beta_s(\text{Å}^{-1}) \approx 2.5\exp(-0.14R_{da})$, varying from ~1.0 Å^{-1} for a system with one intermediate base pair between *d* and *a* sites, to 0.15 Å^{-1} for

systems with five intervening pairs. Delocalization of the hole states over two neighboring guanines causes λ_s to decrease by ~12 kcal/mol, almost independent of the *d*–*a* separation. The internal reorganization energy λ_i for hole transfer between G·C pairs was calculated at 16.6 kcal/mol. The resulting calculated activation barrier for charge transfer, $(\lambda_s + \lambda_i)/4$, increases only slightly with R_{da}, from 16.2 kcal/mol with one base pair between *d* and *a* to 18.0 kcal/mol for six intervening base pairs.

The distance dependence of λ_s found in the present and other recent theoretical studies[13,66] is qualitatively consistent with the Arrhenius analysis of experimental kinetic data for hole transfer between an intercalated acridine derivative (hole donor) and 7-deazaguanine (hole acceptor) separated by one or two (A·T) base pairs.[45] On the other hand, no significant distance dependence of λ_s was indicated in the analysis of isothermal hole transfer in the hairpin duplexes studied by Lewis et al.[6] Furthermore, in the present calculated magnitudes of λ_s are appreciably larger (30–50 kcal/mol for n = 0–3 intervening base pairs and ε_4^{st} = 32) than the estimates inferred from experiment (~10–40 kcal/mol for n = 0–3).[6,45] Our data were calculated for thermoneutral charge transfer in unperturbed DNA duplexes and therefore cannot be compared in a fully quantitative fashion with kinetic data gained for systems with hairpin or intercalated chromophores. Clearly, solvent reorganization energies of charge transfer in DNA deserves further studies, both experimentally and theoretically.

Chapter 6

Environmental Effect on Oxidation Energetics and Driving Forces

6.1 Introduction

As described in Chapter 2, single-step superexchange is relevant for the short-range charge transfer whereas long-range hole migration along a π-stack of DNA (with charge displacements of up to 200 Å) can occur by propagating radical cation states between guanine bases (G) mediated by tunneling through intervening bridges of A·T base pairs ("G-hopping"). A previous study[33] reported oxidation energies of DNA bases calculated for Watson–Crick pairs and their triads. These systems were calculated at regular geometries. For comparable environments, cation states of A^+, T^+ and C^+ were computed at 0.44, 1.28, and 1.55 eV, respectively, higher in energy than G^+.[33] These results agreed well with other quantum mechanical studies[22–24] and experimental estimates.[25] This suggests that G^+ and other nucleobase radical cations are off-resonance and therefore charge transfer between G bases mediated by A·T pairs can occur via superexchange. However, as pointed out in Chapter 2, long-range hole transport over $(A{\cdot}T)_n$ bridges when $n > 4$ can occur via "A-hopping"[7,199,206] which implies thermally induced hole transfer from G^+ to A of a bridge, followed by hole hopping to neighboring adenines. The Boltzmann factor $\exp(-\Delta/k_B T)$ for thermally induced hole hopping from G^+ to A (an energy gap Δ of ~0.4 eV), is 10^{-7} and should be very sensitive to variations of Δ.

Since the states C^+ and T^+ are considerably higher in energy than G^+ (energy gap Δ of ~1 eV), it is highly unlikely that these states can serve as intermediates in charge transfer between guanines.

In DNA, guanine triplets $(G{\cdot}C)_3$ are stronger hole acceptors than single G·C pairs embedded in A·T runs.[34,207] This milestone result demonstrated that hole trapping by guanine is significantly affected by neighboring base pairs. The effects of neighboring pairs on radical cation states in the DNA π-stacks were systematically studied using quantum chemical calculations.[33] Hole trapping by a base B in the duplex sequence $5'\text{-}XBY\text{-}3'$ was shown to be considerably affected by the *subsequent* base pair Y (with relative shifts up to 0.3 eV), whereas the effect of the *preceding* base pair X turned out to be rather small. In addition, recent modeling lead to the conclusion that the localization and energetics of an electron hole state in a DNA strand can be strongly affected by the configuration of neighboring sodium cations.[208,209] Therefore, among other factors, the surrounding electrolyte has to be considered as source for changes in the electrostatic potential created in the interior of DNA that can influence the charge transfer.

In this chapter, using an QM/MD approach we describe in detail the results of a computational study on the energetics of electron hole states in DNA and the role of different dynamical factors which influence the driving force for charge transfer between base pairs.

6.1.1 Dynamics of DNA environment

To understand the dynamics of water molecules and counterions around DNA, it is important to have detailed information about the time that water molecules or ions stay in a solvation shell of specific DNA site, called the residence time. The concept of a residence time of water molecules within a confined area α (e.g. minor groove) can be expressed in terms of a correlation function $C_\alpha(t)$:[112,210–212]

$$C_\alpha(t) = \frac{1}{N_w}\sum_{i=1}^{N_w}\frac{1}{N_{\alpha,i}}\sum_{t_0=1}^{t_{\max}} p_{\alpha,i}(t_0, t_0+t; t^*) \tag{6.1}$$

where $p_{\alpha,i}(t_0, t_0+t; t^*)$ is the survival probability function that takes the value *one* if the *i*th water molecule is in the confined area in the interval from time t_0 to t_0+t, and in the

interim it does not leave the solvation shell of the confined area for any continuous period longer than t^*. Under all other circumstances, it is *zero*. It is not very clear how to choose the time period t^*. Values of 1 ps or 2 ps have been used as water relaxation time, respectively, by Feig and Pettitt,[112] and Impey et al.,[210] whereas García and Stiller,[211] and Rocchi et. al.[212] used $t^* = 0$. $p_{\alpha,i}$ is summed up to the simulation time $t_{\max}$ and normalized by the number of times $N_{\alpha,i}$ at which a water molecule i is found within the confined area. A value $p_{\alpha,i} = 0$ is assumed if $N_{\alpha,i} = 0$. The second summation, over i in Eq. (6.1),[112] is done over all water molecules N_w for which the residence times are calculated. The function $C_\alpha(t)$ representing the average distribution of residence times corresponds to an exponential decay and, therefore, the residence times τ can be calculated by either fitting an exponential function[210]

$$C_\alpha(t) = c\exp(-t/\tau) \tag{6.2}$$

where c and τ are fitting parameters, or a linear fit according to[212,213]

$$\ln C_\alpha(t) = \ln C_\alpha(t_0) - t/\tau \tag{6.3}$$

The residence time of counterions in the solvation shell at specific sites of DNA can be determined in similar fashion.

In studying the dynamics of water molecules, Forester and McDonald[179] have shown that the mobility of water around DNA depends on the strength of DNA-water interaction. Water molecules were classified as strongly coordinated to a DNA site with the general trend of phosphate group > major groove > minor groove. This trend was derived from the radial distribution function (RDF), see Section 3.5, in which the RDF of phosphate-water showed a very sharp peak of the solvation shell in contrast to the peak of the RDF for water in the minor groove.[179] In addition, Feig and Pettitt reported from their simulations that about 8.5 water molecules were found in the first solvation shell around a phosphate group, and 5.5 and 1.7 water molecules per base pair were found in the major groove and the minor groove, respectively.[214] These values are consistent with the trend of water-DNA interaction described above. Feig and Pettitt[214] also reported that the numbers of water molecules for hydrogen bond interactions with G·C and A·T pairs were 20.6 and 19.3 water molecules, respectively. Such hydrogen bonding was considered using the following criterion: the X–H distance (X is nucleobase atom, and H and O are water atoms) should be shorter than 2.5 Å and the X-O-H angle should be larger than 135°. From

this findings, they concluded that water molecules interact stronger with a G·C pair than an A·T pair.

The distribution of counterions around DNA was also studied by Feig and Pettitt.[112] Similarly to the dynamics of water molecules, in average 0.65, 0.25 and 0.03 ions were found around each phosphate, major groove and minor groove of two base pairs, respectively. In the first solvation shell around GG·CC and AA·TT pairs, they found 1.77 and 1.68 ions, respectively. Alternatively, residence times of 123, 280 and 960 ps were determined for sodium ions within the first solvation shell of DNA atoms. Shorter residence times of 14 to 36 ps were found for counterions in the major groove whereas 150 to 280 ps were found at minor groove.[112] The residence times around guanine bases were up to 160 ps, but only 20 to 30 ps near adenine bases. Consequently, the interaction of water molecules and sodium ions with guanine is considerably stronger than with adenine. As will be discussed later this difference is reflected by changing oxidation potentials of guanine and adenine bases.

6.1.2 Oxidation Potentials of DNA Nucleobases

DNA is a highly negatively charged biopolymer immersed in an electrolyte; it displays a remarkable sensitivity to the ionic and polar surroundings due to electrostatic effects.[215] In a study by Kim and VeBreton,[216] gas-phase ionization potentials (IPs) were calculated for anionic clusters of the constitution 2′-deoxyguanosine 5′-phosphate ($5'\text{-}dGMP^-$) and $5'\text{-}dGMP^-\cdot n H_2O\cdot Na^+$ with n = 4, 8, 11, 12 and 14 water molecules; Na^+ was bound to a phosphate. The absolute values of IPs were obtained with Hartree-Fock calculations and shifted by using experimental data for nucleobases. They reported that the IPs of a base in the clusters $5'\text{-}dGMP^-\cdot n H_2O\cdot Na^+$ (7.6–7.8 eV) are larger than that in isolated $5'\text{-}dGMP^-$ (5.8 eV) by 1.8–2.0 eV. They also demonstrate that the direct interaction with water molecules has a much smaller effect on $5'\text{-}dGMP^-$ ionization potentials than the electrostatic stabilization by Na^+. IP of the base in the cluster with 14 water molecules differs by about 0.4 eV from the corresponding IP in the cluster with 4 water molecules.

Schuster and colleagues[208] carried out DFT calculations on (micro-) solvated DNA duplex oligomers, using a local-spin-density approximation. The configurations used in the quantum chemical calculations had been extracted from MD simulations, and these

structure were then re-optimized to the nearest local minimum of the potential energy hypersurface. In each case, the selected model of the duplex $5'\text{-}G_1A_2G_3G_4\text{-}3'$ included the phosphate groups and at least the first water solvation shell of sodium cations assigned to the model cluster. The following models were investigated: (*i*) all Na^+ were near the phosphates, (*ii*) similar (*i*) but one Na^+ was relocated to the major groove near the atom N7 of base G_1, (*iii*) similar (*i*) but one Na^+ was relocated to the major groove near the atom N7 of base G_3, and (*iv*) the same configuration as (*iii*) but without any solvating water molecules. The vertical IPs for these cases are found to be 5.22, 5.69, 5.46 and 4.16 eV, respectively. The increase of IP upon hydration originates from increased electronic binding caused by solvent. The higher vertical IP of the hydrated configuration may, at first, seem surprising in the light of the common notion that hydration serves to screen electrostatic interactions and, thus, reduces the effect of counterions and negatively charged phosphates, whereas the opposite is found by Schuster and co-workers.[208] However, the above screening notion holds when a bulk situation is considered. Nevertheless, it does not apply to the interfacial interactions and molecular length scales that characterize the hydration process of DNA. Underlying the increase of the vertical IP in the hydrated system, i.e. stronger binding of the electrons, is the preferential orientational ordering of the water molecules in the first (and to a smaller extent in the second) hydration shell, with the enhanced binding originating from the added attractive interaction between the electrons and the dipolar charge distribution of the water molecules.[208] Thus, inclusion of the sugar-phosphate backbone, counterions, and water molecules is essential for a proper and accurate description of the energetics and dynamics of charge transport in DNA.[208]

Furthermore, Schuster and colleagues have proposed a new mechanism for charge transfer in DNA which they termed "ion-gated hole-transport".[208,209] According to this mechanism, the ionization potential of a DNA fragment and the localization of the radical cation are strongly modulated by the location of the counterions. These findings led to a model for hole hopping where the rate is controlled, to a large part, by the probability of forming certain counterion configurations that are effective in changing the hole density over the duplex DNA sequence. In other words, hole migration is controlled by the dynamical fluctuations of the arrangements of the counterions (and of the solvation water molecules).

However, in this chapter we will show that fluctuations of configurations of water molecules play the most important role in the charge transfer energetics whereas effects of ions are relatively small. Therefore, it is more appropriate to refer to this mechanism for charge migration in DNA as "solvent-gating" mechanism.

6.2 Details of calculation

6.2.1 Estimation of the oxidation potential of nucleobases in DNA

As already highlighted, the free energy ΔG^{o} of charge transfer reaction is a key characteristic determining the charge migration in DNA. ΔG^{o} can be defined as a difference of oxidation potentials (OxP) of the acceptor and donor nucleobases. So far there has been no experimental data on oxidation potentials of nucleobases within DNA and the values measured for nucleosides in solution[33] are often used by analysis of the CT process. Obviously, in such models many factors affecting OxP of nucleobases are not accounted for. The OxP values can be subtly modulated by surrounding environment. In particular, it was demonstrated that neighboring pairs remarkably effect the oxidation potential of nucleobases.[33] Another very important effect, which should be taken into account by considering the CT kinetics, is structural and environmental fluctuations occurring on the time scale of charge transfer (1–1000 ps). Absolute values of OxP are very difficult to obtain quantitatively from quantum chemical calculations because environmental effects have to be described accurately and one has to consider a sufficient sampling associated with environmental degrees of freedom. In the current work we adopted a microscopic approach that allows to treat the intermolecular electrostatic interaction for adequate configurational sampling. Our model assumes that the ensemble of radical cation states can be approximately characterized by data obtained for the corresponding non-ionized duplex. The driving force for hole transfer between two sites was estimated as the difference of vertical ionization energies calculated along MD trajectories of DNA fragments. We estimated the reaction free energy ΔG^{o} of hole transfer with a quantum chemical method which takes the instantaneous atomic configuration of the environment into account (snapshots at every picosecond). The ionization energies were determined with the semi-empirical NDDO-G method[217] which is specially

parameterized for calculating ionization and excitation energies of organic molecules. This procedure reproduces experimental ionization potentials of nucleobases in the gas phase with an average deviation of 0.09 eV[217] and, therefore, it is well suited for estimating ΔG^{o} of hole transfer between nucleobases. Because the reorganization energies connected with the formation of cation hole states of guanine and adenine bases are very similar (0.40 and 0.35 eV, respectively) and essentially independent of environmental fluctuations, the corresponding contributions cancel when one estimates energy differences of such states. Thus, differences of vertical (instead of adiabatic) ionization energies can be used to estimate ΔG^{o}. The electrostatic effects of the surroundings were accounted for by approximating all atoms in the MD simulation box as point charges with values according to the force field.[89] The effective dielectric constant of the medium was chosen at 2,

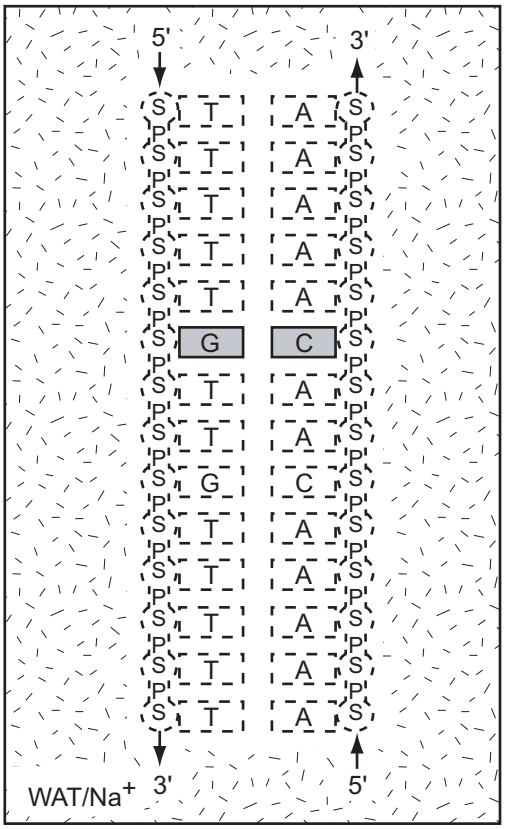

Figure 6.1 Sketch of model used for calculating oxidation potential of guanine within the base pair $(G{\cdot}C)_6$ in the duplex $5'\text{-}T_1T_2T_3T_4T_5G_6T_7T_8G_9T_{10}T_{11}T_{12}T_{13}T_{14}\text{-}3'$. Electrostatic effects of 13 neighboring pairs, 28 sugar (S) and 26 phosphate (P) groups, 26 Na^+ ions and 3747 water molecules (WAT) are accounted for by using point charges (indicated as dashes) and a dielectric constant of 2.

corresponding to fast electronic polarization.[76,218] Hereafter we will refer to ionization potentials of nucleobases calculated with accounting for electrostatic interactions with surroundings as an oxidation potentials. While the absolute value of this quantities can differ from real OxP by a constant, their difference should provide a good estimation ΔG^{o}.

6.2.2 Model

We studied several DNA models with different bridges, namely the duplexes 5′-(T)$_5$G(T)$_n$G(T)$_5$-3′ (n = 1–4), 5′-(T)$_4$G(T)$_4$G(T)$_4$-3′, 5′-(T)$_3$G(T)$_6$G(T)$_3$-3′, and 5′-(T)$_2$G(T)$_8$G(T)$_2$-3′ (corresponding to duplexes 9–16 in Table 3.1).

Long-time (~12 ns) MD simulations of the duplexes have been carried out. Because of the small number of ions compared to water molecules (see Table 3.2) and their very long residence time (0.1–1 ns) around DNA, long simulation times (of the order of 10 ns) are essential for a statistically meaningful representation of the ion distribution.[112] Along the trajectories, we calculated the energetics of hole transfer from G to neighboring bases A and G. The calculated model is shown in Figure 6.1.

6.2.3 Time series analysis of the energy fluctuations

The OxP and ΔG^{o} fluctuations were also analyzed using a normalized autocorrelation function and its Fourier transformation. This allows us to understand the characteristic time of CT energetics changes. Autocorrelation function (also known as autocorrelation coefficient) is a correlation between two values of the same variable at times t_0 and delayed time $t_0 + t$. The lag t of autocorrelation function $C(t)$ of a characteristic F is

$$C(t) = \frac{\langle F(t_0)F(t_0+t)\rangle}{\langle F(t_0)^2\rangle}, \tag{6.4}$$

where $F(t_0)$ and $F(t_0 + t)$ are calculated at time t_0 and $t_0 + t$, respectively.[219] The autocorrelation coefficients vary between −1 and +1. The value of 0 indicates no correlation and the absolute value of 0.9–1.0 indicates a high degree of correlation. The plotted autocorrelations for many lags are used for checking randomness in a data set. If the process is random, the autocorrelation function should nearly be zero for any and all

time-lag separations. If the process is non-random, then one or more of the coefficients will be significantly non-zero. We obtained that in all duplexes the variations of OxP are not random, but correlated within a specific time interval. The calculated autocorrelation functions were then analyzed using Fourier transformation.

It is instructive, to analyze an autocorrelation function $C(t)$ by Fourier analysis, i.e. by identifying (periodic) contributions of various frequencies (1/time). More specifically, a discrete Fourier transform is used because we consider the discrete data points of the autocorrelation function. Formally, the periodicity interval is set to the time over which the autocorrelation function is known. For a data set $C(t_j)$ the Fourier coefficient f_k of index k is defined as[220]

$$f_k = \sum_{j=0}^{N-1} C(t_j) \exp(-i2\pi\omega_k t_j) \tag{6.5}$$

where $\omega_k = k/(N\Delta t)$ is a frequency, $\Delta t = t_{j+1} - t_j$ is the time interval between the data points, and N is the total number of data points. The inverse of the time period, i.e. $1/N\Delta t$, represents a lower limit of the frequencies, whereas the upper limit is defined by half of the inverse of the time interval, i.e. $1/2\Delta t$. The discrete Fourier transformation was carried using the program xmgrace[221] on a Linux workstation.

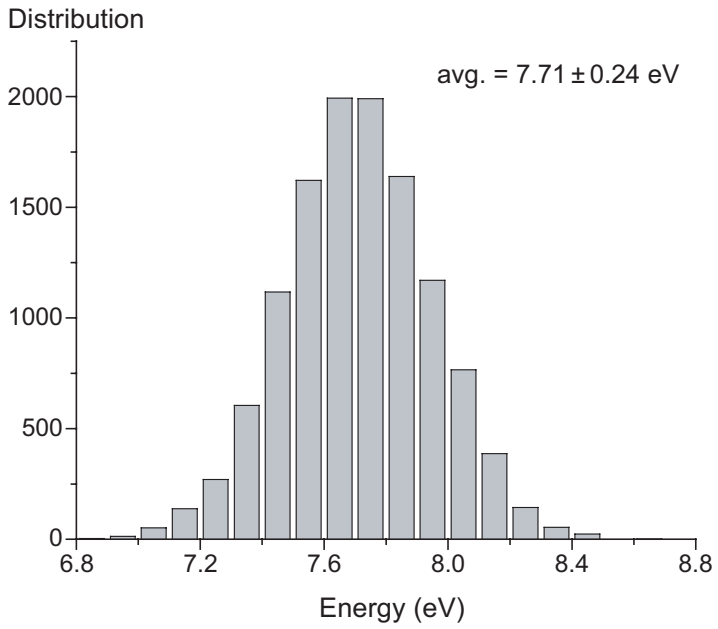

Figure 6.2 Histogram of OxP of G_9 calculated for 12 ns MD simulation of unconstrained duplex 5′- $(T)_5G(T)_2G(T)_5$ -3′ .

Evaluation of data distribution

To determine how well our data are described by a probability distribution function we performed a chi-squared test where we compared observed and expected frequencies.[222] The results show that the energetics of nucleobases can be accepted to be of the normal distribution with a confidence of 98%. As an example, we display a histogram of the OxP of G_9 in the duplex $5'\text{-}(T)_5G(T)_2G(T)_5\text{-}3'$; see Figure 6.2.

6.3 Results and Discussion

6.3.1 Sensitivity of computational results to the choice of the model

As reported by Schuster et al., the ionization potential of nucleobases in aqueous solution is larger than that in the corresponding isolated system.[208,209] As a consequence, the energy for oxidizing a nucleobase strongly depends on the type and number of surrounding molecules. Therefore, we first investigated how sensitive the ionization energy is to type and number of molecules in the environment. As an example, we calculated the oxidation energy of a guanine base, namely G_6, using a snapshot from an MD trajectory of the constrained duplex $5'\text{-}T_1T_2T_3T_4T_5G_6T_7T_8G_9T_{10}T_{11}T_{12}T_{13}T_{14}\text{-}3'$. We started by taking into account 3 base pairs $T_5G_6T_7$, where neighboring pairs $(T{\cdot}A)_5$ and $(T{\cdot}A)_7$ were treated as point charges. In addition, the sugar-phosphate backbone (6 sugar and 4 phosphate groups bound to the fragment $T_5G_6T_7$), the four closest counterions, and water molecules around the surface of the corresponding fragment (varying from 10 to 3747 water molecules) were also considered as point charges. We found that the oxidation energy increases rapidly from 7.00 to 8.48 eV when the number of water molecules increases from ten to a few hundred molecules. In the range of 1600 to 3747 water molecules, the energy increases only by 2% and the energy of 8.89 eV was found when all water molecules were treated. The results are shown in Figure 6.3. Then, the number of neighboring pairs was extended to 4, 6, 8, 10, 12 and all (13) pairs. In each case, the corresponding number of sugar-phosphate groups and sodium ions was taken into account. For instance, 10 sugar and 8 phosphate groups, and 8 sodium ions were accounted for the fragment $T_4T_5G_6T_7T_8$. The increasing number of neighboring pairs lead to a decreasing oxidation energy. Energies of

8.85, 8.40, 8.19, 7.86, 7.82 and 7.80 eV (when accounting for 4, 6, 8, 10, 12 and 13 neighboring pairs, respectively) were obtained when all water molecules were considered (see Figure 6.3).

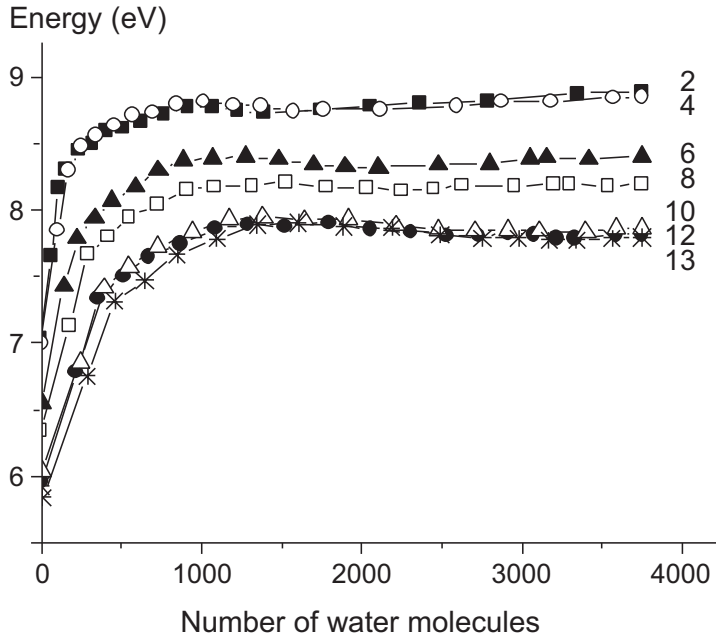

Figure 6.3 Oxidation energy of base G_6 in the constrained duplex $5'$-$(T)_5G(T)_2G(T)_5$-$3'$ calculated with different models. The number of water molecules and neighboring pairs is varied. The following labels are used: *solid square* for two neighboring pairs, *opened circle* for 4, *solid triangle* for 6, *open square* for 8, *open triangle* for 10, *solid circle* for 12 and *star* for 13 pairs.

The sensitivity of the results to the extension of the system was also probed by two further strategies. The number of water molecules in the model was varied as a function of (*i*) the distance *R* around the surface of the DNA fragment and (*ii*) the distance *R*′ around each sodium ion. Again, the number of neighboring pairs, backbone units, counterions and water molecules was varied. The obtained results are similar to the above calculations; the oxidation potential decreases when the number of neighboring pairs increases and an increase of the number of water molecules leads to the increasing oxidation energy. Thus the convergence of the calculated energies is quite weak and it seems to be instructive to take into account point charges of the whole system.

6.3.2 Effects of dynamics of surrounding species on the energetics

Constrained duplex

Analysis of a constrained structure allows us to separate the effect of the dynamics of counterions and water molecules because conformational changes of the duplex are completely excluded. Four duplexes, 5′-$(T)_5G(T)_nG(T)_5$-3′ where n = 1–4, were studied. The averaged energy of both guanines are calculated at 7.76, 7.79, 7.80, and 7.80 eV for duplexes with n = 1 to 4, respectively. For adenine, the obtained energies (averaged over all intervening adenines) were 7.97, 8.00, 8.01, and 8.02 eV for n = 1 to 4, respectively. The energy fluctuations (standard deviations) in both cases are ~0.15 eV (see Table 6.1). One can see that the number of intervening T·A pairs does not have a notable effect on the calculated energies and their fluctuations.

Table 6.1 Oxidation potentials and standard deviations[a] of guanines and bridge adenines of the constrained duplexes 5′-$(T)_5G(T)_nG(T)_5$-3′.

Base	Oxidation potentials (eV)			
	n = 1	2	3	4
5′-G	7.75 ± 0.14	7.79 ± 0.14	7.79 ± 0.16	7.79 ± 0.16
A_7	7.97 ± 0.14	8.00 ± 0.15	7.99 ± 0.15	8.00 ± 0.16
A_8		7.99 ± 0.15	8.02 ± 0.15	8.02 ± 0.15
A_9			8.01 ± 0.15	8.04 ± 0.15
A_{10}				8.04 ± 0.15
3′-G	7.76 ± 0.15	7.79 ± 0.15	7.80 ± 0.16	7.81 ± 0.16

[a] Calculated for 3000 snapshots of 3 ns MD trajectory.

Also the OxP difference between guanine and adenine for each duplex remains almost unchanged for the complexes considered: ~0.21 ± 0.18 eV. Interestingly, when the electrostatic effect of the environment is not accounted for, the energies were calculated at 7.71 and 8.04 eV for guanine and adenine, respectively. Their difference of 0.33 eV is larger; this finding indicates that the environment reduces the energy gap between guanine and adenine.

Unconstrained duplex

Unconstrained MD simulations take into account both the movement of environment molecules and the conformational dynamics and vibrations of the duplex. One expects that oxidation energy fluctuations and their differences will be affected by both types of structural dynamics. Let us first consider the effect of internal vibrations of nucleobases within only one Watson–Crick pair. The energies for guanine and adenine averaged over 12000 snapshots were calculated to be 7.58 ± 0.13 and 8.04 ± 0.10 eV, respectively. Very similar results were obtained for other duplexes (duplexes 9–16 in Table 3.1). In this model the fluctuations of energies are due to internal vibration of the nucleobases. The difference of the energies of guanine and adenine was about 0.46 ± 0.16 eV. Comparing this difference to that found for constrained duplexes without the electrostatic effects of environment (0.33 eV), one can conclude that structural and conformational dynamics of DNA should favorably stabilize the radical cation state of guanine with respect to that of adenine.

Table 6.2 Oxidation energies of guanine and adenine calculated for 12000 snapshots (each ps) of 12 ns unconstrained MD trajectories, in eV.

Duplex	Guanine[a]	Adenine[b]
5′- TTTTTGTGTTTTT -3′	7.70 ± 0.22	8.03 ± 0.21
5′- TTTTTGTTGTTTTT -3′	7.71 ± 0.23	8.04 ± 0.20
5′- TTTTTGTTTGTTTTT -3′	7.70 ± 0.24	8.06 ± 0.20
5′- TTTTTGTTTTGTTTTT -3′	7.66 ± 0.25	8.06 ± 0.21
5′- TTTTGTTTTGTTTT -3′	7.67 ± 0.24	8.06 ± 0.20
5′- TTTGTTTTTTGTTT -3′	7.65 ± 0.28	8.03 ± 0.22
5′- TTGTTTTTTTTGTT -3′	7.64 ± 0.30	8.04 ± 0.21
5′- TTGTTTTTTTTGTT -3′ (methylphosphonate)	7.76 ± 0.25	8.10 ± 0.19

[a] Averaged over two guanines.

[b] Averaged over all adenines intervening between two guanines.

Now taking into account both effect, namely the dynamics of the duplex and its surroundings, we found averaged energies of 7.68 and 8.05 eV for guanine and adenine, respectively (Table 6.2). The energy gap between both nucleobases decreases to 0.37 eV because the interaction of water molecules with G·C pairs is considerably stronger than with A·T pairs (see Section 6.1.2). The fluctuations of OxP of guanines in $5'\text{-}(T)_5G(T)_nG(T)_5\text{-}3'$ for n = 1–4, (Table 6.2), increase slightly from 0.22 to 0.25 eV, whereas those of adenines remain almost unchanged (0.21 eV). This can be explained as the result of the increasing number of water molecules (3793, 3936, 4132 and 4276 water molecules for duplexes n = 1 to 4, respectively). Because guanines are close to the terminal of the duplex, they are stronger affected by solvent than adenines intervening between the guanines. Comparing the fluctuations of the energetics due to internal vibrations within a Watson–Crick pair (0.10–0.13 eV) and due to electrostatic effects (0.20–0.25 eV), one can conclude that the dynamics of water molecules and counterions should play the main role (see also Figure 6.4).

Let us now consider the energetics of normal DNA duplexes comprising 14 Watson–Crick pairs, i.e. duplexes 2, 5, 6 and 7 in Table 6.2. The calculated energies for guanine and adenine are about 7.67 and 8.04 eV, respectively. For these four oligonucleotides, the fluctuation of oxidation energies is calculated to be of 0.21 eV for adenines and ranges from 0.23 to 0.30 eV for guanines. As discussed above, the change of energy fluctuations arises from the difference in number of environment molecules, as well as the position of nucleobases under consideration. Since these four duplexes contain a very similar environment (26 sodium ions, 3940 ± 4 water molecules), one can conclude that the increase of the energy fluctuation is caused by the variation of the guanine units in the duplex. In other words, guanine bases closer to DNA terminals are stronger affected by environment, and thus they exhibit larger fluctuations of oxidation potentials.

The environment effects discussed above are due to both water molecules and counterions. To consider only the effects of water, we studied the neutral fragment $5'\text{-}(T)_2G(T)_8G(T)_2\text{-}3'$ where one phosphoryl oxygen of each phosphate group was replaced by a methyl group. Therefore, no counterions are required for the simulation of the system. We obtained oxidation energies for guanine and adenine at 7.76 and 8.10 eV, respectively (the last duplex in Table 6.2). Comparing the modified and original systems,

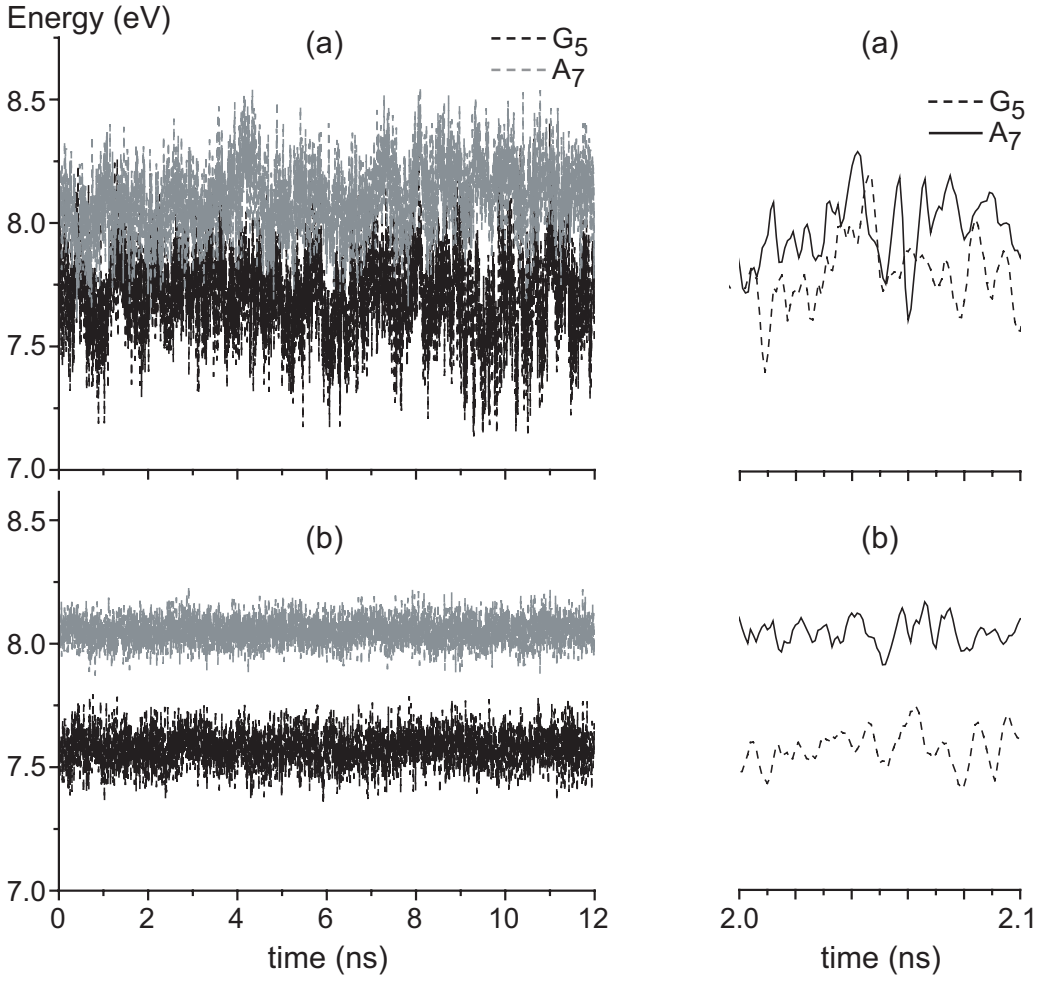

Figure 6.4 At the left, energy profiles of guanine G_5 (gray) and adenine A_7 (black) along 12 ns MD trajectory of unconstrained duplex $5'\text{-}TTTTG_5T_6T_7T_8T_9G_{10}TTTT\text{-}3'$. The energy is calculated (a) with electrostatic effects of environment and (b) without environment. Plots of these systems with shorter time scale are given at the right.

we found that the oxidation energies are 0.12 and 0.06 eV larger in the modified duplexes. This is in agreement with the result calculated by Barnett et al.[208] and can be explained as follows. In a normal duplex, negative charges of phosphate groups have an important effect on the oxidation potentials of nucleobases. Although the system contains positively charged counterions, they cannot completely counterbalance the negative charges of phosphate groups, because of the screening effects of solvent. Consequently, the oxidation energy of a nucleobase within the normal oligonucleotide is lower than within a modified duplex. Furthermore, the OxP fluctuations in the system without sodium ion are 0.25 and

0.19 eV for guanine and adenine, respectively. The fluctuations are smaller as compared to that calculated for the system with sodium ions.

The Fourier transform of OxP autocorrelation function defined for normal duplexes shows 2–3 frequencies (see Table 6.3) depending on the duplex. These frequencies relate to the characteristic times ranging from about 220 to 930 ps. In most cases the characteristic time of the fluctuations is about 300, 400 and 800 ps (Table 6.3).

Table 6.3 Characteristic times (ps) of the OxP fluctuations of guanine and adenine (in bold) for unconstrained duplexes.

Duplex	Guanine	Adenine
5′- T T T T T **G** T G T T T T T -3′ 3′- A A A A A C **A** C A A A A A -5′	302, 414, 718	215, 318, 842
5′- T T T T T **G** T T G T T T T T -3′ 3′- A A A A A C **A** A C A A A A A -5′	341, 433, 750	340, 432, 736
5′- T T T T T **G** T T T G T T T T T -3′ 3′- A A A A A C A **A** A C A A A A A -5′	229, 421, 832	226, 502, 862
5′- T T T T T **G** T T T T G T T T T T -3′ 3′- A A A A A C A **A** A A C A A A A A -5′	332, 433, 895	291, 497, 920
5′- T T T T **G** T T T T G T T T T -3′ 3′- A A A A C A **A** A A C A A A A -5′	328, 787	381, 734
5′- T T T **G** T T T T T G T T T -3′ 3′- A A A C A A **A** A A A C A A A -5′	377, 567, 840	391, 717
5′- T T **G** T T T T T T T G T T -3′ 3′- A A C A A A **A** A A A A C A A -5′	325, 768	374, 528, 934
5′- T T **G** T T T T T T T T G T T -3′ 3′- A A C A A A **A** A A A A A C A A -5′ (methylphosphonate)	87, 166, 337, 858	87, 169, 316, 871

Since in unconstrained simulations both the surroundings movement and DNA vibration are allowed, the characteristic times of the OxP fluctuations on the hundred-picoseconds time scale correspond to structural dynamics of both DNA and its surroundings. Seemingly, our observed time scales are relevant to the movement of sodium ions in molecular dynamics simulation, since the residence times of sodium ions in DNA

system vary from tens to hundreds of picoseconds or even 0.9 ns.[112,180] However, it is not clear whether only counterions are responsible for the long correlation time, because the neutral duplex (without sodium ions) also exhibits long correlation times of about 320 and more than 850 ps (see the last entry in Table 6.3). The Fourier transformed autocorrelation functions of normal and neutral duplexes differ only on a short time scale, i.e. less than 200 ps. The time series analysis of modified duplex $5'\text{-}(T)_2G(T)_8G(T)_2\text{-}3'$ gives additional characteristic times of about 90 and 170 ps, whereas a characteristic time shorter than 200 ps was not identified in the normal duplexes. Therefore, one can conclude that the characteristic times of the OxP fluctuations correspond to DNA structural vibration and the movement of water rather than to the movement of counterions.

As mentioned above, OxP of nucleobases depend on orientation of water molecules. To analyze the short-time effects due to reorientation of water molecule, we studied the molecular dynamics of both constrained and unconstrained duplexes $5'\text{-}(T)_5G(T)_2G(T)_5\text{-}3'$ for 20 ps. The coordinates were stored every 2 fs. This allowed us to investigate the effects of the environment at a femtosecond time scale.[223–225] We found that the autocorrelation functions of OxP of guanines and adenines have similar characteristics. Let us consider this in more detail.

For constrained duplex, the autocorrelation functions of guanines and adenines decrease rapidly to 0.25 at about 20 fs and then increase showing peak at about 40 fs with the coefficients of ~0.55. Then several oscillations can be seen (Figure 6.5a). Fourier transformation of the autocorrelation functions indicate that these oscillations occur with period of about 43 fs, corresponding to the wave number of about 780 cm^{-1}. Other frequencies can be observed at 700–900 cm^{-1} (Figure 6.5b). The OxP fluctuations on the femtosecond time scale can reasonably be assigned to reorientation time of water molecule in line with data reported previously (500–850 cm^{-1}) which were obtained from calculated absorption spectra of rotational modes of water in pure solvent[223,224] and in aqueous solution of metal ion Mn^{2+}.[225]

For unconstrained MD simulation of an oligormer, the energy of nucleobase was calculated in two different ways: by taking into account (*i*) only the Watson–Crick pair under consideration and (*ii*) that Watson–Crick pair and electrostatic interactions of all other species (i.e. neighboring base pairs, sugar-phosphate backbone, water molecules and

counterions). We found that in both cases the autocorrelation functions decrease in a oscillatory fashion with a period of 19 fs (see Figure 6.6a). Their Fourier transformations display the dominant peak at ~1750 cm^{-1} (see Figure 6.6b) which can be assigned to the double bond vibration in nucleobases.[47] Comparing the results obtained with and without the electrostatic effects (dashed and solid lines in Figure 6.6b), the two spectra clearly

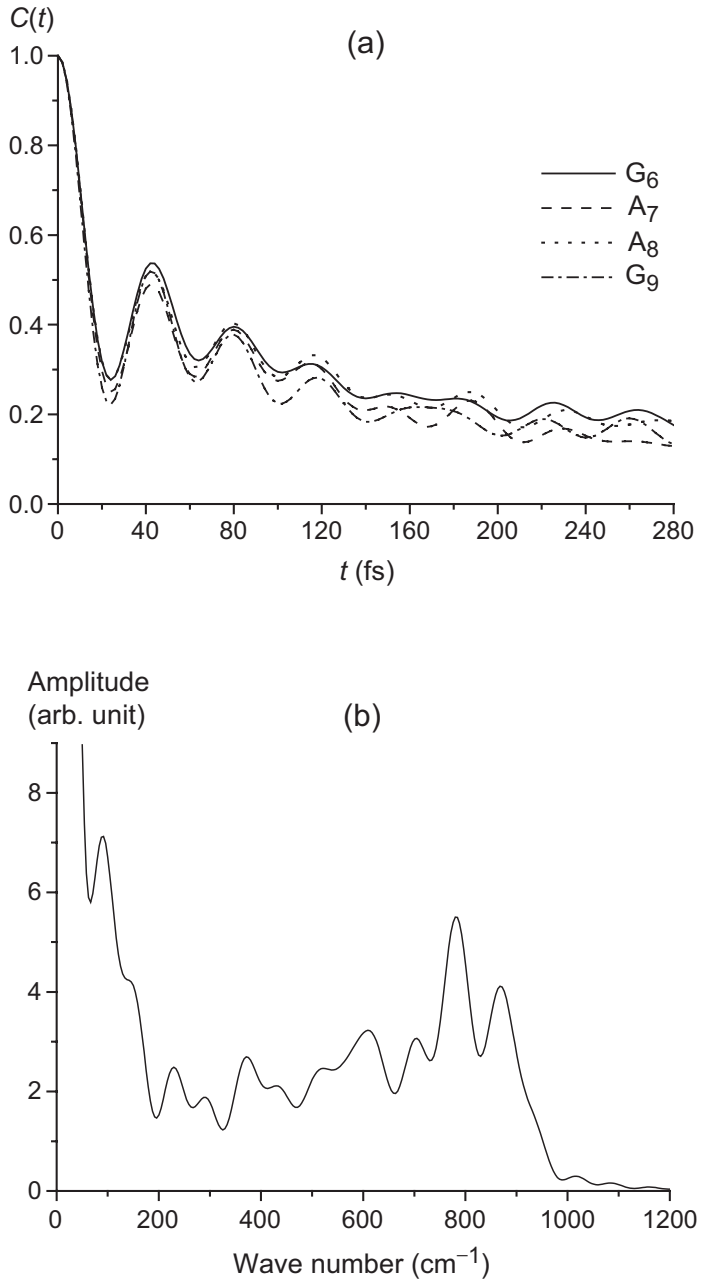

Figure 6.5 Time series analysis of oxidation potentials for the constrained duplex 5′-TTTTTG$_6$T$_7$T$_8$G$_9$TTTTT-3′. (a) Normalized autocorrelation functions of OxP of nucleobases G$_6$, A$_7$, A$_8$, and G$_9$ and (b) Fourier transform of OxP of G$_9$.

differ in the frequency range 700–900 cm^{-1} which is related to the reorientation of water molecule, as found in the constrained system.

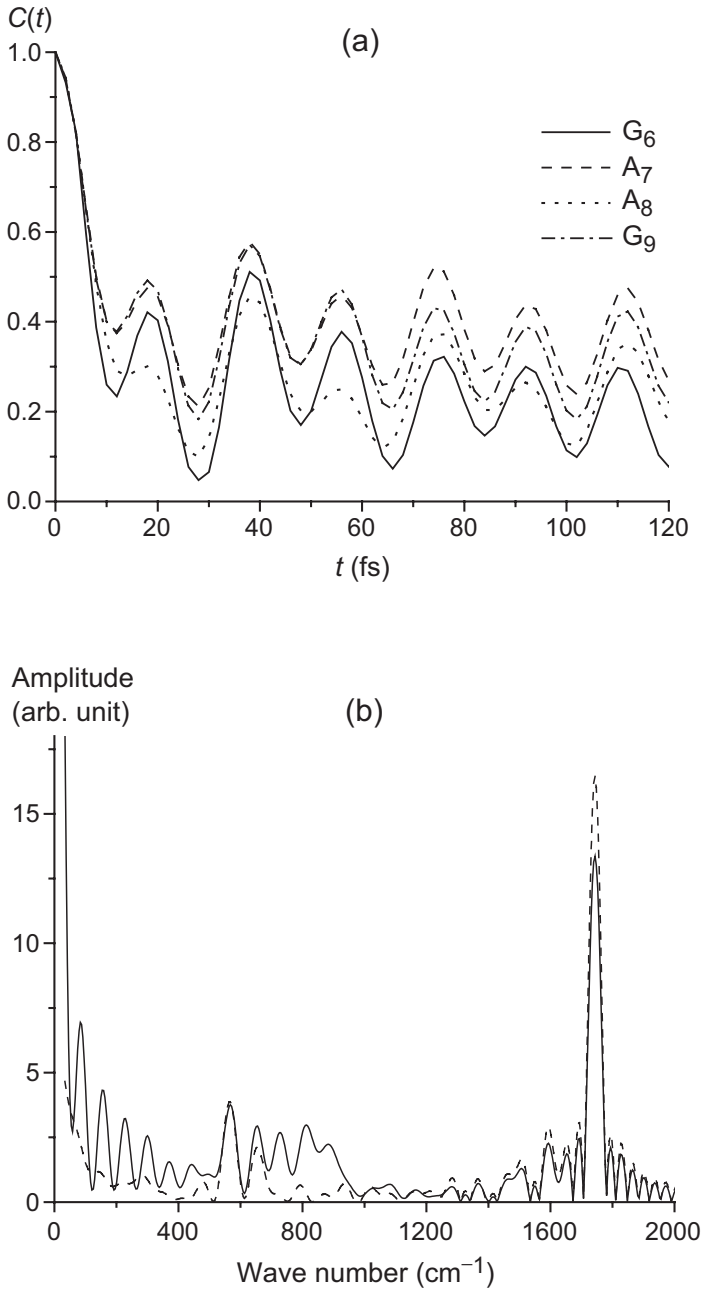

Figure 6.6 Time series analysis of oxidation potentials for the unconstrained duplex 5′-TTTTTG$_6$T$_7$T$_8$G$_9$TTTTT-3′. (a) Normalized autocorrelation functions of OxP of nucleobases G$_6$, A$_7$, A$_8$, and G$_9$ and (b) Fourier transform of OxP of G$_9$ calculated with (solid line) and without (dashed line) electrostatic interactions.

Summarily, because the vibrations of nucleobases occur on the femtosecond time scale, they are not pertinent to the charge migration process which occurs on a time scale ranging from several tens of a picosecond to several nanoseconds. Pertinent CT time scales can be illustrated by rates for hole transfer in DNA. DNA hairpins linked by stilbenedicarboxamide imply characteristic times ranging from 30 ps to 5 ns for bridges of 1 to 3 A·T base pairs.[6] Similar estimates, from 10 to 200 ps, were received for hole transfer from an acridine derivative as chromophore to 7-deazaguanine as acceptor, separated by one and two intermediate A·T pairs.[45] Direct fluorescence probing of the dynamics of water around DNA with femtosecond resolution yielded mean residence times and reorientation times of water molecules at 3–6 ps.[226] Our MD simulations identified three characteristic time scales, 120 ps, 280 ps, and 960 ps, for the dynamics of sodium ions near DNA.[112]

6.3.3 Driving force

In this section we discuss the driving force ΔG^{o} of hole transfer from G^{+} to A in unconstrained duplexes. As the estimated ΔG^{o} values correspond to an ensemble of nuclear configurations, their standard deviations include three contributions, due to (*i*) the nuclear vibrations of the nucleobases, (*ii*) the conformational changes of DNA, and (*iii*) the fluctuations of the solvent environment.

First, we discuss the distribution of hole states in the duplexes along the MD trajectory. We consider charge transfers between two guanine bases via the bridge of intervening A·T pairs. The hole is assumed to be localized on a base with the lowest ionization energy. As can be seen from Table 6.4, the hole is mainly localized on guanines; however, it can also be found on the bridging adenines. The corresponding probability increases with the length of the bridge from 3% to 8% for duplexes with 2 to 8 intervening adenines. Thus, intervening adenines can serve as intermediate hole acceptors. In addition, the probability of hole localization on the bridge is slightly higher in modified duplex $5'\text{-}(T)_2G(T)_8G(T)_2\text{-}3'$ than in the corresponding normal system (7.3%).

As an example, we present results for the 14-mer duplex $5'\text{-}(T)_2G(T)_8G(T)_2\text{-}3'$ in an aqueous solution which includes 26 Na^{+} counterions. Table 6.5 lists relative energies of

Table 6.4 The distribution of hole states in 14-mer duplexes.

Duplex	Bridge length	5′- G	Bridge[a]	3′- G
5′- TTTTTGTTGTTTTT -3′	2	48.7	3.1	48.2
5′- TTTTGTTTTGTTTT -3′	4	40.3	3.3	56.4
5′- TTTGTTTTTTGTTT -3′	6	42.4	5.6	52.0
5′- TTGTTTTTTTTGTT -3′	8	50.7	7.3	42.0
5′- TTGTTTTTTTTGTT -3′ (methylphosphonate)	8	36.7	7.8	55.5

[a] Sum over all adenine bases of the bridge.

hole states localized on purine nucleobases. As expected, the average free energy of hole states A^+ is positive, about 0.4 eV: guanine is a stronger hole acceptor than adenine. As the standard deviations of the ΔG^o values are ~0.3–0.4 eV, configurations of the system should exist where a radical cation state A^+ is more stable than G^+ and, thus, hole transfer from G^+ to A is energetically feasible. The nucleobases G_3 and G_{12} in the duplex have similar surroundings (however, they are not equivalent) and, therefore, the driving force between them is close to zero on average.

In Figure 6.7, we show fluctuations of the CT energy between the bases G_3 and A_6 as a function of time. The characteristic time of such relevant fluctuations is 0.3–0.4 ns. These characteristic times were estimated from the Fourier transform of the autocorrelation function of ΔG^o; they decrease from 380 ps for hole transfer $G_3 \rightarrow A_4$ to 304 ps for $G_3 \rightarrow G_{12}$. As we already discussed fluctuations of OxP and therefore fluctuations of the driving force are due to (1) molecular vibrations of the donor and acceptor sites, and (2) correlated motion of counterions and water molecules. The conformational dynamics of DNA, plays only a minor role in the CT energetics; however, it substantially affects the electronic coupling between base pairs (see Chapter 4).

Counterions in the vicinity of nucleobases have been suggested to strongly affect the energetics of radical cation states; the consequences for electron transfer have been referred to as “ion-gating” mechanism.[208,209] Counterions are solvated and their motion is

Table 6.5 Relative energies ΔG^{o} of radical cation states in the unconstrained duplex 5′-$TTG_3T_4T_5T_6T_7T_8T_9T_{10}T_{11}G_{12}TT$-3′ and its modified neutral derivative[a] calculated for an MD trajectory of 12 ns. Also given is the occupation[b] of the states corresponding to the equilibrium distribution of hole states as measured by the fraction of time (along the trajectory) where the corresponding state has the lowest energy.

Base	Normal DNA		Modified DNA[a]	
	ΔG^{o}, eV	Occupation,[b] %	ΔG^{o}, eV	Occupation,[b] %
G_3		50.7		36.7
A_4	0.38 ± 0.28	1.8	0.30 ± 0.23	0.8
A_5	0.43 ± 0.32	1.3	0.31 ± 0.26	0.8
A_6	0.44 ± 0.35	1.0	0.30 ± 0.29	1.0
A_7	0.45 ± 0.37	0.7	0.29 ± 0.30	1.2
A_8	0.46 ± 0.39	0.6	0.29 ± 0.32	0.9
A_9	0.44 ± 0.40	0.5	0.28 ± 0.33	1.0
A_{10}	0.41 ± 0.42	0.6	0.26 ± 0.35	1.0
A_{11}	0.40 ± 0.45	0.8	0.23 ± 0.36	1.1
G_{12}	0.06 ± 0.49	42.0	–0.12 ± 0.40	55.5

[a] Negatively charged phosphates are replaced by neutral methylphosphonate groups.

[b] In view of the hole transfer from site G_3 to site G_{12}, the distribution was normalized to the range the donor to the acceptor site. Each $(T{\cdot}A)_2$ unit at either end of the duplex beyond the G sites would contribute about 5% (in absolute terms).

correlated by their hydration shell which partially screens their long-range Coulomb effect. To estimate directly the role of environmental fluctuations in modulating ΔG^{o} for hole transfer, we considered a model system where only environmental fluctuations were accounted for, namely the rigid oligomer 5′-$(T)_5G(T)_2G(T)_5$-3′. Standard deviations of ΔG^{o} for hole transfer in such a rigid system (~0.15 eV) do not differ significantly from the standard deviations obtained for the corresponding system with a flexible DNA (~0.19 eV), confirming the key role of movements in the environment for modulating the hole-state energetics.

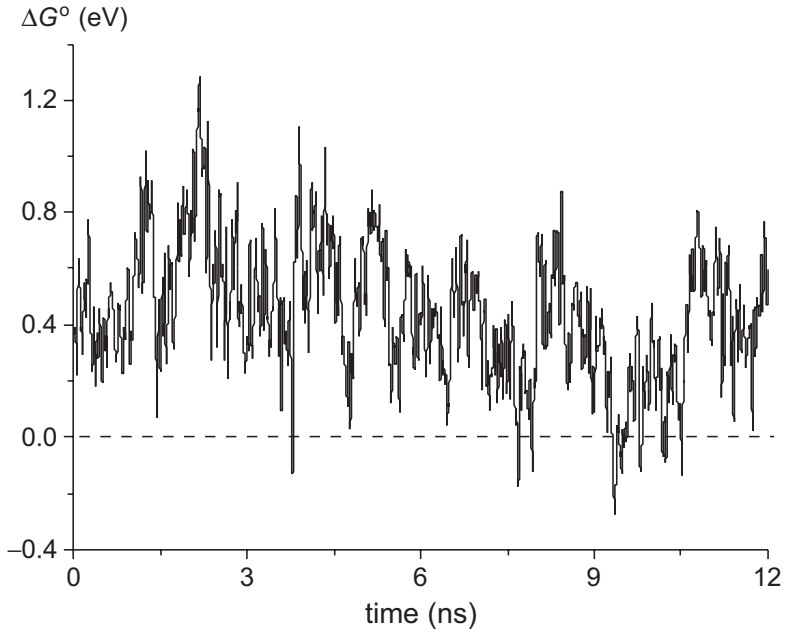

Figure 6.7 Fluctuations of the relative energy ΔG^{o} for hole transfer from G_3 to A_6 in the duplex 5′- $TTG_3T_4T_5T_6T_7T_8T_9T_{10}T_{11}G_{12}TT$ -3′, calculated along a MD trajectory of 12 ns.

To quantify the effect of the counterions directly, we modeled a modified DNA duplex 5′- $(T)_2G(T)_8G(T)_2$ -3′ where the negatively charged phosphate groups of the DNA backbone were replaced by neutral methylphosphonate groups.[208] Unlike the highly charged original duplex (26 e^-), the modified duplex is neutral and, therefore, was modeled without counterions. The energy difference between G^+ and A^+ in modified DNA is reduced to Δ = ~0.3 eV (Table 6.5). Analysis of OxP shows that on going from modified (neutral) DNA to normal DNA, the stabilization of the cations due to the phosphates is not quite compensated by the hydrated counterions and an overall stabilization of the cation states results. Therefore, the reduced energy gap Δ of modified DNA implies that, on average, the A^+ states are better stabilized than G^+ states. Thus, the possibility for hole hopping onto the A sites of the bridge should increase. Also, the standard deviations of the relative energies of modified DNA were calculated only ~20% smaller than for normal DNA. Apparently, the movement of the water molecules causes the dominating contribution (~80%) to the energy variation of hole states.

Changes in the energies of the hole states on different sites of a duplex are correlated. The correlation coefficients between energies of neighboring pairs of

unmodified DNA (~0.5) rapidly decreased with distance to 0.30, 0.11, 0.02 for the second, third, and forth neighbors, respectively. However, energies of states on remote sides tend to be negatively related; for instance, the correlation coefficients between hole states at G_3 on the one hand and A_8, A_{10} and G_{12} on the other were –0.12, –0.20 and –0.31, respectively. The corresponding correlations for the modified duplex are somewhat weaker, –0.02, –0.16 and –0.29. Negative correlations are due to changes in the part of the environment between the sites; for instance, the rotation of the dipole of a water molecule directed along the DNA axis to G_3 by 180° will stabilize of a hole state at G_3 and concomitantly destabilize of a hole state at G_{12}. Also movement of counterions along DNA located between the considered sites contributes to the negative correlation.

Finally, we address the distribution of the hole states for charge transfer from the donor site G_3 to the acceptor site G_{12} in the 14-mer duplex 5′-$TTG_3(T)_8G_{12}TT$-3′ and its methylphosphonate derivative (Table 6.5). We estimated this distribution as the fraction of time (along trajectories of 12 ns) when the corresponding hole state has the lowest energy compared to all other sites under consideration. The longest time interval for a hole resting on one of the G sites (i.e. when this hole state is lower than the cation states at all other sites) is about 100 ps, whereas this resting time is at most 12 ps on a $(T{\cdot}A)_8$ bridge. Non-negligible probabilities (~1%) were determined for events where an electron hole is localized on adenine bases. The total fractions of preferred bridge sites of unmodified and modified DNA (7–8%) are similar, but the distributions over the bridge sites differ in a characteristic way. Thus, the dynamics of water molecules and counterions considerably modulates the relative redox potentials of the nucleobases. As a result, fluctuations of the environment can render hole transfer processes from an G·C base pair to an A·T pair in a DNA duplex energetically feasible.

Analysis of our QM/MD simulations supports the recently proposed ion-gated mechanism for charge transfer in DNA.[208] Note, however, that the thermal movement of the water molecules significantly dominates the variation of the hole-state energies in DNA. The fluctuations of relative energies of radical cation states are significant even in the absence of counterions, as in the case of a modified duplex with methylphosphonate groups instead of phosphates in the backbone. Our results suggest that adenine bases can also act as intermediates of electron hole transfer. Thermal fluctuations of counterions and water molecules around DNA are responsible for configurations where the free energy of

charge transfer from a guanine to an adenine is negative. Such configurations are implied in the recently suggested A-hopping mechanism or, in a wider sense, in a domain mechanism as recently inferred on the basis of experimental data.[37,84] Obviously, fluctuations of the charge transfer driving force should be accounted when estimating the CT rate constants within the thermally induced hole-hopping model.[7,199] Further experiments probing the role of environmental fluctuations in a quantitative way are highly desirable for gaining a quantitative understanding of the CT mechanism in DNA.

6.4 Conclusion

We have carried out the constrained and unconstrained molecular dynamics simulations of duplexes with 13−16 base pairs. The oxidation energies of guanine and adenine bases were then calculated by taking into account electrostatic effects of neighboring base pairs, sugar-phosphate backbone, water molecules and sodium ions as point charges with dielectric constant of 2. Additionally, the model neglecting these effects was also considered for comparison. Without taking into account electrostatic interactions, OxP of guanine obtained for the duplex of ideal structure is lower than that calculated for flexible structure. Also, a larger energy gap between guanine and adenine was found in flexible duplexes. However, when the electrostatic interactions with surroundings were accounted for, the energy gap decreased. For 12 ns MD trajectory, we found the events of remarkable probability (3%−8%) where oxidation potentials of adenines are lower than those of guanines. In contrast, OxP of adenine remain larger than OxP of guanine when the environment effects are excluded. By comparing the oxidation energy of nucleobases in normal DNA and modified DNA, we found that water molecules play a more important role for modulating the energy gap between radical cation states localized on guanine and adenine than sodium ions.

Chapter 7

Summary

In the semi-classical picture of Marcus theory, three main factors control the rate constant of non-adiabatic charge transfer between donor *d* and acceptor *a*: the electronic coupling matrix element V_{da}, the reorganization energy λ, and the driving force ΔG^{o}. In this dissertation, we investigated these parameters for hole transfer in DNA by using a combined QM/MD approach. MD simulations of 16 DNA duplexes in explicit solvent were performed to generate MD trajectories, containing the configurations of DNA and environment molecules. The MD simulations to obtain trajectories of 2–12 ns were carried out by a well-established simulation protocol employing the AMBER force field. Analysis of the distribution of water molecules and counterions around DNA showed good agreement with known theoretical and experimental results. The DNA and environment configurations were then used to study the factors controlling the charge transfer rate in DNA. The results are summarized below.

Electronic matrix elements for hole transfer between adjacent Watson–Crick pairs in DNA have been calculated at the Hartree–Fock SCF level for various conformations of the dimer duplexes [(AT),(AT)], [(AT),(TA)], [(TA),(AT)]. Example configurations of [(TA),(TA)] have also been extracted from molecular dynamics simulations of a decamer duplex. The calculated electronic coupling is very sensitive to variations of the mutual position of the Watson–Crick pairs. The intra-strand A–A interaction is more susceptible to conformational changes than the corresponding inter-strand interaction. The rate of charge migration as measured by the square of the electronic coupling matrix element may vary several hundred-fold in magnitude due to moderate changes of the duplex

conformation. Thus, thermal fluctuations of the DNA structure have to be taken into account when one aims at a realistic description of the electron hole transfer in DNA.

The solvent reorganization energy λ_s can have a significant effect on the activation energy for charge transfer in DNA and its dependence on donor–acceptor distance R_{da}. To estimate λ_s and the resulting effective contribution β_s to the falloff parameter β for the overall transfer rate constant, the Poisson equation was solved numerically for several systems representing DNA duplexes, 5′-GGGT$_n$GGG-3′, n = 0–6, in a realistically structured heterogeneous dielectric, as determined by molecular dynamics simulations. The charge transfer was modeled primarily for holes localized on single guanine bases. Effects of thermal fluctuations on λ_s were taken into account via structures for a given duplex sampled from MD trajectories. Calculated values of λ_s were found to be rather insensitive to thermal fluctuations of the DNA fragments, but depended in crucial fashion on details of the dielectric model (shape and dielectric constants of various zones) that was used to describe the polarization response of the DNA and its environment to the charge transfer. λ_s was calculated to increase rapidly at small R_{da} values (< 15 Å), and accordingly the falloff parameter β_s (defined as a local function of R_{da}) decreases appreciably with increasing R_{da} (from 1.0 Å^{-1} with only one intermediate base pair between *d* and *a* to 0.15 Å^{-1} for systems with five intervening pairs). Calculated λ_s values were accurately fitted (standard deviation of ~0.5 kcal/mol) to a linear function of $1/R_{da}$, including all cases except contact (R_{da} = 3.4 Å), where some overlap of *d* and *a* sites may occur. A linear fit to an exponential (of form $\exp(-\beta_s R_{da})$) gave comparable accuracy for the entire R_{da} range. λ_s based on *d* and *a* holes delocalized over two adjacent guanine bases was uniformly ~12 kcal/mol smaller than the corresponding results for holes localized on single guanines, almost independent of R_{da}. The internal reorganization energy for hole transfer between G·C pairs was calculated at 16.6 kcal/mol at the B3LYP/6-31G* level.

We explored the oxidation energetics of nucleobases in duplexes comprising 13–16 Watson–Crick pairs. Electrostatic effects of the environment were accounted for with a point charge model. By comparing the energy fluctuations obtained from rigid and flexible fragments, we found that the main contribution to the variations of the oxidation potential was caused by water molecules and counterions. Time series analysis indicated that a variety of characteristic times for the energetic fluctuations (220–930 ps) is due to both

vibration of nucleobases and movement of DNA surroundings. In the neutral DNA duplex, characteristic times less than 200 ps were observed.

Our QM/MD study shows that the relative energies of radical cation states on nucleobases in DNA are considerably affected by the local distribution of water molecules and counterions while conformational changes of DNA play only a minor role in modulating the free energy change ΔG^{o} of the charge transfer. Thermal motion of the polar environment induces fluctuations of the redox potentials with a characteristic time of 0.3–0.4 ns. In fact, fluctuations of ΔG^{o} are large enough to render electron hole transfer from G^{+} to A energetically feasible, thus allowing a change-over from the generally accepted G-hopping mechanism to A-hopping. To estimate *directly* the role of counterions a modified duplex was studied where all negatively charged phosphates were replaced by neutral methylphosphonate groups and all counterions were removed. Comparison of the computational results for the normal and modified systems suggests that the dynamics of water molecules strongly dominates the ΔG^{o} fluctuations. The total fractions of time, when bridge sites are energetically preferred, are similar for normal and modified DNA (7–8%).

Appendix

An Interface of QM and MD Approaches

For studying the charge transfer in DNA using a combined quantum mechanics/molecular dynamics method, the AMBER program has been used to perform the MD simulations. The results are MD trajectories consisting of a sequence of snapshots. Each snapshot contains Cartesian coordinates of all atoms in the system. These configurations are then employed to calculate the electronic coupling matrix element V_{da}, the driving force ΔG^{o} (estimated as the difference of vertical ionization potentials (IP) of relevant nucleobases), and the reorganization energy λ. The quantities V_{da} and IP are computed quantum chemically with the semi-empirical NDDO-G method paraterized for calculating ionization potentials and excitation energies of organic molecules.[217] These calculations are performed with the program SIBIQ.[227] The last term, the reorganization energy, is calculated with program DelPhi II, which employs a finite difference solver of the Poisson equation.[196,197]

The program **ETCAT** (**E**lectron **T**ransfer **C**alculation from **A**mber **T**rajectory) is a FORTRAN 77 code used as an interface between the MD trajectories and the programs SIBIQ and DelPhi II. The program **ETCAT** requires three files to operate

(*i*) input file containing several variables that control the program (see below),

(*ii*) topology file containing the information about number, name and type of atoms, their mass and connectivity, residue names, and charges on atoms.

(*iii*) MD trajectory file storing a series of the system configurations (snapshots).

The interface procedure can be presented as a flowchart, see Figure A.1. The program starts with reading the input and topology files. Then a required snapshot is extracted from

the trajectory file. The variables defined in the input file specify snapshots which will be processed (see below). The extracted snapshot is used to prepare either an input file for the program SIBIQ to calculate V_{da} or IP, or for the program DelPhi II to calculate λ. The calculation can be repeated for several snapshots. Finally, the results, as well as the total computational time, are dumped into a output file.

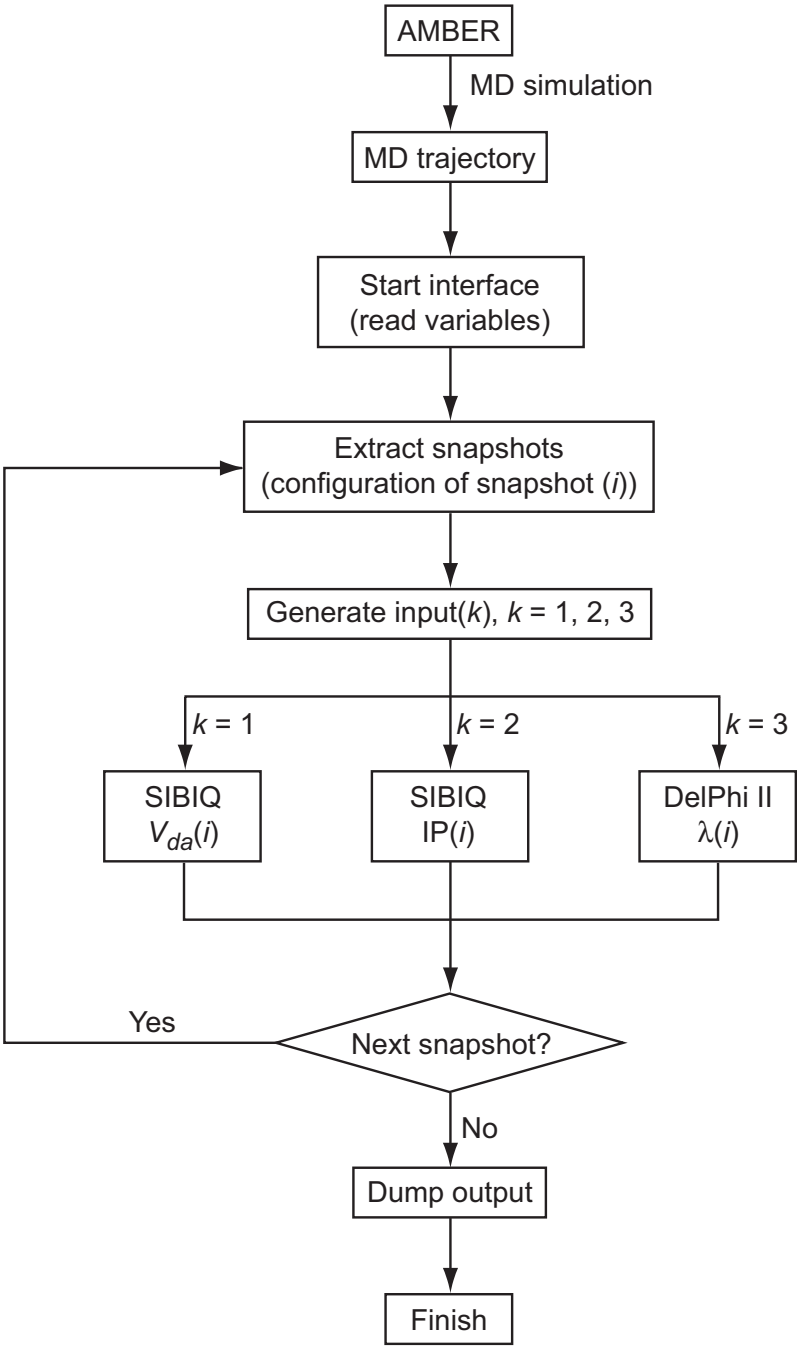

Figure A.1 A flowchart of the program interface.

A.1 Getting started

To run ETCAT, use the following command:

```
% etcat –i input –o output –p topology –x trajectory
```

The input file contains the controlling variables. A sample file, which is just to establish a basic syntax and appearance, is as follows:

```
---------------------------------------------------------------------------------------------------
Title: IP calculation for duplex d(TTGTTTGTT)
&cntrl
     NSTARTSN = 1, NSTOPSN = 100, NOFFSET = 2,
     NSTARTBP = 3, NSTOPBP = 7,
     ISIBIQ = 2, NDDOG = 1, nIP = 1, modeIP = 1,
     iptCRG = 2, CUTOFF = 20.0, IDIELC = 2,
     IdelPhi = 0, modeEPS = 5, NRESD = 3, NRESA = 7,
     fileprm_st = prm.st, fileprm_op = prm.op
&end
---------------------------------------------------------------------------------------------------
```

The first line of input contains a title, which is followed by the namelist identifier `&cntrl`. The variables followed must be one of the namelist variable (a detailed description of the namelist variables is given in the next section). Any order of the variables in the input list can be used (when a variable is specified twice or more, the latest assignment will overwrite the preceding one). Blanks may be anywhere in the input, but not within constants. Variables that are not explicitly presented retain their default values. The input is concluded by the namelist identifier `&end`. Note that the case of the names is ignored in all namelist variables, but the namelist identifier must be specified as `&cntrl`, not `&CNTRL`.

A.2 Namelist variables

The `&cntrl` namelist variables depend on the type of calculation, i.e. V_{da}, IP, or λ. There are several general variables used in all calculations.

A.2.1 General variables

`NSTARTSN` Starting snapshot (default 1)

`NSTOPSN` Final snapshot to be processed (default 1)

`NOFFSET` Offset (snapshot frequency, default 1);

The calculation will perform from `NSTARTSN` to `NSTOPSN` for every `NOFFSET`.

`ISIBIQ` Flag to run SIBIQ

= 0 no calculation (default)

= 1 calculate V_{da}

= 2 calculate IP

= 3 calculate both V_{da} and IP

`NDDOG` Semi-empirical method used for the SIBIQ calculation

= 0 AM1

= 1 NDDO-G (default)

`IDelPhi` Flag to run DelPhi II

= 0 no calculation (default)

= 1 calculate λ

A.2.2 Namelist variables for electronic coupling calculation

`NSTARTBP` Starting base pair (default 1)

`NSTOPBP` Final base pair to be processed (default is the last base pair in DNA sequence);

In case `ISIBIQ` = 1 or 3, the electronic coupling V_{da} between base pair number `NSTARTBP` and number `NSTOPBP`, including the intervening pairs, will be computed.

`modeVda` Model of V_{da} calculation; one can take into account only some parts or a whole pair of nucleotides to calculate V_{da}. Based on such DNA component, three models can be chose:

= 1 base pairs only

= 2 base pairs and sugar

= 3 base pairs, sugar and phosphate

Note that the connecting atoms not taken into account will be replaced by hydrogen atoms, e.g. C1′ atom of sugar, in case `modeVda` = 1, or phosphorus atom, in case `modeVda` = 2.

A.2.3 Namelist variables for ionization potential calculation

`NSTARTBP` Starting base pair (default 1)

`NSTOPBP` Final base pair to be processed (default is the last base pair in DNA sequence);

In the case of `ISIBIQ` = 2, the IP of base pairs from `NSTARTBP` to `NSTOPBP` are considered. For the sample input file above, the program will calculate IP of pairs $(G{\cdot}C)_3$, $(T{\cdot}A)_4$, $(T{\cdot}A)_5$, $(T{\cdot}A)_6$, and $(G{\cdot}C)_7$. In the case of `ISIBIQ` = 3, the first and last base pairs for calculating IP must be the same as set for calculating the V_{da}.

`modeIP` Model of IP calculation; like `modeVda`, one can take into account only some parts or a whole pair of nucleotides. Three models can be chose:

= 1 base pairs only

= 2 base pairs and sugar

= 3 base pairs, sugar and phosphate

The connecting atoms will be replaced by hydrogen atoms, as in the V_{da} calculation.

`nIP` Number of adjacent pairs taken into account in the quantum chemical calculation for calculating the IP of each nucleobase; only odd numbers are valid. When calculating IP of $(G{\cdot}C)_3$, for instance, if `nIP` = 1, only Watson–Crick pair $(G{\cdot}C)_3$ is taken into account. If `nIP` = 3, pairs $(T{\cdot}A)_2$, $(G{\cdot}C)_3$ and $(T{\cdot}A)_4$ are taken into account.

`iptCRG` Number of adjacent pairs taken into account as point charges (integer)

= –1 ignore electrostatic interaction of the point charges (default)

= 0 only sugar and/or phosphate groups which belong to the selected base pair(s) are treated as point charges. There is no additional adjacent pair.

> 0 number of adjacent pairs (for both 5′ and 3′ directions), including their sugar and phosphate groups. For example, if `nIP` = 1 and `iptCRG` = 2, IP of $(G{\cdot}C)_3$ will be calculated, accounting for the base pairs $(T{\cdot}A)_1$, $(T{\cdot}A)_2$, $(T{\cdot}A)_4$ and $(T{\cdot}A)_5$, and the corresponding sugar/phosphate groups as point charges.

`CUTOFF` Water molecules within a cutoff distance (from the center of the base pair) are accounted for as point charges. Default is 999.0 Å, all water molecules are treated.

`ipNa` Flag for treating sodium counterions as point charges (default 1)

= 0 ignore sodium ions. Charge on oxygen atoms of phosphate groups will be reduced to neutralize the net charge of the system.

= 1 sodium ions are taken into account as point charges.

`IDIELC` Dielectric constant, default 1.

A.2.4 Namelist variables for reorganization energy calculation

The variables below (valid only when `IDelPhi` = 1) are required for treating the system as multi-dielectric zones, where different dielectric constants are assigned to different zones. The following abbreviations are used: donor/acceptor (*d*/*a*), nucleobase (NB), sugar-phosphate (SP), and water molecule and sodium ion (WAT/Na^+).

`modeEPS` Flag to treat a dielectric zone.

= 2 two zones: (i) whole DNA, and (*ii*) bulk water

= 3 three zones: (*i*) *d*/*a*, (*ii*) NB and SP, and (*iii*) bulk water

= 41 four zones: (*i*) *d*/*a*, (*ii*) NB and SP, (*iii*) WAT/Na^+, and (*iv*) bulk water

= 42 four zones: (*i*) *d*/*a*, (*ii*) NB, (*iii*) SP, and (*iv*) bulk water

= 5 five zones: (*i*) *d*/*a*, (*ii*) NB, (*iii*) SP, (*iv*) WAT/Na^+, and (*v*) bulk water (default)

Each zone is defined by rolling a probe radius of 1.4 Å (which is a radius of water molecule) on atomic spheres of the corresponding zone.

`CUTDelPhi` Cutoff distance for WAT/Na^+ zone (default 3.0 Å); to generate this zone, atomic spheres of all water molecules and Na^+ ions within this distance far from DNA atoms are taken into account.

`NRESD` Residue number of donor

`NRESA` Residue number of acceptor

`DAepsST` Static dielectric constant of *d*/*a* zone

`DAepsOP` Optical dielectric constant of *d*/*a* zone

`BepsST` Static dielectric constant of NB zone

`BepsOP` Optical dielectric constant of NB zone

`SPepsST` Static dielectric constant of SP zone

`SPepsOP` Optical dielectric constant of SP zone

`WepsST` Static dielectric constant of WAT/Na^+ zone

`WepsOP` Optical dielectric constant of WAT/Na^+ zone

Note that if `modeEPS` = 2, then `BepsX` (X = ST or OP) is assigned for the dielectric constants of the whole DNA. Alternatively, if `modeEPS` = 3 or 41, the `BepsX` is assigned for nucleobases and sugar-phosphate groups.

Furthermore, there are four parameter files required for DelPhi II. The first two are `fileprm_st` and `fileprm_op`, which must be specified in the ETCAT input file (see the sample above). Recall that the reorganization energy is the difference of solvation energies, calculated with the static and optical dielectric constants. The files `fileprm_st` and `fileprm_op` refer to calculations with static and optical constants respectively. File names must contain less than 80 characters and they are case sensitive. The last two files are "DNA.siz" and "DNA.crg", containing atomic radius and atomic charge of each DNA atom, respectively. They are introduced in the first two files. An sample of `fileprm_st` follows here:

```
gsize = 201                                  (grid size)
in(pdb,file = "ETCAT.DelPhi.pdbST"           (pdb file, generated by ETCAT)
in(siz,file = "DNA.siz"                      (atomic radius file name)
in(crg,file = "DNA.crg"                      (atomic charge file name)
exdi = 80.0                                  (static dielectric constant of
                                             bulk water zone)
prbrad = 1.4                                 (probe radius)
```

More details on the input parameters of the program DelPhi can be found at the homepage of the program http://trantor.bioc.columbia.edu/delphi/.

Bibliography

[1] Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

[2] Eley, D. D.; Spivey, D. I. *Trans. Faraday Soc.* **1962**, *58*, 411.

[3] Barbara, P. F.; Olson, E. J. *Adv. Chem. Phys.* **1999**, *107*, 647.

[4] Schuster, G. B. *Acc. Chem. Res.* **2000**, *33*, 253.

[5] Giese, B. *Acc. Chem. Res.* **2000**, *33*, 631.

[6] Lewis, F. D.; Letsinger, R. L.; Wasielewski, M. R. *Acc. Chem. Res.* **2001**, *34*, 159.

[7] Bixon, M.; Jortner, J. *J. Am. Chem. Soc.* **2001**, *123*, 12556.

[8] Ratner, M. A. *Nature* **1999**, *397*, 480.

[9] Berlin, Y. A.; Burin, A. L.; Ratner, M. A. *J. Am. Chem. Soc.* **2001**, *123*, 260.

[10] Voityuk, A. A.; Jortner, J.; Bixon, M.; Rösch, N. *J. Chem. Phys.* **2001**, *114*, 5614.

[11] Priyadarshy, S.; Risser, S. M.; Beratan, D. N. *J. Phys. Chem.* **1996**, *100*, 17678.

[12] Kurnikov, I. V.; Tong, G. S. M.; Madrid, M.; Beratan, D. N. *J. Phys. Chem. B* **2002**, *106*, 7.

[13] Tong, G. S. M.; Kurnikov, I. V.; Beratan, D. N. *J. Phys. Chem. B* **2002**, *106*, 2381.

[14] Olofsson, J.; Larsson, S. *J. Phys. Chem. B* **2001**, *105*, 10398.

[15] Armitage, B. *Chem. Rev.* **1998**, *98*, 1171.

[16] Burrows, C. J.; Muller, J. G. *Chem. Rev.* **1998**, *98*, 1109.

[17] Kelley, S. O.; Jackson, N. M.; Hill, M. G.; Barton, J. K. *Angew. Chem. Int. Ed.* **1999**, *38*, 941.

[18] Fox, M. A. *Acc. Chem. Res.* **1999**, *32*, 201.

[19] Lisdat, F.; Ge, B.; Scheller, F. W. *Electrochem. Commun.* **1999**, *1*, 65.

[20] Halliwell, B.; Gutteridge, J. M. C. *Free Radicals in Biology and Medicine*, 2nd Ed., Clarendon Press: Oxford, 1989.

[21] Demple, B.; Harrison, L. *Annu. Rev. Biochem.* **1994**, *63*, 915.

[22] Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063.

[23] Seidel, C. A. M.; Schulz, A.; Sauer, M. H. M. *J. Phys. Chem.* **1996**, *100*, 5541.

[24] Faraggi, M.; Klapper, M. H. *J. Chim. Phys. PCB* **1994**, *91*, 1054.

[25] Steenken, S.; Jovanovic, S. V. *J. Am. Chem. Soc.* **1997**, *119*, 617.

[26] Hall, D. B.; Holmlin, R. E.; Barton, J. K. *Nature* **1996**, *382*, 731.

[27] Joachim, C.; Gimzewski, J. K; Aviram, A. *Nature* **2000**, *408*, 541.

[28] Tour, J. M. *Acc. Chem. Res.* **2000**, *33*, 791.

[29] Aviram, A.; Ratner, M. A. *Chem. Phys. Lett.* **1974**, *29*, 277.

[30] Grinstaff, M. W. *Angew. Chem. Int. Ed.* **1999**, *38*, 3629.

[31] Kelley, S. O.; Barton, J. K. *Chem. Biol.* **1998**, *5*, 413.

[32] Meggers, E.; Michel-Beyerle, M. E.; Giese, B. *J. Am. Chem. Soc.* **1998**, *120*, 12950.

[33] Voityuk, A. A.; Jortner, J.; Bixon, M.; Rösch, N. *Chem. Phys. Lett.* **2000**, *324*, 430.

[34] Saito, I.; Nakamura, T.; Nakatani, K.; Yoshioka, Y.; Yamaguchi, K.; Sugiyama, H. *J. Am. Chem. Soc.* **1998**, *120*, 12686.

[35] Voityuk, A. A.; Rösch, N.; Bixon, M.; Jortner, J. *J. Phys. Chem. B* **2000**, *104***,** 9740.

[36] Troisi, A.; Orlandi, G. *Chem. Phys. Lett.* **2001**, *344*, 509.

[37] Williams, T. T.; Odom, D. T.; Barton, J. K. *J. Am. Chem. Soc.* **2000**, *122*, 9048.

[38] Henderson, P. T.; Jones, D.; Hampikian, G.; Kann, Y. Z.; Schuster, G. B. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8353.

[39] Wan, C. Z.; Fiebig, T.; Kelley, S. O.; Treadway, C. R.; Barton, J. K.; Zewail, A. H. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6014.

[40] Kumar, K.; Lin, Z.; Waldeck, D. H.; Zimmt, M. B. *J. Am. Chem. Soc.* **1996**, *118*, 243.

[41] Han, H.; Zimmt, M. B. *J. Am. Chem. Soc.* **1998**, *120*, 8001.

[42] Cave, R. J.; Newton, M. D.; Kumar, K.; Zimmt, M. B. *J. Phys. Chem.* **1995**, *99*, 17501.

[43] Cheatham, T. E., III; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 4805.

[44] Duan, Y.; Wilkosz, P.; Crowley, M.; Rosenberg, J. M. *J. Mol. Biol.* **1997**, *272*, 553.

[45] Davis, W. B.; Hess, S.; Naydenova, I.; Haselsberger, R.; Ogrodnik, A.; Newton, M. D.; Michel-Beyerle, M. E. *J. Am. Chem. Soc.* **2002**, *124*, 2422.

[46] Saenger, W. *Principles of Nucleic Acid Structure*, Springer Verlag: New York, 1984.

[47] Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic Acids: Structures, Properties, and Functions*, University Science Books, 2000.

[48] Marcus, R. A. *J. Chem. Phys***. 1956**, *24*, 966.

[49] Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 979.

[50] Marcus, R. A. *Discuss. Faraday Soc.* **1960**, *29*, 21.

[51] Marcus, R. A. *J. Chem. Phys.* **1965**, *43*, 679.

[52] Winkler. J. R.; Gray, H. B. *Chem. Rev.* **1992**, *92*, 369.

[53] Gray, H. B.; Winkler, J. R. *Ann. Rev. Biochem.* **1996**, *65*, 537.

[54] Marcus, R. A.; Sutin, N. *Biochim. Biophys. Acta* **1985**, *811*, 265.

[55] Newton, M. D. *J. Phys. Chem.* **1988**, *92*, 3049.

[56] Farazdel, A.; Dupuis, M.; Clementi, E.; Aviram, A. *J. Am. Chem. Soc.* **1990**, *112*, 4206.

[57] Newton, M. D. *Chem. Rev.* **1991**, *91*, 767.

[58] Rodriguez-Monge, L.; Larsson, S. *Int. J. Quantum Chem.* **1997**, *61*, 847.

[59] Cave, J. R.; Newton, M. D. *Chem. Phys. Lett.* **1996**, *249*, 15.

[60] Cave, J. R.; Newton, M. D. *J. Chem. Phys.* **1997**, *106*, 9213.

[61] Voityuk, A. A.; Rösch, N. *J. Chem. Phys.* **2002**, *117*, 5607.

[62] Rösch, N.; Voityuk, A. A. In *Topics in Current Chemistry*, Schuster, G., Ed.; Springer: Heidelberg, 2004; Vol. 237, pp 37.

[63] Elliott, C. M.; Derr, D. L.; Matyushov, D. V.; Newton, M. D. *J. Am. Chem. Soc.* **1998**, *120*, 11714.

[64] Rust, M.; Lappe, J.; Cave, R. J. *J. Phys. Chem. A* **2002**, *106*, 3930.

[65] Kuznetsov, A. M.; Ulstrup, J. *Electron Transfer in Chemistry and Biology: An Introduction to the Theory*, John Wiley & Sons, Inc., 1998.

[66] Tavernier, H. L.; Fayer, M. D. *J. Phys. Chem. B* **2000**, *104*, 11541.

[67] Doorn, S. K.; Hupp, J. T. *J. Am. Chem. Soc.* **1989**, *111*, 4704.

[68] Myers, A. B. *Chem. Rev.* **1996**, *96*, 911.

[69] Lewis, F. D.; Kalgutkar, R. S.; Wu, Y.; Liu, X.; Liu, J.; Hayes, R. T.; Miller, S. E.; Wasielewski, M. R. *J. Am. Chem. Soc.* **2000**, *122*, 12346.

[70] Nelsen, S. F.; Balckstock, S. C.; Kim, Y. *J. Am. Chem. Soc.* **1987**, *109*, 677.

[71] Rauhut, G.; Clark, T. *J. Am. Chem. Soc.* **1993**, *115*, 9127.

[72] Marcus, R. A. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155.

[73] Liu, Y.-P.; Newton, M. D. *J. Phys. Chem.* **1994**, *98*, 7162.

[74] Liu, Y.-P.; Newton, M. D. *J. Phys. Chem.* **1995**, *99*, 12382.

[75] Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.

[76] Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.

[77] Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

[78] Yonemoto, E. H.; Saupe, G. B.; Schmehl, R. H.; Hubig, S. M.; Riley, R. L.; Iverson, B. L.; Mallouk, T. E. *J. Am. Chem. Soc.* **1994**, *116*, 4786.

[79] Giese, B. *Annu. Rev. Biochem.* **2002**, *71*, 51.

[80] Jortner, J.; Bixon, M.; Langenbacher, T.; Michel-Beyerle, M. E. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12759.

[81] Newton, M. D.; Sutin, N. *Annu. Rev. Phys. Chem.* **1976**, *64*, 4860.

[82] Nunez, M. E.; Barton, J. K. *Curr. Opin. Struct. Biol.* **2000**, *4*, 199.

[83] Bixon, M.; Giese, B.; Wessely, S.; Langenbacher, T.; Michel-Beyerle, M. E.; Jortner, J. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11713.

[84] Giese, B.; Amaudrut, J.; Köhler, A. K.; Spormann, M.; Wessely, S. *Nature* **2001**, *412*, 318.

[85] Felts, A. K.; Pollard, W. T.; Friesner, R. A. *J. Phys. Chem.* **1995**, *99*, 2929.

[86] Okada, A.; Chernyak, V.; Mukamel, S. *J. Phys. Chem. A* **1998**, *102*, 1241.

[87] Segel, D.; Nitzan, A.; Davis, W. B.; Wasielewski, M. R.; Ratner, M. A. *J. Phys. Chem. B* **2000**, *104*, 3817.

[88] Grozema, F. C.; Berlin, Y. A.; Siebbeles, L. D. A. *J. Am. Chem. Soc.* **2000**, *122*, 10903.

[89] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc*. **1995**, *117*, 5179.

[90] MacKerell, A. D.; Wiórkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.

[91] Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596.

[92] Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.

[93] Leach, A. R. *Molecular modelling: Principles and applications*, Longman, 1996.

[94] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, Clarendon Press: Oxford, 1987.

[95] Haile, J. M. *Molecular Dynamics Simulation: Elementary Methods*, John Wiley: New York, 1992.

[96] Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208.

[97] McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.

[98] Cheatham, T. E., III; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435.

[99] Verlet, L. *Phys. Rev.* **1967**, *159*, 98.

[100] Hockney, R. W. *Methods Comput. Phys.* **1970**, *9*, 136.

[101] Swope, W. C.; Anderson, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637.

[102] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

[103] Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501.

[104] van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem. Int. Ed.* **1990**, *29*, 992.

[105] Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667.

[106] Ravishanker, G.; Auffinger, P.; Langley, D. R.; Jayaram, B.; Young, M. A. In *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B., Eds.; Wiley-VCH: New York, 1997; Vol. 11, pp 317.

[107] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

[108] Ewald, P. P. *Ann. Phys.* **1921**, *64*, 253.

[109] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

[110] Cheatham, T. E., III; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434.

[111] Young, M. A.; Ravishanker, G.; Beveridge, D. L. *Biophys. J.* **1997**, *73*, 2313.

[112] Feig, M.; Pettitt, B. M. *Biophys. J.* **1999**, *77*, 1769.

[113] Hamelberg, D.; McFail-Isom, L.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2000**, *122*, 10513.

[114] Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182.

[115] Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **1998**, *8*, 227.

[116] Yang, L.; Pettitt, B. M. *J. Phys. Chem.* **1996**, *100*, 2564.

[117] Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Am. Chem. Soc.* **1997**, *119*, 59.

[118] Cieplak, P.; Cheatham, T. E., III; Kollman, P. A. *J. Am. Chem. Soc*. **1997**, *119*, 6722.

[119] Feig, M.; Pettitt, B. M. *J. Phys. Chem. B* **1997**, *101*, 7361.

[120] Arnott, S.; Selsing, E. *J. Mol. Biol.* **1974**, *88*, 551.

[121] Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86.

[122] MacKerell, A. D., Jr.; Banavali, N. *J. Comput. Chem.* **2000**, *21*, 105.

[123] Tapia, O.; Velazquez, I. *J. Am. Chem. Soc.* **1997**, *119*, 5934.

[124] Daura, X.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **1998**, *19*, 535.

[125] Norberg, J.; Nilsson, L. *Biophys. J.* **2000**, *79*, 1537.

[126] Harvey, S. C. *Proteins* **1989**, *5*, 78.

[127] Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R. *Biophys. J.* **1992**, *63*, 751.

[128] Auffinger, P.; Beveridge, D. L. *Chem. Phys. Lett.* **1995**, *234*, 413.

[129] Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.

[130] Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155.

[131] Mazur, A. K. *J. Am. Chem. Soc.* **1998**, *120*, 10928.

[132] Kimura, R. S.; Brower, R. C.; Zhang, C.; Sugimori, M. *J. Chem. Phys.* **2000**, *112*, 7723.

[133] Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.

[134] Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187.

[135] Beveridge, D. L. In *Encyclopedia of Computational Chemistry*, Schleyer, P. v. R. Ed.; John Wiley & Sons, Inc., 1998; Vol. 3, pp 1620.

[136] Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591.

[137] Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712.

[138] Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 9571.

[139] Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465.

[140] Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.

[141] Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.

[142] Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. *Acc. Chem. Res.* **2000**, *33*, 889.

[143] Cheatham, T. E., III; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265.

[144] Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426.

[145] Ivanov, V. I.; Krylov, D. *Methods Enzymol.* **1992**, *211*, 111.

[146] Franklin, R. F.; Gosling, R. G. *Acta Crystallogr.* **1953**, *6*, 673.

[147] Polh, F. M. *Nature* **1976**, *260*, 365.

[148] Zehfus, M. H.; Johnson, W. C. J. *Biopolymers* **1984**, *23*, 1269.

[149] Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 4658.

[150] Piskur, J.; Rupprecht, A. *FEBS Lett.* **1995**, *375*, 174.

[151] Cheatham, T. E., III; Crowley, M. F.; Fox, T.; Kollman, P. A. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9626.

[152] Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487.

[153] Hud, N. V.; Polak, M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 293.

[154] Schneider, B.; Cohen. D. M.; Schleifer, L.; Srinivasan, A. R.; Olson, W. K.; Berman, H. M. *Biophys. J.* **1993**, *65*, 2291.

[155] Berman, H. M. *Curr. Opin. Struct. Biol.* **1994**, *4*, 345.

[156] Kochoyan, M.; Leroy, J. L. *Curr. Opin. Struct. Biol.* **1995**, *5*, 329.

[157] Mezei, M.; Beveridge, D. L. *Methods Enzymol.* **1986**, *127*, 21.

[158] Rudnicki, W. R.; Pettitt, B. M. *Biopolymers* **1997**, *41*, 107.

[159] Bonvin, A. M. J. J.; Sunnerhagen, M.; Otting, G.; van Gunsteren, W. F. *J. Mol. Biol.* **1998**, *282*, 859.

[160] Nagan, M. C.; Kerimo, S. S.; Musier-Forsyth, K.; Cramer, C. J. *J. Am. Chem. Soc.* **1999**, *121*, 7310.

[161] Kosztin, D.; Gumport, R. I.; Schulten, K. *Nucleic Acids Res.* **1999**, *27*, 3550.

[162] Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 7666.

[163] Jin, R.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 8939.

[164] McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2000**, *304*, 803.

[165] Auffinger, P.; Westhof, E. *J. Mol. Biol.* **2000**, *300*, 1113.

[166] Lu, X.-J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300*, 819.

[167] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revision A.7,* Gaussian, Inc.: Pittsburgh PA, 1998.

[168] Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. R.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 6*, University of California, San Francisco, 1999.

[169] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys*. **1983**, *79*, 926.

[170] Miller, P. S.; Dreon, N.; Pulford, S. M.; McParland, K. *J. Biol. Mol.* **1980**, *255*, 9659.

[171] Bower, M.; Summers, M. S.; Powell, C.; Shinozuka, K.; Regan, J. B.; Zon, G.; Wilson, W. D. *Nucleic Acids Res.* **1987**, *15*, 4915.

[172] Lesnikowski, Z. J.; Jaworska, M.; Stec, W. J. *Nucleic Acids Res.* **1990**, *18*, 2109.

[173] Reynolds, M. A.; Hogrefe, R. I.; Jaeger, J. A.; Schwartz, D. A.; Riley, T. A.; Marvin. W. B.; Daily, W. J.; Vagnefi, M. M.; Beck, T. A.; Knowles, S. K.; Klem, R. E.; Arnold, L. J. *Nucleic Acids Res.* **1996**, *24*, 4584.

[174] Thiviyanathan, V.; Vyazovkina, K. V.; Gozansky, E. K.; Bichenchova, E.; Abramova, T. V.; Luxon, B. A.; Lebedev, A. V.; Gorenstein, D. *Biochemistry* **2002**, *41*, 827.

[175] Hamelberg, D.; Williams, L. D.; Wilson, W. D. *Nucleic Acids Res.* **2002**, *30*, 3615.

[176] Prive, G. G.; Yanagi, K.; Dickerson, R. E. *J. Mol. Biol.* **1991**, *217*, 177.

[177] Lavery, R.; Sklenar, H. J. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63.

[178] Probst, M. M.; Radnai, T.; Heinzinger, K.; Bopp, P.; Rode, B. M. *J. Phys. Chem.* **1985**, *89*, 753.

[179] Forester, T. R.; McDonald, I. R. *Mol. Phys.* **1991**, *72*, 643.

[180] York, D. M.; Darden, T.; Deerfield, D., II; Pedersen, L. G. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1992**, *19*, 145.

[181] Voityuk, A. A.; Michel-Beyerle, M. E.; Rösch, N. *Chem. Phys. Lett.* **2001**, *342*, 231.

[182] Hutter, M.; Clark, T. *J. Am. Chem. Soc*. **1996**, *118*, 7574.

[183] Hunter, C. A.; Lu, X.-J. *J. Mol. Biol.* **1997**, *265*, 603.

[184] Lu, X.-J.; El Hassan, M. A.; Hunter, C. A. *J. Mol. Biol.* **1997**, *273*, 681.

[185] Clowney, L.; Jain, S. C.; Srinivasan, A. R.; Westbrook, J.; Olson, W. K.; Berman, H. W. *J. Am. Chem. Soc.* **1996**, *118*, 509.

[186] Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648.

[187] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

[188] Newton, M. D. *Adv. Chem. Phys.* **1999**, *106*, 303.

[189] Gao, J. In *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH: New York, 1995; Vol. 7, pp 119.

[190] Voityuk, A. A.; Siriwong, K.; Rösch, N. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5421.

[191] Hess, S.; Götz, M.; Davis, W. B.; Michel-Beyerle, M. E. *J. Am. Chem. Soc.* **2001**, *123*, 10046.

[192] Lewis, F. D.; Wu, T.; Zhang, Y.; Letsinger, R. L.; Greenfield, S. R.; Wasielewski, M. R. *Science* **1997**, *277*, 673.

[193] Makarov, V.; Pettitt, B. M.; Feig, M. *Acc. Chem. Res.* **2002**, *35*, 376.

[194] Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. *Biochemistry* **1998**, *37*, 8341.

[195] Yang, L.; Weerasinghe, S.; Smith, P. E.; Pettitt, B. M. *Biophys. J.* **1995**, *69*, 1519.

[196] Rocchia, W.; Alexov, E; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507.

[197] Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128.

[198] Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671.

[199] Bixon, M.; Jortner, J. *Chem. Phys.* **2002**, *281*, 393.

[200] Sharp, K. A. *Biophys. J.* **1998**, *73*, 1241.

[201] Voityuk, A. A.; Rösch, N. *J. Phys. Chem. B* **2002**, *106*, 3013.

[202] LeBard, D. N.; Lilichenko, M.; Matyushov, D. V.; Berlin, Y. A.; Ratner, M. A. *J. Phys. Chem. B* **2003**, *107*, 14509.

[203] Banavali, N. K.; Roux, B. *J. Phys. Chem. B* **2002**, *106*, 11026.

[204] Levine, S.; Wringley, H. E. *Discuss. Faraday Soc.* **1957**, *24*, 43.

[205] Newton, M. D.; Basilevsky, M. V.; Rostov, I. V. *Chem. Phys*. **1998**, *232*, 201.

[206] Ly, D.; Kan, Y.; Armitage, B.; Schuster, G. B. *J. Am. Chem. Soc.* **1996**, *118*, 8747.

[207] Saito, I.; Takayama, M.; Sugiyama, H.; Nakatani, K.; Tsuchida, A.; Yamamoto, M.; *J. Am. Chem. Soc.* **1995**, *117*, 6406.

[208] Barnett, R. N.; Cleveland, C. L.; Joy, A.; Landman, U.; Schuster, G. B. *Science*. **2001**, *294*, 567.

[209] Barnett, R. N.; Cleveland, C. L.; Landman, U.; Boone, E.; Kanvah, S.; Schuster, G. B. *J. Phys. Chem. A* **2003**, *107*, 3525.

[210] Impey, R. W.; Madden, P.A.; McDonald, I. R. *J. Phys. Chem*. **1983**, *87*, 5071.

[211] García, A. E.; Stiller, L. *J. Comput. Chem*. **1993**, *14*, 1396.

[212] Rocchi, C.; Bizzarri, A. R.; Cannistraro, S. *Chem. Phys.* **1997**, *214*, 261.

[213] Brunne, R. M.; Liepinsh, E.; Otting, G.; Wüthrich, K.; van Gunsteren, W. F. *J. Mol. Biol*. **1993**, *231*, 1040.

[214] Feig, M.; Pettitt, B. M. *J. Mol. Biol.* **1999**, *286*, 1075.

[215] Lyubartsev, A. P.; Laaksonen, A. *J. Chem. Phys*. **1999**, *111*, 11207.

[216] Kim, N. S.; LeBreton, P. R. *J. Am. Chem. Soc*. **1996**, *118*, 3694.

[217] Voityuk, A. A.; Zerner, M. C.; Rösch, N. *J. Phys. Chem. A* **1999**, *103*, 4553.

[218] Siriwong, K.; Voityuk, A. A.; Newton, M. D., Rösch, N. *J. Phys. Chem. B* **2003**, *107*, 2595.

[219] Box, G. E. P.; Jenkins, G. *Time Series Analysis: Forecasting and Control*, Holden-Day, 1976.

[220] Smith, J. O., III. *Mathematics of the Discrete Fourier Transform*, W3K Publishing, 2003.

[221] http://plasma-gate.weizmann.ac.il/Grace/

[222] Graham, R. C. *Data analysis for the chemical sciences: A guide to statistic techniques*, VCH Publishers, Inc.: New York, 1993.

[223] Kindt, J. T.; Schmuttenmaer, C. A. *J. Chem. Phys*. **1997**, *106*, 4389.

[224] Gaiduk, V. I.; Gaiduk, V. V. *Mendeleev Commun.* **1997**, *7*, 76.

[225] Schwenk, C. F.; Loeffler, H. H.; Rode, B. M. *J. Am. Chem. Soc.* **2003**, *125*, 1618.

[226] Pal, S. K.; Zhao, L.; Zewail, A. H. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8113.

[227] Voityuk, A. A. *SIBIQ 5.0 A program for semi-empirical calculations*, 1998.