Hybrid QM/MM molecular dynamics simulation in drug-DNA interactions

Thesis submitted for the degree of

Doctor Philosophiæ

Candidate: **Supervisor:**

Katrin Spiegel   Prof. Paolo Carloni

October, 15th 2004

Contents

Introduction

Many anticancer therapies are based on the interactions of drugs with DNA. By binding covalently or non covalently to DNA, drugs may inhibit replication, shield the adducts from nucleotide excision repair, and interfere with transcription by recruiting essential transcription factors from their native binding sites, thus causing the death of cancerous cells. Sequence-specificity and affinity of drugs and proteins towards particular DNA-sequences relies on sequence dependent deviations from canonical B-DNA. For example, DNA sequences containing A-tracts are intrinisically bent as shown by gel-electrophoresis and electron microscopy [1, 2]. These sequence dependent variations can be used as finger prints by proteins and drugs, which recognize and bind selectively to specific DNA-sequences.

Molecular modeling based on effective potentials has proven to be of great help in characterizing the structure, dynamics and energetics of DNA in the free state[3] [4]. and in complex with proteins [5, 6] and small molecules [7]. In fact, the development of realistic force fields and protocols is allowing ever larger systems (including DNA/protein complexes) to be studied for simulation times that can reach the 100-ns time-scale. Still, despite the successes, there are several areas of DNA-modeling in which the use of effective potentials may encounter difficulties. This is due to the fact that there is a large class of phenomena that depend on the electronic structure in such an intricate way that they cannot be modeled via effective potentials. One of such areas is the description of transition metals/DNA complexes, where subtle chemical phenomena play a significant role, such as the ligand field and the fact that the metal ion ligand bond has a partially covalent nature. Another area deals with the description of bond breaking-bond forming processes, which come into play in the binding of drugs which forms covalent bonds to DNA.

An alternative to investigate such systems is offered by QM/MM approaches, in which the lesion is treated at the quantum chemical level and the biomolecular frame, including the solvent and the counter ion, is treated with an effective potential [8, 9, 10]. In the case of DNA, the solvent environment has been shown to be absolutely

essential for the electronic properties [11, 12].

Here we present the first application of such methodologies to drug/DNA adducts. We use the novel Car-Parrinello/MM approach[13, 14], which has been widely used in the context of protein catalysis[15, 16, 17, 18]. This method appears to be very well suited as:

- Drug-DNA interactions are described accurately without the need for new force field parameters. This is particularly important in drugs involving transition metals, such as cisplatin, where the exact coordination geometry depends on the electronic structure of the metal-DNA complex, which is difficult to describe by classical force-fields.

- Chemical reactions between the drug and the DNA can be simulated taking into account the entire biomolecular frame. Inclusion of the entire DNA environment is even more important than in protein catalysis since the electrostatic potential of the highly charged DNA double helix polarizes the drug significantly and may change its reactivity.

- Dispersion interactions are better described by the classical force-field than by DFT. Thus, including stacking energies by the classical force-field has actually an advantage over a full QM treatment.

Thus, a quantum chemical, parameter free approach for the drug-containing moieties in combination with well-established potentials for the macromolecular frame is not only very useful but necessary when discussing the origin of DNA catalytic activity.

The QM/MM method used here has been developed in the lab of Prof. Röthlisberger [19, 13, 14], and it has already been used succesfully to describe chemical and enzymatic reactions by the group in SISSA [15, 20]. The projects discussed in this thesis are carried out in collaboration with the group of Prof. Röthlisberger.

In my thesis I will discuss several aspects of drug-DNA interactions. I will present two conceptually very different anticancer drugs, belonging to two important classes of compounds: The first is cisplatin, which is one of the first anticancer drugs that entered clinical treatment and which has lived an unreached story of success in the treatment of cancer. Since its discovery in the 60's, intense research on cisplatin and its derivative has provided a wealth of experimental and theoretical data giving insights into the kinetic and structural behavior of cisplatin-DNA complexes. Therefore, we can validate our method by comparison to published work.

The second drug belongs to the class of duocarmycins, derived from a natural antibiotic, which binds specifically to the minor groove of DNA. The natural product has shown to have antitumoral activity and currently one derivative has entered clinical test phases [21]. Duocarmycins are very stable in solution, but undergo a remarkable activation upon DNA-binding. The origin of catalytic activity is still under debate and several hypotheses including conformational changes in both the drug and the DNA have been discussed. The QM/MM approach along with a variety of other computational tools allows to investigate both large scale dynamics of the complex, the binding and the alkylation reaction and to elucidate the origin of catalytic activity in DNA. The quantum simulations will be complemented by classical MD in order to investigate conformational changes in DNA upon drug binding before the alkylation reaction, as well as the hydration and structural flexibility of DNA in solution. We will show that these results are useful to explain differences in reactivity and they provide further information for future drug-design.

My thesis is divided into five chapters as follows:

- Chapter I consists of a brief history of DNA structure elucidation followed by an explanation of DNA structure and nomenclature according to the IUPAC definitions.

- In Chapter II I present the methods used, i.e. classical molecular dynamics and hybrid QM/MM molecular dynamics with emphasis on the application to nucleic acids.

- Chapter III deals with the interaction of cisplatin with DNA, which I studied by means of QM/MM molecular dynamics simulation. This study both illuminates differences between the crystal and solution structure of the drug-DNA adduct and validates the new approach for modeling Pt/DNA interactions.

- Chapter IV investigates the origin of DNA catalytic activity towards duocarmycin. Duocarmycin is a minor groove binder, which alkylates DNA at specific adenine sites. Both classical and QM/MM molecular dynamics simulations are applied to explain the reaction mechanism and differences in reactivity between several derivatives.

- Chapter V summarizes the results and present a short outlook of ongoing work, which is based on the results obtained from the study of duocarmycin binding

to DNA. It contains a qualitative picture of the dissociation of different duo-carmycins from the DNA minor groove, obtained by means of steered molecular dynamics.

Chapter 1

Some facts about Nucleic acids

In the early 1950's not much was know about DNA structure. New techniques, such as X-ray crystallography just started to be introduced in structure analysis of biological molecules. Rosalind Franklin, a new research associate in John Randalls lab at Kings College, was given the task to elucidate DNA structure by this technique. She was the first scientist to discover that the phosphate-ribose backbone lies at the outside of the DNA molecule and she also elucidated the basic helical structure of the macromolecule. Her knowledge was a fundamental key in the breakthrough of DNA structure elucidation by Watson and Crick. Another piece in the jigsaw puzzle was Chargaff's 1:1 relationship between the adenine-thymine and guanine-cytosine content, which resulted to be of general validity in all organisms [22]. Crick realized that this quantitave relationship could be caused by complementary base-pairing between adenine and thymine and guanine and cytosine, respectively and he envoked the hydrogen bonding pattern now known as *Watson-Crick hydrogen bonds* (Figure 1). Based on this information, Watson and Crick proposed a model of a right-handed, antiparallel double helix, in which the phosphate-ribose backbone lied at the outside of the helix and the bases pointed towards the center of the helix, held together by these specific hydrogen bonds. This leads to several obsevations:

- The charged phosphate groups are solvent exposed and can therefore easily be neutralized.

- The base-pairing rule offers a simple copying mechanism: One strand serves as a template, and the missing strand can be reconstructed thanks to the recognition of the complentary base by hydrogen bonding.

The structure proposed by Watson and Crick is now known as B-DNA. In ideal B-DNA, the base-pairs are perpendicular to the helix axis and twisted by $\sim 36°$ with
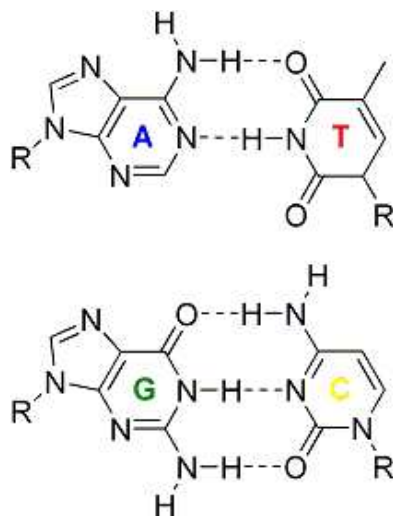
Figure 1.1: Watson-Crick base pairing. Top: Adenine(A) with Thymine (T), Bottom: Guanine(G) with Cytosine(C).

respect to each other. Therefore 10 base pairs constitute an entire turn in the double helix. The two strands of the double helix are separated by two distinct grooves: the *minor groove* and the *major groove*. The narrow minor groove contains ordered water molecules, which form a characteristic hydration pattern, either a *spine of hydration* (A-tracts) or a *ribbon of hydration* (normal DNA-sequences). The correctness of the model was ultimately prooved by X-ray crystallography. In 1952, Rosalind Franklin discovered by the same technique different forms of DNA, the dry and the wet form, now known as A-DNA and B-DNA (Figure 1.2).

The dry A-DNA is a right-handed double helix in which hydrophobic stacking interactions between base-pairs are optimized by an inclination of the base-pair plane with respect to the helix axis of 19°. This leads to a wider minor groove, which lacks a spine of hydration and a very deep major groove (Table 1. The sugar conformation changes from C2'-endo as found in B-DNA to C3'-endo (Figure 1.3).

This conformational change distances the phosphate groups in a way to avoid autocatalysis. Other DNA forms have been detected since such as the left handed Z-DNA (Figure 1.2). The form under which DNA is present depends mainly on environmental effects such as the degree of hydration, salt content and the exact composition of the solution. With the advance of experimental techniques, such as X-ray crystallography, NMR spectroscopy, gel-electrophoresis and electron microscopy, de-
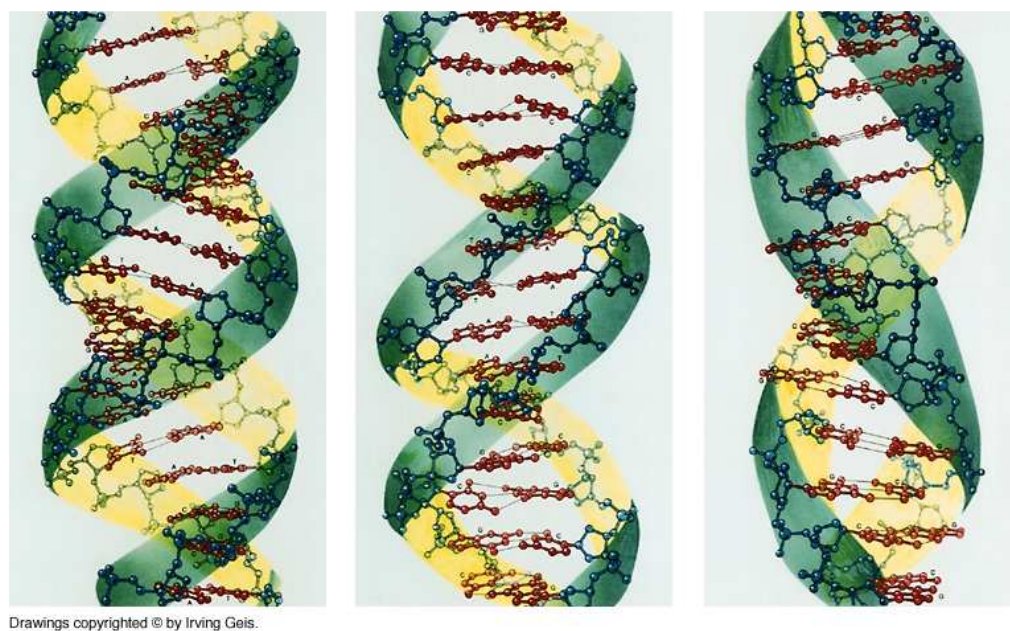
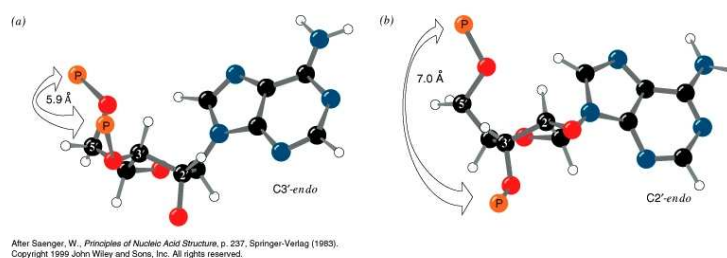Figure 1.2: Schematic drawing of A-DNA (left), B-DNA (middle) and Z-DNA (right).

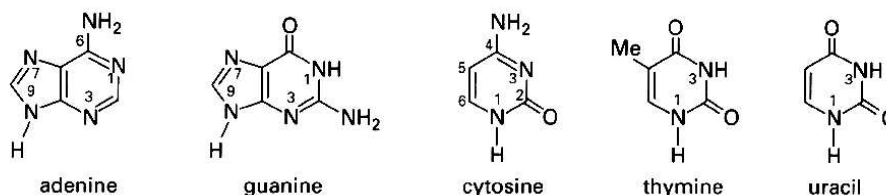Figure 1.3: Sugar pucker as found in A-DNA (left) and in B-DNA (right).

Figure 1.4: Numbering in DNA and RNA bases according to IUPAC.

viations from the ideal A and B-DNA structures were detected. There was growing evidence that these deviations were at least partially sequence dependent, indicating a possible way of sequence specific DNA recognition by proteins and drugs [23, 24]. It also became clear that a universal set of structural parameters was mandatory in order to describe precisely and unequivocally the structural features of a given DNA molecule. In the early 70's an effort has been made to elaborate a nomenclature of general validity for nucleic acids, which is now recommended by IUPAC, the Joint Commission on Biochemical Nomenclature [25, 26, 27, 28] (see also the following helpful links [29, 30]) and in the following I will give the definitions, which are most relevant to the discussion that follows.

The nucleobases can be divided into the purines adenine (A) and guanine (G) and the pyrimidines cytosine (C) and thymine (T) or uracyl in RNA (U). The numbering of the atoms in the nucleobases is given in Figure 1.4.

The atoms in the phosphate-ribose backbone are distinguished from the ones in the bases by a prime. The sugar ring is attached to the N9 position in purines and to the N1 position in pyrimidines via a glycosylic C-N bond. The torsional angle around the glycosylic bond is denoted χ and defines the orientation of the nucleobase with respect to the sugar ring, which can be eiter *syn* or *anti*. In standard B-DNA the bases are usually in the anti-conformation, which is sterically more favorable.

The sugar ring is not planar but can adopt different conformations, in which one or two atoms are displaced from the plane described by the other atoms. If the "out-of-plane" atom lies on the same side as C5' it is called *endo*, if it lies on the other side it is called *exo*. The most common conformations are C2'-endo, present in canonical B-DNA, and C3'-endo, which is usually met in A-DNA. A more complete picture is given by the *pseudorotation cycle* as described for example in [31]. The obtained angles can be linked to the more approximate endo-exo definition by means of the sugar wheel (Figure 1.5).
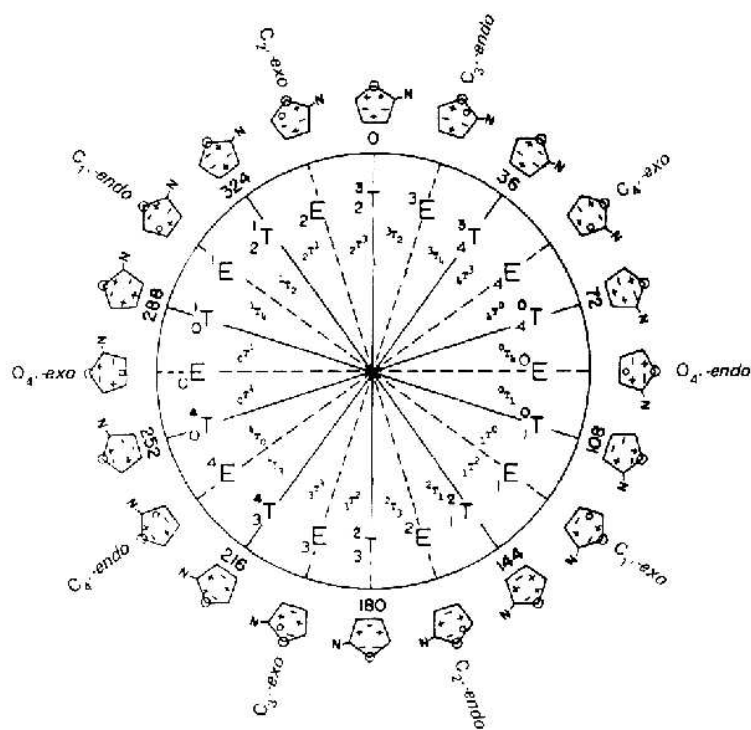
Figure 1.5: Pseudorotation cycle of the furanose ring in nucleosides. Values of the phase angles are given in multiplets of $36°$.[31]
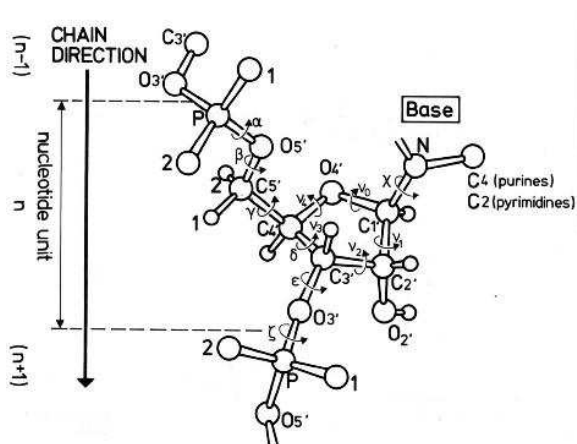
Figure 1.6: Atomic numbering scheme and definitions of torsional angles in phosphate-sugar moiety in RNA (valid also for DNA, besides that O2' is substituted by H2').[31]

Further torsional angles along the phosphoester bonds are called α, β, γ, δ, ϵ and ζ shown in Figure 6.

The orientation of the bases and base-pairs with respect to each other is described by the helical parameters (Figure 1.7). These parameters can be divided into groups according to the geometrical features they describe:

- Parameters describing the relative orientation of two bases in the same base-pair: Shear, Stretch, Stagger, Propeller twist, Buckle and Opening.

- Parameters, which consider the relative orientation of two adjacent base-pairs: Roll and Tilt angle, Shift, Slide and Rise as well as helical Twist angle.

- Parameters which describe the orientation of a base-pair to a primarily defined global helix-axis: X-and Y displacement, Inclination and Tip

The measurement of these structural parameters require the definition of an axis system for each base and base-pair as well as the determination of a global helix axis. This task is solved by several analysis programs but usually the determination of the local and global reference frames is not unequivocal, leading to slightly different results concerning the above parameters. In particular, the determination of a global axis requires special attention since large distortions in the double helix can lead to difficulties in the determination of a straight helix axis. In this work, we use mainly the analysis program Curves, written by Richard Lavery and Heinz Sklenar [32, 33].
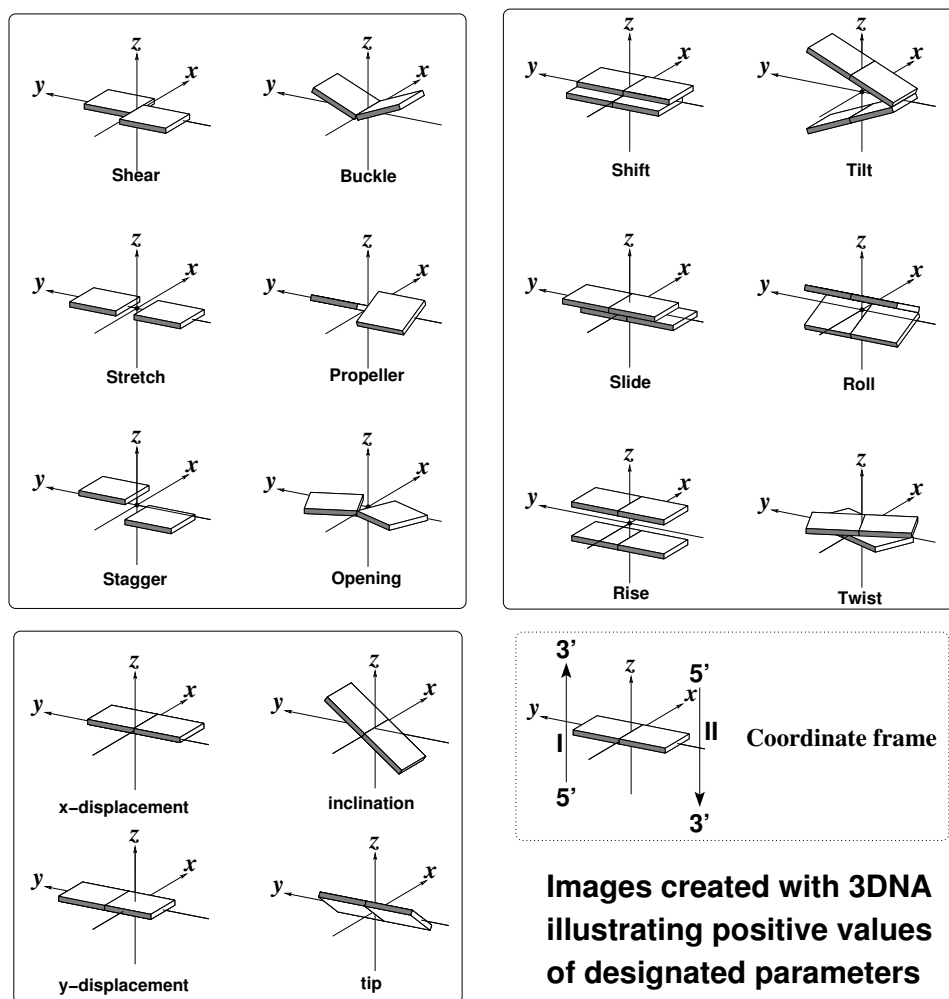
Figure 1.7: Schematic drawing of distortion between bases and base-pairs described by helical parameters.

	Residues per turn	Twist per base-pair [°]	Rise per base-pair [Å]	Sugar pucker
B-DNA	10	36	3.3-3.4	C2'-endo
A-DNA	11	32.7	2.9	C3'-endo
	Minor groove		Major groove	
	width	depth	width	depth
B-DNA	5.7	7.5	11.7	8.8
A-DNA	11.0	2.8	2.7	13.5

Table 1.1: Selected parameters for right handed DNA double helixes

In this program, an optimal curvilinear axis is determined by optimizing a suited geometrical function containing four variables for each base, namely the X-displacement, Y-displacement, Inclination and Tip.

Changes in the structural parameters of DNA are often correlated due to geometrical restrictions of the sugar/phosphate backbone. This becomes for example evident in case of a large helical twist, which stretches the backbone. The stretch is partially relieved by bringing the base-pairs closer together, lowering thus the rise. This correlation has been quantified by Yanagi et al.[34] in the profile sum:

$$Profilesum = (\omega - 36) - 16.24(D_z - 3.36) + 0.744\chi - 0.703\rho \qquad (1.1)$$

In case of large twist and low rise one is in the region of High Twist Profile (HTP) and in case of low twist and large rise one is in the region of Low Twist Profile (LTP). Finally, minor and major groove width and depth are useful parameters to look at. The width can be measured between the phosphorus atoms or C4' atoms, the latter usually yielding better results. Since the phosphate groups being closest together are usually separated by several nucleotides in sequence, the assignment is not always unambiguous. The Curves algorithm solves the problem by projecting the minor groove width on the global helix.

In Table 1 we report some standard parameters of canonical A and B-DNA.[35]

Chapter 2

Methods

2.1 Classical Molecular Dynamics

Until 1995, Classical Molecular Dynamics (MD) simulation of nucleic acids were plagued by instabilities due to approximation methods in the treatment of long range electrostatics and limited accuracy in the force-field [36]. Thanks to the effort made by several research groups, stable multi-ns simulations of oligonucleotides are now possible [3, 37, 38, 39]. It has long been known that a correct treatment of long range electrostatic interactions is crucial in highly charged polyelectrolytes and approximate truncation methods lead inevitably to severe defects in the simulation, such as local heating, distortion of the double helix and complete rupture of Watson-Crick base-pairs [40]. Increasing computational power allows now for a more rigorous treatment of electrostatics, using faster algorithms such as the particle mesh Ewald method. Together with more accurate force fields, simulations of nucleic acids are possible without imposing artificial structural constraints and these simulations provide useful insight into the dynamical behavior of nucleic acids.

Here, we describe shortly the force-field used here, the integration method of the equations of motion (EOM) and the Ewald summation method.

2.1.1 Force Field

The classical Force Field describes the structure and covalent connectivity of molecules. For large biological systems, typically a pairwise additive potential function is used, which takes the form:

$$U = \sum_{bonds} k_b(r - r_{eq})^2 + \sum_{angles} k_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \sum_n \frac{V_n}{2} \times [1 + \cos(n\phi - \gamma)]$$

$$\tag{2.1}$$

$$+ \sum_i \sum_{j>i} [(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}) + \frac{q_i q_j}{\epsilon \cdot r_{ij}}] + \Lambda$$

where r_{eq} and θ_{eq} refer to equilibrium bond lenghts and angles, k_b and k_θ are the vibrational constants and V_n is the torsional barrier corresponding to the n_{th} barrier of a given torsional angle with phase γ. The last two terms in equation (1) refer to non-bonded interactions described by a Lennard-Jones potential, containing an attractive term due to dispersion ($1/r^6$) and a repulsive term due to exchange repulsion ($1/r^{12}$) and an electrostatic interaction term, where ϵ is the dielectric constant.

The bond and angle parameters can be quite easily derived from ab initio calculations of small fragments in vacuo. The derivation of the torsional parameters especially in case of the floppy sugar-phosphate backbone is a delicate task. In this work I use the modified Cornell et al. force field (parm98) [41]. Cheatham et al. introduced some subtle modifications in the Cornell et al force field in order to better reproduce the equilibrium between C2'-endo and C3'-endo sugar puckers and the helical repeat for DNA, which is slightly underestimated by the parm94 force-field [42]. The changes are shown in Table 2.1 The changes do not lead to a systematic improvement over the original force field. One drawback for example is the slower conversion from A to B-DNA or vice-versa, which could hamper effective sampling in short MD runs.

2.1.2   Integration of equation of motion

During the molecular dynamics simulations, the system evolves according to the classical equation of motion, which can be expressed by the Lagrangian as follows:

$$\frac{d}{dt}(\partial L/\partial \dot{q}_k) - (\partial L/\partial q_k) = 0 \tag{2.2}$$

where $L = K - V$ with $K = \frac{1}{2}m\dot{q}_k^2$ and V is given by the force field as defined above. q_k are generalized coordinates and $\dot{q}_k$ the corresponding velocities. From equation 2.2 Newtons equation of motion for the nuclei can be derived:

	n	Cornell *et al.* V_n [kcal/mol]	$\gamma(°)$	parm98.dat V_n [kcal/mol]	$\gamma(°)$
CT-OS-CT-N*	3	1.15/3	0.0	0.383	0.0
CT-OS-CT-N*	2	0.0	0.0	1.0	0.0
OS-CT-CT-OS	3	0.144	0.0	0.144	0.0
OS-CT-CT-OS	2	1.0	0.0	1.5	0.0
OH-CT-CT-OS	3	0.144	0.0	0.144	0.0
OH-CT-CT-OS	2	1.0	0.0	1.5	0.0
OH-CT-CT-OH	3	0.144	0.0	0.144	0.0
OH-CT-CT-OH	2	1.0	0.0	1.5	0.0
OS-CT-N*-CK	2	0.5	180.0	0.0	0.0
OS-CT-N*-CK	1	2.5	0.0	2.5	0.0
OS-CT-N*-CM	2	0.5	180.0	0.0	0.0
OS-CT-N*-CM	1	2.5	0.0	2.5	0.0

Table 2.1: Changes in the Cornell et al. force field introduced by Cheatham et al. [42].

$$F_k = -\frac{\partial E}{\partial q_k} = M\ddot{q}_k \qquad (2.3)$$

Computing the classical trajectory exactly would require to solve a system of 3N second order differential equations, where N is the number of atoms.

In practice, these equations are never solved exactly but rather approximated by a suitable algorithm. To this aim, the positions, of each coordinate are expressed by a Taylor expansion in $r(t)$. Suitable combination of these expansions yield an expression for $r(t + \delta t)$ and in a similar way, one obtains expressions for $v(t + \delta t)$ and $a(t + \delta t)$.

The most widely used algorithm is the *Verlet algorithm*, which uses the positions at time t and time $t - \delta t$ as well as the acceleration at time t to estimate the positions at time $t + \delta t$:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{F(t)}{m}\Delta t^2 + O(\Delta t^4) \qquad (2.4)$$

In this original algorithm the velocities do not appear explicitly, which inhibits temperature control by velocity scaling. The *leap-frog algorithm*, which is equivalent to the Verlet algorithm, updates both positions and velocities using the forces F(t) determined by the positions at time t:

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{F(t)}{m}\Delta t$$

$$r(t + \Delta t) = r(t) + v(t + \frac{\Delta t}{2})\Delta t$$

$$(2.5)$$

The precision of the algorithm is up to third order (in the Taylor expansion) and fulfills the condition of time-reversibility. The ensemble generated by an MD simulation is NVE. By coupling the system to a thermostat [43, 44, 45], one obtains the NVT ensemble, and by introducing pressure coupling, one obtains the NPT ensemble. Even though the Berendsen weak coupling algorithm is extremely efficient for relaxing the system to the target temperature, it does not allow for sampling of the canonical ensemble (even though the differences might be negligible). Canonical ensemble simulation is enabled by employing a Nosé-Hoover thermostat [44, 45].

2.1.3 Thermostats

Constant temperature can be obtained by coupling the system to a thermostat. Several methods are usually available in typical MD packages:

- Weak coupling method (Berendsen thermostat) [43].

- Coupling to an external bath (Nosé Hoover thermostat) [44, 45].

In the weak coupling method, the temperature of the system is kept close to the target equation by the equation:

$$\frac{dT}{dt} = \frac{1}{\tau_T}[T_0 - T(t)] \tag{2.6}$$

where T is the instantaneous value of the temperature and τ_T is the temperature coupling time. The instantaneous temperature T of a system with N_{df} degrees of freedom is related to the kinetic energy $E_{kin}(t)$:

$$E_{kin}(t) = \sum_{k=1}^{N} m_k \dot{q}_k^2(t) = \frac{1}{2} N_{df} k_b T(t) \tag{2.7}$$

The atomic velocities can then be scaled by a factor λ as follows:

$$\delta E_{kin}(t) = [\lambda(t)^2 - 1]\frac{1}{2} N_{df} k_b T(t) \tag{2.8}$$

The factor $\lambda(t)$ is used to scale the velocities $\dot{q}_k$ at each integration step, in order to relax the temperature towards the target value T_0. The relaxation rate is controlled by the time coupling constant τ_T.

In the Nosé-Hoover approach, the ensemble is extended by introducing a thermal reservoir and a friction term in the equations of motion. The friction force is proportional to the product of each particle's velocity and a friction parameter ξ. This friction parameter is a fully dynamic quantity with its own equation of motion. The time derivative is calculated from the difference between the current kinetic energy and the reference temperature. Thus, the equation of motion for each particle reads now:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - \xi \frac{dr_i}{dt} \tag{2.9}$$

and the equation of motion for the bath parameter ξ becomes:

$$\frac{d\xi}{dt} = \frac{1}{Q}(T - T_0) \tag{2.10}$$

where T_0 is the reference temperature and Q determines the strength of coupling to the heat bath.

2.1.4   Ewald summation

Ewald summation is a technique for computing the interaction of a charge and all its periodic images [46, 47, 48]. The charge distribution ρ(r) in the system is an infinite set of point charges, mathematically represented as delta functions:

$$\rho = q_i \delta(r - r_i) \tag{2.11}$$

Each point charge q_i is now surrounded by an isotropic Gaussian charge distribution of equal magnitude and opposite sign.

$$\rho_i^G(r) = -q_i (\alpha/\sqrt{\pi})^3 e^{-\alpha 2 |r - r_i|^2} \tag{2.12}$$

This smeared charge screens the interaction between the point charges, so that the interaction calculated using the screened charge distribution becomes short-ranged due to the appearance of the complementary error function. The smeared electrostatic interaction energy reads now:

$$E_s = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{n=0}^{\infty} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{|r_{ij} + n|} efrc(\alpha |r_{ij} + n|) \tag{2.13}$$

$$erfc(x) = 2\pi^{1/2} \int_x^\infty e^{-y^2} dy \tag{2.14}$$

E^s can now be well approximated using a finite summation. A purely Gaussian charge distribution E^G must be added in order to recover the original charge distribution. The interaction of these Gaussian distributions can be calculated as a lattice sum in reciprocal space as follows:

$$E^G = \frac{1}{4\pi\epsilon_0} 2\pi L^{-3} \sum_{k \neq 0}^{\infty} k^{-2} e^{-k^2/(4\alpha^2)} \sum_{j=1}^{N} q_j e^{-k \cdot r_j^2}$$

$$- \frac{1}{4\pi\epsilon_0} \pi^{-1/2} \alpha \sum_{j=1}^{N} q_j^2 \tag{2.15}$$

The parameter α can be tuned in order to optimize convergence properties of both summations. In particular, in the particle mesh Ewald method, α is chosen sufficiently large so that atom pairs for which r_{ij} exceeds a specified cutoff are negligible in the direct space sum. The reciprocal space sum is then approximated by a multidimensional piecewise interpolation approach, in which energy and forces can be expressed as convolutions and can thus be evaluated by the fast Fourier transforms (FFT).

2.2 Quantum mechanics

Classical mechanics do not allow for studying chemical reactions, since electrons are not taken explicitly into account. Inclusion of electrons require however a quantum mechanical description. In quantum chemistry each particle is described by a wave function $\Psi(r, t)$, where $\Psi^* \Psi$ represents the probability of finding the particle at position r at time t. To each classical observable, such as position and momentum, there exists a corresponding quantum mechanical operator, which when acting on the wave function, yields the expectation value of this operator. The most important operator, the Hamiltonian, was introduced by Schrödinger and enables to calculate the energy of a given system, constituted of M nuclei and N electrons. The Schrödinger equation in its time-independent form reads:

$$H\Psi = E\Psi \tag{2.16}$$

where the Hamilton operator can be divided into several terms as:

$$H = T + V_{ext} + V_{ee} \tag{2.17}$$

T is the kinetic energy operator, V_{ext} is the external potential produced by the nuclei acting on the electrons and V_{ee} is the electron-electron potential. The precise form of the Hamilton is obtained by substituting the classical form of observables by their quantum mechanical operator:

$$
\begin{aligned}
x &= \hat{x} \\
p_x &= \frac{\hbar}{i}\frac{\partial}{\partial x}
\end{aligned}
\tag{2.18}
$$

Other variables, such as spin have no classical counterpart but can also be introduced in the Hamiltonian. Since in this work we are only interested in closed shell system, we do not need to consider spin variables.

2.2.1 The variational principle

In practice, the wave function of an interacting many-body system is not know a priori. To solve the Schrödinger equation one starts from a trial wave function Ψ_T, which is usually a linear combination of basis functions:

$$\Psi_T = \sum_n c_n \psi_n \tag{2.19}$$

where ψ_n are orthonormal and normalized wave functions, such that $\sum_n |c_n|^2 = 1$. An approximation to the true ground-state wave function is found by applying the variational principle, which states:

"The expectation value of a Hamiltonian, $\hat{H}$, calculated with a trial wave function Ψ_T is never lower in value than the true ground state energy, ϵ_0 , value calculated with the true ground-state wave function Ψ_0.

By minimizing ϵ_0 with respect to ψ_n, which means basically optimizing the linear combination coefficients, one obtains the true ground-state wave function with the corresponding ground-state energy.

2.2.2 Born-Oppenheimer Approximation

The Born-Oppenheimer Approximation is the assumption that the electronic and nuclear motion can be separated. The many-body wave function depends both on electronic and nuclear degrees of freedom, but since the electrons are by orders of magnitude lighter than the nuclei, they move much faster and relax very rapidly to the instantaneous ground-state configuration given by the nuclear positions. The nuclei can therefore be considered as stationary points and their coordinates enter as parameters in the wave function for the electrons:

$$\tilde{\Psi}(r_i, R_\alpha) = \Psi(r_i; R_\alpha)\phi(R_\alpha) \tag{2.20}$$

where r_i are the electronic variables and R_α are the nuclear positions. The electronic wave function can now be reintroduced in the Schrödinger equation for the electrons for a given nuclear configuration:

$$[-\frac{1}{2}\sum_i \nabla_i^2 - \sum_i \sum_\alpha \frac{Z_\alpha}{|r_i - R_\alpha|} + \frac{1}{2}\sum_i \sum_{j\neq i} \frac{1}{|r_i - r_j|}]\Psi(r_i; R_\alpha) = \epsilon_e(R_\alpha)\Psi(r_i; R_\alpha)$$
$$\tag{2.21}$$

where $\epsilon_e(R_\alpha)$ is called the adiabatic contribution of the electrons to the energy of the system. This is usually a very good approximation, since the nuclear and electronic frequencies are well separated.

2.2.3 Density Functional Theory

Density Functional Theory (DFT) allows to calculate ground-state properties, and in its extension also excited state properties, from the electron density of an N-electron

system. As a result of the Born-Oppenheimer approximation, the Coulomb potential arising from the nuclei is treated as a static external potential V_{ext}:

$$V_{ext} = -\sum_{\alpha} \frac{Z_{\alpha}}{|r - R_{\alpha}|} \qquad (2.22)$$

The remainder of the electronic Hamiltonian includes the kinetic part and the electron-electron interaction:

$$\hat{F} = -\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_i \sum_{j \neq i} \frac{1}{|r_i - r_j|} \qquad (2.23)$$

The ground-state $\Psi(r_i; R_{\alpha})$ for this Hamiltonian gives rise to a ground-state electronic density $\rho_0(r_i)$:

$$\rho_0 = \int \Psi_0^* \Psi_0 dr \qquad (2.24)$$

Both the ground-state wave function and density are functionals of the number of electrons N and the external potential $V_{ext}(R)$. Density Functional Theory is based on two theorems proposed by *Hohenberg, Kohn* and *Sham* in the early 60's [49, 50].

- The external potential V_{ext} is uniquely determined by the corresponding ground-state electronic density, to within an additive constant. The opposite is true as well: the external potential V_{ext} determines exactly the electron density $\rho(\vec{r})$. Furthermore, the ground state expectation value of any observable is a unique functional of the ground state density.

- The variational principle is also valid for the electron density. The total energy is minimal for the ground-state density $\rho_0(\vec{r})$ of the system.

The Schrödinger equation can now be reformulated by using the density ρ:

$$E[\rho] = \hat{F}[\rho] + \int dr V_{ext}(r) \rho(r) \qquad (2.25)$$

$\hat{F}$ is a universal functional of ρ and the ground-state density is found by minimizing the functional $E[\rho]$ with respect to ρ. DFT would in principle be exact if the functional would be known exactly. This is however not the case. Kohn and Sham separated $\hat{F}$ into three distinct parts as follows:

$$E[\rho(r)] = T_s\rho(r) + \frac{1}{2}\int\int\frac{\rho r \rho r'}{|r_i - r_{i'}|}dr\,dr' + E_{xc}[\rho r] + \int\rho(r)V_{ext}(r)dr \quad (2.26)$$

The kinetic energy term and the exchange-correlation functional are not solvable in this form. The former can however be calculated if the electron density $\rho(r)$ is build up from wavefunctions $\Psi_i(\vec{r})$ which are then used to calculate the kinetic energy T_s

$$T_s[\rho(r)] = \frac{1}{2}\sum_1^N\int\Psi_i^*(r)\nabla^2\Psi_i dr \quad (2.27)$$

Kohn and Sham proposed a coupled set of differential equations, the Kohn-Sham equations:

$$[\nabla^2 + V_H + V_{xc} + V_{ext}]\Psi_i(\vec{r}) = \epsilon_i\Psi_i(\vec{r}) \quad (2.28)$$

where V_H is the Hartree potential, V_{xc} the exchange-correlation term, V_{ext} the external potential and $\Psi_i(\vec{r})$ are the Kohn-Sham orbitals. These equations are then solved iteratively until self consistency is attained.

The exchange correlation potential V_{xc} follows from equation 2.29:

$$V_{xc} = \frac{\delta E_{xc}[\rho(r)]}{\delta\rho(r)} \quad (2.29)$$

This potential, since not known exactly, has to be approximated. The most common functionals are:

- Local Density Approximation (LDA)[51]: the exchange-correlation energy at point $\vec{r}$ is approximated by the exact exchage-correlation energy of a homogeneous electron gas with density $\rho[r]$:

$$E_{xc}^{LDA}[\rho(r)] = \int \epsilon(r)[\rho(r)]\rho(r)dr \tag{2.30}$$

The LDA method is often surprisingly accurate and gives generally very good results for systems with slowly varying charge densities.

- General Gradient Corrected methods (GGA): take also the first gradient of the density into account:

$$E_{xc}^{GGA}[\rho(r)] = \int \epsilon(r)[\rho(r)]\rho(r)dr + \int F_{xc}[\rho(r), |\nabla\rho(r)|]dr \tag{2.31}$$

where F_{xc} is a suitable functional, which satisfies certain known limits of E_{xc}.

One of the most popular gradient corrected functionals is given by the combination of the Becke exchange functional and the Lee, Yang and Parr correlation functional (BLYP) [52, 53]. The Becke exchange functional is given by:

$$
\begin{aligned}
E_x[\rho] &= E_x^{LDA}[\rho] - \beta \int \rho^{4/3}(r)\frac{x^2}{1 + 6\beta sinh^{-1}x}dr \\
&with \\
x &= \frac{|\nabla\rho(r)|}{\rho^{4/3}(r)}
\end{aligned}
\tag{2.32}
$$

The parameter β is determined by a fit on the exact HF data and was fixed by Becke to be 0.0042au.
The Lee, Yang and Parr correlation function was derived from the Colle-Salvetti formula to calculate the correlation energies from HF second order density matrix:

$$E_c[\rho] = -a \int \frac{1}{1+d\rho^{1/3}}(\rho + b\rho^{-2/3}[C_F\rho^{5/3} - 2t_W + (\tfrac{1}{9}t_W + \tfrac{1}{18}\nabla^2\rho)]e^{-c\rho^{1/3}})dr$$

$where$ \hfill (2.33)

$$C_F = \tfrac{3}{10}(3\pi^2)^{2/3}, t_W(r) = \tfrac{1}{8}\frac{|\nabla\rho(r)|^2}{\rho(r)} - \tfrac{1}{8}\nabla^2\rho(r),$$

$$a = 0.04918, b = 0.132$$

One drawback of the BLYP functional is its poor description of long-and medium range exchange-correlation effects. A further improvement is achieved with hybrid functionals, such as B3LYP where some of the exact HF exchange is added:

$$E_{xc} = a_0 E_x^{HF} + (1 - a_0) E_x^{LDA} + a_x \Delta E_x^B + (1 - a_c) E_c^{VWN} + a_c \Delta E_C^{LYP} \quad (2.34)$$

where the parameters a_0, a_x and a_c are fitted to thermodynamic and spectroscopic properties of various systems and take the typical values of 0.20, 0.72 and 0.81 respectively.

In this work we use mainly the BLYP functional, which has shown to give in general very good results for biological systems [54, 55, 56].

2.2.4 Car-Parrinello Molecular dynamics

In the 80's *Car* and *Parrinello* developed a method, which combines molecular dynamics and density functional theory [57]. The electronic and nuclear degrees of freedom evolve simultaneously according to a modified set of classical equations of motion, in which a fictitious electron mass is assigned to the electronic Kohn-Sham (KS) orbitals Ψ_i. Forces are evaluated both on the nuclei and the electrons applying the Kohn-Sham equations. The Lagrangian of the combined system reads then:

$$L^{CP} = \frac{1}{2} \sum_i^N \mu_i |\dot{\Psi}_i(r)|^2 dr + \frac{1}{2} \sum_\alpha M_\alpha \dot{R}_\alpha^2 - E_{KS}[(\Psi_i), (R_\alpha)] + \sum_{ij} \lambda_{ij}(\Psi_i^*(r)\Psi_j(r)) - \delta_{ij}$$

$$(2.35)$$

where μ is the fictitious electron mass, M_α and R_α refer to the nuclear mass and coordinates and the Lagrangian multipliers δ_{ij} ensure the orthonormality between the KS orbitals. The forces are given by the Euler equation:

$$\mu \ddot{\Psi}_i(r) = -H_{KS}\Psi_i(r) + \sum_j \lambda_{ij}\Psi_j(r)$$

$$(2.36)$$

$$M_\alpha \ddot{R}_\alpha = -\frac{\delta E[(\psi_i), (R_\alpha)]}{\delta R_\alpha}$$

The effect of the nuclei on the electrons is included in the Kohn-Sham Hamiltonian H_{KS}, whereas the nuclear motion depend also on the electron degree of freedom through $E[(\psi_i),(R_\alpha)]$. The Kohn-Sham energy is conserved during the dynamics as long as adiabaticity is ensured. This is the case when the two frequency spectra are well separated. The lowest electronic frequency is estimated by the electron gap E_{gap} between the highest occupied and lowest unoccupied Kohn-Sham orbital.

$$\omega = (2\frac{E_{GAP}}{\mu})^{1/2} \tag{2.37}$$

and thus adiabaticity can be controlled by chosing an appropriate fictitious electron mass. Furthermore, chosing a small integration time step for the equations of motion assures that the electronic wavefunction stays close to the Born-Oppenheimer surface. The integration of the equation of motion is done by the Verlet algorithm.

2.2.5 Plane waves & Pseudopotentials

A particular convenient implementation to solve the Schrödinger equation in periodic system is to use planewaves:

$$\Psi_{nk}(r) = exp(ik \cdot r)u_{nk}(r) \tag{2.38}$$

u_{nk} is a strictly cell periodic function, such that $u_{nk}(r+R) = u_{nk}(r)$ and k denotes a good quantum number, related to translational symmetry. A periodic function can however always be expressed as a Fourier series:

$$u(r) = \sum_G \tilde{u}_G \exp(iG \cdot r) \tag{2.39}$$

Equation [2.39] can be reintroduced in [2.38], which then reads:

$$\Psi_{nk}(r) = \sum_G c_{nk}(G) \exp(i(k+G) \cdot r) \tag{2.40}$$

Applying Bloch's theorem one can show that one needs only to sample k-vectors lying within the first *Brioullin zone*. For each k-vector in the first Brioullin zone, one must calculate the occupied eigenstates of the Hamiltonian in order to construct the density. However, the wave-functions and other properties such as Hamiltonian eigenvalues vary smoothly over the Brillouin zone [58] so that in practice only a finite set of points need to be chosen, and methods for making efficient choices have been developed [59, 60, 61].

The volume of the Brioullin-zone is related to the volume of the supercell as follows:

$$\Omega_{BZ} = \frac{(2\pi)^3}{\Omega_{cell}} \qquad (2.41)$$

In very large systems, the Brillouin zone volume is very small and only a few **k**-points need to be considered to describe the variation across the Brillouin zone accurately. In this work we therefore only calculate the wave-functions at the centre of the Brillouin zone, $\mathbf{k} = 0$, known as the Γ-point. This has the added advantage that at this **k**-point the wave-functions can be chosen to be real without loss of generality.

A further approximation is based on the observation that core electrons are rather unaffected by changes in their chemical environment. We therefore consider only the valence electrons during the dynamics and the effect of the core electrons is considered to be constant and will be included in a *pseudopotential*. One can construct now much smoother wavefunctions, ψ_{PS}, since the rapid oscillations close to the nuclei, which assure orthogonality between core and valence states are no longer needed. This has the advantage that the number of planewaves in the expansion of the wave-function can be greatly reduced. The new pseudo-valence states are constructed from an all-electron calculation:

$$[\hat{H} + \hat{V}_{PS}]|\psi_{PS}\rangle = E|\psi_{PS}\rangle$$

$$\hat{V}_{PS} = \sum_n^{core} (E - En)|\chi_n\rangle\langle\chi_n| \qquad (2.42)$$

where χ are the core-electron states and E_n are the energy eigenvalues of these states. The pseuodpotentials should be transferable, which means that they must reproduce the effect of the core electrons in different molecular environments. Further conditions have to be fullfilled:

- The lowest pseudo wavefunction generated by the pseudopotential should not contain any nodes.

- The normalised atomic radial pseudowavefunction with an angular momentum l should be equal to the normalised radial all electron wavefunction outside a given cut-off radius r_{cl}:

$$R_i^{PP}(r) = R_i^{AE} \qquad for \quad r > r_{cl} \qquad (2.43)$$

- Norm conservation: The charge inside of r_{cl} has to be the same for both wavefunctions:

$$\int_0^{r_{cl}} |R_l^{PP}(R)|^2 r^2 dr = \int_0^{r_{cl}} |R_l^{AE}(r)|^2 r^2 dr \qquad (2.44)$$

- The eigenvalues of both wavefunctions should be the same:

$$\epsilon_l^{PP} = \epsilon_l^{AE} \qquad (2.45)$$

2.2.6   Wannier Centers and Boys orbitals

Chemical concepts are often visually depicted in terms of localized orbitals. Boy's orbitals are a useful tool to characterize chemical bonds as well as lone pairs. In a plane waves approach, a generalization of Boys orbitals is introduced to map the electronic density into highly localized molecular orbitals. The maximally localized Wannier functions are derived by an unitary transformation from the KS wavefunctions. The positions of the WC centers allow an accurate description of polarization effects in the simulated system.

According to Marzari and Vanderbilt [62], the Wannier functions are defined as:

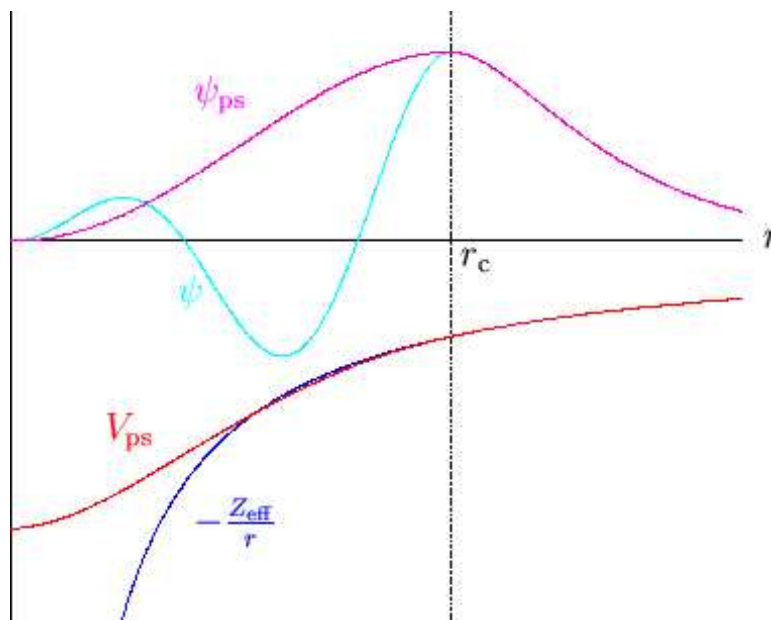$$w_n(r) = u_{nk_0}(r) = \sum_g c_i^{k_0}(g) e^{-ig \cdot r} \qquad (2.46)$$

Figure 2.1: Schematic diagram of the relationship between all-electron and pseudo- potentials and wave-functions.

And they are obtained by a unitary transformation:

$$u_{nk} = \sum_k U_{mn}^k u_{mk} \tag{2.47}$$

under the constraint to minimize the spread function Ω:

$$\Omega = \sum_n (\langle r^2 \rangle_n - \langle r \rangle_n^2) \tag{2.48}$$

The Wannier Centers correspond to the location of maximal charge density of the Boys orbitals and can be used to vizualize the polarization of chemical bonds and lone pairs.

2.2.7 Electrostatic decoupling in clusters

The simulation of isolated clusters within a periodic boundary condition scheme needs some care, as self-interaction among replicas has to be cancelled. In our calculations, we have used the procedure developed by Martyna and Tuckerman that allows treating the system as isolated. This algorithm is based on the general possibility of writing a potential Φ as the sum of two arbitrary functions:

$$\Phi(\mathbf{r}) = \phi^{long}(\mathbf{r}) + \phi^{short}(\mathbf{r})$$

These functions are defined so that $\phi^{short}(\mathbf{r})$ vanishes exponentially quickly at large distances from the system, while $\phi^{long}(\mathbf{r})$ contains all the long-range components of the physical potential. It can be demonstrated that the average potential energy for a cluster can be written in the reciprocal space as:

$$\langle \Phi \rangle = \frac{1}{2V} \sum_{\mathbf{g}} |\bar{\rho}(\mathbf{g})|^2 \left[\tilde{\Phi}(-\mathbf{g}) + \hat{\phi}^{screen}(-\mathbf{g}) \right] \tag{2.49}$$

The screen function $\hat{\phi}^{screen}$, defined as:

$$\hat{\phi}^{screen}(-\mathbf{g}) = \bar{\phi}^{long}(\mathbf{g}) - \tilde{\phi}^{long}(\mathbf{g}) \tag{2.50}$$

is the difference between the Fourier series $\bar{\phi}^{long}(\mathbf{g})$ and the Fourier transform $\tilde{\phi}^{long}(\mathbf{g})$ of the long-range potential, and its computation in the reciprocal space is efficient for all $\mathbf{g}$ vectors ($\mathcal{O}(N \log N)$). . The screen function and has the meaning of "screening" the interaction of the system with an infinite array of periodic images.

2.3 QM/MM

Hybrid QM/MM techniques have become very fashionable in the last years [8, 10, 63]. In these schemes the biological system of interest is divided into two parts. The region of biological interest, typically the active site of a protein or the binding site in drug-DNA complexes, is treated at the ab-initio level (QM-part) while the remaining protein or DNA residues as well as the solvent and counter ions are treated classically (MM-part). A wide variety of schemes have been developed based on Hartree-Fock (HF), Density functional theory (DFT) or also on the empirical valence bond method (EVB) [64]. Here we use a method based on DFT and Car-Parrinello molecular dynamics as implemented in the CPMD program [65] interfaced with the classical molecular dynamics program GROMOS [66]. Special attention has to be payed on the way the QM and MM systems are interfaced. The next paragraphs explain how this is done.

2.3.1 Non-bonded interactions

The interactions can be divided into bonded and non-bonded interactions. The non-bonded mixed Hamiltonian reads:

$$H_{non-bonded} = \sum_{i \in MM} q_i \int dr \frac{\rho(r)}{|r - r_i|} + \sum_{i \in MM, j \in QM} V_{vdw}(r_{ij}) \qquad (2.51)$$

where r_i is the position of the MM atom i (with charge q_i), ρ is the total charge of the QM-system and $V_{vdw}(r_{ij})$ is the van der Waals interaction between the MM-atom i and the QM-atom j.

The evaluation of the electrostatic interaction causes several problems:

- Positive classical charges deprived of the Pauli repulsion term act as electron traps when close to the QM-region. This leads to the electron spill-out problem and this effect is particularly pronounced in highly delocalized wave functions such as planewaves.

- The full evaluation of the electrostatic interaction is computationally very costly and of the order $N_r N_{MM}$ where N_r is the number of real space grid points.

The first problem is solved by applying a modified classical Coulomb potential [13], the charge of which is smeared out and goes to a finite value close to the nuclei:
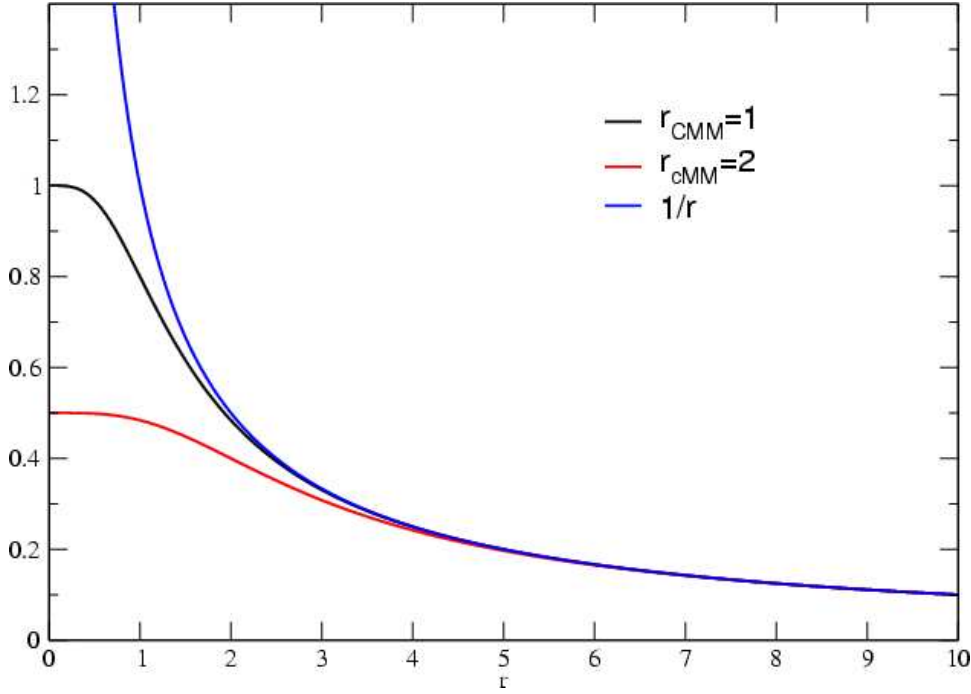
Figure 2.2: Schematic drawing of modified Coulomb-Potential (black and red) and pure Coulomb-Potential (blue).

$$v_{MM}(r) = \frac{r_{cMM}^n - r^n}{r_{cMM}^{(n+1)} - r^{(n+1)}} \tag{2.52}$$

v_{MM} is the modified classical Coulomb-potential and r_{cMM} is the covalent radius of the MM-atom.

The second problem is more intriguing. The computational cost for evaluating the electrostatic interaction energy is minimized, applying a hierarchical scheme in which the classical system is divided into three shells around the QM-system:

- In the first shell, closest to the QM-part, the electrostatic interaction energy is calculated as integral over the QM charge density interacting with the NN classical point charges in the first shell through the modified Coulomb-potential described in [2.52]:

$$H_{el} = \sum_{j \in NN} q_j \int dr \rho v_j(|r - r_j|) \tag{2.53}$$

- In the second shell, the electrostatic interactions are calculated between the classical point charges and QM *D-RESP* point charges, obtained by a fit to reproduce the electrostatic potential on the NN-atoms [14].

- In the outermost shell, the electrostatic interactions are calculated between a multipolar expansion of the QM-charge density and the MM classical point charges contained in this last shell:

$$
\begin{aligned}
H_{mp} &= C \sum_{MM} \frac{q_j}{|r_j - \bar{r}|} + \sum_{\alpha} D^{\alpha} \sum_{j \in MM} q_j \frac{(r_j^{\alpha} - \bar{r}^{\alpha})}{|r_j - \bar{r}|^3} \\
&+ \frac{1}{2} \sum_{\alpha\beta} Q_{\alpha\beta} \sum_{j \in MM} q_j \frac{(r_j^{\alpha} - \bar{r}^{\alpha})(r_j^{\beta} - \bar{r}^{\beta})}{|r_j - \bar{r}|^5}
\end{aligned}
\tag{2.54}
$$

2.3.2 D-RESP charges

In the evaluation of the electrostatic interaction, atomic charges on the QM-atoms are used in the second shell as described above. These charges are obtained by a fit to the electrostatic field at the classical atoms close to the QM-region ($j \in NN$) and restrained to the corresponding Hirshfeld charges with a quadratic penalty function [14]. Thus, the charges will reproduce the electrostatic potential due to the QM-charge density, which is polarized by the MM-part including thus automatically polarization effects. Furthermore, since they are evaluated at every MD-step, fluctuations during the MD can be monitored easily.

In order to evaluate the electrostatic interactions within the first shell, the potential at each MM-atom in this part has to be calculated according to equation 2.53 and thus at each MM-atom the electrostatic potential V_i is given by:

$$
V_i = \int dr \rho v_j(|r - r_j|)
\tag{2.55}
$$

The D-RESP charges are then obtained by minimizing the norm of the following restraining function by a least squares fit:

$$
E = \sum_{j \in NN} \left(\sum_{i \in QM} q_i^D / r_{ij} - V_j \right)^2 + W(q_i^D)
\tag{2.56}
$$

where W is a restraining function, which makes use of the Hirshfeld charges and is of the form:

$$W(q_i^D) = w_q \sum_{j \in QM} (q_j^D - q_j^H)^2 \qquad (2.57)$$

where w_q is an adjustable parameter and the Hirshfeld charges are given by:

$$q_j^H = \int dr \rho^{el}(r) \frac{\rho_j^{at}(|r - r_j|)}{\sum_k \rho_k^{at}(|r - r_k|)} - Z_j \qquad (2.58)$$

where ρ_j^{at} is the atomic (pseudo) valence charge density of the atom j and $Z_j = \int dr \rho_j^{at}(r)$ is its valence. The Hirshfeld charges provide a physically sound restraining set, since they are directly related to the charge distribution of the system and therefore more physical than a restraint to neutral charges.

2.3.3   Bonded Interactions

The bonded interactions consider bonds, angles and dihedrals, which involve atoms belonging both to the QM and to the MM region. Bond lengths, angle and dihedral values are taken into account by the classical force-field such that the MM atoms connected to the QM-part adjust themselves according to the classical force-field during the MD.

The most tricky part is the cut between QM and MM system since the valence shell of the QM atom, which is part of a mixed QM-MM bond is not saturated. There exist several ways to remedy this problem, such as the link-atom approach, frozen orbitals or monovalent pseudopotentials. The QM/MM method used here [13, 14] offers two possibilities:

- For the last QM-atom a particular monovalent pseudopotential is used, which has the same mass and charge as the normal pseudopotential but only one valence electron, being able to form one bond towards the QM-system.

- The last QM-atom is saturated by additional hydrogen atoms. These hydrogen atoms have to be excluded from the classical Hamiltonian, but are however present in the QM-part.

	entire nucleotide	Capping Hydrogens	Monovalent Pseudo
		C1'	
HOMO	-4.4402908	-5.2309521	-3.5288509
LUMO	-1.3777974	-1.4715600	-2.2638069
GAP	3.0624934	3.7593921	1.265044
		C2'/C4'	
HOMO		-5.36069	-3.74698
LUMO		-1.60986	-2.43657
GAP		3.7511	1.31041

Table 2.2: Energies in [eV] for HOMO and LUMO KS-orbitals for Adenosine in water. Different cutting positions are tested and the electron gap has been monitored.

The first approach is conceptually rather simple and does not introduce unphysical atoms in the system. It is however difficult to create a transferable pseuodopotential, which can be used to cut all bond-types. So far, the best results are obtained when cutting an apolar C-C bond.

The additional hydrogen atoms on the other hand can be introduced wherever needed, leaving the electronic structure almost unchanged. The interaction with the classical force-field can be excluded rather straight-forward. The QM-charge density will however be slightly perturbed due to the presence of additional hydrogen atoms.

We perform several tests both with the monovalent pseudopotential and capping hydrogens to see how they influence the electronic structure of nucleic acids. The results show, that it is difficult to use a monovalent pseudopotential since DNA and RNA contain almost exclusively polar bonds. In Table 2.2 we report the electronic gap in the QM/MM system of adenosine in classical water. The adenosine is cut at different positions both with capping hydrogens and with the monovalent pseudopotential. The latter decreases the electron gap considerably and changes also the nature of KS-orbitals (Figure 2.3). In particular, the LUMO is centered on the last QM-atom described by the monovalent pseudopotential, which is clearly an artefact due to the pseudopotential. Thus, for our purposes, additional hydrogen atoms placed at C1' are best suited since only a small portion of the nucleoside has to be included in the system and the electronic structure is well reproduced.
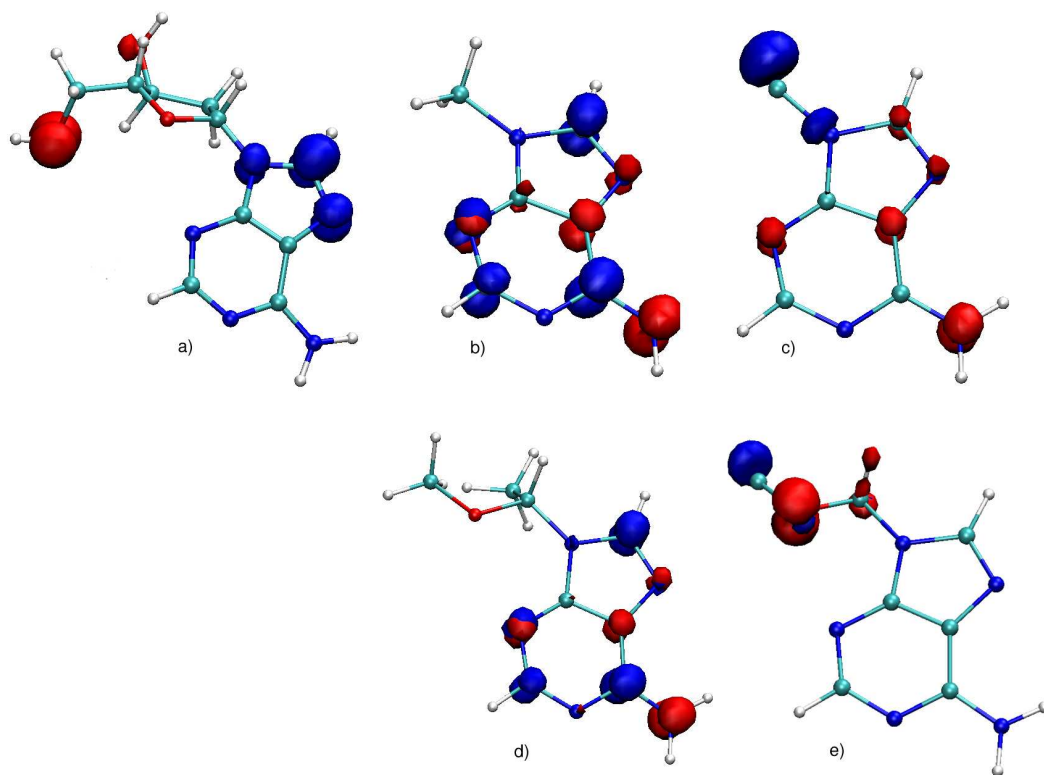
Figure 2.3: Isodensity surfaces of HOMO (red) and LUMO (blue) for different QM/MM systems (only QM-part shown). a) entire nucleoside is included in the QM-part. b) Capping hydrogens in C1' position. c) Cut with monovalent pseudopotential in C1'-position. d) Capping hydrogens in C4' and C2' position. e) cut at C4' and C2'.

2.4 Calculation of chemical shift

Progress in the field of quantum chemistry has made possible caluclations of useful accuracy and reliability on fragments of biomolecules and transition state metal complexes, containing dozens of atoms. Here, a brief explanation of the chemical shift is given, followed by a short outline of the method used to calculate ^{195}Pt chemical shift, i.e. the zeroth order approximation (ZORA) as implemented in the ADF code[67].

2.4.1 Origin of chemical shift

NMR active nuclei have a non-zero spin I, which can take 2I+1 orientations. In an external magnetic field the degeneracy of these orientations is lifted due to the splitting of the energy levels. For I=1/2 the energy splitting is:

$$\Delta E = \frac{ghB}{2\pi} \tag{2.59}$$

where g is the *gyromagnetic ratio*, an intrinsic constant of every nuclei, and h is the Planck constant. B is the magnetic field at the nuclei.

The effective magnetic field at the nucleus is not equal to the applied magnetic field since the latter is shielded by the electrons moving around the nucleus. The resulting effective field is:

$$\vec{B}^{eff} = \vec{B}^{ext}(1 - \vec{\sigma}) \tag{2.60}$$

where $\vec{\sigma}$ is the nuclear shielding tensor, which in the isotropic case simplifies to the isotropic shielding constant Tr$\vec{\sigma}$/3. The shielding tensor depends on the chemical environment of the nucleus. Nuclei of the same kind placed in a different chemical context are shielded differently and can therefore be distinguished. The chemical shift is then defined as:

$$\delta/ppm = 10^6 \frac{v_L - v_{ref}}{v_{ref}} \tag{2.61}$$

where v_L is the *Larmor frequency* of the sample and v_{ref} is the Larmor frequency of a reference compound. The Larmor frequency is related to the shielding constant as follows:

$$v_L = (1 - \sigma)\frac{g \cdot B^{eff}}{2\pi} \tag{2.62}$$

The shielding constant can be calculated from the mixed second-order derivative of the energy with respect to the nuclear magnetic moment and the external field assuming that in lowest order the energy of the system is linear to both $\vec{B}^{ext}$ and μ:

$$\sigma_{st} = \frac{\partial^2 E}{\partial B_s^{ext}\partial \mu_t}|_{\vec{B}ext=\vec{\mu}=0} \tag{2.63}$$

The calculation of the shielding constant requires an accurate description of the wavefunction close to the nuclei. In case of heavy nuclei such as Pt, a relativistic treatment of the electrons is necessary to obtain at least qualitatively reasonable results. In this work we used the zero-order regular approximation as implemented in the ADF code [67, 68, 69] in order to treat scalar relativistic terms.

2.4.2   The Zero-Order Regular Approximation

A relativistic treatment requires that both space and time coordinates are equivalent. This is not the case in the Schrödinger equation (SE), in which spacial coordinates appear as 2nd-order derivative, whereas the time variable as first-order derivative. Dirac replaced the time-dependent SE by a four dimensional equation:

$$[c\boldsymbol{\alpha} \cdot \boldsymbol{p} + \boldsymbol{\beta}mc^2] = i\frac{\partial\psi}{\partial t} \tag{2.64}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are 4×4 matrices. In particular, $\boldsymbol{\alpha}$ can be written in terms of 2×2 Pauli matrices and $\boldsymbol{\beta}$ in terms of a 2×2 unit matrix **I**.

$$\alpha_{x,y,z} = \begin{pmatrix} 0 & \sigma_{x,y,z} \\ \sigma_{x,y,z} & 0 \end{pmatrix} \tag{2.65}$$

$$\beta = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} \tag{2.66}$$

where $\sigma_{x,y,z}$ are the Pauli matrices except for a factor 1/2.

The wavefunction, according to the four-dimensional Dirac equation, contains four components, which can be attributed to two different particles (electrons and positrons) and two different spins (α and β):

$$\Psi = \begin{pmatrix} \Psi_{L\alpha} \\ \Psi_{L\beta} \\ \Psi_{S\alpha} \\ \Psi_{S\beta} \end{pmatrix} \tag{2.67}$$

where Ψ_L and Ψ_S correspond to the large and the small components of the wavefunction, and α and β indicate the usual spin. For electrons, the large component reduces to the solution of the Schrödinger equation where c$\rightarrow \infty$ and the small component disappears. The small component of the electronic wavefunction corresponds to the coupling with the positronic states.

In the presence of an electric magnetic field, the Dirac equation is modified as:

$$[c\boldsymbol{\alpha} \cdot \boldsymbol{\pi} + \boldsymbol{\beta}' mc^2 + V]\Psi = i\frac{\partial \Psi}{\partial t} \tag{2.68}$$

where V=electrostatic potential

and $\boldsymbol{\pi}$ is a generalized momentum operator, including the vector potential A:

$\boldsymbol{\pi} = \boldsymbol{p} + \boldsymbol{A}$.

$\mathbf{B} = \nabla \times \boldsymbol{A}$ and $\boldsymbol{\beta}'$ is the same as above but corrected by the free electron rest energy (mc^2).

In the time independent case, the Dirac equation may be written as:

$$[c\boldsymbol{\alpha} \cdot \boldsymbol{\pi} + \boldsymbol{\beta}' mc^2 + V]\Psi = E\Psi \tag{2.69}$$

which can be factored out (due to the block-matrix nature of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$)

$$c(\boldsymbol{\alpha} \cdot \boldsymbol{\pi})\Psi_S + V\Psi_L = E\Psi_L \qquad (2.70)$$

$$c(\boldsymbol{\alpha} \cdot \boldsymbol{\pi})\Psi_S + (2mc^2 + V)\Psi_S = E\Psi_S \qquad (2.71)$$

Equation 2.71 can be solved for Ψ_S and then reintroduced in equation 2.70 to yield:

$$[\frac{1}{2m}(\boldsymbol{\alpha} \cdot \boldsymbol{\pi})K(\boldsymbol{\alpha} \cdot \boldsymbol{\pi}) + (-E + V)]\Psi_L = 0 \qquad (2.72)$$

where K is defined as:

$$K = (1 + \frac{E - V}{2mc^2}) \qquad (2.73)$$

Equation 2.72 is then usually expanded around $(E - V)/2mc^2$ to give the first-order Pauli Hamiltonian. Such an expansion works fine as long as the velocity of the electrons is everywhere small compared to the speed of light. This condition is not satisfied close to the nucleus. Thus, in the approach described by van Lenthe et al.[70, 71] K in equation 2.72 is replaced by K$'$=$E/(2mc^2 - V)$. This factor is always much smaller than 1 and can now be expanded in powers of $E/(2mc^2 - V)$. Keeping only the zero-order term gives the Zero-Order Regular Approximation (ZORA) Hamiltonian:

$$[\boldsymbol{\pi}\frac{c^2 2m^2 c^2 - V}{\boldsymbol{\pi}} + (\frac{c^2}{2m^2 c^2 - V})^2\frac{Z\boldsymbol{s} \cdot \boldsymbol{L}}{r^3} + V]\Psi_L = E\Psi_L \qquad (2.74)$$

From the ZORA-Hamiltonian the energy and energy-derivatives can be calculated to give the isotropic shielding constant.

Chapter 3

Cisplatin

3.1 Introduction

The antitumoral activity of cisplatin (cis-diamminedichloroplatinum(II)) has been fortuitosly discovered by Rosenberg in 1969. Nowadays, the drug is widely used in clinic treatment against a variety of cancer deseases [72, 73, 74, 75]. Its beneficial effects arise from its binding to DNA nucleobases, preferentially to two N7 atoms belonging to two adjacent guanine nucleobases or, to a lesser extent, to adenine guanine (AG) sequences [76, 77, 72, 78, 79]. Other adducts such as G-G interstrand compounds have also been observed. The binding of cisplatin to DNA is believed to act as a signal giving rise to a complex cellular response, which in susceptible cells, activates programmed cell death, i.e. apoptosis. Different adducts have different biological effects such as mutagenesis [80, 81], inhibition of DNA polymerases [82], and repair of specific DNA adducts [83, 84]. Based on the hypothesis that the biological effects of the platinum-DNA adducts are related to specific modifications of the helical DNA structure, the latter has been extensively studied over the last decades, making use of different methods, such as X-ray crystallography, NMR spectroscopy, and molecular dynamics simulation.

All structural studies have shown that cisplatin binding induces an overall curvature of the DNA double helix towards the major groove [85, 86, 87, 88, 89, 90] associated to an opening of the minor groove. The platinated DNA moieties bind to the High Mobility Group (HMG) domains of proteins [91, 92, 93, 94, 95, 96, 97, 98], impeding replication and cell repair processes [87, 99], leading eventually to cell death [100, 87, 101, 102, 103].

The available structures of cisplatin in complex with B-DNA oligomers in the solid

pdb-entry	Experiment/ Resolution or # structures	Sequence	Specification	Axis bend [°]
1A2E	X-ray/1.63	5'd(CCTCG*CTCTC) 5'-d(GAGAG*CGAGG)	interstrand	-
1A84	NMR/1	5'-d(CCTCTG*G*TCTCC)	intrastrand	84.5
1AIO	X-ray/2.6	5'-d(CCUCTG*G*TCTCC)	intrastrand	39.7
1AU5	NMR/1	5'-d(CCTG*G*TCC)	intrastrand	65.8
1CKT	X-ray/2.5	5'-d(CCUCTCTG*- G*ACCTTCC)	intrastrand HMG A	65.7
1DDP	NMR/10	5'-d(GATAG*CTATG) 5'-d(CATAG*CTATC)	interstrand	-
1I1P	Xrau/1.63	5'-d(CCCTCG*CTCTC) 5'-(GAGAG*CGAGG)	interstrand	-
1KSB	NMR/1	5'-d(CTCCG*G*CCT)	intrastrand	55.6
5BNA	X-ray/2.6	5'-d(CGCG*AATTCG*CG) 5'-d(CGCG*AATTCGCG)	mono- substituted	-

Table 3.1: Available structures of cisplatin-DNA compelxes where G* denotes the platinated site.

state (solved by X-ray diffraction) [91, 86, 88] and in aqueous solution (solved by NMR) [90, 104, 89, 105] provide detailed information on the mode of binding of cisplatin but evidences also large differences between the available structures (Table 3.1).

The metal ion, by binding to DNA causes an axis bend of the double helix ranging from $40°$ to $84°$ [85, 90, 104, 86, 89]. Furthermore, the rise is larger at the platinated site, the propeller twist is increased with respect to B-DNA, and in some cases, some of the Watson-Crick hydrogen bonds are disrupted [89, 90, 88, 106]. The Pt-coordination geometry is slightly distorted from its typical square planar conformation: the metal ion moves out of the plane defined by the purine rings (deviation 0.4 to 1.3 Å), with an N7-Pt-N7 angle as small as $69°$ [107].

The DNA double helix around the platinated site is distorted in a way that renders it more A-DNA like: (i) The minor groove is wider and shallower, (ii) the helix twist is remarkably lower than in standard B-DNA, leading to an unwinding of the double
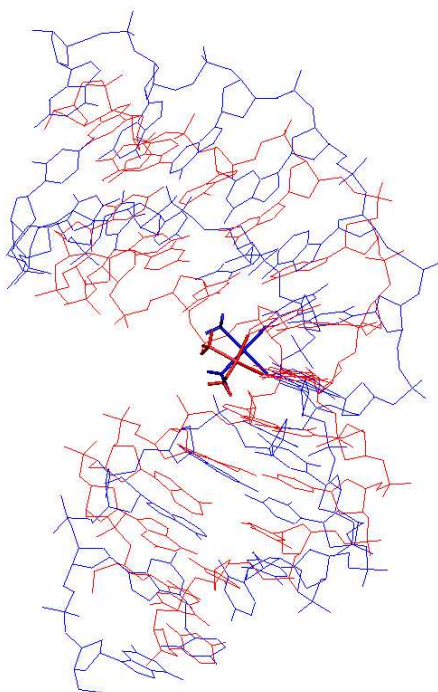
Figure 3.1: Superpostion of NMR (blue) and X-ray structure (red) of the same adduct

helix, and (iii) the sugar moiety in 5'-position of the platinated guanine is in C3'-endo conformation.

A comparison between the X-ray [85] and NMR [90] structures of cispt-d(CCTC-TG*G*TCTCC)-d(GGAGACCAGAGG) (**A**, Figure 3.1) has allowed to pinpoint the structural differences on passing from the solid state to aqueous solution, namely:

- An increase of the axis bend and roll angle.

- A smaller displacement of the Pt atom from the guanine purine rings.

- A different puckering of the 3'-sugar at the platinated lesion with a 3'-endo conformation in the X-ray structure similar to A-DNA and a 2'-endo conformation in solution.

The structural basis of the molecular recognition between the protein domain and the platinated DNA has been provided by Lippard and coworkers, who solved the X-
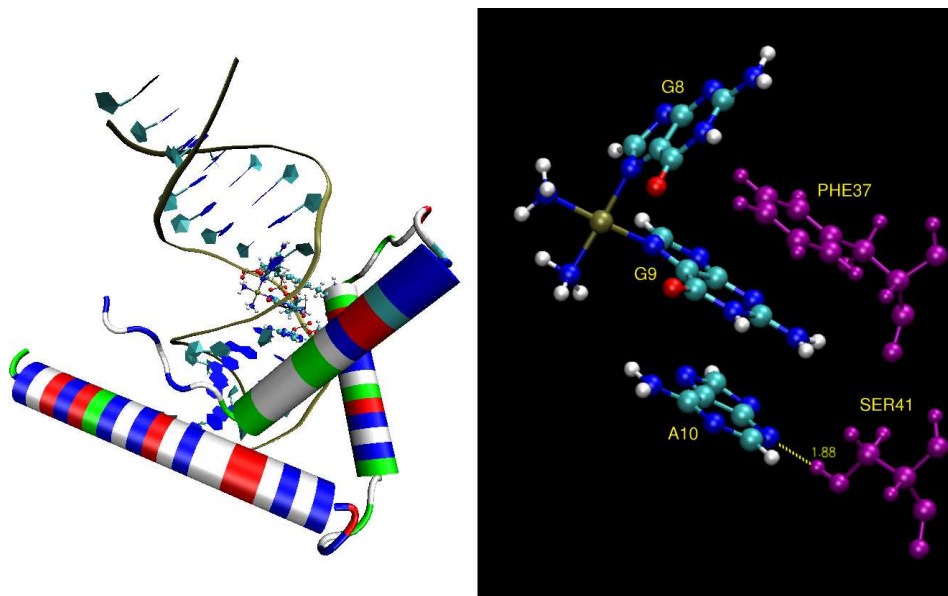
Figure 3.2: left: cisplatin-DNA adduct (16-mer) in complex with HMG A domain of HMG1 protein. right: Detail view of cisplatin-moiety with PHE37 and SER41 interacting with guanine 8 and 9 and adenine 10.

ray structure of a platinated 16-mer, cispt-d(CCUCTCTG*G*ACCTTCC)-d(GGAGA-GACCTGGAAGG) in complex with HMG A (**B**, Figure 3.2) [91].

The HMG domain contains three helixes connected by loop regions of variable lengths. The X-ray structure of the ternary cisplatin-DNA/HMG A complex shows that the hydrophobic side-chains of Helix I and Helix II function as a hydrophobic wedge, which interacts with the minor groove of the DNA (Figure 3.3). Comparison between NMR-structure of the protein alone and the X-ray structure of the protein in complex with the cisplatin-DNA adduct, yields an rmsd of 2.1 Å, indicating only slight structural rearrangement of the protein upon DNA-binding. One of these rearrangments involves a positional shift of Phe37 enabling its side-chain to intercalate between the platinated base-pairs, nameley the G8-C25/G9-C24 base-pair step. Another key structural determinant for molecular recognition is a hydrogen bond between SER41 and the N3 atom of A10, the base directly adjacent to the cisplatin-DNA cross-link (Figure 3.2). Comparison between the X-ray structure of the protein-cispt-DNA complex and the uncomplexed cisplatin-DNA adduct reveal that the characteristic structural distortions are conserved, but augmented in magnitude: (i) the overall

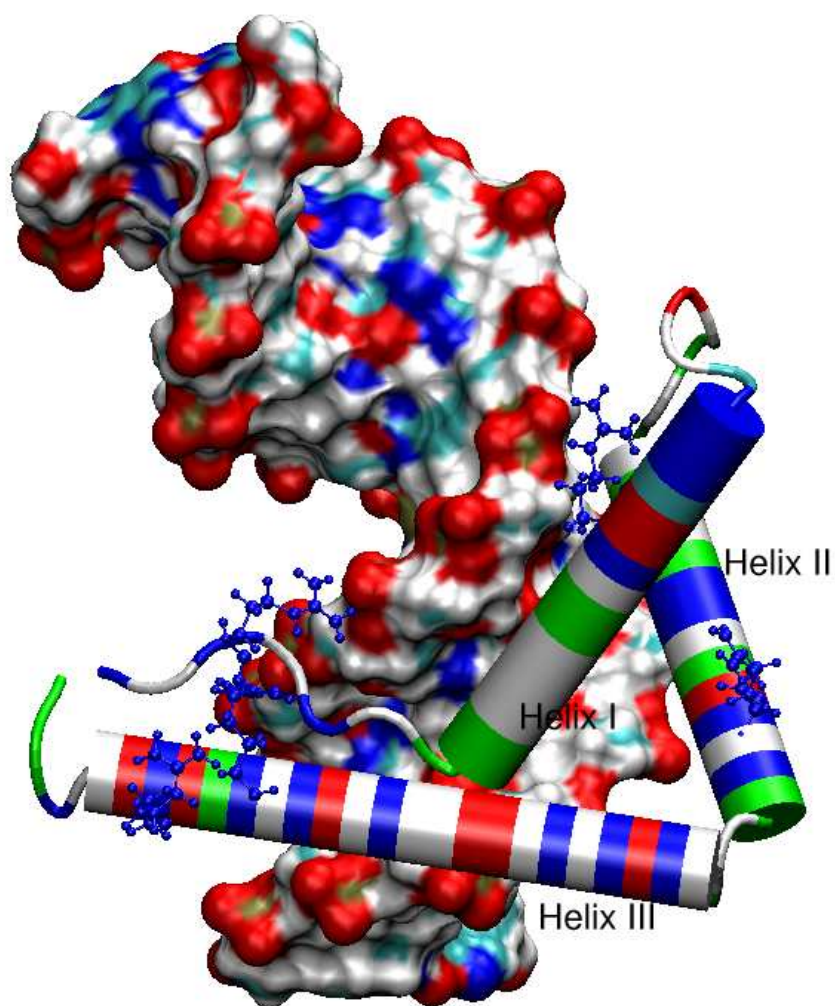Figure 3.3: HMG A protein in complex with cisplatin modified DNA, showing interactions between Helix I and Helix II with DNA minor groove. Arginine residues are displaced in ball and stick and the solvent accessible surface of DNA is shown with electrostatic color-code such as to evidence interactions between sugar-phosphate backbone and the arginine residues of the N-terminal and Helix III.

curvature is as large as 61° with respect to 40° in the free cispt-DNA adduct. (i) The roll angle and rise between the platinated guanines are even larger than in platinated DNA oligomers. In complex with HMG, the hydrophobic DNA-surface is shielded against the solvent by the intercalated Phenylalanine side-chain, which forms $\pi - \pi$ stacking to G9 and *edge to face* interactions with G8 (Figure 3.2). (iii) The minor groove is even wider and shallower than in the free cisplatin-DNA adduct and the helical twist is as low as 9° at the platinated site.

Force field based molecular dynamics simulation have provided complementary information on the dynamics and flexibility of cisplatin and other Pt-based drug-DNA complexes [108, 109, 110]. In particular, they have shown that platinated DNA structures are highly flexible, with large fluctuations in the axis bend and curvature of the DNA double helix. On the other hand, the Watson-Crick H-bond pattern is completely maintained [111, 112, 113, 108]. Due to the very specialized character of cisplatin-DNA parameters, the transferability of the force field models is limited [114, 112]. For instance, the intrinsic dependence of platinum coordination on the electronic structure requires a special set of torsional parameters to be introduced in the force field (Figure 3.4) [111]. As a result, force field parameters are available for very few Pt-based ligands [110, 112, 115].

Here, we present hybrid QM/MM simulations studies on cisplatin/DNA complexes using the approach presented in references [13, 14]. The platinum moiety is treated at the density functional level, with the Car-Parrinello approach [65], which has proven to reliably describe structure, dynamics and electronic properties of platinum-nucleotide complexes in the solid state as well as in water solution [116, 117]. The remaining DNA residues, solvent molecules and Na$^+$ counter ions are treated classically and therefore, the steric and electrostatic effects of the surrounding are taken explicitly into account.

Our calculations are based either on the X-ray structures of platinated DNA alone (**A** [85]), or in complex with the HMG A domain of HMG1 protein (**B** [91]) and on a canonical B-DNA structure, where we docked manually cisplatin on the GG site (**C**). All structures are shown in Figure 3.5.

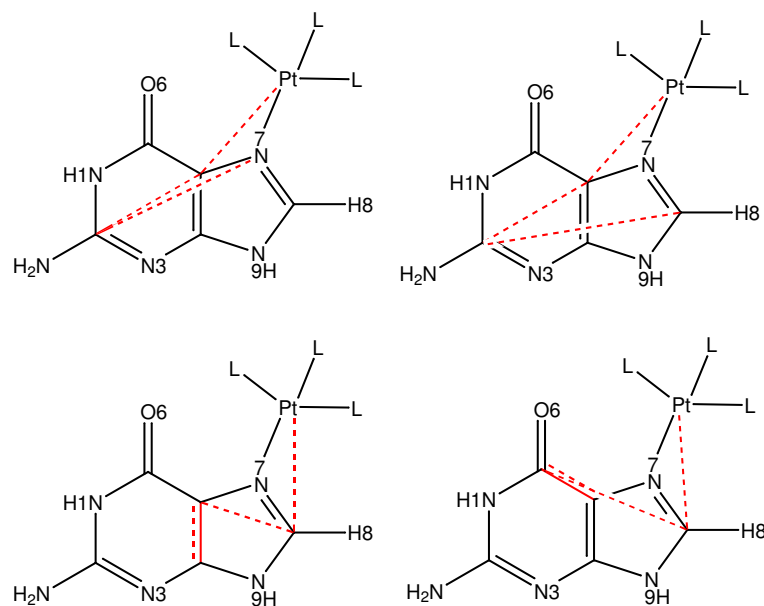Figure 3.4: Additional improper dihedral angles shown in red dashed lines, which are used to reproduce the Pt-coordination geometry in case of a Pt-Purine bindig: Pt-C5-C2-N7 ($k_2/2 = 4.7 kcal/mol$), Pt-C5-C2-C8 ($k_d/2 = 4.7 kcal/mol$), Pt-C8-C5-C4 ($k_d/2 = 30.4 kcal/mol$), and Pt-C8-C6-C5 ($k_d/2 = 27.7 kcal/mol$).
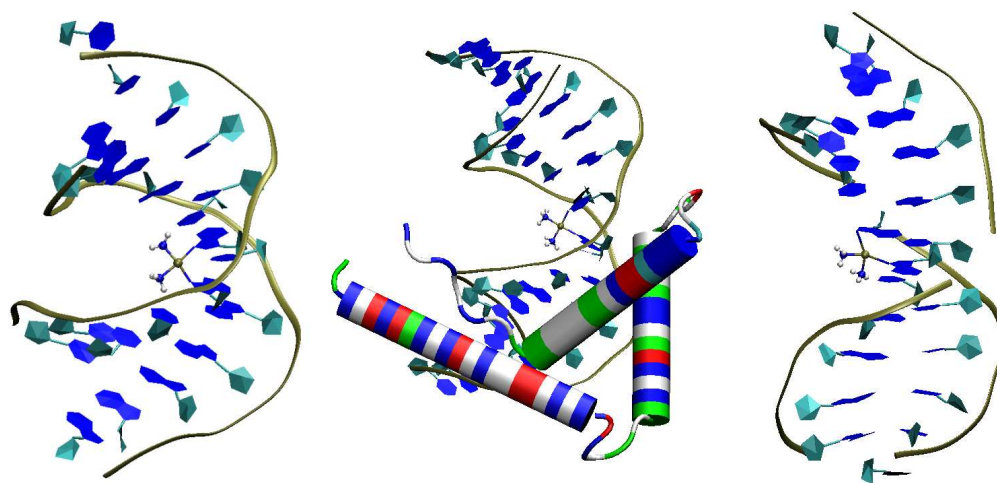

Figure 3.5: Initial QM/MM structures. left: Complex **A**, middle: Complex **B** , right: Complex **C**

	Complex **A**			Complex **B**			Complex **C**		
Number of atoms	12,280			29,535			12,349		
Number of waters	3,877			9,103			3,904		
Number of counter ions	20			20			22		
	x	y	z	x	y	z	x	y	z
Box size [Å]	59.7	45.0	43.4	67.0	72.0	61.8	63.0	44.8	44.2

Table 3.2: System Sizes of the three classical simulations at the end of equilibration.

3.2 Computational Methods

3.2.1 Structural Models and classical equilibration

The calculations are based on the following three structures:

1. Complex **A**: cisplatin-d(CCTCTGGTCTCC) X-ray structure at 2.6 Å resolution (pdb-entry 1AIO1).

2. Complex **B**: cisplatin-d(CCUCTCTGGACCTTCC)/HMG1 A domain X-ray structure at 2.5 Å resolution (pdb-entry 1CKT2). The positions of 71 out of 89 residues of HMG1 structure have been detected. The two histidine residues (HIS20 and HIS24) present in the structure were assumed to be protonated on the δ-N atom based on their putative hydrogen-bonding pattern.

3. Complex **C**: Cisplatin docked onto MD equilibrated d(CCTCTGGTCTCC)-B-DNA.

Sodium counter ions are added so as to counterbalance the total charge of **A-C**: the ions were located in the positions of lowest electrostatic potential, calculated with the AMBER electrostatic term [118]. The complexes are immersed in a periodic box filled with water molecules, at a density of $1.0 g/cm^3$. Water molecules with their oxygen atoms closer than 1.76 (i.e. the van der Waals radii of oxygen) to any atom of the complex are discarded. The final number of atoms and the sizes of the simulation boxes are reported in Table 3.2.

   All structural models undergo first an equilibration phase via classical MD, carried out with the AMBER6 program package [118]. The AMBER parm98 force field [42, 41] was adopted for DNA, the HMG A domain and the sodium counter ions. For $[Pt(NH_3)_2 - d(G^*G^*)]^{2+}$ and the guanine bases, the parametrization of Herman et al., modified by Kozelka is used [111, 112], along with a harmonic

constraint of 5 kcal/(mol·Å^2) on all the atomic positions of the cisplatin moiety $[Pt(NH_3)^2 - d(pG6^*pG7^*)]$ and, in **B**, on the C24, C25 nucleobases and residue PHE37. The latter constraint turns out to be necessary to avoid unphysical distortions of these groups. Solvent water molecules are treated with the TIP3P model [119]. Electrostatics are evaluated with the Particle Mesh Ewald (PME) method [47, 48]. A cutoff of 10 Å is used for the van der Waals interactions and the real part of the electrostatic interactions. A time step of 1 fs is applied. Room temperature simulations are achieved by coupling the systems to a Berendsen thermostat [43]. The initial structures are relaxed by short steepest descent minimization runs of 1000 steps. 100 ps MD at constant volume are performed during which the system is heated up to 300K. 2 ns of MD at constant pressure (1 Atm) and temperature (300K) is then performed. The final MD structures have an overall rmsd of 3.2Å (**A**), 1.8Å (**B**) and 3.0 Å (**C0**) with gyration radius of 11.5 (**A**), 10.2 (**B**) and 11.5 (**C0**) and final densities of the system of 1.05, 1.03, 1.02g/cm^3, respectively. No significant rearrangement is observed during the dynamics. In particular, in **C0**, 13 sugar puckers kept their C2'-endo conformation whereas the other sugars are in the closely related C1'-exo conformation.

3.2.2   QM/MM calculations

The final **A**, **B** and **C** MD structures are partitioned into two regions.

- The QM-region, treated at the DFT-level, contains the cisplatin moiety and the two guanine bases coordinated to it. Furthermore, in complex **A** and complex **B** also the phosphate group of G6 (**A**) or G8 (**B**) is included since it forms a hydrogen bond to the NH$_3$ group and is therefore fundamental to reproduce the electronic properties in the QM-region (Figure 3.6).

- The MM-region consists of the remaining nucleobases, the protein in case of complex **B**, the solvent and counter ions and is described by the modified Cornell et al. force field parm98 [41, 42].

The QM-part is cut at C1' and when the phosphate group is present also at C3' and C5' and saturated by positionally unrestrained, additional 'capping' hydrogen atoms [19].
The plane wave basis set is extended up to an energy cutoff of 70 Ry. Norm conserving pseudopotentials of the Martins-Troullier type are used [120]. Integration of the non-local parts of the pseudopotential is obtained via the Kleinman-Bylander

Figure 3.6: Schematic drawing of QM-region. Carbon atoms, where capping hydrogens are introduced are shown in red.
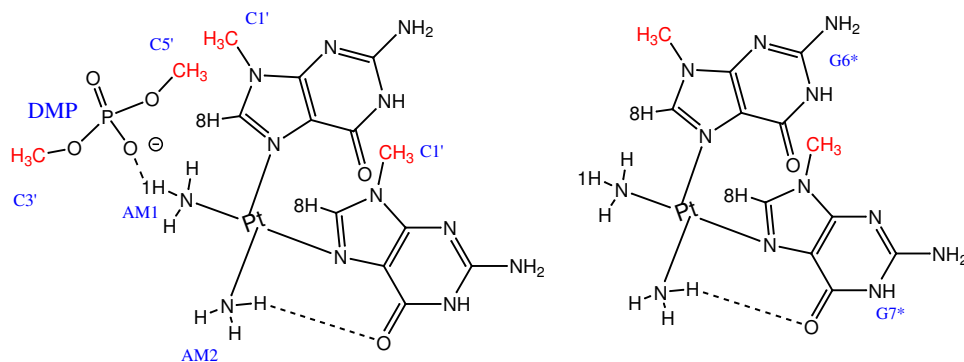
scheme [121] for all the atoms except platinum, for which a Gauss-Hermite numerical integration scheme is used. The gradient corrected Becke exchange functional and the Lee-Yang-Parr correlation functional (BLYP) were used [52, 53]. Isolated system conditions [122] are applied. D-RESP charges are calculated at each MD step as described in ref. [14]. In all complexes, the QM/MM-averaged charges are lower in absolute value than the values in the parametrization of Herman [111], possibly because of the presence of the DNA environment (Table 3.3).

The electrostatic interactions between QM and MM atoms are calculated as described in the Chapter Methods.

For the three systems the MM regions are first relaxed during 1,000 steps of MD, while keeping the QM part frozen. Then 1,000 steps of simulated annealing are performed on the QM system in order to relax the constraints on the QM-atoms. The systems are slowly heated up to 300K. A time step of 0.073fs is used. NVT simulations are carried out by coupling the systems to a Nosé-Hoover thermostat [44, 45]. The distance constraints between N(AM1)-O1P(DMP) and N7(G6*)-N7(G7*) are applied in complex **A** and **B** during the heating and released at room temperature. In Complex **C** the annealing procedure is repeated several times, decreasing step by step the distance between Pt and N7 from 2.5Å to a final value of 2.17Å. Then, the system is heated up while the distance between Pt and N7 atoms is kept fixed. At 300K, the constraints are released. The simulations cover 5ps (**A** and **B**) and 7ps (**C**), respectively.

	Complex **A**		Complex **B**		Complex **C**		Herman
Pt	0.86 (0.02)		0.91 (0.06)		0.91 (0.08)		0.86
	AM1	AM2	AM1	AM2	AM1	AM2	AM1/AM2
N	-0.18 (0.02)	-0.25 (0.06)	-0.20 (0.02)	-0.07 (0.03)	-0.17 (0.02)	-0.16 (0.02)	-0.525
H1	0.24 (0.07)	0.20 (0.04)	0.09 (0.1)	0.22 (0.06)	0.20 (0.03)	0.23 (0.05)	0.24
H2	0.22 (0.04)	0.31 (0.04)	0.072 (0.09)	0.10 (0.06)	0.27 (0.03)	0.16 (0.03)	0.24
H3	0.11 (0.06)	0.14 (0.04)	0.25 (0.04)	0.22 (0.04)	0.10 (0.03)	0.19 (0.06)	0.24
	G6*	G7*	G8*	G9*	G6*	G7*	G6/G7
N1	0.12 (0.02)	0.11 (0.02)	0.08 (0.04)	0.10 (0.02)	0.09 (0.02)	0.10 (0.02)	-0.758
H1	0.27 (0.02)	0.30 (0.02)	0.29 (0.03)	0.24 (0.03)	0.24 (0.02)	0.23 (0.02)	0.381
C2	0.14 (0.02)	0.15 (0.02)	0.11 (0.02)	0.12 (0.02)	0.08 (0.02)	0.13 (0.02)	0.939
N2	-0.04 (0.02)	-0.03 (0.03)	-0.09 (0.05)	-0.05 (0.02)	-0.07 (0.03)	-0.05 (0.03)	-0.772
H21	0.21 (0.03)	0.26 (0.02)	0.22 (0.03)	0.23 (0.02)	0.18 (0.02)	0.25 (0.02)	0.369
H22	0.12 (0.03)	0.17 (0.02)	0.16 (0.04)	0.15 (0.03)	0.15 (0.11)	0.04 (0.03)	0.358
N3	-0.24 (0.6)	-0.24 (0.07)	-0.24 (0.06)	-0.18 (0.01)	-0.35 (0.06)	-0.29 (0.05)	-0.682
C4	0.02 (0.02)	-0.01 (0.02)	-0.07 (0.04)	0.09 (0.02)	0.10 (0.02)	0.04 (0.03)	0.564
C5	-0.07 (0.03)	-0.06 (0.02)	-0.13 (0.03)	0.03 (0.03)	-0.04 (0.02)	-0.02 (0.02)	-0.682
C6	0.08 (0.01)	0.02 (0.02)	0.03 (0.03)	0.09 (0.02)	0.09 (0.02)	0.05 (0.02)	0.721
N7	-0.35 (0.03)	-0.26 (0.03)	-0.31 (0.03)	-0.19 (0.03)	-0.23 (0.02)	-0.20 (0.02)	-0.660
C8	-0.16 (0.03)	-0.11 (0.02)	-0.10 (0.02)	-0.05 (0.02)	-0.08 (0.02)	-0.03 (0.02)	0.332
H8	-0.15 (0.04)	-0.18 (0.05)	-0.02 (0.04)	-0.17 (0.06)	-0.50 (0.03)	-0.13 (0.05)	0.113
N9	0.03 (0.03)	0.04 (0.02)	0.02 (0.04)	0.10 (0.03)	0.04 (0.03)	0.11 (0.04)	-0.081
O6	-0.40 (0.04)	-0.50 (0.03)	-0.43 (0.04)	-0.55 (0.07)	-0.42 (0.03)	-0.50 (0.05)	-
O1P	-0.47 (0.06)		-0.40 (0.09)				
O2P	-0.58 (0.06)		-0.49 (0.06)				

Table 3.3: QM/MM-averaged D-RESP charges of QM regions of **A**-**C**. Comparison with RESP charges of ref. [111] is made. Standard deviations in parathesis. Labeling as in Figure 4.4
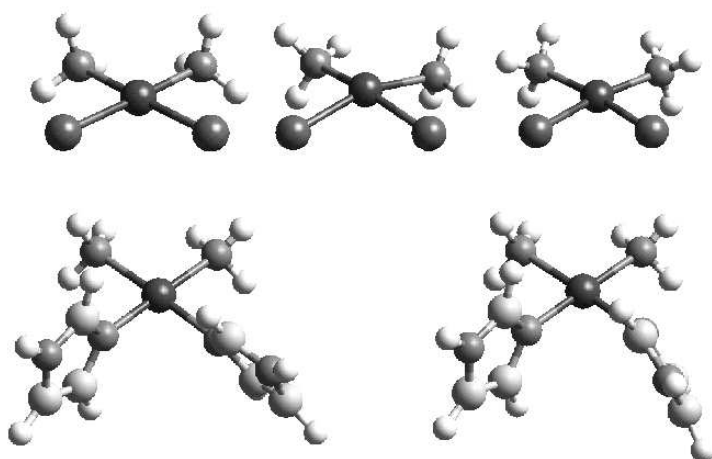
Figure 3.7: Model compounds with different types of distortions as observed in the X-ray structure. Top left: fully optimized cisplatin, Top middle: *Pt out-of plane* distortion, Top right: Cl-Pt-Cl angle =77°, Bottom left: fully optimized cispt-imidazole$_2$, Bottom right: cispt-imidazole$_2$ with Pt atom displaced from imidazole ring.

3.2.3 Properties Calculation

The QM/MM trajectories are saved every 50 steps, which corresponds to every 3.6 fs. The following structural properties are calculated:

- DNA structural parameters such as sugar pucker, axis bending and roll angle are evaluated using the program Curves [32, 33].

- Radial distribution functions (g(r)), rmsd's and gyration radii are calculated using the ptraj module of the AMBER6 package [118].

- Energy levels of HOMO and LUMO are calculated by diagonalization of the Kohn-Sham states both on the QM regions of complexes **A-C** and of gas phase models of cisplatin (Figure 3.7) in different conformations.

- ^{195}Pt chemical shifts are calculated by the ADF code [67, 123, 124]

For the ADF calculations we use a localized, Slater-type basis set, in more detail, a triple zeta basis set with one polarization function (TZP) for Pt and a double zeta basis set with one polarization function (DZP) for the other atoms along with the BP approximation [52, 125, 126] of the exchange correlation functional are used.

Bond Leght [Å]	Chemical shift	Angle [°]	Chemical shift	Displacement from ring plane	Chemical shift
1.9	0	90	0	0	520
1.95	190	88	-3	0.4	494
2.00	404	86	-4	0.8	458
2.05	650	84	-4		
2.10	913	82	-5		
2.15	1200	80	-6		

Table 3.4: ^{195}Pt chemical shifts of cisplatin and $[Pt - (NH_3)_2(C_3H_4N_2)_2]^{2+}$ in vacuo [ppm].

The BP functional is known to be more accurate than the BLYP approximation for the NMR properties, in particular for heavy atoms [69, 127]. Scalar relativistic effects are taken into account with the zero order regular approximation (ZORA) approach [123, 68, 69]. Test calculations on cisplatin and $[Pt(NH_3)_2(C_3H_4N_2)_2]^{2+}$ $[cispt - (imidazole_2)]^{2+}$ (Figure 3.7) are performed to estimate the effect of distortions on the calculated chemical shift. It is found that the latter is very sensitive to bond lengths but to a lesser extent to a distortion of the angles and out of plane displacement of the metal ion Table 3.4. The ^{195}Pt chemical shift, which are calculated on selected snapshots of the CPMD/classical MD trajectory, are expected to experience extremely large variations, because of the fluctuations of the Pt-ligand bond lengths during the dynamics. Thus, calculations can provide only qualitative results here. Notice that spin orbit terms [127] account for 10% of the chemical shift, as shown by test calculations on the cisplatin molecule and cisplatin-guanine complexes $[Pt(NH_3)_2G_2]^{2+}$. Indeed, as already discussed previously [123], the ^{195}Pt chemical shift is mainly determined by the paramagnetic contribution [124]. Thus, these terms are not expected to affect significantly the results and they are not included here because of their large computational cost. The difference between these chemical shifts and that of cisplatin in water (geometry taken from the MD-averaged structure of Ref. [116]) are finally calculated.

3.3   Results

QM/MM calculations are carried out with a platinated DNA dodecamer (**A**, Figure 3.5) and a platinated 16-mer in complex with the HMG domain (**B**, Figure 3.5), based
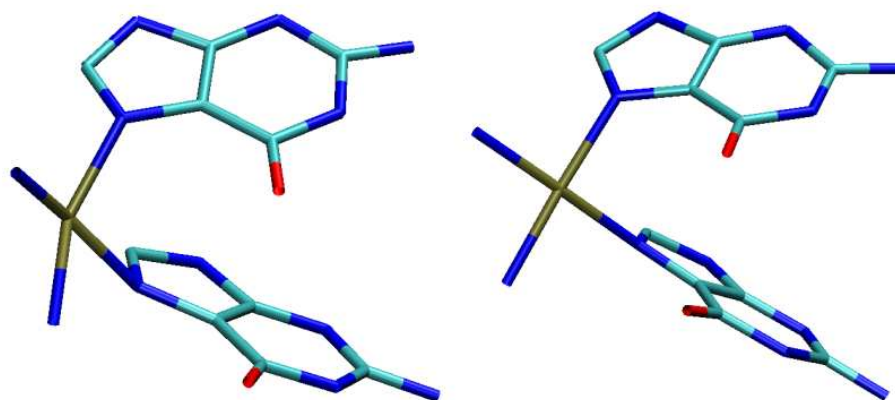
Figure 3.8: Coordination geomerty of cisplatin moiety in complex **A**. Left: initial structure (from X-ray) and right: QM/MM structure (average).

on the X-ray structures determined by Lippard and co-workers [85, 90, 91]. The calculations were preceded by 2.0ns classical MD simulations allowing for a relaxation of the DNA structure, while keeping the cisplatin moiety close to its inital position. Comparison is made with NMR structural data with ^{195}Pt chemical shifts of cisplatin-DNA adducts [128, 129] This section is concluded by a docking of the $[Pt(NH_3)_2]^{2+}$ moiety on a B-DNA dodecamer (**C**, Figure 3.5) and subsequent QM/MM simulation.

3.3.1   Complex A

Coordination geometry and electronic structure: The cisplatin moiety is very similar to the initial X-ray structure (rmsd: 0.9 Å, Table 3.5). However, the Pt-ligand distances are slightly longer than in the X-ray or in the NMR-structure, as already noticed in calculations with this computational setup [116]. The platinum ion, which is moved out of the square planar coordination plane in the X-ray structure, lies now in the plane defined by its four ligands (Figure 3.8).

The displacement of the metal ion from the purine ligands is smaller than in the X-ray structure. It is instead rather similar to that in the NMR structure of the same molecule in water (Table 3.5). The distortion of the Pt-N7-purine ring angle is known to be related to the destacking of the G6*G7* bases [90]: a larger roll angle between the platinated residues allows for a less distorted Pt-coordination geometry [91], favoring in this way complex formation. On the other hand, because of the increased roll angle, the hydrophobic bases are more solvent exposed than in undistorted B-

	A	**B**	**C**	X-ray **A** [85]	X-ray **B** [91]	NMR **C** [90]
	Platinum coordination geometry					
Pt-N7(G6*)	2.06(0.007)	2.02(0.15)	2.07(0.06)	1.91	1.98	1.94
Pt-N7(G7*)	2.03(0.06)	2.00(0.14)	2.07(0.04)	1.90	1.89	1.89
Pt-N(AM1)	2.11(0.06)	2.09(0.15)	2.07(0.02)	1.89	1.98	1.98
Pt-N(AM2)	2.09(0.06)	2.08(0.15)	2.08(0.03)	1.96	1.96	1.92
N7-Pt-N7	86(3)	89(3)	86(3)	100	90	90
AM1-Pt-AM2	89(5)	91(4)	92(3)	92	89	90
N7-Pt-AM1	90(5)	92(5)	89(4)	88	89	90
N7-Pt-AM2	93(5)	88(4) 93(4)	72	92	90	
Displacement of Pt from G6*	0.8(0.1)	0.4(0.1)	0.74(0.1)	1.3	0.48	0.7
Displacement of Pt from G7*	0.34(0.1)	0.09(0.1)	0.8(0.1)	0.8	0.0	0.64

Table 3.5: Average bond lengths [Å] and angles [°] at platinum coordination site from QM/MM simulations compared to experimental values.

DNA. Consistently, in our final QM/MM structure the N7-Pt-N7 angle increases from $77°$ in the X-ray structure to $86°$ (Table 3.7) and the roll angle is $42°$ (Table 3.5).

In order to estimate the effect of these distortions on the electronic structure and on the energetics, we carry out calculations on model systems, namely cisplatin and the imidazole containing complex $[Pt(NH_3)_2(C_3H_4N_2)_2]^{2+}$ in vacuo (Figure 3.7). Different conformations, with geometries similar to that of the $[Pt(NH_3)_2 - d(G6 * G7*)]^{2+}$ geometry in the X-ray structure, are compared to the fully optimized geometries. The energy of the molecule on passing from the 'out of plane' conformation, as observed in the X-ray structure, to the optimized geometry is 8.0 kcal/mol . Such an energy barrier suggests that the 'out of plane' conformation found in the X-ray structure has very low probability. However, environmental effects, which are neglected in these calculations, could alter significantly this picture. In addition, removal of this distortion causes an increase of the HOMO-LUMO gap from 2.5 eV to 2.8 eV. Distortion of the Cl-Pt-Cl angle has a much less significant effect (Table 3.6). In all conformations, the HOMO is mainly localized on the corresponding metal d_{xy} orbital, as in the geometry-optimized structure [117] and the LUMO is on the $d_{x^2-y^2}$ orbital, which is to be expected for a d^8 diamagnetic square planar complex. The same holds true for cisplatin-Guanine$_2$ complexes in vacuo (Figure 3.9).
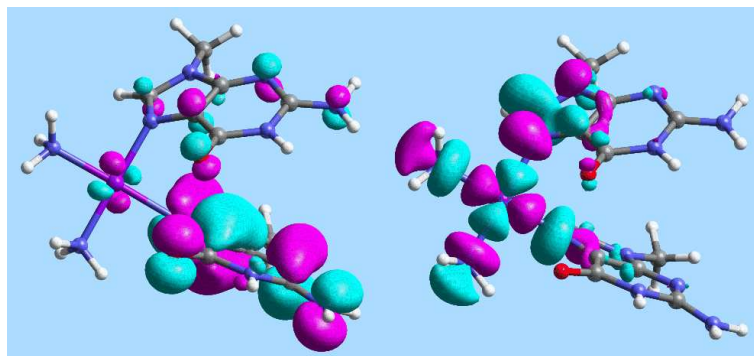
Figure 3.9: HOMO (right) and LUMO (left) KS-orbitals for cisplatin-Guanine$_2$ complex in vacuo.

The distortion causes also a slight elongation of the metal-ligand bonds (Table 3.6). The same holds true for a displacement of the Pt atom from the imidazole rings. The presence of the biomolecular frame modifies the nature of the chemically relevant orbitals. In the QM region of complex **A** the HOMO lies in the π-system of G7* whereas the LUMO is localized on the phosphate group (Figure 3.10).

The HOMO-1 corresponds to the d_{xy}-orbital, but all the d-levels are clustered within 1eV and can also be switched during the simulation (Figure 3.11). The $d_{x^2-y^2}$ level, on the other hand is shifted to the LUMO+1 level. Reordering of the electronic energy levels is probably due to the perturbation by O6 of the guanine ligands (isodensity surfaces are shown in Figure 3.10) and has already been observed in Pt-nucleotide complexes [116].

The ^{195}Pt NMR chemical shift relative to the cisplatin in vacuo with the average MD structure is 637 ($\pm$353) ppm, as compared to the experimental value of 310ppm [128] (Table 3.6) [129]. Typical errors in calculated ^{195}Pt chemical shifts for energy-minimized square planar complexes involving Cl and N ligands range from 30ppm to 160ppm [124]. The larger error found here (over 300ppm) might be at least in part caused by the fact that the calculations are based on QM/MM snapshots where bond lengths are highly fluctuating. This in turn dramatically affects the ^{195}Pt NMR chemical shifts so that only qualitative agreement with experimental values is obtained.

DNA distortion due to cisplatin binding: As mentioned above, cisplatin binding to DNA causes not only a distortion in the cisplatin coordination geometry but also considerable structural changes in the DNA. During the QM/MM simulation, the average DNA curvature increases to an intermediate value between that in the X-ray
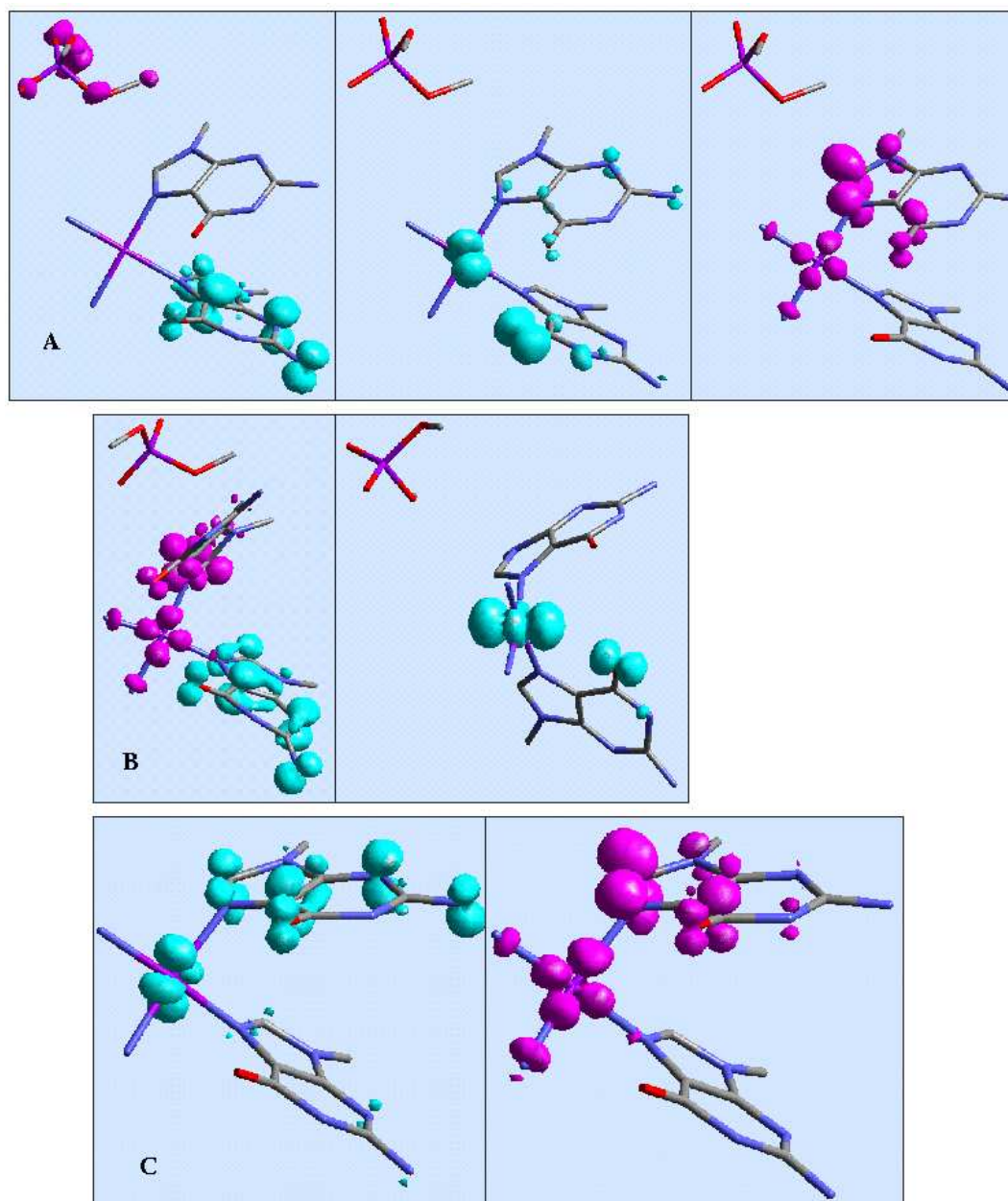
Figure 3.10: Isodensity surfaces of KS-orbitals of complexes **A**-**C**, calculated on a snapshot during the QM/MM simulation. (**A**): Left: HOMO (light blue) and LUMO (pink), Right: HOMO-1 (light blue), LUMO+1 (pink). (**B**): Left: HOMO (light blue) and LUMO (pink), Right: HOMO-1. (**C**): Left: HOMO, Right: LUMO.

A	I	II	III	IV	V
Geometry					
Pt-Cl/Pt-IMI	2.34	2.35	2.36	2.05	2.05
Pt-N	2.10	2.12	2.10	2.09	2.09
Cl-Pt-Cl/	96	93	76	88	88
IMI-Pt-IMI					
N-Pt-Cl/	83	82/83	94	96	91
N-Pt-IMI					
N-Pt-N	98	96	94	90	90
out of plane	0	29.3	2.8	0.0	0.0
Displacement	-	-	0.0	0.8	
from imidazole					
HOMO-LUMO energy gap					*[eV]*
CPMD	2.826	2.478	2.672		
ADF	2.667	2.306	2.490	3.42	3.38

B	A	B	C
HOMO-LUMO energy gap [eV]			
CPMD	2.4	2.6	3.0
ADF	2.3	2.3	2.3
^{195}Pt *NMR chemical shift with respect to cisplatin in vacuo*			
ADF	637 (353)	534 (243)	532 (140)

Table 3.6: Geometric and electronic properties of A: cisplatin (CPT) and $[Pt(NH_3)_2(N_2C_3H_4)]^{2+}$ (CPT-IMI). I: CPT, optimized structure, II: CPT with Pt 'out-of-plane' distortion as in crystal structure, III: CPT with Cl-Pt-Cl set to $77°$, IV: CPT-IMI optimized structure, V: CPT-IMI, Pt displaced from imidazole ring. B: complexes **A-C**
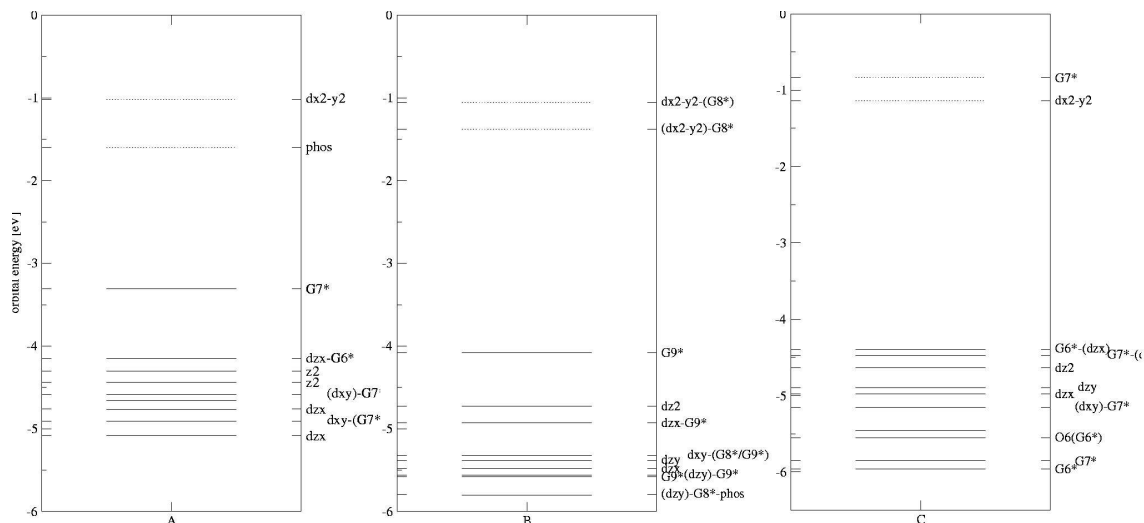
Figure 3.11: Energy level diagram of KS-orbitals for Complex **A** (left), Complex **B**, and Complex **C**, calculated on a snapshot during the QM/MM simulation.

structure and that in the NMR-structure (Table 3.7), remaining however far from the $84°$ observed in the solution structure [90]. During the short timescale explored, the overall axis curvature covers values from $30°$ to $76°$. The measurement of the overall curvature can be affected by the way the overall axis is defined. It is therefore useful to consider the local axis bend instead, measured between normal vectors of two adjacent base pairs. We see that the curvature originates in a large axis bend at the G6-G7 step. The local axis bend at G6-G7 steps is still close to the one observed in the X-ray structure, but the flanking residues have now as well an increased axis bend (Figure 3.12). Also in the NMR structure, the curvature is not completely localized at the platinated base-pair step, but extends over 5 base pair steps. The local axis bend does not carry any information about the directionality of the curvature. This information can be obtained when looking at the roll or tilt angle. In our case, the increased axis bend is mainly due to an increased roll angle. A positive roll corresponds to a bend towards the major groove, whereas a negative roll indicates a bend towards the minor groove. Here, the roll angle is positive at the G6-G7 step, and thus the DNA is bent towards the major groove (Figure 3.13).

 The square planar coordination geometry imposed by cisplatin forces the otherwise stacked guanine bases to open up in a way that leads to a larger roll and rise (Figure 3.13). The latter measures the vertical distance between the two bases. The cisplatin lesion affects also the opposite strand, in particular the minor groove width and depth

Figure 3.12: Left: Local Axis bend for the platinated base pair step and the two flanking base pair steps towards either side of the cisplatin lesion, Right: Helical Twist again for the 5 central base pair steps.
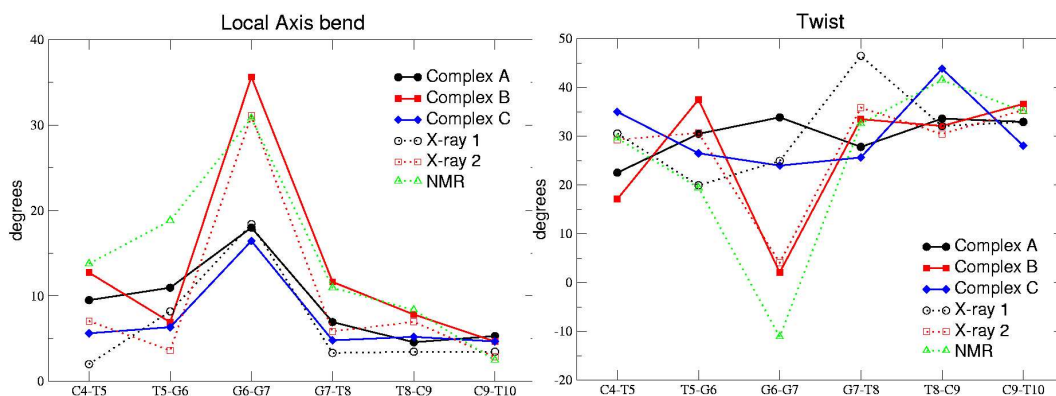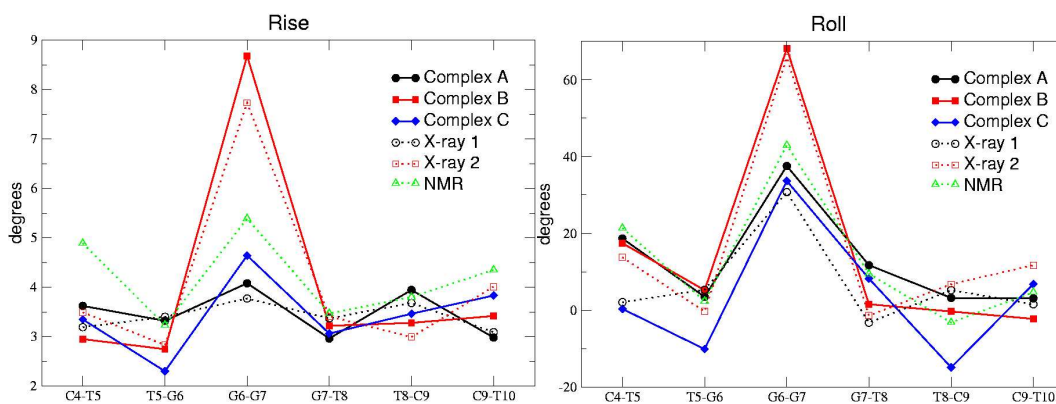

Figure 3.13: Left: Rise for the platinated base step and the two flanking base pair steps towards either side of the cisplatin lesion, Right: Roll between platinated bases and for the two adjacent base steps.
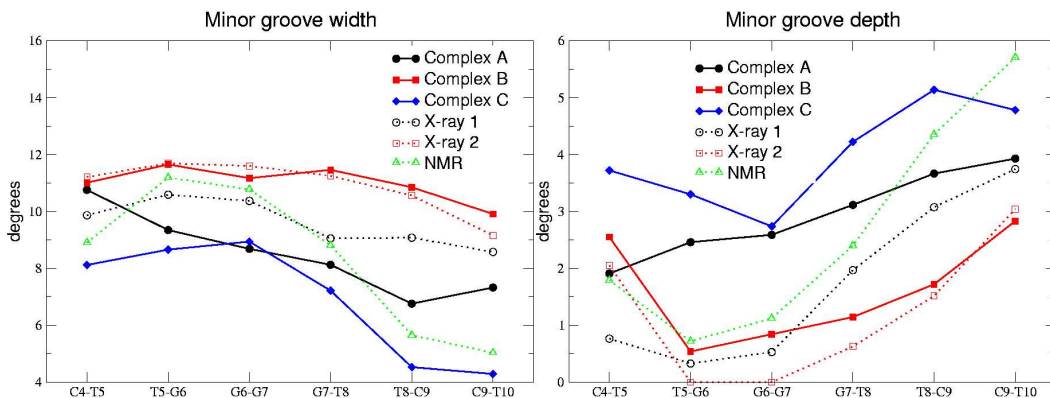
Figure 3.14: Left: Minor groove width, Right: Minor groove depth.

and the helical twist. The minor groove, being now very wide and shallow, resembles more standard A-DNA (Figure 4.5). The helical twist at the platinated base pair step is extremely low and thus the double helix highly unwound (Figure 3.12).

The Watson-Crick hydrogen bonds around the platinated site are expected to be perturbed due to the large distortion induced by the cisplatin moiety. The G6*-C19, which is intact in the X-ray structure, breaks and reforms several times. This feature may be consistent with the NMR data, which indicate that the G6*-C19 base pair is perturbed, retaining only the G6*(O6)-C19(N4) hydrogen bond [90]. The simulation suggests that also the G7*-C18 base pair is highly flexible and the Watson-Crick H-bond is broken and reformed several times. This bond is intact in the solution structure, but distorted in the crystal structure [85]. Since our QM/MM simulation is too short to obtain converged results, we monitor the presence of H-bonds over a 2ns classical molecular dynamics simulation and compare it to a classical MD of the free DNA with the same sequence. The Watson-Crick hydrogen bonds are disturbed in the complexed form as can be seen by the lower occupancy. Propeller twists are increased with respect to standard B-DNA, having values as high as $30°$ (Table 3.7). Another important effect of cisplatin binding to DNA regards the conformation of the sugar-phosphate backbone. Indeed, it is found experimentally, that the ribose moiety of G6 undergoes a repuckering from C2'-endo, typical for B-DNA, to C3'-endo, which is typical for A-DNA. The QM/MM averaged values of the key backbone torsional angles are reported in Table 3.7. The experimentally observed repuckering from A-DNA in the solid state to B-DNA in aqueous solution occurs on a time scale of 200-500ps [4]. Here, as expected, it is not observed during our simulation time.

	A	**B**	**C**	X-ray **A** [85]	X-ray **B** [91]	NMR **C** [90]
Curvature	51(10)	57(5)	48(8)	40	66	85
sugar pucker G6	C3'-endo	C2'-exo	C2'-endo	C3'-endo	C3'-endo	C2'-exo
sugar pucker G7	O1'-endo	C2'-endo	O4'-endo	C4'-exo	C2'-endo	C2'-endo
minor groove	8.3-8.7	11.0-11.7	8.6-9.1	10.8	12.0	12.0
propeller twists						
T5-A20	-14 (11)	1 (10)	-2 (9)	7	3	30
G6-C19	-34 (13)	-6 (10)	-18 (10)	-24	-1	-7
G7-C18	-16 (11)	5 (9)	-12 (11)	-16	20	-9
T8-A17	-11 (13)	-9 (11)	-11 (9)	-30	2	-3
torsional angles						
G6*χ	-149 (9)	-98 (10)	-80 (10.5)	-147	-108	-109
G6*δ	82 (8)	91 (9)	132 (11)	91	105	92
G6*ξ	-54 (11)	-64 (9)	-84 (42)	-56	-81	-92
G7*χ	-149 (14)	-113 (15)	-126 (15)	-168	-110	-141
G7*δ	85 (13)	140 (10)	107 (19)	82	137	136
G7*ξ	-76 (11)	-103 (60)	-98 (16)	-82	-148	-91

Table 3.7: Selected Structural Parameters Describing the DNA structure of QM/MM simulation compared to experimental structures. In case of **B**, the following substitutions in base-numbering occur: T5=T7,A20=A26,G6=G8,C19=C25,G7=G9,C18=C24,T8=A10,A17=T23

	A	B	C	NMR struct.
A	0.9/1.8	1.04/-	0.9 / 2.98	1.50/4.20
B		0.8/1.8	1.16/-	1.32/-
C		1.34/2.02	1.47/3.96	

Table 3.8: Mutual rmsd's in [Å] of complexes **A-C** and in the NMR structure of ref. [90]. The first value refers to the Pt-moiety s shown in Figure 3.6, the second value refers to the entire double helix. Rmsd between the same structures refer to the inital and final average QM/MM structure.

The final rmsd of complex **A** with respect to the initial crystal structure and the NMR structure is 1.8 Å and 4.0 Å, respectively including all atoms. When considering only the Pt-moiety the rmsd is 0.9 Å and 1.5 Å for the X-ray structure and NMR structure, respectively (Table 3.8).

Ligand H-bonding: The AM1-G6 phosphate hydrogen bond (Figure 4.4), present in the X-ray structure, is first broken and then reformed several times until, after about 3 ps it is definitively lost. This feature may be consistent with the NMR structure of the complex in solution, which does not exhibit this H-bond [90]. The relatively weak AM2-O6(G7*) H-bond, present in the X-ray and NMR structure, is instead fully conserved during the dynamics. The other hydrogen belonging to the two ammonia ligands form H-bonds with water molecules (treated classically). In the first solvation shell there are on average 2.8 water molecules (Table 3.9). The radial distribution function between ammonia and water and between Pt and water for the QM/MM simulation and the classical MD of the same complex are compared in Figure 3.15.

3.3.2 Complex B

Pt-coordination geometry and electronic structure: The initial platinum coordination geometry, which is less distorted than in the X-ray structure of the platinated DNA in the free state (Table 3.5), is well maintained. The Pt-atom lies in the square planar coordination plane in agreement with the X-ray structure. The HOMO lies mainly in the π^*(G9*) -system as in **A** and the LUMO lies on the $d_{x^2-y^2}$ orbital. The HOMO-1 is concentrated in the d_{z^2}-metal orbital. The latter is destabilized due to the almost axial position of the O6(G9*) oxygen with respect to the Pt-coordination plane (Figure 3.10).
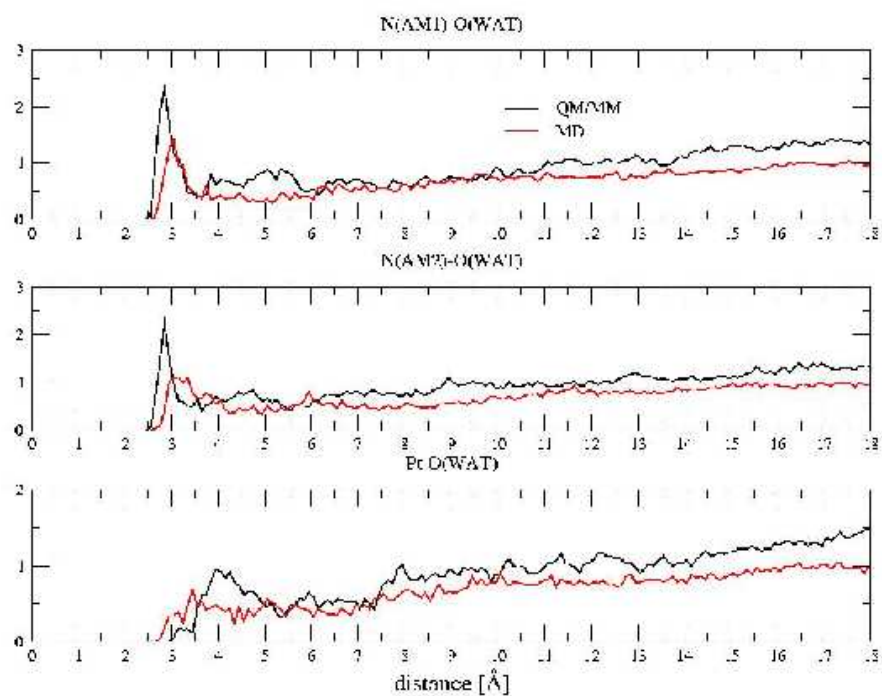
Figure 3.15: Water Radial distribution functions (g(r)) for AM1 (upper panel), AM2 (middle) and Pt (lower panel) of complex **A**.
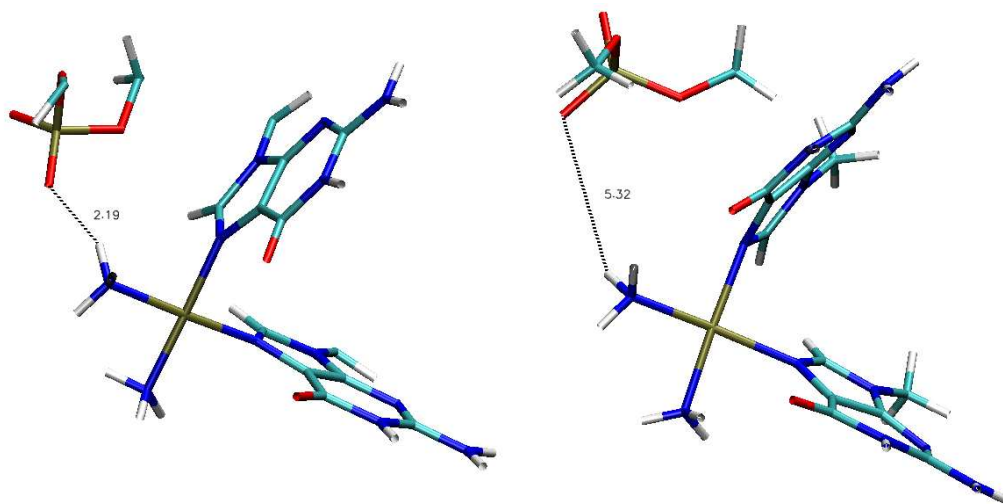
Figure 3.16: Left: Initial Pt-coordination geometry in the X-ray structure. The hydrogen bond between the phosphate group of G8 and AM1 is still intact. Right: Final Pt-coordination geometry. The hydrogen bond between phosphate group and AM1 is not present anymore.

Again, the d-levels are clustered within 1eV (Figure 3.11).

DNA distortion due to cisplatin and protein: The overall axis curvature in the DNA decreases slightly during the simulation from $66°$ to $57°$. Again, the increased roll angle at the G8-G9 step, which is $61°$ in agreement with the X-ray structure, accounts mainly for the overall curvature (Figure 3.13). The rise at the G8*G9* step is significantly larger than in complex **A** (7.7 Å) and in fair agreement with the crystal structure (7.4 Å) (Figure 3.13). Larger destacking of the two bases leads to a smaller distortion of the Pt-N7 bond [91]: the displacement of the Pt-atom from the purine planes is smaller than in **A**. On the other hand, the hydrophobic faces of the guanine bases are even more solvent exposed than in **A**. The sugar puckers and the Watson-Crick H-bond patterns are maintained (Table 3.7, Figure 3.18). In terms of torsional angles, the structure has B-DNA character except for G8* and C11 which exhibit A-DNA character (Table 3.7). G8*, binding to Pt in 5' position has A-DNA conformation also in the X-ray structure [91], but C11 has B-DNA character in the X-ray structure. The repuckering could be due to thermal fluctuations. The minor groove width is similar to the initial X-ray structure (Figure 4.5, Table 3.7).

Ligand Hydrogen Bonding: The hydrogen bond between the phosphate group

	A	B	C
AM1	2.8	4.3	4.2
AM2	2.3	3.4	3.4
Pt	6.1	8.3	8.8

Table 3.9: Number of water molecules in the first solvation of ammonia (within 3.5 Å) and Pt /within 5.0Å

of G8* and AM1 is lost at the beginning of the dynamics (Figure 3.16), when the restraint is released whereas the N(AM1)-O6(G8*) and N(AM2) and O6(G9*) hydrogen bonds are maintained. In contrast to **A**, all three hydrogen atoms of AM1 are able to form hydrogen bonds to the solvent and therefore there are more water molecules in the first solvation shell than in complex **A** (Table 3.9). There are more than three waters in the first solvation shell of AM1. As mentioned in ref. [130], this feature is due to the presence of mobile water molecules in the first solvation shell, which are not directly hydrogen bonded to ammonia. AM2 has less water molecules in its first solvation shell than AM1 since one hydrogen is involved in a hydrogen bond towards O6 of G9*.

3.3.3   Complex C: Docking of $[Pt(NH_3)_2]^{2+}$ on B-DNA

The predictive power of our computational setup is investigated by constructing a structural model of platinated DNA. The $[Pt(NH_3)_2]^{2+}$ moiety is docked onto the two adjacent guanines of B-DNA with the same sequence as **A**. The B-DNA structure in solution is previously equilibrated by 2 ns of classical dynamics **C0**.

Pt-coordination geometry and electronic structure:
The initial Pt-coordination geometry after docking is highly distorted (Figure 3.17). The initial N7-Pt-N7 angle is $135°$ and the initial N7-Pt bond lengths are fixed to 2.5Å. The stabilization energy due to relaxation of the system, calculated by comparing the energy of $[Pt(NH_3)_2G_2]^{2+}$ moiety in its initial and final structure, is as high as 150kcal/mol. Most of this energy resides in the rather strong coordination bonds [131]. It is therefore not surprising that the bond-lengths adjust themselves very fast. At the end of our simulation, the square planar coordination geometry is slightly more distorted than in **A**, as shown by the displacement of the Pt atom from the purine ring planes (Table 3.5). The calculated HOMO and LUMO gap is slightly larger than in complex **A** (Figure 3.11). The HOMO has large parts of the density on the π-system
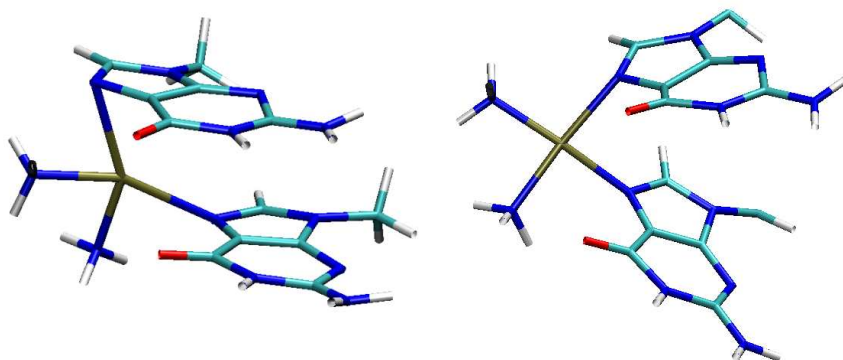
Figure 3.17: Pt-coordination geometry in complex **C** at the beginning (left) and at the end (right) of our simulation.

of the guanine ligand, whereas the LUMO corresponds to the $d_{x^2-y^2}$ as in **B**. The d-levels lie just below the HOMO clustered within 1eV. The ^{195}Pt NMR chemical shift relative to cisplatin is 532 ($\pm$116) ppm, also in this case in qualitative agreement with experiments [128, 129]. The very large standard deviations do not allow establishing whether the agreement is better here or in complex **A**.

DNA distortion due cisplatin moiety: The distortion in the DNA structure induced upon cisplatin binding is less immediate. Still, on the very short time-scale explored, we observe considerable conformational changes in the DNA double helix. A relaxation of the initially strongly distorted N7-Pt-N7 angle forces the two coordinated guanines to open, which leads to an increase in the roll angle (Table 3.5, Figure 3.13). The final value is still below the one of **A**, but the running average shows an increasing trend. Since in DFT dispersion interaction are poorly reproduced, one could argue that the opening of the two bases is an artefact of DFT. We perform therefore a short classical molecular dynamics simulation and verify that also in this case, the two guanines are getting destacked. Both local axis bend and roll angle at the G6-G7 base-step increase with final values of $26°$ and $50°$, respectively.

The increased positive roll at the G6-G7 step leads to a local axis bend, which translates into an overall curvature in the double helix. The average curvature is $48°$ ($\pm 8°$), lower than in the reported NMR structure [90], but in good agreement with **A**. The local axis bend and roll angles are shown in Figure 3.13 and 3.12 and propeller twists are reported in Table 3.7.

The torsional angles and sugar puckering are maintained (Table 3.7).
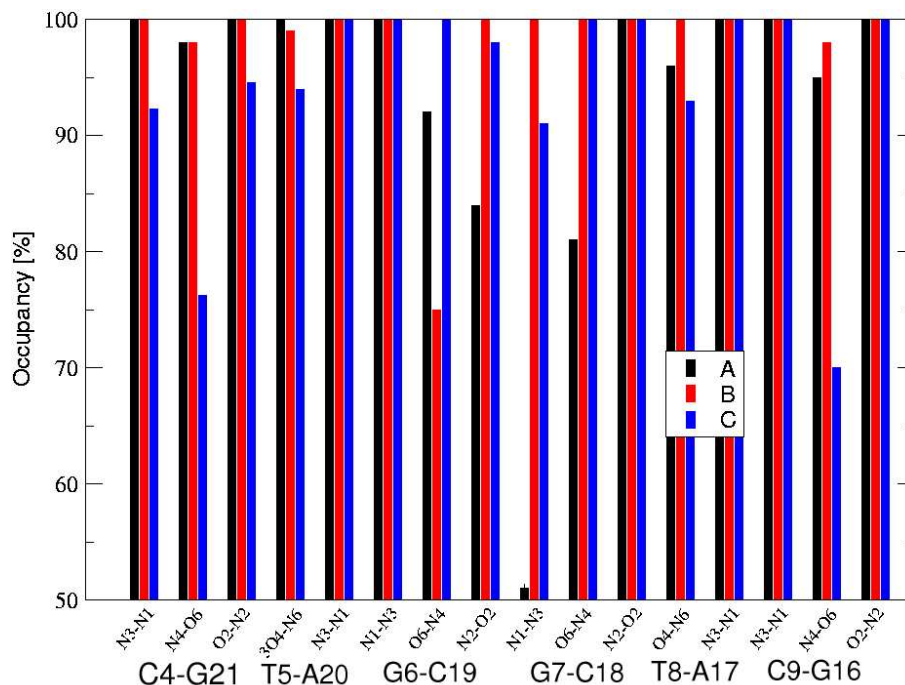
Figure 3.18: occupancy of Watson-Crick hydrogen bonds around the central 6 base-pairs for the three simulations **A**-**C**.

The AM1 ammonia ligand does not interact with the phosphate group of G6* during the duration of the QM/MM simulation, the distance between the groups decreases from 9 to 7 Å, which corresponds to an indirect hydrogen bond, mediated by two water molecules. Classical MD simulations have shown several transitions from a direct phosphate-AM1 hydrogen bond to a water mediated hydrogen bond [108] as observed when passing from crystal [85] to solution structure [90]. During a very short interval in the MD-dynamics [108] also the two water mediated hydrogen bond has been observed as in **C**. A direct hydrogen bond favors however N-conformation of the sugar in 5'-position of the Pt-lesion, whereas with increasing distance between AM1 and the phosphate group, S-conformation becomes accessible as well. The Watson-Crick hydrogen bonds are not disrupted during the simulation, at most, there is a slight perturbation at the base-pairs adjacent to the Pt-lesion (Figure 3.18). In general, the structure maintains overall B-DNA character and is less distorted than complex **A**. Nevertheless, the final structure is quite similar to complex **A** with an rmsd of 0.9 Å including the atoms at the Pt-site, and 3 Å including all atoms (Table 3.7). The main structural differences include the sugar ring of G6* and the G6*-phosphate group
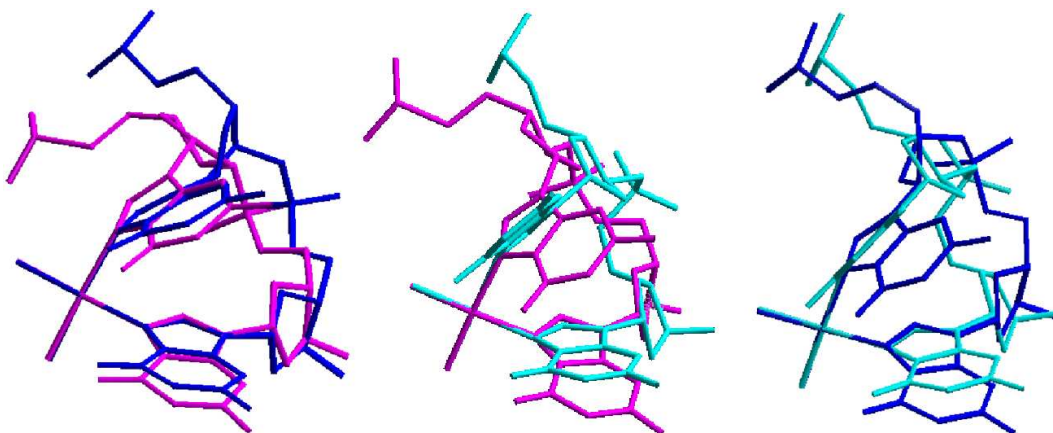
Figure 3.19: Left: superposition of platinated moiety of complex **A** and complex **C**, Middle: Superposition of Pt-moiety of complex **A** with NMR-structure, Right: Superposition of Pt-moiety of Complex **C** with NMR-structure.

(Figure 3.19). The latter is located further away from the AM1 ligand than in complex **A**. The rmsd of complex **C** with respect to the solution structure is 1.5 Å at the platinated site and 4.0 Å, including all atoms. Thus the final **A** and **C** structures, even though generated from completely different structures, are in fair agreement with each other.

Ligand hydration: The hydrogen bond between AM2 and O6 of G7* is present as in complexes **A** and **B**. AM1 solvation properties are similar to that of **B** (Table 3.9), given the better solvent accessibility of AM1 with respect to complex **A**.

3.4 Conclusions

We have presented a mixed quantum/classical simulation of cisplatin binding to DNA in the free state and in complex with HMG A (Figure 3.2).

In **A**, the coordination geometry of the cisplatin moiety relaxes considerably during the QM/MM simulation, which is based on the X-ray structure, suggesting that the 'out-of-plane' position of the platinum atom is due to crystal packing effects. The angle spanned by the two purine bases and the platinum increases on average by $10°$ and both roll angle and DNA curvature increase. In consequence, the Pt displacement from the purine ring planes decreases to 0.8Å and 0.4Å, respectively. These values are comparable to the solution structure [90]. The orientation of the coordinated guanines still differs from solution structure and compares actually better to the X-ray

structure. We check the dependency of our final results upon the initial structure by performing simulation **C**. In this simulation we start from a highly distorted structure, where the cisplatin moiety is docked on canonical, straight B-DNA. In spite of the short time-scale investigated, the docking process causes a rather dramatic change in the DNA structure: The axis bend, the roll angle and the rise increase towards the values of **A** and are still increasing in the end of our simulation. The good agreement between **A** and **C** at the platinated moiety is a strong validation of our model.

A second important point for the reliability of our QM/MM approach is the correct description of the DNA distortion observed in cisplatin-DNA complexes. Stacking interactions are not well described in DFT methods and this could lead to a trend to larger roll angles. We compare therefore our results to a fully classical simulation starting from the cisplatin docked on canonical B-DNA (**C**). Local axis bend and roll angle at the platinated step are in good agreement with the QM/MM structure **C** and even slightly larger, showing that the destacking of the two guanine bases is not an artefact of DFT.

We also attempted to calculate ^{195}Pt chemical shifts. The final results are only in qualitative agreement with experimental data [128, 129] and the too large errors in our data do not allow to establish a structural relationship.

In complex **B** the structure remains very close to the X-ray structure. In this case, crystal packing effects play probably a less important role and therefore no large conformational changes are observed when passing from solid state to solution structure, showing that the QM/MM method reproduces well the overall structural parameters.

The binding of the HMG A domain leads to a higher axis bend.[89] Still, as in **A** and **C**, the DNA maintains its high flexibility. The DNA adopts a larger rise and roll angle at the platinated site, leading to a smaller displacement of the Pt atom from the purine rings. The strain on the Pt-coordination geometry is therefore less than in complex **A**. Nevertheless, the two structures compare well, given that the orientation of the two guanine ligands is the same, which confirms again the correct reproduction of the Pt-coordination geometry (Figure 3.19).

The sugar repuckering experimentally observed on passing from the solid state [85] to the solution structure [90] is not seen during the QM/MM simulation because of the short simulation time. Hydrogen bonding between one of the Pt-ammonia ligands and a phosphate group is not stable through the entire simulation time. This is consistent with experiment where the same H-bond is lost when passing from solid state [85] to solution [90]. It has been argued that this hydrogen bond might be important for stability. Our simulation and other MD simulation suggest that once the complex is formed, the direct hydrogen bond is in equilibrium with an indirect hydrogen bond

mediated by water molecules. Still, H-bonding to phosphate groups could be important during complex formation.

We conclude that the QM/MM approach described here is a valuable tool for metal-DNA models, especially if combined with classical molecular dynamics simulation. It allows for a parameter free treatment of the metal-coordination geometry within the full DNA-solvent environment treated classically and it can be used in the future to model the interaction of other platinum-based compounds with DNA oligomers and DNA nucleobases, for which a valuable force field parametrization has not yet been developed.

Chapter 4

DNA Alkylation by duocarmycins

4.1   Introduction

Duocarmycins make part of a class of exceptionally potent antitumor-antibiotics and recently, the first duocarmycin derivative has been introduced into clinical trials [21]. Since their disclosure in the 1990's [132] and the demonstration that they derive their biological properties through a sequence selective alkylation of duplex DNA, extensive efforts have been devoted to defining the fundamental principles underlying the relationships between structure, chemical reactivity, and biological properties. These molecules, which contain a variable number of substituted aromatic ring systems (Figure 4.1), bind preferentially to A-tracts, forming a covalent bond with N3 of adenine at the end of an A-tract. The moiety reactive towards DNA consists of cyclopropyl-indole unit condensed with a pyrrole ring (Ring **A**, Figure 4.1). The least substituted carbon atom in the cyclopropyl unit of ring **A** performs a stereoelectronically controlled addition to adenine at the N3 position [133]. Intriguingly, DNA acts as a catalyst for the alkylation reaction, increasing k_{obs} by at least five orders of magnitude relatively to adenine in water k_{obs} depends also strongly on the nature and extent of the substituents, usually indole rings, which are connected to ring **A** by an amide link.

The mode of binding of two derivatives (**DSA** and **DSI**, Figure 4.1) to DNA has been provided by the recent NMR structures of the complexes between these drugs and d(pGpApCpTpApApTpTpGpApC)- d(pGpTpCpApApTpTpA*pGpTpC)(**I**) [134, 135]. The drugs are accommodated in the minor groove, forming a covalent bond between its cyclopropyl unit and the last adenine (indicated by a star) of the so-called
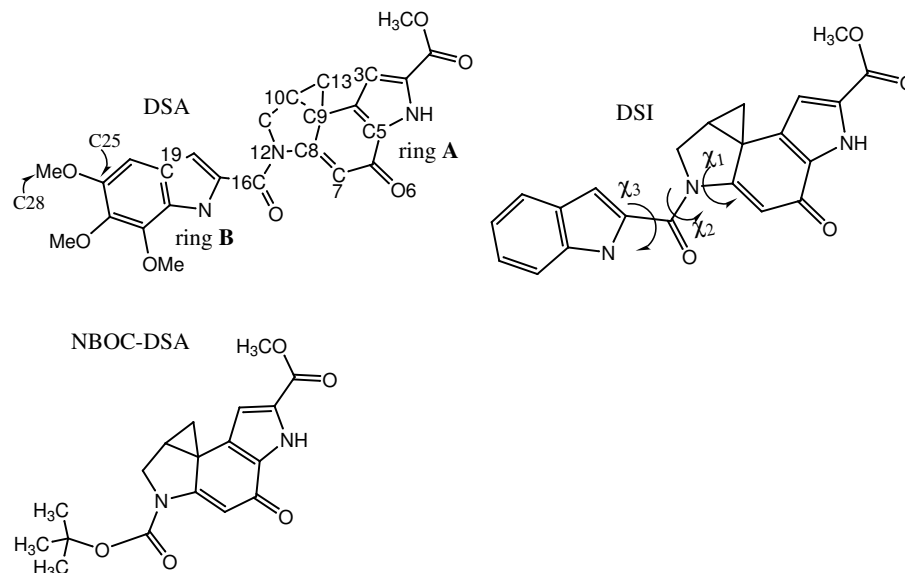
Figure 4.1: Duocarmycin anticner drugs investigated in this work. For **DSA** ring **A** and ring **B** are indicated and the numeration scheme is provided. For **DSI**, the torsional angles around the vinylogous amide bond are displayed

'A-tract'. The latter contains the A-A, T-T and A-T (but not the T-A steps) and exhibits a conformation other than that of canonical B-DNA [35], such as a narrow minor groove, highly negative propeller twists and a sizeable curvature of the double helix [1, 2, 136, 137, 138, 139, 140]. Several specific non-bonded interactions are formed (Table 4.7, Figure 4.2), which are believed to be responsible for the affinity and the sequence-selectivity of the drugs towards the A-tract with guanine as a flanking residue s[134]: (i) hydrophobic interactions between ring **B** and the sugar hydrogens in the minor groove wall. (ii) A hydrogen-bond between the C2-carbonyl group and the N2-amino group of G20 [141]. No water molecules are present between the drugs and DNA [134].

DNA acts as a catalyst for this reaction for most duocarmycin derivatives at physiological pH [142]. In particular **DSA** (Figure 4.1), which features an indole unit substituted in 5,6,7 positions by three methoxy groups (ring **B**), reacts five orders of magnitude faster with adenine in double stranded DNA than with adenine in aqueous solution. **DSI**, in which ring **B** is not substituted, is less reactive than **DSA**. The structurally related **NBOC-DSA** compound, in which ring **B** is replaced by a tert-butyl group (Figure 4.1), is 3-4 orders less reactive towards DNA compared to **DSA** [143]. Understanding the key factors governing duocarmycins' reactivity is clearly of great

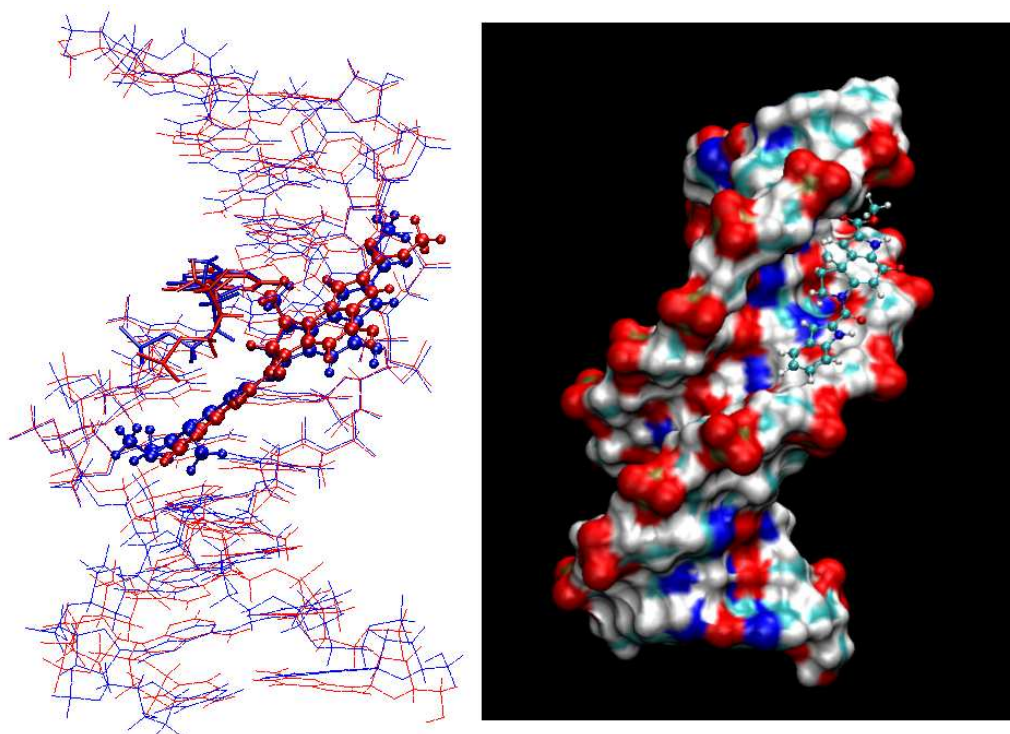Figure 4.2: Left: Superposition of **DSA-I** (blue) with **DSI-I** (red). The drug is shown in ball and stick whereas the reactive adenine nucleotide is shown as cylinders. Right: Solvent accessible surface of DNA double helix with electrostatic color code showing the snug fit of **DSI** into the minor groove.

pharmacological interest as it could provide the basis for more active drugs.

Several hypotheses regarding the catalytic effect of DNA on the reactivity of the drug have been put forward: (i) Hurley et al. have proposed that a *conformational change in the DNA double helix* may lead to a more reactive adenine-base [144] at the end of A-tracts. NMR studies reveal particular junction sites at the 3'-end and 5'-end of the A-tract [136, 137] and duocarmycins could specifically target these unusual junctions. Alternatively, duocarmycins by binding to DNA could also trap a particular, more reactive DNA conformation [145]. (ii) Boger et al have suggested that the higher reactivity of the cyclopropane ring in **DSA** relative to **DSI** is caused by its larger distortion observed in the NMR structures around the amid link (Table 4.1) [135, 146]. The distortion is measured by the torsional angles χ_1 (C7-C8-N12-C16) and χ_2 (C8-N12-C16-O16) (Figure 4.1) and is believed to decrease the π-electron conjugation and ultimately the stability of the cylcopropyl unit (*Shape induced activity*). Density functional calculations on a DSA-model (Figure 4.3 [147] and Table 4.2) in the gas phase appear not to support this proposal, although the effects of the solvent and of the macromolecule are expected to alter considerably these findings in vacuo. (iii) A general acid catalysis has also been invoked, in which the drugs are protonated on the O6 group (Figure 4.1) [148, 149]. This hypothesis is based on the presence of a highly acidic environment around the DNA double helix [150, 151].

Here we investigate binding and reaction mechanism by performing molecular simulations of the complexes between the drugs in Figure 1 and the oligonucleotide **I**. We start by looking at the mode of binding of the non-covalent complexes, for which no experimental information is available, by extensive classical MD simulations, in order to study the complex before the alkylation reaction. The drugs have been considered in their neutral form. We find that the duocarmycins investigated here induce only minor distortions in the DNA scaffold, which argues against a conformational change in DNA responsible for its catalytic activity. Calculations on the covalent adduct show excellent agreement between calculated and experimental structure, thus validating the force field used in our classical simulation.

We then turn on to study the DNA alkylation reaction of **DSA**, **DSI** and **NBOC-DSA** with (**I**), and for the smallest compound (**NBOC-DSA**) we also study the correspondent reaction in water (R4 in Figure 4.3), which does not occur at physiological pH, although it reacts at pH=3, $t_{1/2}$=177h[133].

The technique of choice for describing these chemical reactions is the hybrid Car-Parrinello QM/MM molecular dynamics simulation of the three drugs [13]. The ther-
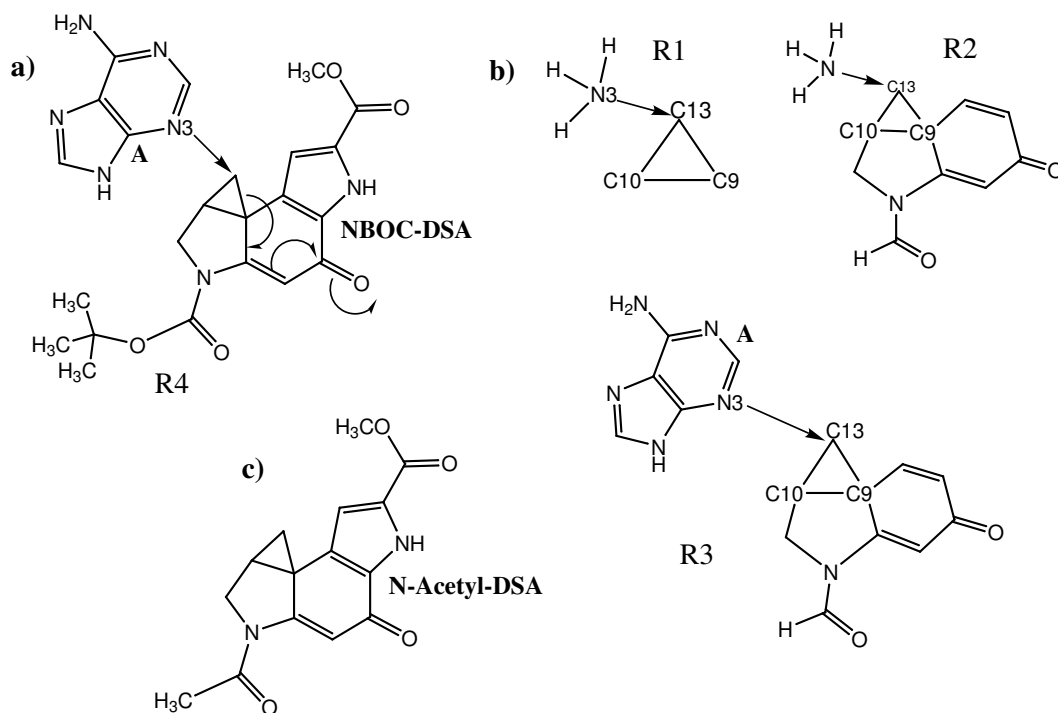
Figure 4.3: Reactions in aqueous solution investigated in this work. a) Reaction between **NBOC-DSA** and adenine (A) either isolated (R4) or embedded in **I**. b) Selected model reactions (R1-R3). The assumed reaction coordinate is shown by an arrow originating in N3 and ointing towards C13. In a) the electronic rearrangement upon alkylation is schematically shown. c) N-Acetyl-DSA used as DSA-model in ref [147] and here for test calculations in vacuo.

	DSA NMR	**DSI** NMR	cov**DSA-I**	cov**DSI-I**
χ_1	22.4(0.7)	14.2(0.6)	14.8(9.4)	15.3(8.7)
χ_2	11.0(0.6)	13.5(0.7)	13.4(10.9)	20.5(8.5)
χ_3	11.4(1.4)	9.7(1.0)	8.4(12.5)	4.7(11.1)
	DSA-I		**DSI-I**	**NBOC-DSA-I**
χ_1	20.9 (12.9)		20.2 (12.4)	11.7(11.6)
χ_2	14.4(12.1)		16.9(10.6)	-2.3(13.2)
χ_3	9.0(14.3)		6.6(14.4)	-

Table 4.1: Average dihedral angles from covalent drug-**I** and non-covalent drug-**I** MD simulations compared to NMR structures [134, 135].Standard deviations are given in parenthesis.

modynamic approach has been used to calculate the activation free energy. This approach has been shown to successfully describe a variety of chemical [152, 153] and biochemical reactions [15, 152]. Comparison is also made with reference reactions R1-R3 between three model compounds in water to get a detailed analysis of the influence of nucleophile and substituents on the reaction profile (Figure 4.3). The role of the mechanical and electrostatic fluctuations for the binding and for the reaction is further explored by classical MD simulations and Poisson-Boltzmann calculations. We find that the biomolecular frame polarizes the reactants significantly and that the nature of the condensed rings in the drugs affects largely their intrinisic reactivity.

4.2   Computational Methods

4.2.1   Structural Models for the classical MD

The starting structures of complexes between the oligonucleotide d(pGpApCpTpAp-ApTpTpGpApC)-d(pGpTpCpApApTpTpA*pGpTpC)(**I**) and **DSA** and **DSI** are based on the NMR structures of the covalent complexes cov**DSA-I** and cov**DSI-I**, respectively (1DSA [134] and 1DSI [135]). The non-covalent **DSA-I** and **DSI-I** are constructed by removing the covalent bond, closing the cyclopropyl ring and deprotonating the C6-O6 group. Close contacts are removed manually. For the non-covalent **NBOC-DSA-I** complex, the drug was fitted on **DSA-I** model and close contacts were again removed manually. 20 sodium counter-ions are added and the systems are solvated by 4722(cov**DSA-I**), 4809(cov**DSI-I**), 4946(**DSA-I**), 4775(**DSI-I**) and 5013(**NBOC-DSA-I**) water molecules. The free DNA 11-mer is constructed with the

nucgen module of the AMBER program package with the same sequence as in the drug-**I** adduct: 5'-d(GpApCpTpApApTpTpGpApC)-3'- 5'(d(GpTpCpApApTpTpAp*GpTpC)-3').

20 counter ions are added and the system is solvated with 3903 water molecules.

The following model systems are constructed: (Figure 4.3):

R1: Complex between cyclopropane and ammonia;

R2: Complex between cyclopropyl-indole and ammonia;

R3: Cyclopropyl-indole and adenine;

R4: Complex between **NBOC-DSA** and adenine.

The four systems are solvated by 853, 1001, 1072 and 2109 water molecules in R1, R2, R3 and R4, respectively, and equilibrated by MD such as to obtain densities of $1.0 \, \mathrm{gcm^{-3}}$. The number of water molecules in the first solvation shell is calculated by integration of the g(r)'s.

The AMBER-parm98 force field [41, 42] is used for the oligonucleotide along with the TIP3P [119] model for water and the Aaqvist parametrization of sodium for the counter ions [154]. In the non-covalent complexes, the parameters of **DSA**, **DSI** and **NBOC-DSA** are adopted from ref. [134], except for the charges, which are derived from ESP charges (RESP) as described in the AMBER program package [118], namely by optimizing the structures at the B3LYP-6-31G* level of theory. For the tert- butyl group of **NBOC-DSA** typical AMBER parameters for the methyl groups are adopted. In the covalent complexes, the charges are recalculated both for the drugs and the attached nucleobase (A19) by the same scheme as described above. Those of ammonia and the smaller duocarmycin model of R2 and R3, are taken from the gaff-database [155], except for the charges, which are calculated as above.

4.2.2 Classical MD calculations

The covalent and non-covalent drug-**I** adducts are minimized by three runs of steepest descent (5000 steps) using the AMBER7 program [118], involving (i) the water molecules, (ii) the water and the counter ions, and (iii) the entire systems. The systems undergo then MD simulation. They are first slowly heated up to 300 K at constant volume, keeping the drug-**I** adduct frozen. At 300 K, constant pressure, constant temperature (NPT) is switched on. A harmonic position constraint on the drug and DNA is steadily decreased starting from a force constant of 2000 kJ/nm. Periodic boundary conditions are applied and the electrostatic interactions are calculated with the particle-mesh Ewald method (PME) [47, 48]. Three separate Nose-Hover thermostats are applied to the DNA, the drug and the solvent, which assure an

even temperature distribution throughout the entire system. Constant pressure at 1 atm is controlled through extended-ensemble pressure coupling implemented by the Parrinello-Rahman method [156]. Bond lengths are constrained using the lincs algorithm and the translational and rotational motion of the center of mass of the drug-**I** adduct is removed every 25 steps of molecular dynamics. A timestep of 0.001ps is used throughout all the classical molecular dynamics simulations described here. 10 ns for cov**DSA-I**, cov**DSI-I**, **DSA-I** , **DSI-I** and **NBOC-DSA-I** and 6 ns for the **I** are collected and analyzed by the GROMACS program package[157, 158]. All the simulations are performed on a Xeon 3.06 GHz dual processors PC.

4.2.3   QM/MM Molecular dynamics

The alkylation reaction is studied by mixed Car-Parrinello-classical molecular dynamics simulation using the scheme developed by Rothlisberger and co-workers[13]. In this approach, the systems are divided into two regions. The first region (QM) is treated within the framework of DFT, as implemented in the program CPMD [65]. Planewaves up to energy cutoff of 70 Ry are used along with pseudopotentials of the Martins-Trouillers type [120]. The second region (MM) contains the rest of the system and is treated with effective potentials described above [41, 42]. The QM-parts include the drugs and the reactive adenine base (A19)(Figure 4.4). The latter is cut at the C1' carbon and the valence shell of this last QM-atom is saturated by two additional hydrogen atoms (capping hydrogens) [159]. The QM-parts contain 76(**DSA-I**), 64(**DSI-I**) and 63(**NBOC-DSA-I**) atoms and are placed in an orthorhombic box with box lengths of $23.8 \times 14.3 \times 21.2 \text{Å}^3$ (**DSA-I**), $23.8 \times 20.0 \times 15.2$ Å^3 (**DSI-I**) and $18.0 \times 14.0 \times 18.9$ Å^3 (**NBOC-DSA-I**) respectively. The overall charge of QM-parts is 0. As a starting point we choose a snapshot of the classical trajectory, in which the drug-adenine distance is about 3.3Å, and the counter ions are not closer than 5Åto either the adenine A19 or ring **A** of the drug.

The systems are quenched in the beginning and then relaxed at 0K. This procedure is repeated several times for the first step along our reaction coordinate, and once for the subsequent steps. The systems are then heated up to 300 K. At 300K, a Nose thermostat is applied to keep the system at constant temperature [44]. A time-step of 5au (0.12fs) and a fictitious electron mass of 600 au is used. The BLYP functional is used for all calculations, which has proven to give accurate results for geometries and energetic features [52, 53]). The QM-part is treated as an isolated system and electrostatic interactions between periodic images are decoupled by the scheme of Hockney [122]. The additional hydrogen atoms are excluded from the classical
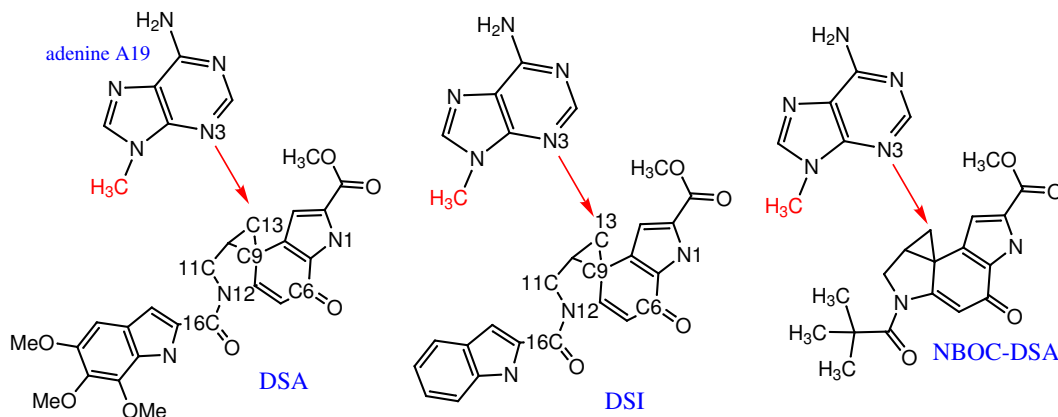
Figure 4.4: QM-region for the three QM/MM calculations: (left): **DSA** with DNA, (middle): **DSI** with DNA, (right): **NBOC-DSA** with DNA. In red: last QM atom C1', which is saturated with capping hydrogens

Hamiltonian. The electrostatic interactions between the QM part and the MM part are treated in a hierarchical scheme as described elsewhere [15, 13]. Thermodynamic integration techniques [160, 161] are used to calculate the activation free energies of the reactions. We choose the N3(A19)-C13(drug) distance to be our reaction coordinate. At each step along this reaction coordinate, the system is simulated for 1.5ps, where 0.5ps is discarded, whereas the final 1.0ps are used for analysis. The force constant on the distance constraint is monitored and the transition state is localized upon change of the sign of the force constant. This occurs at a distance of 2.3 Å in the drug-**I** adducts. The reaction coordinate is sampled at the following distances: 3.3 Å, 3.0 Å, 2.8 Å, 2.6 Å, 2.4 Å, 2.2 Å and 2.2 Å. In the model systems the location of the transition state varies from 2.1 Å for (R1) to 2.6 Å for (R2,R3). The free energy of activation has been calculated by integrating the force constant, f(r) along the reaction coordinate:

$$\Delta F^{\#} = \int \langle f(r) \rangle dr \tag{4.1}$$

The free energy of activation can be related to k_{cat} through transition state theory.

$$k_{cat}(T) = \frac{k_B T}{hc^0} e^{-\Delta G^{\#}/RT} \tag{4.2}$$

$\Delta G^{\#} \sim \Delta F^{\#}$ [162] and under saturation condition the experimental rate, given by:

$$k_{obs} = \frac{k_{cat}[DNA]}{K_M + [DNA]} \tag{4.3}$$

where K_M is the Michaelis constant, simplifies to $k_{obs} \sim k_{cat}$

4.2.4 QM/MM analysis

Boys Orbital Centers (BC) of the fully localized orbitals [62] are calculated on the fly during the QM/MM simulations. BC's are useful to visualize lone pairs and chemical bonds. Here, the polarization index (PI_{AB}) of a chemical bond AB is defined as:

$$BC = \frac{\overrightarrow{A - BC} \cdot \cos \alpha}{\overrightarrow{A - B}} \tag{4.4}$$

where $\cos \alpha$ is the angle spanned between the two vectors $\overrightarrow{A - BC}$ and $\overrightarrow{A - B}$.

PI lower than 0.5 indicates a polarization towards A. A change at the second decimal can be considered significant. D-RESP charges for selected atoms are also calculated on the fly as described in reference [14]. In particular, we look at the D-RESP of the nucleophile-nitrogen. Notice that other D-RESP charges are also calculated, namely the ones on C9, C13 and O6, but they show no clear trend and are not discussed further.

4.2.5 QM calculations of model system in vacuo

We extend the study of the **DSA**-model (Figure 4.3) in vacuo presented in reference [147] to some larger values of χ_1 and χ_2. To this aim we optimize the geometry by quenching the structure while keeping one of the torsional angles fixed. The model system is placed in a box with box lengths of $18 \times 12.96 \times 11.52 \text{Å}^3$ and the electrostatic interactions between periodic images are decoupled by the scheme of Martyna and Tuckerman [163]. All the calculations are done with the CPMD program [65] as described above.

In agreement with reference [147] we do not find significant changes in electronic and structural properties in the cyclopropyl ring upon changes in the torsional angles.

(A) [°]	1	2	3	4	5	6	7	8
χ_1	7.0	7.0	7.0	7.0	0.0	10.0	20.0	30.0
χ_2	5.0	10.0	20.0	30.0	5.0	5.0	5.0	5.0

(B)	Δ Bond length [Å]	Δ PI
C9-C10	0.007	0.004
C9-C13	0.005	0.001
C10-C13	0.003	0.001
N12-C16	0.007	0.006
N12-C8	0.004	0.004
C8-C7	0.003	0.015

Table 4.2: QM-calculations on N-acetyl-**DSA** in vacuo (Figure 4.3 c). (A) Conformations for N-acetyl-**DSA** [°]. (B) Maximal changes observed in bond lengths [Å] and PI_{AB}'s for optimized structures.

4.2.6   Poisson-Boltzman calculations

The non-linear Poisson-Boltzman equation is solved iteratively with the program DelPhi[164, 165, 166], to calculated the field produced by the DNA scaffold onto the drugs. The charges of the drug itself are therefore excluded. The outer dielectric constant of the solvent is set to 80 and the one of DNA to 2. An implicit ionic strength of 0.15M is added. The average potential is calculated for at least 100 snapshots of the QM/MM trajectory or classic trajectory.

4.3   Results

Here, we describe the MD structures of the non-covalent complexes between the oligomer **I** and **DSA**, **DSI** or **NBOC-DSA**. The structures of the DNA scaffold are compared with that obtained with an MD simulation of **I** in the free state. Our computational setup for the MD simulations is established by reference calculations on the covalent **DSA-I** and **DSI-I** adducts, for which the 3D structure has been determined. Next, we discuss the alkylation reaction of **DSA**, **DSI** and **NBOC-DSA** with DNA based on our constrained QM/MM simulations. Comparison is also made for the correspondent reaction with adenine in water for the smallest drug, **NBOC-DSA**. Finally, by comparing our results with QM/MM simulations of models of **NBOC-DSA**/adenine complexes in water we gain insights on the role of the substituted pyrrole ring and of the adenine nucleophile for the reaction.

4.3.1   Classical MD Simulations of the DNA/drug complexes

We focus on aqueous solutions of the complexes between the drugs and I in which the drugs are either non-covalently bound (**DSA-I**, **DSI-I** and **NBOC-DSA-I**, respectively) or covalently bound (cov**DSA- I**, cov**DSI-I**). The systems are built based on the NMR structures of the covalent complexes (Figure 4.2). Comparison is also made with the structure of free 11-mer based on the canonical B-DNA conformation in water.

Validation of the computational setup

Because the force field for the drug is non-standard (see Methods), we first perform MD simulations on the complexes for which structural information is available, namely cov**DSA-I** and cov**DSI-I** and compare our results against experimental data. These calculations are also used to obtain insights on the conformational fluctuations of these adducts.

The structures exhibit only small rearrangements from the initial NMR structures. The rmsd's of the DNA moiety fluctuates around 1.5-1.7Å after 0.5 ns (Table 4.3), whereas those of the drugs fluctuate around 1.8Å. The structural determinants of the simulated and experimental structures are similar: (i) The minor groove width is similar and smaller than that of canonical B-DNA [134, 135] (Figure 4.5); (ii) the propeller twists are similar and largely negative at the A-tract (Figure 4.5); (iii) the overall DNA curvature ranges from $15°$ to $17°$ degrees and is mostly caused by local axis bends at the end of the A-tract. The largest local axis bend for **DSI-I** is found at the T4-A5 step for both MD and NMR structures, whereas that of **DSA-I** is located at the T4-A5 step in the MD structure and at the A5-A6 step in the NMR structure (iv); The twist angles at the T4-A5 step are similar and lower than those of canonical B-DNA (Figure 4.6); (v) The hydrophobic and H-bond contacts are similar, although in the NMR- structure, the contact distances are in general shorter (by 0.2 Å, Table 4.7). This may be the reason for the difference seen for the tilt angle at the alkylation site (Figure 4.6). The torsional angles around C7-C8- N12-C16 (χ_1) and around C8-N12-C16-O16 (χ_2), which measure the degree of distortion of the amide link (Figure 4.1), are fairly similar with the NMR values. However, the rather small difference seen in χ_1 when passing from **DSA** to **DSI**, present in the NMR structure, is not observed in the MD simulations (Table 4.1).

	DNA	Drug	Correlation
covDSA-DNA	1.7(0.3)	1.8 (0.5)	0.69
covDSI-DNA	1.5(0.2)	1.1(0.3)	0.39
Free DNA **I**	2.4(0.5)	-	-
DSA-DNA	1.9(0.4)	2.1(0.6)	0.61
DSI-DNA	2.0(0.5)	2.9(0.8)	0.75
NBOC-DSA-DNA	2.3(0.5)	5.0(1.0)	0.33

Table 4.3: Root mean square deviations for DNA and Drug in [Å], standard deviations in parenthesis, correlation between the two sets of data are also given

Figure 4.5: Propeller Twist (left) and minor groove width (right) for non-covalent MD complexes (upper panel) and covalent MD complexes and NMR-structures (lower panel). Parameters of the MD-structure of **I** are reported in all graphs for comparison. Values are averaged over the entire classical trajectory.

Figure 4.6: Tilt angle (left) and helix twist (right) for non-covalent MD complexes (upper panel) and covalent MD complexes and NMR-structures (lower panel). Parameters of the MD-structure of **I** are reported in all graphs for comparison.

MD Simulation of the free oligonucleotide (I)

It is well-known that the conformation of A-tracts- containing oligonucleotides in solution differ from that of B-DNA [167, 136, 2, 137]. Indeed, the rmsd between the final MD structure of (**I** and the initial B-DNA structure is as large as 2.5Å, pointing to the differences between the two conformations. In particular for the A-tract, the MD structure exhibits: (i) A narrower minor groove, with the narrowest value at the A6 site (4.5 Å vs 5.7 Å for standard B- DNA [35]). (ii) more negative propeller twists in the A-tract (-20° vs. -11° in B-DNA). (iii) Larger local axis bends and roll angles at the end of the A-tract (Figure 4.7). In B- DNA, both local axis bend and roll angle are ideally 0° [35]. The backbone torsional angles are instead similar to those in canonical B-DNA [35](data not shown).
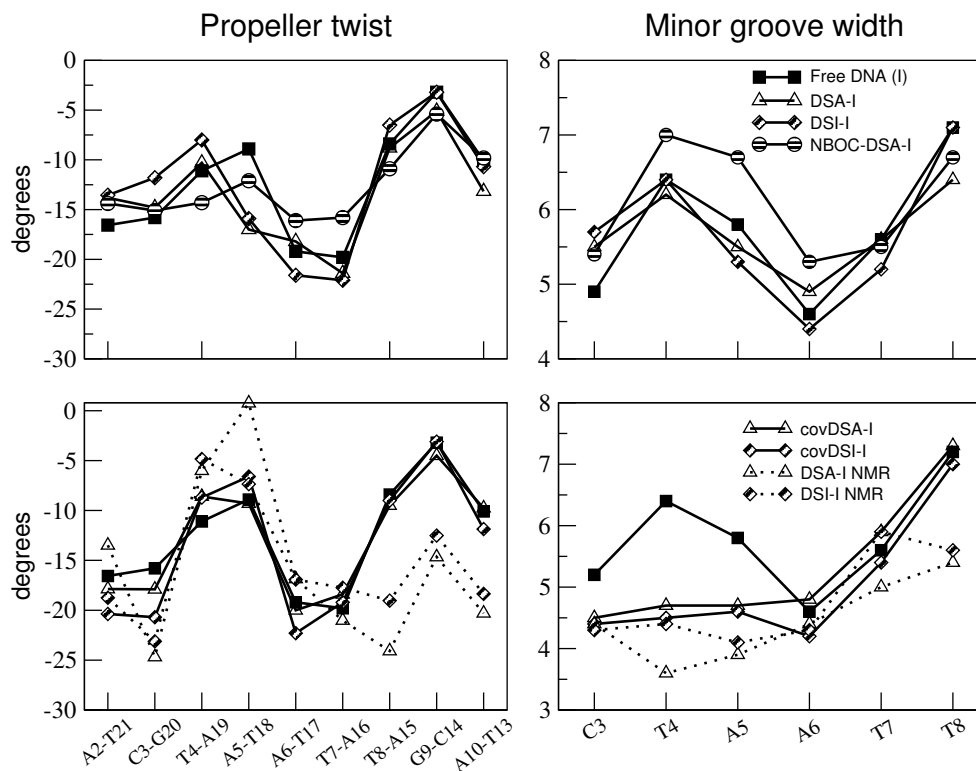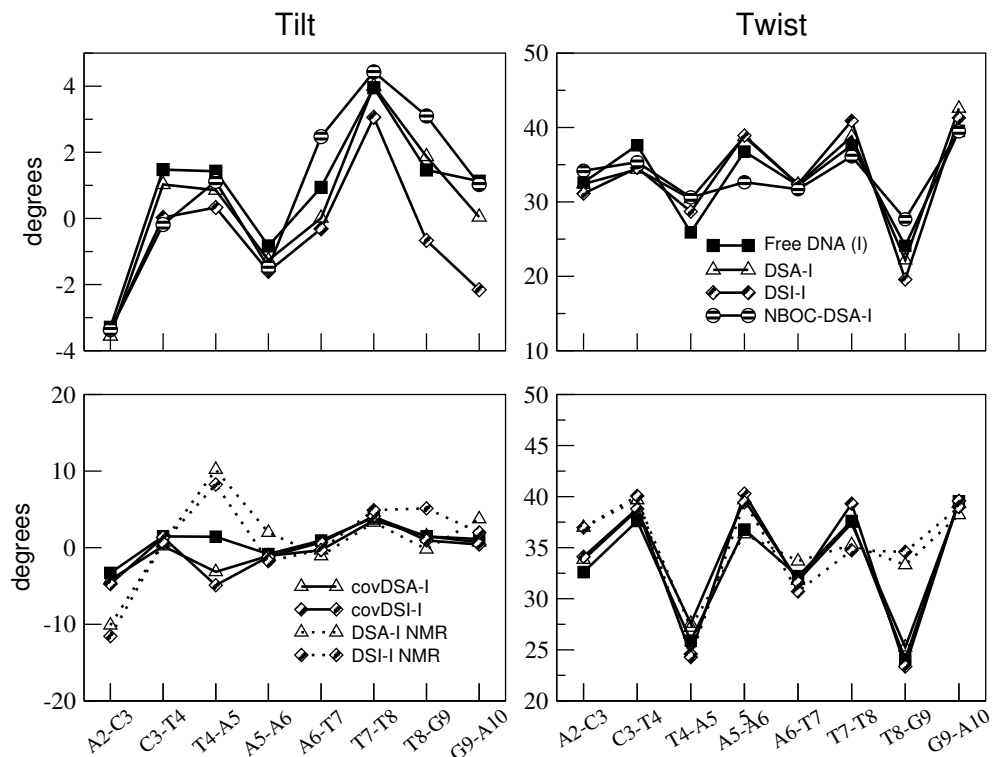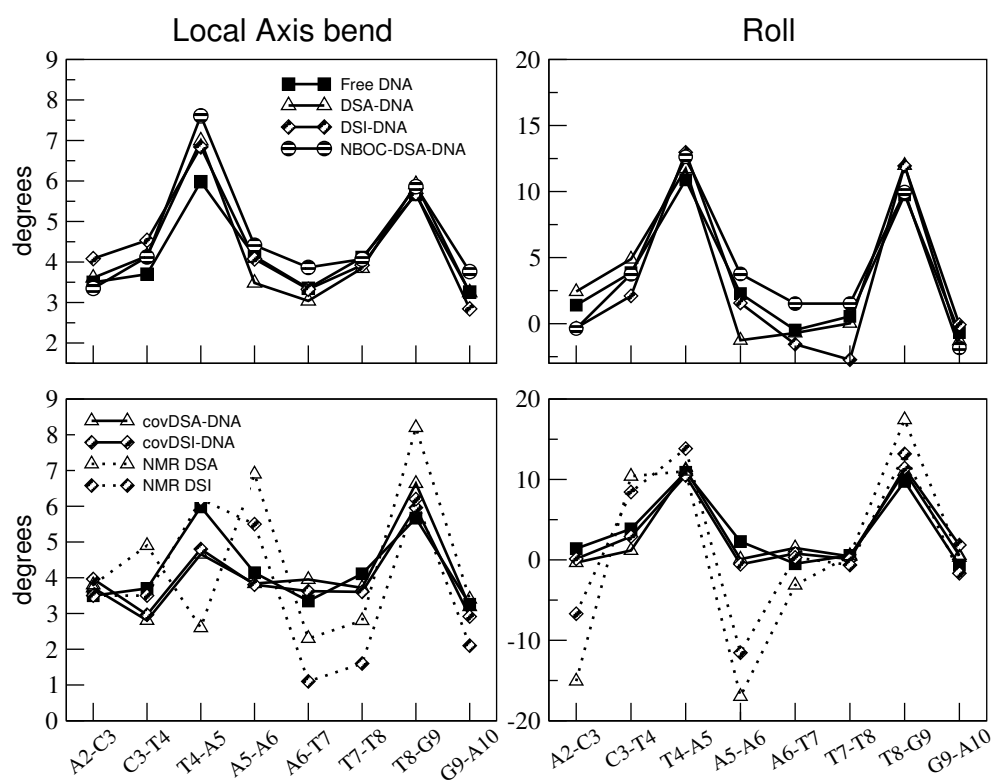
Figure 4.7: Local axis bend (left) and roll angle (right) for non-covalent MD complexes (upper panel) and covalent MD complexes and NMR-structures (lower panel). Parameters of the MD-structure of **I** are reported in all graphs for comparison.

MD Simulations of the Non-covalent Drug-I Complexes

As expected, the non-covalent **DSA-I** and **DSI-I** adducts, constructed from the covalent NMR structures, exhibit slightly larger rmsd's from the initial models than those of the covalent complexes (Table 4.3). The initial MD structure of **NBOC-DSA- I**, for which no experimental structure is available, is derived from the covalent **DSA-I** complex. The tert-butyl unit, which replaces the indole unit of **DSA**, is bulky and penetrates less into the minor groove. Therefore, it is not surprising that the rmsd of **NBOC-DSA-I** is higher than that of **DSA-I** or **DSI-I** (Table 4.3). The mutual rmsd's between the non-covalent drug-**I** complexes and **I** are small, ranging from 0.5 to 0.7 Å. Consistently, all these structures exhibit: (i) A narrow minor groove with the narrowest value at the A6 site (Figure 4.5), (ii) high negative propeller twists in the A-tract (Figure 4.5), (iii) increased local axis bend and roll angles at the ends of the A-tract (Figure 4.7). (iv) A low helix twist at the T4-A5 step (Figure 4.6). The latter is also observed in the covalent drug-**I** complexes, thus this seems to be an inherent feature of the sequence studied here and not a distortion due to the drug as proposed in ref [134]. The overall DNA-curvature ranges from 18° to 29°, similar to that of **I** and higher than that of the covalent drug-**I** complexes (Table 4.4, 4.5).

The local axis bends at the T4-A5 step are slightly larger than in **I** (by 1°-2°, Figure 4.8) and exhibit larger fluctuations (Table 4.4, 4.5). The roll angles are positive at the end of the A-tracts (T4-A5, T8-G9) and account for most of the local axis bends. All other base steps are characterized by a roll angle close to 0°. The propeller twists are larger at the A5-T18 base pair than in **I** (Figure 4.6) in all three drug-**I** complexes. We notice that the fluctuations of these quantities at the drug-binding site are slightly larger compared to the ones of **I**, and, much larger than those of the covalent complexes (Table 4.4,4.5). The minor groove widths of **DSA-I** and **DSI-I** are very similar to **I** and wider at the T4 and A5 step with respect to the covalent drug-**I** adducts (Figure 4.5). On the other hand, the minor groove of **NBOC-DSA-I** is wider at the alkylation site and at the two subsequent steps than those of **DSA-I**, **DSI-I** and **I**. We conclude that the overall oligonucleotide structure is only slightly perturbed by the presence of the drugs as already pointed out in ref. [134], but most by **NBOC-DSA**.

We now investigate the mode of binding of the three drugs. The rmsd's, calculated by fitting only the DNA scaffold, suggest that **NBOC-DSA** rearranges significantly in the minor groove, whereas **DSA** and **DSI** remain close to their initial positions (Table 4.3). For **DSA-I** and **DSI-I**, the distance between the reactants (atoms C13(drug) and N3(A19), d hereafter) fluctuates between 3Å and 4Å, leading to a reactive complex

local axis bend	**DSA-I**	**DSI-I**	**NBOC-DSA-I**
A2-C3	3.6 (1.7)	4.1 (2.0)	3.5 (1.7)
C3-T4	4.1 (2.1)	4.5 (2.5)	4.1 (2.2)
T4-A5	7.0 (3.4)	6.8 (3.8)	7.6 (3.6)
A5-A6	3.4 (1.8)	4.1 (2.4)	4.4 (2.6)
A6-T7	3.0 (1.6)	3.3 (1.7)	3.9 (2.2)
T7-T8	3.8 (1.9)	3.3 (1.9)	4.1 (2.1)
T8-G9	5.9 (2.4)	5.7 (2.5)	5.9 (2.7)
G9-A10	3.2 (1.8)	2.8 (1.6)	3.8 (2.0)
roll angle			
A2-C3	2.4 (4.1)	-0.41 (5.3)	3.5 (1.7)
C3-T4	4.9 (5.2)	2.4 (6.9)	4.1 (2.2)
T4-A5	11.7 (7.6)	13.0 (7.3)	7.6 (3.6)
A5-A6	-1.2 (5.8)	1.6 (6.0)	4.4 (2.6)
A6-T7	-0.7 (4.2)	-1.6 (3.9)	3.9 (2.2)
T7-T8	0.0 (4.8)	-2.7 (4.7)	4.1 (2.1)
T8-G9	12.0 (5.4)	11.9 (5.2)	5.9 (2.7)
G9-A10	-1.3 (4.7)	-0.1 (5.2)	3.8 (2.0)
global curvature			
	17.7 (7.7)	15.3 (8.7)	29.4 (10.1)

Table 4.4: Local axis bend, roll angle and global curvature for non-covalent drug-DNA adducts with standard parameters in parantheses

local axis bend	covDSA-I	cov-DSI-I	Free DNA (I)
A2-C3	3.7 (1.6)	4.0 (1.6)	3.5 (1.6)
C3-T4	2.8 (1.7)	2.9 (1.6)	3.7 (2.1)
T4-A5	4.8 (2.4)	4.8 (2.1)	6.0 (2.8)
A5-A6	3.9 (2.0)	3.8 (1.8)	4.1 (2.2)
A6-T7	4.0 (2.1)	3.6 (1.1)	3.3 (1.8)
T7-T8	3.8 (2.0)	3.6 (1.9)	4.1 (2.0)
T8-G9	6.6 (3.0)	6.0 (2.4)	5.7 (2.5)
G9-A10	3.2 (2.0)	2.9 (1.7)	3.3 (1.7)
roll angle			
A2-C3	-0.3 (4.6)	0.1 (4.0)	3.5 (1.6)
C3-T4	1.2 (4.8)	2.9 (4.4)	3.7 (2.1)
T4-A5	11.2 (6.2)	10.5 (6.0)	6.0 (2.8)
A5-A6	0.1 (5.6)	-0.6 (5.4)	4.1 (2.3)
A6-T7	1.5 (4.2)	0.8 (4.1)	3.4 (1.8)
T7-T8	0.4 (4.6)	0.1 (4.5)	4.1 (2.0)
T8-G9	11.0 (6.7)	11.4 (5.7)	5.7 (2.5)
G9-A10	0.5 (6.1)	1.8 (5.2)	3.3 (1.7)
global curvature			
	16.8 (7.6)	14.7 (6.9)	24.4 (9.6)

Table 4.5: Local axis bend, roll angle and global curvature of covalent drug-DNA adducts and Free DNA (I) with standard deviations in parantheses

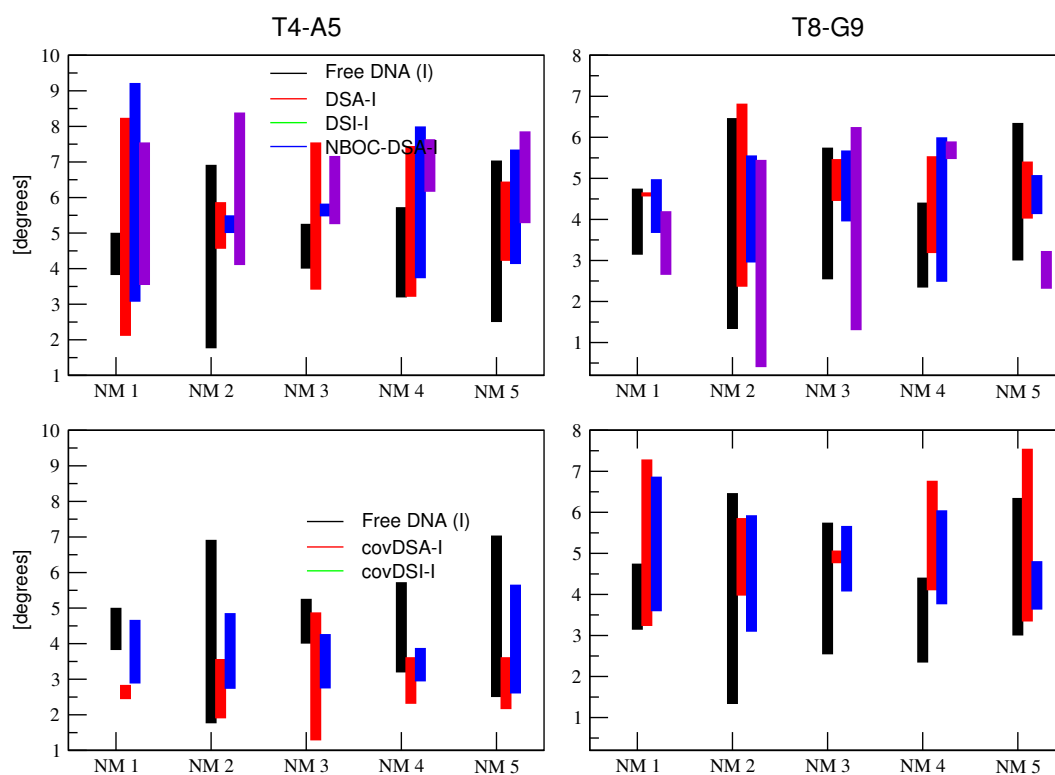Figure 4.8: Variation in local axis bend at the T4-A5 (left) and T8-G9 (right) step along the first five normal modes. Upper panel: non-covalent adducts compared to **I**, lower panel: non-covalent adducts compared to **I**
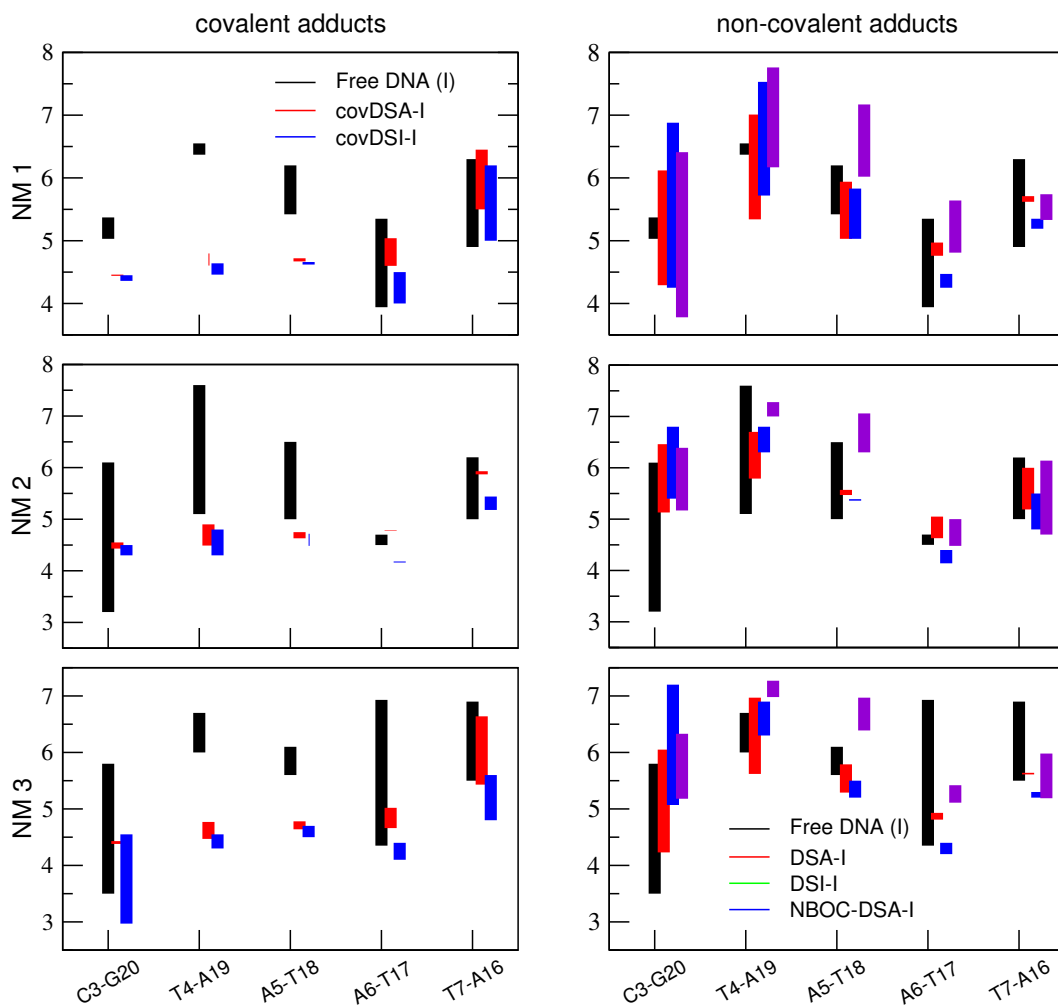
Figure 4.9: Variation in minor groove width along the first three normal modes. To the left: covalent adducts compared to **I**, to the right: non-covalent adducts compared to **I**

	DSA-I	**DSI-I**	**NBOC-DSA-I**
d(C13-N3(A19))	3.6 (0.5)	3.5 (0.4)	4.8 (0.9)
d(C13-N3(G20))	6.3 (0.5)	5.7 (1.1)	4.1 (0.6)

Table 4.6: C13-N3 distance for non-covalent drug-**I** MD structures. Both distance to A19 and G20 are given, standard deviations in parantheses.

between the drug and DNA (Table 4.3, Figure 4.10). Instead, the corresponding distance in **NBOC-DSA-I** fluctuates between 3 Å and 9Å, and d(C13(drug)-N3(G20)) oscillates between 3Å and 6.0Å (Table 4.6, Figure 4.10). Thus, this guanine nucleobase could be an alternative alkylation site for **NBOC-DSA**. So far, alkylation at guanine sites has only been observed for a structurally related compound [168] and not for **NBOC-DSA**.

Ring **B** of **DSA** and **DSI** forms basically the same hydrophobic interactions as the covalent complexes, but is slightly shifted to the 5'-end of the modified strand (Table 4.7, Figure 4.11). The C28 methoxy-group of **DSA** forms additional contacts with the minor groove. **NBOC-DSA**, in which ring **B** is substituted by a tert-butyl group, forms less hydrophobic contacts with **I**. The three methyl groups are equivalent and in contact with the sugar rings of A19 and A6 (Table 4.7). Ring **A** of **DSA** and **DSI** is located further away from the alkylation site and forms less hydrophobic contacts to the minor groove than in the correspondent covalent adducts. The C2-carbomethoxy-N2(G20) H-bond [141] is mediated by one water molecule. Ring **A** of **NBOC-DSA** forms tighter hydrophobic contacts than the other two drugs and a direct H-bond for most of the simulated time.

Comparison among all the models

In this section, we compare selected features of the systems investigated by MD simulations. In the drug-**I** complexes, the hydrophobic contacts between ring **B** and the minor groove are similar; those between ring **A** and DNA are shorter in the covalent complexes. A superposition of covalent and non-covalent complex is presented in Figure 4.11. The only hydrogen bond present between G20 and the methyl-carboxy-group is direct in the covalent complexes whereas it is mediated by one water molecule in the non-covalent complexes. The two carbonyl groups are solvent exposed in all systems (Figure 4.2).

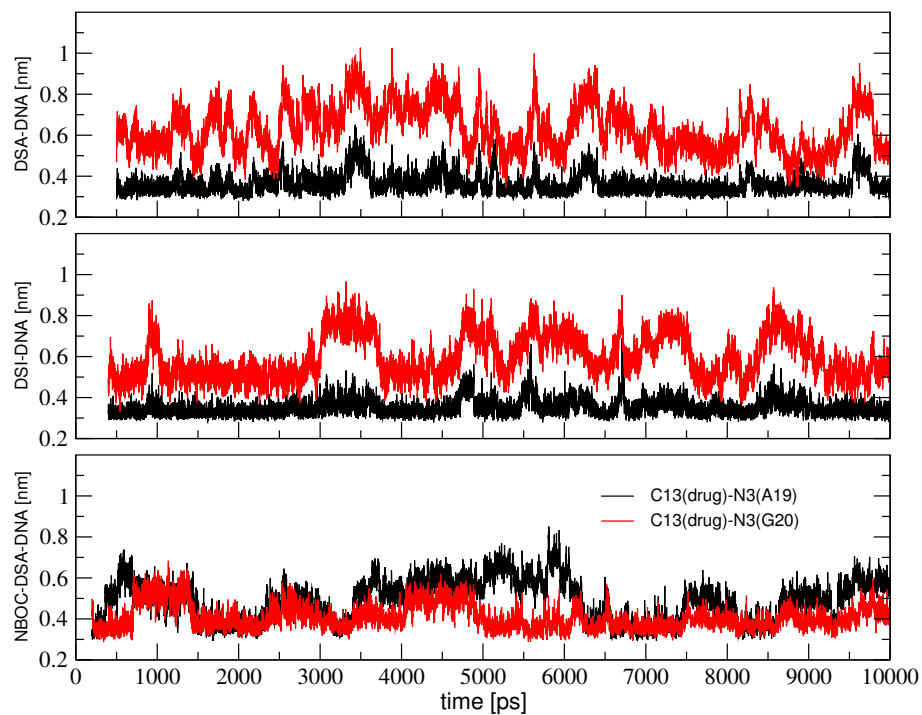The χ_1 and χ_2 torsion angles of **DSA** and **DSI** are very similar and larger than

Figure 4.10: N3-C13 distance during classical MD of non-covalent drug-DNA complexes. No constraints are applied. N3(A19)-C13 distance is shown in black, N3(G20)-C13 distance is shown in red.

	DSA NMR	DSI NMR	cov**DSA-DNA**	cov**DSI-I**
C25-T18	3.78 (0.28)	3.67 (0.03)	3.99 (0.33)	3.86 (0.26)
C25-T7	3.83 (0.35)	3.73 (0.06)	4.02 (0.28)	3.69 (0.23)
C28-T8	3.89 (0.24)	-	4.48 (0.28)	-
C28-T17	4.01 (0.36)	-	5.10 (0.64)	-
C28-T18	3.91 (0.27)	-	4.15 (0.37)	-
C19-T18	4.23 (0.35)	4.18 (0.04)	4.75 (0.64)	4.11 (0.33)
C19-T7	3.67 (0.26)	3.97 (0.08)	4.05 (0.31)	4.04 (0.30)
C3-A5	4.66 (0.34)	4.79 (0.01)	4.35 (0.42)	4.67 (0.36)
C3-G20	4.00 (0.32)	3.89 (0.03)	4.12 (0.24)	4.13 (0.26)
C14-T21	3.76 (0.38)	3.76 (0.03)	3.90 (0.27)	3.90 (0.36)
H-bond O14	4.10 (1.00)	3.30 (0.81)	5.22 (0.28)	3.94 (1.10)
H-bond O15	3.63 (1.11)	4.59 (0.86)	3.21 (0.32)	5.89 (1.10)

	DSA-I	**DSI-I**	**NBOC-DSA-I**	
C25-T18	3.92 (0.28)	3.87 (0.28)	C19-A6	4.94 (0.87)
C25-T7	4.12 (0.41)	3.91 (0.33)	C19-A19	5.98 (1.73)
C28-T8	4.33 (0.50)	-	C20-A6	5.15 (0.84)
C28-T17	4.20 (0.49)	-	C20-A19	6.60 (1.69)
C28-T18	4.58 (0.61)	-	C21-A6	5.48 (0.82)
C19-T18	3.99 (0.39)	4.10 (0.32)	C21-A19	5.86 (1.41)
C19-T7	3.90 (0.30)	3.91 (0.33)		
C3-A5	5.30 (0.71)	4.80 (0.59)	4.3 (0.70)	
C3-G20	5.51 (0.84)	5.55 (0.74)	5.91 (0.67)	
C14-T21	6.80 (1.52)	6.83 (1.62)	5.21 (1.1)	
H-bond O14	9.60 (1.10)	7.29 (1.00)	6.34 (1.3)	
H-bond O15	7.71 (1.00)	9.04 (1.1)	4.31 (1.34)	

Table 4.7: Selected hydrophobic contacts and H-bond distances between drugs and **I** measured as the minimal distance between drug-carbon atom and C1', C2' or C5' of furanose ring. Standard deviations are given in parantheses.
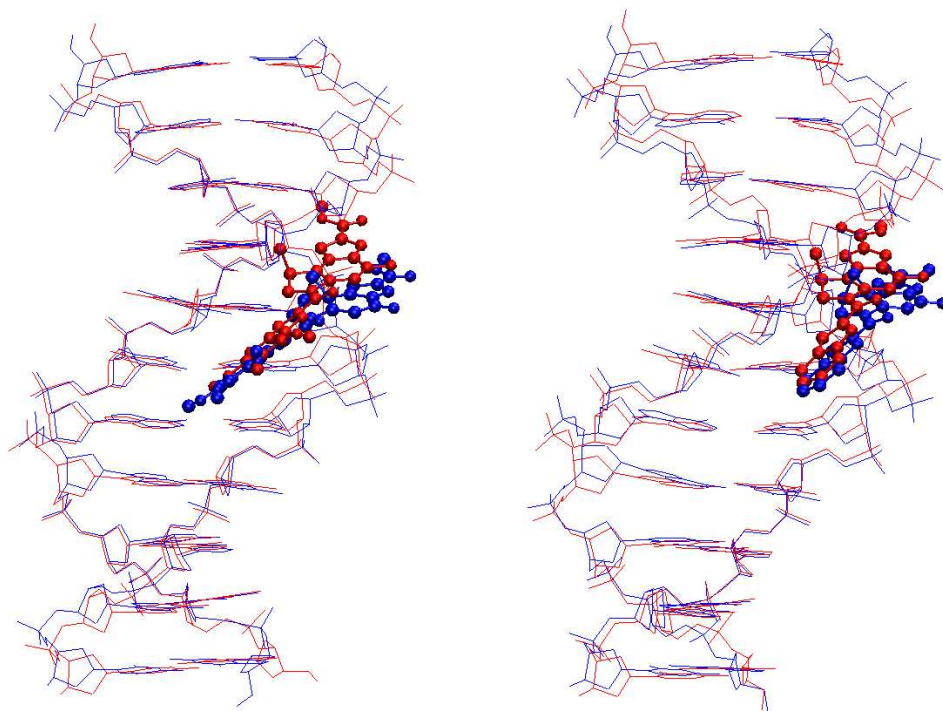
Figure 4.11: Superposition of covalent (red) and non-covalent (blue) drug-**I** adducts. Left: **DSA-I**, Right: **DSI-I**

those of **NBOC-DSA** (Table 4.1). Furthermore, they are larger and fluctuate more than those of the corresponding covalent drug-**I** complexes (Table 4.1). Thus, **DSI** and **DSA** are more distorted and flexible in the free state than when bound to DNA.

We observe that spatial fluctuations of the non-covalent complexes are slightly larger than those of the free nucleotide, and, much larger than those of the covalent complexes. These findings are substantiated by an analysis of the large-scale motions. In fact, the latter are characterized by changes in the minor groove width and the DNA curvature. The minor grooves open up and close again. The widths and the curvature of the non-covalent complexes are fairly similar to those of **I** and they vary by as much as 3Å and 15Å, respectively (Figure 4.9). This further confirms that the drug does not affect very much the structure and dynamics of the DNA scaffold. In the covalent adducts, the variation in minor groove width is much smaller (Δ width: 0.2-0.5 Å), whereas DNA curvature still varies considerably (Δ curvature: 14°). The latter is however due to enhanced fluctuation of the local axis bend at the T8-G9 step, which is outside the drug-binding site (Figure 4.8). Thus, the formation of the covalent bonds leads to a stiffening of the drug-binding site (Figure 4.8,4.9, Table 4.44.5).

We close this section by investigating the hydration of the drug-binding site. The minor groove of the MD structure of **I** is fully hydrated and the water molecules form the typical spine of hydration (Figure 4.12)[3, 169]. The presence of the drug leads to removal of several ordered waters in the minor groove with respect to **I** (Figure 4.12). In **DSA-I** and **DSI-I**, A6 and T18 are fully dehydrated and A19 and A5 are partially dehydrated as seen by the water occupancy (see Methods). In **NBOC-DSA-I**, G20, T4, A19 and A5 have low water occupancy again showing that ring **A** of **NBOC-DSA** is tighter bound to the minor groove than **DSA** and **DSI**. Covalent binding leads to further removal of water molecules from the minor groove in agreement with ref. [134]. In cov**DSA-I**, G20, T4, A19, A5, A6 and T18 are fully dehydrated. In cov**DSI-I** the same sites are fully dehydrated with the exception of T4, which has some residual water occupancy at this site (Figure 4.13).

4.3.2 Alkylation Reactions

Reactions of DSA, DSI and NBOC-DSA with I

The MD structures produced in the previous section are here used to investigate these alkylation reactions. Comparison between the reaction of **NBOC- DSA** with adenine in DNA and adenine in water allows investigating the effect of the DNA scaffold

Figure 4.12: Hydration in minor groove of non-covalent **DSA-I** (red), **DSI-I** (blue) and **NBOC-DSA-I** drug-DNA adducts compared to free DNA (black) Occupancy corresponds to time in which H-bond is present with respect to entire simulation time.
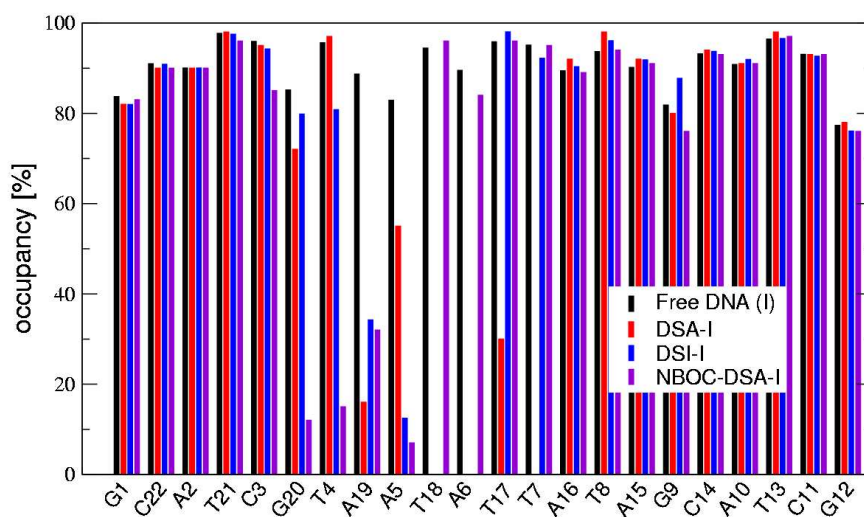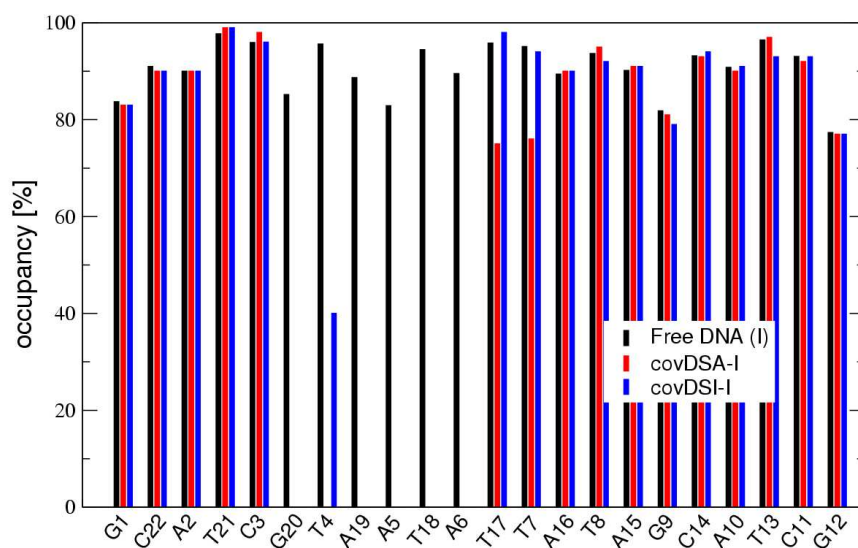

Figure 4.13: Hydration in minor groove of covalent cov**DSA-I** (red) and cov**DSI-I** (blue) drug-DNA adducts compared to free DNA (black) Occupancy corresponds to time in which H-bond is present with respect to entire simulation time.

on the reaction barrier. Our computational tool here is the constrained hybrid Car-Parrinello Molecular Mechanics/Molecular Dynamics method[13, 14]. The reactions are believed to occur via an S_N2 mechanism [133, 170] with the rate limiting step being the attack of the N3 atom of A19 on the least substituted carbon atom of the cyclopropyl unit (C13). Simultaneously, the C9-C13 bond is elongated and eventually broken, whereas the other two bonds in the cyclopropyl ring, d(C9-C10) and d(C10-C13) get shorter. In the second step, which is faster[133], the carbonyl-oxygen in O6 position is protonated. Thus, a plausible reaction coordinate for the rate limiting step is the d(N3(A)-C13(drug)) distance (*d* hereafter), assuming a linear decrease of this distance along the reaction pathway, making the approximation of a mono- dimensional reaction coordinate. We discuss the electronic and structural properties at selected steps along the reaction coordinate in terms of Polarization Index (PI) and D-RESP charges (see Methods)[14], bond-lengths and bond angles involving ring **A**. In particular, we focus on the C9-C13 bond length, as this bond is broken during the reaction. We start from conformations in which d= 3.3 Å. These are frequently encountered in the MD simulations of **DSA-I** and **DSI-I**, whereas they are found only at times in **NBOC-DSA-I** (see above).

At 3.3 < *d* < 2.8Å, the bond lengths of the cyclopropyl group are in the order: (C9-C13), d(C9-C10), d(C10-C13), and the angles in the strained 3-membered ring are close to 60°, as found experimentally in a structurally related compound (Cambridge database, RUGGAB[171]). This suggests that, in this range of *d*, the structure of the reactive moiety of the drug is not significantly perturbed. The C9-C13 bond is polarized towards C9 in the three complexes (mostly in **DSA-I**, Table 4.3.2), while the N3 lone pair is polarized towards C13 (Figure 4.14. The D-RESP charge of N3 is slightly negative.

At *d*= 2.6 Å, the d(C9-C13) becomes slightly longer (by 0.01 Å) and more polarized towards C9 whereas the other two cyclopropyl-bonds get shorter. No significant change is observed in the angles in the cyclopropyl unit. The N3 lone pair is larger and more polarized (Figure 4.14) and its D-RESP charge is slightly decreased (Table 4.3.2).

At *d*= 2.4 Å, the d(C9-C13) gets further elongated and the bond-length fluctuates significantly (Table 4.3.2). The (C13-C9-C10) and (C10-C13-C9) angles are smaller than 50°, whereas ∠(C13-C10-C9) becomes as large as 87°. The charge on N3 is now almost zero.

At *d*= 2.3Å the constraint force is close to 0 and at the next step (*d*= 2.2 Å), the sign of the constraint force changes, indicating that the transition state is passed. Thus, we take the *d*= 2.3 Åas a model of the TS. d(C9-C13) is about 2.0Å, d(C9-C10) is
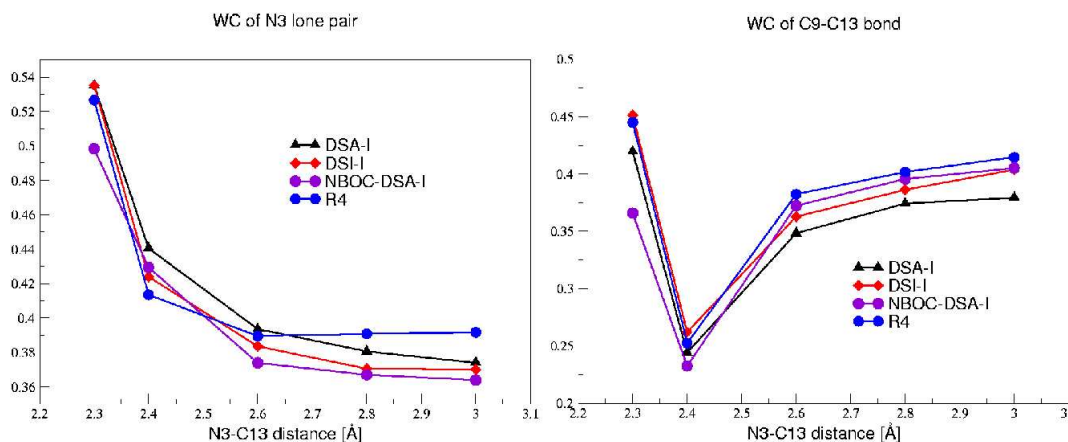
Figure 4.14: Wannier Centers for the three reactions of **DSA**, **DSI** and **NBOC-DSA** in DNA. The reaction of **NBOC-DSA** in water is also shown for comparison.

further shortened and d(C10-C13) is slightly longer. $\angle$(C13-C9-C10) and $\angle$(C10-C13-C9) are significantly smaller than $60°$, whereas $\angle$(C13-C10-C9) gets as large as $87°$. The largest C13-C10-C9 angle is observed in **DSA-I**, which features also the longest C9-C13 bond. The C9-C13 bond is strongly polarized towards C9 and the correspondent Boys orbital center resembles indeed more a lone pair localized on C9 (see Figure 4.18). This negative charge up-build at C9 can be stabilized by delocalization over the aromatic system of ring **A**, as seen in the change of bond- lengths in the 6-membered ring of the indole-unit: d(C9-C8), d(C9-C4), d(C8-C7) and d(C6-C5) get shorter whereas d(C8-C7) and d(C5-C4) get longer. The N3 lone pair is further polarized towards C13 (Figure 4.14) and the D-RESP charge of this atom is less negative. The decrease of negative charge on this center during the reaction illustrates qualitatively the charge transfer and tells about the proceeding of the reaction[14]. In **DSA-I** the polarization of the N3 electron lone pair is the largest, consistent with the highest D-RESP charge on N3 (Table 4.3.2).

Energetics: The free energy profiles are very similar for all three drugs, ranging from 10.5kcal/mol - 13.8kcal/mol (Table 4.3.2). Our calculations underestimate severely the experimental free energies, which range from 22-28kcal/mol, by 10kcal/mol. The several possible causes of such dramatic difference between experimental and theoretical values are discussed in the Discussion Section.

Drug-Conformation: In our classical simulation we observe a shift in the χ_1 and χ_2 distribution when passing from the non-covalent to the covalently bound species. Here, we confront these results to the QM/MM simulation, in which the torsional

DSA-I	N3-C13 distance [Å]			
D-RESP charges	3.0	2.6	2.4	2.3
N3	-0.035	-0.015	0.019	0.032
Bond-lengths and bond angles				
C9-C13	1.590 (.026)	1.660 (.050)	1.816 (.104)	2.057 (0.095)
C9-C10	1.558 (.019)	1.545 (.020)	1.526 (.012)	1.520 (.011)
C10-C13	1.467 (.011)	1.461 (.010)	1.464 (.011)	1.475 (.014)
C9-C8	1.467 (.009)	1.461 (.011)	1.443 (.012)	1.422 (.009)
C8-C7	1.379 (.008)	1.385 (.010)	1.388 (.009)	1.399 (.010)
C7-C6	1.444 (.011)	1.439 (.013)	1.441 (.012)	1.430 (.012)
C6-C5	1.446 (.011)	1.441 (.012)	1.440 (.012)	1.437 (.012)
C5-C4	1.412 (.009)	1.411 (.008)	1.416 (.010)	1.422 (.009)
C9-C4	1.458 (.011)	1.454 (.012)	1.440 (.013)	1.425 (.011)
N3-C13-C9	157 (5)	161 (5)	161 (7)	160 (5)
C13-C9-C10	56 (1)	54 (1)	51 (2)	46 (2)
C10-C13-C9	61 (1)	59 (2)	54 (3)	48 (3)
C13-C10-C9	63 (1)	67 (3)	75 (5)	87 (5)
DSI-I	N3-C13 distance [Å]			
D-RESP charges	3.0	2.6	2.4	2.3
N3	-0.014	0.009	0.044	0.030
Bond-lengths and bond angles				
C9-C13	1.592 (.033)	1.645 (.041)	1.786 (.112)	1.971 (.109)
C9-C10	1.541 (.018)	1.540 (.017)	1.532 (.013)	1.520 (.009)
C10-C13	1.480 (.014)	1.473 (.010)	1.466 (.007)	1.474 (.015)
C9-C8	1.474 (.009)	1.464 (.011)	1.447 (.009)	1.433 (.009)
C8-C7	1.383 (.007)	1.383 (.010)	1.388 (.007)	1.393 (.008)
C7-C6	1.441 (.013)	1.447 (.012)	1.443 (.009)	1.431 (.008)
C6-C5	1.444 (.011)	1.449 (.014)	1.443 (.009)	1.440 (.009)
C5-C4	1.407 (.008)	1.412 (.010)	1.414 (.008)	1.420 (.006)
C9-C4	1.458 (.013)	1.455 (.014)	1.440 (.009)	1.432 (.010)
N3-C13-C9	158 (7)	171 (5)	169 (5)	164 (4)
C13-C9-C10	56 (1)	55 (1)	52 (2)	48 (2)
C10-C13-C9	60 (1)	59 (1)	55 (3)	50 (3)
C13-C10-C9	64 (2)	66 (2)	73 (6)	82 (5)
NBOC-I	N3-C13 distance [Å]			
N3	-0.056	-0.055	-0.007	0.018
Bond-lengths and bond angles				
C9-C13	1.591 (.032)	1.626 (.032)	1.791 (.085)	1.990 (.105)
C9-C10	1.573 (.022)	1.559 (.021)	1.536 (.015)	1.529 (.009)
C10-C13	1.470 (.009)	1.462 (.012)	1.456 (.006)	1.468 (.012)
C9-C8	1.447 (.011)	1.464 (.011)	1.439 (.015)	1.424 (.007)
C8-C7	1.374 (.008)	1.377 (.008)	1.388 (.008)	1.395 (.005)
C7-C6	1.444 (.011)	1.439 (.011)	1.431 (.010)	1.427 (.008)
C6-C5	1.451 (.013)	1.446 (.012)	1.441 (.010)	1.439 (.009)
C5-C4	1.409 (.008)	1.411 (.007)	1.416 (.009)	1.421 (.006)
C9-C4	1.466 (.013)	1.462 (.012)	1.443 (.010)	1.433 (.009)
N3-C13-C9	166 (6)	157 (4)	156 (4)	156 (4)
C13-C9-C10	56 (1)	55 (1)	51 (2)	47 (2)
C10-C13-C9	62 (1)	60 (1)	55 (3)	50 (3)
C13-C10-C9	62 (2)	65 (2)	74 (4)	83 (5)

Table 4.8: D-RESP charges, bond-lengths [Å] and angles [°] for selected distances along the reaction coordinate

	DSA-I	DSI-I	NBOC-DSA-I
$\Delta G^{\#}_{theoretical}$	10.5 ±3.4	13.8±3.6	12.1±2.4
$\Delta G^{\#}_{experimental}$	22.7	24.5	26.8-28.2

Table 4.9: Activation free energy for reactions of **DSA**, **DSI** and **NBOC-DSA** in DNA [kcal/mol]

angle distribution does not rely on classical parametrization. Even though the total sampling time is very short (7.5ps), we observe a distribution, which is similar to the classical one and almost as broad (Figure 4.15). The distribution does not change significantly along the reaction.

Reaction between adenine and NBOC-DSA (R4)

As a reference reaction in water, we choose the smallest compound **NBOC-DSA**, which is stable in water at pH 7 [172]. The simulation is based on a classical MD simulations in which d is constrained at 3.3Å. In the MD-simulation, the N3(A) atom has no neat solvation shell because of its vicinity to **NBOC-DSA**, whereas N1, which has a similar charge (-0.6e), is fully solvated with two water molecules within 3.5Å. We follow the reaction using the same reaction coordinate as for the reaction towards **I**, analyzing the same steps along the reaction coordinate in terms of PI, D-RESP charges and bond lengths and angles (Table 4.3.2). In particular, ∠N3-C19-C9 can be taken as a measure for the ideal atom positioning for S_N2 reactions, in which attacking and leaving group are in-line and the central carbon atom adopts a trigonal bipyramidal structure. This angle varies significantly during the reaction. Instead, in the reaction of **NBOC-DSA-I**, the ∠N3- C19-C9 angle changes very little (Table 4.3.2), because of geometrical constraints.

We focus on the difference between the two reactions. The latter is manifest in the activation free energy, which is 4kcal/mol higher in water than in DNA. This corresponds to a rate acceleration of roughly 3 orders of magnitude. Although the reaction barrier is largely underestimated, our finding is in qualitative agreement with the fact that DNA catalyzes the reaction.

At d= 3.0 Å, the C9-C13 bond is less polarized in R4 (Figure 4.14), suggesting that this polarization is a key feature for the increase in reactivity of the drug caused by **I**. Poisson-Boltzman calculation of the electrostatic potential at the C9 and C13 sites supports this picture (Table 4.11): In **I**, the electrostatic potential is more negative at the C13 position than at the C9 position, leading to a field, which polarizes this bond
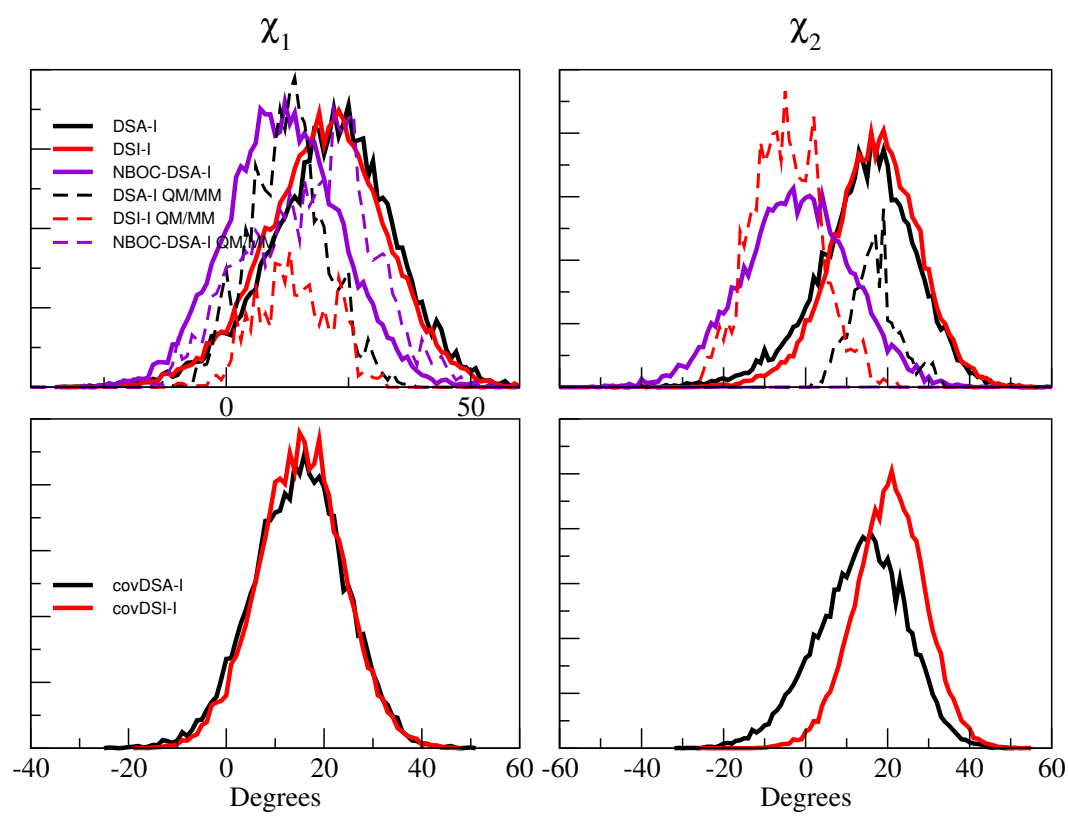
Figure 4.15: Dihedral angle distribution for the classical MD simulation in the non-covalent (upper panel) and covalent (lower panel) drug-DNA complexes.

R4	N3-C13 distance [Å]			
D-RESP charges	3.0	2.6	2.4	2.3
N3	-0.270	-0.229	-0.313	-0.154
Bond-lengths and bond angles				
C9-C13	1.580 (.036)	1.626 (.036)	1.768 (.087)	2.168 (.147)
C9-C10	1.559 (.027)	1.554 (.022)	1.535 (.020)	1.524 (.012)
C10-C13	1.478 (.012)	1.467 (.009)	1.466 (.011)	1.486 (.014)
C9-C8	1.477 (.015)	1.466 (.010)	1.449 (.015)	1.412 (.008)
C8-C7	1.379 (.009)	1.377 (.009)	1.388 (.011)	1.407 (.008)
C7-C6	1.441 (.012)	1.440 (.011)	1.441 (.014)	1.426 (.009)
C6-C5	1.447 (.020)	1.446 (.010)	1.444 (.015)	1.432 (.009)
C5-C4	1.408 (.010)	1.412 (.007)	1.413 (.010)	1.423 (.009)
C9-C4	1.464 (.017)	1.461 (.010)	1.447 (.015)	1.421 (.010)
N3-C13-C9	165 (9)	155 (7)	153 (6)	146 (6)
C13-C9-C10	56 (1)	55 (1)	52 (2)	43 (3)
C10-C13-C9	61 (2)	60 (2)	56 (3)	45 (4)
C13-C10-C9	63 (2)	65 (2)	72 (4)	92 (7)

Table 4.10: D-RESP charges, bond-lengths [Å] and angles [°] for selected distances along the reaction coordinate for the reactions R1-R4 in water

and renders it more reactive. In R4, the electrostatic potential is at both sites very similar and fluctuates around 0 (Table 4.11). The N3-lone pair of adenine is on the other hand more polarized in R4.

In the following d= 2.8 Å, 2.6 Åand 2.4 Åthe polarization of this electron lone pair remains almost constant in R4 whereas it increases in **I** and at d= 2.4 Å, the polarization of N3 towards C13 is very similar in both reactions. d(C9-C13) undergoes similar changes in both environments and increases steadily whereas $\angle$(C13-C9-C10) and $\angle$(C10-C13-C9) angles decrease, $\angle$(C13-C10-C9) increases and the absolute values compare very well.

At d=2.3 Å, d(C9-C13) is around 2Å, and longer in R4 than in **I**. In both reactions, the N3 lone pair is largely polarized towards C13, indicating the formation of a bond between these two atom centers. The D-RESP charge on N3 gets less negative during both reactions. In R4, the D-RESP is however still negative at the TS whereas in **I**, N3 gets slightly positive.

QMMM calculations on model systems in water

We conclude this Section by reporting QM/MM simulations of selected systems in aqueous solution, using the same computational scheme. Starting from R4, discussed

	NBOC-DSA-I		R4	
d(N3-C13) [Å]	3.3	3.0	3.3	3.0
C9	-6.65 (0.32)	-6.72 (0.22)	0.004 (0.047)	-0.111 (0.02)
C13	-7.62 (0.32)	-7.86 (0.24)	0.005 (0.050)	-0.27 (0.04)
Δ	0.97	1.14	~ 0.0	0.16

	DSA-I	**DSI-I**	**NBOC-DSA-I**
C9	-5.3 (1.0)	-5.6 (1.1)	-4.3 (0.8)
C13	-6.1 (1.4)	-6.4 (1.5)	-4.8 (1.1)
Δ	0.89	0.84	0.50

Table 4.11: Electrostatic Potential [kT/e] at C9 and C13 sites of drugs. A constant dielectric constant for DNA and water and an even counter ion distribution are assumed, results are only of qualitative nature. (i) QM/MM calculation of **NBOC-DSA** in **I** and with adenine in water (R4). (ii) MD-simulation of non-covalent drug-**I** complexes.

above, we decrease the complexity of the system gradually (R3-R1, Figure 4.3). First, we reduce the **NBOC-DSA** unit by omitting the pyrrole ring condensed to the indole unit (R3). Then we substitute the adenine by ammonia (R2) and finally, we investigate the reaction of cyclopropane with ammonia (R1). Comparison between these reactions provide insights on the role of the pyrrole ring, adenine and the indole unit for the reactivity. Some features of R4, which are related to the influence of the pyrrole ring, are also discussed in this section.

The systems undergo first classical MD and they are fully solvated. In particular, the nucleophile nitrogen of ammonia and adenine interacts on average with 1.0 and 0.8 waters, as obtained from the radial distribution functions.

In the QM/MM simulation, at d= 3.0 Å, the nucleophile (either NH_3 or A) moves freely around the cyclopropyl unit with little orientational preference in the timescale investigated as can be seen in $\angle$(N3-C13-C9), which ranges from 105°-163°. The standard deviations are large, indicating high mobility in this degree of freedom. The cyclopropyl-unit present in R2 and R3 renders the C-C bonds of the cyclopropyl unit asymmetric: the C9-C13 and C9-C10 bonds are longer than C10-C13 (Table 4.3.2). Instead, as expected, the cyclopropyl-unit is almost symmetrical in R1: the slight differences among the three C-C bond-lengths are possibly caused by inhomogeneities in the solvent electric field and the presence of the ammonia. As a result, the bonds in the cyclopropyl unit are slightly polarized in R2, R3 and R4, whereas in R1 they are not polar (Figure 4.16).

In the three systems, the C-C angles are close to 60°, which leads to a considerable
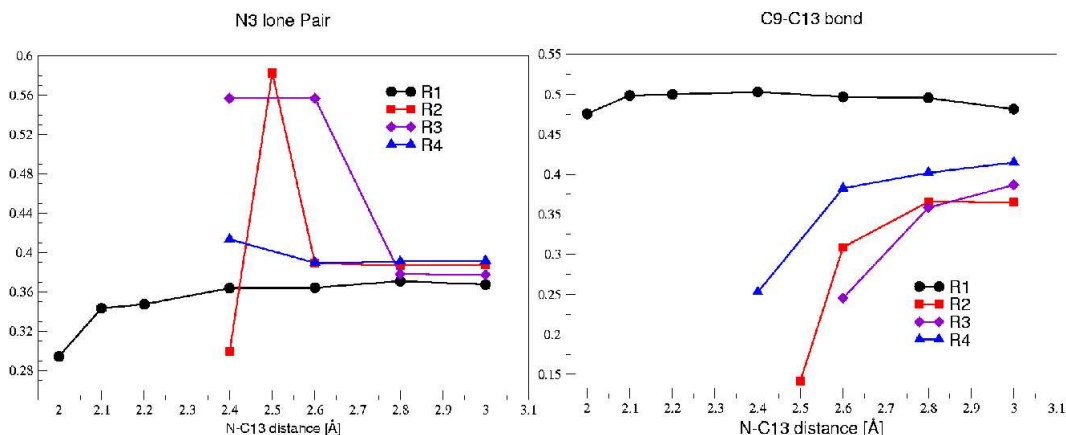
Figure 4.16: PI for nitrogen nucleophile (left) and C9-C13 bond (right) for R1-R4

ring strain. This strain is partially lifted thanks to so-called banana shaped orbitals, which are displaced outwards with respect to the straight bond (Figure 4.17). The lone pair of the ammonia-nitrogen (R1, R2) is less polarized than the N3(A) lone pair (R3, Fig. 6). The D-RESP charges at this center are largely negative in all reactions.

At d= 2.6 Å, the TS state is reached for R2 and R3. The nucleophile adopts an orientation in which the N3 or the N3 electron lone pair (represented by the Boys-orbital represented in Figure 4.18) points towards the C13 atom and all three atom centers involved in the reaction, namely N3, C13 and C9 form a close to ideal angle, which ranges from 158°-170°.

The standard deviations of this angle are also decreasing, showing that the fluctuations are getting smaller in this degree of freedom (Table 4.3.2). In R2 and R3, the C9-C13 bond is further elongated and the other two C-C bonds get slightly shorter. The C-C bonds are instead almost equivalent in R1. The C9-C13 bond is also more polarized towards C9, whereas the N of ammonia and N3(A) are more polarized towards C13 (Figure 4.16). The D-RESP charge of N3(A) in R3 has decreased significantly and is close to 0 whereas in R2, the charge on the ammonia-nitrogen is still highly negative. However, at the next step after the TS, d= 2.5 Å the D-RESP charge is positive. At d= 2.1Å, the TS is reached also for R1. The C-C bond and C-C-C angels are much less distorted than those of R2, R3 (Table 4.3.2). The N3 lone pair is much less polarized towards C13 and the C9-C13 bond towards C13 than in R2-R3 (Figures 4.16, 4.17, and 4.18). The D-RESP charge is still highly negative at the TS.

Energetics: The activation free energy decreases dramatically from R1 to R3 and R2 (Table 4.13). That of R4 is larger than R3 by few kcal/mol. No experimental data
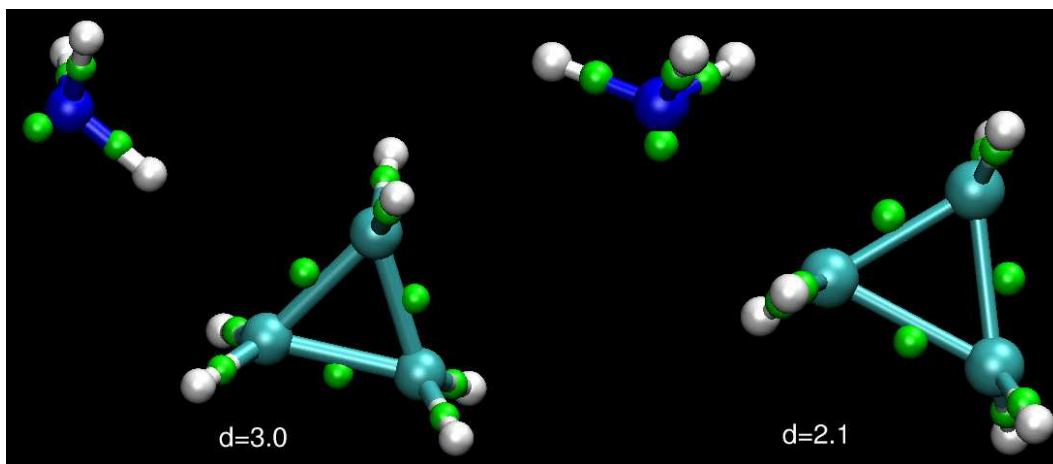
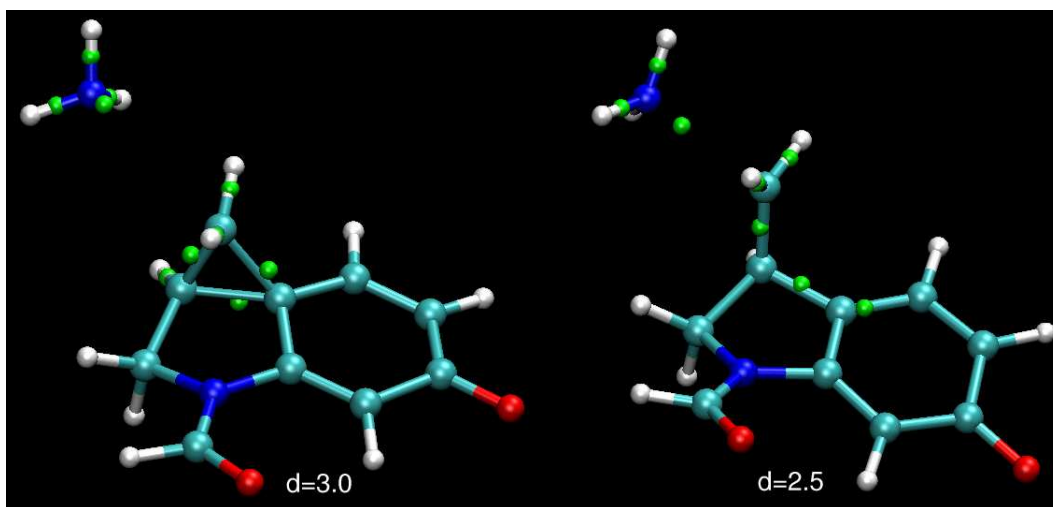Figure 4.17: Boys-orbital centers in R1 at two snapshots for d=3.0Å (left) and d=2.1Å.

Figure 4.18: Boys-orbital centers in R2 at two snapshots for d=3.0Å (left) and d=2.5Å.

R1	N3-C13 distance [Å]		
D-RESP charges	3.0	2.6	2.1
N3	-0.490	-0.320	-0.236
Bond-lengths and Angles			
C9-C13	1.520 (.029)	1.526 (.020)	1.603 (.035)
C9-C10	1.531 (.030)	1.523 (.014)	1.525 (.035)
C10-C13	1.517 (.025)	1.510 (.011)	1.487 (.010)
N3-C13-C9	159 (8)	161 (5)	161 (4)
C13-C9-C10	59 (2)	59 (1)	57 (1)
C10-C13-C9	60 (1)	60 (1)	59 (1)
C13-C10-C9	60 (1)	60 (1)	64 (2)

R2	N3-C13 distance [Å]		
D-RESP charges	3.0	2.6	2.5
N3	-0.322	-0.362	0.039
Bond-lengths and Angles			
C9-C13	1.598 (.040)	1.689 (.066)	2.184 (.232)
C9-C10	1.560 (.027)	1.557 (.021)	1.547 (.016)
C10-C13	1.473 (.019)	1.462 (.011)	1.472 (.014)
C9-C8	1.470 (.016)	1.453 (.012)	1.424 (.011)
C8-C7	1.370 (.010)	1.376 (.009)	1.389 (.007)
C7-C6	1.453 (.018)	1.444 (.012)	1.429 (.008)
C6-C5	1.467 (.018)	1.456 (.014)	1.433 (.010)
C5-C4	1.356 (.012)	1.365 (.007)	1.392 (.012)
C9-C4	1.451 (.017)	1.435 (.018)	1.403 (.009)
N3-C13-C9	107 (10)	170 (4)	157 (11)
C13-C9-C10	56 (1)	53 (1)	42 (6)
C10-C13-C9	61 (2)	59 (2)	45 (7)
C13-C10-C9	63 (2)	68 (3)	93 (12)

R3	N3-C13 distance [Å]		
D-RESP charges	3.0	2.6	2.5
N3	-0.061	-0.007	0.124
Bond-lengths and Angles			
C9-C13	1.628 (.062)	1.719 (.085)	2.069 (.138)
C9-C10	1.570 (.030)	1.556 (.024)	1.542 (.011)
C10-C13	1.462 (.019)	1.463 (.014)	1.467 (.013)
C9-C8	1.457 (.022)	1.446 (.012)	1.427 (.006)
C8-C7	1.372 (.020)	1.376 (.010)	1.389 (.006)
C7-C6	1.445 (.020)	1.443 (.012)	1.437 (.008)
C6-C5	1.455 (.184)	1.449 (.012)	1.437 (.007)
C5-C4	1.363 (.161)	1.367 (.086)	1.385 (.009)
C9-C4	1.444 (.022)	1.433 (.014)	1.408 (.007)
N3-C13-C9	163 (5)	158 (10)	156 (4)
C13-C9-C10	54 (2)	53 (2)	46 (3)
C10-C13-C9	60 (2)	58 (3)	49 (4)
C13-C10-C9	66 (3)	69 (4)	85 (8)

Table 4.12: D-RESP charges, bond-lengths [Å] and angles [°] for selected distances along the reaction coordinate for the reactions R1-R4 in water

	R1	R2	R3	R4
$\Delta G^{\#}$	46.8 $\pm$2.3	6.8$\pm$3.8	10.6 $\pm$4.7	16.1 $\pm$4.0

Table 4.13: Activation free energy for R1-R4 in water [kcal/mol]

is available. Based on the previous analysis, we can identify the features responsible for such ordering. The higher reactivity of R2, R3 and R4 with respect to R1 can be explained by the efficient delocalization of the negative charge in the aromatic system at the TS. This is illustrated again by the increase in aromaticity in the 6-membered ring of the indole unit, as shown by the similarity of the C- C bond lengths at the TS; specifically, the C9-C8, C7-C6, C6-C5 and C9-C4 bonds are shortened, whereas the C8-C7 and C5-C4 bonds get elongated (Table 4.3.2). The larger activation free energy of R4 relative to R3 is caused by the less aromatic character of the 6- member ring of **NBOC-DSA** at the TS. This is evident by the fact that the C9-C4 bond is considerably shorter in R3 than R4 (Table 4.3.2). The decrease in aromaticity, in turn, delocalizes less efficiently the negative charge. This decrease is caused by the presence of the condensed aromatic pyrrole ring, which stabilizes the electron delocalization depicted in Figure 4.19, relative to other resonance structures. R2 is more reactive than R3 possibly because ammonia is a slightly better nucleophile than adenine. R2 is far more reactive than R1 as the indole unit stabilizes significantly the transition state by aromatic conjugation, and better delocalization of the negative charge on its electronegative groups (O6).

4.4 Discussion

We have presented a theoretical investigation of the key steps of the binding of duo-carmycins to DNA. Classical MD simulations on the 10 ns timescale have provided the mode of binding of non-covalent complexes (i.e. before the reaction takes place) between the **DSA**, **DSI**, and **NBOC-DSA** and the oligonucleotide **I**. Our computational setup for the MD has been validated by performing MD simulations on the NMR adducts, which compare well with the experimental data. The investigation has been complemented by a QM/MM study of the rate-limiting step of the alkylation reaction performed by the drugs. We find that **NBOC-DSA** binds less tightly than **DSA** and **DSI** to **I** because of the presence of the tert- butyl unit, which is shorter and bulkier than the indole unit of the other two drugs (Figure 4.1). The tert- butyl unit penetrates therefore less into the minor groove and the interactions involve also fewer
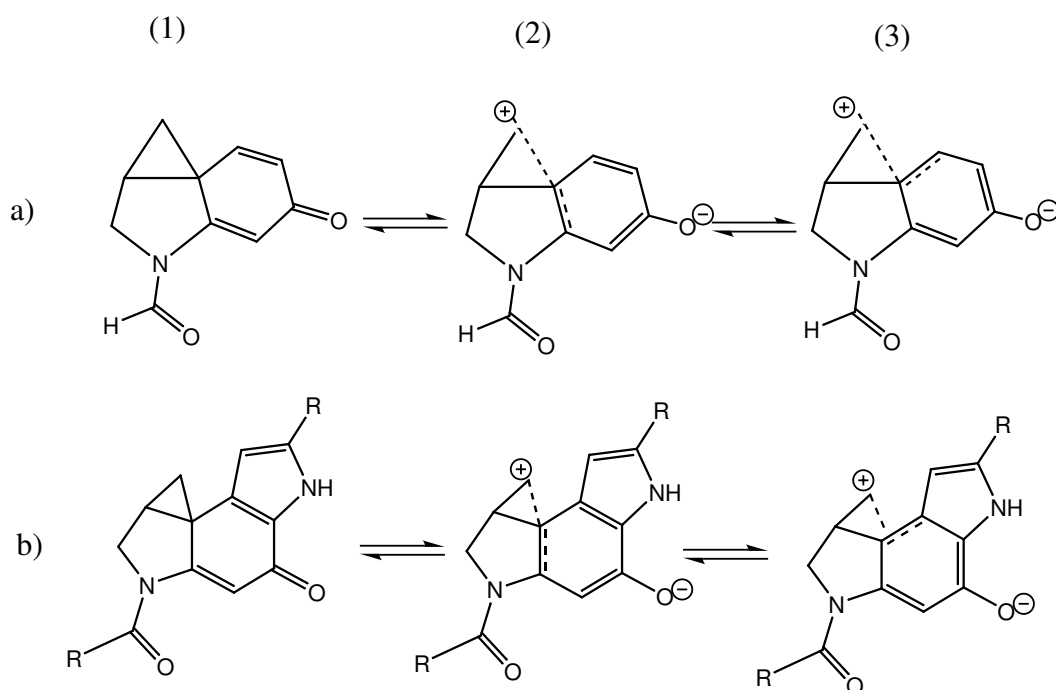
Figure 4.19: Resonance structures, which are able to delocalize the negative charge up-build at C9. (1) ground-state resonance structure, (2) first resonance structure, delocalization over C9-C8, (3) second resonance structure, delocalization over C9-C4. In **NBOC-DSA** the second resonance structure seems to be hampered by the presence of the condensed pyrrole ring.

DNA residues. As a result, reactive conformations, in which the cyclopropyl moiety of the drug is in close contact (about 3 Å) with the reactive adenine nucleobase, are frequently encountered only for **DSA** and **DSI**. In addition, **NBOC-DSA** binds rather unspecifically to **I** relative to the other two, as the drug cyclopropyl interacts also with N3(G20).

The structures of all the systems investigated here are similar and differ substantially from that of B- DNA, as expected for an A-tract containing oligonucleotide. The drugs bind to **I** by removing two to three ordered water molecules (occupancy < 20%, Figure 4.12,4.13). Formation of the covalent adduct causes the removal of a further number of water molecules (5 to 6 fully dehydrated sites). The DNA structure is not affected by drug binding. Thus, we expect that in the binding process a free energy cost is associated to DNA dehydration (counterbalanced by the free energy of drug binding), whilst that associated to conformational changes of the DNA scaffold is probably negligible. In all the DNA systems considered here, the large-scale motions involve the bending of the entire DNA double helix and the opening and closing of the DNA minor groove. The latter motion is slightly larger in the non-covalent complexes than in **I**, and it is much smaller in the covalent ones (Figure 4.8). The drugs' conformation changes when passing from the non-covalent to the covalent adducts. The torsional angles χ_1 (C7-C8-N12-C16) and χ_2 (C8-N12-C16-O16) are used to describe the conformational properties of the drugs. In the non-covalent adducts, the average χ_1 and χ_2 angles in **DSA** and **DSI** are very similar and larger than those in **NBOC-DSA**. In the corresponding covalent drug-DNA adducts, the average χ_1 is smaller and the standard deviation are also much lower. Thus, the drug is more flexible and distorted in the non-covalent than in the covalent complex. At the speculative level, we argue here that the different mobilities of covalent and non covalent complexes could be relevant for molecular recognition processes and the drug/DNA reactivity. In the covalent complexes, the smaller minor groove width relative to the free oligonucleotide, along with the relative rigidity of the entire complexes might play a role for their interaction with the duocarmycin-DNA adduct recognizing protein, DARP[173]. On the other hand, the relatively large fluctuations of the non-covalent complex could help to trap the DNA in a reactive conformation, as suggested by Ref. [145] The χ_1 and χ_2 values of the non-covalent complexes differ by only few degrees. This observation argues against the *'shape induced activation'* mechanism, in which the difference between such angles play a crucial role for the alkylation [143, 133].

The alkylation reactions have been described using constrained hybrid Car-Parrinello Molecular Mechanics/Molecular Dynamics simulation [13, 159]. The reaction is fol-

lowed using the d(N3(A)- C13(drug)) distance as reaction coordinate. The calculations are based on reactive conformations of the non-covalent drug-DNA structures, in which the drugs are in a favourable orientation and in close contact with the oligonucleotide for the S_N2 reaction. These are frequently found in the MD simulations of **DSA-I** and **DSI-I** (Figure 4.10), whereas they are found only at times in **NBOC-DSA-I**. Thus, it appears that **DSA** and **DSI** have been designed so as to have an optimal interaction with the DNA scaffold.

During the reaction, the C9-C13 bond of the cyclopropyl unit gets progressively longer and more polarized towards the C9, while the other two bonds in the 3-membered rings get slightly shorter. The N3(A19) electron lone pair gets more polarized towards the C13 and the negative charge at N3 slowly decreases until it eventually becomes positive at the transition state. At the transition state the C9-C13 bond is broken (Table 4.3.2) and the corresponding electronic pair is mainly localized on C9. The negative charge up-build at this center can be efficiently stabilized by delocalisation over the aromatic system. As a result, the aromaticity of the ring system increases (Table 4.3.2) following closely the scheme in Figure 4.3. The distribution of the torsional angles χ_1 and χ_2 in **DSA** and **DSI** are very similar and the average values are slightly larger than in **NBOC-DSA** (Figure 4.15). Despite the short simulation time, they are similar with the classical MD simulation, suggesting that the parametrization of the drug provide a reliable description of the drugs' conformational properties.
The calculated activation free energy profiles are very similar for all three drugs and severely underestimate the experimental values. Several can be the factors leading to such large discrepancy between theory and experiment: (i) The accuracy of the methodology used. It is well known that the BLYP approximation for the exchange correlation functional may largely affect the transition state energies. However, this does not seem to be the case for several chemical and enzymatic reactions[15] so far investigated with this method. To further address this issue, we have performed test calculations on R2 with a different recipe for the exchange-correlation functional, the B3LYP, with a localized basis set and implicit model of the solvent (Table 4.14). The calculated free energy is still off by 7 kcal/mol. The severe underestimation is however consistent with other theoretical studies on S_N2 reactions, in which different DFT functionals are compared to higher level ab initio calculations. Indeed, the transition state, where two electrons are delocalized over three atom centers resulting in a highly delocalized exchange hole, is poorly described by pure DFT methods and the TS results to be too stable, which leads eventually to a lower reaction barrier[174, 175, 176, 177]. Consistently, Barone and coworkers[178], performing the reaction on a very similar, but protonated NBOC-DSA-derivative in solution, ob-

	B3LYP	BLYP	B3LYP (PCM)	BLYP (PCM)
$\Delta E^{\#}$	18.8	16.4	9.6	6.3

Table 4.14: Gaussian calculations for R2 at BLYP and B3LYP level in vacuo and with implicit solvent (PCM). Geometry optimization at B3LYP level with single point calculations at BLYP level. Activation free energy in [kcal/mol].

tain an error within experiment of 12 kcal/mol for the reaction with methanol. (ii) The time-scale investigated. Because the QM/MM simulation is necessarily short, the results depend on the conformational properties and flexibility of the DNA scaffold as well as on the location of the counter ions [12].

Comparison with the correspondent reaction in water for the smaller of these drugs, **NBOC-DSA**, provides insights on the effect of the DNA scaffold on its reactivity. **NBOC-DSA** reacts with DNA but it is stable at pH 7 towards adenine. We find that: (i) The C9-C13 bond is more polarized in DNA than in water and therefore more reactive in the biomolecular frame, (ii) N3 of adenine is less negative in DNA than in water, which could render the charge transfer more easy in DNA and (iii) the activation free energy is 4kcal/mol higher in water than in DNA. This corresponds to a rate acceleration of roughly 3 orders of magnitude. Thus, although the also the activation free energy of the model reaction is smaller than the experimental data, our calculations are qualitatively in agreement with experiment, in that the DNA frame catalyzes the reaction.

Finally, comparison of the reaction in water with those between the simple reactions in Figure 4.3 (R1-R3) suggests that: (i) the presence of the condensed pyrrole ring leads to a less efficient aromatic stabilization of the negative charge over the aromatic indole ring. A similar energetic effect has been seen in a recent study on the acid catalysed reaction of duocarmycin analogues [178]. Indeed, in that study, the electronegative pyrrole ring attracts the positive charge from the protonated C6-carbonyl group, lowering so the electrophilicity of the cyclopropyl unit. (ii) Ammonia is a better nucleophile than adenine. (iii) The indole unit stabilizes significantly the transition state by aromatic conjugation, as seen by the variation in the bond lengths in the 6- membered ring of the indole unit, which tend to become all of the same length. The same effect is seen in the study of ref [178], even though in this study, the compound is already protonated. The good agreement can be due to the fact the number of electrons in the π-system is conserved.

4.5 Conclusions

Molecular simulation approaches have been used to investigate the key steps of the alkylation reaction of three duocarmycins. Several conclusions can be drawn by our study: (i) The experimentally observed DNA catalytic power might be due, at least in part, to a polarization of the biomolecular scaffold over the drugs. This effect may be reminiscent of the electrostatic preorganization concept in enzymatic catalysis, with the important difference that here the polarization seems to affect more the ground state than the TS [8]. Instead, our calculations do not support the so-called *'shape induced activation'* mechanism, in which the conformational properties of the drug play a pivotal role for catalysis [143]. (ii) The chemical nature of the drug influences the structure of the complex, thus affecting its reactivity; (iii) the DNA scaffold does not rearrange significantly upon accommodating the substrate.

Chapter 5

Final Remarks & Outlook

5.1 Conclusions

Molecular simulations of DNA have now reached a rather mature phase. On the one hand side, classical molecular dynamics of simulations have revealed itself as an extremely powerful tool to describe DNA structure, dynamical properties and energetics of nucleic acids [3, 5, 37, 38]. On the other hand, quantum chemical calculations on nucleodides/nucleosides in the gas [179, 180, 181, 182] and crystal phase [11] have provided insights on the energetics along with details of the electronic structure.

QM/MM studies have first been applied to investigate the hydrolysis of the glycosylic bond of uracyl residues erroneously incorporated in DNA [183]. In this thesis, we have used such a technique applied to drug/DNA complexes as an approach which could further expand the scope of molecular simulations of drug/target interactions. By focusing on two very important anticancer drugs we have addressed rather different issues.

In the first part of my study, I have investigated the interactions between the drug cisplatin and DNA. Cisplatin induces severe distortions in the DNA structure, which lead to a large kink at the platinated site along with a local conformational change from B-DNA to A-DNA. Platinated DNA-complexes are recognized by the HMG-domain of HMG-containing proteins. In my QM/MM simulation, the structure of the platinated DNA dodecamer rearranges significantly towards structural determinants of the solution structure as obtained by NMR spectroscopy [90]. The calculated ^{195}Pt chemical shifts of the QM/MM structure relative to cisplatin in aqueous solution are in qualitative agreement with the experimental data [128, 129]. The QM/MM structure of the platinated/DNA HMG complex, on the other hand, remains rather

similar to the X-ray structure, consistent with its relatively low flexibility. Docking of $[Pt(NH_3)_2]^{2+}$ onto DNA in its canonical B-conformation causes a large axis bend and a rearrangement of DNA as experimentally observed in the platinated adducts. Our study shows that the QM/MM approach proves to give reliable results for Pt-DNA complexes. Thus, this method could be very useful in cases where the incorporation of new compounds into classical force fields is combersome as for cisplatin and its derivatives. Short QM/MM molecular dynamics simulation can be used as refinement technique in order to reproduce the correct coordination geometry and allow for studying the electronic structure of the complex bound to DNA. A more accurate ab-initio treatment of the Pt-moiety can be very helpful in understanding slight electronic and structural differences between different compounds.

My second study has exploited the power of QM/MM method in the possibility to study chemical reactions while including the entire biomolecular frame. Many methods exist, which take solvent and environmental effects into account. However, most of these approaches do not take into account explicitly either the DNA scaffold, and/or temperature effects and/or the solvent. These effects are automatically incorporated in the hybrid Car-Parrinello QM/MM. In case of duocarmycin, DNA-binding follows a two step mechanism, the first one yielding the non-covalent drug-DNA complex and the second one producing the covalent drug-DNA adduct. Even though a wealth of kinetic data exists about hydrolysis in neutral and acidic environment, little is known about the non-covalent binding process. We have performed therefore also extensive classical molecular dynamics on both the non-covalent and the covalent drug-DNA complexes. Our classical mechanics simulation suggest that non-covalent binding and positioning of the drug in the minor groove is of great importance to the drug activity. Our QM/MM calculations, on the other hand, show that the biomolecular frame polarizes the drugs significantly and renders them more reactive with respect to the reaction in water. Furthermore, QM/MM simulation of model compounds in water suggest that the nature of the condensed rings in the drugs affects largely the intrinsic reactivity of the drug.

The accuracy of the activation free energies depends largely on the exchange-correlation functional. The BLYP gradient correction [52, 53], used here, has proven to be very reliable for a variety of enzymatic reactions, yet it underestimates the barriers in this work. This might be caused by the poor description of the exchange hole at the TS, in which 2 electrons are delocalized over 3 atom centers [176, 177]. Other parametrizations, notably the B3LYP performs only slightly better, whereas the PBE0

[184, 185] yields results comparable to BLYP for a similar reaction [178].

5.2   Future Research

Several lines of research would be interesting to pursue:

1. **Action of second generation drugs**
   The use of cisplatin is severely limited due to their intrinsic toxicity and re-
   sistance problems. A great effort is devoted therefore to the design of new
   Pt-drugs. The QM/MM docking of 2nd - and 3rd generation Pt-based drugs
   on DNA structure may help to gain insights on the structural properties of the
   platinated lesion. In this respect, I would like to mention that during the last
   year, our group has started a collaboration with the group of Prof. Ruggerone
   of the *Dipartimento di Fisica* in Cagliari, and the experimental group of Prof.
   Reedijk at the *Leiden Institute of Chemistry*, on the interaction between drugs
   other than cisplatin and DNA.

2. **Use of different functionals**
   Use of other functionals than BLYP, such as BP and PBE, might improve both
   the structural properties (e.g.the bond-lengths in Pt complexes, which could
   lead to an improvement in the calculated ^{195}Pt chemical shifts) as well as the
   energetics. Especially in case of chemical reactions, such as the $S_N 2$ reaction
   studied here, an improvement in the estimate of the TS is desirable.

3. **Non-covalent Drug Binding**
   Our MD simulations of the non-covalent duocarmycin-DNA adducts have shown
   that correct binding of the drug in the minor groove prior to alkylation con-
   tributes to the different reactivity observed in the three drugs studied here. The
   QM/MM simulations of the alkylation reaction are very similar for all the three
   drugs. Within the limitations of the methodology as discussed above, we sug-
   gest that the difference in reactivity arises in part also from the non-covalent
   binding process.
   The diffusion of the drugs towards DNA could be studied by means of meta-
   dynamics simulations [186]. This method allows for inclusion of collective
   coordinates, which characterize the conformational transition of the system un-
   der investigation. The system evolves according to these collective coordinates,

exploring the conformational space around them as in conventional MD. A history dependent potential term fills the minima of the free energy surface (FES) during the dynamics and allows for an efficient exploration of the FES as a function of the collective coordinates.

Future work could include the definition of a suitable set of collective coordinates and the simulation of the binding and dissociation process of duocarmycin to the DNA minor groove. These simulations could give insight into the recognition and binding process and they could help to identifying the structural features of the drug, which are most important for the binding process.

4. **Drug-DNA/protein interactions**

The antitumor activity of cisplatin-derived drugs seems to be correlated to the binding affinity ot the platinated DNA towards HMG domain containing proteins [99]. Both X-ray and NMR structure of the oxaliplatin-DNA complex exist (pdb-entries 1PG9, 1IHH, respectively [187, 188]). Docking of the HMG A domain on these structures followed by MD-simulations could give insight into structural and energetic differences responsible for the eperimentally known binding constants.

Similar studies for duocarmycin-DNA adducts would also be of large interest but they are still outside reach, since no structural data about the nature of DARP exists so far.

Chapter 6

Acknowledgments

The hundred pages of thesis do not tell a lot about these four years of PhD. This PhD is most of all the story of a naive girl - me - with a very vague idea to go to Italy in order to do a PhD - or let's tell the truth - compensate for the first 15 years of my life in which I didn't see the sea. Maybe the taste of pizza was playing a non-negligible role as well in the decision-process.

Anyway, after an initial period during which I overcame the cultural shock and acquired a rudimentary knowledge of the italian language (never improved since) I started to enjoy myself quite a lot. I therefore want to thank all the great people for the unforgettable time we had together:

My supervisor Paolo Carloni, who already knew something about swiss women but nevertheless gave me the opportunity to do a PhD in his group. He always stayed on my side and for this I am extremely grateful!

Anna, for being my friend and the most crazy flatmate I ever had, for all the discussions, which were never boring and for organizing the nicest birthday surprise when she actually had one of the worst hangovers of her life!

Michele (Ciras) and Andrea, who were the only persons letting me drive their cars and therefore showing that they really trust in me (or that they have a good car insurance...) Thanks for being such good friends and for helping me all the times it was necessary! Life at Trieste would have been much more boring and difficult without them!

Michele (Cascella), the only person who shared the office with me during all these four years and who got so fond of the swiss way that he moved to Lausanne. Besides some hysteric crisis, which I attribute to his sensibility and his intelligence, he is a wonderful friend, and he spent a lot of (wasted) time to diminuish my cultural igno-

rance.

Alejandro and Roxanna and also small Santiago, for their friendship, their easy way of dealing with problems and for always being in a good mood!

Angelo, who is the most honest and modest person I ever met. For example, he tought me a lot of useful italian words for conversation. Then when people noticed, he pretended that I didn't learn them from him. Come on, don't be shy! It's really your merit!!!

Cristina, who stayed with us at SISSA for a too short time but cheered up life considerably during this period. It is a pity that you left again but you did not get rid of us - The week-end spent in Barcelona is one of the best recordings I have!

Alessandra: I don't remember life at SISSA before she joined our group. I learned more about female psychology than in the 27 years before and according to me, I did a big step forward in diplomacy thanks to her. I'm happy to leave this rather complicated issue behind me and turn back as before, but I'm sure that this will not hinder our friendship!

Daniele: for his friendship and his innumerable good advices, which aimed at infecting me by his perfectionsm. If the world would be ideal, I would have taken all his good advices and I would have tried to follow them. Unfortunately nothing is ideal and therefore Daniele remains one of the few persons fighting constantly for perfection. Thanks for being my friend anyway and see you in Philadelphia!

Nico, for all the interesing discussions we had even thoug we never had the same opinion. Also for his music collection - I will miss the car-rides!

Stefano, for being so typical italian - always in a good mood and always late...

My new flatmates, namely Gioia, one of the most lovely person I met at SISSA, unfortunately too late. Thanks for all the nice dinners and for bringing Spritz into our appartment! And Gorana, who basically came and left again - escaping to London.

All the actual and former grouper members: Lore and Gianni, for housing me when I didn't have any place to stay; Marco Punta and Valeria; Sergio and Claudia; Francesco Musiani; Matteo and Lucia with new-entry Francesco and Marco and Francesca: thanks for all the scientific and non-scientific help and for being great friends! Pietro for his wonderful cooking and for being the most quiet office-mate (he's never around); Giovanna for being the way she is; Giacomo for his endless patience to explain obvious computer-tricks to rather limited people like me; Marco Berrera; Our senior group members: Simone, Yurai, Andrea, Claudio for useful discussions. The young members: Marilisa, Vicenzo and Attilio: Special thanks to Attilio who brought up questions about DNA that made me actually THINK.

All the tough women at SISSA with whom we organized "cena delle donne": Mon-

ica, Silvia, Valentina, Paola and Manuela.

All the other guys from SISSA: Luca, Pedro, Ruben and Valentina, Roberto and Chiara, Lorenzo Stella, Claudio, Michele Casula, Claudia and Osvaldo, Adriano, Luca Marsella, Federico, Vittoria, Francesca (the electric one), Moritz and Jeanne...

All the people I have worked with: Prof. Ursula Rothlisberger, Prof. Paolo Ruggerone, for the hospitality and the opportunity to taste snails and Paola Melis who is carrying on the Pt-project.

I also wish to thank all my friends from Switzerland, from Norway and all around Europe, who came to visit me: Susan and Christina - survivors of little car-crashes, Alice, Veronique and Severine, Andreas and Rike, Olivier - who never forgets my birthday, Anja, Dag, Cathrine and Raad and my host family: Lisbeth and Arne.

A special thank to my family, who stayed close to me during all this time and provided moral and material support in form of chocolate, cheese and many other things, which I might have missed from home. They are the best that could happen to me!

My last thank goes to Lorenzo, my wonderfully funny and extremely pedantic boyfriend, who has convinced me that blonds are boring. You make me always laugh, and sometimes cry, but I'm happy for every moment that we spend together. Thanks for being the way you are and accepting me (more or less) the way I am. You are probably the person who suffered most with me during these four years - freezing on the motor bike at -10° or hanging around on parties where everybody speaks only swiss german... Thanks!

Thank you!

October 31th, 2004

Bibliography

[1] Hagerman P.J. Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. *Biochem.*, 24:7033–7037, 1985.

[2] Haran T.E. and Crothers D.M. Cooperativity in a-tract structure and the bending properties of composite tnan blocks. *Biochem.*, 28:2763–2767, 1989.

[3] Beveridge D.L. and McConnell K.J. Nucleic acids: theory and computer simulation, Y2K. *Curr Opin Struct Biol.*, 10:182–196, 2000.

[4] Cheatham III T.A. and Kollman P.A. Observation of the A-DNA to B-DNA Transition During Unrestrained Molecular Dynamics in Aqueous Solution. *J.Mol.Biol.*, 259:434–444, 1996.

[5] Kombo D.C., McConnell K.J., Young M.A., and Beveridge D.L. Molecular dynamics simulation reveals sequence-intrinsic and protein-induced geometrical features of the OL1 DNA operator. *Biopolymers*, 59:205–225, 2001.

[6] Beveridge D.L., Dixit S.B., Barreiro G., and Thayer K.M. Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. *Biopolymers*, 73:380–403, 2004.

[7] Saito M. and Sarai A. Free energy calculations for the relative binding affinity between DNA and lambda-repressor. *Proteins*, 52:129–136, 2003.

[8] Warshel A. COMPUTER SIMULATIONS OF ENZYME CATALYSIS: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.*, 32:425–443, 2003.

[9] Warshel A. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J.Mol.Biol*, 103:227–49, 1976.

[10] Schlegel B.H. Exploring Potential Energy Surfaces for Chemical Reactions: An Overview of Some Practical Methods. *J.Comp.Chem.*, 24:1514–1527, 2003.

[11] Gervasio F.L., Carloni P., and Laio A. Electronic structure of wet DNA. *Phys.Rev.Lett.*, 89:108102, 2002.

[12] Barnett R.N., Cleveland C.L., Joy A., Landman U., and Schuster G.B. Charge Migration in DNA: Ion-Gated Transport. *Science*, 294:567–571, 2001.

[13] Laio A., VandeVondele J., and Rothlisberger U. A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J.Chem.Phys.*, 116:6941, 2002.

[14] Laio A., VandeVondele J., and Rothlisberger U. D-RESP: Dynamically Generated Electrostatic Potential Derived Charges from Quantum Mechanics/Molecular Mechanics Simulation. *J.Phys.Chem.B*, 106:7300–7307, 2002.

[15] Sulpizi M., Laio A., VandeVondele J., Rothlisberger U., Cattaneo A., and Carloni P. Reaction Mechanism of Caspases: Insights from Mixed QM/MM Car-Parrinello Simulations. *Proteins- Structure, Function and Genetics*, 52:212–224, 2003.

[16] Piana S., Bucher D., Carloni P., and Rothlisberger U. Reaction Mechanism of HIV-1 Protease by Hybrid Car-Parrinello MD/Classical MD Simulations. *J. Phys. Chem. B*, in press.

[17] Dal Peraro M., Vila A.J., and Carloni P. Substrate binding to mononuclear metallo-β-lactamase from Bacillus cereus. *Proteins*, 54:412–423, 2004.

[18] Röhrig U.F., Frank I., Hutter J., Laio A., VandeVondele J., and Rothlisberger U. QM/MM Car-Parrinello Molecular Dynamics Study of the Solvent Effects on the Ground State and on the First Excited Singlet State of acetone in water. *ChemPhysChem*, 4:1177–1182, 2003.

[19] Rothlisberger U. and Carloni P. *Theoretical and computational chemistry*, chapter Simulations of enzymatic systems: Perspectives from Car-Parrinello Molecular dynamics simulations., pages 215–251. Politzer P. and Maksic Z.B., 2001.

[20] L. Guidoni, Maurer P., Piana S., and Rothlisberger U. Hybrid Car-Parrinello/Molecular Mechanics Modelling of Transition Metal Complexes: Structure, Dynamics and Reactivity. *Quant. Struct.-Act. Rel.*, 21:119–127, 2002.

[21] Small E.J., Figlin R., Petrylak D., Vaughn D.J., Sartor O., Horak I., Pincus R., Kremer A., and Bowden C. A phase II pilot study of KW-2189 in patients with advanced renal cell carcinoma. *Invest New Drugs.*, 18:193–197, 2000.

[22] Vischer, Zamenhof, and Chargaff. Microbial nucleic acids: The Desoxypentose Nucleic acids of avian tubercle bacilli and yeast. *J.Biol.Chem*, 177:429–438, 1949.

[23] Dickerson R.E. and Chiu T.K. Helix Bending as a Factor in Protein/DNA Recognition. *Biopolymers*, 44:361–404, 1997.

[24] Olson W.K., Gorin A.A., Lu X.J., Hock L.M., and Zhurkin V.B. *Proc.Natl.Acad.Sci.USA*, 95:11163–11168, 1998.

[25] Commission on Biochemical Nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. *Eur.J.Biochem.*, 17:193–201, 1970.

[26] Commission on Biochemical Nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. *Eur.J.Biochem.*, 131:9–15, 1982.

[27] Dickerson R.E., Bansal M., Calladine C.R., Diekmann S., Hunter W.N., Kennard O., von Kitzing E., Lavery R., Nelson H.C.M., Olson W.K., Saenger W., Shakked Z., Sklenar H., Soumpasis D.M., Tung C.S., Wang A.H.J., and Zhurkin V.B. Nomenclature of nucleic acids. *EMBO J.*, 8:1–4, 1989.

[28] Olson W.K., Bansal M., Burley S.K., Dickerson R.E., Gerstein M., Harvey S.C., Heinemann U., Lu X.J., Neidle S., Shakked Z., Slkenar H., Suzuki M., Tung C.S. Westhof E., Wolberger C., and Berman H.M. A standard reference frame for the description of nucleic acid base-pair geometry. *J.Mol.Biol.*, 313:229–237, 2001.

[29] Moss G.P. Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents. http://www.chem.qmul.ac.uk/iupac/misc/naabb.html.

[30] Image Library of Biological Macromolecules IMB JENA. Nucleic Acid Nomenclature and Structure. http://www.imb-jena.de/ImgLibDoc/nana/IMAGE_NANA.html.

[31] *Principles of Nucleic Acid Structure*, chapter 2. Springer Verlag, New York, Berlin Heidelberg Tokyo, 1984.

[32] Lavery R. and Sklenar H. The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids. *J.Biomol.Struct.Dyn.*, 6:63–91, 1988.

[33] Lavery R. and Sklenar H. Defining the Structure of Irregular Nucleic Acids: Conventions and Principles. *J.Biomol.Struct.Dyn.*, 6:655–667, 1989.

[34] Yanagi K., Prive G.G., and Dickerson R.E. Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J.Mol.Biol.*, 217:201–204, 1991.

[35] *Nucleic Acids in Chemistry and Biology*. Oxford University Press, 2 edition, 1996.

[36] Cheatham T.E.III and Kollman P.A. Molecular Dynamics Simulation of Nucleic Acids. *Annu.Rev.Phys.Chem.*, 51:435–471, 2000.

[37] Cheatham T.E.III and Young M.A. Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations, and Promise. *Biopolymers*, 56:232–256, 2001.

[38] MacKerell A.D.Jr., Banavali N., and Foloppe N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56:257–265, 2000.

[39] Auffinger P. and Westhof E. Water and ioin binding around RNA and DNA (C,G) oligomers. *J.Mol.Biol.*, 300:1113–11131, 2000.

[40] Cheatham T.E.III, Miller J.L., Fox T., Darden D.A., and Kollman T.A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method leads to Stable Trajectories of DNA, RNA, and Proteins. *J.Am.Chem.Soc.*, pages 4193–4194, 1995.

[41] Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M.Jr., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., and Kollman P.A. A

second-generation force field for the simulation of proteins and nucleic acids. *J.Chem.Phys.*, 117:5179–5197, 1995.

[42] Cheatham T.E.III, Cieplak P., and Kollman P.A. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J.Biomol.Struct.Dynam.*, 16:845–862, 1999.

[43] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., and Haak J.R. Molecular dynamics with coupling to an external bath. *J.Chem.Phys.*, 81:3684–3690, 1984.

[44] Nosé S. A molecular dynamics method for simulations in the canonical ensemble. *Mol.Phys.*, 52:255–268, 1984.

[45] Hoover W.G. Canonical dynamics: equilibrium phase-space distributions. *31*, 31:1695–1697, 1985.

[46] van Gunsteren W.F. and Berendsen H.J.C. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angew.Chem.Int.Ed.Engl.*, 29:992–1023, 1990.

[47] Darden T.A., York D.M., and Pedersen L.G. Particle Mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J.Chem.Phys.*, 98:10089–10092, 1993.

[48] York D.M., Darden T.A., and Pedersen L.G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and trucated list methods. *J.Chem.Phys.*, 99(10):8345–8348, 1993.

[49] Hohenberg P. and Kohn W. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964.

[50] Kohn W. and Sham L.J. Self-consistent equations including exchange and correlation effects. *Phys.Rev.*, 140:A1133–A1138, 1965.

[51] Langreth D.C. and Perdew J.P. The exchange-correlation energy of a metallic surface. *Solid State Comm.*, 17:1425–, 1975.

[52] Becke A.D. Density-functional exchange-energy approximation with corrct asymptotic behavior. *Phys.Rev.A*, 38:3098–3100, 1988.

[53] Lee C., Yand W., and Parr R.G. Development of the Colle-Salvetti correlation-energy formulainto a functional of the electron density. *Phys.Rev.B*, 37:785–789, 1988.

[54] Sulpizi L., Folkers G., Rothlisberger U., Carloni P., and Scapozza L. Applications of Density Functional Theory-Based Methods in Medicinal Chemistry. *Quant. Struct.-Act. Rel.*, 21:173–181, 2002.

[55] Colombo M., Guidoni L., Laio A., Magistrato A., Maurer P., Piana S., Röhrig U., Spiegel K., Sulpizi M., VandeVondele J., Zumstein M., and Rothlisberger U. Hybrid QM/MM Car-Parrinello Simulations of Catalytic and Enzymatic Reactions. *CHIMIA*, 56:13–19, 2002.

[56] Carloni P., Rothlisberger U., and Parrinello M. The Role and Perspective of Ab-initio Molecular Dynamics in the Study of Biological Systems. *Acc. Chem. Res. 35*, 35:455–464, 2002.

[57] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, 1985.

[58] Bouckaert L.P., Smoluchowski R., and Wigner.E. Theory of Brillouin zones and symmetry properties of wave functions in crystals. *Phys.Rev.*, 50:58, 1936.

[59] Monkhorst H.J. and D. Pack J.D. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13:1588, 1976.

[60] Chadi D.J. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 16:1746, 1977.

[61] Evarestov K.A. and Smirnov V.P. Special points of the Brillouin zone and their use in the solid state theory. *Phys. Stat. Sol*, 119:9, 1983.

[62] Marzari N. and Vanderbilt D. Maximally localized generalized Wannier functions for composite energy bands. *Phys.Rev.B*, 56, 1997.

[63] Gogonea V., Westerhoff L., and Merz K.M. Jr. QM/QM Methods: 1. A Divide and Conquer Strategy for Solving the Schrödinger Equation for Large Molecular Systems using a Composite Density Functional - Semiempirical Hamiltonian. *J. Chem. Phys*, 113:5604–5613, 2000.

[64] Bala P., Grochowski P. AND. Lesyng B., and McCammon J.A. Quantum-Classical Molecular Dynamics Simulations of Proton Transfer Processes in Molecular Complexes and in Enzymes. *J.Phys.Chem.*, 100:2535–2545, 1996.

[65] Hutter J., Alavi A., Deutsch T., Silvestri W., Parrinello M., Rothlisberger U., Marx D., Focher P., Tuckerman M., Andreoni W., Curioni A., Fois E., Giannozzi P., Sebastiani D., Laio A., VandeVondele J., Seitsonen A., and Billeter S. *CPMD*. MPI fuer Festkoerperforschung Stuttgard and IBM Zurich Research Laboratory (3.5.), 3.5. edition, 2000.

[66] van Gunsteren W.F. *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* Zuerich, Groningen, 1996.

[67] Baerends E.-J., Autschbach J., Berces A., Bo C., Boerrigter P.M., Cavallo L., Chong D.P., Deng L., Dickson R.M., Ellis D.E., Fan L., Fischer T.H., Fonseca Guerra C., and van Gisbergen S.J.A. ADF2002.3,SCM, Theoretical Chemistry. *Phys.Rev.B*, 37:785–789, 1988.

[68] Schreckenbach G.; Wolff S.K.; Ziegler T. NMR Shielding Calculations across the Periodic Table: Diamagnetic Uranium Compounds. 1. Methods and Issues. *J.Phys.Chem.A*, 104:8244–8255, 2000.

[69] G. Schreckenbach. NMR shielding calculations across the periodic table: diamagnetic uranium compounds. 2. Ligand and metal NMR. *Inorg.Chem*, 41:6560, 6572 2002.

[70] van Lenthe E., Baerends E.J., and Snijders J.G. Relativistic regular two-component Hamiltonians. *J.Chem.Phys.*, 99:4957–4610, 1993.

[71] van Lenthe E., Baerends E.J., and Snijders J.G. Construction of the foldy-wouthuysen transformation and solution of the dirac equation using large components only. *J.Chem.Phys.*, 105:2373–2377, 1996.

[72] Reedijk J. Bioinorganic Chemistry Special Feature: New clues for platinum antitumor chemistry: Kinetically controlled metal binding to DNA. *Proc.Natl.Acad.Sci.U.S.A*, 100:3611–3616, 2003.

[73] Comess K.M. and Lippard S.J. *Molecular Aspects of Anticancer Drug-DNA Interactions*, pages 134–168. Macmillan Press, London, 1993.

[74] Lippert.B. *Cisplatin: Chemistry and Biochemistry of a leading anticancer drug.* Viley-VCH, 1. edition, 1999.

[75] G.L. Eichhorn and L.G. Marzilli. *Metal Ions in Genetic Information Transfer.* Elsevier, New York, 1981.

[76] F. Legendre, J. Kozelka, and J.C. Chottard. GG versus AG Platination: A Kinetic Study on Hairpin-Stabilized Duplex Oligonucleotides. *Inorg.Chem.*, 37:3964–3967, 1998.

[77] Lippert B. Platinum nucleobase Chemistry. *Prog.Inorg.Chem.*, 37:1–9, 1989.

[78] Reedijk J. Why does Cisplatin reach Guanine-N7 with competing S-donor ligands available in the cell? *Chem.Rev*, 99:2499–2510, 1999.

[79] Fichtinger-Schepman A.M., van der Veer J.L., den Hartog J.H., Lohman P.H., and Reedijk J. Adducts of the antitumor drug cis-diaminedichloroplatinum(II) with DNA: formation, identification and quantitation. *Biochem.*, 24:707–713, 1985.

[80] Bradley L.J.N., Yarema K.J., Lippard S.J., and Essigman J.M. Mutagenicity and Genotoxicity of the Major DNA Adduct of the Antitumor Drug cis-diamminedichloroplatinum(II). *Biochemistry*, 32:982–988, 1993.

[81] Burnouf D., Daune M., and Fuchs R.P.P. Spectrum of cisplatin-induced mutations in *escherichia coli*. *Proc.Natl.Acad.Sci.USA*, 84:3758–3762, 1992.

[82] Comess K.M., Burstyn J.N., Essigmann J.M, and Lippard S.J. Replication Inhibition and Translesion Synthesis on Templates Containing Site-Specifically placed cis-diamminedichloroplatinum(II) DNA Adducts. *Biochemistry*, 31:3975–3990, 1992.

[83] Page J.D., Husain I., Sancar A., and Cheney S.G. Effect of the Diaminocyclohexane carrier ligand on Platinum Adduct Formation, Repair and Lethality. *Biochemistry*, 29:3975–3990, 1990.

[84] Zamble D.B., Mu D., Reardon J.T., Sancar A., and Lippard S.J. Repair of cisplatin-DNA adducts by the mammalian excision nuclease. *Biochemistry*, 35:10004–10013, 1996.

[85] P.M. Takahara, A.C. Rosenzweig, and S.J. Frederick, C.A.AND Lippard. Crystal structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin. *Nature*, 377:649–652, 1995.

[86] S.E. Sherman, D. Gibson, A.H. Wang, and S.J. Lippard. X-ray structure of the major adduct of the anticancer drug cisplatin with DNA: cis-[pt(nh$_3$)$_2$(d(pgpg))]. *Science*, 230:412–417, 1985.

[87] E.R. Jamieson and S.J. Lippard. Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chem.Rev.*, 99:2467–2498, 1999.

[88] P.M. Takahara, C.A. Frederick, and S.J. Lippard. Crystal Structure of the Anticancer Drug Cisplatin Bound to Duplex DNA. *J.Am.Chem.Soc.*, 118:12309–12321, 1996.

[89] Yang D., van Boom S.S.G.E., Reedijk J., van Boom J.H., and Wang A.H.-J. Structure and Isomerization of an Intrastrand Cisplatin-Cross-Linked Octamer DNA Duplex by NMR Analysis. *Biochem.*, 34:12912–12920, 1995.

[90] A. Gelasco and S.J. Lippard. NMR solution structure of a DNA dodecamer duplex containing a cis- diammineplatinum(II) d(GpG) intrastrand cross-link, the major adduct of the anticancer drug cisplatin. *Biochem.*, 37:9230–9239, 1998.

[91] U.M. Ohndorf, Rould M.A., Q. He, Pabo C.O., and S.J. Lippard. Basis for recognition of cipslatin-modified DNA by high-mobility-group proteins. *Nature*, 399:708–712, 1999.

[92] J. Kasparkova, O. Delalande, M. Stros, M.A. Elizondo-Riojas, M. Vojtiskova, J. Kozelka, and V. Brabec. Recognition of DNA interstrand cross-link of antitumor cisplatin by HMGB1 protein. *Biochem.*, 42:1234–1244, 2003.

[93] Y. Jung and S.J. Lippard. Nature of full-length HMGB1 Binding to Cisplatin-Modified DNA. *Biochem.*, 42:2664–2671, 2003.

[94] Yoshioka K., Saito K., Tanabe T., Yamamoto A., Ando Y., Nakamura Y., Shirakawa H., and Yoshida M. Differences in DNA Recognition and confomational change activity between boxes A and B in HMG2 Protein. *Biochem.*, 38:589–595, 1999.

[95] Q. He, U.M. Ohndorf, and S.J. Lippard. Intercalating residues determine the mode of HMG1 domains A and B binding to cisplatin-modified DNA. *Biochem.*, 39:14426–14435, 2000.

[96] S.M. Cohen, Mikata Y., He Q., and S.J. Lippard. HMG-Domain Protein Recognition of Cisplatin 1,2-intrastrand d(GpG) Cross-Links in Purine-Rich Sequence Contexts. *Biochem.*, 39:11771–11776, 2000.

[97] E.R. Jamieson, M.P. Jacobson, C.M. Barnes, C.S. Chow, and S.J. Lippard. Structural and kinetic studies of a cisplatin-modified DNA icosamer binding to HMG1 domain B. *J.Biol.Chem.*, 274:12346–12354, 1999.

[98] S.U. Dunham and S.J. Lippard. DNA sequence context and protein composition modulate HMG-domain protein recognition of cisplatin-modified DNA. *Biochem.*, 36:11428–11336, 1997.

[99] E.A. Pasheva, I. Ugrinova, N.C. Spassovska, and I.G. Pashev. The binding affinity of HMG1 protein to DNA modified by cis-platin and its analogs correlates with their antitumor activity. *Int.J.Biochem.Cell Biol.*, 34:87–92, 2002.

[100] K.M. Comess, J.N. Burstyn, J.M. Essigmann, and S.J. Lippard. Replication inhibition and translesion synthesis on templates containing site-specifically placed cis-diamminedichloroplatinum(II) DNA adducts. *Biochem.*, 31:3975–3990, 1992.

[101] Chu G. Cellular responses to cisplatin. The roles of DNA-binding proteins and DNA repair. *J.Biol.Chem.*, 269:787–790, 1994.

[102] V.M. Gonzalez, M.A. Fuertes, C. Alonso, and J.M. Perez. Is cisplatin-induced cell death always produced by apoptosis? *Mol.Pharmacol.*, 59:657–663, 2001.

[103] Bose R.N. Biomolecular targets for platinum antitumor drugs. *Mini Rev.Med.Chem.*, 2:103–111, 2002.

[104] Marzilli L.G., Saad J.S., Kuklenyik Z., Keating K.A., and Xu Y. Relationship of Solution and Protein-Bound Structures of DNA Duplexes with the Major Intrastrand Cross-Link lesions formed on Cisplatin Binding to DNA. *J.Am.Chem.Soc.*, 123:2764–2770, 2001.

[105] S.U. Dunham, Turner C.J., and S.J. Lippard. Solution structure of a DNA duplex containing a nitroxide spin-labeled platinum d(GpG) intrastrand cross-link refined with NMR derived long-range electron-proton distance restraints. *J.Am.Chem.Soc.*, 120:5395–5406, 1998.

[106] A. Pelmenschikov, I. Zilberberg, J. Leszczynski, A. Famulari, M. Sironi, and M. Raimondi. cis-[pt(nh$_3$)$_2$]$^{2+}$ coordination to the N7 and O6 sites of a guanine-cytosine pair: disruption of the Watson-Crick H-bonding pattern. *Chem.Phys.Lett.*, 314:496–500, 1999.

[107] Parkinson J.A., Chen Y., Murdoch P.d.S., Guo Z., Berners-Price S.J., Brown T., and Sadler P.J. *Chem.-Eur.J.*, 6:3636–3644, 2000.

[108] M.A. Elizondo-Riojas and J. Kozelka. Unrestrained 5 ns molecular dynamics simulation of a cisplatin-DNA 1,2- GG adduct provides a rationale for the NMR features and reveals increased conformational flexibility at the platinum binding site. *J.Mol.Biol.*, 314:1227–1243, 2001.

[109] M.A. Elizondo-Riojas, Gonnet F., Auge-Barrere-Mazouat P., Allain F., Berges J., Attias R., Chottard J.-C., and Kozelka J. *Molecular Modeling of Platinum Complexes with Oligonucleotides:Methodological Lessons and Structural Insights*, pages 131–160. Kluwer Academic Publishers, Dordrecht, 1997.

[110] Yao S., Plastaras J.P., and Marzilli L.G. A Molecular Mechanics AMBER-Type Force Field for Modeling Platinum Complexes of Guanine Derivatives. *Inorg.Chem.*, 33:6061–6077, 1994.

[111] F. Herman, J. Kozelka, V. Stoven, E. Guittet, J.P. Girault, T. Huynh-Dinh, J. Igolen, J.Y. Lallemand, and J.C. Chottard. A d(GpG)-platinated decanucleotide duplex is kinked. An extended NMR and molecular mechanics study. *Eur.J.Biochem.*, 194:119–133, 1990.

[112] J. Kozelka. Molecular modeling of transition metal complexes with nucleic acids and their constituents. *Met.Ions.Biol.Syst.*, 33:1–28, 1996.

[113] M.A. Fuertes, Castilla J., Alonso C., and Perez J.M. Novel concepts in the development of platinum antitumor drugs. *Curr.Med.Chem.Anti-Canc.Agents*, 2:539–551, 2002.

[114] Banci L.C.P. *Molecular Modeling and dynamics of bioinorganic compounds.* Kluwer Academic Publishers, Dordrecht,Boston, 1997.

[115] V. Brabec, M. Sip, and M. Leng. DNA conformational change produced by the site-specific interstrand cross-link of trans-diamminedichloroplatinum(II). *Biochem.*, 32:11676–11681, 1993.

[116] Carloni P., Sprik M., and Andreoni W. Key Steps of the cis-Platin-DNA Interaction:Density Functional Theroy-Based Molecular Dynamics Simulations. *J.Phys.Chem.B*, 104:823–835, 2000.

[117] Carloni P. and Andreoni W. Platinum-Modified Nucleobase Pairs in the Solid State: A Theoretical Study. *J.Phys.Chem.*, 100:17797–17800, 1996.

[118] Pearlman D.A., Case D.A., Caldwell J.W., Ross W.S., T E Cheatham III, De-Bolt S., Ferguson D., Seibel G.L., and Kollman P.A. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free enery calculations to simulate the structural and energetic properties of molecules. *Comp.Phys.Commun.*, 91:1–41, 1995.

[119] Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., and Klein M.L. *J.Chem.Phys.*, 79:926–935, 1983.

[120] Trouillier N. and Martins J.L. Efficient pseudopotentials for plane-wave calculation. *Phys.Rev.B*, 43:1943–2006, 1991.

[121] Kleinman L. and Bylander D.M. Efficacious Form for Model Pseudopotentials. *Phys.Rev.Letter*, 48:1425–1428, 1982.

[122] Barnett R.N. and Landman U. Born-Oppenheimer molecular dynamics simulations of finite systems. Structure and dynamics of $(H_2O)_2$. *Phys.Rev.B.*, 48:2081–2097, 1993.

[123] Bouten R.AND Baerends E.J.AND van Lenthe E., Visscher L., Schreckenbach G., and Ziegler T. Relativistic Effects for NMR shielding constants in transition metal oxides using the Zeroth-Order Regular Approximation. *J.Phys.Chem.A*, 104:5600–5611, 2000.

[124] Gilbert T.M. and Ziegler T. Prediction of [195]Pt NMR chemical shifts by Density Functional Theroy Computations: The Importance of magnetic coupling and relativistic effects in explaining trends. *J.Phys.Chem.A*, 103:7535–7543, 1999.

[125] Perdew J.P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys.Rev.B*, 33, 1986.

[126] Perdew J.P. Accurate density functional for the energy: Real-space cutoff of the gradient expansion for the exchange hole. *Phys.Rev.Lett.*, 55, 1985.

[127] Wolff S.K.; Ziegler T.; van Lenthe E.; Baerends E.-J. Density functional calculations of nuclear magnetic shieldings using the zeroth-order regular approximation (ZORA) for relativistic effects: ZORA nuclear magnetic resonance. *J.Chem.Phys.*, 110:7689–7698, 1999.

[128] Bancroft D.P., Lepre Ch.A., and S.J. Lippard. [195]Pt NMR Kinetic and Mechanistic Studies of cis-and trans-Diamminedichloroplatinum(II) Binding to DNA. *J.Am.Chem.Soc.*, 112:6860–6871, 1990.

[129] Miller S.K. and Marzilli L.G. Interaction of Platinum Antitumor Agents with Guanine Nucleosides and Nucleotides. [195]Pt and 1H NMR Spectroscopic Characterization of Compound III. *Inorg.Chem.*, 24:2421–2425, 1985.

[130] Brugé F., Bernasconi M., and Parrinello M. Ab initio Simulation of Rotational Dynamics of Solvated Ammonium Ion in Water. *J.Am.Chem.Soc.*, 121:10883–10888, 1999.

[131] J.V. Burda, J. Sponer, and J. Leszczynski. The influence of square planar platinum complexes on DNA base pairing. An ab initio DFT study. *Phys.Chem.Chem.Phys.*, 3:4404–4411, 2001.

[132] Ichimura M., Ogawa T., Takahashi K., Kobayashi E., Kawamoto I., Yasuzawa T., Takahashi I., and Nakano H. Duocarmycin sa, a new antitumor antibiotic from streptomyces sp. *J.Antibiot.*, 43:1037–1038, 1990.

[133] Boger D.L. and Garbaccio R.M. Catalysis of the CC-1065 and Duocarmycin DNA Alkylation Reaction: DNA Binding induce conformational change in the agent results in activation. *Bioorg.Med.Chem*, 5:263–276, 1997.

[134] Eis P.S., Smith J.A., Rydzewski J.M., Case D.A., Boger D.L., and Chazin W.J. High Resolution Solution Structure of a DNA Duplex Alkylated by the Antitumor Agent Duocarmycin SA. *J.Mol.Biol*, 272:237–252, 1997.

[135] Schnell J.R., Ketchem R.R., Boger D.L., and Chazin W.J. Binding-Induced Activation of DNA Alkylation by Duocarmycin SA: Insights from the Structure of an Indole Derivated-DNA Adduct. *J.Am.Chem.Soc.*, 121:5645–5652, 1999.

[136] Katahira M., Sugeta H., and Kyogoku Y. A new model for the bending of DNAs containing the oligo(da) tracts. *Nucleic Acids Res.*, 18:613–618, 1990.

[137] Nadeau J.G. and Crothers D.M. Structural basis for DNA bending. *Proc.Natl.Acad.Sci.U.S.A*, 86:2622–2626, 1989.

[138] Goodsell D.S., Kaczor, Grzeskowiak M., and Dickerson R.E. The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps. *J.Mol.Biol.*, 239:79–96, 1994.

[139] Mack D.R., Chiu T.K., and Dickerson R.E. Intrinsic Bending and Deformability at the T-A Step of CCTTTAAAGG: A Comparative Analysis of T-A and A-T Steps within A-tracts. *J.Mol.Biol.*, 312:1037–1049, 2001.

[140] McConnell K.J. and Beveridge D.L. Molecular Dynamics Simulations of B'-DNA: Sequence Effects on A-tract-induced Bending and Flexibility. *J.Mol.Biol.*, 314:23–40, 2001.

[141] Lin C.H. and Patel D.J. Solution structure of the covalent Duocarmycin a-DNA Duplex Complex. *J.Mol.Biol.*, 248:162–179, 1995.

[142] Warpehoski M.A. and Harper D.E. Enzyme-like rate acceleration in the DNA minor groove, cyclopropylpyrroloindoles as mechanism-based inactivators of DNA. *J.Am.Chem.Soc.*, 117:2951–2952, 1995.

[143] Boger D.L. and Garbaccio R.M. Shape-Dependent Catalysis: Insights into the Source of Catalysis for the CC-1065 and Duocarmycin DNA alkylation reaction. *Acc.Chem.Res.*, 32:1043–1052, 1999.

[144] Sun D., Lin C.H., and Hurley L.H. A-tract and (+)-cc-1065-induced bending of DNA. comparison of structural features using non-denaturing gel analysis, hydroxyl-radical footprinting, and high-field nmr. *Biochem.*, 32:4487–4495, 1993.

[145] Thompson A.S., Sun D., and Hurley L.H. Monoalkylation and Cross-Linking of DNA by Cyclopropapyrroloindoles Entraps Bent ant Straight Forms of A-Tracts. *J.Am.Chem.Soc.*, 117:2371–2372, 1995.

[146] Boger D.L., Bollinger B., Hertzog D.L., Johnson D.S., Cai H., Mesini P., Garbaccio R.M., Jin Q., and Kitos P.A. Reversed and sandwiched analogs of

duocarmcyin sa: Establishment of the origin of the sequence-selective Alkylation of DNA and new insights into the source of catalysis. *J.Am.Chem.Soc.*, 119:4987–4998, 1997.

[147] Kirschner K.N., Lee M., Stanley R.C., and Bowen J.P. Density Functional and ab initio studies on N-Acetyl-Duocarmycin SA: Insight into its DNA interaction properties. *Bioorg.Med.Chem.*, 8:329–335, 2000.

[148] Lamm G., Wong L., and Pack G.R. DNA-Mediated Acid Catalysis: Calculations of the Rates of DNA-Catalyzed Hydrolyses of Diol Epoxides. *J.Am.Chem.Soc*, 118:3325–3331, 1996.

[149] Warpehoski M.A. and Harper D.E. Acid-dependent electrophilicity of cyclopropylpurroloindoles. nature's masking strategy for a potent DNA alkylator. *J.Am.Chem.Soc.*, 116:7573–7580, 1994.

[150] Jayaram B., Sharp K.A., and Honig B. The electrostatic potential of B-DNA. *Biopolymers*, 28:975–993, 1989.

[151] Lamm G. and Pack G.R. Acidic Domains Around Nucleic Acids. *Proc.Natl.Acad.Sci.U.S.A.*, 87:9033–9036, 1990.

[152] Rothlisberger U. 15 years of Car-Parrinello Simulations in Physics, Chemistry and Biology. In Leszczynski J., editor, *Computational Chemistry: Reviews of Current Trends*, volume 6, pages 33–68. World Scientific, 2001.

[153] Guidoni L., Spiegel K., Zumstein M., and Rothlisberger U. Green Oxidation Catalysts: Computational Design of High-Efficiency Models of Galactose Oxidase. *Angew.Chem.*, 116:3348–3351, 2004.

[154] Aaqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. *J.Phys.Chem.*, 94:8021–8024, 1990.

[155] Wang J., Wolf R.M., Caldwell J.W., and Kollman P.A. A Development and testing of a general Amber force field. *J.Comp.Chem.*, 25:1157–1174, 2004.

[156] Parrinello M. and Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J.Appl.Phys.*, 52:7182–7190, 1981.

[157] Berendsen H.J.C., van der Spoel D., and Vandrunen R. Gromacs: a message-passing parallel molecular-dynamics implmentation. *Comp.Phys.Comm*, 91:43–56, 1995.

[158] Lindahl E., Hess B., and van der Spoel D. GROMACS 3.0: a packag for molecular simulation and trajectroy analysis. *J.Mol.Mod.*, 7:306–317, 2001.

[159] *Medicinal Quantum Chemistry*, chapter Advances in Density Functional Based Modelling Techniques: Recent Extensions of the Car-Parrinello Approach. Methods and Principles in Medicinal Chemistry. Wiley-VCH, Weinheim, 2003.

[160] Carter E.A., Ciccotti G., Hynes J.T., and Kapral R. *Chem.Phys.Lett.*, 156:472, 1989.

[161] Ivanov I. Klein M.L. Deprotonation of a Histidine Residue in Aqueous Solution Using Constrained Ab Initio Molecular Dynamics. *J.Am.Chem.Soc.*, 124:13380–13381, 2002.

[162] *Protein Physics: A Course of Lectures*. Soft Condensed Matter Complex Fluids and Biomaterials. Academic Pres, New York, 2002.

[163] Martyna G.J. and Tuckerman M.E. A reciprocal space based method for treating long-range interactions in ab initio and force-field based calculations in clusters. *J.Chem.Phys.*, 110:2810–2821, 1999.

[164] Rocchia W., Alexov E., and Honig B. Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J.Phys.Chem.B*, 105:6507–6514, 2001.

[165] Gilson M.K., Sharp K., and Honig B. Calculating the Electrostatic Potential of Moleucles in Solution: Method and error assessment. *J.Comp.Chem.*, 9:327–335, 1987.

[166] Honig B. and Nicholls A. Classical electrostatics in biology and chemistry. *Science*, 268:1144–1149, 1995.

[167] M.A. Young, Ravishanker G., Beveridge D.L., and Berman H.M. Analysis of local helix bending in crystal structure of DNA oligonucleotides and DNA-protein complexes. *Biophys. J.*, 68:2454–2468, 1995.

[168] Boger D.L., Johnson D.S., and Yun W. (+)-and (-)-Duocarmycin sa and (+)- and(-)-n-boc-dsa DNA Alkylation properties. Alkylation site models that accommodate the offset at-rich adenine n3 Alkylation selectivity of the enantiomeric agents. *J.Am.Chem.Soc.*, 116:1635–1656, 1994.

[169] Madhumalar A. and Bansal M. Structural Insights into the Effect of Hydration and Ions on A-tract DNA: A Molecular Dynamics Study. *Biophys.J.*, 85, 2003.

[170] Boger D.L., Goldberg J., and McKie J.A. A comparative study of the solvolysis reactivity, regioselectrivity, and stereochemistry of the Duocarmycin a and sa Alkylation subunits. *Bioorg.Med.Chem.Lett.*, 6:1955–1960, 1996.

[171] Parrish J.P., Kastrinsky D.B., and Boger D.L. Synthesis and X-ray Analysis of an unprecedented and stable 2-aza-4,4-spirocyclopropacyclohexadienone. *Org.Lett.*, 5:2577–2579, 2003.

[172] Boger D.L. Machiya K., Hertzog D.L., Kitos P.A., and Holmes D. Total synthesis and preliminary evaluation of (+)- and ent-(-)-duocarmycin sa. *J.Am.Chem.Soc.*, 115:9025–9036, 1993.

[173] Asai A., Yano K., Mizukami T., and Nakano H. Characterization of a Duocarmycin-DNA Adduct-recognizing Protein in Cancer Cells. *Cancer Res.*, 59:5417–5420, 1999.

[174] Lunch B.J. and Truhlar D.G. How well can Hybrid Density Functional Methods predict Transition State Geometries and Barrier Heights? *J.Phys.Chem.A*, 105:2936–2941, 2001.

[175] Gritsenko O.V., Ensing B., Schipper P.R.T., and Baerends E.J. Comparison of the accurate Kohn-Sham solution with the generalized gradient approximations (GGA's) for the $s_2 2$ reaction $f^- + ch_3 f \rightarrow fch_3 + f^-$: a qualtitative rule to predict successs or failure of GGA's. *J.Phys.Chem.A*, 104:8558–8565, 2000.

[176] Ensing B., Meijer E.J., Blöchl P.E., and Baerends E.J. Solvation effects on the $S_2 2$ reaction between CH_3Cl and Cl^- in water. *J.Phys.Chem.A*, 105:3300–3310, 2001.

[177] Raugei S., Cardini G., and Schettino V. An ab-initio molecular dynamics study of the $S_2 2$ reaction $Cl^- + CH_3Br \rightarrow CH_3Cl + Br^-$. *J.Chem.Phys.*, 24:10887–10894, 1999.

[178] Cimino P., Improta R., Bifulco G., Riccio R., Gomez-Paloma L., and Barone V. Nucleophilic Cyclopropane Ring Opening in Duocarmycin SA Derivatives by Methanol under Acid Conditions: A Quantum Mechanical Study in the Gas-Phase and in Solution. *J.Org.Chem.*, 69:2816–2824, 2004.

[179] Sponer J., Leszczynski J., and Hobza P. Electronic Properties, Hydrogen Bonding, Stacking, and Cation Binding of DNA and RNA Bases. *Biopolymers*, 61:3–31, 2001.

[180] Sponer J., Gabb H.A., Leszczynski J., and Hobza P. Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum chemical study. *Biophys. J.*, 73:76–87, 1997.

[181] Sponer J., Jurecka P., and Hobza P. Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J.Am.Chem.Soc.*, 126:10142–10151, 2004.

[182] Popelier P.L. and Joubert L. The elusive atomic rationale for DNA base pair stability. *J.Am.Chem.Soc.*, 124, 2002.

[183] Dinner A.R., Blackburn G. M., and Karplus M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature*, 413:752 – 755, 2001.

[184] Perdew J.P., Burke K., and Ernzerhof M. Generalized Gradient Approximation made simple. *Phys.Rev.Lett.*, 77:3865–3868, 1996.

[185] Adamo C. and Barone V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J.Chem.Phys.*, 110:6158–6170, 1998.

[186] A. Laio and M. Parrinello. Escaping Free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99:12562–12566, 2002.

[187] Spingler B., Whittington D.A., and Lippard S.J. 2.4 a Crystal Structure of an Oxaliplatin 1,2-d(GpG) Intrastrand Cross-Link in a DNA Dodecamer Duplex. *Inorg.Chem.*, 40:5596–5602, 2001.

[188] Wu Y., Pradhan P., Havener J., Boysen G., Swenberg J.A., Campbell S.L., and Chaney S.G. NMR solution structure of an oxaliplatin 1,2-d(GG) intrastrand cross-link in a DNA dodecamer duplex. *J.Mol.Biol.*, 341:1251–1269, 2004.