Molecular recognition of ligands targeting DNA: a computational approach

Thesis submitted for the degree of

Doctor Philosophiæ

Candidate:
Attilio Vittorio Vargiu

Supervisors:
Prof. Paolo Carloni
Prof. Paolo Ruggerone
Dr. Alessandra Magistrato

January, 2008

Contents

IV   Conclusions and Perspectives   143

Appendices   147

A   The structure of DNA   147

B   Statistical Mechanics of non-Hamiltonian systems   155

C   Ewald sum method   159

Bibliography   166

Part I

Introduction and Methods

Chapter 1

Introduction

1.1 DNA and Cancer

Cancerous cells divide without control, growing also in the absence of the proper signals and ignoring the pressure from neighboring cells. The search of a cure against cancer progression is one of the big challenges in medicine, mainly because of two relevant issues [1]. The first is that several factors appear to be related with (and be the cause of) the disease. These include:

- The alteration in the DNA sequence of cancer-related genes [1]: *i)* tumour suppressor genes (e.g. p53, RB and APC), which encode for proteins restraining cells' ability to survive and divide; *ii)* oncogenes (e.g. BRAF, c-fos, cerbb3), whose mutation causes oncoproteins to become hyperactive prompting the cell to grow ignoring stopping signals.

- The inactivation of one or more genes encoding for proteins that synthesize or repair the DNA [2]. This leads to the introduction of a high number of random mutations during cellular replication, which remain unrepaired and accumulate, and eventually hit cancer-related genes short-circuiting apoptotic mechanisms.

- The silencing of "master" genes regulating cell division [3], which endowes daughter cells with mangled chromosomes (wrong in number, cut, mixed or containing extra segments).[1] These aneuploid cells can accumulate chromosomal aberrations with generations, finally altering the relative amounts of thousands of genes and eventually the normal dosage of tumor suppressor and on-

---

[1]It is worth to point out that the search for a restricted set of oncogenes, tumour suppressor genes, or "master genes" has failed so far. For example, in the 2003, 116 oncogenes and 15 tumour suppressor genes were already identified, inducing scientist to not exclude *a priori* the idea that each tumour is unique in the pattern of its genetic disarray.

coproteins. It is not clear if aneuploidy is related to mutations in genes or it is itself a cause of cancer [4].

- The epigenetic modifications, such as DNA methylation and histone-modifications [1, 5].

The second and most life-threating reason is that primary tumours often evolve in metastases, for which no effective cure exists (they are cause of nine of every ten deaths from cancer). Metastatic cells are in fact able to enter blood and limphatic vessels, spreading out the malignancy in different parts of the body. This renders uneffective the treatment of the primary tumour.

The most common strategies for the cure of cancer are surgery, chemotherapy, radiotherapy, immunotherapy and others. The choice depends on the type and the stage of cancer and on the status of the patient. Obviously only localized tumours can be efficaciously removed by surgical procedures (known examples are breast and prostate cancer), which besides need to be extremely precise to avoid recurrence phenomena. The preferred alternative to surgery has become today chemotherapy [6, 7, 8], not only for the treatment of leukaemias and lymphomas, which are spread in the body and thus untractable by surgery and radiotherapy, but for the cure of almost all kinds of cancers. In addition, chemotherapy is often used after surgical intervention to lower the risk of recurrence.

Most chemotherapeutics in use today rely on a cytotoxic action, which is a direct consequence of their reversible or covalent binding to cellular DNA [9, 7, 10, 11]. This affects the propensity of DNA to interact with proteins [12, 13], eventually inhibiting vital processes such as replication, transcription, repairing enzymes action, and it may lead ultimately to cell apoptosis (Figure 1.1) [12, 14, 15, 16, 17]. Figure 1.2 reports a (partial) list of of drugs that have been assayed for their propensity to interact with transcription factor/DNA systems [12]. Although effective against virtually all kinds of tumours, genotoxic chemotherapeutics approved so far in the clinics act indiscriminately on wealth and malignant cells. Thus, understanding at a molecular level the mechanism of action of these drugs is a key step in the design new and more effective pharmaceuticals.

1.2   Structural determinants of drug/DNA adducts

Most drugs exert their action by binding to DNA. Therefore their rational design requires a detailed understanding of which interactions are responsible for affinity, reactivity and selectivity towards nucleic acids, as well as the comprehension of the molecular recognition events. Such a knowledge is the prerequisite for the investigation of molecular routes triggering cellular responses to genotoxic agents (like ATM/p53 activation – Figure 1.1). For these purposes, structural information on
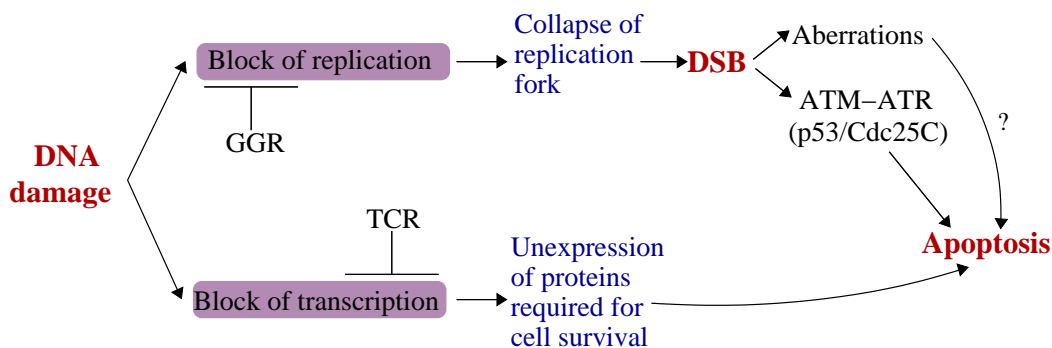
Figure 1.1: Responses provoked by replication- and transcription-blocking DNA lesions. Replication-blocking DNA lesions are repaired by global genomic repair (GGR), which is part of the Nucleotide Excision Repair (NER). If not repaired, they block DNA replication during S phase, which can lead to collapse of the replication fork and the formation of lethal Double Strand Breaks (DSBs). Bulky DNA lesions can also block transcription. Transcription-blocking lesions are repaired by Transcription-Coupled Repair (TCR), which also is part of NER. If not repaired, proteins required for survival might not be transcribed and the cell will undergo apoptosis (adapted from Ref. 17).

drug/DNA adducts is firstly required. To date ~ 1500 ligand/oligonucleotide X-ray or NMR structures have been reported on Nucleic Acids and Protein DataBases (NDB and PDB respectively, Table 1.1).[2] Ligands can bind DNA both covalently (to nucleobases or to phospate sugar-backbone) or non covalently. They can be accomodated in the minor or in the major groove, with or without intercalating between base pairs. Most compounds lie in the minor groove, where they form favorable packing interactions with the walls of the DNA. Organic covalent binders usually alkylates purines in the minor groove. The structures of about 35 adducts between DNA and minor groove ligands have been solved (Table 1.2), of which 6 are DNA anticancer alkylating agents [9, 12]. The dominance of non-covalent binders is mainly due to the difficulties in obtaining pure crystallographic quantities of adducts. This contrasts with the simple mixing procedures used for non covalent complex formation [18, 19].

Molecular modeling of drug/DNA adducts

Computer simulations have become an essential tool for investigating the behavior of biological systems, unraveling the subtle relationship involving structural and dynamical features and activity of biomolecules at an atomic level of detail. However,

---

[2]For a short summary of DNA structure see Appendix A.
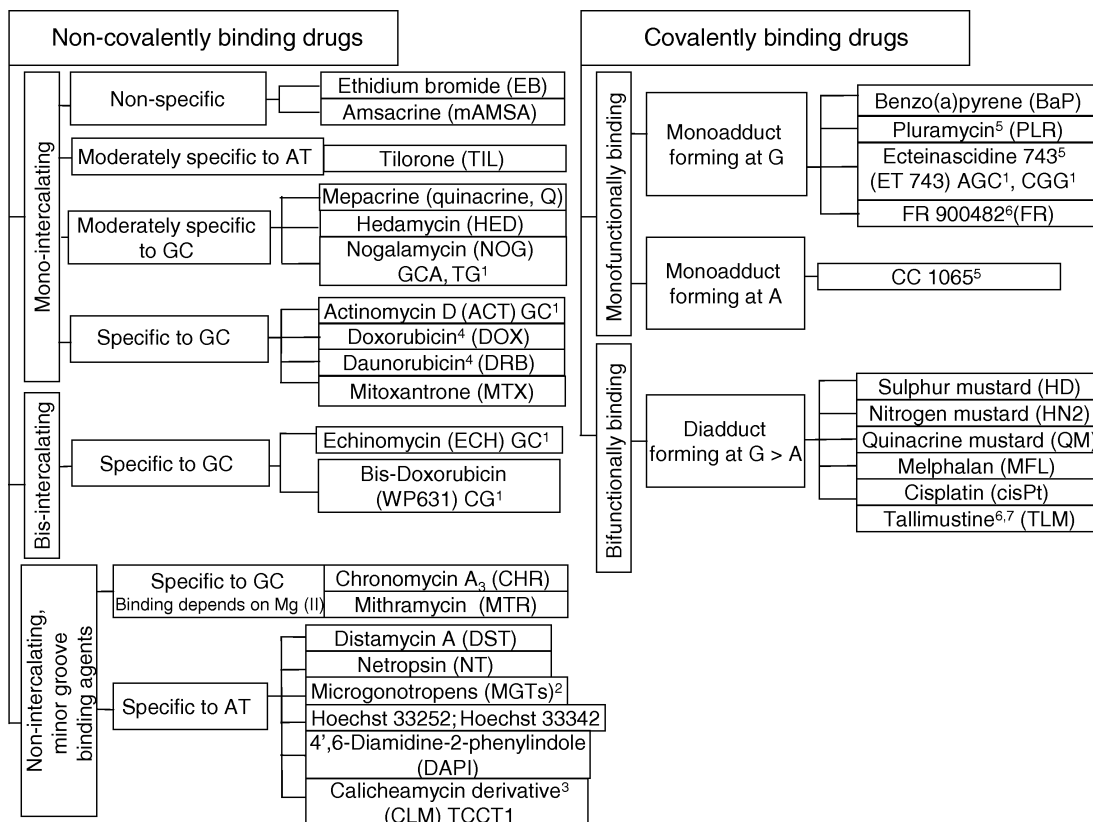
Figure 1.2: DNA-binding targets: mode of interactions and their region and sequence specificity (taken from Ref. 12). [1] Sequence preference. [2] A series of related drugs with distamycin A moiety and a polyamine tail, which interacts in the major groove and exhibits preference to GC. [3] Methyl-derivative of aryltetrasaccharide moiety of calicheamicin. The drug lacks the enedyne component, hence, it does not cleave DNA. [4] In some studies, anthracycline antibiotics covalently bound to DNA in the presence of dithiothreitol and Fe (III) ions or formaldehyde was assayed. [5] Interactions in the minor groove. [6] A cross-linking agent related to mitomycin C. [7] Preferential binding to A.

Biomolecule contains	Number of hits
DNA	2567
DNA and ligands	1477
Proteins	45946
Proteins and ligands	11045
DNA and proteins	1350

Table 1.1: Number of hits corresponding to structures containing DNA, proteins and ligands. The search has been performed on the PDB database over a total of 47805 structures (december 2007), after removal of sequences with similarity larger than 95%.

the development of models (*force fields*) for the simulation of DNA has revealed itself much more challenging than for proteins, because of several factors. These include: *i)* the polyanionic nature of oligonucleotides, which requires a very accurate treatment of the balance between the oligonucleotides' conformational energies and their interactions with the aqueous solvent [20] (i.e. a very accurate tuning of bonded *vs* non bonded *force field* terms); *ii)* the relatively low potential energy barrier associated to the torsion angles governing the conformation of the DNA phosphate-sugar backbone [21, 22, 23]. Reliable *force fields* for the simulation of oligonucleotides in acqueous solvent have been developed since the mid 1990s [24, 25, 26, 27, 28, 29, 30], and are under continuous development (pathologies still exist which are being cured nowadays [31]). Despite these difficulties, molecular modeling allowed to study structural, kinetic and thermodynamic aspects of oligonucleotides alone or interacting with proteins and ligands [32, 33, 34, 35, 36, 37, 38]. Among the biologically relevant issues which have been addressed there are:

- Conformational properties of DNA duplexes: the structural stability and flexibility of A, B, Z-DNA in solution [39, 40, 41, 42], the transitions between different oligonucleotide conformations [43, 44, 45]; sequence effects on the structure [46, 47, 48], the population of different conformational substates [49, 50, 51, 52]; the opening of base-pairs [53, 54, 55, 56, 57], the unwinding and stretching of DNA duplexes [58, 59], the minor groove deformability [60]; the ion and water distribution around DNA [61, 62].

- Electrostatic properties of DNA [63, 64, 65, 35].

- Protein/DNA interactions and molecular recognition [66, 67, 68, 69].

- Interactions, selectivity and binding of small ligands to DNA [70, 71, 72, 19, 73, 74, 75, 76].

Such studies facilitate interpretation of experimental data and allow for information not readily accessible to experimental methods to be obtained [20, 35, 34].

1.3   Motivation of the work

The first and largest part of my thesis focuses on ligands which bind to DNA, in particular alkylating agents. Despite the latest methodological advances and the pharmacological relevance of alkylating drugs [77, 78, 79, 80, 81], their interaction with DNA has been less characterized compared to non covalent binders. Only few recent theoretical studies have appeared in the literature (including some from our group)[3] concerning the structure, the thermodynamics, the kinetics and the molecular recognition of minor groove alkylating binders [82, 83, 84, 85, 86, 71]. This contrasts with the relatively large number of studies on non covalent ligands (see e.g. Refs. 87, 76, 88, 37, 75, 89, 74, 90, 91, 92, 93, 94, 70, 95, 96, 97). More importantly, many issues are still not fully understood, such as the contributions of different interactions and structural features to DNA sequence selectivity, the routes of molecular recognition by ligands and proteins and the catalytic role of nucleic acids in bond covalent modification. This information is essential for structural biology and drug design, as the latter relies on the assumption that beneficial effects of drugs origin from the interaction with its target.

My project during this PhD is thigtly placed within the scenario depicted above. I firstly characterized the structure and the energetics of adducts between the anticancer covalent binder anthramycin [98] and DNA, for which x-ray structure is available [99]. This study follows the spirit of previous investigations performed in the group on other two anticancer compounds (the major groove binder cisplatin [100, 101, 102, 103, 104] and the minor groove alkylating agent duocarmycin [105, 106, 107, 83]). Subsequently, I extended my investigation to the molecular recognition events involving both the latter compound as well as anthramycin. Comparison has also been made with the prototype of non covalent minor groove binders, distamycin A [108, 109, 110, 111]. The focus has been on the calculation of free energies, the evaluation of enthalpic *vs* entropic contributions and the role of the solvent, the estmation of the weigths of various interactions. This information has provided insights on similarities and differences in the molecular recognition process between the different kind of DNA ligands.

The second part of the thesis focused on the hydrolysis of two ruthenium-based compounds, ICR [112] and NAMI-A [113, 114, 115]. These are very promising for the treatment of primary tumours and metastases, respectively [116, 117]. The identification of most stable hydrolytic products is important as they would likely interact with

---

[3]I do not include in this list MD simulations performed to refine NMR structures

Compound	bond type	total number
duocarmycin		
anthramycin		
mitomycin		
bizelesin		
esperamycin	covalent, minor groove	9
calicheamicin		
tamoxifen (anti-estrogen)		
epidoxorubicin (also intercalator)		
psolaren (also intercalator)		
netropsin		
distamycin		
lexitropsins (3)		
polyamides (6)		
hoechst33258		
hoechst33342		
proamine		
bis-benzimidazole		
tris-benzimidazole		
propamidine		
pentamidine	non covalent, minor groove	29
furamidine		
chromomycin		
mithramycin		
DAPI		
AR-1-144		
DB819		
DB884		
DB921		
SN6999		
SN7167		
UCH9		
cisplatin		
oxaliplatin	metal-containing	3
bisplatin		

Table 1.2: Minor groove covalent and non covalent binders, and metal-based DNA drugs for which the structure of an adduct with DNA has been resolved (from PDB databank, updated december 2007).

biological targets. This work has started thanks to the collaboration here in Trieste with E. Alessio and G. Sava's labs, who discovered NAMI-A in the mid 1990s.

1.4   Outline of the Thesis

The outline of the thesis is the following:

- Chapter 2 introduces the theoretical background needed to understand molecular simulations and free energy calculations, which are the main tools used in the work exposed here.

- Chapters 3 to 5 constitute the first part of my thesis.

  – Chapter 3 focuses on the interaction between the anticancer alkylating drug anthramycin and DNA. Following the spirit of previous works here in SISSA, I characterized both the covalent complex and the non-covalent adducts between two supposed reactive forms of the drug and the DNA by performing classical MD. Then I used Car-Parrinello quantum MD to gain insigths on the reactivity of the two forms and to investigate the catalytic role of DNA for the reaction.

  – Chapter 4 addresses the problem of molecular recognition. This is a complex phenomenon, that goes beyond the simple formation of a stable adduct. In fact, small molecules exploit not only binding, yet sliding into and unbinding from the minor groove until they reach a preferred sequence on DNA. In this Chapter is reported the evaluation of free energy profiles associated to the sliding of anthramycin and duocarmycin into the minor groove of their non covalent complexes with DNA. The drugs are different in size, selectivity and affinity, which reflects in two significantly different profiles associated to the process. A dissection of the profiles is also undertaken, which allows for a understanding of the sources of similarities and differences in the mechanisms of the two drugs.

  – Chapter 5 describes the mechanism of unbinding from DNA of anthramycin and distamycin A, again starting from the non covalent complexes. The behaviors are compared between a covalent binder, whose specificity to guanines is due to covalent linkage, and a non covalent ligand that achieves its high selectivity towards AT-tracts through specific h-bonds. The routes of detachment turn out to be very different. Indeed unbinding of distamycin goes through the formation of a relatively stable intermediate. In contrast, the detachment of anthramycin appears as a monophasic process triggered by local DNA structural changes and involving a role of

waters. For both the compounds, the last contacts with DNA are formed by the hydrophilic tails, and the main cost of detachment arise from the disruption of hydrophobic packing interactions.

- Chapter 6 concerns the second Part of my thesis. It is a quantum mechanical study on the hydrolysis of NAMI-A and ICR. Despite the structural similarity, ICR is active mainly on the primary tumours, while NAMI-A is only active against metastases. Significant differences are seen in the hydrolysis paths of these two drugs, which could be relevant to understand the reason of their diverse pharmacological activity.

- In the last Part I draw some conclusions of these works and I outline future perspectives.

Chapter 2

Methods

The role of molecular modeling for life sciences has had a boomy increase over the past few decades thanks to the constant expansion in computer performances and the development of fast and robust simulation algorithms. In particular, computer simulations has been more and more being employed to investigate features not directly accessible to experiments. Among these are multiple conformations of ligand-DNA complexes [75], multi nanosecond dynamics at atomistic level [62, 31], and discovery of unpredicted conformations of chemical compounds [118]. The bridge between molecular simulations and macroscopic world has deep bases in statistical mechanics. Indeed, thermodynamic quantities are determined in computer simulations as *ensemble averages* over a large number of microscopic configurations assumed by the system under study. Common techniques to *sample* the configurations assumed by a system at equilibrium are Molecular Dynamics (MD), Monte Carlo (MC) sampling and Stochastic/Brownian dynamics. The work described in this thesis mainly benefit of MD techniques, with both a quantum and classical description of the system under study. As the use of MD in Statistical Biology is based on the ergodic hypothesis and the existence of shadow orbits, I will start briefly recalling these assumptions. Then, I will describe how to approach the many-body problem within Schrödinger and Newton system-evolution representations. Finally, I will describe methods I used for the calculations of free energies and for the evaluation of its various contributions.

2.1 Statistical foundations of MD: Ergodic hypothesis and shadow orbits

The idea behind MD simulations echoes the way real-life experiments are performed. The equilibrium behavior of a complex system is studied by following its time evolution in the absence of external impulses and thermodynamic properties are calculated from averages over a sufficient long trajectory. Such a procedure is well founded only

for the so-called *ergodic* systems, which are assumed to fully sample the accessible phase space (hyper)volume during the "observation" (i.e. "simulation") time. This is equivalent to assume that time averages are independent from MD initial conditions. Let consider a system of N particles in a volume V and denote $\Omega(N, V, E)$ the number of microstates with energy E. Within the ergodic hypothesis, if we perform a number $\#\{r^0, p^0\}$ of MD simulations starting from different initial conditions (r^0, p^0) in the space of coordinates and momenta (*phase space*), the time average of any properties $f(r, p; t')$ over a time interval t can be expressed as:

$$\overline{f(r,p)} = \lim_{t \to \infty} \frac{1}{t} \int_0^t dt' \, f(r,p;t') = \frac{\sum_{\{r^0,p^0\}} lim_{t \to \infty} \frac{1}{t} \int_0^t dt' \, f(r,p;r^0,p^0;t')}{\#\{r^0,p^0\}} \quad (2.1)$$

The last term in eq. 2.1 can be rewritten as:

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t dt' \sum_{\{r^0,p^0\}} \frac{f(r,p;r^0,p^0;t')}{\#\{r^0,p^0\}} \quad (2.2)$$

Considering the limiting case in which we sum a function $g(r^0, p^0)$ over *all* the possible initial conditions, we obtain the *ensemble* average:

$$\frac{\sum_{\{r^0,p^0\}} g(r^0,p^0)}{\#\{r^0,p^0\}} \longrightarrow \frac{\int_E dr dp \, g(r^0,p^0)}{\Omega(N,V,E)} = \langle g(r^0,p^0) \rangle_{NVE} \quad (2.3)$$

where $\Omega(N, V, E) = c \int_E dr dp \, g(r^0, p^0)$ and c is a constant.[1]
Thus, we have:

$$\overline{f(r,p)} = \lim_{t \to \infty} \frac{1}{t} \int_o^t dt' \langle f(r,p;r^0,p^0,t') \rangle_{NVE} = \langle f(r,p) \rangle_{NVE} \quad (2.4)$$

where the last equality comes from the independence of ensemble averages on time [119]. It is generally assumed that complex systems, such as the majority of those biologically relevant, are ergodic. Although this is a plausible assumption, it should be pointed out that the *ergodic hypothesis* is not always true (for some example see e.g. Ref. [119]). A further drawback affecting (in principle) all kinds of MD simulations is the so-called Liapunov instability [120, 119]. For chaotic systems, like almost all those simulated by MD, the trajectory is extremely sensitive to initial conditions. This implies that any error in the integration of the equation of motion, no matter how small, will always cause the simulated (numerical) trajectory to diverge exponentially from the "true" trajectory starting from the same initial conditions. However, for the purpose of MD simulations, what is needed is the equivalence

---

[1]We assume here $c = 1$, that is correct for quantum systems. In the classical limit, $c = 1/\text{h}^{3N}$ and $c = 1/N! \text{h}^{3N}$ for distinguishable and indistinguishable particles respectively [119].

of the numerical and "true" trajectories in a *statistical* rather than *deterministic* sense. Thus, even if the numerical trajectory diverges from the true one, as long it conserves the total energy E within a given accuracy ΔE, the sampled configurations will belong to a constant energy hypersurface E^* for which $|E^* - E| < \Delta E$. This means that the numerical trajectory will be, within the error ΔE, representative of some true trajectory in the phase space, although it is not known a priori which one. The true trajectory, to which the numerical one overlaps for a certain period of time, is called *shadow orbit*. At the time the Liapunov instability raises up, the numerical trajectory will get far from that specific shadow orbit, but there always will be another one of these to which it is superimposed.

2.2 MD in various Statistical Ensembles

2.2.1 NVE Ensemble

In the last section we introduced the ergodic hypothesis for a system described by the microcanonical distribution NVE. The dynamics of such system follows the Hamiltonan (or the equivalent Newtonian) laws of motion:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i}, \tag{2.5}$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i. \tag{2.6}$$

However, the conditions of constant volume V, number of particles N and total energy E do not fit those in which experiments are usually made. Thus, it is necessary to define schemes allowing for the evolution of systems under conditions of constant volume and temperature (NVT), or constant pressure and temperature (NPT), corresponding to typical real-life situations. In order to devise MD schemes for these ensembles, two main strategies have been developed [119], one based on the extended Lagrangian formalism, the other on mixing MD with some MC moves which transport the system between different shells of constant energy. Here I describe the Nosé-Hoover thermostat [121, 122] and the Andersen barostat [123], which are implemented in the MD codes I used and belong to the extended Lagrangian picture.

2.2.2 NVT: Nosé-Hoover thermostat

A way to sample the NVT ensemble within the framework of MD was introduced about twenty years ago by Nosé [121] and reformulated by Hoover [122], who modify Newton equation of motion by adding two non physical variables, thus introducing

the following non-Hamiltonian dynamical system:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i}, \tag{2.7}$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \frac{p_\eta}{Q}\mathbf{p}_i, \tag{2.8}$$

$$\dot{\eta} = \frac{p_\eta}{Q}, \tag{2.9}$$

$$\dot{\mathbf{p}}_\eta = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{m_i} - LkT. \tag{2.10}$$

where $\{\mathbf{r}_i\}, \{\mathbf{p}_i\}$ are coordinates and and momenta of the N particles with masses m_i, the forces $\mathbf{F}_i$ are derived from the N-particle potential and L is a parameter to be determined. The two nonphysical variables η and p_η in eq. 2.7 regulate the fluctuations in the total kinetic energy of the physical variables, and can be thus regarded as an effective "thermostat" for the physical system. The parameter Q controls the strength of the coupling to the thermostat: high values result into a low coupling and viceversa. These equations of motion conserve the following quantity:

$$C = H(\mathbf{p}, \mathbf{r}) + \frac{p_\eta^2}{2Q} + LkT\eta \tag{2.11}$$

where $H(\mathbf{p}, \mathbf{r})$ is the physical Hamiltonian. Assuming that C is the only constant of motion, the microcanonical partition function at temperature T can be written as:

$$\Omega_T(N, V, C) = \int d^N\mathbf{p}\, d^N\mathbf{r}\, dp_\eta\, d\eta\, e^{3N\eta}\delta\left(H(\mathbf{p}, \mathbf{r}) + \frac{p_\eta^2}{2Q} + LkT\eta - C\right)$$
$$= \frac{e^{3NC/LkT}}{LkT} \int dp_\eta e^{-3Np_\eta^2/2QLkT} \int d^N\mathbf{p}\, d^N\mathbf{r}\, e^{-3NH(\mathbf{p},\mathbf{r})/LkT} \tag{2.12}$$

that, with $L = 3N$, is proportional to the canonical partition function for the physical system. Thus, the Nosé "extended" scheme gives conformations of the physical systems that belong to the canonical ensemble. In eq. 2.12, the metric $e^{3N\eta}$ has been used as $d^N\mathbf{p}\, d^N\mathbf{r}\, dp_\eta\, d\eta\, e^{3N\eta}$ is the phase-space conserved measure for this set of equation of motion (see Appendix B).

Nosé-Hoover chains

When there is more than one constant of motion a single Nosé-Hoover thermostat is unable to get the correct canonical sampling for the physical system [119]. It has been indeed shown that this is the case for small or stiff systems [124], where a single Nosé-Hoover thermostat can give rise to high oscillations in the temperature. Examples of

pathological systems are a catalytic site of a protein or a drug-nucleobase adduct in quantum MD simulations. To allow the simulation of more general systems Martyna *et al.* [125] devised a scheme in which additional dynamical variables are inserted in the Lagrangian. This corresponds to couple the first thermostat to another one, and so on, generating a whole chain of thermostats (here the name Nosé-Hoover chains). In fact, as the momentum of the extended variable entering in the first Nosé-Hoover thermostat is distributed canonically, it can also be thermostated. The chains take into account for additional conservation laws, and generates the correct canonical distribution. The equations of motion for a system coupled to N_c thermostats are:

$$\dot{\mathbf{r}}_i = \frac{\dot{\mathbf{p}}_i}{m_i}, \tag{2.13}$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \frac{p_{\eta_1}}{Q_1}\mathbf{p}_i, \tag{2.14}$$

$$\dot{\eta}_k = \frac{p_{\eta_k}}{Q_k}, \qquad k = 1, ..., N_c, \tag{2.15}$$

$$\dot{\mathbf{p}}_{\eta_1} = \left(\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i} - LkT\right) - \frac{p_{\eta_2}}{Q_2}p_{\eta_1}, \tag{2.16}$$

$$\dot{\mathbf{p}}_{\eta_k} = \left(\frac{\mathbf{p}_{\eta_{k-1}}^2}{Q_{k-1}} - kT\right) - \frac{p_{\eta_{k+1}}}{Q_{k+1}}p_{\eta_k}, \tag{2.17}$$

$$\dot{\mathbf{p}}_{\eta_{N_c}} = \left(\frac{\mathbf{p}_{\eta_{N_c-1}}^2}{Q_{N_c-1}} - kT\right). \tag{2.18}$$

The dynamics generated by these equations conserves the following "energy":

$$H_{NHC} = H(\mathbf{p}, \mathbf{r}) + \sum_{k=1}^{N_c}\frac{p_{\eta_k}^2}{2Q_k} + LkT\eta_1 + \sum_{k=2}^{N_c}kT\eta_k \tag{2.19}$$

2.2.3 NPH: Andersen barostat

At constant pressure, the volume V of the system fluctuates and has to be treated as a dynamical variable. Andersen introduced a dynamical scheme in which the time-average of any property is equal to the isoenthalpic-isobaric (NPH) ensemble average [123]. Subsequently, Parrinello and Rahaman [126] generalized the algorithm to allow the simulation of solids under stress, where different pressures can be applied along the three box edges. The Andersen algorithm considers a physical system of N particles, with coordinates $\{\mathbf{r}_i\}$ and momenta $\{\mathbf{p}_i\}$, and introduces a new variable Q

and a set of scaled coordinates $\{\rho_i\}$ which enter in the following Lagrangian:

$$\mathcal{L}_{And}(\rho^N, \dot{\rho}^N, Q, \dot{Q}) = \frac{1}{2}mQ^{2/3}\sum_{i=1}^{N}\dot{\rho}_i \cdot \dot{\rho}_i - \sum_{i<j=1}^{N}\phi(Q^{1/3}\rho_{ij}) + \frac{1}{2}M\dot{Q}^2 - \alpha Q, \quad (2.20)$$

where m is the mass of the (identical) N particles, ϕ is the potential. α a parameter having the value of the pressure P of the system to be simulated, and M is the mass of the "piston" coupled to it. The corresponding Hamiltonian is:

$$\mathcal{H}_{And}(\rho^N, \pi^N, Q, \Pi) = (2mQ^{2/3})^{-1}\sum_{i=1}^{N}\pi_i \cdot \pi_i + \sum_{i<j=1}^{N}\phi(Q^{1/3}\rho_{ij}) + \frac{\Pi^2}{2M} + \alpha Q,$$

$$(2.21)$$

where $\pi_i = mQ^{2/3}\dot{\rho}_i$ and $\Pi = M\dot{Q}$ are the momenta conjugate to ρ_i and Q, respectively. A trajectory generated by this Hamiltonian has the property that the time average of a function of the original system variables $F(\mathbf{r}^N, \mathbf{p}^N; V)$ is equal, except for a negligible error, to the isoenthalpic-isobaric ensemble average of F, at pressure $P = \alpha$:

$$\bar{F} = F_{NPH}(N, P, H) = \frac{\int d^N\mathbf{r}\, d^N\mathbf{p}\, dV\, \delta\left(\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N; V) + PV - H\right) F(\mathbf{r}^N, \mathbf{p}^N; V)}{N!\,\Gamma(N, P, H)},$$

$$(2.22)$$

where $\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N; V)$ is the original system Hamiltonian, and

$$\Gamma(N, P, H) = \frac{1}{N!}\int d^N\mathbf{r}\, d^N\mathbf{p}\, dV\, \delta\left(\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N; V) + PV - H\right) \quad (2.23)$$

The above relation is true providing that the following correspondences between the original and scaled dynamical variables are defined:

$$V(t) = Q(t), \quad (2.24)$$
$$\mathbf{r}_i(t) = Q^{1/3}\rho_i(t), \quad (2.25)$$
$$\mathbf{p}_i(t) = Q^{-1/3}\pi_i(t). \quad (2.26)$$

This mapping identifies the instantaneous value of the volume of the system with the introduced variable Q and allows to express $\mathcal{H}_{And}$ in terms of $\mathcal{H}$.

$$\mathcal{H}_{And}(\rho^N, \pi^N, Q, \Pi) = \mathcal{H}(Q^{1/3}\rho^N, Q^{-1/3}\pi^N; Q) + \frac{\Pi}{2M} + \alpha V. \quad (2.27)$$

Now, expressing F through the scaled coordinates as

$$F(\mathbf{r}^N, \mathbf{p}^N; V) = F(Q^{1/3}\rho^N, Q^{-1/3}\pi^N; Q) \equiv G(\rho^N, \pi^N, Q, \Pi), \qquad (2.28)$$

it follows that the average of F is equivalent to the average of G in the micro-canonical ensemble NE corresponding to $\mathcal{H}_{And}$:

$$\bar{F} = \bar{G} = \int_1 d\rho^N \int d\pi^N \int dQ \int d\Pi \, \delta \left(\mathcal{H}_{And}(\rho^N, \pi^N, Q, \Pi) - E \right) G(\rho^N, \pi^N, Q, \Pi)$$

$$= \int_1 d\rho^N \int d\pi^N \int dQ \int d\Pi \, \delta \left(\mathcal{H}_{And}(\rho^N, \pi^N, Q, \Pi) - E \right) F(Q^{1/3}\rho^N, Q^{-1/3}\pi^N; Q)$$

$$(2.29)$$

The integration in ρ^N is done over a unitary volume, since the coordinates are scaled. Returning to the variables $\mathbf{r}^N$, $\mathbf{p}^N$, and V one obtains:

$$\bar{F} = \frac{\int d\Pi \int d^N\mathbf{r} \, d^N\mathbf{p} \, dV \, \delta \left(\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N; V) + \frac{\Pi^2}{2M} + \alpha V - E \right) F(\mathbf{r}^N, \mathbf{p}^N; V)}{\int d\Pi \int d^N\mathbf{r} \, d^N\mathbf{p} \, dV \, \delta \left(\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N; V) + \frac{\Pi^2}{2M} + \alpha V - E \right)}$$

$$(2.30)$$

which resembles F_{NPH}, except for the integration in Π. Thus, we can write:

$$\bar{F} = \frac{\int d\Pi \, \Gamma(N, \alpha, E - \frac{\Pi^2}{2M}) \, F_{NPH}(N, \alpha, E - \frac{\Pi^2}{2M})}{\int d\Pi \, \Gamma(N, \alpha, E - \frac{\Pi^2}{2M})}, \qquad (2.31)$$

and expanding the ensemble average F_{NPH} in the numerator as a power series of $\Pi^2/2M$ up to the first order one finds:

$$\bar{F} \approx F_{NPH}(N, \alpha, E - \overline{\Pi^2/2M}). \qquad (2.32)$$

As the momentum Π appears in $\mathcal{H}_{And}$ only as quadratic term, the average $\overline{\Pi^2/2M}$ is equal to $kT/2$, where T is the temperature of the scaled system (at fixed N and E). The equations of motion for the original system are obtained from those of the scaled one (derived from $\mathcal{H}_{And}$) by using the correspondence rules 2.24:

$$\dot{\mathbf{r}}_i = \frac{\dot{\mathbf{p}}_i}{m_i} + \frac{1}{3}\mathbf{r}_i\frac{d\ln V}{dt}, \qquad (2.33)$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \frac{1}{3}\mathbf{p}_i\frac{d\ln V}{dt}, \qquad (2.34)$$

$$\frac{M d^2V}{dt^2} = -\alpha + \left(\frac{2}{3}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m} - \frac{1}{3}\sum_{i<j=1}^{N} r_{ij}\frac{d\phi(r_{ij})}{dr_{ij}} \right)/V. \qquad (2.35)$$

Summarizing, when the trajectory of the physical system is being determined by that of the scaled system through the correspondence rules 2.24, the time average of any property F coincides with its NPH ensemble average.

2.3   Force calculation

Once the equations of motion have been devised to sample the different statistical ensembles, all we need is to calculate the forces acting on each "particle" (i.e. the constitutive entity in the system). At the first level of approximation after the neglection of the subnuclear particles, these particles are identified as nuclei and electrons, which are rigorously described by quantum mechanics.[2] However, for most of the purposes of biological simulations, the quantum nature of nuclei can be neglected, leading to the so-called semi-classical description of matter. In this framework, we describe the Born-Oppeheimer and Car-Parrinello MD schemes. The next level of approximation consist in integrating out the electronic degrees of freedom, describing the interaction among atoms though mean-average *force fields*. These are used to evolve systems through classical MD.

2.3.1   The semi-classical approximation

In the non-relativistic limit, the time evolution of a system of M atoms and N electrons are described by the time dependent Schrödinger equation [127]:

$$i\hbar\frac{\partial}{\partial t}\Psi = \mathcal{H}\Psi\,,\tag{2.36}$$

where the wavefunction Ψ depends (in the space representation) on the $3(M+N)$ coordinates, the nuclear and electronic spin states and on time. In absence of external

---

[2] For the sake of clarity, in the following we will often restrict to closed-shell systems, for which the sum over spins can be factorized out. Only when no additional complexity is added the discussion will be taken as most as general considering explicitly spin-orbitals instead of spatial orbitals.

fields, the Hamiltonian $\mathcal{H}$ reads:

$$
\begin{aligned}
\mathcal{H} =& T + V = T_e + T_N + V_{eN} + V_{ee} + V_{NN} \\
=& -\sum_{i=1}^{M} \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_{I=1}^{N} \frac{\hbar^2}{2M_I} \nabla_I^2 + \sum_{I,i} \frac{Z_I \mathrm{e}^2}{|\mathbf{R}_I - \mathbf{r}_i|} \\
& + \sum_{i<j}^{M} \frac{\mathrm{e}^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J}^{N} \frac{Z_I Z_J \mathrm{e}^2}{|\mathbf{R}_I - \mathbf{R}_J|} \\
=& -\sum_{i=1}^{M} \frac{\hbar^2}{2M_I} \nabla_I^2 + \mathcal{H}_e(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})
\end{aligned}
\tag{2.37}
$$

At room temperature the thermal wavelength λ is about $0.1\,\text{Å}$, while typical interatomic distances, in liquids and solids, are of the order of $1\,\text{Å}$. Thus, a good approximation is to neglect quantum correlations between wave functions of different nuclei, *i.e* to consider the nuclear wavefunction as an incoherent superimposition of individual nuclear wave packets. In addition, nuclear masses are large enough that such individual wave packets are usually well localized. These two observations lead to the conclusion that the quantum nature of nuclei can be safely neglected in many circumstances. In fact the majority of the so-called *first-principle* or *ab initio* methods consider the nuclei as *classical particles* moving in a (effective) potential due to electrons. To achieve a classical description of nuclei, it is first necessary to separate electronic from nuclear degrees of freedom. Formally, this can be done by writing the wavefunction as product of terms depending only on electronic or nuclear positions ("one-determinant" Ansatz [128]):

$$
\Phi(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}; t) \approx \Psi(\{\mathbf{r}_i\}; t) \chi(\{\mathbf{R}_I\}; t) \exp\left[\frac{i}{\hbar} \int_{t_0}^{t} dt' \langle \mathcal{H}_e(t') \rangle \right]
\tag{2.38}
$$

This separation leads to a set of self-consistent equations for the electronic and nuclear wavefunctions, in which each system move in the *average* field of the other class of degrees of freedom:

$$
i\hbar \frac{\partial \Psi}{\partial t} = -\sum_{i=1}^{M} \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + \left\{ \int d\mathbf{R}\, \chi^*(\{\mathbf{R}_I\}; t)\, V_{eN}(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})\, \chi(\{\mathbf{R}_I\}; t) \right\} \Psi
\tag{2.39}
$$

$$
i\hbar \frac{\partial \chi}{\partial t} = -\sum_{I=1}^{N} \frac{\hbar^2}{2M_I} \nabla_I^2 \chi + \left\{ \int d\mathbf{r}\, \Psi^*(\{\mathbf{r}_i\}; t)\, \mathcal{H}_e(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})\, \Psi(\{\mathbf{r}_i\}; t) \right\} \chi
\tag{2.40}
$$

Eqs. 2.39 and 2.40 define the Time-Dependent Self-Consistent Method (TDSCF), first introduced by Dirac in 1930. A classical description of nuclei dynamics is achieved by expressing χ in exponential form [129],

$$\chi(\{\mathbf{R}_I\}; t) = A(\{\mathbf{R}_I\}; t) \exp\left[iS(\{\mathbf{R}_I\}; t)/\hbar\right] , \tag{2.41}$$

where A and S are real and $A > 0$. Substituting this expression in eq. 2.40 gives, in the classical limit $\hbar \to 0$, the following set of equations:

$$\frac{\partial S}{\partial t} + \sum_{I=1}^{N} \frac{(\nabla_I S)^2}{2M_I} + \int d\mathbf{r} \, \Psi^* \, \mathcal{H}_e \, \Psi = 0 \tag{2.42}$$

$$\frac{\partial A^2}{\partial t} + \sum_{I=1}^{N} \nabla \cdot \left(A^2 \frac{\nabla_I S}{M_I}\right) = 0 . \tag{2.43}$$

Using the connection $\mathbf{P}_I \equiv \nabla_I S$ eq. 2.42 is isomorphic to the Hamilton-Jacobi equation of classical motion for action S and Hamiltonian $\mathcal{H} = T(\{\mathbf{P}_I\}) + V(\{\mathbf{R}_I\})$, while eq. 2.43 is a continuity equation for the density probability $A^2 = |\chi|^2$ of nuclei, which move with classical velocities $\nabla_I S/M_I = \mathbf{p}_I/M_I$. Thus, the motion of nuclei can be described using Newton's law:

$$\frac{d\mathbf{P}_I}{dt} = -\nabla_I \int d\mathbf{r} \, \Psi^* \, \mathcal{H}_e \, \Psi = -\nabla_I V^E(\{\mathbf{R}_I(t)\}) . \tag{2.44}$$

In particular, nuclei are driven by a *mean-field* potential due to electrons (and containing also a contribution from their kinetic energy) and depending on nuclear configuration at time t. To get off the nuclear wavefuntion also from eq. 2.39 one replaces the nuclear density $|\chi(\{\mathbf{R}_I\}; t)|^2$ by a product of delta functions $\prod_I \delta(\mathbf{R}_I - \mathbf{R}_I(t))$ (i.e. incoherent wave packets extremely localized) [128]. Summarizing, in the semi-classical picture the nuclei evolve classically, and the only requirement is to solve the Schrödinger equation for the electrons, whose wave function depends on time only *parametrically* on nuclear positions at time t:

$$i\hbar \frac{\partial \Psi}{\partial t} = -\sum_{i=1}^{M} \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + V \Psi = \mathcal{H}_e \Psi \tag{2.45}$$

$$M_I \ddot{\mathbf{R}}(t) = -\nabla_I \int d\mathbf{r} \, \Psi^* \, \mathcal{H}_e \, \Psi . \tag{2.46}$$

The set of equations 2.45 and 2.46 describe the so called "Ehrenfest molecular dynamics" scheme. It is clear now that the motion of the nuclei is dictated by the Hamiltonian $\mathcal{H}_e$, which basically contains the quantistic information on the electronic

system. Thus, the major task of quantum mechanics concerns the solution of the Schrödinger equation for the electrons, whose solution allow to know the dynamical behaviour of the system.

2.3.2 Empirical force-fields and Classical MD

The large computational costs in the implementation of *ab-initio* molecular dynamics end up into two major limitations: *i)* the relatively small size of the systems that can be simulated (hundreds of atoms) and *ii)* the relatively short simulation timescale (tenths of picoseconds). As many relevant biological processes usually involve large systems (thousands of atoms or more), and occur in relatively long timescales (nanoseconds or more), it is necessary to develop effective parametrized potentials, which are faster to integrate, albeit less accurate, in order to study this kind of systems. Force-fields based simulations originate from the assumption that the Born-Oppenheimer potential energy surface can be approximated by additive parametrized many-body terms that can be obtained by fitting experimental and high-level quantum chemical data into simple functional forms. The term "force field" indicates a functional form for this approximation, which relates the configuration of the system ($\{\mathbf{R}_i\}, i = 1, ..., N$) to its internal energy U, along with the set of parameters used in that function. In this work, the AMBER parm94 force field [24] for description of macromolecules in solution has been used, which is particularly well-suited for nucleic acids modeling. Namely, two sligthly modified versions of the original Cornell et al. force field have been employed. The first, called parm99 [25], was introduced by Cheatham et al. in 1999 to better reproduce the equilibrium between C2'-endo and C3'-endo sugar puckers and the helical repeat of DNA, which is slightly underestimated by the parm94 force field. The second was introduced this year by Perez et al. to improve the description of α/γ conformers [31]. It is important to remind that changes in a force field do not guarantee a systematic improvement over the original version. For example, one drawback of parm99 with respect to parm94 is its slowness in converting the DNA from A to B form, which could hamper effective sampling in short MD runs. The AMBER force fields belong to the so-called ClassI, whose the functional form is the following:

$$U = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\vartheta(\vartheta - \vartheta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos\left(n\phi - \gamma\right)\right]$$
$$+ \sum_{i<j}\left[4\varepsilon_{ij}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 + \frac{q_i q_j}{\varepsilon r_{ij}}\right]$$

(2.47)

Atom bond stretching and angle bending are represented as harmonic terms, while dihedrals or torsionals are described by a sinusoidal term. Non-bonded interactions

comprise two terms, the first is a Lennard-Jones 6-12 which describes atom-atom repulsion and dispersion interactions, the second is the Coulomb electrostatic term. In eq. 2.47, r and ϑ are respectively the bond length and valence angle; ϕ is the dihedral or torsion angle and r_{ij} is the distance between atoms i and j. Parameters include the bond force constant and equilibrium distance, K_r and r_{eq}, respectively; the valence angle force constant and equilibrium angle, K_ϑ, and ϑ_{eq}, respectively; the dihedral force constant, multiplicity and phase angle, V_n, n, and γ, respectively. The functional form used for out-of-plane distorsions (e.g. in planar groups) is different in different force fields [130]. For instance, in the AMBER force field [24, 25, 131] this term has the same form as that used for proper dihedrals, while in CHARMM [132, 133] an harmonic term is used. Collectively, these parameters represent the internal or intramolecular ones. Nonbonded parameters between atoms i and j include the partial atomic charges, q_i, along with the LJ well-depth, ε_{ij}, and σ_{ij}, the (finite) distance at which the interparticle potential is zero. These terms are also referred to as the interaction or external parameters. Typically, ε_{ii} and σ_{ii} are obtained for individual atom types and then combined to yield ε_{ij} and σ_{ij} for the interacting atoms via combining rules. The dielectric constant ε is typically set to 1 (corresponding to the permittivity of vacuum) in calculations that incorporate explicit solvent representations [130]. Van der Waals and electrostatic interactions are calculated between atoms belonging to different molecules or for atoms in the same molecules separated by at least three bonds.

Force-field optimization.   Force field parameters are fitted to reproduce chemical-physical properties of a class of model compounds representative of the biomolecules of interest. To this end quantum mechanics geometry optimizations are used to obtain bond and valence angle equilibrium constants and the dihedral phase and multiplicity, whereas vibrational spectra calculations are used to adjust force constants [130]. Lennard-Jones parameters are fitted to reproduce observable such as enthalpies of vaporization, free energies of solvation and densities of molecular liquids. Atomic charges are optimized to reproduce the QM-determined electrostatic potential (ESP) on a grid surrounding the molecule. As ESP charges tend to be undetermined, a widely used approach is to use restraints during fitting (usually to Hirshfeld charges), a method referred as Restrained ESP (RESP) [134]. Of course gas-phase calculations do no properly represent some of the condensed phase properties, thus a further refinement based on available experimental data is necessary. ESP charges are calculated, in most the cases, at the HF/6-31G(d) level as this is known to overestimate dipole moments and interaction energies [24]. This overestimation leads to partial charges that include the implicit polarization proper for condensed phase simulations.

2.3.3 Long-range interactions

In simulations of biological systems it is highly desirable to avoid the calculation of all non-bonded pair interactions, since the computational cost would be proportional to the square of the number of atoms. These interactions primarily dictates the dynamics of biomolecules, and cannot be merely truncated beyond a given cutoff [119] when long-ranged. An interaction can be roughly defined long-range if the associated energy falls off slower than r^{-d}, where d is the dimensionality of the system. Thus, Coulomb ($\sim \phi^{-1}$) and dipole-dipole ($\sim \phi^{-3}$) should be considered long-range when dealing with three dimensional systems. Indeed, cutting the interaction on a sphere of radius r_c makes it is easy to see that the tail correction, proportional to $\int_{r_c}^{\infty} \phi(r)\, 4\pi r^2$ diverges if the potential function goes to zero with $r \to \infty$ faster than r^{-3}. [3] Fortunately efficient methods to handle the evaluation of long-range interactions were developed. Among these, the most used are reaction field methods and lattice methods [135]. The first class assumes that beyond a given distance the interaction can be treated using an "average field" approach, derived from macroscopic electrostatic. The second class of methods, much widely used, is represented by the Ewald summation algorithm [136], which efficiently handles long-range interaction in periodic systems (and thus require Periodic Boundary Conditions - PBC - to be used for the system being simulated).

Ewald sums

Within PBC each particle interacts with all the other N-1 particles into the simulation box and with all the N particle images in an infinite 3D array of periodic cells. The electrostatic potential energy U^{Coul} of the infinite system, therefore, takes the form:

$$U^{Coul} = \frac{1}{2} \sum_{\mathbf{n}=0}^{}{}' \sum_{ij}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}\mathbf{L}|} \tag{2.48}$$

where L is the length of the periodic box (assumed cubic for simplicity), N is the total number of atoms, and **n** are the direct lattice vectors. The prime on the summation indicates that the sum is over all the images **n**, and $i \neq j$ if $\mathbf{n} = 0$. The convergence of this series is extremely slow, and thus a very large number of images is required to achieve a reliable estimate of U^{Coul}. A simple recipe to cope with this problem consist in splitting the electrostatic energy in two parts, a short-range and a

---

[3] The LJ interactions, instead, can be cut at a given cutoff (tipically 10-12 Å), since they decay very rapidly. Methods have been developed that treat the region beyond the cutoff radius r_c as being homogeneous (i.e. with LJ parameters averaged over all the LJ atom types) [119]. These interactions are thus effectively treated as short-range interaction, with the addition of a constant correction term.

long-range term. This means to find some function β such that the Coulomb potential can be rewritten ($x \equiv |\mathbf{r}_{ij} + \mathbf{n}\mathbf{L}|$):

$$\frac{1}{x} = \frac{\beta(x)}{x} + \frac{1 - \beta(x)}{x} \tag{2.49}$$

If β is properly chosen, the first term will be negligible beyond a given cutoff x_c, and the latter will be a slowly varying function of x, that can be expressed as Fourier Series with a low number of **k** points. Ewald [136] made for $\beta(x)$ the natural choice of the complementary error function $\mathrm{erfc}(x) = (2/\sqrt{(\pi)}) \int_x^\infty \exp(-u^2) du$, such that $1 - \beta(x) = 1 - \mathrm{erfc}(x) = \mathrm{erf}(x)$. Thus, in eq. 2.49 the term $\mathrm{erfc}(x)$ is evaluated cutting the interactions at a given cut-off, while the long-range term is Fourier transformed in the reciprocal space, where it becomes short-ranged and can be thus accurately estimated by summing over a limited number of reciprocal vectors (both the sums in real and reciprocal space becomes in fact exponentially converging). The parameter α tunes the relative weights of real and reciprocal sums, although the final result is independent of it. An optimal choice makes the Ewald sum converge as $N^{3/2}$, which can be further improved to $N \ln N$ with the use of Particle-Mesh methods (as the Particle-Mesh Ewald, PME [137] or the Particle-Particle Particle-Mesh, PPPM [138]), making advantage of the Fast Fourier Transform (FFT) [139] algorithm. For a detailed derivation of the electrostatic energy in the Ewald method and a brief discussion on Particle-Mesh Approaches see Appendix C.

2.4   The electronic structure problem

In the previous section I described the recipes that are used to evaluate interactions in dynamical systems. Concerning the force fields, once a proper parametrization is available, nothing more is requested that solving classical Newton equation of motion. Within the semi-classical approximation, the knowledge of the electronic structure is needed to calculate forces acting on the nuclei. In the following, I describe the main methods used for solving the electronic structure problem.

2.4.1   Time-space separation

Within the TDSCF scheme derived in the previous section, it is relevant noticing that $\mathcal{H}_e$ depends on time only *parametrically* through the positions of nuclei. Whenever the Hamiltonian does not depend explicitly on time, it is possible to formally separate the variables and reduce to a time-independent eigenvalue problem. In fact, we can cast the electronic wavefunction Ψ as simple product:

$$\Psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}, t) = \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) \, f(t) \,, \tag{2.50}$$

where $\mathbf{R}_I$ are instantaneous positions at time t, and ψ and f satisfy the following set of equations (obtained substituting the above expression in eq. 2.45):

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} f(t) = E f(t) \,, \tag{2.51}$$

$$\mathcal{H}_e \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) = E\psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) \,. \tag{2.52}$$

A *particular* solution of the Time-Dependent Schrödinger Equation (TDSE) is thus the product of a sinusoidal wave in time and a function satisfying the eigenvalue equation 2.52, which is called time-independent Schrödinger equation (TISE):

$$\Psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}, t) = \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) \left[f \, \mathrm{e}^{-\mathrm{i}Et/\hbar} \right] \,. \tag{2.53}$$

where $E = \langle \mathcal{H}_e \rangle$ is the energy of the electronic system in the nuclear configuration $\{\mathbf{R}_I\}$. At this point we have a *whole series* of solutions because generally there will be multiple values of E for which eq. 2.52 has solutions for ψ. As the time-dependent Schrodinger equation is linear in time, the general solution will be simply given by a linear combination of the various independent solutions:

$$\Psi = \sum_{k=0}^{\infty} f_k(t) \, \psi_k(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}) \,, \tag{2.54}$$

where $f_k(t) \equiv f_k \, \mathrm{e}^{-\mathrm{i}E_k t/\hbar}$ and ψ_k is the eigenfunction corresponding to energy E_k. Thus, system time evolution is entirely described by the set of coefficients $f_k(t)$. In fact, the occupation of eigenstates at any time is given by $|f_k(t)|^2$ (with $\sum_k |f_k(t)|^2 = 1$), while transitions are described *via* the cross-terms $f_k f_{l \neq k}$.

2.4.2 Methods for solving Time Independent Schrödinger Equation

Eq. 2.52, describing the behaviour of the electronic subsystem, is still too complicated, thus approximations have to be introduced to tackle it. A common approach consist in writing the total electronic wavefunction as a product (symmetrized or not) of *single-particle* wavefunctions. The work presented in this thesis has profited from Density Functional Theory [140] (DFT) and Hartree-Fock (HF) with Møller-Plesset (MP) 2nd order corrections. Both of these methods include electronic correlation effects, and allows treatment of relatively large systems with a reasonable computational cost.

Hartree-Fock Methods

The Hartree-Fock method takes into account the Pauli principle for electrons writing the total electronic wavefunction as *single* (antisymmetric) Slater determinant of the spin-orbitals $\psi_i(\mathbf{x}) = \phi_i(\mathbf{r})\,\sigma(s)$ [141], where $\sigma(s) = \alpha(s)$ or $\beta(s)$:

$$\Psi_{HF} = \frac{1}{\sqrt{(N!)}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix}. \tag{2.55}$$

Minimizing the expectation value of the Hamiltonian $\mathcal{H}_e$ with respect to the set $\{\psi_i\}$ subject to the orthonormalization conditions $\langle\psi_i|\psi_j\rangle = \int d\mathbf{x}, \psi_i^*(\mathbf{x})\psi_j(\mathbf{x}) = \delta_{ij}$ gives, after diagonalization through a unitary operator $\mathbf{U}$, the canonical Hartree-Fock system of equations:

$$F_i\phi_i = \epsilon_i\phi_i, \tag{2.56}$$

with

$$F_i = \underbrace{-\frac{\nabla_i^2}{2} + \sum_I \frac{Z_I e}{|\mathbf{r_i} - \mathbf{R_I}|}}_{h_i} + \sum_{j=1}^N (J_j - K_j). \tag{2.57}$$

The so-called Fock operator F_i is an effective one-electron operator describing the kinetic energy of an electron, the attraction to all the nuclei, and the repulsion between electrons, through the Coulomb and exchange operators:

$$J_j(\mathbf{x}_1)\,\phi_i(\mathbf{x}_1) \equiv \sum_{k=1}^N \int d\mathbf{x}_2 \phi_k^*(\mathbf{x}_2)\phi_k(\mathbf{x}_2)\,\mathbf{g}_{12}\,\phi_i(\mathbf{x}_1), \tag{2.58}$$

$$K_j(\mathbf{x}_1)\,\phi_i(\mathbf{x}_1) \equiv \sum_{k=1}^N \int d\mathbf{x}_2 \phi_k^*(\mathbf{x}_2)\phi_i(\mathbf{x}_2)\,\mathbf{g}_{12}\,\phi_k(\mathbf{x}_1), \tag{2.59}$$

with

$$\mathbf{g}_{ij} \equiv \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{2.60}$$

The exchange operator has a non-local character, and is the term that accounts for the exclusion principle of Pauli. The expectation value of the Fock operator

$$\varepsilon_l = \langle\phi_l|F_l|\phi_l\rangle \tag{2.61}$$

can be interpreted as the energy of the *l*-th MO, which in the limit of frozen orbitals is equal to minus the ionization energy I_l of the *l*-th electron (Koopmans' theorem [142]). The total energy

$$E = \sum_{i=1}^{N} \varepsilon_i - \frac{1}{2} \sum_{ij=1}^{N} (J_{ij} - K_{ij}) + V_{nn} \tag{2.62}$$

is not simply the sum of MO energies, because the Fock operator contains terms describing the repulsion of each MO to *all* other electrons, and thus the sum over ε_i counts the electron-electron repulsion twice, which have to be corrected by the second term. Moreover, this total energy cannot be exact, as the electron-electron repulsion is only accounted for in an average fashion, due to the approximation of a single Slater determinant as the trial wave function. The absence of correlation among electrons can be included within a perturbative scheme, like the one of Møller-Plesset.

Møller-Plesset perturbation theory

In the Møller-Plesset scheme [143, 144] the unperturbed Hamiltonian H_0 is taken to be a sum over Fock operators. As this sum counts twice the (average) electron-electron repulsion, the perturbation H_1 becomes the exact V_{ee} operator minus twice $\langle V_{ee} \rangle$ (also called *fluctuation potential*:

$$H_0 = \sum_{i=1}^{N} F_i = \sum_{i=1}^{N} \left(h_i + \sum_{j=1}^{N} (J_j - K_j) \right) = \sum_{i=1}^{N} h_i + 2 \langle V_{ee} \rangle \tag{2.63}$$

$$H_1 = V_{ee} - 2 \langle V_{ee} \rangle \tag{2.64}$$

The zero-order wave function is the HF determinant, while the first (MP1) order correction to the energy is given by the average electron-electron repulsion changed in sign. Electron correlation enters at the MP2 level, and involves only a sum over doubly excited determinants (if canonical HF orbitals are used) [144]:

$$E^{\text{MP2}} = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\iint d\mathbf{r}_1 d\mathbf{r}_2 \phi_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2) \mathbf{g}_{12} \left[\phi_a(\mathbf{r}_1) \phi_b(\mathbf{r}_2) - \phi_b(\mathbf{r}_1) \phi_a(\mathbf{r}_2) \right]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} . \tag{2.65}$$

The MP2 correction typically recovers 80-90% of the correlation energy, at a cost roughly twice as that for solving HF equations in practical calculations (this because only two-electron integrals corresponding to two combination of two occupied and two virtual MOs are required in eq. 2.65). Moreover, for "well-behaved" systems, MP2 usually gives better results than MP3 [144]. Including higher terms in

the perturbation is not very common, as other methods become competitive [144], like Configuration Interaction (CI), which has the advantage of being intrinsically multi-reference.

Density Functional Theory

Density Functional Theory (DFT) is a rigorous method to find the ground state of many particle system [140]. The main idea lies in assumption that the ground-state properties of a quantum system of N particles can be described starting from its density $\rho(\mathbf{r})$

$$\rho(\mathbf{r}) = N \int |\Psi(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N)|^2 \, d\mathbf{r}_2 d\mathbf{r}_3...d\mathbf{r}_N \,. \tag{2.66}$$

This has two main advantages over the other techniques:

- density is an observable that can be easily measured and visualized.

- the dimensionality of the problem is reduced from $3N$ to 3, as the density is a function of space.

The Hohenberg-Kohn theorems. The use of the density as fundamental quantity is based on the two Hohenberg-Kohn theorems, enunciated in the early sixties [140]. The first theorem demonstrates that, given a Hamiltonian characterized by a general external potential V_{ext}, the ground-state density $\rho(\mathbf{r})$ associated to it is unique. As V_{ext} univocally determines the Hamiltonian of the system, it follows that the ground state wavefunction, and thus all the observables are functionals of the density ρ. The second theorem provides a variational principle for the ground state density: given any trial density $\bar{\rho} > 0$ for which $\int \bar{\rho}(\mathbf{r})d\mathbf{r} = N$, it follows that $E[\bar{\rho}] \geq E[\rho]$. From this result, one can get a variational equation to obtain the ground-state energy. Let apply the Hohenberg-Kohn theorem to a system of N electrons in which the external potential is due to the nuclei. The energy in terms of the electronic density reads:

$$E[\rho] = T_e[\rho] + V_{ee}[\rho] + V_{eN}[\rho] + V_{NN} = F[\rho] + \int d\mathbf{r}\rho(\mathbf{r})V_{ext} \tag{2.67}$$

where $F[\rho] = T_e[\rho] + V_{ee}[\rho] = \langle \psi | T_e + V_{ee} | \psi \rangle$ is a universal functional independent from the external potential $V_{ext}[\rho] = V_{eN}[\rho] + V_{NN}$.

Applying to ρ the stationary principle

$$\delta \left\{ E[\rho] - \mu \left[\int \rho(\mathbf{r})d\mathbf{r} - N \right] \right\} = 0, \tag{2.68}$$

we obtain the Euler-Lagrange equation for the multiplier μ:

$$\mu = V_{ext}(\mathbf{r}) + \frac{\partial F[\rho]}{\partial \rho} \tag{2.69}$$

Although DFT is formally a rigorous method, the application of the variational principle requires in practice an explicit form of the functional F. Kohn and Sham suggested to decompose it in parts whose only the most important need to be treated exactly [145].

Kohn-Sham equations. The main idea of the Kohn-Sham method lies in the possibility of mapping a system of N interacting particles into an equivalent one of non-interacting bodies, characterized by the same ground state density [145]. For such systems, the density can be written as a summation over single-particle contributions:

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} \left| \varphi_i^{KS}(\mathbf{r}) \right|^2 \tag{2.70}$$

and the kinetic energy functional has an analytical expression:

$$T_0[\rho] = \sum_{i=1}^{N} \left\langle \varphi_i^{KS} \left| -\frac{1}{2}\nabla^2 \right| \varphi_i^{KS} \right\rangle \tag{2.71}$$

The functional $F[\rho]$ can be rewritten as:

$$F[\rho] = T_0[\rho] + V_H[\rho] + E_{xc}[\rho] \tag{2.72}$$

where $V_H = \frac{1}{2} \int d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ is the classical part of the particle-particle interaction, and the 'exchange-correlation' functional E_{xc} is defined as:

$$E_{xc}[\rho] = T[\rho] - T_0[\rho] + V_{ee}[\rho] - V_H[\rho] \tag{2.73}$$

Thus all the unknowns of the problem are put into E_{xc}, which sums the corrections in the kinetic energy and of the non-classical part of the particle-particle interaction. The echange-correlation term describes the lowering in energy gained by a system of interacting electrons with respect to the Fermi gas, and has therefore a negative sign. Formally E_{xc} can be written in terms of an exchange-correlation energy per particle ε_{xc}, which is itself functional of the total density:

$$E_{xc}[\rho] = \int d\mathbf{r}\rho(\mathbf{r})\varepsilon_{xc}[\rho] \tag{2.74}$$

Equation 2.69, turns out to be:

$$\mu = V^{KS}(\mathbf{r}) + \frac{\partial T_0[\rho]}{\partial \rho} \tag{2.75}$$

with

$$V^{KS} = V_{ext}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{xc}[\rho] \tag{2.76}$$

where we have defined the exchange correlation potential

$$V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho} \tag{2.77}$$

Eq. 2.75 says that we can solve the original problem by finding the ground state energy for a system of non-interacting electrons in a effective potential. The single-particle orbitals describing these electrons solve the self-consistent Kohn-Sham (KS) equations:

$$[\frac{1}{2}\nabla^2 + V_{KS}(\mathbf{r})]\varphi_i^{KS} = \varepsilon_i \varphi_i^{KS} \qquad i = 1, ..., N \tag{2.78}$$

The total energy of the system is not the sum of KS eigenvalues, but can be expressed as

$$E = \sum_{i=1}^{N} \varepsilon_i - \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[\rho] - \int d\mathbf{r} V_{xc}(\mathbf{r}) \, \rho(\mathbf{r}) \tag{2.79}$$

The method of Kohn and Sham shifts the complexity of the problem on finding a suitable analytical formulation of the exchange-correlation functional. In practice this is not possible, so once again approximated expressions have been derived for E_{xc}.

Exchange-Correlation functionals

Local Density Approximation.   An approximation for the exchange-correlation functional has been proposed already in the original paper by Hohenberg and Kohn [140, 141]. They recover the idea beyond Thomas-Fermi approximation of the kinetic energy for an homogeneous electron gas, and apply it to the evaluation of $E_{xc}[\rho]$. The exchange-correlation energy density in $\mathbf{r}$ is assumed to be *local*, i.e. only depends on the value of ρ in $\mathbf{r}$ itself (here the name Local Density Approximation, LDA): $\varepsilon_{xc}[\rho] = \varepsilon_{xc}(\rho(\mathbf{r}))$. In addition, ε_{xc} is approximated by that of an homogeneous gas

of electrons of density $\rho^{hom} = \rho(\mathbf{r})$ (in a uniform background of positive charge). Thus

$$\varepsilon_{xc}^{LDA}[\rho] = \varepsilon_{xc}^{hom}(\rho(\mathbf{r})) \tag{2.80}$$

The simplification introduced by LDA becomes clear if one divides ε_{xc} into exchange and correlation contributions ε_x and ε_c. In fact, for a homogeneous electron gas ε_x is known exactly [140, 141, 146], and is proportional to the cubic square of the density:

$$\varepsilon_x^{LDA}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3}(\mathbf{r}) \tag{2.81}$$

The situation is more complicated for the correlation term, which has been determined analytically in the high and low density limit [147, 148], and by Quantum Monte Carlo calculations for intermediate states [149]. Suitable (approximate) analytical formulas have been derived by Vosko, Wilk and Nusair (VWN correlation functional) [150] and by Perdew and Wang (PW) [151]. The main reason behind success of LDA is most probably a partial cancellation of errors. In fact, LDA typically underestimates E_c but overestimates E_x, resulting in unexpectedly good values of E_{xc}. However, for molecular systems ε_x is underestimated by a factor of 10, leading to errors larger than the whole correlation energy (overestimated by a factor 2), and bond energies up to 25 kcal/mol larger than experimental values [144]. In addition, LDA exhibits heavy deficiencies in describing hydrogen-bonds, which are crucial for studies on biologically relevant systems [152, 153].

Generalized Gradient Approximation. The General Gradient Approximation (GGA) successfully improves the accuracy of DFT by introducing the gradient[4] of the density in the functional form of E_{xc}:

$$E_{xc}^{GGA}[\rho] = \int d(\mathbf{r})f(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})) \tag{2.82}$$

Quite generally, GGAs functionals give good results for all the main bond types (covalent, ionic, metallic and hydrogen bonds) (see for example [146] and Refs. therein). For Van der Waals interactions, however, common GGAs and LDA fail. To treat these weak interactions more specialized approaches have been developed [154, 155] but I will not treat them in detail here. The GGA functionals used in biochemistry typically are derived by fitting parameters on the properties of sets of molecules.

---

[4] Notice here the difference from the gradient-expansion-approximation (GEA), where one tries to systematically calculate gradient-corrections to LDA of the form $|\nabla\rho|$, $\nabla^2\rho$, $|\nabla\rho|^2$. In contrast to GGA, GEA shown no improvement with respect to LDA, because of the loss of some important properties of the exchange-correlation hole [144].

Nowadays the most popular and reliable GGA functional used in biochemistry is BLYP, which also has been used in the work exposed in this thesis. The acronym BLYP denotes the combination of Becke's exchange functional B88 [156], and the correlation functional of Lee, Yang and Parr (LYP) [157], both introduced in 1988.

Becke exchange functional. This correction was introduced by Becke [156] to reproduce the exact asymptotic behaviour of the exchange energy. The analytical formula reads:

$$E_x[\rho]^{B88} = E_{xc}^{LDA} + \Delta E_x^{B88} \tag{2.83}$$

where

$$\Delta E_x^{B88} = -\beta \int d\mathbf{r} \rho^{\frac{4}{3}}(\mathbf{r}) \frac{x^2}{1+6\beta \sinh^{-1} x} \tag{2.84}$$

$$x = \frac{|\nabla \rho|}{\rho^{\frac{4}{3}}} \tag{2.85}$$

x gives a measure of the local inhomogeneity of the system, and the parameter β has been fixed to 0.0042 a.u. by a fit on known Hartree-Fock data for the rare gas atoms.

Lee-Yang-Parr correlation functional. The LYP functional [157] for the correlation energy was derived from the Colle-Salvetti formula, and computes correlation energies from HF second order density matrices. Its expression is the following:

$$E_c^{LYP}[\rho] = -a \int d\mathbf{r} \frac{1}{1+d\rho^{-\frac{1}{3}}} \left\{ \rho + b\rho^{-\frac{2}{3}} \left[C_F \rho^{\frac{5}{3}} - 2t_W + \left(\frac{1}{9}t_W + \frac{1}{18}\nabla^2 \rho \right) \right] e^{-c\rho^{-\frac{1}{3}}} \right\} \tag{2.86}$$

where $C_F = \frac{3}{10}(3\pi^2)^{\frac{2}{3}}$, $t_W = \frac{1}{8}\frac{|\nabla \rho|^2}{\rho} - \frac{1}{8}\nabla^2 \rho$. The parameters a=0.04918, b=0.132, c=2533 and d=0.349 are obtained by fitting the functional formula on HF calculations for the Helium atom.

Hybrid functionals. The main problem in DFT practical implementations arise from the approximation inherent in the exchange-correlation potential introduces a Self Interaction Error (SIE) that does not conveniently cancels as in HF theory [144]. This leads to serious drawbacks in calculations of thermochemistry of molecules using the LDA or GGA standard functionals. Relative energies of states with different spin multiplicity are often poorly described, radicals or atom transfer transition structures are predicted to be too stable, and activation barriers are severely underestimated [83, 158]. These considerations have pushed Becke to include *exact* exchange

into the functional form of E_{xc}. Theoretical instrument for this proposal is the adiabatic connection method [159], that gives a link between the non-interacting reference system and the real fully interacting system:

$$E_{xc} = \int_0^1 \frac{dE_{xc}^\lambda}{d\lambda} d\lambda \qquad (2.87)$$

Here λ is an inter-electronic coupling strength parameter taking the system from no Coulomb interaction at $\lambda = 0$ to full correlation at $\lambda = 1$. In the limit $\lambda = 0$ we have only exchange energy, *exactly* the one given by the HF theory [5]. Following this idea people tried to decompose E_{xc} as a combination of LDA, exact exchange and a gradient correction term [160]. Coefficients have been both calculated by fitting of some properties [161, 162] or derived from requirements on the correct behaviour of V_{xc} [163]. Among these hybrid functionals the B3LYP [161], employed in the work exposed here, is probably the most used in biological calculations:

$$E_{xc}^{B3LYP} = (1-a)E_{xc}^{LDA} + aE_x^{exact} + b\Delta E_x^{B88} + (1-c)E_c^{LDA} + cE_c^{LYP} \quad (2.88)$$

The a, b and c parameters are determined by fitting to experimental data and depend on the chosen forms for E_x^{GGA} and E_c^{GGA}, with typical values being $a \sim 0.2$, $b \sim 0.7$ and $c \sim 0.8$. Hybrid functionals represent today the better choice in terms of performance vs accuracy.

2.4.3 Basis Set approximation

In actual implementations of (Post)HF or DFT-Kohn-Sham schemes, the MOs are usually expanded in terms of M_b basis functions of well-known behavior

$$\varphi_i = \sum_\alpha^{M_b} c_{i\alpha} \chi_\alpha \qquad (2.89)$$

The mathematical problem is thus transformed into that of solving a secular matrix equation, in which the matrix elements are calculated from arrays of integrals evaluated for the given basis functions. Taking HF equations 2.56 as example, and expanding the eigenfunctions as above, one obtain the famous Roothaan-Hall equations (for a closed shell system) [144]:

$$\mathbf{FC} = \mathbf{SC}\varepsilon \qquad (2.90)$$

---

[5]This energy would be equal to that calculated with HF wavefunction methods if the the KS orbitals were identical to the HF ones.

where $\mathbf{S}$ is the matrix describing basis set functions overlap ($S_{\alpha\beta} = \langle\chi_\alpha|\chi_\beta\rangle$) and the matrix elements of the Fock operator are written as sum of one-electron integrals and products of a *density* matrix with two-electron integrals

$$F_{\alpha\beta} = \langle\chi_\alpha|\mathbf{F}|\chi_\beta\rangle = h_{\alpha\beta} + \sum_{\gamma\delta} G_{\alpha\beta\gamma\delta}\, D_{\gamma\delta} \tag{2.91}$$

$$h_{\alpha\beta} = \langle\chi_\alpha|\mathbf{h}|\chi_\beta\rangle = \int \mathrm{d}\mathbf{r}_1\chi_\alpha(1)\frac{-\nabla^2}{2}\chi_\beta(1) + \sum_a^N \int \mathrm{d}\mathbf{r}_1\chi_\alpha(1)\frac{Z_a}{|\mathbf{R}_a - \mathbf{r}_1|}\chi_\beta(1) \tag{2.92}$$

$$G_{\alpha\beta\gamma\delta} = \langle\chi_\alpha(1)\chi_\beta(2)|\mathbf{g}_{12}|\chi_\gamma(1)\chi_\delta(2)\rangle = \int \mathrm{d}\mathbf{r}_1\chi_\alpha(1)\chi_\beta(2)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\chi_\beta(1)\chi_\delta(2) \tag{2.93}$$

$$D_{\gamma\delta} = \sum_j^{occ.MO} c_{\gamma j}c_{\delta j} \tag{2.94}$$

Essentially two philosophies exist for the construction of a basis set, one using localized atomic orbitals (AO), the other delocalized plane-waves (PW) for the expansion of MOs.

Localized basis sets

The basis sets are composed by localized and usually atomic-centered functions [6], firstly introduced by Slater, that resemble hydrogen-like eigenfunctions [164]

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{n-1}\mathrm{e}^{-\zeta r} \tag{2.95}$$

Slater Type Orbitals (STO) are certainly the best suited for electronic structure calculations. However, the evaluation of especially three and four center two-electrons integrals is very expensive using such an exponential functions. So, for practical purposes STO are almost universally replaced by Gaussian Type Orbitals (GTO) [165]

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{2n-2-l}\mathrm{e}^{-\zeta r^2} \tag{2.96}$$

In fact, the product of two Gaussians located at different centers have the (nice) property of being a *Gaussian* centered at the intermediate position, that greatly improves the efficiency in calculating two-electrons integrals. Obviously a single GTO does not reproduce as well as an STO the proper behavior of the wavefunction (in particular near to the nucleus and for large r), so three times as many GTOs as STOs

---

[6]For a recent and exhaustive description of AO type basis sets see e.g. the book by Jensen [144].

are roughly required to reach a given accuracy. This theoretical disadvantage is more than compensated by the overall gain in computational time. The quality of a calculation depends obviously on the number of functions used in the expansion. A double *zeta basis* (DZ) set, in which the number of basis functions is twice what is needed to contain all the electrons of neutral atoms, is considered "good" for organic molecules. Most often only valence orbitals are doubled, while core states are described with the smallest number of functions possible (minimum basis set), which gives the *double zeta valence split basis* (VDZ). Often polarization and diffuse functions are added to the basis set to improve the description of electronic correlation and polarization, and systems with loosely bound electrons (see Ref. [144] for further details). MOs are thus expanded as linear combination of a given number of GTOs with different exponents ζ (*primitives*, PGTO). As these are determined by an energy-based variational procedure, most of them are "well-tuned" on core-states, which are energetically but not chemically relevant. To improve efficiency *contracted* basis sets have been introduced. The idea is to combine a given set of primitives into a smaller set of (contracted, CGTO) orbitals, each one being a linear combination of a given number of PGTO with *fixed* coefficients. The acronyms DZ, VDZ, etc. always refer to the number of contracted basis functions. Calculations reported in this thesis have been performed using the VDZ Pople-style k-nlmG basis set described in the next paragraph.

Pople-style *k-nlmG* **basis sets.** In this basis set the k indicates the number of PGTOs used for representing the core orbitals, while nlm indicates both how many functions the valence orbitals are split into, and how many PGTOs are used for their representation. Two values (nl) indicate a split valence, three (nlm) a triple split valence. Polarization functions are specified after the G. The most used basis set of this kind is the 6-31G one [166], in which the core orbitals are a contraction of six PGTOs, the inner part of valence orbitals is contraction of three, and the outer part is represented by one PGTO.

Plane waves

Following the Bloch theorem [167] for periodic systems, a one-particle wave-function can be written as Fourier's series:

$$\varphi^{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{g}} c_j^{\mathbf{k}}(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}} \tag{2.97}$$

where V is the volume of the cell, $\mathbf{k}$ vectors belong to the first Brillouin zone, $\mathbf{g}$ is a reciprocal lattice vector, c is the first Fourier component of the plane waves expansion, and the summation is extended to infinite lattice vectors In the treatment of

isolated clusters with a low symmetry, such as, organic molecules or the active sites of enzymes, the Γ-point approximation ($\mathbf{k}$=0) still guarantees a good accuracy, leading to a relevant reduction of the computational cost. 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 [168], which linked the expression of the electrostatic potential energy of a cluster to that of an infinitely periodic system, allowing thus to take advantage of the Ewald method (see Section 2.3.3 and Appendix C) for isolated systems treated with a plane-wave orbital expansion. The algorithm is based on the assumption that the electrostatic potential Φ can be written as the sum of a short and long-range contributions:

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

In a plane-wave expansion, the potential energy due to Φ can be expressed as

$$U = \frac{1}{2V}\sum_{\mathbf{g}}|\bar{\rho}(\mathbf{g})|^2\,\bar{\Phi}(-\mathbf{g}) = \frac{1}{2V}\sum_{\mathbf{g}}|\bar{\rho}(\mathbf{g})|^2\left[\bar{\phi}^{short}(-\mathbf{g}) + \bar{\phi}^{long}(-\mathbf{g})\right] \quad (2.98)$$

where $\bar{\rho}(\mathbf{g})$ and $\bar{\Phi}(-\mathbf{g})$ are the *finite* Fourier series of the density and the potential, respectively, the finite Fourier series of $f(\mathbf{r})$ being:

$$\bar{f}(\mathbf{r}) = \int_V d\mathbf{r}\, e^{-i\mathbf{g}\cdot\mathbf{r}} f(\mathbf{r}) \tag{2.99}$$

Requiring $\phi^{short}(\mathbf{r})$ to vanish exponentially quickly at large distances from the center of the cluster ($d \sim 2d_s$, where d_s is the typical system size), we can express this function as a Fourier transform plus an (exponentially vanishing) error term:

$$\begin{aligned}
\bar{\phi}^{short}(\mathbf{g}) &= \int_V d\mathbf{r}\, e^{-i\mathbf{g}\cdot\mathbf{r}} \phi^{short}(\mathbf{r}) \\
&= \int_{all\ space} d\mathbf{r}\, e^{-i\mathbf{g}\cdot\mathbf{r}} \phi^{short}(\mathbf{r}) + \varepsilon(\mathbf{g}) \\
&= \tilde{\phi}^{short}(\mathbf{g}) + \varepsilon(\mathbf{g})
\end{aligned} \tag{2.100}$$

where $\tilde{\phi}^{short}(\mathbf{g})$ is the Fourier transform of $\phi^{short}(\mathbf{r})$. Thus, neglecting the error term:

$$\begin{aligned}
\bar{\phi}(\mathbf{g}) &= \tilde{\phi}^{short}(\mathbf{g}) + \bar{\phi}^{long}(\mathbf{g}) \\
&= \tilde{\phi}^{short}(\mathbf{g}) + \bar{\phi}^{long}(\mathbf{g}) + \tilde{\phi}^{long}(\mathbf{g}) - \tilde{\phi}^{long}(\mathbf{g}) \\
&= \tilde{\phi}(\mathbf{g}) + \tilde{\phi}^{screen}(\mathbf{g})
\end{aligned} \tag{2.101}$$

The function

$$\tilde{\phi}^{screen}(\mathbf{g}) = \bar{\phi}^{long}(\mathbf{g}) - \tilde{\phi}^{long}(\mathbf{g}) \tag{2.102}$$

screens the interaction of the cluster with an infinite array of periodic images, realizing the link between isolated and replicated systems (its value is zero in this latter case). Thus, the potential energy can be written as:

$$U = \frac{1}{2V} \sum_{\mathbf{g}} |\bar{\rho}(\mathbf{g})|^2 \left[\tilde{\Phi}(-\mathbf{g}) + \tilde{\phi}^{screen}(-\mathbf{g}) \right] \tag{2.103}$$

In this expression $\tilde{\phi}^{long}(\mathbf{g})$ is supposed to be known, while $\bar{\phi}^{long}(\mathbf{g})$ can be evaluated efficiently over a grid using Fast Fourier Transform [139], which scales as $\mathcal{O}(N \ln N)$.

Pseudopotentials. The greatest drawback in using a plane-wave basis-set comes from the impossibility, from a practical point of view, of describing core electrons within a reasonable computational expense. Indeed, the sharp spatial oscillations of their wave-functions near to the nuclei would require an extremely high number of plane-waves for an accurate characterization. On the other hand, the core levels are well separated in energy from valence electrons, and, at a first level of approximation, do not play any role in the chemical properties of molecular systems. Thus, the core electron orbitals can be frozen in the KS equations and only the valence electrons are described explicitly. The core-valence electron interactions are implicitly included into the nuclear potential, which becomes an "effective-potential" or "pseudopotential". Pseudopotentials are usually derived from all electron (AE) atomic calculations, and several recipes have been proposed to date. In the work presented in this thesis "norm-conserving" pseudopotentials derived from the Martins-Troullier method [169] have been used. Pseudopotentials have to satisfy the following conditions:

- The valence pseudo-wave-function should not contain any radial nodes.

- The valence AE and pseudopotential eigenvalues from the radial KS equations must be the same:
$$\varepsilon_\ell^{PP} = \varepsilon_\ell^{AE} \tag{2.104}$$
where ℓ is the angular momentum.

- The pseudo and AE atomic radial wave-functions must be equal for r greater than a chosen cut-off distance r_{cut}.

These three conditions ensure that the pseudo-atom behaves like the real one in the region of interaction with other atoms while forming chemical bonds. Other conditions are the following:

- The integrated electron density within the cut-off radius for the two wave-functions must be the same. This requirement guarantees the transferability and the norm conserving rule of the MT pseudopotential.

- At $r = r_{cut}$, the pseudo wave-function and its first four derivatives should be continuous.

- The pseudopotentials should have zero curvature at the origin.

With these conditions, the general form for a pseudopotential wave-function is:

$$\varphi_\ell^{PP}(r) = \begin{cases} \varphi_\ell^{AE}(r) \; ; \; r > r_{cut} \\ r^\ell e^{p(r)}; \; r \le r_{cut} \end{cases} \qquad (2.105)$$

where $p(r) = c_0 + \sum_{i=1}^{6} c_i r^{2i}$, and the coefficients are obtained by imposing the first three conditions.

The functional form of the pseudopotential is

$$V_{pseudo} = V_{val}(r) + \sum_{m,l} |Y_{l,m}\rangle V_l(r) \langle Y_{l,m}| \qquad (2.106)$$

where $|Y_{l,m}\rangle$ are spherical harmonics. The "semilocality" of this functional form (local in the radial coordinate, non local in the angular ones), implies an increase in the computational cost. This difficulty can be overcome by using the method of Kleinman-Bylander [170], which implies addition and subtraction of an "ad-hoc" radial function $V_L(r)$ to the pseudopotential, leading to a new functional form, where the local and non-local parts can be completely separated.

2.5 Born-Oppenheimer approximation

We have seen that the Ehrenfest molecular dynamics scheme, eqs. 2.45 and 2.46, allows to propagate the electronic system by solving the time-dependent Schrödinger equation "on the fly", as the nuclear configuration changes under the force $\nabla_I \langle \mathcal{H}_e \rangle$. Unfortunately there is a major problem with practical implementation of the Ehrenfest scheme: the time scale and thus the time step used to integrate eqs. 2.45 and 2.46 simultaneously is dictated by the intrinsic dynamics of electrons. Now, typical vibrational and angular frequencies in biological systems rise up to $3000-4000$ cm^{-1} (for example bond frequencies in water are ~ 3500 cm^{-1} [171]), which correspond to a

timescale of $\tau_N \sim 10^{-14}$. This time interval is two order of magnitude larger than the maximum time step Δt_e^{max} necessary to integrate correctly electron dynamics. Thus, there is a bottleneck limiting the efficient implementation of such a simultaneous evolution of electronic and nuclear systems. A solution to this problem is the Born-Oppenheimer (BO) approximation, which was proposed in the early days of quantum mechanics (1927) [172]. In the BO scheme the strong dynamical separation between electronic and nuclear motions is exploited to increase the maximum time step used to propagate the nuclei. Since atoms are about three orders of magnitude heavier than electrons, the latter are supposed to follow *instantaneously* the motion of the nuclei, staying always in the same stationary state of the Hamiltonian. This stationary state will vary with the configuration of nuclei because of the Coulomb coupling between the two sets of degrees of freedom, but non-radiative transitions like those from *phonon-electron* interactions are negligible. This is obviously true only if the energy separation between the ground and the first excited state is larger than typical phonon energies. In this case, one can solve the time-independent Schrödinger equation at *fixed* positions of the nuclei, move them under the action of the effective electronic potential, and iterate the process. The equations describing the so-called Born-Oppenheimer approximation (for the ground state) are then [7].

$$\mathcal{H}_e \Psi_0 = E_0 \Psi_0 \tag{2.107}$$
$$M_I \ddot{\mathbf{R}}(t) = -\nabla_I \, min_\Psi \left\{ \langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle \right\} \tag{2.108}$$

At opposite to Ehrenfest dynamics, now time dependence of the electronic structure is only *implicit* through the motion of nuclei. This allows for time steps $\Delta t_N^{max} \sim \tau_N/10$. However, the bottleneck of Born-Oppenheimer dynamics is that at each MD step the electronic wavefunction needs to be relaxed.

2.6 Car-Parrinello molecular dynamics

The drawbacks of both Ehrenfest and Born-Oppenheimer molecular dynamics schemes stimulated the search for new algorithms that could overcome their limitations. In 1985 Car and Parrinello developed a new scheme, based on the extended Lagrangian formalism and avoiding the optimization of the electronic wavefunction by introducing a second order *fictious* dynamics on the electrons. These latter are kept sufficiently close to the adiabatic surface, allowing for an increase of the time step by a factor ~ 10 with respect to Ehrenfest dynamics. The method is based on the observation that $\langle \Psi_0 | \mathcal{H} | \Psi_0 \rangle$ can be viewed not only as a function of $\{\mathbf{R}_I\}$, but also as a

---

[7]Note that Born-Oppenheimer approximation slightly differs from the so-called "adiabatic" one for presence in the latter of a diagonal correction term containing the expectation value of the nuclear kinetic energy operator on the electronic wavefunction [144].

functional of the wavefunction Ψ_0, and thus of the set of one-electron orbitals $\{\psi_i\}$ used to build it. In this case, the force acting on these orbitals can be obtained from a functional derivative of a suitable Lagrangian containing $\langle\Psi_0|\mathcal{H}|\Psi_0\rangle$, like in classical mechanics for the nuclear motion. The Lagrangian $\mathcal{L}$ proposed by Car and Parrinello has the form

$$\mathcal{L} = \sum_{i=1}^{M} \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_{i=1}^{N} \frac{1}{2}\mu_i \left\langle \dot{\psi}_i \middle| \dot{\psi}_i \right\rangle - \langle\Psi_0|\mathcal{H}_e|\Psi_0\rangle + constraints \qquad (2.109)$$

where the first term is the kinetic energy of nuclei, and $\mu_i = \mu$ are the "fictious masses" assigned to orbitals; the second term represents the fictious kinetic energy associated to them (the sum is on the occupied orbital only). The (holonomic) constraints act in general on both the orbitals (e.g. to guarantee orthonormality) and on the nuclei (e.g. if one would perform molecular dynamics with geometric restraints). The dynamics is described by the Euler-Lagrange equations associated to $\mathcal{L}$

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\mathbf{R}}_I} = M_I\ddot{\mathbf{R}}_I = \frac{\partial\mathcal{L}}{\partial\mathbf{R}_I} = -\frac{\partial}{\partial\mathbf{R}_I}\langle\Psi_0|\mathcal{H}_e|\Psi_0\rangle + \frac{\partial}{\partial\mathbf{R}_I}\{constraints\} \text{ (2.110)}$$

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\psi}_i^*} = \mu_i\ddot{\psi}_i^* = \frac{\partial\mathcal{L}}{\partial\psi_i^*} = -\frac{\partial}{\partial\psi_i^*}\langle\Psi_0|\mathcal{H}_e|\Psi_0\rangle + \frac{\partial}{\partial\psi_i^*}\{constraints\} \text{ (2.111)}$$

Note that if $|\mu\ddot{\psi}_i| \to 0$ eq. 2.111 reduces to a stationary problem, and the electronic system will stay on the Born-Oppenheimer surface (no forces acting on orbitals), corresponding to the true equilibrium dynamics. Higher is the fictious kinetic energy $T_e = \sum_{i=1}^{N} \frac{1}{2}\mu_i \left\langle \dot{\psi}_i \middle| \dot{\psi}_i \right\rangle$, more the electrons will be far from the minimum energy configuration. In particular, a ground state wavefunction optimized at time t_0 will stay close to its ground state if it is kept at sufficiently low temperature. The only quantity one can change to ensure this condition is μ, often called "adiabacity parameter" [128]. If μ and τ are chosen consistently the energy flow between electronic and nuclear subsystems is slow enough to cause no drift in T_e [8], thus conserving the "physical" energy E_{phys}:

$$E_{phys} = E_{tot} - T_e = T_I + V_e = \sum_{i=1}^{M} \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \langle\Psi_0|\mathcal{H}_e|\Psi_0\rangle \qquad (2.112)$$

---

[8] T_e actually performs two-frequency *bound* oscillations around a constant value. The first frequency is associated to the drag exerted by the nuclei, and it is in anti-phase with V_e oscillations, while the second is a small-amplitude high-frequency oscillation intrinsic to the fictious electronic dynamics. Note that having a nonvanishing masses, also the electrons dampen nuclear motion, causing a renormalization of the nuclear masses which can be important in the case of light atoms.

The choice of a reasonable fictious mass μ meets two opposite requirements. In fact, considering a simple harmonic model for the electronic system around the BO surface, (discrete) excitations frequencies are given by

$$\omega_{ij}^e = \sqrt{\frac{2(\varepsilon_i^* - \varepsilon_j)}{\mu}} \tag{2.113}$$

where ε^* and ε are energy levels of unoccupied and occupied orbitals, respectively. If ω_{max}^n is the maximum vibrational frequency of the nuclear system, in order to perform adiabatic dynamics it must be $\omega_{gap}^e >> \omega_{max}^n$. As the only tunable parameter is μ, one could decrease it arbitrarily to increase the frequency of the gap ω_{min}^e. However, decreasing μ stretches the entire spectrum $\{\omega_{ij}^e\}$ and in particular increases ω_{max}^e, which is inversely proportional to the maximum time step. Typical values of μ are in the range $500 - 1500$ a.u., which allow for a time step of about $5 - 10$ a.u. $(0.12 - 0.24\,\text{fs})$. For calculations discussed here we used $\mu = 600$ a.u. and a time step of 5 a.u.

2.7 Hybrid Models

Pure Quantum calculations are today restricted to the treatment of at most a few hundreds of atoms. Classical Molecular Mechanics, on the other hand, can deal with systems containing $10^5 - 10^6$ atoms, but cannot take into account the quantum nature of chemical bonds. Since most of times the relevant chemistry of a biological process is restricted to a small subset of atoms, hybrid schemes have been developed that model different parts of the system at a different level of modeling [173,174,175,176]. These schemes allow to evaluate the effect of the biological environment on chemical processes, and represents thus an improvement over a quantum calculation in vacuo. In particular a widely adopted approach is to partition the system into two regions and to treat one at Quantum Mechanics and the other at Molecular Mechanics levels. Such approach, as implemented in the CPMD code [177], has been used in the works reported in this thesis, and is based on a single hybrid Hamiltonian:

$$H = H_{QM} + H_{MM} + H_{QM/MM} \tag{2.114}$$

where H_{QM} is the quantum Hamiltonian, H_{MM} is the Molecular Mechanics Hamiltonian and $H_{QM/MM}$ is the Hamiltonian describing the interaction between the two subsystems. For the purpose of describing each term in eq. 2.114, let start by considering the total system (QM+MM) as described uniformly at quantum level and by (artificially) partitioning the system in the QM and MM regions. According to the Hohenberg-Kohn theorem, the total energy of the system is given by the following

functional:

$$E[\rho] = T[\rho] + \int_\Omega V^{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int\int\frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}d\mathbf{r}_1 d\mathbf{r}_2 + \frac{1}{2}\sum_{I,J}\frac{Z_I Z_J}{R_{IJ}} + E_{xc}[\rho]$$

(2.115)

where T and E_{xc} are the kinetic and exchange-correlation energy functionals, respectively; V^{ext} is the electrostatic potential of the nuclei, r_{12} and R_{IJ} the inter-electronic and internuclear distances and Z_I and Z_J the nuclear charge of atom I and J, respectively. By partitioning the total electronic density into the two contributions, $\rho_{QM+MM} = \rho_{QM} + \rho_{MM}$, the total energy can be rewritten as:

$$E[\rho_{QM+MM}] = E[\rho_{QM}] + E[\rho_{MM}] + \int\int\frac{\rho_{QM}(\mathbf{r}_1)\,\rho_{MM}(\mathbf{r}_2)}{r_{12}}d\mathbf{r}_1 d\mathbf{r}_2$$
$$+ \sum_{\substack{I\,\in\,QM \\ J\,\in\,MM}}\frac{Z_I Z_J}{R_{IJ}} + E_{xc}^{NL} + T^{NL}$$

(2.116)

where $E_{xc}^{NL} = E_{xc}[\rho_{QM}+\rho_{MM}] - E_{xc}[\rho_{QM}] - E_{xc}[\rho_{MM}]$ and $T^{NL} = T[\rho_{QM} + \rho_{MM}] - T[\rho_{QM}] - T[\rho_{MM}]$ arise from the nonlinearity of the kinetic and exchange and correlation functionals. In eq. 2.116, the term $E[\rho_{QM}]$ is treated at the quantum level, while each contribution to $E[\rho_{MM}]$ is approximated by using a force field, function of the nuclear coordinates only. In particular, as force fields are parametrized at a fixed value of the electronic density, the kinetic energy functional is an additive constant which can be neglected. In this context, the energy E_{xc} is approximated by a Lennard-Jones pair-additive potential:

$$E_{xc} \approx \sum_{I,J\in MM} 4\epsilon_{IJ}\left(\left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^6\right).$$

(2.117)

The remaining three term of eq. 2.115 describe the nuclear-electronic, electronic-electronic, and nuclear-nuclear charge densities electrostatic energies. In the force-field spirit, the total contribution to the energy is represented by the interaction energy among effective point charges located at the nuclear positions:

$$\int_\Omega V^{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int\int\frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}d\mathbf{r}_1 d\mathbf{r}_2 + \frac{1}{2}\sum_{I,J\in MM}\frac{Z_I Z_J}{R_{IJ}}$$
$$\approx \frac{1}{2}\sum_{I,J\in MM}\frac{q_I q_J}{R_{IJ}}$$

(2.118)

The chemical bonding cannot be described by solely Lennard-Jones and point charges electrostatic interaction energy, thus bonded terms as in eq. 2.47 have to be added to the MM energy. The most interesting part of eq. 2.116 concerns the interaction between the two subsystems. Using for T^{NL} and E_{xc}^{NL} the same approximation as above, we can express the interaction energy as:

$$
\begin{aligned}
E[\rho_{QM/MM}] = &\sum_{I \in MM} \int_\Omega \frac{q_I}{|\mathbf{R_I} - \mathbf{r}|} \rho_{QM}^{el+nuc}(\mathbf{r}) d\mathbf{r} \\
&+ \sum_{\substack{I \in QM \\ J \in MM}} 4\epsilon_{IJ} \left(\left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{6} \right) \\
&+ \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\vartheta(\vartheta - \vartheta_{eq})^2 \\
&+ \sum_{dihedrals} \frac{V_n}{2} \left[1 + \cos\left(n\phi - \gamma\right)\right]
\end{aligned}
\tag{2.119}
$$

where bonds, angles and dihedrals involve at least one QM atom. Notice that in eq. 2.119 the term T^{NL} has been neglected while for E_{xc}^{NL} the force-field approximation has been used. In this formulation the electrostatic potential provided by the effective classical point charges polarizes the QM electronic charge density. It should be noted that the presence of a discontinuous QM/MM interface introduces a series of artifacts. One of the most serious is the so-called *link atom problem*. When a chemical bond involves atoms on the two subsets, the QM system will contain by construction unsaturated valencies and has to be made chemically inert. Two approaches are mainly used to deal with this problem. The first consists of the use of a monovalent pseudopotential situated at the position of the MM involved in the bond crossing the QM/MM interface. In the CPMD code an analytical non-local pseudopotential of the Goedecker type [178] is used. The second approach introduces capping atoms (usually hydrogens) to saturate chemical bonds at the interface. It should be pointed out that the latter strategy introduces additional artifacts and a correction for the interactions between the "ghost" atoms and the classical environment is required. Furthermore the approximation of E_{xc}^{NL} by a purley classical term, i.e. not involving QM electronic degrees of freedom, results in the so-called *electron spill out* problem. Due to the fact that the MM region contains no electrons, those of the QM part are no longer repelled by closed-shell cores of the atoms belonging to the MM region. The effect of the missing Pauli repulsion is to artificially localize electrons on MM positive point charges. In order to avoid this artifact a pseudopotential-like approach can be applied by replacing the classical point charges Coulomb potential with a suitable function $v_I(|\mathbf{r} - \mathbf{R_I}|)$, which ensures the correct $1/r$ behaviour for large r and

goes to a finite value for $r \rightarrow 0$. The first term of eq. 2.119 is thus replaced by the following quantity:

$$E_{QM/MM}^{elec} = \sum_{I \in MM} q_I \int_\Omega \rho_{QM}^{el+nuc}(\mathbf{r}) v_I(|\mathbf{r} - \mathbf{R_I}|) d\mathbf{r} \,. \tag{2.120}$$

In particular an appropriate form for $v_I(|\mathbf{r} - \mathbf{R_I}|)$ is:

$$v_I(|\mathbf{r} - \mathbf{R_I}|) = \frac{r_{cI}^4 - r^4}{r_{cI}^5 - r^5} \,, \tag{2.121}$$

where r_{cI} is the covalent radius of atom I. This functional form resembles that obtained by smearing the MM point charges into Gaussian charge distributions of finite width. In the context of plane-waves, the QM/MM scheme devised above cannot be used for practical purposes without an additional approximation. Indeed, the quantum charge distribution is distributed on a grid of $N_r \sim 100^3$ points, so that an exact evaluation of $E_{QM/MM}^{elec}$ would involve $N_r \times N_{MM}$ operations, with $N_{MM} \geq 10^5$. Therefore, this interaction term is split into a short and a long-range part, in a way reminishent of the Ewald method [136]. The direct evaluation of the integral in eq. 2.120 is done only for a subset (NN) of MM atoms. The latter is defined in such a way as to include all non neutral atoms belonging to charge groups with at least one atom inside a shell of thickness R_c around any QM atom. The rest of MM atoms belong to the second shell. For those, the electrostatic interaction with the QM system is calculated using for the charge density of the QM system a multipolar expansion around the geometrical center of the quantum system $\bar{\mathbf{r}}^\alpha = 1/N_{QM} \sum_I \mathbf{r}_I$ up to the quadrupole order. In particular, the electrostatic interaction Hamiltonian can be expressed as:

$$H_{elec} = \sum_{j \in NN} q_j \int d\mathbf{r} \rho(\mathbf{r}) v_j(|\mathbf{r} - \mathbf{r}_j|) + H_{lr} \tag{2.122}$$

where H_{lr} is defined as:

$$H_{lr} = C \sum_{j \notin NN} \frac{q_j}{\tau_j} + \sum_\alpha D^\alpha \sum_{j \notin NN} \frac{q_j}{\tau_j^3} \tau_j^\alpha + \frac{1}{2} \sum_{\alpha\beta} Q^{\alpha\beta} \sum_{j \notin NN} \frac{q_j}{\tau_j^5} \tau_j^\alpha \tau_j^\beta \tag{2.123}$$

with $\tau_j^\alpha = r_j^\alpha - \bar{r}^\alpha$; C, D^α and $Q^{\alpha\beta}$ are the total charge, the dipole and the quadrupole of the electronic charge distribution, respectively. The potential entering into the Khon-Sham Hamiltonian is given by the functional derivative of H_{elec} with

respect to the density ρ_{el}:

$$
V(r) = \frac{\delta H_{elec}}{\delta \rho_{el}} = \sum_{j \notin NN} \frac{q_j}{\tau_j} + \sum_{\alpha} (r^\alpha - \bar{r}^\alpha) \sum_{j \notin NN} \frac{q_j}{\tau_j^3} \tau_j^\alpha
$$
$$
+ \frac{1}{2} \sum_{\alpha\beta} \left[3(r^\alpha - \bar{r}^\alpha)(r^\beta - \bar{r}^\beta) - \delta^{\alpha\beta} |r - \bar{r}|^2 \right] \sum_{j \notin NN} \frac{q_j}{\tau_j^5} \tau_j^\alpha \tau_j^\beta
$$

$$(2.124)$$

The forces on the atom arising from H_{elec} are obtained by taking the derivatives with respect to the atomic positions. These are for the QM, NN and classical atoms not belonging to the NN set:

$$
F_j^\gamma = \frac{1}{N_{QM}} \left[-\frac{5}{2} \sum_{\alpha\beta} Q^{\alpha\beta} \sum_{k \notin NN} \frac{q_k}{\tau_k^7} \tau_k^\alpha \tau_k^\beta \tau_k^\gamma + \sum_{\alpha} Q^{\alpha\gamma} \sum_{k \notin NN} \frac{q_k}{\tau_k^5} \tau_k^\alpha \right],
$$
$$
\text{for } j \in QM
$$

$$(2.125)$$

$$
F_j^\gamma = q_j \int d\mathbf{r} \rho(\mathbf{r}) g_j(|\mathbf{r} - \mathbf{r}_j|) \frac{r^\gamma - r_j^\gamma}{|\mathbf{r} - \mathbf{r}_j|},
$$
$$
\text{for } j \in NN
$$

$$(2.126)$$

$$
F_j^\gamma = -q_j \left[\left(-\frac{C}{\tau_j^3} - \frac{3}{\tau_j^5} \sum_{\alpha} D^\alpha \tau_j^\alpha - \frac{5}{2\tau_j^7} \sum_{\alpha\beta} Q^{\alpha\beta} \tau_j^\alpha \tau_j^\beta \right) \tau_j^\gamma + \frac{D^\gamma}{\tau_j^3} + \frac{1}{\tau_j^5} \sum_{\alpha} Q^{\alpha\gamma} \tau_j^\alpha \right]
$$
$$
\text{for } j \notin NN, QM
$$

$$(2.127)$$

where $g_j(r) = dv_j/dr$. This two level coupling scheme can also be refined introducing an intermediate third layer in which the charge density of the QM system is replaced by variational D-RESP charges [179]. In the work exposed in the thesis I have used the implementation of QM/MM realized in the code CPMD [177], which has been interfaced to the AMBER [24, 25, 131] force field. The code is available for free with the CPMD package (www.cpmd.org).

2.8 Free Energy calculations

The calculation of free energies and the molecular-level description of rare events are probably the fields where biomolecular simulation has most demonstrated its power, often giving insights which are not directly accessible to experiments [72, 180, 181]. A number of elegant and powerful methods [182, 183, 184, 185, 186, 187, 188, 189, 180] have been developed allowing to characterize the thermodynamics of processes like chemical reactions [104, 190, 191, 192], ligand binding [193, 194, 195, 196, 197, 198], "alchemical" mutations [70, 76, 199, 200] and conformational changes between

relevant states of biomolecules (e.g. B and Z conformations of DNA [201, 202]). It is well-known that thermodynamics is intimately related to Statistical Mechanics: the achievement of a reliable estimate of the free energy change accompanying a process depends on how well the phase-space is sampled in a computer simulation. Thus, enhancing the sampling of thermodynamically unlikely states is the primary scope of methods for the calculation of free energies. This represent a particular challenge in biology, due to the ruggedness of free energy profiles of complex macromolecules. In the following I review the methodologies used here to calculate the free energies of drug sliding within the minor groove and unbinding from oligonucleotides. These schemes rely on the definition of few coordinates $\{s_j\}$, called *collective variables* (CVs) or *reaction coordinates*, which are believed to be relevant for the process under study. This means that they are slow, i.e. change very little during the course of a standard MD or MC run. Thus, enhancing their sampling will allow to simulate the rare event and to accurately calculate the associated free energy ΔF as a function of the CVs, which takes the name of Potential of Mean Force (PMF) [203]. Among the various ways to enhance the exploration of the phase space, one consist in adding to the original potential energy function $U(\mathbf{x})$ a biasing term in the CVs space, which forces the system to visit interesting configurations. The following scheme depicts the passage from the absolute to this reduced representation:

$$\mathbf{x} \equiv \{x_i\} \,, \, i = 1, ..., 3N \quad \longleftrightarrow \quad F = -\tfrac{1}{\beta} \ln \mathcal{Z}$$
$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$
$$\mathbf{s} \equiv \{s_j\} \,, \, j = 1, ..., N_{CV} \quad \longleftrightarrow \quad F(\mathbf{s}) = -\tfrac{1}{\beta} \ln P(\mathbf{s})$$

where $\beta = k_B T$, N is the number of particles, N_{CV} is the number of collective variables, $\mathcal{Z}$ is the canonical configuration partition function, and P is the probability to find the system in a state corresponding to a given value s of the reaction coordinates. For ergodic systems, this is equal to the distribution function along $\mathbf{s}$:

$$P(\mathbf{s}) = \lim_{t \to \infty} \frac{1}{t} \int_0^t dt \, \delta \left(\mathbf{s} - \mathbf{s}(t)\right) \equiv \rho(\mathbf{s}) = \frac{\int d\mathbf{x}\delta \left(\mathbf{s} - \mathbf{s}(\mathbf{x})\right) e^{-\beta U(\mathbf{x})}}{\mathcal{Z}} \qquad (2.128)$$

where the dependence of s on the coordinates x has been explicited. At this point it is important to say that also in the case s is a "good" reaction coordinate, estimating ΔF as a PMF is an intrinsically ill-defined approximation. Indeed, considering a reversible change of variable $s \to q$, the PMF as a function of these two CVs will not be the same. In fact, due to the invariance property of the probability distribution,

whose integral is indipendent from the chosen coordinate[9]:

$$P(s)ds = P(q)dq \Rightarrow F(q) = F(s) + \frac{1}{\beta} \ln \left(\frac{dq}{ds} \right) \tag{2.129}$$

This relation confirms that the choice of the CVs is fundamental in all the processes being studied within a PMF approach.

2.8.1 Umbrella Sampling

The Umbrella Sampling method [183] realizes a non-Boltzmann sampling in the phase space by introducing a modified distribution function. In general, the free energy difference between two systems, the "interesting" and "reference" one, with internal energies $U(\mathbf{x})$, $U_0(\mathbf{x})$ and inverse temperatures β and β_0, can be written as [183, 135]:

$$\Delta F^* = \beta F - \beta_0 F_0 = - \ln \frac{\int d\mathbf{x} \left[e^{-\beta U(\mathbf{x}) + \beta_0 U_0(\mathbf{x})} \right] e^{-\beta_0 U_0(\mathbf{x})}}{\mathcal{Z}_0} = - \ln \left\langle e^{-\Delta U^*(\mathbf{x})} \right\rangle_0 \tag{2.130}$$

where the subscript 0 means that the average is on the reference system, and $F^* = \beta F$ and $U^*(\mathbf{x}) = \beta U$ are the *reduced* free and internal energies, respectively. Now, ΔU^* will be large (and $e^{-\Delta U^*}$ small) for the configurations corresponding to the interesting system (i.e. those for which the reduced internal energy is close to U^*), unless this is very similar to the reference one. Indeed, the average ensemble is performed on this latter, which very rarely will sample configurations of the interesting system. Thus, we are lacking almost one half of the sampling needed to obtain a reliable estimate of F^*. This can be better understood if one rewrites 2.130 as:

$$\Delta F^* = - \ln \int_{-\infty}^{\infty} d\,\Delta U^* \, \rho_0 \left(\Delta U^* \right) e^{-\Delta U^*} \tag{2.131}$$

with $\rho_0 \left(\Delta U^* \right)$ being the probability density of ΔU^* in the *reference* system. By definition, ρ_0 will be peaked at a value of ΔU^* given by the difference between the average energies of the two systems, $\langle U^* \rangle - \langle U_0^* \rangle_0$, and will decay to zero for small ΔU^*. However, the exponential in Eq. 2.131 will give the largest contributions just for small ΔU^*. Therefore, the most important contribution to the integral will be poorly evaluated in every finite-length simulation, rendering the estimation of ΔF^*

---

[9] This equation can be derived from the expression of $F(\mathbf{s})$ using the relation

$$\delta \left(q(s') - q(s(x)) \right) = \frac{\delta \left(s' - s(x) \right)}{|dq/ds|}$$

inaccurate. Torrie and Valleau [183] introduced a bias to sample configurations in a wider ΔU^* range, through a new non-Boltzmann probability function:

$$\rho_w(\mathbf{x}) = \frac{w(\mathbf{x})e^{-\beta_0 U_0(\mathbf{x})}}{\int d\mathbf{x}\, w(\mathbf{x})e^{-\beta_0 U_0(\mathbf{x})}} \tag{2.132}$$

$w(\mathbf{x}) = w(\Delta U^*)$ is a weighting function chosen to sample configurations corresponding to significative values ΔU^* in the integral 2.131. Notice that *all* of these significative values should be sampled, corresponding to both the reference and interesting systems. In other words, the requirement for the weighting function is to *flatten* the profile of the distribution function in order to cover the range between them like an umbrella (hence the name). The unbiased average of any observable O can be easily recovered from the biased ensemble:

$$\langle O \rangle = \frac{\int d\mathbf{x}\, [O(\mathbf{x})/w(\mathbf{x})]\, w(\mathbf{x})e^{-\beta_0 U_0(\mathbf{x})}}{\int d\mathbf{x}\, [1/w(\mathbf{x})]\, w(\mathbf{x})e^{-\beta_0 U_0(\mathbf{x})}} = \frac{\langle O/w \rangle_w}{\langle 1/w \rangle_w} \tag{2.133}$$

where $\langle\ \rangle_w$ indicates averaging over the distribution ρ_w. From the biased sampling one can also recover the unbiased probability density

$$\rho_0(\Delta U^*) = \frac{\rho_w(\Delta U^*)/w(\Delta U^*)}{\langle 1/w(\Delta U^*) \rangle_w} \tag{2.134}$$

The improved ρ_0, evaluated through eq. 2.133 in a wider range than from a Boltzmann sampling, can be used to calculate with a greater accuracy the reduced free energy difference ΔF^*. As rule of thumb ρ_w should extend the range of energies sampled by at least three times.

In terms of PMF, the enhanced sampling of the configuration space can be obtained by adding to $U(\mathbf{x})$ a biasing potential $U_w(\mathbf{s}(\mathbf{x}))$ that depends on $\mathbf{x}$ only through the CVs.

The unbiased probability distribution $\rho(\mathbf{s})$ can be derived from $\rho_w(\mathbf{s})$ through the following relation:

$$
\begin{aligned}
\rho_w(\mathbf{s}) &= \frac{\int d\mathbf{x}\, \delta\left(\mathbf{s} - \mathbf{s}(\mathbf{x})\right) e^{-\beta U_w(\mathbf{s}(\mathbf{x}))} e^{-\beta U(\mathbf{x})}}{\mathcal{Z}_w} \\
&= e^{-\beta U_w(\mathbf{s})} \frac{\int d\mathbf{x}\, \delta\left(\mathbf{s} - \mathbf{s}(\mathbf{x})\right) e^{-\beta U(\mathbf{x})}}{\mathcal{Z}} \frac{\mathcal{Z}}{\mathcal{Z}_w} \\
&= \rho(\mathbf{s}) e^{-\beta U_w(\mathbf{s})} e^{f_w}
\end{aligned}
\tag{2.135}
$$

where

$$f_w = \frac{1}{\beta} \ln \frac{\mathcal{Z}}{\mathcal{Z}_w}$$

Thus, the unbiased PMF $F(\mathbf{s})$ is given by:

$$F(\mathbf{s}) = -\frac{1}{\beta} \ln\left[\rho_w(\mathbf{s})\right] - U_w(\mathbf{s}(\mathbf{x})) - f_w \tag{2.136}$$

Note that f_w is a constant independent of $\mathbf{s}$, which can be safely neglected when calculating free energy differences.

Error evaluation

We derive an expression for the standard deviation of $F(s)$ and $\rho_w(s)$ in a finite-length simulation (considering a one-dimensional profile for simplicity). In an umbrella sampling simulation $\rho_w(s)$ is estimated after discretization of the CVs space as:

$$\rho_w(s) \sim \frac{1}{m\Delta s} \sum_{k=1}^{m} \chi_s(s_k) \tag{2.137}$$

where $s_1, ..., s_m$ are the values assumed by the collective variable, and $\chi_s = 1$ if $s_k \in [s, s + \Delta s]$ and zero otherwise. Assuming that the s_k are uncorrelated, if $p = \rho_w(s)\Delta s$ is the probability that $s_k \in [s, s + \Delta s]$, the probability to observe i out of m entries within the interval $[s, s + \Delta s]$ is given by the Poisson distribution ($m \to \infty$):

$$\mathcal{P}_{p,m}(i) = \frac{\nu^i e^{-\nu}}{i!} \tag{2.138}$$

with $\nu = pm$. For this distribution, we have $\left\langle \frac{i}{m} \right\rangle = p$ and $\left\langle \frac{i^2}{m^2} \right\rangle = p^2 + \frac{p}{m}$ [10], hence the variance on $\rho_w(s)$ is:

$$\sigma^2(\rho_w(s)) = \frac{\rho_w(s)}{m\Delta s} = \frac{1}{m^2\Delta s} \sum_{k=1}^{m} \chi_s(s_k) \tag{2.139}$$

A constant variance is obtained if $\rho_w(s)$ itself is a constant, which implies (eq. 2.136) $U_w(s) = -F(s)$ for each value of the CVs (up to a constant). The variance on the free energy will be:

$$\sigma^2(F(s)) = \frac{1}{\beta^2} \frac{\sigma^2(\rho_w(s))}{(\rho_w(s))^2} = \frac{1}{\beta^2} \frac{1}{(m\Delta s)^2} \frac{1}{\sigma^2(\rho_w(s))} \tag{2.140}$$

---

[10] For example

$$\left\langle \frac{i}{m} \right\rangle = \sum_{i=1}^{m} \frac{i}{m} \frac{\nu^i e^{-\nu}}{i!} = \frac{e^{-\nu}}{m} \sum_{i=1}^{m} \nu \frac{\nu^{i-1}}{(i-1)!} \xrightarrow{m\to\infty} \frac{e^{-\nu}}{m} \nu\left[e^\nu\right] = p$$

and a similar derivation can be made for $\left\langle \left(\frac{i}{m}\right)^2 \right\rangle$.

which is also independent on s if one requires $\sigma^2(\rho_w(s))$ to be constant.

2.8.2 Weighted Histogram Analysis Method

A single biased simulation is usually not enough to obtain a reliable $F(\mathbf{s})$ over the required range of $\mathbf{s}$. Indeed, we have seen that to have a flat $\sigma^2(F(s))$ we would know the free energy itself, and this can be only be done approximatively or iteratively within a limited interval. To cope with this problem, a number N_w of simulations can be performed each with different bias potentials, covering adjacent windows in the CVs space. The results from each window then need to be unbiased and glued together into a single PMF. Among the various algorithms proposed, the Weighted Hystogram Analysis Method (WHAM) [184, 185] has proven to be very efficient and almost free of information-loss. The main idea, which goes back to the hystogram method developed by Ferrenberg and Swensden [204, 205], consist in constructing $\rho(\mathbf{s})$ as *weighted* sum of the unbiased distribution functions extracted from each window

$$\rho(\mathbf{s}) = A \sum_{i=1}^{N_w} \pi_i(\mathbf{s})\rho_i(\mathbf{s})\,. \tag{2.141}$$

The weigths are functions of $\mathbf{s}$, and are chosen as to minimize $\sigma^2(\rho(\mathbf{s}))$, subject to normalization $\sum_i \pi_i(\mathbf{s}) = 1$. Thus they are determined using the Lagrange λ multiplier method:

$$\frac{\delta}{\delta\pi_j(\mathbf{s})} \left[\sigma^2(\rho(\mathbf{s})) - \lambda \left(\sum_i \pi_i(\mathbf{s}) - 1 \right) \right] =$$
$$\frac{\delta}{\delta\pi_j(\mathbf{s})} \left[A^2 \sum_i \pi_i^2(\mathbf{s})\sigma^2(\rho_i(\mathbf{s})) - \lambda \left(\sum_i \pi_i(\mathbf{s}) - 1 \right) \right] \tag{2.142}$$
$$= 2A^2\pi_j(\mathbf{s})\sigma^2(\rho_j(\mathbf{s})) - \lambda = 0$$

which, after writing $\lambda/2A^2 = 1/\sum_i \left[\sigma^2(\rho_i(\mathbf{s}))\right]^{-1}$ (from normalization), gives:

$$\pi_j(\mathbf{s}) = \frac{\left[\sigma^2(\rho_j(\mathbf{s}))\right]^{-1}}{\sum_i \left[\sigma^2(\rho_i(\mathbf{s}))\right]^{-1}} \tag{2.143}$$

Thus, the most accurate ρ_j will have the largest weigth when composing the total distribution function. In a real simulation is clearly useful to express the weigths in terms of the known biasing potentials $U_w i$. From eq. 2.135, one has

$$\sigma^2(\rho_i(\mathbf{s})) = \left[e^{\beta[U_{wi}(\mathbf{s})-f_{wi}]}\right]^2 \sigma^2(\rho_{wi}(\mathbf{s}))\,, \tag{2.144}$$

which by insertion of eq. 2.139 gives:

$$\pi_j(\mathbf{s}) = \frac{m_j e^{-\beta[U_{wj}(\mathbf{s})-f_{wj}]}}{\sum_i m_i e^{-\beta[U_{wi}(\mathbf{s})-f_{wi}]}} \tag{2.145}$$

where m_j is the number of sampled points in the j-th window. Thus, the weigths depends on the parameters f_{wj}, that in turn are function of the π_j:

$$
\begin{aligned}
e^{-\beta f_{wj}} = \frac{\mathcal{Z}_{wj}}{\mathcal{Z}} &= \int d\mathbf{s}\, \rho_j(\mathbf{s}) e^{-\beta U_{wj}(\mathbf{s})} \\
&= A \int d\mathbf{s}\, e^{-\beta U_{wj}(\mathbf{s})} \sum_{k=1}^{N_w} \pi_k(\mathbf{s})\rho_k(\mathbf{s}) \\
&= A \int d\mathbf{s}\, e^{-\beta U_{wj}(\mathbf{s})} \frac{\sum_{k=1}^{N_w} m_k \rho_{wk}(\mathbf{s})}{\sum_i m_i e^{-\beta[U_{wi}(\mathbf{s})-f_{wi}]}}
\end{aligned}
\tag{2.146}
$$

Thus, eqs. 2.145 and 2.146 have to be solved with a self-consistent procedure. The variance of the free energy can be obtained straigthforwardly from eq. 2.140:

$$\sigma^2(F(s)) = \frac{1}{\beta^2} \frac{\sum_i \pi_i^2(\mathbf{s})\sigma^2(\rho_i(\mathbf{s}))}{[\sum_i \pi_i(\mathbf{s})\rho_i(\mathbf{s})]} = \frac{1}{\beta^2} \frac{1}{\sum_i m_i \rho_{wi}(\mathbf{s})} \tag{2.147}$$

Notably, the error on the free energy can be made as constant as the cumulative histogram in the denominator of eq. 2.147 is uniform, and it is independent on the form of the biasing potential (no approximate knowledge of the free energy on the explored range is required). The only need is to have $\rho_{wi}(\mathbf{s})$'s from adjacent windows whose tails are overimposed in order to get a uniform histogram: this can be always done by adjusting the sampling on a given window or introducing a new one to fill a hole in the overall probability distribution.

2.8.3 Metadynamics

The main idea behind metadynamics [189] is to drive the evolution in the space of CVs by adding to the thermodynamic force, coming from the free energy $F(\mathbf{s})$, a force due to a history-dependent biasing potential $F_G(\mathbf{s}, t)$. The bias potential is constructed as a sum of Gaussians deposed along the trajectory of the CVs up to time t. The method can be seen as the finite temperature extension of the Wang and Landau algorithm [206], and is related in the spirit to taboo search [207], local elevation [208] and adaptive bias force [187]. The most important property of metadynamics is that the biased trajectory proceeds by filling valleys in the free energy surface [189], so that the system tends to escape from every stable state. After a long enough time t

the sum of the Gaussians deposed along the trajectory will counterbalance the free energy landscape, allowing to estimate the free energy itself:

$$\lim_{t \to \infty} F_G(\mathbf{s}, t) = F(\mathbf{s}) \tag{2.148}$$

It can be shown that the above relation is true under rather general assumptions [209]. In real cases, the time needed for considering eq. 2.148 to be valid can be estimated by visual inspection of the trajectory of the CVs: when the $F_G(\mathbf{s}, t)$ counterbalances the free energy $F(\mathbf{s})$, the CVs have a diffusive behaviour. If one uses simultaneously 2 or more CVs is not necessary to know *a priori* the reaction path in metadynamics. The force due to the history-dependent potential naturally drives the system through the Lowest Free Energy Path (LFEP), i.e. the most likely reaction path [210]; after the crossing of a barrier the system naturally goes towards a new and possibly unpredicted metastable state. For this reason the method has found a large use not only for predicting free energies, but also for accelerating rare events and investigating molecular mechanism of biological processes. Obviously, as for all the methods based on dimensional reduction, the chosen CVs must describe somewhat the process of interest. Nevertheless, many application have shown that choosing general and flexible CVs allows discovering unknown stable states and reaction mechanisms [211, 212]. The work exposed in this thesis has been done choosing as CVs the distance between the centers of mass of molecular groups and coordination numbers. Here I first describe the "discrete" version of the algorithm, which has been introduced by Laio and Parrinello in 2002 [189], along with some of the modifications introduced later to enhance its efficiency and accuracy [213, 209]. The so-called "continuous" version, which has been used in the work exposed in this thesis, can be simply derived from the discrete one by discarding the evaluation of the thermodynamic force (speeding-up the algorithm).

In discrete metadynamics the CVs are evolved step by step. A multidimensional Gaussian of width $\delta \mathbf{s} = (\delta \mathbf{s_1}, ..., \delta \mathbf{s_{N_{CV}}})$ and height w is deposed at position $\mathbf{s}(\mathbf{x}, t)$ every time *metastep* τ_G. Thus, at time t the free energy underlying the dynamics of the CVs is given by:

$$\tilde{F}(\mathbf{s}(\mathbf{x}), t) = F(\mathbf{s}(\mathbf{x})) + F_G(\mathbf{s}(\mathbf{x}), t) = F(\mathbf{s}(\mathbf{x})) + w \sum_{j=1}^{n_G^t} \prod_{i=1}^{N_{CV}} e^{-\left[\frac{s_i(\mathbf{x}) - s_i(\mathbf{x}(j\tau_G))}{\sqrt{2}\delta s_i}\right]^2} \tag{2.149}$$

where the number n_G^t of Gaussians deposed at time t is given by the integer closest to t/τ_G. In order to simplify matter the CVs space can be rendered approximately spherical by rescaling all the CVs s_i to their thermal fluctuations $\sqrt{(s_i - \langle s_i \rangle)^2}$, which can be evaluated at the starting minimum through an unbiased dynamics. In this way spherical Gaussians can be used, for which the width is the same in all the

directions, $\delta s_i = \delta s, \forall i$. From eq. 2.149 it can be seen that the dynamics of the CVs is driven by two forces:

- the thermodynamic force, evaluated at $\mathbf{s}^t = \mathbf{s}(\mathbf{x}(t))$:

$$f_i^{th}\big|_t = -\frac{\partial}{\partial s_i^t} F(\mathbf{s}) . \tag{2.150}$$

Following Sprik and Ciccotti [214], these forces are estimated through the Constrained Reaction Coordinate Dynamics (CRCD) algorithm, adding to the normal Lagrangian of the system a restraining term $\sum_{i=1}^{N_{CV}} \lambda_i(s_i - s_i(\mathbf{x}(t))$. By averaging over the time, in the absence of inertial terms, the components of the thermodynamic force are given by $f_i^{th} = \langle \lambda_i \rangle$. Thermodynamic forces are evaluated in the time between two subsequent hill depositions.

- the history-dependent force at time t, whose components are:

$$f_i^G\big|_t = -\frac{\partial}{\partial s_i^t} w \sum_{j=1}^{n_G^t} \prod_{i=1}^{N_{CV}} e^{-\left[\frac{s_i(\mathbf{x})-s_i(\mathbf{x}(j\tau_G))}{\sqrt{2}\delta s_i} \right]^2} \tag{2.151}$$

which discourage the system to visit the same region in the CVs phase space.

The *metadynamics* of the walker in the CVs space thus is regulated by the following discrete equation of motion (I removed the superscript t for simplicity):

$$\mathbf{s}(t + \tau_G) = \mathbf{s}(t) + \delta \mathbf{s} \cdot \frac{\tilde{\mathbf{f}}}{\left| \tilde{\mathbf{f}} \right|} \tag{2.152}$$

$$\tilde{\mathbf{f}} = \mathbf{f}^{th} + \mathbf{f}^G \tag{2.153}$$

Equation 2.152 was introduced firstly in Ref. [189]. Subsequently, three corrections were applied in order to enhance accuracy and reduce sistematic errors:

1. To improve efficiency, the Gaussians are shifted with respect to the position of the walker. In fact, if the thermodynamic force at time t is evaluated at the same point where the Gaussian is placed, the total force felt by the walker will be the same as in the previous metastep. In order to compensate the thermodynamic force is better to depose the Gaussian at a distance $\delta \mathbf{s}$ from $\mathbf{s}(t)$ in the direction of the thermodynamic force.

2. To reduce the correlation induced by depositions with constant step, at every iteration the metastep is chosen randomly from a uniform distribution with two limiting values (e.g. $\delta \mathbf{s}$ and $1.5\,\delta \mathbf{s}$).

3. When the metadynamics is terminated F_G will present a bump in the region around the last hill; the spread of this bump depends on the correlation time of the metadynamics. In order to reduce these spatial correlations in the free energy the contributions of the Gaussians placed at the end of the dynamics are weighted less. In Ref. [213] each term of the sum in eq. 2.149 has been multiplied by $\tanh\left(\frac{(n_G^t - j)\tau_G}{\tau_c}\right)$, where τ_c is larger than the typical time required to sweep the "filled" CVs space, j is the metastep number and n_G^t is the total number of metasteps at time t.

With the modification listed above, eqs. 2.152, 2.153 and 2.149 become:

$$\mathbf{s}(t + \tau_G) = \mathbf{s}(t) + \Delta\mathbf{s} \cdot \frac{\tilde{\mathbf{f}}}{\left|\tilde{\mathbf{f}}\right|}, \ \Delta\mathbf{s} \in [\delta\mathbf{s}, \ \alpha\delta\mathbf{s}], \ \alpha > 1 \tag{2.154}$$

$$\tilde{f}_i = f_i^{th} - \frac{\partial}{\partial s_i^t} w \sum_{j=1}^{n_G^t} \prod_{i=1}^{N_{CV}} e^{-\left|\frac{s_i(\mathbf{x}) - s_i(\mathbf{x}(j\tau_G)) - \delta s_i \frac{\tilde{f}_i}{|\tilde{\mathbf{f}}|}}{\sqrt{2}\delta s_i}\right|^2} \tag{2.155}$$

$$\tilde{F}(\mathbf{s}(\mathbf{x}), t) = F(\mathbf{s}(\mathbf{x})) + w \sum_{j=1}^{n_G^t} \prod_{i=1}^{N_{CV}} \tanh\left(\frac{(n_G^t - j)\tau_G}{\tau_c}\right) e^{-\left|\frac{s_i(\mathbf{x}) - s_i(\mathbf{x}(j\tau_G)) - \delta s_i \frac{\tilde{f}_i}{|\tilde{\mathbf{f}}|}}{\sqrt{2}\delta s_i}\right|^2}$$

$$\tag{2.156}$$

In the continuous version of metadynamics the system is not constrained between two successive metasteps in order to evaluate thermodynamic forces; these latter indeed are not calculated at all. This is equivalent [209] to evolve the CVs continuosly (at every MD step, altough this is not done in practice for reasons of efficiency). The system evolve under the action of the history-dependent forces only, which act directly on the cartesian coordinates:

$$f^G(t) = -\frac{\partial}{\partial x} \frac{w}{\tau_G} \int_0^t dt' \prod_{i=1}^{N_{CV}} e^{-\left[\frac{s_i(\mathbf{x}) - s_i(\mathbf{x}(t'))}{\sqrt{2}\delta s_i}\right]^2} \tag{2.157}$$

where the parameter $\frac{w}{\tau_G}$ controls the Gaussians deposition rate and $\mathbf{x}(t)$ is the trajectory of the system. It has been shown [209] that the reconstructed Free Energy does not vary for constant ratios $\frac{w}{\tau_G}$ along a quite large interval for τ_G (in classical MD the invariance has been demonstrated up to $\tau_G = 5 \ ps$). The reconstructed free energy is given by:

$$F_G(\mathbf{s}, t) = \frac{w}{\tau_G} \int_0^t dt' \prod_{i=1}^{N_{CV}} e^{-\left[\frac{s_i - s_i(\mathbf{x}(t'))}{\sqrt{2}\delta s_i}\right]^2} \tag{2.158}$$

and it is an approximation of the $F(\mathbf{s})$ in the CVs region explored up to time t.

Efficiency and Accuracy. It can be demonstrated [209] that the error on the calculated $F(\mathbf{s})$ is proportional to the height barrier w; furthermore, with the improvements achieved with eqs. 2.154,2.155,2.156, a single metadynamics run has shown to already give a very good estimate of the profile, with an error approximately constant in the region where the number of accumulated Gaussians is significant (conventionally this means that $F(\mathbf{s})/w \geq 5$. Obviously the width of the Gaussian also influences the efficiency and the accuracy of the method. In particular, using hills of width larger than typical thermal fluctuations can lead to "bury" some thermodynamic state corresponding to narrow minima. Finally, the accuracy in the evaluation of the thermodynamic force enters in the overall error.

Part II

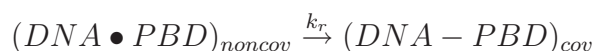# Drug/DNA interactions and molecular recognition

Chapter 3

Characterization of Anthramycin/DNA adducts

3.1 Introduction

Pyrrol[1,4]benzodiazepines (**PBD**'s) cytotoxins [215] exert a powerful antitumoral activity [98, 216] by covalently binding to the minor groove of B-DNA. All **PBD**'s share a condensed three ring moiety featuring a right-handed twist between the phenol and pyrrol rings (rings A and C in Chart I), which allows them to snugly fit into the backbone of B-DNA. In spite of their low molecular weight, **PBD**'s are quite sequence-selective, recognizing codes $3 \div 7$ base pairs (bps) long [217, 218]. Drug binding is believed to proceed by formation of a non-covalent complex (Step I) [219]:

$$DNA + PBD \overset{K_B}{\rightleftharpoons} (DNA \bullet PBD)_{noncov}$$
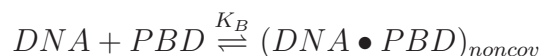
followed by a covalent linkage (Step II):

$$(DNA \bullet PBD)_{noncov} \overset{k_r}{\rightarrow} (DNA - PBD)_{cov}$$

The molecular recognition process is supposed to be driven by (non-specific) non-bonded interactions and sequence-specific structural features, rather than by specific H-bond patterns [18]. However, the molecular details of this process are not yet fully understood. A **PBD** derivative, anthramycin [99] (Chart I), is one of the very few organic minor groove covalent binders for which the X-ray structure of the covalent complex with an oligonucleotide is available. This drug covalently binds to the exocyclic amino group of guanine through the $C11$ carbon [99, 220], showing a modest sequence selectivity for guanines embedded in Pu-G-Pu sequences (namely AGA-5'>AGG-5'≈GGA-5'>GGG-5') [217, 219, 218, 221]. The investigation of DNA/PBD complexes may provide useful insights on the factors affecting both reactivity and binding of minor groove covalent binders. In fact, although anthramycin
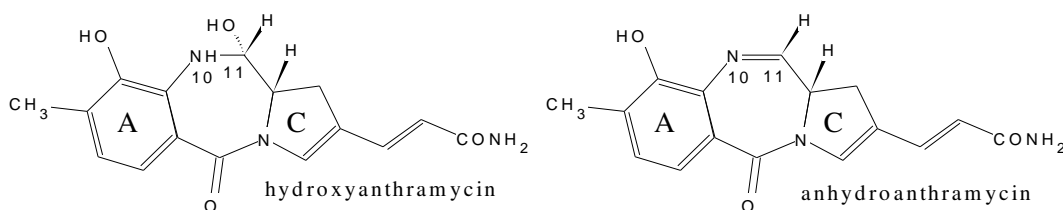
Chart I

is cardiotoxic [98, 222], a number of its derivatives have shown improved antitumoral activity. In particular, the dimer SJG-136 [223, 224] is currently in Phase I clinical trials in both UK and USA. In addition, **PBD**'s have also being used as components of a gene targeting strategy [225, 15] aiming at designing molecules able to target a specific cancer-related gene. Prompted by the biological and pharmacological relevance of **PBD**'s, we performed molecular dynamics (MD) simulations on both the hydroxy and anhydro putative reactive forms of anthramycin (Chart I) in complex with DNA. Previous computational studies on **PBD**s have been performed without inclusion of explicit waters and counter-ions [222, 226, 227, 228], which are very important to accurately simulate nucleic acids and their complexes [36, 38, 35], as well as their reactivity [229]. The main aim of our investigation is to obtain insights on Step I of the reaction. The drugs are initially positioned in front of the reactive guanine G^* of d[GCCAACGTTG*GC]d[GCCAACGTTGGC], based on the X-ray structure of the covalent complex [99]. Our simulations show that both the drugs, docked in front of the reactive site, induce structural deformations on the DNA frame at its central region, while for the covalent adduct the most significant distortions are seen at the binding region. Interestingly, anhydro-anthramycin slides by one base pair step towards the centre of the oligonucleotide after $\sim 10\,ns$, where it is stable for the rest of the simulation; this process causes both an energetic and structural relaxation of the complex. Instead, the hydroxy form oscillates around the reactive site for the whole dynamics. Our MD simulations are complemented by a quantum chemistry investigation of environmental effects on the reactivity of the drug and $G10$ in the two adducts. By using a hybrid QM/MM method [176], we show that the DNA frame influence significantly the electronic distribution on the reactants; in particular, it appears to polarize the anhydro form of the drug in a way favourable to the alkylation reaction.

3.2 Systems and Methods

3.2.1 Classical MD simulations

The following system have been investigated in aqueous solution:

Covalent complex: We built our model (**ANT-DNA** hereafter) starting from the X-ray structure of two anthramycin molecules covalently linked to the oligonucleotide d[CCAACGTTG*G]$_2$ [99]. To asses the structural impact of binding, we remove one of the two drugs present in the experimental structure. To limit artifacts associated to terminal effects, we added one CG base pair (bp) at each end.

Non-covalent complexes: Hydroxy-anthramycin•d[GCCAACGTTG*GC]d[GCC-AACGTTGGC] (**ANT•DNA**) and anhydro-anthramycin•d[GCCAACGTTG*GC]-d[GCCAACGTTGGC] (**IMI•DNA**) complexes were built: *i)* cutting the bond between the $C11$ of the drug and the $N2$ of deoxyguanine, and manually pulling out the drug perpendicularly to the minor groove, until distance $N2$-$C11$ was ~ 3.5; *ii)* adding the OH group on $C11$ in the case of **ANT•DNA**, or removing an H atom from $N10$ in the case of **IMI•DNA** (see Chart I); *iii)* optimizing *in vacuo* the geometries of the drugs at the B3LYP-6-31G(d) [156, 161, 157] level of calculations with Gaussian [230].

Reference systems: i) **DNA**, which is the 12-mer d[GCCAACGTTGGC]d[GCCA-ACGTTGGC] in the B conformation, and was constructed with the *nucgen* module of the AMBER package [231], *ii)* hydroxy-anthramycin (**ANT**) and *iii)* anhydro-anthramycin (**IMI**). Because we observed a sliding of **IMI** along the oligonucleotide (see Results), we constructed an additional complex between **IMI** and the 14-mer d[CAACGTTG*GCCAAC]d[GTTGGCCAACGTTG] (**IMI•DNAc**). This latter has been built taking **IMI•DNA** as template (so that the mode of binding of the drug is the same as in **IMI•DNA**), removing the first two d[GC]d[GC] bps and adding the d[CAAC]d[GTTG] duplex at the end near to the drug. The simulation of **IMI•DNAc** allowed us to check whether such a sliding occurs also if the ligand is located in the central part of the oligonucleotide.

The DNA-containing systems were neutralized by adding sodium counter-ions. All the systems were immersed in water boxes allowing for a solvent shell extending for at least 12 Å around each solute atom (see Table 3.1 for detailed information).

The oligonucleotide (except for the nucleoside G* in **ANT-DNA**) was parametrized using the refined Cornell *et al.* force field [24, 25, 131], while TIP3P [232] and Aaqvist [233] models were used to represent water and sodium ions respectively. The

	ANT-DNA	ANT•DNA	IMI•DNA	IMI•DNA(c)	DNA	ANT	IMI
box (Å^3)	$58 \times 72 \times 58$	$60 \times 62 \times 66$	$64 \times 59 \times 66$	$66 \times 67 \times 78$	$56 \times 55 \times 56$	$45 \times 36 \times 35$	$47 \times 47 \times 46$
waters	6000	7700	8000	8900	5500	1900	2900
$Na+$	22	22	22	26	22	-	-
time (*ns*)	21	19	19	20	21	6	10*s*

Table 3.1: Box dimensions, number of waters, number of counterions, and simulated time for the various system.

structural parameters of the drugs were taken from the gaff [234] database, or constructed with the *parmcal* module of the AMBER package [231], while RESP [134] charges were obtained using the *resp* utility of AMBER after minimization of the electronic structure with Gaussian [230]. The nucleoside G* in **ANT-DNA** was parametrized using the same scheme. Periodic boundary conditions were used, and the electrostatic interactions were calculated with the particle-mesh Ewald method (PME) [137], using a $12\,\text{Å}$ cutoff for the real part, as for the van der Waals interactions. NPT simulations at $300K$ and $1\,atm$ were performed using the Nosé-Hoover [121, 122] thermostat and the Andersen-Parrinello-Rahman [123, 126] pressure-coupling scheme. A time step of $1.5\,fs$ was set for all the simulations. H-bond lengths were constrained using the lincs algorithm [235], and the translational and rotational motions of the center of mass of the solute were removed every 25 MD steps.

Before starting the dynamics, all the systems underwent geometry optimization in two steps: the first with an harmonic restraint of $k = 150\,kcal/mol$ imposed to the solute; the second without any restraint. Then, the systems were heated up linearly to $300\,K$ in $100\,ps$ of NVT MD (namely the temperature was increased by $15\,K$ each $5\,ps$), and, as a last step preceeding the productive dynamics, $200\,ps$ of NPT runs were carried out. Multi-ns trajectories were finally collected (Table 3.1), for a total of $\sim 140\,ns$. Although it has been show that the sampling of counter-ions motion requires very long times [52], according to recent publications [236, 62, 34] the time intervals spanned in our simulations provide good statistic on DNA conformational and helicoidal parameters.

All the simulations were performed using the GROMACS package [237, 238]. The structural parameters of DNA were calculated with the program Curves [239, 239]. The minor groove widths were defined as the distances between sugar $C4'$ atoms, subtracted by two carbon van der Waals radii [99]. A propeller twist and a buckle angle between the pyrrol and the phenol rings of the drug were also defined, according to Ref. [99]. The relative energies of non-bonded interactions were estimated using the terms in the AMBER force field [231, 25, 131]. The molecular Solvent Accessible Surface Area (SASA) [240] of various molecules was calculated with GROMACS [238], using a probe radius of 1.4 Å. The variation in the hydrophobic SASA, which is proportional to the non-polar free energy of solvation, was evaluated as $\Delta\text{SASA} = \text{SASA}_{complex} - \text{SASA}_{DNA} - \text{SASA}_{drug}$. All contributions to the binding are calculated limitedly to the bps $G7 - C18\ldots G11 - C14$, directly involved in the binding and covering the sliding path of anhydro-anthramycin (a test calculation shown that the interaction of the drug with the rest of the oligonucleotide amounts to less than 5% of the total - data not shown).

3.2.2 QM/MM simulations

The QM/MM scheme allows to investigate at a quantum level the reactive site of the system, taking into account the electrostatic effects of the biomolecular frame, and to dissect between various sources of polarization. In this work we used the scheme developed by Rothlisberger and co-workers [176], implemented in the program CPMD [177]. The drug/DNA adducts were partitioned in quantum and classical regions. The first (QM hereafter) comprised the drug and the nucleoside of the reactive guanine $G10$ ($56 \div 59$ atoms), and was treated at the DFT level [241, 242], using the BLYP [243, 157] exchange-correlation functional. The guanine was cut at the $C1'$ carbon, and the valence of this atom was saturated with two capping hydrogens [244]. The remaining part of the system (MM hereafter) was treated using the AMBER [231, 25, 131] effective potential.

The QM part is contained within a rectangular box of $6000 \div 7000\,\text{Å}^3$, and was treated as an isolated system [245]. Quantum calculations were performed using plane waves decomposition of up to a kinetic energy of $70\,Ry$, along with pseudopotentials of the Martins-Trouillers type [169]. The electrostatic interactions between QM and MM atoms were calculated using a hierarchical scheme [176]. In this approach, the short range electrostatic interactions between the QM and MM part were taken explicitly into account within a sphere of radius $5.3\,\text{Å}$ around every QM atom using an appropriately modified Coulomb potential that ensures that no electron spill-out occurs. Beyond this first shell and within $10.6\,\text{Å}$, the electrostatic interactions were calculated using the D-RESP charges for the QM atoms. In the outermost region, a multipole expansion scheme was employed. This hierarchical method is fully Hamiltonian and ensures high accuracy in the core region and efficient computation in the region further away from the QM part.

The starting structures were taken from the last MD snapshots, and then slowly quenched to reach a temperature close to $0\,K$, in order to relax the constraints on QM atoms. Subsequently, the systems were slowly heated up to $300\,K$, and NVT simulations were carried out for $3\,ps$, coupling the systems to a Nosé-Hoover thermostat [121, 122]. A time step of $5\,au$ ($\sim 0.12\,fs$) and a fictitious electronic mass of $600\,au$ were used. To estimate the amount of polarization within the QM region, Boys orbitals [246] (BO's) were calculated during the QM/MM simulations, and, based on their position, we estimated the ionicity of selected bonds. To this aim, following [247], we defined a bond ionicity of the A-B bond as $\text{BI}_{AB} = \frac{|\vec{d_A}|\cos\theta}{|\vec{d_{AB}}|}$, where $\vec{d_A}$ is the position vector of the BO along A-B with respect to A, $\vec{d_{AB}}$ is the position vector of B with respect to A, and θ is the angle between the two position vectors. With this definition, BI's identify electron lone pairs and can be used to visualize the polarization of selected bonds. For example, a value of BI_{AB} much lower than 0.5 indicates a strong polarization of the electronic charge towards A.

3.3   Results and Discussion

3.3.1   MD simulations

Covalent complex

We summarize here our results from a $20\,ns$ MD simulation of the **ANT-DNA** covalent complex in aqueous solution, and compare its features with those of the X-ray structure [99] in order to validate our computational setup. The RMSD between the average structure of the 10-mer d[CCAACGTTG*G]d[CCAACGTTGG] and the X-ray structure is $3.3\,\text{Å}$. The average RMSD of the whole DNA frame with respect to **DNA** is $1.6\,\text{Å}$: consistently, binding by the drug does not induce any appreciable bending of the DNA axis (Figure 3.1), in agreement with experimental findings [99, 220]. The structural determinants (Figures. 3.2, 3.3) and the B-factors of the complex agree well with the experimental ones, except around the region $C24$-$G1$....$T20$-$A5$, from where we removed a second drug bonded to $G22$ in the X-ray structure (See Methods). It turns out that the bonding by the drug affects the DNA frame essentially in the binding region. The H-bond pattern between the drug and the oligonucleotide however differs significantly with respect the experiment. In MD simulation, one to four *direct* H-bonds are formed and broken during the dynamics (average number 1.7), involving mostly the OH_{Ant} group, that interacts with both NH_{G11} and O_{C15}, and the acrylamid tail of the drug (Figure 3.8 and Table 3.2). In addition, one to three water-mediated H-bonds are established between $HN15_{Ant}$ and $O2_{T9}$, $O4'_{G10}$, $N3_{A17}$. These waters push the drug tail towards the strand containing $A17$. As a consequence, $HN15_{Ant}$ links to N_{A17} and looses the contact with $O2_{T9}$ and $O4'_{G10}$. Note that these bridging waters are not present in the X-ray structure. Three of these *direct* H-bonds are also present in the X-ray structure, that exhibits six H-bonds (Table 3.3). This discrepancy might be due to the effect of packing forces in the crystallographic structure, which are particularly important at the ends of the oligonucleotide [18]. According to our results the role of the H-bond pattern connecting the drug and the DNA

	IMI•DNA 1	IMI•DNA 2	ANT•DNA	ANT-DNA	DNA	IMI	ANT
drug	3	2	3	4		6	9
	3.6 ± 1.0	2.2 ± 0.8	4.3 ± 1.1	3.6 ± 0.9	–	2.3 ± 1.1	5.4 ± 1.3
DNA	9	8	13	8	14		
	8.6 ± 1.5	7.5 ± 1.4	8.8 ± 1.7	5.6 ± 1.1	11.4 ± 1.6	–	–

Table 3.2: Total number of persistent H-bonds (present for more than 10% of the simulation time) formed by the drug and the five bps d($G_7T_8T_9G_{10}G_{11}$)d($C_{18}A_{17}A_{16}C_{15}C_{14}$) involved in the binding. Averages and standard deviations of the total number of H-bonds are also reported in the second row of each entry.
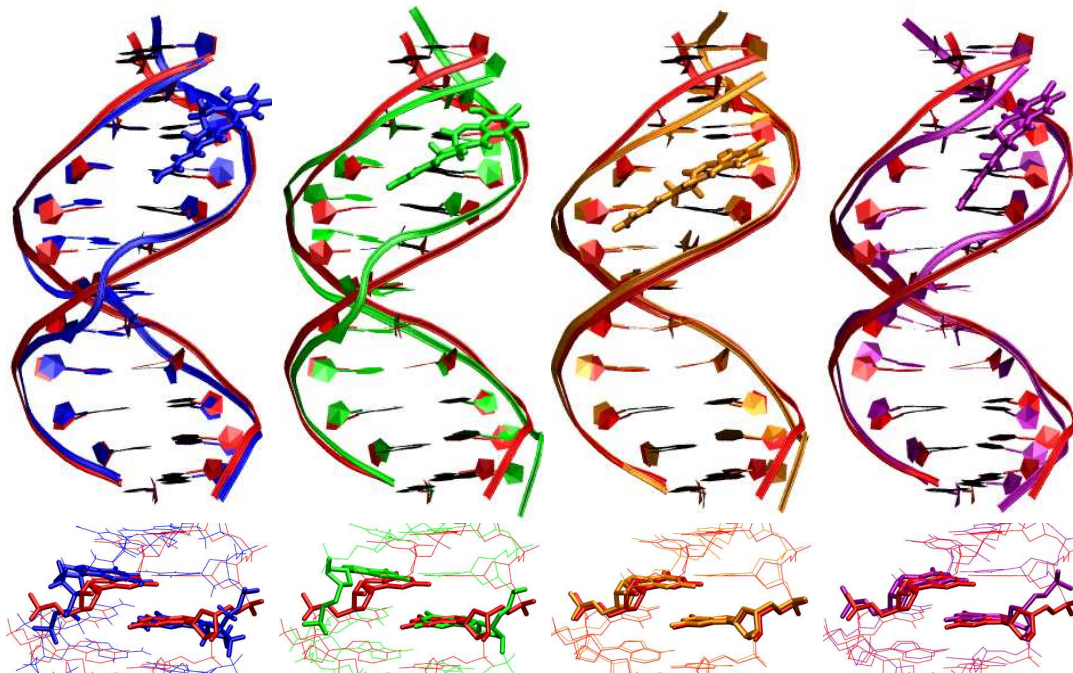
Figure 3.1: Superimposition between the MD-averaged structure of **DNA** (red) and those of **ANT•DNA** (blue), **IMI•DNA 2** (orange), **IMI•DNA 1** (green) and **ANT-DNA** (magenta). The details show the sugar flips that arise upon non-covalent binding.

have a minor role in stabilization, as compared to Ref. [99]. However, the aqueous solvent, not detected by the X-ray experiment, might assist significantly the formation of water-mediated H-bonds. We conclude that the MD structure of **ANT-DNA** reproduces most of the experimental data, apart from the drug-DNA H-bond pattern, which may differ mostly from the X-ray structure because of environment effects.

	$N3_{G11}..O9_{An}$	$O2_{C15}..O9_{An}$	$O2_{C15}..N10_{An}$	$N3_{G11}..N10_{An}$	$O2_{T9}..N15_{An}$	$O4'_{G10}..N15_{An}$
X − ray	3.9	3.1	3.4	3.7	2.9	3.3
MD	4.0 ± 0.3	2.9 ± 0.3	2.9 ± 0.2	3.3 ± 0.2	5.3 ± 0.4	4.9 ± 0.9

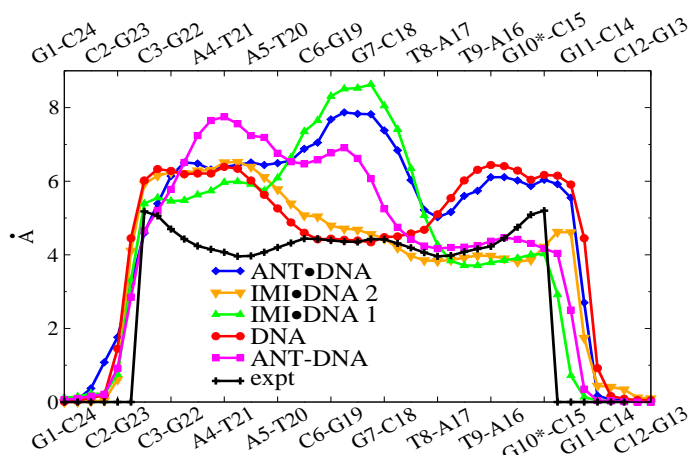Table 3.3: Comparison between X-ray and calculated acceptor-donor distances involving drug and DNA.

Figure 3.2: Average values of the minor groove width of **DNA** (red), **ANT•DNA** (blue), **IMI•DNA 1** (green) , **IMI•DNA 2** (orange), **ANT-DNA** (magenta), calculated between C4' sugars.

Non-covalent complexes

In this section, we provide insights on aspects of the molecular recognition process between anthramycin and DNA, through MD simulations of the non-covalent complexes **IMI•DNA** and **ANT•DNA**, and we compare them with the isolated reactants in water solution (**ANT**, **IMI**, and **DNA**), as well as that of the covalent adduct **ANT-DNA**, discussed above.

Structural analysis and sliding of anhydro-anthramycin. The structural stability of the systems has been firstly analyzed in terms of deviations from their averages, as well as from those of canonical B-DNA and A-DNA. Figure 3.4 shows that the RMSD of the oligonucleotides assumes two different values for several ns in the non-covalent complexes, while it is almost constant in **DNA** and **ANT-DNA**. Consistently, while for **DNA** and **ANT-DNA** the RMSD is 4 Å with respect to B-DNA and 5 Å with respect to A-DNA for the whole simulation, the non-covalent complexes assume for several ns conformations having lower RMSD relative to A-DNA than to B-DNA. The most interesting event observed is a sliding of the drug in **IMI•DNA** along the minor groove, towards the central part of the oligonucleotide (from site **1** - configuration **IMI•DNA 1** - to **2** - configuration **IMI•DNA 2** - in Figure 3.5), which occurs after $10\,ns$. Interestingly, shuffling and translocation have been observed experimentally for other DNA-binding drugs [248, 249]. The drug moves by one base pair (bp) step, so that $C11_{Ant}$ is located in the vicinity of $T9$ in **IMI•DNA 2**, and keeps the new position for the rest of the simulation. In contrast, in **ANT•DNA** the
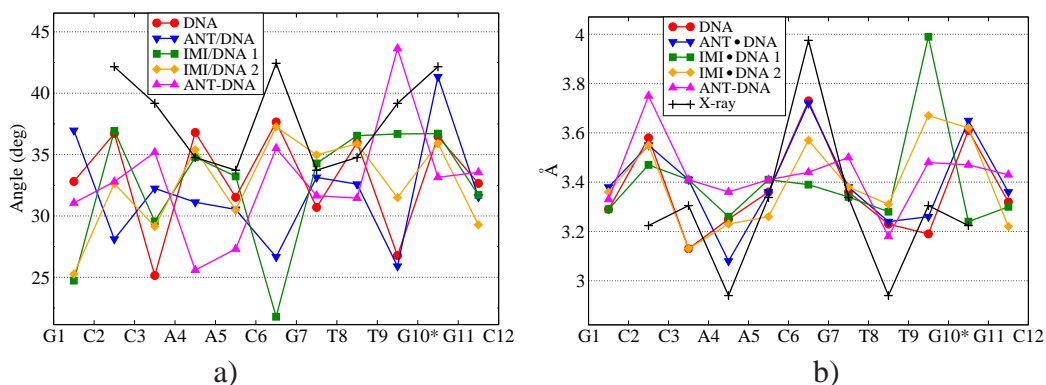
Figure 3.3: Average values of selected helical parameters (*a:* twist, *b:* rise) of **DNA** (red), **ANT•DNA** (blue), **IMI•DNA 1** (green), **IMI•DNA 2** (orange), **ANT-DNA** (magenta).

drug is stable around its binding site for the entire dynamics ($20 \, ns$). It is interesting to test whether **ANT** would move back if initially located in the "slipped" position of **IMI**. To this aim, we performed MD simulations in which **ANT** is docked in the non-reactive site (with the reactive carbon between bases T8 and T9). Two slightly different configurations were chosen. Both MD simulations, carried out for $2.5 \, ns$, show that the drug does not move back to the reactive site: in one MD, the average distance between $C11_{ANT}$ and $N2_{G10}$ passes from $6.3 \, \text{Å}$ to $5.6 \pm 0.3 \, \text{Å}$ (average value over the last $2 \, ns$), in the other from $7.4 \, \text{Å}$ to $7.6 \pm 0.3 \, \text{Å}$.

These results suggest that movement of **ANT** from one "site" to another are most probably not kinetically driven. A key issue to be addressed is on whether the sliding process is due (totally or in part) to the choice of the model, i.e. by having positioned the drug near to the end of the oligonucleotide rather than in its center: in fact, nucleobases and ligands in such positions might be subjected to different electrostatic and dynamical stresses. To quantify the relevance of end-effects on the properties of the biomolecular frame, we have compared features of **DNA**, i.e. the oligonucleotide in aqueous solution, in its central (d[ACGT]$_2$) and terminal (d[TGGC]$_2$) parts. The RMSD's with respect to the average structure turns out to be very similar ($1.4 \pm 0.3 \, \text{Å} \, vs \, 1.3 \pm 0.3 \, \text{Å}$), and the flexibilities, measured as RMSF's, are identical ($0.9 \pm 0.3 \, \text{Å}$). The differences in electrostatic forces acting on the base atoms are very small, within 15%. Furthermore, we have construct two more models other than **IMI•DNA**. In the first, the system is the same, but the initial conditions (in particular the docking of the drug, done manually) are different; $20 \, ns$ MD simulation of the model in aqueous solution show that a sliding of the drug occurs after $10 \, ns$ (note that this is quite long time in MD simulations with explicit solvent and counter-ions [250, 236, 35, 34]). In the second, which also has been simulated for $20 \, ns$, **IMI** is docked in the centre of the 14-mer
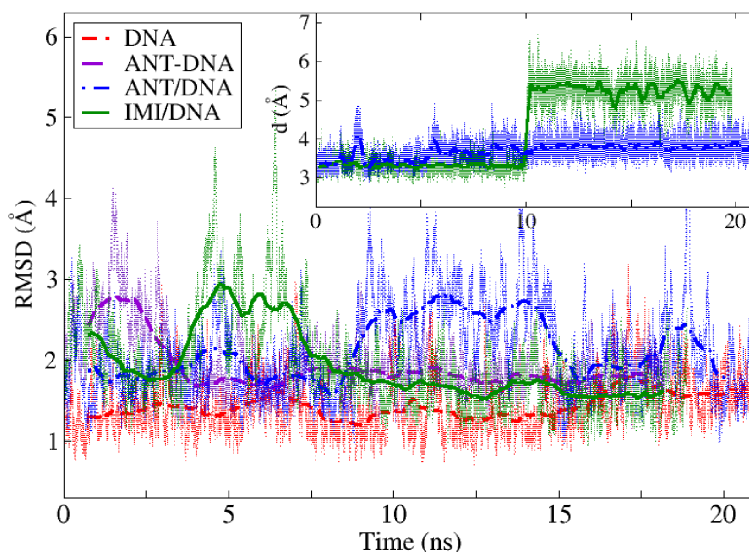
Figure 3.4: Mass weighted RMSDs of the oligonucleotides in **ANT•DNA**, **IMI•DNA**, **ANT-DNA**, and **DNA**, calculated for heavy atoms after least square (LSQ) fit to the average structures in solvent. Running averages are reported as bold lines. The inset reports the distance between reactive atoms $N2_{G11}$ and $C11_{Ant}$ in the non-covalent complexes.

d[CAACGTTG*GCCAAC]d[GTTGGCCAACGTTG] (**IMI•DNAc**). In this case the same sliding process occurs after $15\,ns$; in addition, flexibility, minor groove width and electrostatic field acting on the drug, turn out to be very similar to those of **IMI•DNA** (Table 3.4 and Figure 3.6). Based on results from the structural analysis of **DNA** and from three independent simulations, we suggest that the sliding of the drug is caused by a stabilisation of the drug in a position other than the "reactive" one, and we proceed to a structural and energetic analysis of the process. Structurally, the sliding of the drug is accompanied by a significant rearrangement of the DNA frame (mirrored by a decrease of $8\,kcal/mol$ of the internal energy): in fact, the average structure of the complex before the sliding (**IMI•DNA 1**) shows large deformations with respect to those of **DNA** and **ANT-DNA**, in particular at the central region (Fig-

system\step	C12-C15	G11-C16	G10-A17	T9-C18
IMI•DNA	6.8 ± 0.7	4.9 ± 0.4	4.4 ± 0.4	6.1 ± 0.6
IMI•DNAc	4.4 ± 0.7	5.1 ± 0.4	4.4 ± 0.4	4.0 ± 0.6

Table 3.4: Average values (Å) of minor groove widths calculated between C4' atoms at d[TG*GC]$_2$.
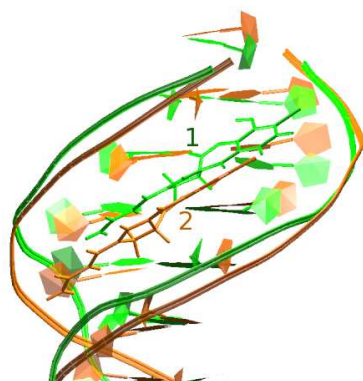
Figure 3.5: Superimposition of the **IMI•DNA** MD-averaged structures before (green, **1**) and after (orange, **2**) $10\,ns$. The drug moves by 1bp step. The mean distances between reactive atoms $N2_{G11}$ and $C11_{Ant}$ are respectively 3.3 ± 0.2 Å in **1** and 6.5 ± 0.5 Å in **2**.

ure 3.1). The plot of the average minor groove widths (Figure 3.2) shows the presence of a large unwinding in this region relative to **DNA** and **ANT-DNA**, which is correlated to a flip of the χ glycosidic dihedral at $G7$ and $G19$ from *anti* (-120) to *high anti* configuration (-70 - Figure 3.1). The same findings come out from the analysis of helicoidal inter-base-pairs parameters (Figure 3.3). The oligonucleotide after the sliding (**IMI•DNA 2**) becomes structurally similar to **DNA** and the minor groove narrows, in particular in the central region of the oligonucleotide. The other non-covalent complex (**ANT•DNA**) assumes also a conformation differing from those of **DNA** and **ANT-DNA** particularly in its middle (Figures. 3.1, 3.2, 3.3). However, the minor groove has almost the same width as in **DNA**, due to the presence of the $OH - 11$ group of **ANT** (Chart I). Summarizing, both the non-covalent complexes with the drug positioned in front of the reactive site exhibit significant rearrangements
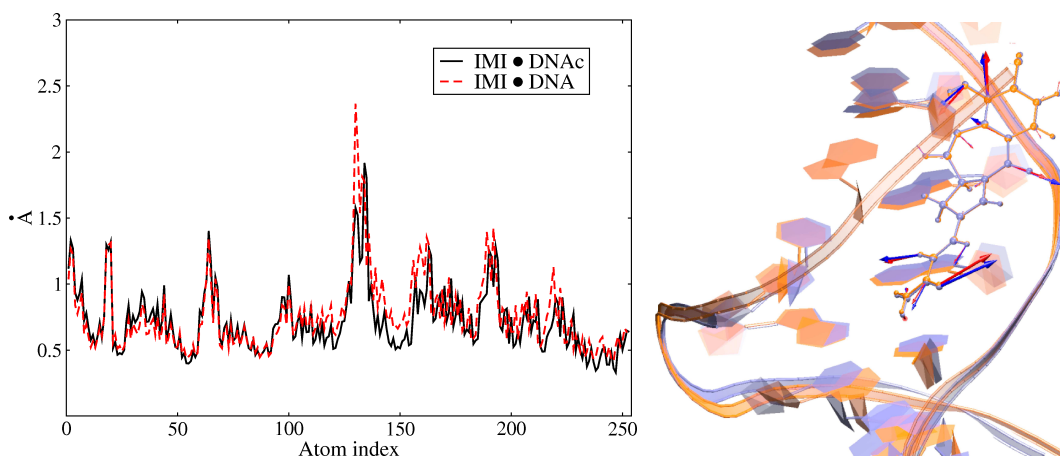

Figure 3.6: *Left*: Comparison between RMSFs around the average structure at d[TG*GC]$_2$ in **IMI•DNA** and **IMI•DNAc**. *Right*: Average electrostatic forces acting on the drug in **IMI•DNA** (cyan) and **IMI•DNAc** (orange).
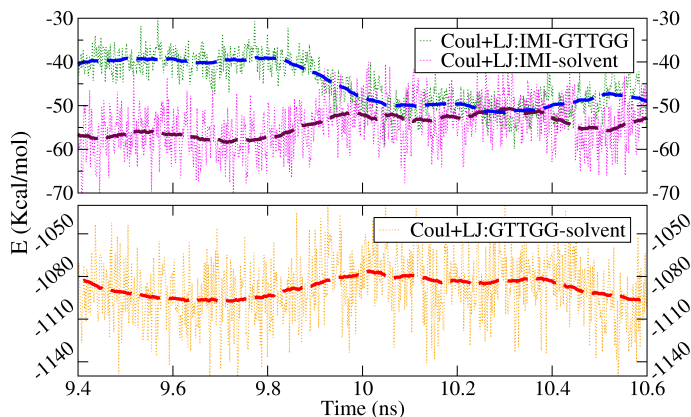
Figure 3.7: Variations of the non-bonded (electrostatic+LJ) energy of the drug-$d(G_7T_8T_9G_{10}G_{11})d(C_{18}A_{17}A_{16}C_{15}C_{14})$, drug-solvent, and solvent-$d(G_7T_8T_9G_{10}G_{11})d(C_{18}A_{17}A_{16}C_{15}C_{14})$ moieties, upon sliding of anhydro-anthra-mycin (dashed lines correspond to running averages performed on blocks of 100 steps). The sliding process is associated to a increase of $\sim 10\,kcal/mol$ in the interaction energy between the drug and the dodecamer.

in the center of the oligonucleotide with respect to **DNA** and **ANT-DNA**. Upon sliding of **IMI** the conformational parameters recover values close to those of **DNA**. Some values of the helical parameters in Figure 3.3, like high twist and rise, and negative roll, are typical of BII-DNA backbone conformation [50]. BI/BII equilibria can be influenced by ligands, in order to optimize contacts between partners and to allow the formation of water-mediated H-bonds [251, 90]. In fact, we find that the binding of the drug affects these equilibria (data not shown), even if a correlation between binding and conformational changes of the backbone is not easy to rationalize. For example, binding either imposes freezing of certain parameters, or allows a major variability of others. In addition, the presence of a low twist (and rise) at CG steps (like in C_6G_7) in MD simulations has been recently correlated to unusual α/γ backbones at the adjacent 5' step [236]. We found indeed along the backbone some flips of α/γ to stable states other than the canonical g^-/g^+, but they appear not to be cor-

	ANT•DNA	IMI•DNA 1	IMI•DNA 2	ANT-DNA
LJ	-30 ± 2	-33 ± 3	-40 ± 3	-34 ± 3
COULOMB	-12 ± 3	-8 ± 3	-11 ± 3	-20 ± 4

Table 3.5: LJ and coulomb interaction energies (kcal/mol) extracted from the non-bonded terms of the AMBER force field [25, 131].

	ANT-DNA	ANT•DNA	ANT	IMI•DNA 1	IMI•DNA 2	IMI
PROPELLER	31 ± 7 (37.1)	26 ± 8	22 ± 8 (34.4^*)	29 ± 7	22 ± 7	26 ± 9
BUCKLE	13 ± 6 (17.6)	10 ± 5	12 ± 6 (8.2^*)	8 ± 4	8 ± 4	9 ± 5

Table 3.6: Propeller twist and buckle of the drugs (degrees). In parenthesis are reported available experimental results. Data labelled with the asterisk refers to the $O11$-methyl ether of the drug.

related with the low twist at C_6G_7. The lifetimes of these conformations are supposed to be larger than typical MD timescales [52], yet our and other simulations do observe them [236], even if in almost all the cases reverse transitions are absent. The origin of their appearance has not been enlighted yet, although it might be related to the force field parametrization. Also the drugs exhibit a slight distortion upon binding, as shown by the average values of the propeller and twist angles reported in Table 3.6. Despite the high standard deviations, twist seems to be anti-correlated with the degree of binding, with the only exception of **IMI•DNA 2**. The buckle instead does not change significantly upon binding. The energetic analysis has been performed using the non-bonded terms of the AMBER force field [24, 25, 131] for an approximate estimate of the interaction energies. Although the number of drug-DNA H-bonds is reduced in going from **1** to **2** the overall interaction energy increases by few $kcal/mol$, (Figure 3.7). This mirrors a better accommodation of the drug into the DNA minor groove after the sliding, increasing the hydrophobic contact area by 4% (Table 3.7). As a result, the twist of the drug decreases ($29°$ *vs* $22°$ in Table 3.6), allowing a better
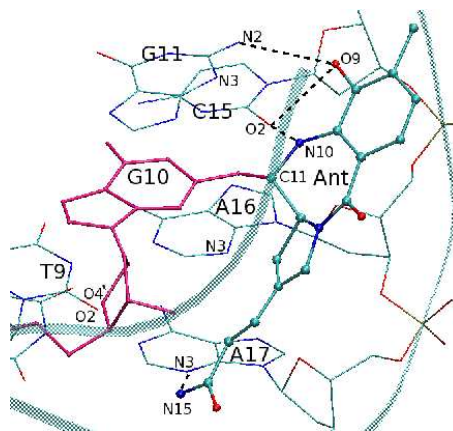

Figure 3.8: Drug/oligonucleotide H-bonds in **ANT-DNA** (MD structure).

	IMI	ANT	GTTGG	IMI•CTTGG	ANT•GTTGG	ANT-GTTGG
SASA	191 ± 5	161 ± 5	764 ± 17	$667 \pm 18\,(\mathbf{2})$ $709 \pm 17\,(\mathbf{1})$	746 ± 25	739 ± 21
ΔSASA				$-288\,(30\%,\,\mathbf{2})$ $-246\,(26\%,\,\mathbf{1})$	$-179\,(19\%)$	$-216\,(26\%,\,\mathbf{IMI})$ $-186\,(20\%,\,\mathbf{ANT})$

Table 3.7: Calculated hydrophobic SASA's (Å^2). The hydrophobic effect plays a role in the formation of the complex: a negative value of ΔSASA upon solvation in water corresponds to a gain in the non-polar part of the solvation free energy.

delocalization of π orbitals on the drug. The entropy change associated to the sliding process turns out to be very small, as: *(i)* the first shell of waters surrounding the drug and the binding site does not change, as shown by a calculation of the radial distribution function of waters around the binding region (data not shown); *(ii)* the conformational entropy of the DNA (calculated with the Schlitter's formula [252]) is unchanged. Within the limitations of these simple estimations, we conclude that the sliding of the drug is mainly due to a gain in non-specific interactions with the DNA frame, accompanied by a release of the structural stress accumulated by this latter under the presence of the drug.

Other features of IMI•DNA 2 and ANT•DNA. The LJ interactions account for $\sim 70-80\%$ of the non-bonded energy for both the systems, while electrostatics plays a minor role (Table 3.5). Note that the LJ (Coulomb) interaction is slightly stronger (weaker) in **IMI•DNA 2** than in **ANT•DNA**, possibly because of the presence in the latter of the $OH11$ group, that prevents the drug to be very close to the DNA backbone. Consistently, the relative decrease of the hydrophobic SASA [240] upon formation of the complex is larger in **IMI•DNA 2** than in **ANT•DNA** (Table 3.7), corresponding to a larger relative gain in the non-polar contribution to the free energy of solvation. Thus, **IMI** establishes a more favourable non-bonded interaction with the DNA frame than **ANT**. In the covalent complex the electrostatic energies become comparable to the LJ ones (Table 3.5): in fact, the formation of the covalent bond shortens the distances between charged groups, inducing a different distribution of the electronic density on the reactants. Interestingly, the H-bonds between the solvent and **IMI**, **ANT**, **DNA** (calculated only for the atoms at the drug-DNA interface) are only partially replaced in the non-covalent complexes (Table 3.2). Specifically, drug-DNA H-bonds are formed between $G10$, $G11$, $A16$ nitrogens, $O4'_{G11}$, and $O9$, $O11$ and $N10$ on the drug (Figure 3.8). No water molecules are present between the reactive guanine $G10$ and the drugs, as should be expected for extended aromatic systems making favourable hydrophobic interactions with the DNA backbone. However, the presence of the $OH11$ group in **ANT•DNA** allows the entrance of water(s) in the
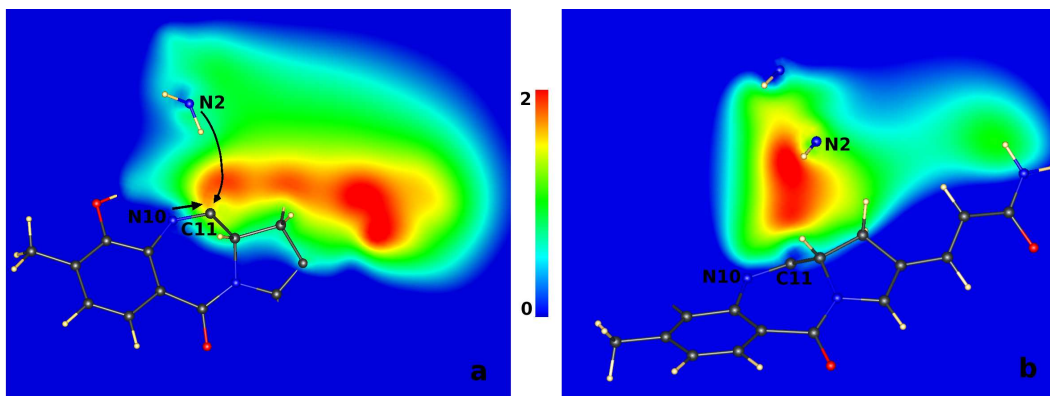
Figure 3.9: Contour plots of the electrostatic potential (Volts) within the QM region, through the plane defined by atoms $N2$, $N10$, $C11$. The plots report the difference in the electrostatic potential calculated *with* and *without bioframe*. Note that only in **IMI•DNA** (*a*) the largest difference is *localized* near to the electrophilic carbon, while in **ANT•DNA** (*b*) it is *spread* in front of the reactive amino group of the guanosine. The $N10$-$C11$ arrow indicates the polarization of the bond, while the arrow between $N2$ and $C11$ indicates the gradient in ΔV.

region between the drug and the DNA backbone (Table 3.2). Based on these results we suggest that the formation of the complexes has an important entropic contribution, and their stabilization, even in a highly charged system like the minor groove of DNA, is mainly due to hydrophobic contacts.

3.3.2 Reactant Polarization

DNA is a polyelectrolite and it may produce large electric fields on the reactants. In this section, we investigate the electrostatic effect of the biomolecular frame on the electronic structure of the reactants (the drug and its guanosine target) by $3\,ps$ QM/MM simulations on **ANT•DNA** and **IMI•DNA 1**. Within this scheme the drug and the reactive guanosine are treated at a quantum level, while the rest of the system is parametrized using AMBER [231, 25, 131] effective potential. Thus the QM/MM scheme allows to dissect the source of polarization into various contributions, namely those due to the DNA and the solvent, by switching off their charges. We also evaluate the amount of reciprocal polarization of the reactants by performing additional QM calculations *in vacuo*.

Our QM/MM simulations suggest that the biomolecular frame alters significantly the electric field of the reactants, in particular the region between the reactive atoms $N2_{G10}$ and $C11_{drug}$ (Figure 3.9). The polarizing field created by the environment turns out to be due almost completely to the DNA frame, since identical results are

|		bioframe	no bioframe	no solvent	vacuo
ANT•DNA	$C11 - H11$	0.66 ± 0.01	0.68 ± 0.04	0.66 ± 0.01	0.68
	$C11 - N10$	0.57 ± 0.01	0.54 ± 0.01	0.58 ± 0.01	0.55
	$C11 - N10$	0.58 ± 0.01	0.55 ± 0.01	0.59 ± 0.01	0.56
	$C11 - C11a$	0.51 ± 0.01	0.51 ± 0.01	0.50 ± 0.01	0.51
	$N2 - H21$	0.57 ± 0.01	0.59 ± 0.01	0.57 ± 0.01	0.59
	$N2 - H22$	0.59 ± 0.01	0.58 ± 0.01	0.57 ± 0.01	0.58
	$N2 - C2$	0.36 ± 0.04	0.37 ± 0.04	0.38 ± 0.06	0.33
	$N2$ lone pair	0.21 ± 0.05	0.19 ± 0.05	0.21 ± 0.07	0.24
IMI•DNA	$C11 - H1$	0.66 ± 0.01	0.68 ± 0.04	0.66 ± 0.01	0.67
	$C11 - N10$	0.57 ± 0.01	0.57 ± 0.01	0.58 ± 0.01	0.57
	$C11 - C11a$	0.50 ± 0.01	0.50 ± 0.01	0.50 ± 0.01	0.51
	$C11 - O11$	0.63 ± 0.01	0.62 ± 0.01	0.62 ± 0.01	0.62
	$N2 - H21$	0.57 ± 0.01	0.59 ± 0.01	0.57 ± 0.01	0.59
	$N2 - H22$	0.58 ± 0.01	0.58 ± 0.01	0.59 ± 0.01	0.58
	$N2 - C2$	0.37 ± 0.04	0.38 ± 0.04	0.38 ± 0.04	0.33
	$N2$ lone pair	0.20 ± 0.05	0.15 ± 0.06	0.15 ± 0.06	0.24

ANT

IMI

GUA
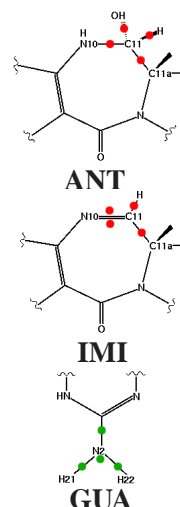
Table 3.8: Bond ionicity of $X\text{-}N2_{G10}$ and $X\text{-}C11_{drug}$ bonds. Last column refers to calculations *in vacuo* on the drug or the guanosine. Two rows with the same label indicate a double bond, while the last row in each system represents the BI associated to the lone pair of the guanine nitrogen. The figures on the right side depict the BO's considered for the BI calculations (see Methods).

obtained in calculations in which the charges of the solvent are switched off (data not shown). The DNA field is directed from $N2_{G10}$ to $C11_{drug}$ in **IMI•DNA 1**. Thus, this field may assist the alkylation reaction by rendering $C11_{drug}$ more electrophilic, i.e creating electrostatic conditions very favourable for a nucleophilic attack. Instead, in **ANT•DNA** the largest effect occurs in proximity of the amino group of $G10$. Thus, in this latter case, the DNA field is expected to enhance the reactivity neither of the drug nor of the DNA. These findings are in agreement with experimental findings [253, 254], indicating the anhydro form as the most reactive.

A measure of the polarity of chemical bonds is given by bond ionicity indexes (BI's). For the bond between atoms A and B, a BI<0.5 (>0.5) indicates shift of the center of mass of the electronic charge towards A (B) (see Methods). Such analysis can be extended also to investigate the polarization of electron lone-pairs of a given atom. In this case, the BI is defined in such a way that larger is the index, stronger are the polarization effects on the lone-pair. Here we investigate the changes in BI's, averaged along the QM/MM trajectory, due to the inclusion of the electric field of the biomolecular frame. It is worth to notice that also little differences between these values are significant, because of the restricted range of values of the BO projection onto the bonds [247]. The $N10\text{-}C11$ π bond is the most polarized by the external electric field in **IMI•DNA 1** (where the drug is located in front of the reactive guanine [1]),

---

[1] Note that, due to the delocalization of π bonds [247], these results should be taken as an indication.

whilst in **ANT•DNA** the largest effect is seen on the $N2_{G10}$ lone pair (Table3.8). Consistently, a large field is localized on the $C11$ atom in the first system (Figure 3.9). Thus, in **IMI•DNA 1** the $C11$ atom is more electron poorer than *in vacuo*, suggesting that DNA renders the complex more reactive. These results agree with experiments, which have shown a marked preference of anthramycin to bind double-stranded DNA (dsDNA), rather than single nucleobases [98, 255, 99].

3.4 Concluding Remarks

Whilst a large number of ns-scale simulations of non-covalent drug/DNA complexes have appeared in the literature (see e.g. Refs. [74, 75, 256, 257, 76]), very few MD studies have been reported on alkylating drug/DNA systems, in spite of their pharmacological relevance [258, 104, 259]. Here we have investigated by molecular mechanics simulations the non-covalent interaction between two forms of anthramycin present in solution (hydroxy and anhydro) and the 12-mer d[GCCAACGTTGGC]$_2$. Our MD simulations suggest that the formation of the covalent complex does not affect DNA helical axis, in agreement with experiments [99]. However, local structural changes are evident in both the DNA and the drug, consistently with the "induced fit mechanism" of binding proposed by Kizu *et al.* [221]. The non-covalent complexes are mainly stabilized by non-specific hydrophobic interactions, consistently with the modest sequence selectivity shown by these drugs [217,218,260,15]. In this respect, it is worth to note the remarkable difference observed in dynamical properties of the two forms of the drug. Hydroxy-anthramycin remains in the same site during the whole length of our simulation. In contrast, anhydro-anthramycin moves after $\sim 10\,ns$ towards an adjacent site, and the system keeps the new configuration for the rest of the dynamics. Note that similar shuffling of minor groove binders has been already detected experimentally [248]. To the best of our knowledge, however, this is the first time that such a behaviour is observed for minor groove alkylating drugs in a MD simulation. The sliding of the anhydro form is associated with an energetic and structural relaxation of the non covalent complex, and is seen in a variety of additional simulations of the complex. The DNA conformation of the reactive complexes (i.e. with the drugs in front of its target guanine) is distorted relative to that assumed in solution or in the covalent complex. In particular, significant changes are observed in sugar, backbone, and helical parameters at the central region of the oligonucleotide. These values are associated with conformations higher in energy than that of canonical B-DNA. Interestingly, after the sliding of anhydro-anthramycin (non-reactive complex), the above structural parameters become similar to those of the free DNA. We conclude therefore that the oligonucleotide in the reactive complex is more distorted than in solution, in the observed non-reactive complex, and in the final product of the reaction. This requirement of a sizeable structural distortion of the DNA frame and, to a lesser

extent, that of the drug, could play a role for the experimentally observed slower reactivity of anthramycin relative to that of other alkylating agents [98]. Classical simulations are complemented by QM/MM calculations. These show that the dsDNA electric field polarizes the reactive carbon of anhydro-anthramycin so as to render it more electrophilic. On the contrary, in the complex with hydroxy-anthramycin, the DNA field is unable to increase either the drug electrophilicity or the guanine nucleophilicity (although it clearly affects the electronic structure of the reactants). The field due the water solvent appears instead to play a negligible role in both the complexes. Thus, although the exact role of the electrostatic field of the environment can be established only if one simulates the entire alkylation reaction, our results suggest that the DNA frame assists the alkylation reaction by polarizing one of the reactants, at least in the case of the complex with anhydro-anthramycin. This is consistent with the experimental observation that the drug is able to covalently bind G-containing oligonucleotides but not guanine in aqueous solution [98, 255, 99].

Chapter 4

Mechanism and thermodynamics of drug sliding along the minor groove

4.1   Introduction

The design of sequence-selective anticancer molecules which bind to regions of DNA involved in replication and transcription processes is a key goal of rational drug design [261, 262, 263, 264]. In this respect sequence-specific read-out codes have been deciphered for the class of non-covalent minor groove binders [73, 265]. In contrast, the important class of DNA-alkylating agents has received less attention, particularly from a theoretical point of view, probably also due to the paucity of experimentally determined structural data [19, 9]. It is commonly accepted that molecular recognition and formation of the non-covalent complex by these drugs are driven by non specific interactions and sequence-specific structural features along the minor groove [19, 9]. Recently, molecular dynamics studies performed in the group [83, 84] have provided information at the molecular level on the non-covalent interaction between the DNA and two covalent binders, anthramycin [253, 99, 254, 221] and duocarmycin [107, 266]. Anthramycin alkylates guanines, showing a modest sequence selectivity for PuG*Pu sequences [221], while duocarmycin binds to adenines and is very selective towards AT-rich sequences which have to be at least 4 bps long [107, 266]. Both compounds are powerful cytotoxic agents interfering with transcription and replication processes, and some of their derivatives have entered clinical test phases [223, 267]. The natural twist of 35° between phenol and pyrrol rings (respectively A and C in Chart I), gives to anthramycin the ideal shape to fit into the minor groove [99]. In contrast, the largest duocarmycin is formed by two moieties connected via an amide link and to fit into the minor groove it needs a 40° twist around the amide link relative to its conformation in water [266].

For the purposes of the present work, we focus on the putative reactive imine form

of anthramycin [hereafter IMI] [253, 99, 254, 221], and on duocarmycin-SI [hereafter DSI] [107, 266]. Non-covalent complexes of both IMI and DSI with DNA are mainly stabilized through hydrophobic interactions [83, 84]. However, while IMI could form a relatively strong H-bond network with DNA [84, 99], DSI is able to form one H-bond only [83, 107]. As seen in the previous Chapter, IMI turned out to slide along the minor groove of the d[GCCAACGTTGGC]$_2$ duplex, leading to the formation of a non reactive stable complex. Such a sliding is independent on the initial location of the drug, occurring when IMI sits either at the end or in the middle of a DNA duplex. The same kind of displacement has been observed after docking the molecule to its preferred site, the triplet AGA [221], within a 14-mer duplex (data not shown). Instead, the complex between DSI and the duplex d[GACTAATTGAC]$_2$ is stable during the whole dynamics [83].

As the non-covalent recognition of the preferred DNA sequence may involve a sliding of drugs along the minor groove [248, 249, 268, 269], knowledge of the energetics of such process may provide useful information on ligand selectivity and molecular recognition. In this chapter this issue is addressed by investigating quantitatively the mechanism of sliding of IMI and DSI. The free energy is then dissected at a qualitative level into its enthalpic and entropic contributions, the rank of which has been shown to depend strongly on the chemical structure of the drug and on the DNA sequence [270, 271].

It is found that binding of IMI to the reactive site is less favored than at nearby base-pair, whereas for DSI the non-covalent binding site coincides with the reactive one, in agreement with previous MD simulations [83, 84]. While the PMF associated to sliding of IMI can be roughly rationalized through a simple analysis of drug/DNA enthalpic interactions, solvation effects appear to be much more relevant for DSI, as a consequence of more stringent requirements for optimal fit of drug into the minor groove. These findings suggest the need of considering multiple binding pathways in drug design, and provide a rationale for the modest selectivity of anthramycin relative to duocarmycin [266, 217].

4.2   Systems and Methods

Free energy profiles have been calculated as a function of the position of the drug along the minor groove, for the non-covalent complexes of IMI with d[GCCAACGTT-G*GC]-d[GCCAACGTTGGC] (hereafter **IMI•DNA**) and DSI with d[GACTAATTG-AC]-d[GTCAATTA*GTC] (hereafter **DSI•DNA**). These are built starting from experimental structures [99, 266] by cutting the covalent bond between the carbon of the drug and the nitrogen of the nucleobase, and manually pulling out the drugs from the minor groove, until the distance d[C-N] was about 3.3 Å. As DSI moves toward the nearest end of the duplex upon sliding, test calculations have been to investi-
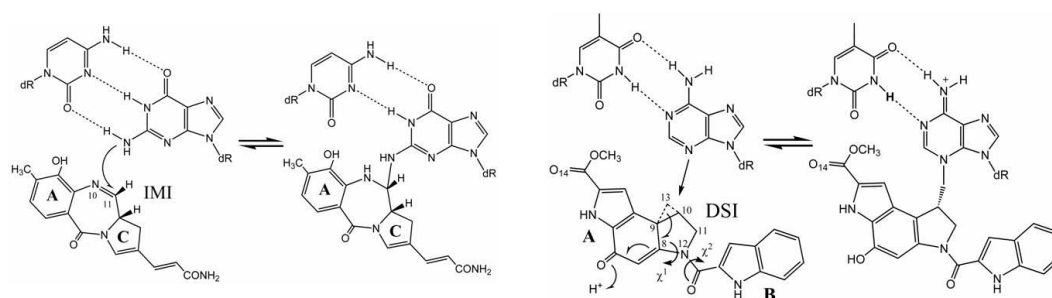
ChartI: Schematic view of alkylation of guanine by IMI (left), and adenine by DSI (right).

gate the possible influence of end-effects [84]. To this end, an additional simulation was performed on the 14-mer d[GACGACTAATTGAC]-d[GTCAATTA*GTCGTC] [hereafter**DSI•DNAc**] in complex with DSI. The drug was placed with its reactive carbon C13 in front of the base T24 (corresponding to T21 in **DSI•DNA**). A 7 ns MD simulation was performed on this system, and analyses were performed on the last 2 ns.

All simulations have been carried out using the GROMACS package [237, 238, 272]. AMBER/gaff force fields [24, 25, 131, 234] were used for the parameterization of oligonucleotides and drugs [83, 84]. Drug structures were optimized by means of DFT calculations at B3LYP/6-31G(d,p) [161, 157] level, using the Gaussian03 package [273]. Atomic RESP [134] charges were derived using the *resp* module of AMBER after wavefunction relaxation. Potassium ions, modelled with the AMBER-adapted Aaqvist potential [233], have been added to achieve charge neutrality (22, 20 and 28 in **IMI•DNA** and **DSI•DNA** and **DSI•DNAc**, respectively). Systems have been solvated with a cubic box of TIP3P water molecules [274] ensuring solvent shell to extend for at least 12 Å around the DNA. Periodic boundary conditions were used, and constant temperature-pressure (T=300 K, P=1 atm) dynamics have been performed through the Nosé-Hoover [121, 122] and Andersen-Parrinello-Rahman [123,126] coupling schemes ($t = 1$ ps). Electrostatic interactions were treated using the particle mesh Ewald (PME) algorithm [275] with a real space cutoff of 10 Å, the same as for Van der Waals interactions. The pairlist was updated every 10 steps, and Lincs constraints [235] were applied to all bonds involving hydrogen atoms, allowing to use a time step of 2 fs. Coordinates were saved each 500 steps, corresponding to 1000 snapshots per ns. DNA minor groove width was defined as the distance between sugar C4' atoms, and it was calculated with the program Curves [276, 239].

The potentials of mean force (PMF's) [277] associated to drug sliding along the minor groove of **IMI•DNA** and **DSI•DNA** were calculated with the Umbrella Sampling method [183]. The distance between the reactive atoms of drug and DNA base (specifically the C11 atom of IMI to the N2 atom of guanine 10 in the DNA and atom

C13 of DSI to atom N3 in adenine 19 of the DNA) was chosen as reaction coordinate. This simple choice is well suited to describe movements of the drugs that are few bps long, since both compounds cause no appreciable bending of DNA [83, 84]. In order to sample the various conformations corresponding to different positions of the drug along the minor groove, we have imposed a harmonic constraint of $15\,kcal/(mol \cdot Å^2)$ to the distance d[C-N], starting from $2.9\,Å$ to $12.1\,Å$ with a step of $0.4\,Å$. Doing so the total movement of $9.2\,Å$ (slightly less than 3 bps steps) is partitioned into 23 windows. Initial configurations of each window were generated starting from the reactive one, and increasing the distance d[C-N] by steps of $0.2\,Å$, after $70\,ps$ of equilibration for each d[C-N] value. The Weighted Histogram Analysis Method (WHAM) [278, 185] method has been used to recombine PMF obtained from different windows. As can be seen from insets in Figures 4.1 and 4.2, the error on the calculated PMFs is almost constant, meaning that the reaction coordinate has been sampled in a fairly uniform manner.

Extended MD simulations of oligonucleotides in water [62] indicate that some of the DNA conformational and helicoidal parameters have relaxation times of 0.5 ns. Thus, at least multi-ns trajectories must be collected to obtain well-converged free energy profiles [236]. Systems investigated here were equilibrated for $\sim4\,ns$, albeit the PMF landscape was roughly converged after $2\,ns$. A further check on the RMSD of drug●DNA complexes indicates that convergence has been reached for all of the windows after $\sim4\,ns$.Therefore, $6\,ns$ of MD simulation were performed on each window (for a total of $\sim140\,ns$), and the last $2\,ns$ were used to estimate the PMF and for a structural analysis of the complexes.

The PMF has been decomposed as the sum of individual components, each with a physical meaning, and evaluated *via* the force field terms. In particular, variations in the drug-DNA electrostatic and Van der Waals interactions, [1] solute adaptation energy, solute configurational entropies, and solvation free energies were examined. Solute vibrational entropies were calculated within the harmonic approximation [279], using the *nmode* module of AMBER 9 [280]. As customary [281, 39, 282], a subset of structures was selected to be analyzed (20 for each relevant configuration, extracted every 0.1 ns from the last $2\,ns$). These were minimized in the absence of explicit solvent, the latter being mimicked by a dielectric constant $\epsilon = 4r$ (r is the interatomic distance in $Å$). Then, up to 20000 steps of minimization with no cutoff for all the interactions were performed, of which the first 100 are steepest descent followed by conjugate gradient, until the RMS of gradient drops below $10^{-4}kcal/(mol \cdot Å)$. The estimate of relative solute conformational entropies is not a trivial task [39, 282, 119], in particular for very flexible systems such as oligonucleotides. Moreover, it is not

---

[1]Van der Waals interactions were roughly estimated by summing the number of hydrophobic contacts for each atom type pair. For each of these, the equilibrium distance d_0 of the Lennard-Jones potential was calculated and contact was counted when $d<(d_0+0.2\,Å)$.

possible to guess *a priori* how the "spreading" of the dynamics among the various basins of the free energy surface depends on the conformation (i.e. the position of the drug). Therefore, results on conformational entropies should be taken only as suggestive.

Solvent contributions to free energy were calculated using the MM-PBSA methodology [283]; for each relevant configuration 200 snapshots from the last 2 ns of each MD were saved. Since $\Delta G_{solv} = G_{solv}^{complex} - G_{solv}^{drug} - G_{solv}^{DNA}$, and the last two terms in the right-hand side are the same for each configuration of the complex, ΔG_{solv} is directly proportional to $G_{solv}^{complex}$. Based on this proportionality, only this latter term was evaluated for each relevant configuration of the two complexes. The electrostatic contribution to solvation was evaluated using the Poisson-Boltzmann continuum method [284], as implemented in the module *pbsa* [285] of AMBER 9. Dielectric constants for the internal and external parts were set to 1 and 80 respectively. Grid dimensions were set as long as twice the tickness of the solute, and a grid spacing of 0.25 Å was used. The dielectric boundary is the molecular surface defined by a 1.4 Å probe sphere and by spheres centered on each atom with radii taken from the PARSE [286] parameter set (H=1.0, C=1.7, N=1.5, O=1.4 Å with a value of 2.0 Å for the phosphorus). The boundary dielectric constants were set as the harmonic sum of solvent and solute Debye-Hückel values. Salt effects were not included implicitly in the continuum model. The hydrophobic component of solvation free energy is assumed proportional to the change of the solvent accessible surface area (SASA), $\Delta G_{np} = \gamma \Delta SASA + b$ where γ=0.00542 kcal/Å^2 and b=0.92 kcal/mol [286]. For comparison, we performed MM-PBSA calculations also using a second approach available in the *pbsa* module of the AMBER 9 package [285, 287]. In this method the non polar contribution is cast into two terms, a repulsive one (cavity), correlated to the SASA, and an attractive one (dispersion), calculated through a surface-integration approach [288].

4.3   Results and Discussion

In this section a description of PMF profiles corresponding to the sliding of IMI and DSI inside the minor groove of oligonucleotides d[GCCAACGTTG*GC]-d[GCCAA-CGTTGGC] and d[GACTAATTGAC]-d[GTCAATTA*GTC] (**IMI•DNA** and **DSI•DNA** respectively), considering d[C-N] as reaction coordinate. Note that the standard deviation is very small along the entire sampling interval ($\approx$0.2 kcal/mol, see insets in Figures 4.1, 4.2), which makes the estimation of the free energy quantitatively reliable. Then, the solute enthalpic and entropic contributions to the PMF are analyzed, along with the effects of the solvent. Finally, the relevance of end-effects [84] on the energetics of the sliding process is assessed.
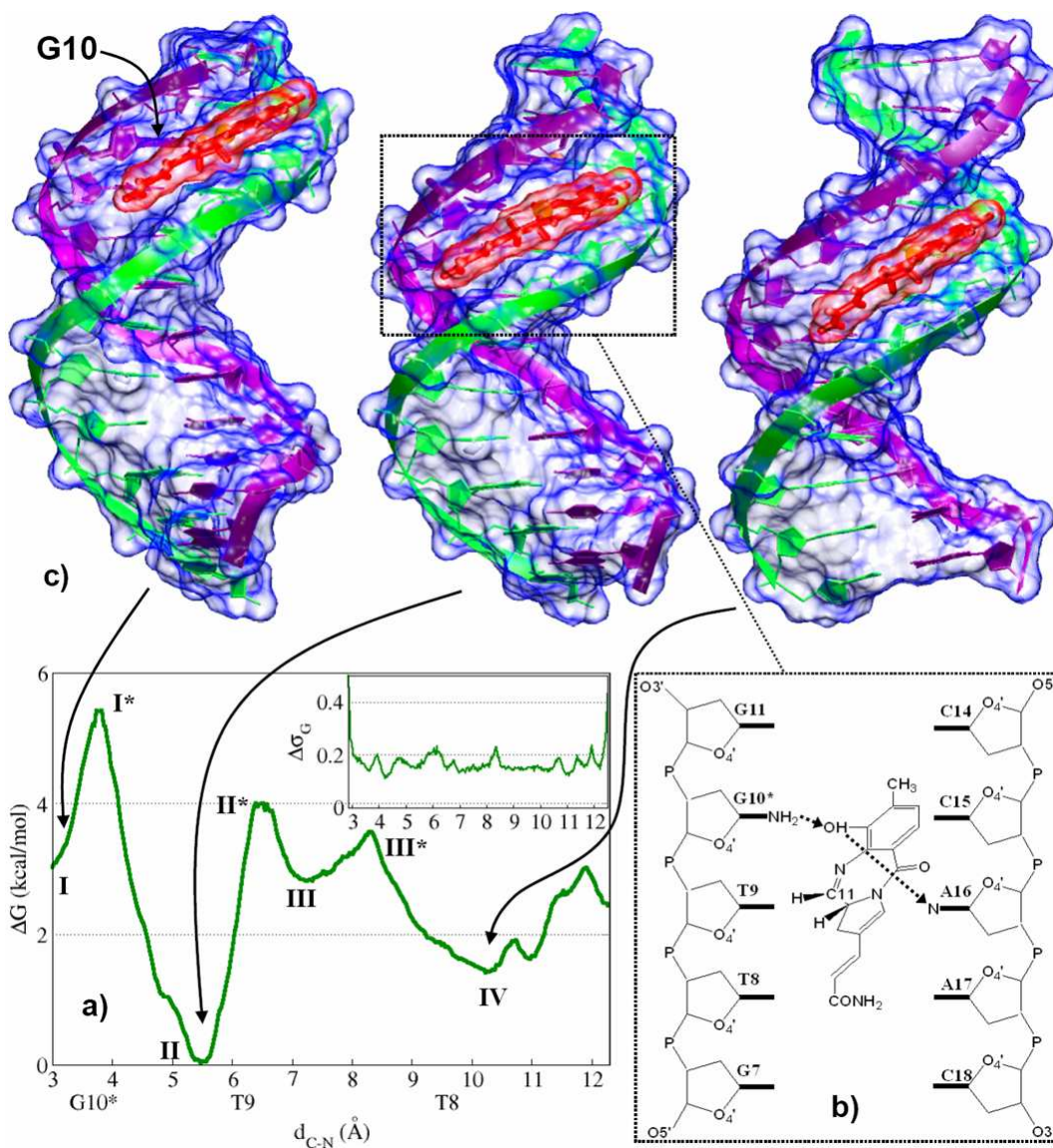
Figure 4.1: (a) Potential of Mean Force (PMF) associated to IMI sliding along the DNA minor groove. Standard deviation is reported in the inset. (b) Schematic view of the complex in the configuration corresponding to the free energy minimum. (c) Structures of three relevant conformations of the drug/DNA complex. The second and third structures (from the left) has been rotated respectively by $\sim30°$ and $\sim100°$ about the helical axis, to center the drug for visualization. Molecular surfaces of drug and DNA are depicted in transparent red and blue, respectively.
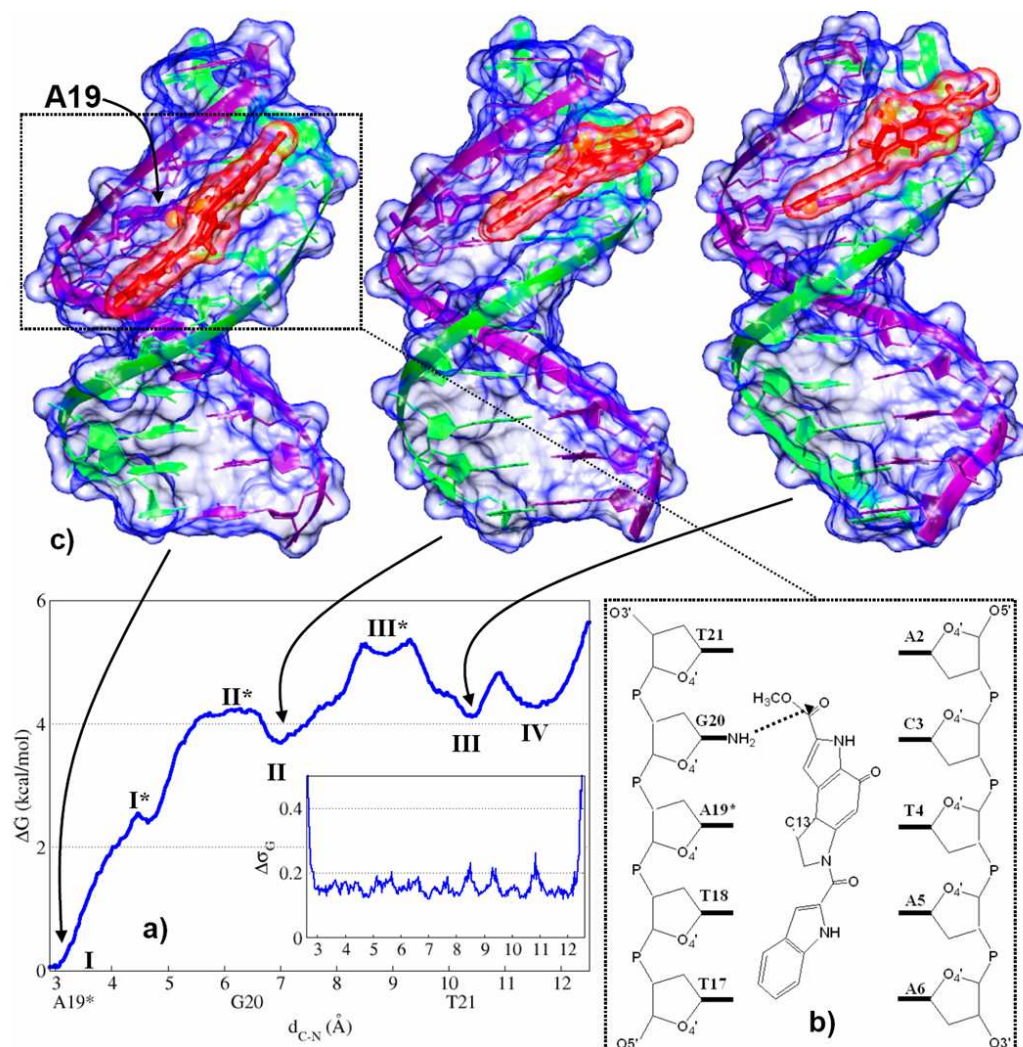
Figure 4.2: (a) PMF associated to DSI sliding along the DNA minor groove. Standard deviation is reported in the inset. (b) Schematic view of the configuration corresponding to the free energy minimum. (c) Structures of three relevant conformations of the complex. The third structure (from the left) has been rotated $\sim$40° about the helical axis to center the drug for visualization. Molecular surfaces of drug and DNA are depicted in transparent red and blue, respectively.

4.3.1 PMF profiles

The PMF of **IMI•DNA** features four minima (I-IV, Figure 4.1). In I, the complex is in its reactive configuration, i.e. carbon C11@IMI faces N2@G10 (d[C-N]$\approx$3 Å). As seen in the previous Chapter, I corresponds to the average structure assumed by

the complex during the first 10 ns of a 20 ns long MD simulation. From I, a small barrier of 1.5 kcal/mol has to be overcome at transition state I^* to reach the absolute minimum II, where d[C-N]≈5.6 Å. In II, C11@IMI sits slightly before the T9 plane (along the direction 3'→5'), and IMI sits at the same location as in the last 10 ns the dynamics mentioned above. The barrier from II to I (5.5 kcal/mol) is much larger than that from I to II, which again is consistent with results from Chapter 3. At larger d[C-N] distances other two minima are found: III (7.2 Å) and IV (10.1 Å), in which C11@IMI is located immediately beyond T9 and in front of O2@T8, respectively.

The PMF landscape of **DSI•DNA** is remarkably different (Figure 4.2). Indeed, sliding is hindered by a barrier of ∼4 kcal/mol which traps the drug in the reactive configuration I (corresponding to the absolute minimum in the PMF, along the investigated path). This minimum is followed by three other ones (higher in energy by ∼4 kcal/mol) virtually isoenergetic among them. In such minima, the reactive group of the drug (in particular C13@DSI) is in front of G20 (II, a very shallow minimum), T21 (III), and between T21 and C22 (IV).

The PMF of **IMI•DNA** appears to be much rougher than that of **DSI•DNA**; in particular it features two configurations corresponding to a larger gain in binding free energy than in the reactive conformation. Results obtained for the two adducts are in line with those extracted from unbiased MD simulations [84, 83], and help to explain the higher selectivity of duocarmycins as compared to anthramycins [266, 217]. For instance, in **IMI•DNA** the reactive configuration should be markedly less populated with respect to II (and competitive with III and IV). In contrast, in **DSI•DNA** there is a net gain in the binding free energy at the reactive configuration, the PMF being almost flat elsewhere.

4.3.2   Dissection of the PMF's

The PMF can be interpreted as the change in binding free energy ΔG_b upon drug sliding. To identify the relevant contributions to this process, such a changes can be expressed as a sum of individual terms [263, 37, 289, 290]. Here ΔG_b is decomposed in a term due to solute conformation and interactions between the two moieties (solute terms), and a second one due to the presence of water and counter ions (solvent effects):

$$\Delta G_b = \Delta G_{solute} + \Delta G_{solv} \tag{4.1}$$

where:

$$\Delta G_{solute} = \Delta H_{adapt} + \Delta H_{el} + \Delta H_{vdW} - T\Delta S_{vib} - T\Delta S_{r+t} \tag{4.2}$$

$$\Delta G_{solv} = \Delta G_{solv,p} + \Delta G_{solv,np} \tag{4.3}$$

In eq. 4.2, ΔH_{adapt} represents the enthalpic term due to DNA and drug structural deformations upon binding, while ΔH_{el} and ΔH_{vdW} are contributions from elec-

	d[C-N](Å)	ΔH_{adapt}	ΔH_{el}	# Contacts	# H-bonds	$\Delta G_{solv,p}$	$\Delta G_{solv,np}$
IMI•DNA	3.2 (I)	0.0(22.7)	0.0(1.6)	15(4)	2.0(0.8)	0.0(34.2)	0.0(2.0)
	3.9 (I*)	+9.6(22.3)	+3.3(1.3)	12(3)	0.9(0.6)	-38.7(50.2)	-2.1(2.4)
	5.6 (II)	-7.0(21.6)	-2.4(1.1)	19(3)	1.2(0.6)	-8.8(47.3)	-4.6(1.8)
	6.5 (II*)	-3.8(21.6)	-1.0(1.8)	13(3)	0.6(0.5)	-17.5(35.4)	+2.0(2.2)
	7.2 (III)	-4.9(22.2)	+1.2(1.1)	11(3)	1.0(0.2)	+11.4(36.7)	+4.2(2.4)
	8.2 (III*)	-1.4(21.8)	+2.2(1.3)	9(2)	0.5(0.3)	-16.4(31.6)	+4.8(1.7)
	10.1 (IV)	-1.5(22.4)	-1.4(1.0)	15(3)	0.5(0.5)	-1.5(53.2)	-2.1(3.0)
DSI•DNA	3.0 (I)	0.0(18.1)	0.0(1.9)	63(10)	0.5(0.3)	0.0(36.0)	0.0(2.0)
	4.3 (I*)	+18.3(18.2)	+6.3(2.0)	47(9)	0.0(0.0)	-60.2(45.6)	+7.4(2.6)
	6.0 (II*)	+18.6(18.1)	+3.6(1.7)	63(11)	0.0(0.0)	-41.3(29.6)	+0.3(2.8)
	7.0 (II)	+17.6(18.2)	+2.2(5.2)	53(12)	0.8(0.4)	-25.6(35.0)	+1.3(2.4)
	8.9 (III*)	+19.9(18.5)	+3.7(2.4)	56(11)	0.0(0.0)	-34.5(40.0)	+1.4(2.5)
	10.3 (III)	+15.7(18.1)	-1.2(2.2)	58(10)	0.9(0.6)	-44.4(41.4)	+3.9(2.0)
	11.9 (IV)	+14.1(18.5)	+5.3(1.5)	54(11)	0.0(0.0)	-49.5(39.1)	+2.9(2.0)

Table 4.1: Selected values of solute ΔH_{adapt} (column II), drug/DNA ΔH_{el} (c. III), number of hydrophobic contacts (c. IV), average number of H-bonds (c. V), polar (c. VI) and non polar (c. VII) contributions to ΔG_{solv}. Energies are in kcal/mol and referred to configuration I. Standard deviations are reported in brackets.

trostatic and Van der Waals interactions, respectively. ΔH_{adapt} and ΔH_{el} are estimated through the terms in the AMBER force field, while ΔH_{vdW} is assumed to be roughly proportional to the variation in the number of hydrophobic contacts. $T\Delta S_{vib}$ and $T\Delta S_{r+t}$ are the free energy contributions from vibrational and translational+rotational entropy changes upon binding. The former was evaluated through a normal mode analysis [279], while $T\Delta S_{r+t}$ was assumed to be sequence independent [271] and was thus not calculated. In eq. 4.3. $\Delta G_{solv,p}$ and $\Delta G_{solv,np}$ are respectively the polar (electrostatic) and non-polar (hydrophobic) contributions to solvation, evaluated here with the MM-PBSA method. It is worth to point out that whilst the PMFs are quantitatively reliable, their dissection into enthalpic, entropic, and solvation terms has been carried out using approximated and/or strongly sampling-dependent methods. Thus, these contributions have to be considered qualitatively, and are here used to gain insights into the sources of drug selectivity. In the following, all the values refer to the reactive configuration of the respective complex, which is indicated as I.

Solute terms

The analysis of enthalpic terms in **IMI•DNA** is summarized in Tab. 4.1 and in Figures 4.3a, 4.3b, 4.3c for ΔH_{adapt}, ΔH_{el}, and number of contacts, respectively. Consistently with previous unbiased dynamics [84], at I the DNA is distorted in the central

C6G7 tract with respect to its structure in bulk solvent. The adaptation energy further increases going from I to the maximum I*, while drug/DNA interactions weaken. Both the present and the previous works [84, 83] indicate that the whole structure relaxes at II (indeed, the conformation assumed by the oligonucleotide is very close to that of canonical B-DNA in aqueous solution), whilst electrostatic and Van der Waals interactions strengthen. Actually, such interactions are the strongest along the investigated path, while the adaptation expense is the lowest. In contrast, the (average) number of *direct* H-bonds decreases. The subsequent increase in free energy at III* is partially due to the steric clash between the drug acrylamide tail and the amino group of G7. The latter hinders the sliding of the drug, causing a rotation of its principal axis with respect to the oligonucleotide, and slightly exposing the molecule to the solvent (4.4).Correspondingly, the number of contacts reaches a minimum in III*, and the minor groove widens in the tract C6...T8 (average width 8 Å with respect to 4 Å at II). Once the guanine amino group has been overcome by the drug, a wide free energy minimum is reached in proximity of T8 (IV). Here, the drug recovers some of the contacts with the DNA backbone. The (solute) enthalpic contributions fairly correlate with the PMF (Figures 4.3a, 4.3b, 4.3c), albeit deviations are present, pointing to the role of solute entropic and/or solvation terms. This is particularly evident at PMF transition states: for example, the sum of the enthalpic terms indicates II* to be lower than III*, whereas in the PMF III* is slightly lower than II* (Tab. 4.1). Concerning solute vibrational entropies, the calculated values at minima of the PMF turn out to differ scarcely,consistently with results of the previous Chapter. Despite the limitations of the methodology used here [39], these values indicate that solute entropy is likely to not vary significantly at different configurations.

In **DSI•DNA**, the free energy absolute minimum I, corresponding to the reactive configuration, shows the lowest ΔH_{adapt} (Figure 4.3e and Tab. 4.1) and the largest number of hydrophobic contacts (Figure 4.3g and Tab. 4.1), while ΔH_{el} is very similar to the values in II and III (Figure 4.3e and Tab. 4.1). The major source of stabilization in I is the extended pattern of hydrophobic contacts formed between the drug and the DNA, in agreement with experimental suggestions [107, 266]. The drug is significantly distorted when leaving the reactive configuration: the twist between the two drug moieties increases from ∼20° at I to ∼110° at I*, causing a significant widening of the minor groove (4.4). Consequently, the block A, along with the minor groove floor (see also next section and Figure 4.3), is more solvent exposed, the electrostatic interactions weaken and the number of drug/DNA hydrophobic contacts decreases drastically. Additionally, the H-bond which forms discontinuously at the reactive configuration is here definitively lost. At II*, both the DNA and the drug are less distorted than in I*, and the number of hydrophobic contacts is the same as in I. Surprisingly, the adaptation cost is essentially the same as in I*. II is characterized by a small gain in electrostatic and adaptation energies as compared to II* and III*, along with the recover of one H-bond. However, it features less hydrophobic contacts

than II* and III*. Apart from the reactive configuration, III is the lowest enthalpic minimum; however, its free energy value is almost identical to those at II and IV (Figure 4.2). In summary, the (solute) enthalpic analysis shows that several points lack correspondence with the PMF, indicating a key role of entropy and/or solvent effects for the sliding process of DSI. Nevertheless, enthalpy contributions pinpoint unambiguously the configuration I as the most stable one, with ΔH_{adapt} playing a more important role than in **IMI•DNA** (max $\Delta \Delta H_{adapt} = 18$ and 7 kcal/mol for DSI and IMI, respectively).

Finally, $T\Delta S_{vib}$ was evaluated at the minima of the PMF. The calculated values show a range of variation comparable to that found for **IMI•DNA**: thus, also in this case solute entropic contributions to the sliding process are neglected.

Solvent effects

It is well known that solvent is crucial for nucleic acid structure and stability, and consequently it is important for binding of drugs to DNA. Indeed, the analysis of ΔG_{solute} clearly points to the role of solvent for the sliding of the two investigated drugs, in particular for DSI. The development of reliable implicit solvent models and fast finite-difference numerical procedures [285, 286, 287, 288] has made feasible the calculations of solvation free energy *via* MD simulations. In particular, binding free energies of ligands to DNA calculated with the MM-PBSA method agree fairly well with experimental values [75, 37, 282]. Electrostatic $\Delta G_{solv, p}$ and hydrophobic $\Delta G_{solv, np}$ contributions to ΔG_{solv}, that can be interpreted as the free energy cost for drug desolvation upon binding, are evaluated using either the "standard" setup [39, 286] and the alternative approach available in AMBER 9 [285, 286, 287, 288] (see Systems and Methods). It is worth to recall tha that binding is here always associated to an unfavorable – positive – change in solvation free energy. The profiles of ΔG_{solv} evaluated with the two methodologies are very similar, albeit the second approach highlights the sequence dependence of non-polar contributions and is thus reported here (Figures 4.3d, 4.3h)

The changes in ΔG_{solv} appears very large as compared to the PMF; large standard deviations are seen, as already reported by other authors [37, 282]. Nevertheless, some interesting differences between IMI and DSI are observed: *i)* The cost in desolvation upon formation of the complex, ΔG_{solv}, is the highest at the absolute PMF minimum in **DSI•DNA** (Figure 4.3h), in agreement with the better fit of the drug into the minor groove at I. Already at I*, where drug distortion results in its enhanced solvation, ΔG_{solv} suddenly decreases, and reaches a rough plateau for the other configurations. This behavior mirrors the PMF landscape, where the minimum corresponding to the reactive configuration is the lowest; *ii)* In contrast, ΔG_{solv} does not show any clear correlation with the PMF in **IMI•DNA**. In particular, the deepest minimum in the PMF is not associated with the highest value of ΔG_{solv} (highest sol-
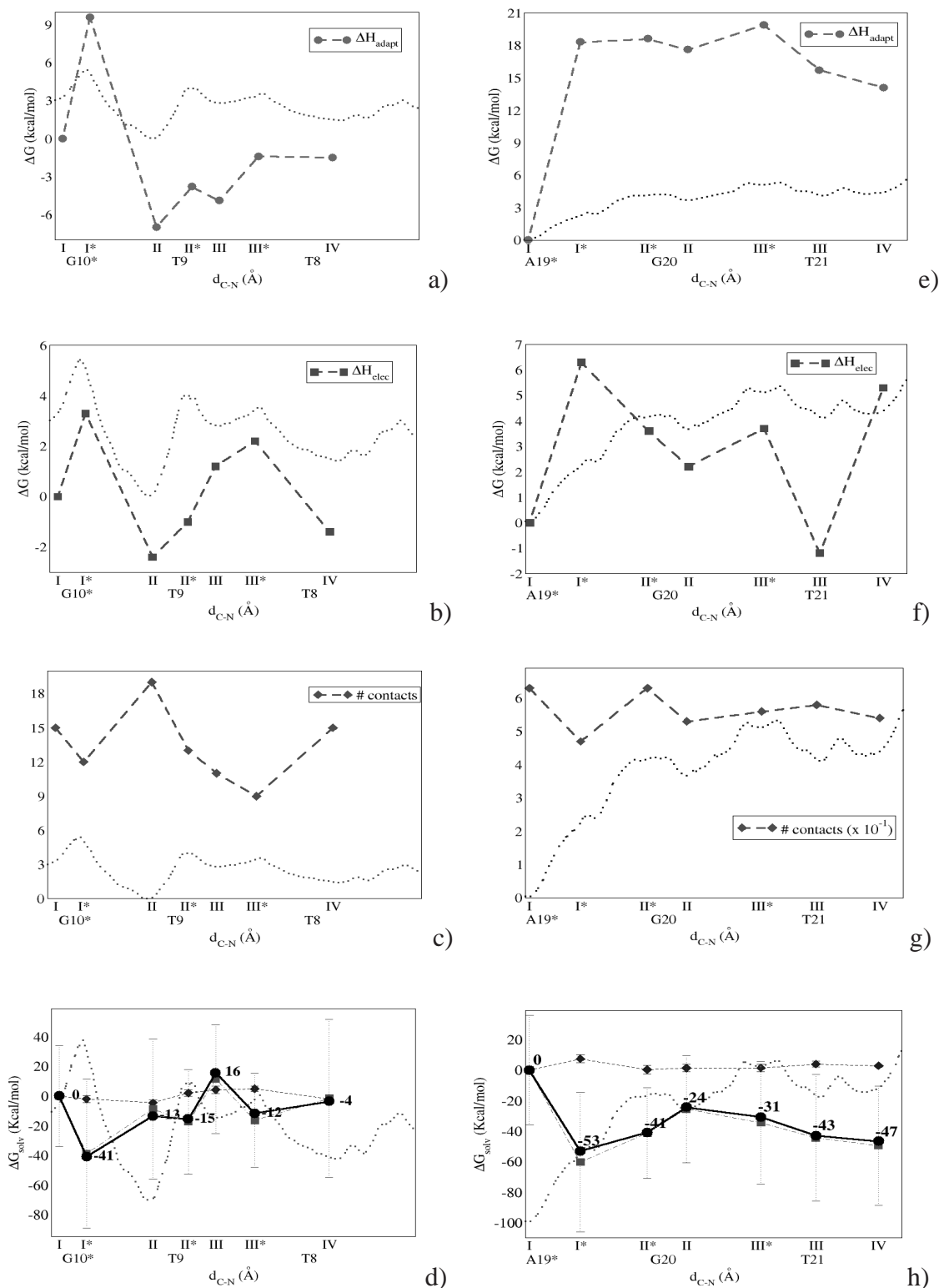
Figure 4.3: Selected values of: a,e) ΔH_{adapt}; b,f) ΔH_{elec}; c,g) number of hydrophobic contacts; d,h) electrostatic (squares, dotted-dashed line), hydrophobic (rhombus, dashed line), and total (circles, solid line) ΔG_{solv} along the reaction coordinate d[C-N], compared to the PMF (dotted line, and rescaled in c), d), g), h) to allow for an easier comparison with ΔH_{solv} profiles). Data on the left (a-d) refer to **IMI•DNA**, those on the right (e-h) to **DSI•DNA**. Solvation free energies have been obtained using the method of Luo and co-workers.
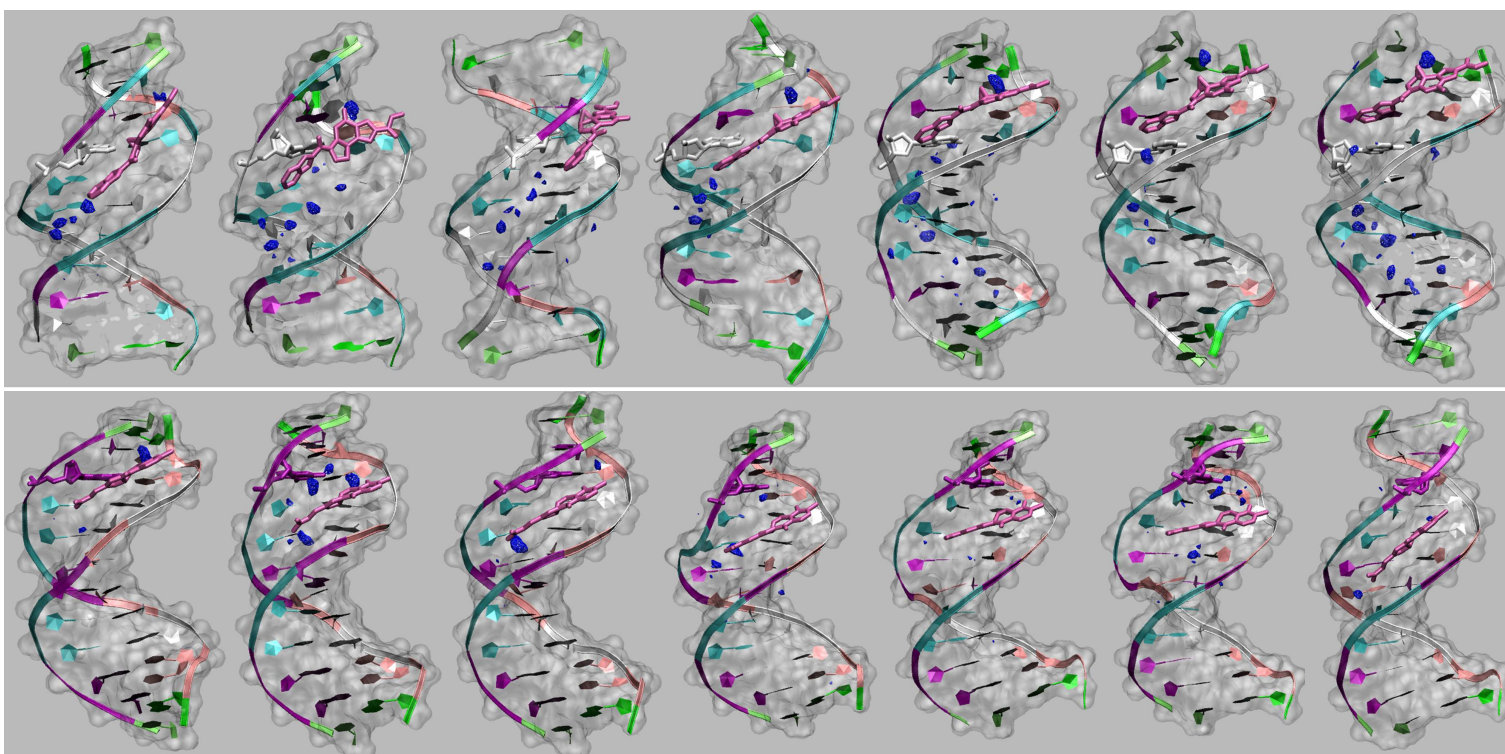
Figure 4.4: Water density distributions extracted from the last 2 ns of dynamics at relevant points of the PMF in **IMI•DNA** (upper panel) and **DSI•DNA** (lower panel). Densities are calculated with the $g_{spatial}$ utility of GROMACS, using a 3D grid with step 0.5 Å. Water densities at values ~ 3 times larger than that of bulk water are shown as blue wireframes.

vation cost). Density plots of water (Fig 4.4) exhibit qualitatively the same trend of the profile of ΔG_{solv}. Both **IMI•DNA** and **DSI•DNA** I* feature more waters inside the groove with respect to I, and in **IMI•DNA** III has the largest desolvation cost, consistently with its very poor hydration. Interestingly, the number of hydrophobic contacts in III is lower than that of II, indicating that this number does not necessarily increase with a lower degree of hydration. This might be due, at least in part, to the "bridging" of the two DNA strands by the drug, which prevents water to access the groove around the binding region.

Within the limitations of this analysis, it is suggested that solvent effects play a key role in **DSI•DNA**, consistently with the finding that ΔG_{solute} does not correlate with the PMF. Instead, no clear correlation exists between the PMF and ΔG_{solv} in **IMI•DNA**, which is also consistent with the rough correlation found between ΔG_{solute} and the PMF.

4.3.3   Evaluation of end-effects

All simulations reported here have been started with IMI and DSI positioned in front of their reactive sites [99, 266] (nucleobases G10 and A19 in **IMI•DNA** and **DSI•DNA** respectively).[2] When building-up the sliding windows as described in Systems and Methods, DSI turned out to slide towards the 3'-end of the strand containing A19, while in **IMI•DNA** the drug moves towards the oligonucleotide center. As a result, the regions explored by the two drugs are slightly different (although it should be noticed that they are comparable, as can be seen from Figures 1 and 2), and diverse influence of "end-effects" might hamper a thorough comparison between the calculated PMF's. In this respect, in the previous Chapter it has been already shown that end-effects play only a minor role on the dynamics of **IMI•DNA**. Here, additional MD calculations were performed to assess the possible influence of such effects on the interaction between DSI and DNA. Specifically, the features of the tract **DSI•d[GACT]**$_2$ either embedded in the 11-mer **DSI•DNA** (**DSI•d[G1A2C3T4]**$_2$, see Systems and Methods) or in the 14-mer **DSI•DNAc** (**DSI•d[G4A5C6T7]**$_2$ were compared along the last 2 ns of the respective MD simulations. The structure, the conformational flexibility, the hydration of **DSI•d[G1A2C3T4]**$_2$ and **DSI•d[G4A5C6T7]**$_2$ turned out to be rather similar, along with the interactions between DSI and the tracts d[GACT]$_2$. In fact:

1. The structures of **DSI•d[G1A2C3T4]**$_2$ and **DSI•d[G4A5C6T7]**$_2$ overlap very well (Figure 4.5): moreover the widths of minor groove at the binding region are comparable with or without the presence of three additional nucleobases

---

[2] Indicating such nucleobases with the asterisk the sequence are d[GCCAACGTTG*GC]-d[GCCAACGTTGGC] and d[GACTAATTGAC]-d[GTCAATTA*GTC] for IMI•DNA and DSI•DNA, respectively.
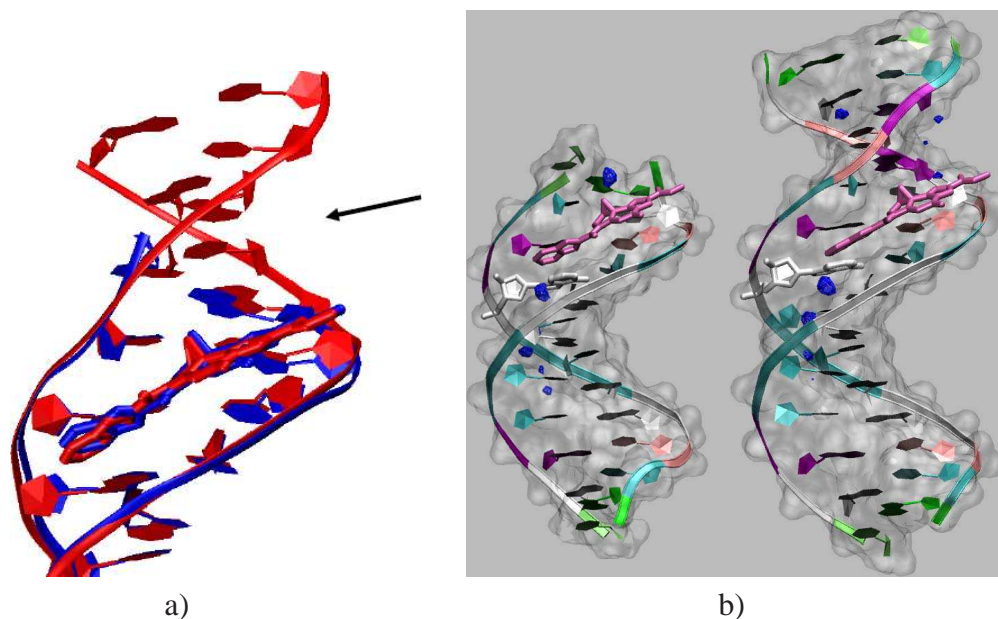
Figure 4.5: *a)* Superimposition between structures (averaged over the two last 2ns) of **DSI•DNA** (blue) and **DSI•DNAc** (red). The tracts d[GACT]$_2$ and DSI have a very similar arrangement in the two complexes. In particular, although the terminal base-pair in **DSI•DNA** is clearly distorted relative to **DSI•DNAc**, the position of the drug in the minor groove is essentially unchanged. The arrow indicates the opening of the minor groove in **DSI•DNAc** *b)* Water density distribution extracted from the last 2ns of dynamics for **DSI•DNA** (left) and **DSI•DNAc**. Densities are calculated with the g_spatial utility of GROMACS, using a discrete 0.5 Å step in a 3D grid. For each frame, the system has been rototranslated to obtain the best superimposition of DNA onto a reference structure; the density function is then calculated. Water densities at values 3 times larger than that of bulk water are shown as blue wireframes.

(data not shown). Consistently the calculated RMSD is very low, 0.6 Å. Furthermore, independently of the duplex length, DSI does not fit perfectly into the minor groove. As a consequence of these structural similarities, in both **DSI•DNA** and **DSI•DNAc** the methoxyl-carbonyl ester of the drug is fully solvated.

2. The number of hydrophobic contacts and H-bonds between DSI and d[GACT]$_2$ is almost the same (data not shown).

3. The DSI-O$_{wat}$ and DSI-H$_{wat}$ Radial Distribution Functions (RDFs), which provide information on the solvation of the drug, are also rather similar. Consistently, the first and second solvation shells around the drug contain virtually

the same number of waters. This is not unexpected due to the excellent super-imposition between the two tracts (Figure 4.5) and the comparable hydration of the DSI methoxyl-carbonyl ester. In particular, $O14_{DSI}$ (Chart I), which in **DSI•DNA** is the drug acceptor (pointing to the minor groove) closest to the duplex end, forms in both cases one H-bond with similar lifetime. Interestingly, also tracts $d[G1A2]_2$ and $d[G4A5]_2$ do not show relevant differences in their H-bond patterns. Finally, density plots of waters extracted from MD simulations are also rather similar (4.5).

4. The values of the Root Mean Square Fluctiations (RMSF) do not differ significantly, apart for the terminal sugar moieties, which are obviously more flexible in **DSI•DNA** (data not shown).

5. The electrostatic energy between DSI and tracts $d[GACT]_2$ is exactly the same in both systems.

Summarizing, end-effects seem to play a minor role on the interactions between the two drugs and the DNA duplexes, although they do slightly influence the structure of water within the minor groove. These findings confirm that steric hindrance and optimal interactions with the groove have a prominent role in determining the preferred binding sequence of duocarmycin.

4.4   Concluding Remarks

The PMFs associated to the sliding of IMI and DSI along the DNA minor groove differ remarkably The reactive configuration is the most favourable for **DSI•DNA**, whereas in **IMI•DNA** significant activation energy is required to reach the reactive site moving from the absolute PMF minimum (corresponding to a non-reactive configuration). Results are consistent with previous MD simulations of **IMI•DNA** [84], showing that the reactive configuration becomes unstable after few ns, and with those of **DSI•DNA** [83], in which the drug is stable in the reactive configuration for the whole dynamics. Moreover, there is an indication that the higher specificity of DSI compared to IMI correlates with a higher cost for moving the drug from the preferred site, at least for the sequences considered here. End-effects turn out not to play a relevant role for the binding.

For both complexes, configurations associated with the absolute minima of the PMF are those with the smallest adaptation cost and the better packing. This is consistent with the usual assumption of negligible ΔH_{adapt} upon binding of the drug in the minor groove to the preferred sequence [263, 37, 271]. Apart from this common feature, various contributions to the PMF are found to have different relative weights in the two adducts. In **IMI•DNA** the changes in enthalpic terms correlate fairly well

with the PMF, particularly at the four minima, while the ΔG_{solv} landscape is apparently unrelated to the free energy profile. In **DSI•DNA** a rough anti-correlation between ΔG_{solv} and the PMF profile is found instead, whereas enthalpy contributions do not have an overall correspondence with the free energy profile, except for the coincidence of absolute minima.

In summary, this work suggest that, also for alkylating agents, shape complementarity and packing are significant factors in determining the preferred site of non-covalent binding [37, 289, 290, 291]. In addition, insights are gained on the way differences in chemical structure, size and flexibility may influence solvation and molecular recognition along the minor groove. DSI, due to its larger size relative to IMI, covers more DNA base pairs and needs to rotate around the linkage between groups A and B to fit in the preferred sequence. The cost of such a rotation (which has been supposed to be a key factor for reactivity) depends on flexibility and structure of the target DNA sequence. Furthermore, the lack of optimal docking results in a larger drug exposure to the solvent, at opposite with findings for **IMI•DNA**, in which drug desolvation cost is uncorrelated to sequence selectivity. This is consistent with the fact that IMI has a natural right-handed conformation allowing for a snug fit into the minor groove at different sequences (indeed the reactive configuration does not correspond to a minimum in ΔH_{adapt}), and features a lower sequence selectivity as compared to DSI.

Finally, although speculative, different molecular recognition mechanisms can be proposed for the two drugs investigated here. DSI might firstly bind to DNA at those sequences characterized by the lowest desolvation cost (the non-preferred ones), and then easily slide towards the preferred site, corresponding to a funnel in the PMF. In this scenario, solvent effects could be critical for the molecular recognition of DNA by duocarmycins. In contrast, recognition of DNA by anthramycin appears to be more complicated, since no preferred route exists exist according to results reported here. Moreover, the reactive site can be reached only upon crossing a significant free energy barrier.

Clearly, calculations of free energy barriers to form non-covalent complexes at different sequences, also with drugs other than those considered here, are needed to achieve a general picture of the molecular recognition. However, the findings reported here help to rationalize the higher selectivity of duocarmycins as compared to anthramycins, and point out clearly the possibility of multiple patterns for molecular recognition.

Chapter 5

Unbinding of distamycin and anthramycin from DNA investigated by metadynamics

The mode of binding of minor groove (MG) binders to their target DNA sequences has been well characterized at an atomic-detail level using both experimental biophysical techniques as well as molecular simulations [292,289,76,265,37,271,15,293,19,263, 84,83,261,262]. In contrast the structural, thermodynamic and kinetic features of the molecular recognition event are mostly lacking [294, 295, 296, 297, 269, 292, 298], despite their key role in rational drug design [264, 9]. Recognition goes beyond the formation of the final adduct: both non covalent [249, 248, 268, 269] and co-valent [296, 299] binders exploit mechanisms like shuffling sliding within the MG and translocation among different sites until a preferred sequence is reached. In the last years the development of novel methods to accelerate rare events [180], along with the increased reliability of force fields [24, 25, 31, 300, 27, 30] and simulation protocols [137, 275], allowed to investigate non-equilibrium dynamics of complex biological systems [301, 302, 303]. In particular, the recently introduced metadynam-ics method [189] has been applied to investigate mechanism and thermodynamics of biologically relevant processes [304, 305, 306], providing insights on aspects of the molecular recognition events of proteins by small ligands [307, 193].

Here we have used such an approach to investigate the unbinding from DNA of two ligands, representative of the main classes of non covalent and covalent MG binders. We focus on the unbinding process because of its importance in selectivity and molecular recognition [297, 269, 308] and because it can provide insights on the mechanism ligands interfere with protein/DNA interactions and dynamics [298]. The prototype of non-covalent binders is distamycin A [108] (hereafter DST, Chart I). DST is highly selective towards AT-rich tracts [111, 109], its amide nitrogen forming H-bonds with N3(A) and/or O2(T) of DNA [9, 73, 309, 310]. The alkylating agent

is the already studies (see previous Chapters) anthramycin [99, 98], here considered in its most reactive anhydrous form [84, 253] (hereafter IMI, Chart I). In contrast to DST, anthramycin shows only a modest preference towards PuG*Pu triplets (the asterisk indicating the alkylated guanine) [221]. Both these antibiotics have shown to exert an antitumoral action by interfering with transcription [15, 9, 311, 312, 313] and replication [98, 314, 315] processes, and have been the prototypes for anticancer drugs which are currently undergoing clinical trials [78, 80, 316, 317]. For both, experimental structural information is available of the adducts with DNA[1] [99, 109, 111]. In addition, for DST experimental information on the kinetics of the binding and unbinding events is available in literature [292]; this enables a test on accuracy of the method used.

From our metadynamics simulation we extract the Lowest Free Energy Path (LFEP) [210] associated to the unbinding, which is equivalent to an Intrinsic Reaction Pathway (IRC) at a finite temperature and gives the minimum-cost path of detachment, making it possible to identify relevant conformations (and associated free energies) of the process. Our simulations reveal a different mechanism of IMI and DST unbinding. In the first case only metastable states are observed between the bound and unbound configurations, and the overall activation free energy is ∼12 kcal/mol. The unbinding occurs through a loss of contacts between the ligand and one DNA strand, followed by the detachment of the drug head (ring A in Chart I), which leads to the complete unbinding. In contrast the detachment of DST passes through an intermediate which, along the LFEP, is reached after crossing a barrier of ∼18 kcal/mol and it is characterized by the full hydration of the ligand hydrophobic moiety, with the charged tail of the drug anchored into the MG. Importantly, from this complex the rearrangement towards the bound and unbound configurations are equally likely. The reliability of our computational scheme is assessed by the finding of virtually the same mechanism of detachment with different choices of the CVs and of the metadynamics parameters. The activation free energies calculated in different simulations are also within ∼1 kcal/mol in the case of IMI and ∼4 kcal/mol in the case of DST. Furthermore, a good agreement has been found between calculated and experimentally determined [292] activation free energies of DST.

Our results are consistent with experimental findings on the mechanisms of unbinding of the Hoechst dye from DNA [269], helping at rationalize results on the kinetic features of these processes at a molecular level.[2] In addition, the characterization of the unbinding proposed here could be useful to achieve a deeper understanding of the dynamics between binary ligand/DNA complexes and proteins involved in vital cellular processes (e.g. topoisomerases [15, 9, 311, 312, 313, 319], or heli-

---

[1]DST has been found to bind to DNA as a side-by-side dimer at high drug/DNA ratios [9, 318]. Here we focus exclusively on the monomeric binding mode as to make a direct comparison with the 1:1 IMI/DNA complex.

[2]Notice that no attempt has been made here to estimate binding free energies (See Methods).
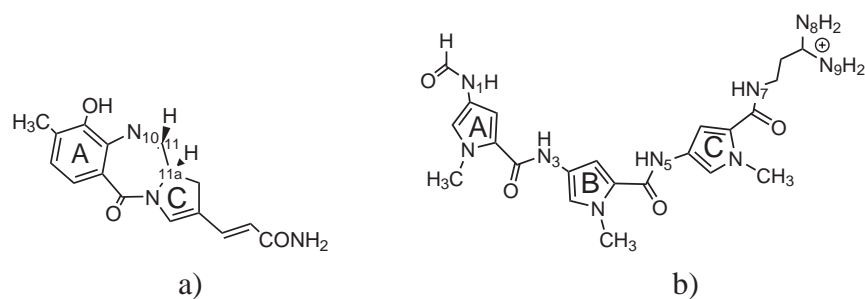
Chart I

cases [320, 320, 321, 322]).

5.1 Methods

Systems. The DST·d[5'-CGC**ATTA**CGC-3']$_2$ complex[3] was built by taking the ligand from the X-ray structure of the adduct with the duplex d[GGCC**AATT**GG]$_2$ (PDB CODE 1JTL [111]), while the 10-mer was built using the *nucgen* module of AMBER [231, 280]. The initial structure of IMI·d[5'-CAACG**TTGG**CCAAC-3']$_2$ was extracted from a previous MD simulation [84], where the adduct was built starting from the X-ray structure of the covalent complex between anthramycin and the 10-mer d[5'-CCAACGTTG*G-3']$_2$ (PDB code 274D [99]). Sodium counter-ions (18 in DST•DNA and 22 in IMI•DNA) were added to achieve system neutrality, and the structures were solvated in a monoclinic box with 10246 (DST•DNA) and 13611 (IMI•DNA) water molecules. Periodic boundary conditions were applied. The initial size of the box were $75\times80\times72$ Å^3 and $60\times84\times69$ Å^3 in IMI•DNA and DST•DNA respectively. With this choice the distance between the solute and its first image is larger than 25 Å during the whole unbinding process.

MD calculations. The recently refined AMBER force field [25,31] was used for the parameterization of the oligonucleotide moieties. The force field parameters for IMI were taken from Ref. 84; those of DST were calculated in a similar fashion. Structural parameters were taken from the *gaff* force field [234] while the charges were calculated following the RESP procedure [134]: the ligand was firstly optimized at the HF/6-31G(d) level up to a convergence in energy of 10^{-5} *au*. The presence of the solvent, which was required to avoid formation of unphysical intermolecular H-bonds, was simulated with the CPCM [323] implicit model using the Gaussian03 [273] package. The electrostatic potential map generated by the optimized structure was then

---

[3]Bold bases are those covered by the ligand, while the asterisk indicated the base forming the covalent bond.

calculated *in vacuo* at the same level of theory, and atomic RESP [134] charges were derived using the *antechamber* module of AMBER. Sodium and water were modeled with the AMBER-adapted Aqvist [233] and TIP3P [274] potentials, respectively. Electrostatic interactions were treated using the particle mesh Ewald [137] (PME) algorithm with a real space cutoff of $10\,\text{Å}$, the same as for van der Waals interactions. The pairlist was updated every 10 steps, and Lincs constraints [235] were applied to all bonds involving hydrogen atoms, permitting a time step of $2\,fs$. NPT conditions were achieved by coupling the systems with a Nosé-Hoover [121, 122] thermostat ($\tau = 1\,ps$) and the Andersen-Parrinello-Rahman [123, 126] barostat ($\tau = 1\,ps$).

Systems were equilibrated using the following computational procedure: *i)* structural optimization of the solvent, using the conjugate gradient algorithm up to a convergence of $10^{-4}\,\text{kcal/mol}$; *ii)* energy minimization of the entire system, using the same convergence criteria as in *i)*; *iii)* linear heating of the system up to $300\,\text{K}$ (in $100\,ps$) while putting harmonic restraints on the solute (k=5 kcal/Å^2); *iv)* $5\,ns$ of constant temperature-pressure (T=300 K, P=1 atm). Starting structures for metadynamics were taken from the last *ns* of MD, selecting those with the lowest RMSD with respect to the average structure of the solute (calculated also in this last *ns*). DNA MG width was defined as the distance between sugar C4' atoms, and it was calculated with the program Curves [276, 239]. All simulations have been carried out using a modified version of the GROMACS [272] package implemented with metadynamics [305].

Metadynamics. This recently introduced method [189] allows to compute free energies and to accelerate rare events. It is based on a dimensional reduction: a set of Collective Variables (CVs) s_i ($i = 1, \ldots, \text{N}_{\text{CV}}$), function of the system coordinates $\mathbf{x}$, is evolved with a standard MD supplemented by a history-dependent potential which prevents the system from visiting conformations already sampled. Namely, the history-dependent potential is built-up by Gaussians of dimensionality N_{CV}, height w and widths δs_i ($i = 1, \ldots, \text{N}_{\text{CV}}$), deposed at time intervals τ_G along the CVs trajectory. In the limit of a long metadynamics run, the sum of the penalty terms tends to compensate exactly the underlying FES in the reduced space permitting a reconstruction of this FES explored up to time t [209]. To investigate the unbinding mechanism of IMI and DST the following CVs were chosen:

1. The distance d_{CMs} between the centers of mass of the ligand and of the DNA tracts d[**G**TT**GG**]$_2$ and d[**CATTA**C]$_2$ (for IMI and DST, respectively). A similar choice of CV has been applied in Ref. 193. A reflective wall has been applied at a distance of 20 (IMI) and 22 (DST) Å in order to avoid the sampling of conformations where the ligand is fully solvated.

2. The number of hydrophobic contacts n_{hph} between non polar carbons of the ligand and the bases it covers in the starting structure, modeled as a coordination

number:

$$n_{hph} = \sum_{ij} \frac{1 - (r_{ij}/r_0)^a}{1 - (r_{ij}/r_0)^b} \qquad (5.1)$$

The parameters a and b have values of 6 and 12, respectively, while r_0=6 Å accounts for the typical carbon-carbon distance (4/4.5 Å) and the thermal motions' amplitude (1.5/2 Å) [307]. A similar CV has been used in Refs. [307, 324].

3. The number n_{hb} of H-bonds between the ligand and the bases it covers in the starting structure, described also by Eq. 5.1 and values $a = 8$ and $b = 12$, and $r_0 = 2.5$ Å. A similar CV has been chosen in Refs. 304, 305.

Gaussian parameters were $w = 0.1$ Kj/mol, $\delta s_{CMs} = 0.5$ Å, $\delta s_{hph} = 8$ and $\delta s_{hb} = 1$ in the case of IMI, $w = 0.3$ Kj/mol, $\delta s_{CMs} 0.5$ Å, $\delta s_{hph} = 6$ and $\delta s_{hb} = 3$ for DST. Gaussian widths were selected so as to be almost half of the thermal fluctuations. The time interval between two successive Gaussian depositions is $0.5\,ps$ in all the simulations. *Ad hoc* CVs were further built to avoid "walking" of the ligand in the solvent after unbinding, forcing the re-binding at the same site of detachment: the angles N1$_{G5}$-N3$_{T7}$-C1$_{IMI}$, C2$_{G9}$-N1$_{A22}$-C1$_{IMI}$ in IMI•DNA and C2$_{A4}$-O2$_{T6}$-N5$_{DST}$, N3$_{G8}$-O2$_{T6}$-N5$_{DST}$ in DST•DNA. Such CVs were kept inactive (i.e. Gaussian width was set to zero), and only a wall was applied to force $|\cos\alpha| \leq 0.3$, where α is the angle selected as CVs [193].

Free energies surfaces were calculated as a function of (d_{CMs}, n_{hph}) and (d_{CMs}, n_{hb}). In addition, simulations were performed in which the three CVs were kept active. In every metadynamics run the unbinding event was seen after a few ns (see Results). Furthermore, the unbinding mechanism of each ligand turned out to be very similar in all simulations, which means that the relevant slow motions of the systems are captured by our CVs. However, no attempt was made to estimate binding free energies, due to the generality of the CVs itself. Indeed, when the ligand exits the groove and is fully solvated, it assumes different orientations relatively to the duplex before a re-binding event occurs. As a consequence, an entropic hole is scavenged in the FES by all those unbound conformations corresponding to very close points in the CVs space (for example different head to tail orientations of the drug gives very similar number of hydrophobic contacts). For this reason, we only report here the activation free energies, which can be extracted reliably by stopping the summation over Gaussians just after the complete detachment of the drug. For the sake of clarity, we show here only the free energy profiles as a function of d_{CMs} and n_{hph} in the case of IMI•DNA and as a function of d_{CMs} and n_{hb} in the case of DST•DNA. On these surfaces the LFEPs, which allows a precise identification of minima, intermediates and transition states, were calculated by adapting a program developed by B. Ensing [210].

Cluster analysis. A cluster analysis [325] has been performed on the ensemble of conformations corresponding to relevant minima and transition states of LFEP. In each state a dominant cluster has been found which is larger in size at least one order of magnitude than the others. Representatives of each principal cluster are shown in Figures 5.1 and 5.2. A dissection of free energy differences into various energetic contributions was performed through averaging over each principal cluster.

5.2   Results

In this section we analyse the mechanism of unbinding of IMI and DST from their respective DNA duplexes. For the sake of simplicity, we report here free energy surfaces $G(d_{CMs}, n_{hph})$ for IMI•DNA and $G(d_{CMs}, n_{hb})$ for DST•DNA, along with associated Lowest Free Energy Paths (LFEPs) and relevant conformations of the complexes (Figures 5.1 and 5.2). The CVs d_{CMs}, n_{hph} and n_{hb} are respectively the distance between the centers of mass of the ligand and of the DNA tracts it covers, the number of hydrophobic contacts and the number of hydrogen bonds (modeled through a coordination number, see Methods). However, virtually the *same* mechanism has been found using other collective variables (see methods), although the free energy may vary up to 4 kcal/mol. Let us start describing the unbinding process of the alkylating agent IMI to DNA.

5.2.1   IMI

The initial structural model has been taken from the final configuration of a previous MD simulation [84] of the adduct IMId[5'-CAACGTTG*GCCAAC-3']$_2$. In this configuration the binding site of IMI, located in d[5'-TTG*-3']$_2$, does not coincide with its alkylation site d[5'-TG*G-3']$_2$. The free energy surface $G(d_{CMs}, n_{hph})$ corresponding to the unbinding of IMI from the duplex, along with associated LFEP and relevant conformations of the complex, are shown in Figure 5.1). There are two virtually isothermic global minima, which are separated by a small barrier of ∼3 kcal/mol (Table 5.1). In the first (Min1, the absolute one), located at $d_{CMs} = 0.54$, $n_{hph} = 188$, the drug is bound at the tract d[5'-TTG*-3']$_2$ and it is stabilized by hydrophobic contacts with both DNA strands and by electrostatic interactions (Table 5.2); on average, one H-bond is formed, involving bases G7, A21 and the hydroxyl group and N10 of IMI, in agreement with the results from the unbiased dynamics performed in Ref. 84.

The structure of the complex in this minimum also corresponds to that of Ref. 84 after sliding of IMI from its alkylation site. Min2 differs only for the local deformation of the DNA duplex in the IMI binding region, not for the relative position of the drug into the MG (compare Figure 5.3a with Figure 5.5). Correspondingly, these two minima feature very similar drug/DNA interaction strength and hydration properties
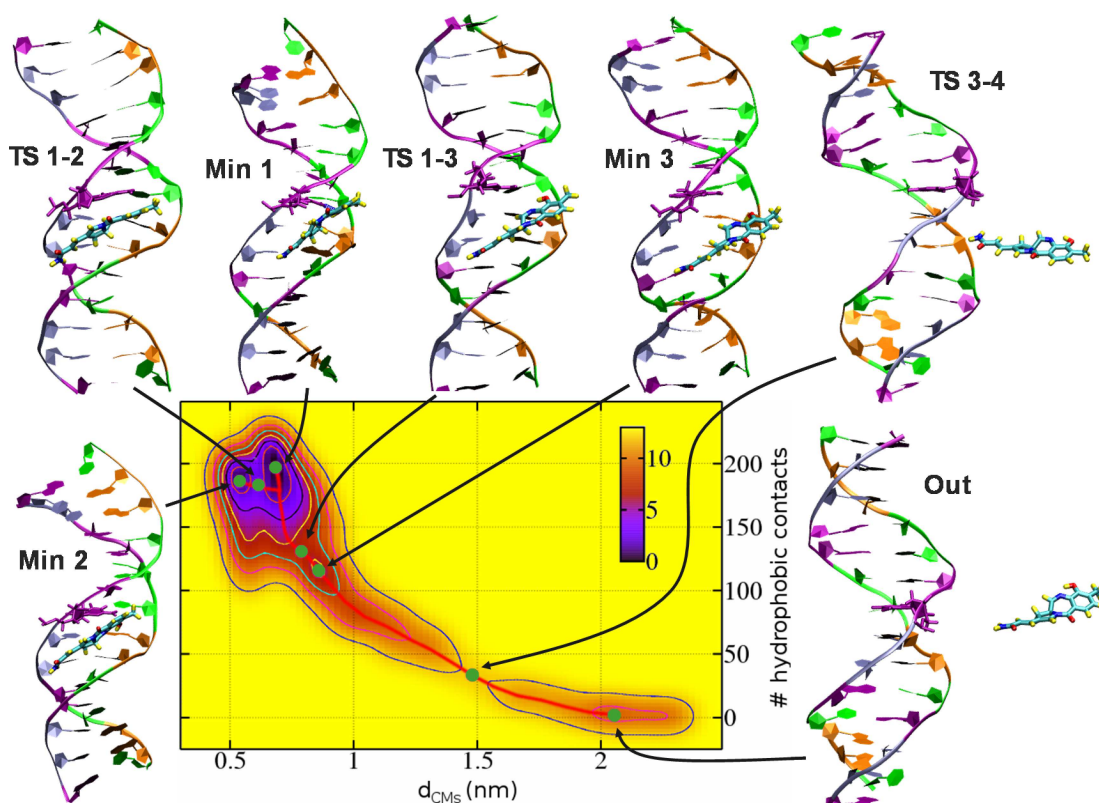
Figure 5.1: Free energy surface $G(d_{CMs}, n_{hph})$ associated to the detachment of IMI from DNA. A Lowest Free Energy Path (LFEP) is also traced connecting relevant minima and transition states (whose representative structures are shown). Isosurfaces are drawn one per 2 kcal/mol.

	Min1	**TS1-2**	**TS1-3**	**Min2**	**Min3**	**TS3-4**
d_{CMs}, n_{hph}	0.54, 188	0.61, 186	0.69, 133	0.68, 199	0.86, 117	1.47, 35
ΔG	0.0	3.0	6.7	0.7	5.7	12.3

Table 5.1: Free energies (kcal/mol, Min1 taken as reference) associated to relevant configurations in the unbinding of IMI from DNA. These define the Lowest Free Energy Path (LFEP).

(Tables 5.2, 5.3). A simple visual inspection of Figure 5.1 reveals that is very unlikely that Min2 could be directly involved in the unbinding: from it the system will evolve most likely towards Min1. The two minima are separated by a transition state TS1-2 of about 3 kcal/mol. Figure 5.3a shows that in this conformation the MG widens by ~ 1 Å in the tract d[5'-TG*G-3']$_2$, and the whole width profile virtually overlaps to

that in Min2, suggesting that this opening is related to the transition from Min1.

The detachment process begins at TS1-3, higher in free energy by $\sim$7 kcal/mol with respect to Min1. The region of the MG covered by the drug opens (Figures 5.1 and 5.3a) and a few water molecules enter the region between IMI and the strand containing the alkylating guanine (hereafter strand1, while strand2 refers to the opposite one, Figure 5.6). Consistently, the free energy of hydration ΔG_{hyd} (calculated from the Solvent Accessible Surface Area [326] - SASA - relatively to the unbound complex, Out in Figure 5.1) undergo a change from the common value in Min1, Min2 and TS1-2 (Table 5.6). In addition, the enhanced hydration causes a weakening of drug/DNA van der Waals interactions (Table 5.2), which mainly involves strand1 (Figure 5.4a).

After TS1-3 the LFEP leads to the metastable minimum Min3, associated to a ΔG of $\sim$6 kcal/mol relative to Min1: more waters enter the same region and the ligand lays now on strand2 (Figures 5.4a, 5.6). The relief of the steric hindrance due to the DNA backbone permits a slight relaxation of the ligand, which can be quantified through the values assumed by dihedral angles between benzene and pyrrol rings and between the pyrrol and the tail $CONH_2$ (respectively Dihe1 and Dihe2 in Table 5.4).

At $d_{CMs} = 1.47$, $n_{hph} = 35$ we reach the highest transition state of the plot (TS3-4, $\sim$12 kcal/mol relative to Min1), in which the drug has still a weak contact with the duplex through its hydrophilic tail while IMI head is completely solvated (Figure 5.1, Tables 5.2, 5.3). From this transition state the complete unbinding of the ligand takes place (Out in Figure 5.1).

During the whole unbinding process the profile of the MG width significantly changes; in particular two subsets can be identified: I, comprising Min1, Min2 and TS1-2, in which the groove is narrow in the tract d[5'-GTT-3']$_2$ and wide in the adjacent tract d[5'-G*GC-3']$_2$; 2); II, encompassing the remaining conformations, in which the situation is reversed. Despite these structural rearrangements, the oligonucleotide remains quite close to the canonical B-DNA conformation, and IMI also undergoes only minor deformations during the whole process(Table 5.4). In conclusion, our results suggest that the detachment process takes place in the following way. The drug first loses contacts with the strand containing the alkylating guanine G*, allowing waters to insert in between (Min3); then it experiences a complete detachment crossing a transition state in which the head of the drug (A, in Chart I) leaves the MG, while IMI tail keeps small residual interactions with the duplex. In this conformation the principal axis of the drug is almost perpendicular to the one of the DNA. The overall calculated barrier for the exit of the drug is $\sim$12 kcal/mol, and two metastable states are found along the LFEP (Figure 5.1, Table 5.1).

Importantly, the detachment of the drug occurs with the same mechanism as above when the active CVs are d_{CMs} and n_{hb} or d_{CMs},n_{hph},n_{hb}, and the overall barrier calculated as a $\Delta G^{\ddagger}(d_{CMs}, n_{hb})$ is $\sim$11 kcal/mol, very similar to that found described here (Figure 5.7).

	Min1	TS1-2	TS1-3	Min2	Min3	TS3-4	Out
Coulomb	-8.4 (2.5)	-7.3 (2.4)	-4.8 (3.0)	-7.9 (2.4)	-2.9 (2.3)	-2.3 (1.3)	-0.1 (0.1)
LJ	-38.1 (3.4)	-37.7 (4.9)	-22.2 (3.5)	-38.0 (4.3)	-19.4 (2.4)	-6.2 (2.2)	-0.4 (0.1)
$\Delta C_p{}^{\text{a}}$	-0.06 (0.13)	-0.06 (0.13)	-0.04 (0.13)	-0.06 (0.12)	-0.03 (0.13)	0.00 (0.12)	-
$T\Delta S_{hyd}{}^{\text{b}}$	4.5 (9.8)	4.5 (9.8)	4.5 (9.8)	4.5 (9.1)	2.3 (9.8)	0.0 (9.1)	-
$\Delta G_{hyd}{}^{\text{c}}$	-4.8 (10.5)	-4.8 (10.5)	-3.2 (10.5)	-4.8 (9.7)	-2.4 (10.5)	0.0 (9.7)	-

a. Heat capacity changes are evaluated as $\Delta C_p = 0.00032\Delta A_{np} - 0.00014\Delta A_p \,\text{kcal mol}^{-1}\,K^{-1}$, where ΔA_{np} and ΔA_p are the changes in the hydrophobic (non-polar) and hydrophilic (polar) surface areas, respectively.
b. Evaluated as $\Delta C_p\,(T - 295)\,\text{kcal mol}^{-1}$.
c. Evaluated as $\Delta C_p\,T\,\ln(T/386)\,\text{kcal mol}^{-1}$.

Table 5.2: Interaction energies (kcal/mol, calculated between IMI and the tract d[5'-CGTTGGC-3']$_2$) and changes in the heat capacity C_p, hydration entropy and hydration free energy (referred to the unbound system, Out) at minima and TS of the LFEP.

	Min1	TS1-2	TS1-3	Min2	Min3	TS3-4
# H-bonds	0.9 (0.6)	0.7 (0.5)	0.5 (0.8)	0.7 (0.5)	0.2 (0.7)	0.0 (0.2)
IMI	9.0 (1.7)	9.1 (1.6)	11.0 (2.3)	8.8 (1.7)	12.4 (2.2)	13.4 (2.4)
d[5'-GTTGG-3']$_2$	6.6 (1.3)	6.7 (1.1)	7.2 (1.1)	6.7 (1.2)	8.9 (1.2)	10.6 (1.0)

Table 5.3: Average number of IMI·d[5'-GGTTGGC-3']$_2$ H-bonds and number of waters within the first hydration shell of IMI and d[5'-GGTTGGC-3']$_2$, calculated through relevant configurations of the LFEP.

5.2.2   DST

The initial configuration has been built by manually docking the ligand on the symmetric 10-mer d[**GCGATTAGCG**]$_2$, using as a template the X-ray structure of the adduct DST·d[**GGCCAATTGG**]$_2$ (see Methods). Figure 5.2 shows the free energy surface $G(d_{CMs}, n_{hb})$ corresponding to the unbinding of DST, along with associated LFEP and relevant conformations assumed by the complex. The bound conformation is associated with a single broad minimum (Min1) centred at $d_{CMs} = 0.45$, $n_{hb} = 57$. Here DST makes on average 4 H-bonds with the DNA duplex (Table 5.6), involving mostly the nitrogen-donors of the ligand and atoms O2 and N3 on the nucleobases. The complex is further stabilized by Van der Waals interactions between the three pyrrol rings of DST and the DNA backbone, and by favourable electrostatic interac-

	Min1	TS1-2	TS1-3	Min2	Min3	TS3-4	Out
Dihe1	26.8 (6.3)	27.4 (7.1)	30.2 (7.0)	27.2 (6.6)	31.0 (7.1)	28.6 (8.2)	31.1 (6.8)
Dihe2	11.8 (6.6)	12.8 (6.9)	14.5 (7.8)	13.3 (7.2)	15.9 (8.7)	13.2 (7.2)	17.8 (10.8)

Table 5.4: Dihedral angles between benzene and pyrrol rings (Dihe1) and between the pyrrol and the tail CONH$_2$ (Dihe2) of IMI along the LFEP path.
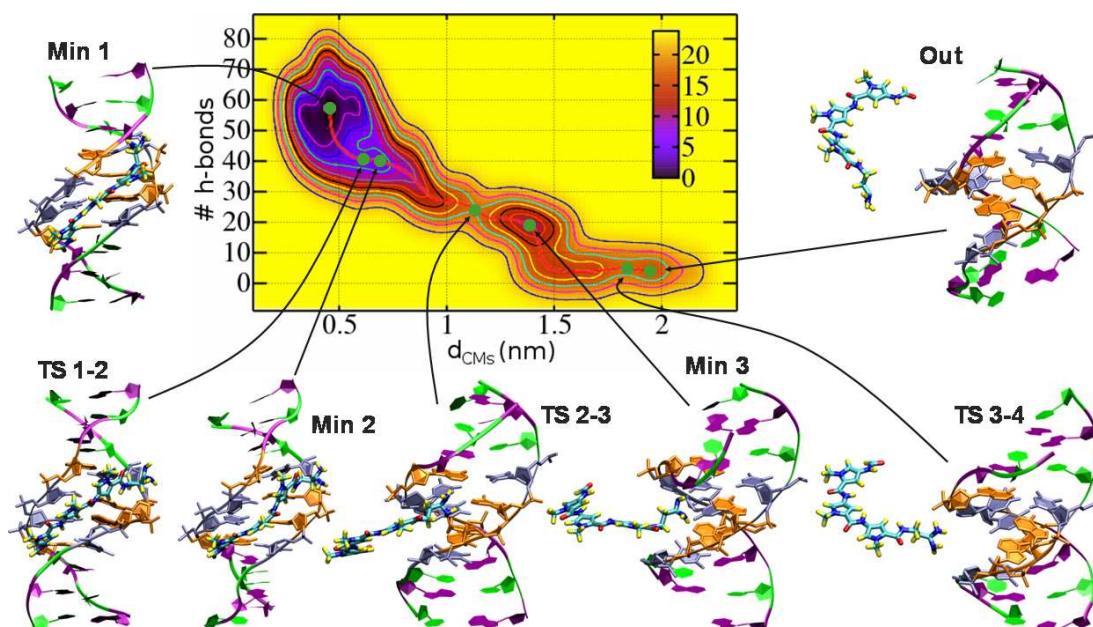
Figure 5.2: Free energy surface $G(d_{CMs}, n_{hb})$ associated to the detachment of DST from DNA. See Figure 5.1 for details.

	Min1	**TS1-2**	**Min2**	**TS2-3**	**Min3**	**TS3-4**
d_{CMs}, n_{hb}	0.45, 57	0.61, 40	0.69, 40	1.13, 24	1.38, 19	1.84, 5.3
ΔG	0.0	5.4	4.5	17.8	10.2	17.2

Table 5.5: LFEP relevant points (kcal/mol, referred to Min1) in the unbinding of DST from DNA.

tions between the charged amidinium tail and the tract d[5'-AG-3']$_2$ (Table 5.6). The free energy of hydration is about -14 kcal/mol with respect to the unbounded configuration and is almost entirely entropic in nature, consistently with: *i)* the presence of hydrophobic rings in the ligand; *ii)* the fact that the DST amide protons are all H-bonded to the nucleobases atoms. The pyrrol rings A and B (Chart I) are slightly rotated with respect to each other to match the curvature of the groove, while the amidinium terminal is rotated with respect to the optimized structure of DST in the solvent in order to maximize contacts with the duplex. To quantify the distortion we define dihedral angles between pyrrol rings (Dihe1 and Dihe2 for the dihedral angles between rings A and B and B and C, respectively) and between the amidinium tail and ring C (Dihe3) (Chart I, Table 5.8). In the bound conformation, Dihe1 and Dihe2 have values of ∼20° (in agreement with experimental [111] values observed for the

	Min1	**TS1-2**	**Min2**	**TS2-3**	**Min3**	**TS3-4**	**Out**
Coulomb	-35.4 (5.3)	-20.6 (2.8)	-23.5 (5.6)	-20.8 (3.7)	-13.9 (5.2)	-18.2 (1.2)	-0.1 (0.1)
LJ	-75.3 (4.6)	-59.5 (4.5)	-56.1 (2.4)	-21.1 (4.0)	-12.2 (2.3)	-0.9 (2.1)	-0.1 (0.2)
ΔC_p	-0.17 (0.12)	-0.15 (0.14)	-0.12 (0.11)	-0.07 (0.15)	-0.04 (0.13)	0.00 (0.06)	-
$T\Delta S_{hyd}$	12.8 (9.1)	11.3 (10.6)	9.1 (8.3)	5.3 (11.3)	3.0 (9.8)	0.2 (5.4)	-
ΔG_{hyd}	-13.6 (9.7)	-12.0 (11.3)	-9.6 (8.9)	-5.5 (12.1)	-3.1 (10.5)	-0.2 (6.3)	-

Table 5.6: Interaction energies (kcal/mol, calculated between DST and the and tract d[5'-CGTAATCG-3']$_2$) and change in the heat capacity C_p, hydration entropy and hydration free energy (referred to the unbound system, Out) at minima and TS of the LFEP. See Table 5.2 for details.

adduct DST·d[GGCC**AATT**GG]$_2$), while Dihe3 changes by $\sim$180°. Interestingly, we also found that DST influences asymmetrically the MG width [111], which is larger at the 3' end of strand1 (Figure 5.3b), whereas it does not influence the overall structural parameters of the duplex. The detachment process starts at the transition state TS1-2 ($\Delta G^{\ddagger} = 5.4$,kcal/mol). Here the drug is significantly distorted with respect to Min1; in particular the amidinium tail now points outside the MG floor, which is reflected by changes of 80° and 50° in Dihe3 and Dihe2, respectively (Table 5.8). The tail reorientation causes a widening of the MG by $\sim$4 Å in the neighbouring region (Figure 5.3b), and the complex gains $\sim$10 waters within the first ydration shell (Table 5.7). Consistently, the average number of DST-DNA H-bonds, the van der Waals and the electrostatic interactions decrease with respect to Min1 (Table 5.6).

The next relevant conformatin we encounter along the LFEP is a relative minimum, Min2, 4.5 kcal/mol higher in free energy than Min1. Here the values of Dihe1 and Dihe2 are very similar to those found in Min1, while Dihe3 further decreases by $\sim$20° with respect to TS1-2 (Table 5.8). The number of drug/DNA H-bonds as well as the overall interaction energy is virtually the same as in TS1-2. A few additional waters enter the hydration shell of the ligand and the duplex; indeed a lower desolvation free energy is found as compared to Min2.

The successive relevant conformation is associated to the highest state in free energy, TS2-3, $\sim$18 kcal/mol relative to Min1. Here DST loose contacts with the MG

	Min1	**TS1-2**	**Min2**	**TS2-3**	**Min3**	**TS3-4**
# H-bonds	3.9 (1.0)	1.7 (0.8)	1.7 (0.8)	1.5 (0.7)	1.5 (0.7)	1.0 (0.2)
DST	11.4 (1.9)	16.4 (1.9)	18.0 (1.5)	18.9 (3.2)	19.8 (2.5)	22.3 (3.7)
d[5'-CGTAATCG-3']$_2$	5.4 (0.8)	8.9 (1.6)	12.5 (0.7)	10.7 (1.1)	12.5 (1.1)	15.3 (1.3)

Table 5.7: Average number of DST·d[5'-CGTAATCG-3']$_2$ H-bonds and number of waters within the first hydration shell of IMI and d[5'-CGTAATCG-3']$_2$, calculated through relevant configurations of the LFEP.

floor but still bounds to the DNA backbone: as can be seen in Figure 5.2, the ligand "follows" the strands of the duplex while its charged tail is still into the groove. As the main body of DST exits the MG this latter recovers the same overall profile it has in Min1 (Figure 5.3b). The loss of hydrophobic contacts between the ligand pyrrole rings and the MG walls in Min3 is mirrored by a significant decrease of van der Waals interactions, while leaving unaltered the other non-bonded interactions (Tables 5.6, 5.7) and the number of waters within first hydration shells. As for the ligand conformation, only Dihe3 features a significant change with respect to Min1.

TS2-3 is followed by the intermediate Min3, in which the main axis of the ligand is almost perpendicular to that of the duplex. Both the electrostatic and van der Waals interactions weaken with respect to TS2-3, while the number of H-bonds between the ligand and the duplex does not change(Tables 5.6, 5.7). As for TS2-3, there is a significant decrease of the solvation free energy gain with respect to Min2, yet the hydration of the complex is quite similar to that found there. This might be traced back to two compensating factors: *i)* the narrowing of the MG (Figure 5.3b), which causes a decrease of the value of the SASA contrasting the increase due to the unbinding of the main DST body; *ii)* the burying of the charged amidinium tail, which compensate for the complete solvation of DST rings and polar hydrogens. Regarding the structure of the ligand, the rings are almost planar among them and Dihe3 also recovers a value very similar to that found in Min1.

From Min3 the ligand exits the MG (configuration Out) passing through TS3-4 ($\Delta G^{\ddagger} \sim 7$ kcal/mol with respect to Min3 and ~ 17 kcal/mol relative to Min1, where it is stabilized by electrostatics interactions with strand1 (Table 5.6, Figure 5.4b). Both the duplex and DST gain a few water molecules in their hydration shells, and the free energy of hydration is virtually zero.

On passing from Min1 to Out the DNA remains close to the canonical B-DNA conformation, and there is a progressive decrease in the hydration free energy, consistently with the largest buried area in Min1 with respect to the other relevant configurations. Summarizing, the unbinding of DST proceeds in three steps *i)* a partial rearrangement of amidinium tails occurs, with breaking of ligand/DNA H-bonds; *ii)* DST pyrrol rings are solvated, while the positively charged tail anchors the drug to

	Min1	**TS1-2**	**Min2**	**TS2-3**	**Min3**	**TS3-4**	**Out**	**QM**[a]
Dihe1	21 (12)	16 (15)	21 (3)	4 (22)	-4 (18)	-9 (4)	-10 (9)	-19
Dihe2	22 (12)	-26 (25)	29 (18)	17 (22)	-20 (16)	-11 (7)	-17 (10)	23
Dihe3	89 (27)	9 (22)	-16 (14)	26 (42)	108 (35)	-71 (21)	81 (24)	-86

a. Taken from the structure optimized at the HF/6-31G(d)/CPCM level (see Methods).

Table 5.8: Dihedral angles between rings A and B (Dihe1), B and C (Dihe2), and between ring C and the amidinium tail of (Dihe3) of DST along the LFEP path.
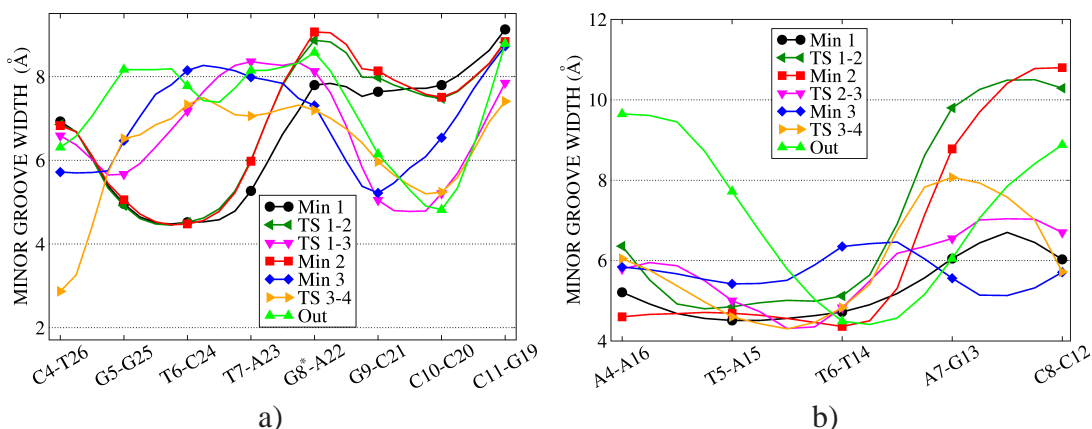
Figure 5.3: a) Variations of the minor groove width in the region d[5'-CGTTGGCC-3']$_2$, calculated along relevant points of the LFEP associated to the detachment of IMI from the DNA duplex; b) Variations of the minor groove width in the region d-[5'-ATTAC-3']$_2$, during detachment of DST.
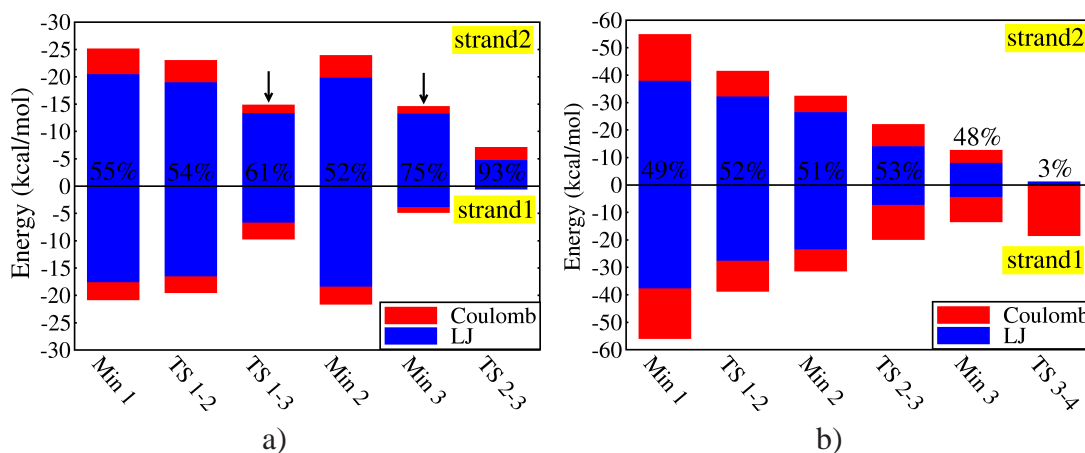

Figure 5.4: a) Interaction energies of IMI with the tracts d[5'-CGTTG*GC-3'] (strand1) and d[5'-CGTTG*GC-3'] (strand2) along relevant points of the LFEP. Arrows indicate configurations where waters enter the region between IMI and strand1, weakening hydrophobic interactions; b) Interaction energies of DST with the tracts d[5'-CGTAATCG-3'] (strand1) and d[5'-CGTAATCG-3'] (strand2). In contrast to IMI, DST interactions with the two strands are almost identical in strength during the entire unbinding process.

MG, *iii)* a complete detachment occurs with the removal of the ligand tail from the groove. The largest activation free energy we found, $\Delta G^{\ddagger} \sim 13$ kcal/mol, is required by the transition from Min2 to Min3 through TS2-3. However, considering the neg-

Figure 5.5: Close view of the difference between conformations Min1 and Min2 in IMI•DNA. The two structures globally overlap, although the ribose sugar and the nucleoside G8 are rotated relatively to each other.
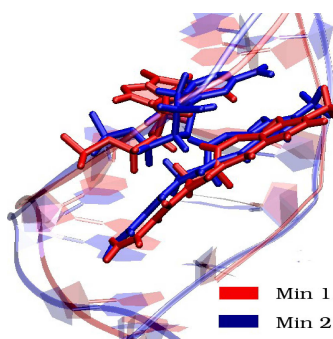
ligible barrier of 1 kcal/mol associated to TS2-1, we can assume an overall activation free energy for the pyrrol rings detachment of $\sim$18 kcal/mol. In addition, the free energy barrier for the complete DST detachment (from Min3 to Min4) is practically identical to that found for the reverse path from Min3 to Min2 ($\Delta G^{\ddagger} \sim$7 kcal/mol).

As for IMI, the detachment of the drug occurs with the same mechanism described above when the active CVs are d_{CMs}, d_{CMs},n_{hph} or d_{CMs},n_{hph},n_{hb} (see Figure 5.8). The largest free energy barrier ($\Delta G^{\ddagger} \sim$14 kcal/mol) corresponds again to the disruption of hydrophobic ligand/DNA contacts and to the solvation of the DST main body (from Point 1 to 7 in Figure 5.8). However, some differences in the free energy surface $G(d_{CMs}, n_{hph})$ are found with respect to Figure 5.2: the overall barrier in the latter is higher by $\sim$4 kcal/mol than that calculated as a function of d_{CMs},n_{hb}, and the intermediate is shifted towards lower values of d_{CMs} and less stable than Min3. The larger discrepancies found in the free energy surfaces as compared to IMI are probably due to the larger number of internal degrees of freedom in DST. The enhanced flexibility of the ligand renders free energy profiles obtained trhough a *dimensional reduction* (as those calculated here) much more "CVs-sensitive", despite a very similar underlying mechanism.
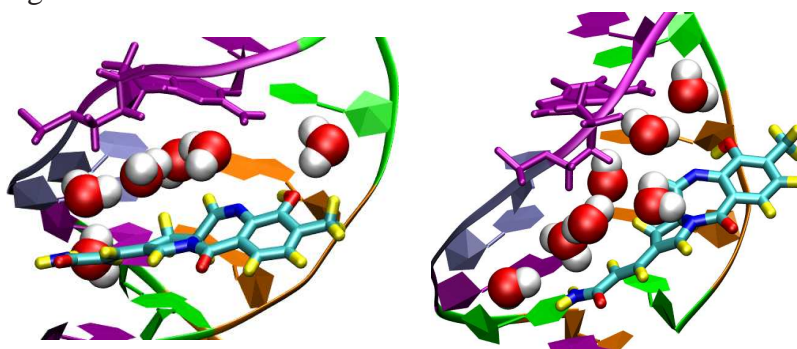
Figure 5.6: View of the MG widening in configurations Min3 (left) and TS1-3 (right) of IMI•DNA. The opening of the groove enhances the hydration of the region between the drug and strand1, which appears to be an important pre-requisite for the exit of the compound.

5.3 Discussion

In this work we have characterized the unbinding of anthramycin and distamycin starting from the non covalent complexes between the ligands and the DNA. Although catching the free energy changes in the process by dimensional reduction is a delicate task, especially for ligands featuring internal flexibility as DST, the mechanism appears to be much less sensitive to the choice of the CVs, at least in this specific case. DNA duplexes do not experience bending or other global deformations upon unbinding of both ligands. This is the case also for the quite rigid molecule anthramycin, consistently with previous results indicating that the drug already possesses a natural conformation to fit into the MG [84, 99]. In contrast, the more flexible DST undergo some relevant conformational change along the process. The barrier for the removal of DST, $\sim$18 kcal/mol, turns out to be significantly larger than that for the detachment of IMI ($\sim$12 kcal/mol), consistently with the higher molecular weight of the former. The calculated activation free energy for DST•DNA compares well with that (16.6 kcal/mol) extracted within the Arrhenius theory from stopped-flow experiments [292].

The primary source of destabilization in DST•DNA is the loss of specific H-bonds, which in turn causes a weakening of hydrophobic interactions. Disruption of these latter corresponds to the highest barrier ($\sim$13 kcal/mol) along the LFEP, in agreement with previous suggestions [9], while electrostatic plays a relevant role in the last step of detachment. These results indicate that, also for this prototype of sequence-readers based on H-bond matching [73, 327, 9], the hydrophobic interactions provide the main contribution to DNA affinity [37, 271, 261], while the charged tail could be the main responsible for the targeting of negative potential AT tracts. Hydrophobic interactions give the largest contribution to the affinity also in IMI•DNA, whereas H-bonds or electrostatics do not play a key role.

Both IMI and DST anchor to the MG *via* their hydrophilic tails before the complete detachment takes place. However, the strength of this interaction is different in the two complexes: DST forms an intermediate by interacting with both DNA strands, because its positively charged amidinium tail complements the negative potential of AT tracts. In contrast, in IMI•DNA such a conformation corresponds to a transition state because the neutral ligand's tail is not able to establish such an interaction. A further difference between the two compounds regards the role of DNA conformational changes and waters in the detachment process. The detachment of IMI is triggered by a local deformation of the duplex and a corresponding hydration of the region between the drug and one DNA strand. On the contrary, for DST no intermediate or metastable states where waters interpose between the ligand and one DNA strand have been identified in our simulations. Thus, water molecules seem not able to interfere with the tight packing between DST and DNA and to play an active role in the unbinding.

Breusegem *et al.* have performed kinetic measurements on DNA binding and unbinding by Hoechst33258 (HST) [297, 269], representative of the class of MG binders containing benzimidazole groups. The comparison of their data with our results, along with those of previous MD simulations of anthramycin and duocarmycin [328], seems to suggests the existence of common aspects in the DNA molecular recognition by MG binders. Indeed they found that dissociation of HST from DNA is characterized by a rather complex behavior, necessitating an intermediate that requires structural changes with respect to the bound conformation [269]. Thus, two different states are apparently involved while the drug is bound within the MG, that can be interpreted either as a sliding of the drug along the MG (as for IMI [84]) or as the presence of an intermediate (as for DST). In addition, as in the case of DST, they suggest that rates of binding and unbinding from the intermediate (here Min3) are very similar, so that the overall dissociation rate is due to the breaking of Van der Waals contacts and hydrogen bonds in the bound state. Although this discussion is in part speculative and requires a further thorough analysis, it is still important to pinpoint common features among the different classes of MG binders.

Limitations of the present calculations. Although appropriate to investigate the detachment, the CVs used here may be too general to distinguish among many different possible orientations that the drug would assume to enter the groove. Particular attention may be payed in the future to identify CVs well suited to characterize the binding process. In addition, no evaluation of the error has been performed as we stop estimating the FES after the complete detachment of the drug (see Methods). Nevertheless, the accuracy of metadynamics has been shown to depend only on the height and widths of the CVs and on their intrinsic diffusion coefficient, along with the deposition time [209]. The choices we made for metadynamics parameters are in line with previous studies, for which the average error has been shown to be less than $\sim$2 kcal/mol [304, 307].

5.4 Conclusions

The mechanism of unbinding of IMI and DST from DNA has been here investigated via metadynamics in the framework of classical molecular dynamics. Our simulations reveal that IMI unbinding occurs *via* the detachment of the ligand from strand containing the alkylating guanine, which leads to the exit of the IMI benzodiazepine head. The process has an overall activation free energy barrier $\sim$12 kcal/mol, and groove hydration appears to be relevant to the weakening of ligand/DNA interactions.

Diversely, DST unbinding occurs through three steps: initially the amidinium tail of the drug rearranges, determining a loss of ligand/DNA H-bonds contacts and increased hydration and MG width; successively, the main body of the drug exits from

the groove; the charged amidinium tail keeps the last contacts with the DNA before unbinding. Due to the stronger packing and electrostatic stabilization (associated with the higher effective contact size within the minor groove) as well as the larger number of specific H-bonds, the overall activation free energy is larger for DST than for IMI ($\Delta G\ddagger \sim 18$ kcal/mol, in good agreement with the experimental value). The accordance confirms the reliability of metadynamics, here applied for the first time to drug/DNA complexes, in the characterization of rare events. The mechanism found here for DST compares quite well with the one suggested for the dye Hoechst33258 on the basis of kinetics measurements. Furthermore, our study provides interesting molecular-level features on the unbinding of representatives of two large and pharmacologically relevant families of MG binder, unveiling their similarities and peculiarities.

Pos	ΔG
1	0.0
2	0.3
3	7.5
4	7.7
5	0.8
6	8.8
7	10.9
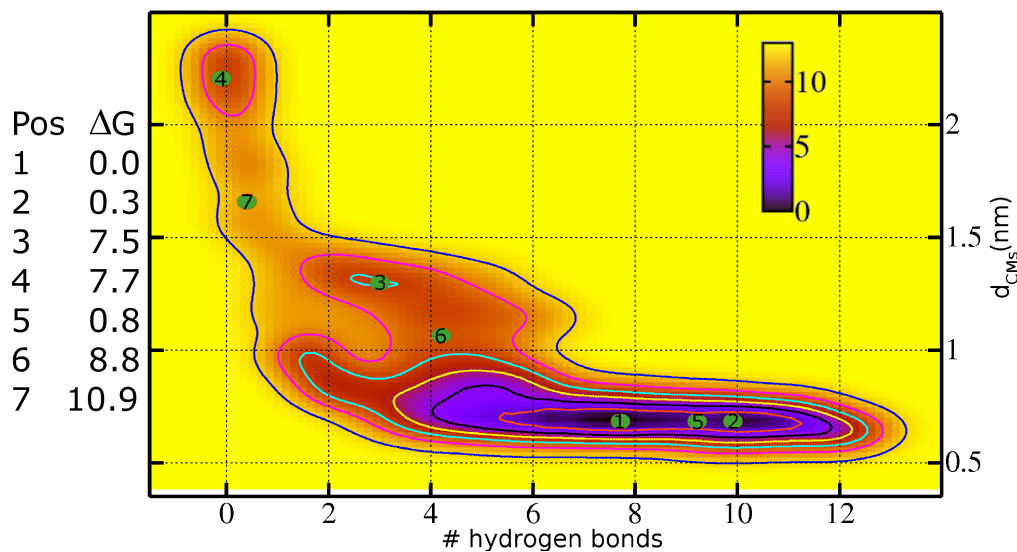
Figure 5.7: Free energy surface of detachment of IMI from DNA calculated as a function of d_{CMs} and n_{hb}. Selected minima and transition states are indicated by green points and corresponding free energies are reported in kcal/mol. Isosurfaces are drawn one per 2 kcal/mol.

Pos	ΔG
1	0.0
2	2.8
3	3.9
4	12.3
5	21.6
6	8.6
7	13.7
8	22.3
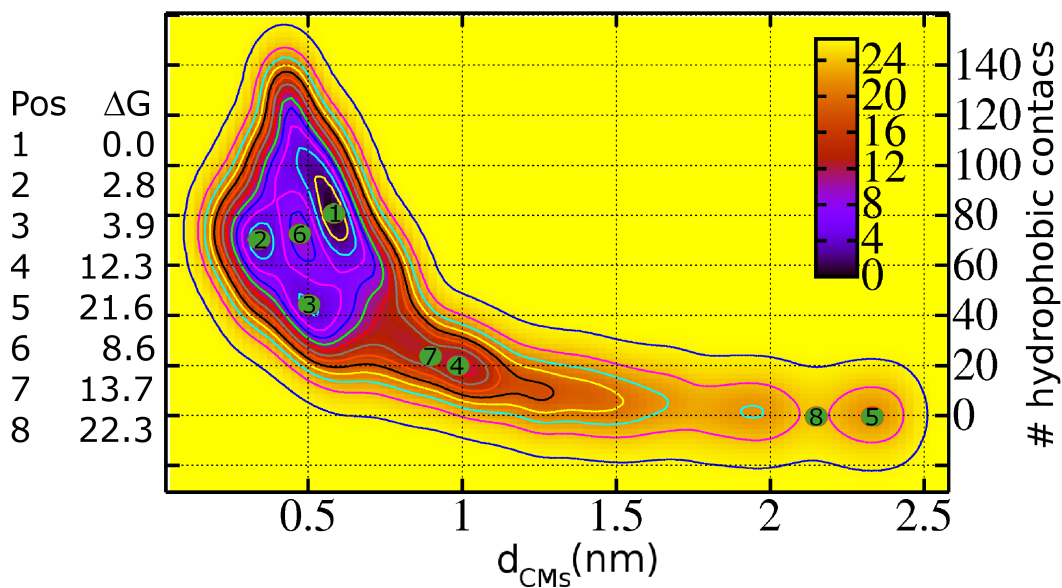
Figure 5.8: Free energy surface of detachment of DST from DNA calculated as a function of d_{CMs} and n_{hph}. Selected minima and transition states are indicated by green points and corresponding free energies are reported in kcal/mol. Isosurfaces are drawn one per 2 kcal/mol.

Part III

Hydrolysis of Ru-based drugs

Chapter 6

Hydrolysis of anticancer ruthenium drugs NAMI-A and ICR

6.1 Introduction

Ru-based complexes are today one of the most promising alternatives to platinum-containing anticancer drugs [329]. Among this class of compounds, (ImH)[*trans*-RuCl$_4$(DMSO-S)(Im)] (Im=imidazole, DMSO-S=S-bonded dimethylsulfoxide), NAMI-A (Chart I), is a powerful antimetastatic agent that already passed phase I clinical trials [330]. Despite its promising pharmacological activity, nature of NAMI-A chemical species responsible for the antimetastatic activity [331, 332] and the mechanism of action [329, 116] have not been unveiled. *In vitro* experiments have shown the drug to readily undergo hydrolysis at physiological conditions (pH=7.4) [333, 334, 335], suggesting that NAMI-A *in vivo* may form a mixture of hydrolyzed species, whose nature strongly depends on pH. NAMI-A disappears after 15-20 min [333, 334, 335] due to both Cl$^-$ and DMSO hydrolyses; 45 minutes are required to complete the first hydrolysis step, while the second occurs within 2 hours [334]. Under acidic conditions (pH=3-6), Cl$^-$ dissociation is much slower and only a partial DMSO dissociation occurs [331, 332, 333, 334, 336]. Hydrolysis eventually leads to formation of polymeric oxo- or hydroxo-bridged species [334, 335], albeit this phenomenon is less likely *in vivo* due to the bio-ligands availability, *e.g.*, presence of albumin precludes formation of such polymeric species [337]. NAMI-A metabolites have shown to bind strongly to several proteins, including integrins, transferrins and human serum albumin [337, 338, 339, 340, 341, 342]. In contrast to platinum drugs, NAMI-A binds to DNA weakly [343], indicating that its antimetastatic activity may be not related to DNA damage. The compound (ImH)[*trans*-RuCl$_4$(Im)$_2$], ICR (or KP418), (Chart I) and its indazole analogue KP1019 [116, 344, 345, 346], albeit structurally similar to NAMI-A, have promising anticancer activity against colorectal can-

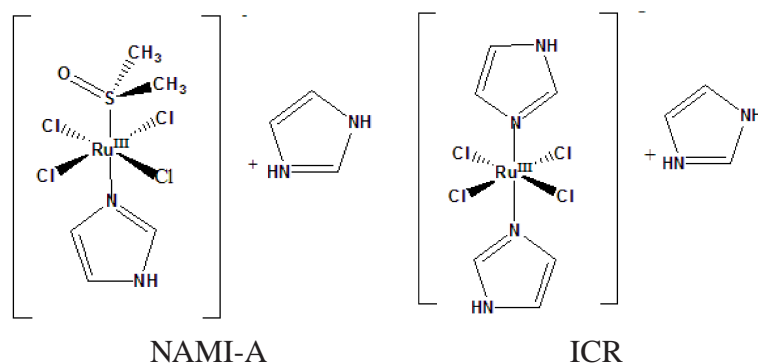NAMI-A   ICR

Chart I

cers [342, 345, 346], but no relevant antimetastatic effects [344]. ICR reacts with DNA following a cisplatin-like kinetics, inhibiting its template-primer properties for nucleic acid synthesis [347]. Experimental data indicate that *in vitro* ICR undergoes step-wise hydrolysis of chloride ions [345, 348], but nature and oxidation state of the active ruthenium metabolites are still unknown. The hydrolysis is supposed to be a crucial step as only aged ICR solutions how the drug binding to DNA [349].

The biological diversity between NAMI-A and ICR may be partially due to their different redox properties: reduction potentials *Em* (*vs.* normal hydrogen electrode, NHE) are 0.235 V and -0.275 V, respectively [350, 351]. Such a difference is caused by the DMSO, which is a moderate π-acceptor and stabilizes RuII [332, 333]. Because of its positive *Em*, NAMI-A may be easily reduced *in vivo* by biological reductants, even before undergoing dissociation. Indeed, the reduced form of NAMI-A and/or its metabolites have been demonstrated to be of crucial importance [331, 332, 333, 334], having a robust antimetastatic activity on tumor models [334]. In contrast, ICR would need stronger reductants which may be more rare in a biological environment [331, 348, 352]. Consistently, ICR has proved to be more inert than NAMI-A, and its hydrolysis path is less affected by reductants [353].

Recently, Chen *et al.* have reported a theoretical investigation on NAMI-A and ICR hydrolyses [354, 355], focusing on Cl^- dissociation. Calculated structural properties and kinetic constants are in good agreement with the available experimental data. However, dissociations of DMSO and imidazole have not been studied, despite the proved biological relevance of DMSO (partial) hydrolysis [333, 334]. Furthermore, the hydrolysis of the reduced compounds has not been considered, even though it may be important for both drugs due to the high reducing power of solid tumor tissues [348, 356].

In order to understand the mechanism of action of NAMI-A and ICR at the molecular level, an extensive investigation of the hydrolysis paths in both Ru oxidation states, along with an evaluation of the reduction potentials of the two drugs and their

metabolites is needed. In this work we tackle this issue and report a systematic study of the hydrolyses of the two drugs up to the third aquation, combining DFT calculations and an implicit description of the solvent through the Conductor-like Polarizable Continuum Model (CPCM) [323]. Reduction potentials of NAMI-A, ICR, and of their most relevant hydrolyzed species were estimated within the same setup, to explore the effects of biological reductants.

This study is a first attempt to rationalize the large amount of available experimental findings, that in some cases led to discrepant conclusions [332, 348]. In addition, it confirms the ability of computational studies to qualitatively predict the hydrolysis of different substituents from metal-containing compounds.

6.2   Computational Details

The hydrolysis of NAMI-A and ICR in both RuIII and RuII oxidation states was investigated in the framework of the Density Functional Theory (DFT), employing the Gaussian03 package [273]. In order to consider all possible hydrolytic paths, dissociation of each ligand (namely Cl^-, DMSO-S and Im), was studied in the presence of one explicit incoming water. To achieve a trade-off between accuracy and computational costs the following methodology was adopted [357, 358]:

1. Full *in vacuo* optimization without symmetry constraints of reactants (R), products (P) and transition states (TS) using the B3LYP exchange-correlation functional [157, 161], with 6-31G(d,p) and Lanl2DZ [359, 360, 361] basis sets on N, S, C, O, H, Cl and Ru atoms, respectively.

2. Harmonic frequency calculations at the same level of theory, *i)* to confirm that optimized structures were minima or transition states, *ii)* to estimate zero-point vibrational energies, thermal and entropic corrections.

3. Single point calculations with larger basis sets on optimized structures. Namely, 6-311+G(3d) on Cl and S atoms, 6-311+G(d,p) on all the other non-metal atoms, and Lan2DZ(f) (ζf = 1.235) [362] on Ru were used in order to obtain an improved value of the internal energy [357, 358].

4. Single point calculations (same set-up as in *c*) on optimized structures mimicking hydration effects *via* an implicit solvation model (CPCM) [323]. We are aware that PCM calculations of small ions such as Cl^- may be problematic. However, a very good agreement was found between experimental and theoretical solvation free energies of Cl^- calculated here [323, 363] (data not shown).

Free energies were then calculated adding to the internal energy the zero-point vibrational energy, the entropic term (using Wertz correction) and the solvation free energy [364, 365]. Correction for the standard concentration was also considered [354, 355].

Activation free energies are usually calculated as the difference between the free energy of the transition state (TS) and that of the reaction intermediate RI, i.e. the stable complex between the reactant and a water molecule (the same terminology is used also for the product intermediate, indicated by PI). However, Lau *et al.* have cast doubts on the physical basis of considering reaction intermediates RI and PI as reference states for calculating activation and reaction free energies of hydrolysis processes [358]. Indeed, they have shown that the free energy of an isolated reactant and an isolated water molecule (named as R, and P for the reactant and product, respectively) significantly differs from that of RI (and PI). This discrepancy is believed to arise from the failure of implicit solvent models at reproducing the correct solvation entropy in bimolecular reactions. To overcome this issue in Ref. [358] the entropy was corrected by Wertz's approach [364, 365], obtaining almost identical free energies for RI and R. Unfortunately, the same procedure does not provide consistent results when applied to the PIs, whose free energies are systematically higher than those of Ps. In particular, all hydrolysis steps turned out to be endothermic (up to 10 kcal/mol) when comparing the free energy of PI to that of RI. Considering the free energy differences between reactants R and products P led to reaction free energies in close agreement with experiments.

Therefore, we calculated the activation free energies by referring each TS to the corresponding RI (which are likely to be affected to the same extent by the inaccuracy of our computational model), while reaction free energies were calculated by comparing P with R, as in Ref. [358]. We believe this protocol to be more suitable to correctly calculate TS barriers, and to minimize the uncertainty when comparing different hydrolysis products (e.g. those resulting from Cl^- or DMSO dissociation). Notably, this scheme gave results in good agreement with previous theoretical and experimental results (Tabs. 6.1, 6.3), *vide infra* [350, 354, 355, 366]. We are thus confident that our set-up may predict thermodynamic and kinetic data, not experimentally available, and may be useful to rationalize both NAMI-A and ICR hydrolyses.

For the sake of clarity, products of different steps were labeled as *PNa,b,c* where *a,b,c* refers to the ligand exchanged during hydrolysis, N indicating the hydrolysis step. Similarly, transition state structures were reported as *TSNa*, where *a* is the leaving group, N indicating the hydrolysis step. We would like to stress that reactions involving small groups or atoms are routinely characterized, while those in which the leaving groups are as large as Im (or DMSO) present many issues. In particular, due to the large number of degrees of freedom the search for their TS may be awkward. We indeed found accurate TS structures corresponding to Cl^- and DMSO dissociation. Unfortunately, such a satisfactory accuracy was not achieved for Im dis-

sociation. Due to these difficulties, we considered as (pseudo) TS for Im dissociation the structure corresponding to the highest energy while breaking the Ru-N_{Im} bond in the presence of one explicit water. This should allow a qualitative comparison of the Im dissociation barrier with activation free energies of Cl^- and DMSO. Indeed, our results suggest, in line with experimental findings [333], that Im dissociation is not likely under neutral conditions.

A final comment on the reactants concentrations is necessary. It has been shown that cell chloride ion concentration strongly affects the hydrolysis of other metal-based drugs, such as cisplatin [367, 368, 369]. This may be expected to occur also for NAMI-A and ICR, but our theoretical approach cannot consider such an important aspect. Nevertheless, Reedijk and co-workers [333] have shown that Cl^- concentration affects both Cl^- and DMSO dissociations from NAMI-A^{III} in a similar manner, suggesting that also other hydrolysis pathways may be similarly affected *in vivo*.

All possible hydrolytic paths[1] were characterized employing the approach discussed above, yet, for the sake of clarity, we only reported energetics, along with selected structural and electronic properties of the most relevant complexes (for more details, see Supporting Information). Remarkably, structures of NAMI-A and ICR obtained with our computational set-up are very similar to those reported in previous theoretical studies [354, 355] and in fair agreement with experimental results [350, 112] (Tab. 6.1). Discrepancies may be caused by the intrinsic inaccuracy of the computational approach as well as to crystal packing forces present in the X-ray structures [354, 355]. To further test our computational set-up, we performed calculations with larger basis sets with and without the inclusion of solvent during optimization, finding virtually the same geometries as with the previous setup. Furthermore, structural and energetic properties of NAMI-A^{II} were calculated at the MP2 level [143], founding a satisfactory agreement with DFT results.

Molecular properties were calculated using Bader *et al.*'s theory of Atoms In Molecules (AIM) [370, 371], in order to analyze the electron density (ρ) at Bond Critical Points (BCP). Indeed, the electron density at such critical points correlates well with the strength of chemical bonds and interactions [372, 373, 374, 375]. Topologies were built using AIMPAC series of programs [376]. Charges were calculated with the full Natural Bond Orbital (NBO) analysis [377]. Both NBO and AIM analyses were carried out considering the solvent effects. Following previous works [378, 379], reduction potentials (*Em*) at standard conditions were estimated from the following equation:

$$E_m = \frac{\Delta G}{nF} \tag{6.1}$$

---

[1]For example, about 15 geometry optimizations were performed only for the first hydrolysis step of NAMI-A^{III}, or more than 400 calculations for describing the entire hydrolysis paths for NAMI-A and ICR in both oxidation states.

where $F = 96485\,C$ (the Faraday constant), E_m is the reduction potential and n the number of electrons involved in the redox process. To compare reduction potentials with experimental data, E_m were referred to the normal hydrogen electrode, NHE. As suggested by previous benchmark studies [380], to reproduce redox potential is not a trivial task and medium-large basis sets, zero-point-energy and entropic corrections should be employed to provide a semi-quantitative agreement with experimental results.

6.3 Results and Discussion

Energetic, structural and electronic properties of the most relevant metabolites of NAMI-A and ICR in both RuIII and RuII oxidation states (hereafter as NAMI-A$^{III/II}$, ICR$^{III/II}$) are reported. As NAMI-A is likely to be reduced *in vivo*, and NAMI-A^{II} metabolites have shown antimetastatic effects [331, 332, 334, 345, 346, 352, 380, 114], the chemical behavior of reduced species is crucial to understand the biological activity of this drug. Reduction of ICR in biological environments is less likely, but may take place in cancerous tissues. For these reasons, and since theoretical studies on NAMI-A^{III} and ICR^{III} have recently been reported [354, 355] discussion of RuII species is more detailed and precedes that of RuIII ones.

6.3.1 Hydrolysis Mechanism of NAMI-A

Fig. 6.1 displays the most likely hydrolysis paths of NAMI-A in both Ruthenium oxidation states, up to the third hydrolysis step. At the end of the Chapter is also reported a complete thermodynamic characterization of all the hydrolysis steps investigated here (Figs. 6.4, 6.5, 6.6 6.7). Dissociation of Im typically requires an activation free energy ($\Delta G^{\ddagger}$) larger than that of Cl$^-$ and DMSO, and products of Im hydrolysis have the largest endothermic character. Therefore, Im-water exchange is unlikely under neutral conditions in both Ru oxidation states, in agreement with experimental results [116]. This may be partly due to the strength of their coordination bonds, e.g., the Ru-N bond of NAMI-A^{II} is somewhat stronger than the Ru–S bond ($d_{Ru-N} = 2.10\,\text{Å}$ and $\rho_{Ru-N} = 0.0701$ au, $d_{Ru-S} = 2.32\,\text{Å}$ and $\rho_{Ru-S} = 0.0576$ au, where ρ is the density at the critical point, Tab. 6.1). Thus, the following discussion is limited to Cl$^-$ and DMSO hydrolyses.

NAMI-A^{II}

Step I. In line with previous theoretical studies [354] and experimental findings [332] our data suggest that Cl$^-$ dissociation is the most favorable process. The activation
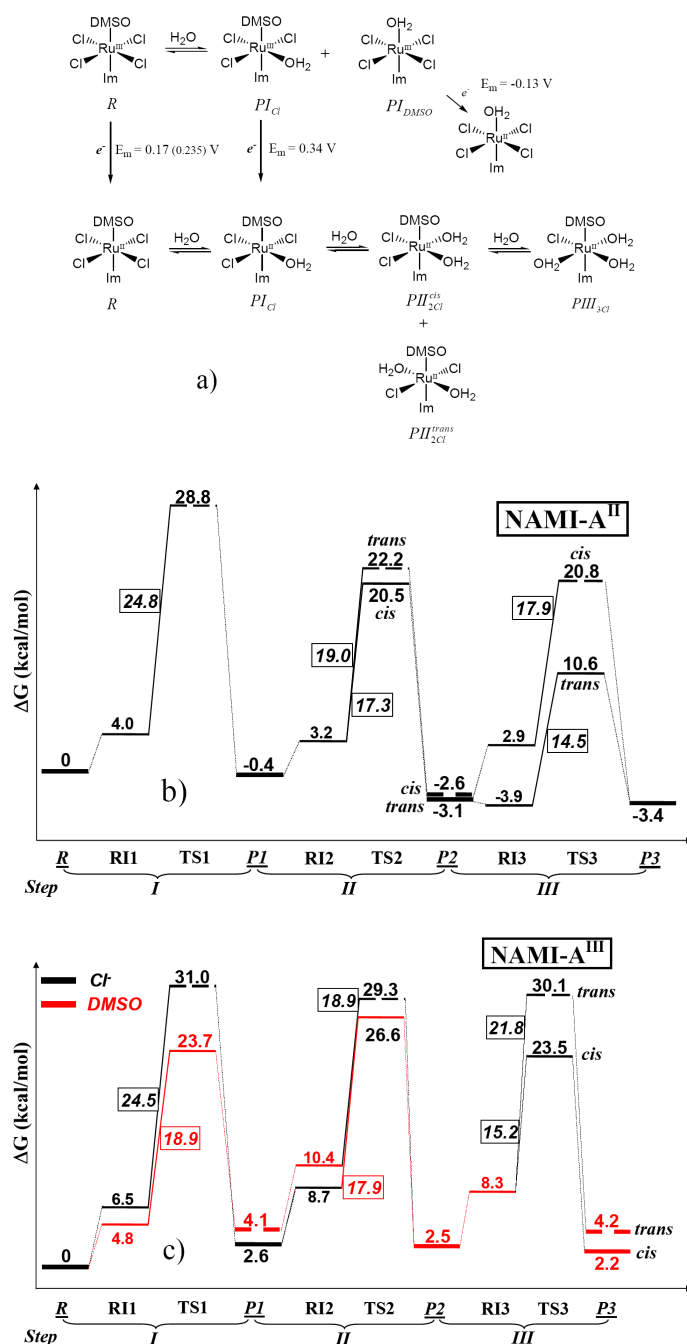
Figure 6.1: *a)* Proposed reaction paths for the hydrolysis of NAMI-A. Calculated and experimental (in brackets) reduction potentials *vs* NHE are shown. *b)* Free energy profile associated to the most likely hydrolysis path of NAMI-A^{II}, up to the third substitution. The most favorable transition states and products are represented by solid lines, while dotted lines are used for unlikely transition state and intermediates. Activation energies (kcal/mol) are displayed in the squares along the reaction path. Reaction free energies are calculated with respect to R. *c)* Free energy profile associated to the most likely hydrolysis path of NAMI-A^{III}. Black and red lines correspond to Cl$^-$ and DMSO hydrolysis paths.

free energies of DMSO and Cl^- dissociations are comparable (Fig. 6.4), but thermodynamics largely favors chloride ion hydrolysis (PI_{Cl} is virtually isothermic to R, whilst PI_{DMSO} is endothermic by 10.7 kcal/mol). Indeed, the NBO analysis confirmed that after DMSO dissociation, the drug will maintain the overall -2 charge, losing a large group on which the electron density may delocalize. In contrast, the leaving Cl^- has partial charge of about -0.9 au, i.e., the overall negative charge on the metal complex is reduced to -1, promoting the reaction. The larger stability of PI_{Cl} as compared to PI_{DMSO} may be partly caused by an intra-molecular O-H$\cdots$O$_{DMSO}$ H-bond between the coordinated water molecule and the DMSO oxygen ($d_{H\cdots O}$=1.96 Å and ρ=0.0189 au), as well as an intra-molecular O-H$\cdots$Cl H-bond, ($d_{H\cdots Cl}$=2.30,Å and $\rho_{H\cdots Cl}$=0.0148 au). We verified that the isomer in which the Ru-OH$_2$ bond is rotated by 180∘ (no H-bonding) is less stable by ∼6 kcal/mol, i.e. PI_{Cl} should be preferred over PI_{DMSO} also in case the intra-molecular H-bond to DMSO-S is broken. Thus, both kinetic and thermodynamic features suggest that chloride ion hydrolysis is the most favored at the first step.

Step II. Reaction proceeds faster from step I to II, as the activation free energy of the second Cl^--water exchange decreases by ∼6-8 kcal/mol. As in previous step, this dissociation is kinetically competitive with that of DMSO, yet relative products stability favors dissociation of a second Cl^-. Indeed, the product of DMSO hydrolysis is endothermic by 10.4 kcal/mol, while PII_{2Cl}^{cis} and PII_{2Cl}^{trans} sit respectively at -2.6 and -3.1 kcal/mol with respect to R. This would rule out the possibility of DMSO dissociation at the second step, in line with *in vitro* studies at neutral pH [332, 335]. In addition, our calculations suggest formation of a mixture of *cis* and *trans* diaqua isomers, supporting experimental hypotheses [332].

Step III. The analysis of the third hydrolysis reveals intriguing features. In particular, DMSO dissociation becomes kinetically unlikely with respect to that of Cl^-, yet

	NAMI-A[II]		ICR[II]		NAMI-A[III]		ICR[III]	
	d	ρ	d	ρ	d	ρ	d	ρ
Ru-Cl	2.55	0.0604	2.58	0.0529	2.43(2.34)	0.0627	2.44(2.36)	0.0609
Ru-S	2.32	0.0576	-	-	2.40(2.30)	0.0520	-	-
Ru-N	2.10	0.0701	2.10	0.0772	2.10(2.08)	0.0854	2.10(2.08)	0.0761

Table 6.1: Selected bond lengths (d, Å) and electron densities (ρ, au) of NAMI-A and ICR in both oxidation states. Experimental values (in brackets) of NAMI [330, 331] (the analogous sodium salt of NAMI-A) and ICR [332, 335] bond lengths are displayed (Ru-Cl is reported as average).

the free energy difference between $PIII_{3Cl}$ and $PIII_{2Cl,DMSO}$ reduces as compared to the first and second steps. A simple electrostatic analysis may provide a rationale for this behavior. During the first two hydrolysis steps, the metal complex bears an overall negative charge of -2 and -1, respectively, i.e. dissociation of the negatively charged chloride is favored with respect to neutral DMSO-S ligand. After the second Cl^- dissociation, PII_{2Cl} is neutral, with DMSO dissociation becoming more favorable. Thus, the hydrolysis of NAMI-A[II] mainly consists of a stepwise Cl^--water exchange: all Cl^- dissociations are exothermic with decreasing activation free energy barriers as the reaction proceeds. Findings of the first, second and third hydrolysis steps are in qualitative agreement with experiments by Alessio and co-workers, who extensively characterized the hydrolytic behavior of NAMI-A[II] at pH=7 [332, 351].

NAMI-A[III] *vs.* **NAMI-A**[II] **hydrolysis**

At the first hydrolysis step, the activation free energy of Cl^- dissociation is larger (by ∼6 kcal/mol) than that of DMSO (Fig. 6.1). Notably, the calculated barrier for Cl^- dissociation ($\Delta G^{\ddagger}$=24.5 kcal/mol) is in good agreement with experimental measurements ($\Delta G^{\ddagger}$=22.8 kcal/mol) and recent, more accurate and computationally expensive, theoretical predictions ($\Delta G^{\ddagger}$=23.2 kcal/mol) [350, 354]. Since PI_{Cl} and PI_{DMSO} are virtually isothermic, DMSO substitution, thermodynamically hindered for NAMI-A[II], is very likely to take place during NAMI-A[III] hydrolysis. Indeed, geometrical properties and electron density analysis confirm that the Ru-S bond is stronger in NAMI-A[II] than in NAMI-A[III] , i.e. DMSO-S is a moderate π-acceptor and the π component of the RuII-S bond is expected to be stronger when the complex has a negative charge [334] (see Tab. 6.1). Consistently with these results, experiments by Sava *et al.* showed that a slow DMSO hydrolysis occurs [334]. This is also in line with NMR, spectroscopic and X-ray data, which indicate that DMSO-S binds more strongly to RuII than to Ru[III] [350, 331].

Irrespective of the path followed, the product of the second hydrolysis step is $PII_{Cl,DMSO}$: the most likely leaving group from PI_{Cl} is DMSO, with a $\Delta G^{\ddagger}$=17.9 kcal/mol, while, the release of Cl^- from PI_{DMSO} occurs with a kinetic barrier of 18.9 kcal/mol. This scenario is compatible with experimental findings of a partial DMSO dissociation occurring simultaneously to Cl^- hydrolysis [333]. The loss of a further Cl^- from $PII_{Cl,DMSO}$ is kinetically more favored in *cis* than in *trans*, with the entire process being slightly endothermic.

These results suggest that reduction significantly affects the hydrolysis of NAMI-A: NAMI-A[II] hydrolysis is overall exothermic with slightly lower activation free energies barriers compared to those of NAMI-A[III], this difference increasing as the hydrolysis proceeds. Importantly, our data suggest that neither NAMI-A reduction nor DMSO dissociation can be neglected *a priori*, since they may both play a fundamental role for NAMI-A antimetastatic activity.

6.3.2 Hydrolysis Mechanism of ICR

In order to identify those features responsible for the different pharmacological activity of NAMI-A and ICR, a broad analysis of the ICR hydrolysis mechanism is proposed. In agreement with experimental studies [333] we find that imidazole dissociation is typically unlikely in both oxidation states.

ICR^{II}

The most likely group to dissociate from ICR is Cl$^-$ (Figs. 6.2b, 6.6) with $\Delta G^{\ddagger}$=23.2 kcal/mol, PI_{Cl} being virtually isothermic to R. Interestingly, the activation free energy is similar to that of 24.8 kcal/mol estimated for TSI_{Cl} of NAMI-A^{II}, suggesting that kinetics of the first Cl$^-$ dissociation are virtually unaffected by the presence of DMSO-S or Im ligands. At the second hydrolysis step PII_{2Cl}^{trans} and PII_{2Cl}^{cis} may originate with the following energetics: $\Delta G^{\ddagger} = 21.6$ and $\Delta G = 0.7$ kcal/mol, $\Delta G^{\ddagger} = 16.1$ and $\Delta G = -1.8$ kcal/mol respectively. Importantly, our calculations strongly suggest that formation of *cis* diaqua complexes is kinetically preferred to that of *trans* isomers. This may be due to an intricate interplay of several factors, including *trans* influence and H-bonding. Firstly, as consequence of the *trans* influence, Ru-Cl in *trans* to water is the strongest in PI_{Cl}, and Cl*trans* therefore is the least likely to leave (d$_{Ru-Cl}^{trans}$=2.46 Å and $\rho_{Ru-Cl}^{trans} = 0.0666$ au, as compared to d$_{Ru-Cl}^{cis}$ =2.56 Å and ρ_{Ru-Cl}^{cis}=0.0560 au, averaged over the other two Ru-Cl bonds).

Because of a strong H-bonding (Tab. 6.2), both entering water and leaving Cl$^-$ interact more strongly with the metal center compared to $TSII_{Cl}^{trans}$. These effects may provide the extra stabilisation of the transition state that leads to the *cis* diaqua isomer. The kinetic preference for PII_{2Cl}^{cis} with respect to PII_{2Cl}^{trans} is remarkable as may be related to different pharmacological activity of NAMI-A and ICR. Yet, calculations that include dynamical properties of explicit solvent molecules are needed to confirm this point beyond doubt. Nevertheless, we would like to remark that strong intramolecular H-bonds should be less affected by the presence of explicit water molecules [373, 381]. In agreement with experimental findings [349], thermodynamics of *cis* and *trans* diaqua complexes are virtually equivalent i.e., on a long time scale both complexes are formed. Similarly to NAMI-A^{II} hydrolysis, the third Cl$^-$ substitution is exothermic and most likely to take place from $TSII_{Cl}^{trans}$.

ICR^{III} *vs.* ICR^{II} hydrolysis

As suggested by the experiments of Mestroni *et al.* [353], a change in Ru oxidation state affects ICR hydrolysis to a minor extent (Figs. 6.2c, 6.6). Indeed, the ICR^{III} hydrolysis consists of a stepwise Cl$^-$-water exchange with Im dissociation being unlikely at every step. The activation free energy relative to the first hydrolysis is virtu-
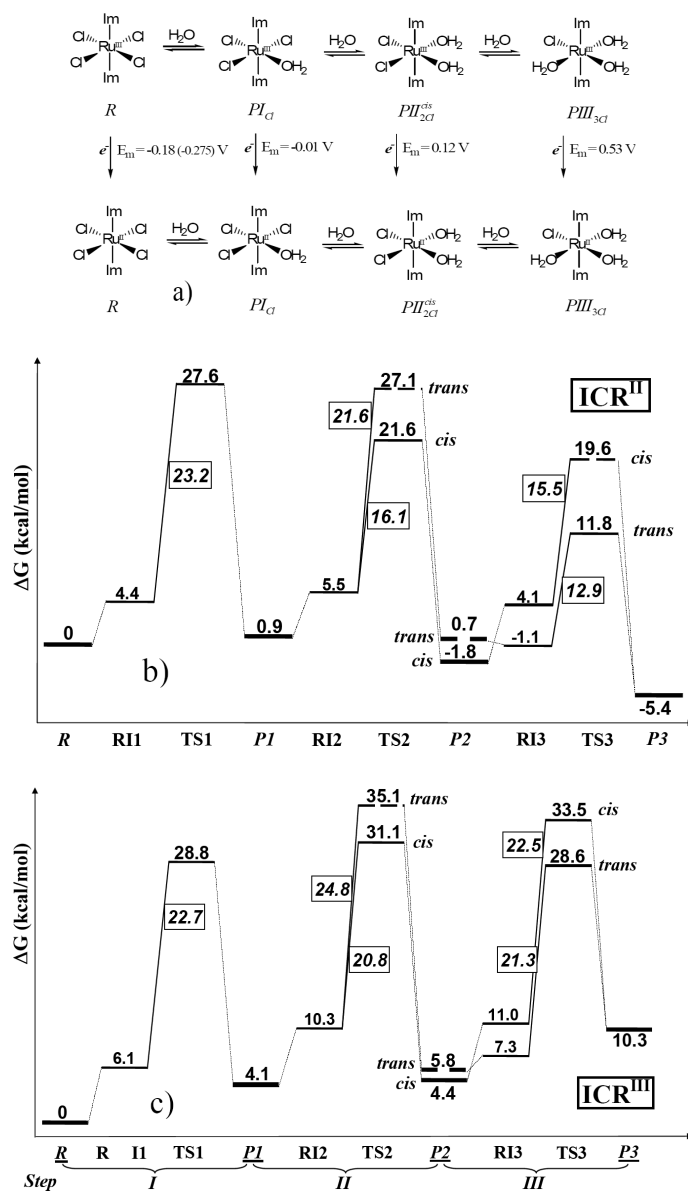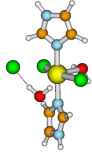
Figure 6.2: *a)* Proposed reaction paths for the hydrolysis of ICR. Calculated and experimental (in brackets) reduction potentials *vs* NHE are shown. *b)* Free energy profile associated to the most likely hydrolysis path of ICR^{II}, up to the third substitution. The most favorable transition states and products are represented by solid lines, while dotted lines are used for unlikely transition states and reaction intermediates. Activation energies (kcal/mol) are displayed in the squares along the reaction path. Free energies are calculated with respect to R. *c)* Free energy profile associated to the most likely hydrolysis path of ICR^{III}. Only *cis* diaqua complexes are shown to remark the ICR kinetic preference for these isomers.

		O*-H $\cdots$ Cl‡	Ru-O-H $\cdots$ Cl‡	Ru $\cdots$ Cl‡	Ru $\cdots$ *OH$_2$
PII_{2Cl}^{cis}		2.12(0.0300)	1.95(0.0437)	3.20(0.0173)	2.84(0.0189)
PII_{2Cl}^{trans}		2.05(0.0347)	- -	3.28(0.0135)	2.91(0.0144)

* entering water molecule; ‡ leaving Cl ion.

Table 6.2: Selected geometrical (Å) and electron density (in brackets, au) properties of TS leading to PII_{2Cl}^{cis} and PII_{2Cl}^{trans} products of ICR hydrolysis.

ally insensitive to the Ru oxidation state, but the process is thermodynamically more favored. Calculated activation free energy ($\Delta G^\ddagger$=22.7 kcal/mol) is in good agreement with that recently found by a theoretical investigation based on more computationally demanding calculations (24.0 kcal/mol) [355] and to that measured experimentally (24.1 kcal/mol) [366].

As for the first step, Ru oxidation state hardly affects the second hydrolysis. Indeed, chloride ion dissociation is the most likely event, kinetically preferred (by 4 kcal/mol) in *cis* rather than in *trans*. The activation free energies, $\Delta G^\ddagger$=20.8 and 24.8 kcal/mol for $TSII_{Cl}^{cis}$ and $TSII_{Cl}^{trans}$, respectively, are in good agreement with the values of 20.0 and 25.4 kcal/mol reported recently [355]. By contrast, thermodynamics of *cis* and *trans* diaqua complexes are similar, suggesting that as for reduced species, both products may be found in solution [349]. The third hydrolysis step has a similar barrier to that of previous aquations, but, unlike ICR^{II}, reaction is endothermic.

6.4   Estimation of Reduction Potentials

As the electrochemical behavior of NAMI-A and ICR is crucial for their biological activity [331, 334, 348], reduction potentials of these drugs as well as those of their most relevant hydrolysis products were estimated (Tab. 6.3). Calculated and experimentally measured reduction potentials of NAMI-A and ICR are in good agreement, as they differ by $\sim$0.08 V on average. Thus, the estimated difference between reduction potentials of NAMI-A and ICR was found to be close to the experimental value, $\Delta\Delta E_m \sim$ 350mV *vs.* $\Delta\Delta E_m^{\exp} \sim$ 500mV. Unfortunately, data about redox potentials of their hydrolysis products are not available in literature. However

	NAMI-A		ICR	
	Calc.	Expt.	Calc.	Expt.
R	0.17	0.235^a	-0.18	-0.275^b
PI_{Cl}	0.34	$0.337^{a,c}$	-0.01	
$PI_{DMSO(Im)}$	-0.13		-0.13	
PII_{2Cl}^{cis}	0.67		0.12	
PII_{2Cl}^{trans}	0.57		0.07	
$PII_{Cl,DMSO(Im)}$	-0.17		-0.17	
$PIII_{3Cl}$	1.02		0.53	
$PIII_{2Cl,DMSO(Im)}$	0.09		0.09	

a: from Ref. [350]; b: from Ref. [331]; c: experimental reduction potential refers to *mer*-RuCl$_3$(DMSO)$_2$(Im), see text for further details.

Table 6.3: Calculated and experimental reduction potentials (E_m, *vs*. NHE - V) of NAMI-A and ICR metabolites.

electrochemical measurements on *mer*-RuCl$_3$(DMSO)$_2$(Im), which contains a Ru-O$_{DMSO}$ bond similar to Ru-O$_{water}$ of the first hydrolysis product of NAMI-A, are available [350]. Indeed, the experimental reduction potential (E_m=0.337 V, *vs* NHE), compares very well with our calculated value (0.34 V). These data confirm the reliability of our computational set-up, allowing us to predict reduction potentials for NAMI-A and ICR metabolites possibly with a similar accuracy.

As the hydrolysis of Cl$^-$ proceeds to PI, PII and $PIII$, reduction potentials increase to $\sim$0.3, $\sim$0.7 and $\sim$1.0 V and to $\sim$0.0, $\sim$0.1 and $\sim$0.5 V for products of NAMI-A and ICR, respectively (Fig. 6.3). This is in qualitative agreement with experimental findings showing that for each Cl$^-$-water exchange a constant increase of the reduction potential occurs [352, 382]. In contrast, reduction potential decreases upon dissociation of DMSO (or Im for ICR), to -0.13 V and -0.17 V for $PI_{DMSO(Im)}$ and $PII_{Cl,DMSO(Im)}$ (Tab. 6.3). This suggests that DMSO dissociation may preclude NAMI-A reduction, slowing down the hydrolysis and possibly the formation of the active metabolites, confirming the crucial role played by DMSO for NAMI-A pharmacological activity [331, 334].

Reduction of ICR *in vivo* is unlikely as reduction potential is negative, yet this may occur in specific regions where strong biological reductants are available, such as proliferating cells [348]. Alternatively, as the hydrolysis proceeds, E_m increases and the products of the second aquation have a positive reduction potential. Thus, reduction may easily occur in particular for PII_{2Cl}^{cis}, which has a similar reduction potential to NAMI-A. These results highlight the role of reduction in driving the hydrolysis of NAMI-A and ICR towards different chemical routes.
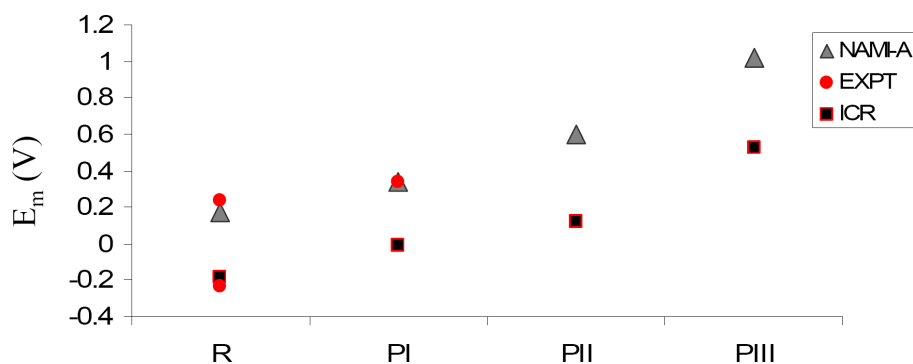
Figure 6.3: Estimated reduction potentials E_m (vs. NHE, V) of NAMI-A, ICR and the products of Cl$^-$ dissociation. Experimental reduction potentials from Ref. [350] and Ref. [332]. See text for further details.

6.5   ICR *vs.* NAMI-A: Biological Implications

The analysis of activation and reaction free energies combined with the estimation of reduction potentials reveals intriguing features of NAMI-A and ICR hydrolyses (Fig. 6.1a, 6.2a).

Under neutral conditions, NAMI-A^{III} hydrolysis leads to the formation of compounds with one DMSO and one Cl$^-$ exchanged by water molecules. In presence of biological reductants, reduction does not take place for the products of DMSO dissociation. On the other hand, NAMI-A and PI_{Cl} have positive reduction potentials and undergo reduction more easily. Therefore, the most abundant NAMI-A metabolites are expected to be the RuII-diaqua and triaqua complexes bearing the DMSO-S ligand. However, we cannot exclude that monoaqua complexes may play a role for its biological activity. Moreover, *in vitro* experiments have shown that poly-oxo species are rapidly formed already during second hydrolysis, whereas *in vivo* the interaction with N- and S- biological donors would prevent polymerization [331]. Either way, products of the third hydrolysis should be less relevant for the biological activity of NAMI-A. Thus, our data stress the relevance of the oxidation state for the hydrolysis path, and support the hypothesis that reduced form of NAMI-A or its hydrolytic products are relevant for the antimetastatic activity [331, 334, 340, 352]. Furthermore, from the analysis of reduction potentials clearly emerged the crucial role of DMSO as it stabilizes RuII species, promoting NAMI-A reduction.

Reduction of ICR is less likely than that of NAMI-A, unless strong biological reductants are available [348]. However, as the hydrolysis proceeds, reduction potential of metabolites increases, i.e. the products of the second hydrolysis have positive reduction potentials. Thus, reduction may require a relatively long time, perhaps long enough for ICR^{III} to undergo hydrolysis, with the most likely products being a

mixture of mono and diaqua complexes (Fig. 6.2a). Once these species are formed, reduction may easily occur, particularly for the *cis* isomer. If reduction occurs, the most abundant metabolites are PII_{2Cl}^{cis} and $PIII_{3Cl}$, holding for the latter similar considerations as for NAMI-A. All these findings are compatible with experimental results, suggesting that ICR binding to biological targets occurs in hours and that only aged ICR solutions can bind to DNA [349].

In addition, the analysis of the second hydrolysis kinetics reveals important differences between ICR and NAMI-A. In particular, ICR (in both RuIII and RuII oxidation states) shows a strong kinetic preference for *cis* diaqua complexes, while both NAMI-A isomers have similar kinetics and thermodynamics. *In vivo* such a kinetic preference may affect the way of binding and also lead the drugs to different biological targets, DNA for ICR and proteins or receptors, for NAMI-A (see for example Ref. [383] and references therein).

6.6 Conclusions

Extensive DFT calculations, with estimation of solvation and entropic effects, were performed to investigate the hydrolysis mechanisms of two promising anticancer drugs, NAMI-A and ICR.

The computational set up has been validated by comparing available experimental and theoretical data with our findings. Predicted and measured reduction potentials of NAMI-A and ICR differ by ~ 0.08 V on average, while calculated activation free energies for the first hydrolysis are within 2 kcal/mol from those determined experimentally. A broad qualitative agreement with experimental findings of NAMI-A and ICR hydrolysis at physiological conditions was achieved.

Our calculations confirmed that NAMI-A is easily reduced *in vivo* and showed that reduction strongly affects the hydrolysis. We suggest *cis* and *trans* RuII-diaqua metabolites to be the most abundant and perhaps the most relevant for the biological activity of NAMI-A. However, we cannot exclude that monoaqua complexes may also be involved.

Reduction of ICR is less likely than that of NAMI-A, yet if occurring would hardly affect the hydrolytic path. Indeed, the most abundant metabolites are the mono and diaqua complexes, with a kinetic preference for *cis* isomer in both oxidation states. Nevertheless, reduction may occur for the products of the second hydrolysis step since reduction potentials increase as the hydrolysis proceeds.

The most relevant NAMI-A and ICR metabolites would retain the main structural and chemical differences of parent compounds, the DMSO-S ligand for the former and Im ligand for the latter. In addition, the kinetic preference for the *cis* diaqua ICR compounds may affect the binding to biomolecules and lead to a different biological behavior. Such features may be of primary importance in determining the diverse

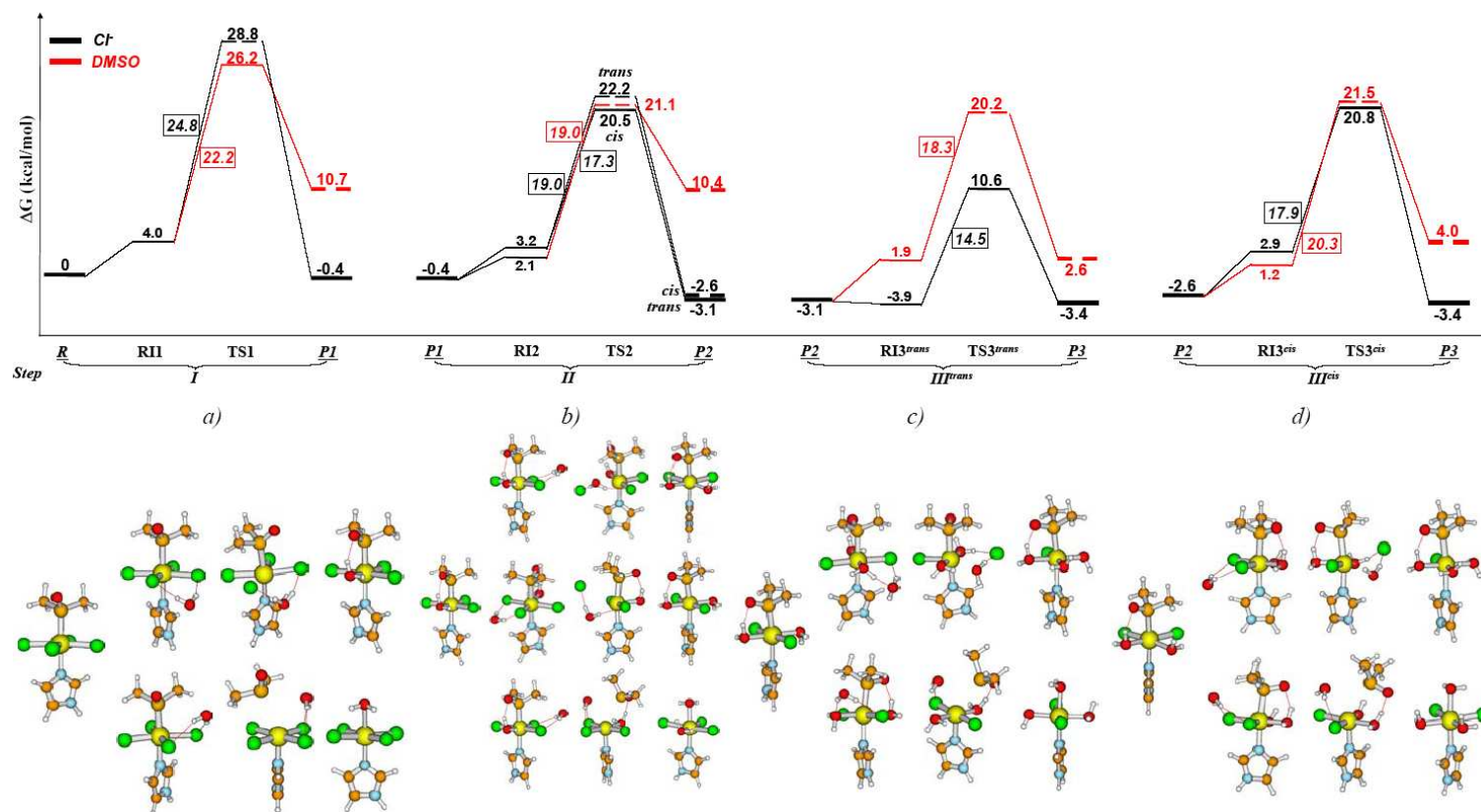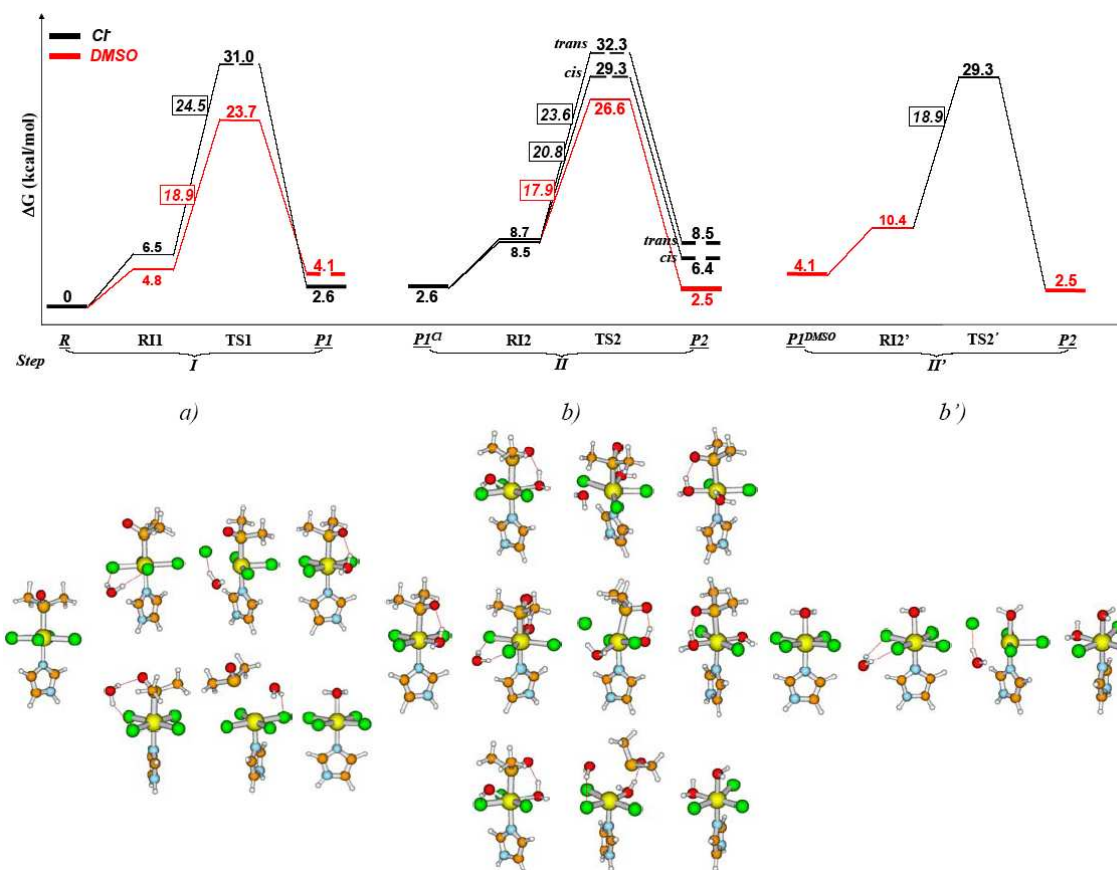pharmacological activity of NAMI-A and ICR anticancer drugs.

Figure 6.4: (continue on the next page) Free energy profiles of Cl$^-$ and DMSO hydrolyses in NAMI-A$^{\mathrm{II}}$: *a)* first hydrolysis; *b)* second hydrolysis; *c)* third hydrolysis starting from PII_{2Cl}^{trans}; *d)* third hydrolysis starting from PII_{2Cl}^{cis}. Black and red lines correspond to Cl$^-$ and DMSO dissociations, respectively, and metabolites without DMSO are indicated by a red line. The most favorable transition states and products are represented by solid lines, while dotted lines are used for the less likely intermediates and TS structures. Activation free energies (kcal/mol) are displayed in the squares along the reaction path. Energies reported on the lines are referred to R.
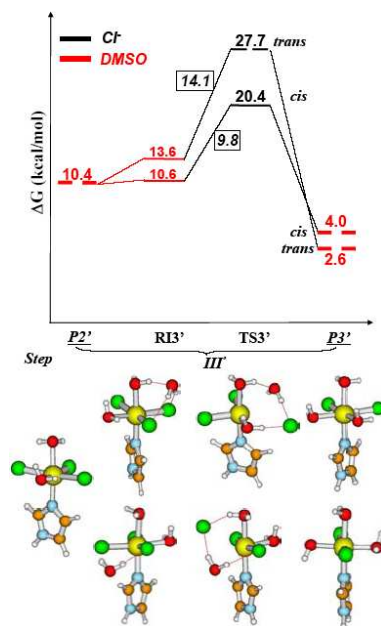
Figure 6.4: (continued) Free energy profiles for the third Cl^- hydrolysis starting from PII_{2Cl}^{trans} NAMI-A$^{\mathrm{II}}$.

Figure 6.5: (continue on the next page) Free energy profiles of Cl^- and DMSO hydrolyses in NAMI-A^{III}: *a)* first hydrolysis; *b)* second hydrolysis from PII_{2Cl}; *b')* second hydrolysis from PII_{DMSO}. For further details see Fig. 6.4.
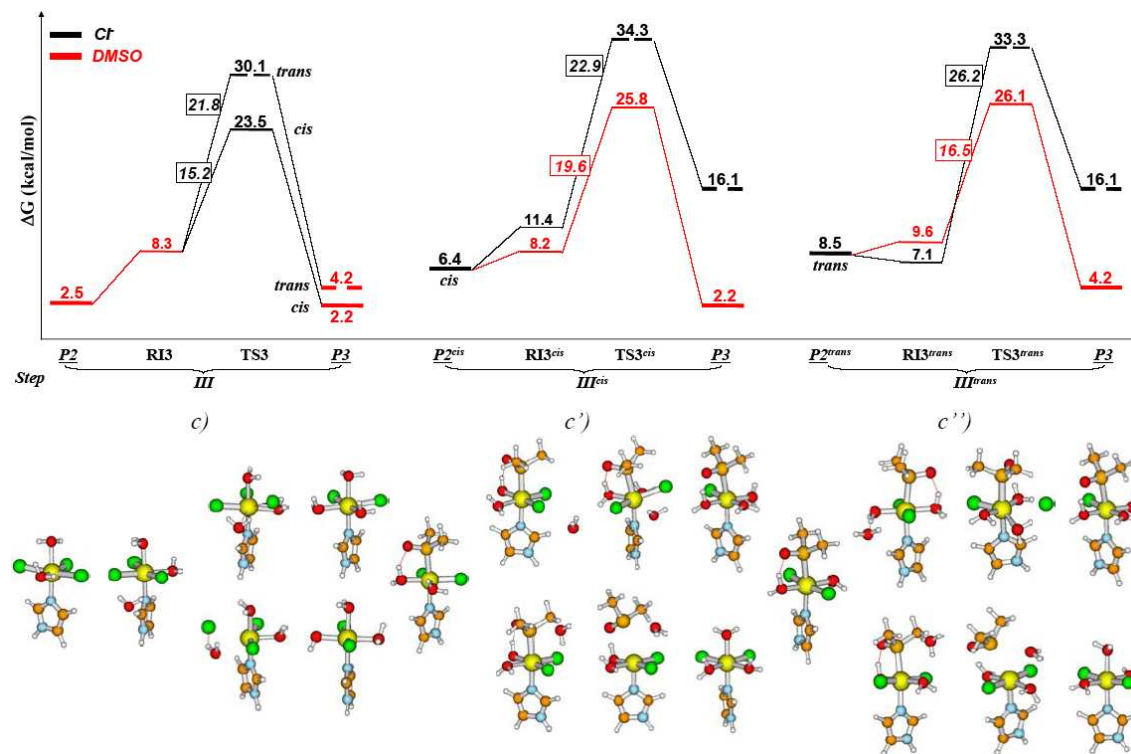
Figure 6.5: (continued) Free energy profiles of Cl$^-$ and DMSO hydrolyses in NAMI-A^{III}: *c)* third hydrolysis from $PII_{2Cl,DMSO}$; *c')* third hydrolysis from PII_{2Cl}^{cis}; third hydrolysis PII_{2Cl}^{trans}.For further details see Fig. 6.4.
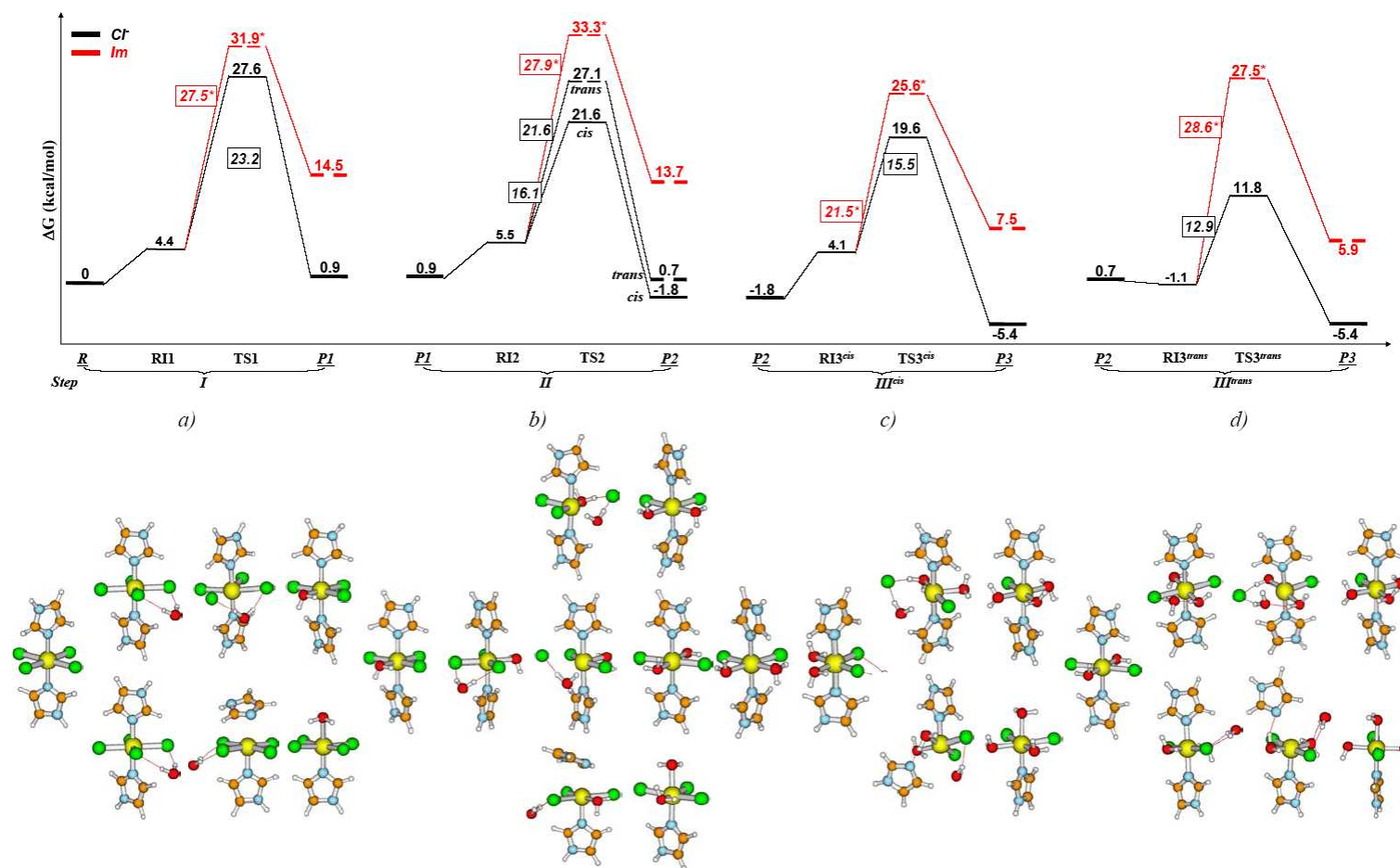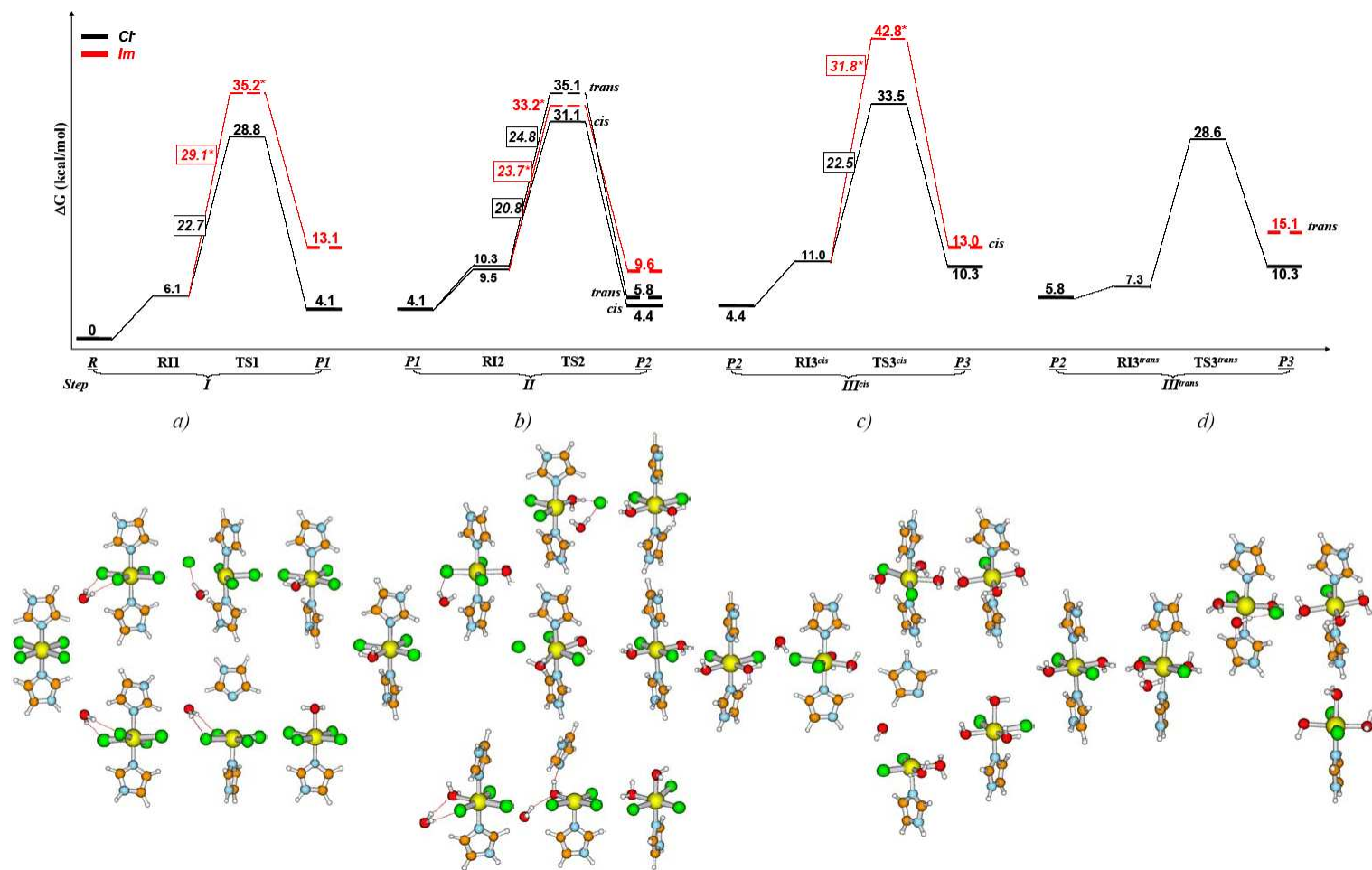
Figure 6.6: Free energy profiles of Cl$^-$ and Im hydrolyses in ICR^{II}: *a)* first hydrolysis; *b)* second hydrolysis; *c)* third hydrolysis starting from PII_{2Cl}^{trans}; *d)* third hydrolysis starting from PII_{2Cl}^{cis}. For further details see Fig. 6.4. * Im pseudo transition states: see Computational Details.

Figure 6.7: Free energy profiles of Cl^- and Im hydrolyses in ICR^{III}: *a)* first hydrolysis; *b)* second hydrolysis; *c)* third hydrolysis starting from PII_{2Cl}^{trans}; *d)* third hydrolysis starting from PII_{2Cl}^{cis}. For further details see Fig. 6.4. * Im pseudo transition states: see Computational Details.

Part IV

Conclusions and Perspectives

In the first part of my thesis work I investigated chemical, thermodynamic and kinetic aspects of drug/DNA molecular recognition and hydrolysis. First, I have focused on the the interaction between DNA and the alkylating agents anthramycin and duocarmycin by classical MD and hybrid Car-Parrinello QM/MM simulations. The electrostatic field due to DNA catalyzes the alkylation reaction by anthramycin, while the solvent plays a minor role in polarizing reactants. Moreover, QM/MM studies allowed to connect this catalytic effect to identification of the most reactive form of this drug. Based on classical MD, it can be suggested that the formation of both the non covalent and covalent drug/DNA adducts does not alter the main structural determinants of oligonucleotide and ligand structures. However, the most reactive form of anthramycin slides by one base-pair step to the site adjacent the reactive guanine. This result prompted for further structural and thermodynamic characterization of the sliding process by umbrella sampling simulations. This is an important step of DNA molecular recognition by small ligands, which can not be restricted to the formation of the final adduct only, but also involves shuffling within the minor groove and translocation between different sites. According to the calculated free energy profile and consistently with results from unbiased simulations, the site adjacent to the reactive one is more stable by a few $kcal/mol$, and the free energy barrier associated to the sliding of the drug is also very small ($\sim 2\,kcal/mol$). Furthermore, free energies corresponding to conformations in which the drug covers other sequences are similar to that associated to the reactive one.

To address the role of the sliding process in recognition and sequence-selectivity, I compared the mechanism exploited by anthramycin to that of the indole derivative of duocarmycin, which is highly selective towards AT-rich tracts. At odd with anthramycin, the alkylation site of duocarmycin is the most stable also in the non covalent adduct, and drug sliding far from it requires to overcome a barrier of $\sim 4\,kcal/mol$. Interestingly, far from the preferred sequence the free energy profile is virtually flat, and the funnel in the free energy landscape mirrors the obtained profile of the solvation free energy. The results allow to speculate about a different molecular recognition mechanisms for the two drugs. Duocarmycin might first bind to DNA at those sequences characterized by the lowest desolvation cost (the non preferred ones) and then easily slide towards the alkylation site. In contrast, recognition of DNA by anthramycin appears to be more complicated as no preferred route exists on the one-dimensional sliding path. Moreover, the reactive site can be reached only upon crossing a significant free energy barrier.

Next, I addressed the unbinding of ligands from DNA. To gain insights into this mechanism, I compared one of the covalent binders previously investigated (anthramycin) to a representative of the main class of non covalent minor groove binders, distamycin A. Free energy calculations using metadynamics pointed out that the detachment of anthramycin from the minor groove involves an active role of waters and a local widening of the minor groove. No stable intermediates are detected during

the simulated unbinding. In contrast, distamycin unbinding involves the formation of an intermediate in which the positively charged tail of the compound lays within the minor groove while its body is solvated. The detachment starts with the disruption of specific ligand/DNA H-bonds and does not involve significant opening of the minor groove or an active role of waters. The calculated activation free energy of detachment of distamycin compares very well with the available experimental data, assessing the reliability of the setup used. In perspective, the natural continuation of this and the previous work involves: *i)* performing similar simulations at different ionic strength to assess its influence on the calculated free energies; *ii* the study of the sliding along different sequences, to assess the effect of sequence-dependent features on the process; the study of the sliding by non covalent binders. *iii)* the study of the binding process (for which different collective variables are needed); *iv)* the study of the way ligands interfere with biological process, starting with the inhibition of helicases by some minor groove binders (the structure of a DNA/helicase complex has been released in june 2007 [384]).

The second part of my thesis concerned the hydrolysis of two anticancer ruthenium compounds, NAMI-A and ICR, the first active against metastases and the other against the primary tumour. Calculations confirmed the experimental fact that the ruthenium center in NAMI-A is easily reduced *in vivo* from oxidation state III to II, with reduction strongly affecting the hydrolysis. The most abundant (and perhaps relevant) hydrolitic products are found to be *cis* and *trans* RuII-diaqua metabolites. Reduction of ICR is less likely than that of NAMI-A, yet if occurring would hardly affect the hydrolytic path. Indeed, the most abundant metabolites are the mono- and diaqua complexes, with a kinetic preference for *cis* isomer in both oxidation states. Nevertheless, reduction may occur for the products of the second hydrolysis step since reduction potentials increase as the hydrolysis proceeds. Importantly, the structural and chemical differences between NAMI-A and ICR, which are at the basis of their different pharmacological activity, are conserved during the hydrolysis steps according to results shown. In addition, the kinetic preference for the *cis* diaqua ICR compounds may affect the binding to biomolecules and lead to a different biological behavior. Such features may be of primary importance in determining the diverse pharmacological activity of NAMI-A and ICR anticancer drugs. A future perspective of this work is the study of the binding of NAMI-A and ICR to their targets [385, 338]. In particular, the interaction between NAMI-A and integrins will be addressed, as these latter have been supposed to be crucial for the development of metastases [339, 386].

Appendix A

The structure of DNA

The structure of DNA fibers was crystallized in the early fifties by Franklin and Wilkins, who discovered that the phosphate-ribose backbone lies at the outside of the DNA and resolved the basic helical structure of the macromolecule. This knowledge was a fundamental key in the breakthrough of DNA structure elucidation by Watson and Crick [387]. 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 [388]. Crick realized that this quantitave relationship could be caused by complementary base-pairing between adenine and thymine or guanine and cytosine, respectively and he envoked the hydrogen bonding pattern now known as Watson-Crick hydrogen bonds (Figure A.1). Based on this information, Watson and Crick proposed a well-known 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 structure lead to the following observations:

- 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 $\sim 36^{\circ}$ with respect to each other. Therefore 10 base pairs realize an entire turn of the double helix. The two strands of the double helix are separated by two distinct grooves: the minor groove and the major groove (Figures A.1 and A.2). The narrow minor groove contains ordered water molecules, which form a characteristic hydration pattern, either a *spine* (A-tracts) or a *ribbon* of hydration (normal DNA-sequences). The
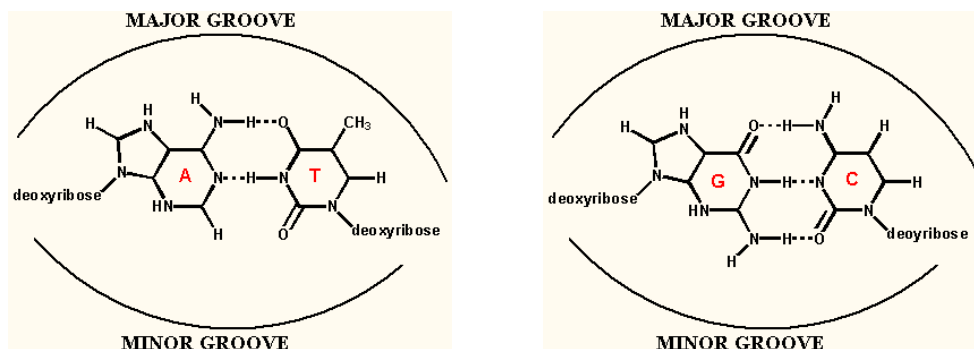
Figure A.1: Watson-Crick base pairing. Left: Adenine (A) with Thymine (T); right: Guanine (G) with Cytosine (C).

	bases per turn	twist per bp [°]	rise per bp [Å]	sugar pucker
B-DNA	10	36	3.3-3.4	C2'-endo
A-DNA	11	33	2.9	C3'-endo
	Minor groove		Major groove	
	width [Å]	depth [Å]	width [Å]	depth [Å]
B-DNA	5.7	7.5	11.7	8.8
A-DNA	11.0	2.8	2.7	13.5

Table A.1: Selected DNA parameters in A and B forms..

correctness of the model was ultimately prooved by X-ray crystallography. In 1952, Franklin and Wilkins discovered by the same technique different forms of DNA, the dry and the wet form, now known as A-DNA and B-DNA (Figure A.2). The dry A-DNA is a right-handed double helix in which hydrophobic stacking interactions between base-pairs are optimized by a $\sim 19°$ inclination of the base-pair plane with respect to the helix axis. This leads to a wider minor groove, which lacks a spine of hydration and a very deep major groove (Table A.1). The conformational change from C2'-endo in B-DNA to C3'-endo in A-DNA 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 A.2). DNA structure has proven to be extremely flexible, and its conformation significantly depends 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, deviations 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
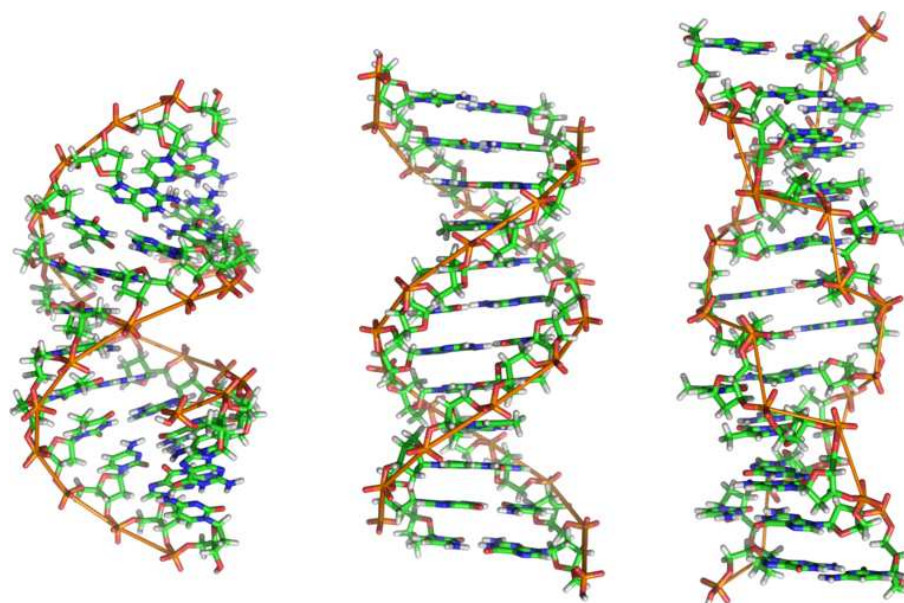
Figure A.2: Schematic drawing of A-DNA (left), B-DNA (middle) and Z-DNA (rigth).

DNA recognition by proteins and drugs [18, 389]. 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 [390] 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 A.3. A nucleobase covalently bound to C1' of a deoxyribose (ribose in RNA) is called a nucleoside, and a nucleoside with one or more phosphate groups attached at C5' is called a nucleotide. 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
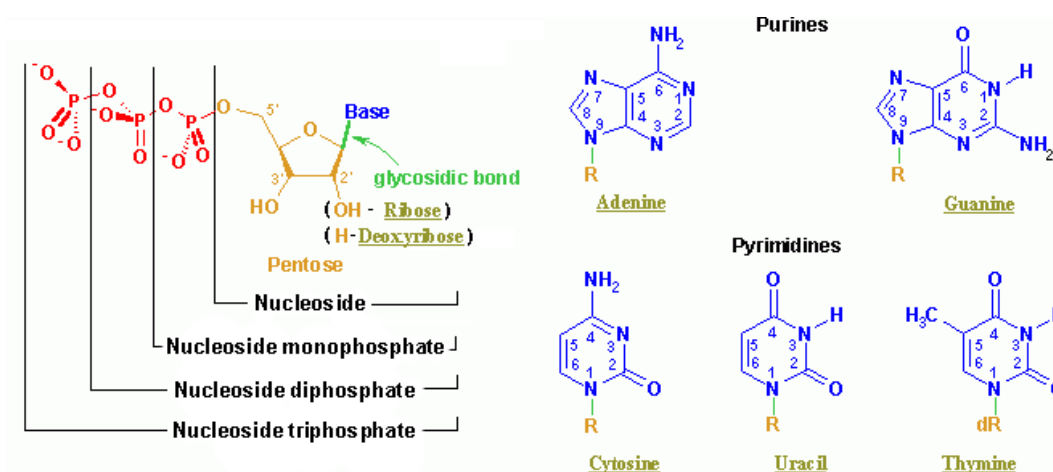
Figure A.3: Structure elements of the most common nucleotides (numbering according to IUPAC). Image from "http://en.wikipedia.org/wiki/Nucleotide".

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 (Figure A.4). A more complete picture is given by the *pseudorotation cycle* as described for example in [19]. Further torsional angles along the phosphoester bonds are called α, β, γ, δ, ϵ and ζ (Figure A.5). The orientation of the bases and base-pairs with respect to each other is described by the *helical parameters* (Figure A.6), which can be divided into groups according to the geometrical features they describe:

- Parameters describing the relative orientation of two bases in the same basepair: 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 a reference 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 diculties in the determination of a straight helix axis. In the work exposed in this thesis I used the program Curves, written by Lavery and Sklenar [276, 239]. In this
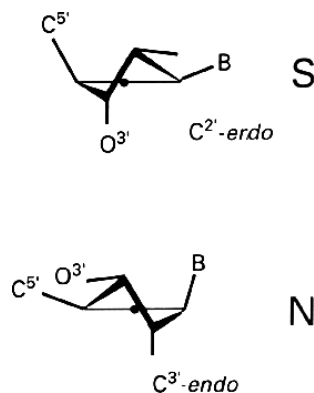
Figure A.4: Sugar puckering in ideal B-DNA (C2'-endo, South) and A-DNA (C3'-endo, North). Taken from "http://www.imb-jena.de/ImgLibDoc/nana/IMAGE_NANA.htm".
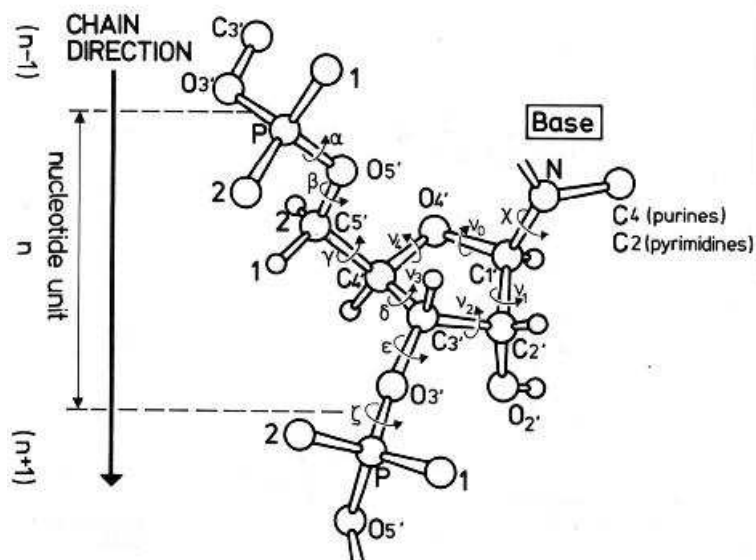

Figure A.5: Atomic numbering scheme and definitions of torsional angles in phospate-sugar moiety in RNA (valid also for DNA, after replacing O2' with H2'). Image from "http://www.imb-jena.de/ImgLibDoc/nana/IMAGE_NANA.htm".
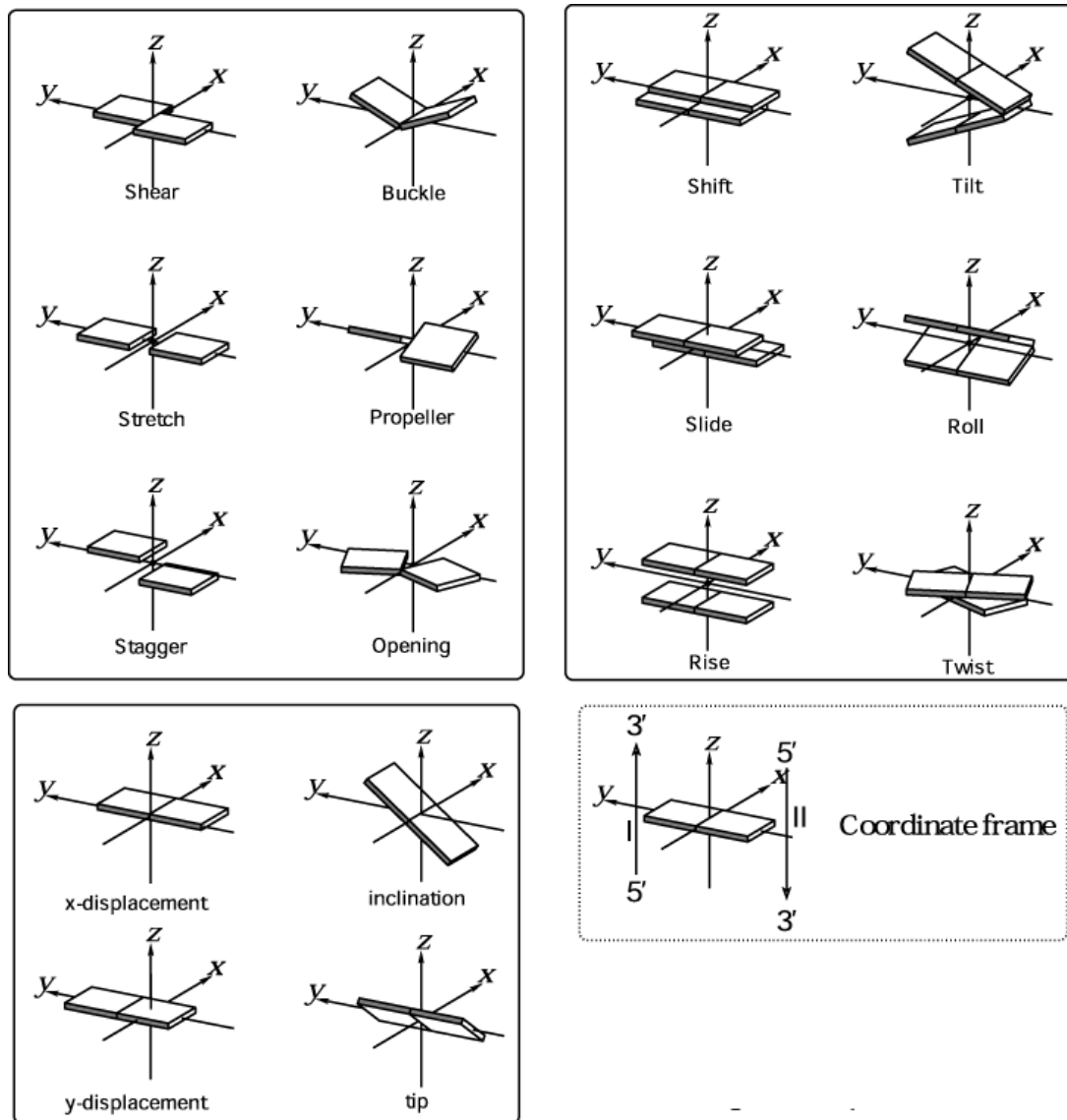
Figure A.6: Schematic drawing of distortion between bases and base-pairs described by helical parameters. Positive values of the designed parameters are shown. Image from "http://rutchem.rutgers.edu/ xiangjun/3DNA/examples.html".

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 quanted by Yanagi *et al* [391] in the *profile sum*. 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, specially in relation to interactions with minor groove binders. The width can be measured between different atoms, P, O4' or C4' atoms being the most used. 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 are reported some standard parameters of canonical A and B-DNA [392].

Appendix B

Statistical Mechanics of non-Hamiltonian systems

Consider a general non-Hamiltonian system whose dynamics is described by the set of equations of motions (e.g. those described by Nosé)

$$\dot{x}^i = \theta^i(x, t) \tag{B.1}$$

subject to initial conditions $x_0 = (x_0^1, ..., x_0^{dN})$, which have solution

$$_t^i = x_t^i(t; x_0^1, ..., x_0^{dN}) \tag{B.2}$$

where N is the number of particles and d the dimensionality. This equation can be seen as a coordinate transformation on the phase space from the initial coordinates at time 0 to those at time t. If we consider a set of trajectories located within the hyphervolume dx_0 at $t = 0$, at time t these will be within dx_t. The variation of the measure with time is given by

$$dx_t = J(x_t; x_0)dx_0 \tag{B.3}$$

where $J(x_t, x_0) = \left| \frac{\partial x_t^1, ..., x_t^{dN}}{\partial x_0^1, ..., x_0^{dN}} \right|$ is the Jacobian of the transformation, *i.e.* the determinant of the matrix T with entries $T_{ij} = \frac{\partial x_t^i}{\partial x_0^j}$. Taking the time derivative of J and using the Jacobi's formula we can write

$$\frac{dJ}{dt} = \text{Tr}(\text{adj}\, T \frac{dT}{dt}) \tag{B.4}$$

where $\text{adj}\, T$ is the adjugate of T, which entries are $(\text{adj}\, T)_{ij} = (-1)^{i+j} M_{ji}$, M_{ij} being the ij-th minor of matrix T. As the dynamics is time-reversable, T will be

invertible, and $\mathrm{adj}\, T = J T^{-1}$. Thus:

$$\frac{dJ}{dt} = J\,\mathrm{Tr}(T^{-1}\frac{dT}{dt}) = J\sum_{i,j=1}^{dN} T_{ij}^{-1}\frac{dT_{ij}}{dt} = J\sum_{i,j=1}^{dN} \frac{\partial x_0^i}{\partial x_t^j}\frac{\partial \dot{x}_t^i}{\partial x_0^j} = J\sum_{i,j,k=1}^{dN} \frac{\partial x_0^i}{\partial x_t^j}\frac{\partial \dot{x}_t^j}{\partial x_t^k}\frac{\partial x_t^k}{\partial x_0^i}$$

(B.5)

where relations $T_{ij}^{-1} = \partial x_0^i/\partial x_t^j$, $dT_{ij}/dt = \partial \dot{x}_t^j/\partial x_0^i$ and the chain rule for the derivative $\partial \dot{x}_t^j/\partial x_0^i$ have been used. Performing the sum over i we get:

$$\sum_{i=1}^{dN} \frac{\partial x_0^i}{\partial x_t^j}\frac{\partial x_t^k}{\partial x_0^i} = \sum_{i=1}^{dN} T_{ij}^{-1} T_{kj} = \sum_{i=1}^{dN} T_{ki}\, T_{ij}^{-1} = (TT^{-1})_{kj} = \delta_{kj} \qquad (\text{B.6})$$

and finally

$$\frac{dJ}{dt} = J\sum_{j,k=1}^{dN} \delta_{jk}\frac{\partial \dot{x}_t^j}{\partial x_t^k} = J\sum_{j=1}^{dN} \frac{\partial \dot{x}_t^j}{\partial x_t^j} = J\nabla_x \cdot \dot{x} = J\kappa(x,t) \qquad (\text{B.7})$$

which has to be solved subject to the initial condition $J(x_0; x_0) = 1$. The quantity $\kappa(x,t) = \nabla_x \cdot \dot{x} = \nabla_x \cdot \theta(x,t)$ is called *phase space compressibility*. Knowledge of the compressibility allows to derive an invariant measure. In the case where κ does not explicitly depends on t, the Jacobian is given by

$$J(x_t; x_0) = \exp[\int_0^t \kappa(x_s)\, ds] \qquad (\text{B.8})$$

By introducing the variable $\omega(x) = \ln J$, which gives the variation in time of the compressibility, $\dot{\omega} = \kappa$, the above equation can be rewritten as:

$$J(x_t; x_0) = \exp[\omega(x_t) - \omega(x_0)] \qquad (\text{B.9})$$

Thus, the following conservation law holds:

$$e^{-\omega(x_t)}dx_t = e^{-\omega(x_0)}dx_0 \qquad (\text{B.10})$$

This is a generalized version of the Liouville theorem of incompressibility of the phase space. It can be shown that the above formalism holds also if κ depends explicitly on time. Thus, the measure $e^{-\omega(x_t,t)}dx_t$ is conserved by the dynamics, not simply dx_t. The exponential $e^{-\omega(x_t,t)}$ can be viewed as a metric factor $\sqrt{g(x,t)}$, satisfing the metric transformation rule

$$\sqrt{g(x_0,0)} = J(x_t; x_0)\,\sqrt{g(x_t,t)} \qquad (\text{B.11})$$

The existence of an invariant measure means that time and *ensemble* averages are still equivalent for a non-Hamiltonian dynamics (under the assumption of ergodicity),

with the proper metric taken into account when evaluating integrals. Furthermore, a *generalized Liouville equation* holds:

$$\frac{\partial f \sqrt{g}}{\partial t} + \nabla_x \cdot (f \sqrt{g} \dot{x}) = 0 \tag{B.12}$$

which, noting that $d\sqrt{g(x,t)}/dt \equiv \partial \sqrt{g}/\partial t + \nabla \sqrt{g} \cdot \dot{x} = -\kappa \sqrt{g}$, gives the conservation law for f equivalent to the Hamiltonian case:

$$\frac{\partial f(x,t)}{\partial t} + \nabla_x \cdot \dot{x}\, f(x,t) \equiv \frac{df}{dt} = 0 \tag{B.13}$$

Finally, this equation is combined with the conservation law for the measure, we obtain a general statement of Liouville's theorem:

$$f(x_0, 0)\sqrt{g(x_t, t)}dx_0 = f(x_t, t)\sqrt{g(x_t, t)}dx_t \tag{B.14}$$

In the absence of external and explicitly time-dependent forces, a non-Hamiltonian microcanonical ensemble can be defined on the basis of the existence of the invariant measure. If the dynamical system has a set of M conserved quantities $K_\lambda(x) = \bar{K}_\lambda, \lambda = 1, ..., M$, the microcanonical distribution function f can be expressed as:

$$f(x) = \Pi_{\lambda=1}^{M} \delta(K_\lambda(x) - \bar{K}_\lambda) \tag{B.15}$$

and the corresponding partition function is

$$\Omega(N, V, \bar{K}_1, ..., \bar{K}_M) = \int dx \sqrt{g(x)} \Pi_{\lambda=1}^{M} \delta(K_\lambda(x) - \bar{K}_\lambda) \tag{B.16}$$

Statistical averages can be thus computed with respect to coordinates x at any time:

$$\langle A \rangle_t = \frac{\int dx \sqrt{g(x,t)} A(x) f(x,t)}{\int dx \sqrt{g(x,t)} f(x,t)} \tag{B.17}$$

Consider now the non-Hamiltonian Nosé equations of motion. The compressibility is given by

$$\kappa = \nabla_x \cdot \dot{x} = \sum_{i=1}^{dN} \frac{\partial \dot{p}_i}{\partial p_i} = -dN\dot{\eta} \tag{B.18}$$

so that

$$\omega = \int_0^t \kappa(s)\, ds = -dN \int_0^t \dot{\eta}(s)\, ds = -dN\eta \tag{B.19}$$

and the metric factor is

$$\sqrt{g} = e^{dN\eta} \tag{B.20}$$

Thus, the conserved measure in the phase space of Nosé dynamics is

$$d\mu = d^{dN}r\, d^{dN}p\, d\eta\, dp_\eta\, e^{dN\eta} \tag{B.21}$$

For further details and generalization to Nosé-Hoover chains, see e.g. Refs. [393, 394, 395].

Appendix C

Ewald sum method

Consider a neutral system composed of N positively and negatively particles of charge q_i and placed at $\mathbf{r}_i$, enclosed in a volume V. Within *pbc* each particle interacts with all the other $N-1$ particles into the simulation box and with all the N particle images in an infinite 3D array of periodic cells. The electrostatic potential energy U^{Coul} of the infinite system, therefore, takes the form:

$$U^{Coul} = \frac{1}{2}\sum_{i=1}^{N} q_i \phi(r_i) = \frac{1}{2}\sum_{\mathbf{n}=0}^{\prime}\sum_{ij}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij}+\mathbf{n}L|} \tag{C.1}$$

where $L = V^{-1/3}$ is the length of the periodic box (assumed cubic for simplicity) and $\mathbf{n}$ are the direct lattice vectors. The prime on the summation indicates that the sum is over all the images $\mathbf{n}$, and $i \neq j$ if $\mathbf{n} = 0$. The convergence of this series is conditional, its value depending on the way the cluster of simulation boxes is built to fill the space, *i. e.* from its shape, and the boundary conditions (vacuum or dielectric [135]) The slow convergence of the series requires a very large number of images to achieve a reliable estimate of U^{Coul}. The idea behind the Ewald method is to surround every charge q_i with a density cloud ϱ_- of opposite sign and total charge $-q_i$. In this way an efficient screening is realized, so that interactions rapidly go to 0 and direct summation is possible. Then, a *cancellation* charge density ϱ_+ of total charge q_i centered at $\mathbf{r}_i$ is also added to recover the original distribution. In order to exclude self-interactions the contributions of these three charge densities should be not evaluated in $\mathbf{r}_i$. However, it is convenient to keep self-interactions due to the cancelling charge distribution ϱ_+, since ϱ_+ is in this way periodic and can be represented as a rapidly converging Fourier sum. The spurios self-interaction can be easily subtracted separetely. In the original formulation [136] Ewald choose two Gaussian charge distributions, $\varrho_-^G(\mathbf{r})$ and $\varrho_+^G(\mathbf{r})$ of width $\sqrt{2/\alpha}$ and total charge $-q_i$ and q_i, respectively.

Thus, the total charge distribution of the system $\varrho(\mathbf{r})$ may be rewritten as:

$$\varrho(\mathbf{r}) = \varrho(\mathbf{r}) + \varrho_-^G(\mathbf{r}) + \varrho_+^G(\mathbf{r})$$

$$= \sum_{\mathbf{n}} \left[\left(\sum_{i=1}^{N} q_i \delta(\mathbf{r} - (\mathbf{r_i} + \mathbf{nL})) + \sum_{i=1}^{N} (-q_i) \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha|\mathbf{r} - (\mathbf{r_i} + \mathbf{nL})|^2} \right) \right. $$
$$\left. + \sum_{i=1}^{N} q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha|\mathbf{r} - (\mathbf{r_i} + \mathbf{nL})|^2} \right] \qquad (C.2)$$

At this point, $\varrho + \varrho_-^G$ and ϱ_+^G contributions to the total charge distribution can be considered separately. From the first term one gets a potential energy U^{dir} rapidly going to 0 at large r, due only to the fraction of q_i unscreened by ϱ_-^G. From ϱ_+^G the long-range interaction energy U^{rec} comes, that can be evaluated in the reciprocal space. To this aim, we derive the electrostatic potential in this space through Poisson's equation $\nabla^2 \phi(\mathbf{r}) = -4\pi\varrho(\mathbf{r})$. Fourier transforming the members of this equation gives:

$$\nabla_{\mathbf{r}}^2 \phi(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{k}} \tilde{\phi}(\mathbf{k}) \nabla_{\mathbf{r}}^2 e^{i\mathbf{k}\cdot\mathbf{r}} = -\frac{1}{V} \sum_{\mathbf{k}} k^2 \tilde{\phi}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \qquad (C.3)$$

$$\varrho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{k}} \tilde{\varrho}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \qquad (C.4)$$

where

$$\tilde{f}(\mathbf{k}) = \int_V d\mathbf{r} f(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \qquad (C.5)$$

and $\mathbf{k} = (2\pi/L)\mathbf{l}$, $\mathbf{l}$ being reciprocal lattice vectors. The Poisson's equation takes the particularly simple form:

$$k^2 \tilde{\phi}(\mathbf{k}) = 4\pi \tilde{\varrho}(\mathbf{k}) \qquad (C.6)$$

Substituting the expression for $\varrho_+^G(\mathbf{r})$ in eq. C.5 we have:

$$\tilde{\varrho}_+^G(\mathbf{k}) = \int_V d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{n}} \sum_{j} \left(\frac{\alpha}{\pi}\right)^{3/2} q_j e^{-\alpha|\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|^2}$$

$$= \sum_{j} \left(\frac{\alpha}{\pi}\right)^{3/2} q_j \int_V d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{n}} e^{-\alpha|\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|^2} \qquad (C.7)$$

$$= \int_{allspace} d\mathbf{r} \sum_{j} \left(\frac{\alpha}{\pi}\right)^{3/2} q_j e^{-i\mathbf{k}\cdot\mathbf{r}} e^{-\alpha|\mathbf{r} - \mathbf{r}_j|^2}$$

where the relation $\sum_n \int_{-a}^{a} f(x+na) = \int_{-\infty}^{\infty} f(x)$ for periodic functions has been used. With the change of variables $\mathbf{r} \rightarrow \mathbf{r}' = \mathbf{r} - \mathbf{r}_j$ we obtain finally the expression for $\tilde{\varrho}_+^G$ in the reciprocal space:

$$
\begin{aligned}
\tilde{\varrho}_+^G(\mathbf{k}) &= \sum_j q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} \left[\left(\frac{\alpha}{\pi}\right)^{3/2} \int_{allspace} d\mathbf{r}' e^{-i\mathbf{k}\cdot\mathbf{r}'} e^{-\alpha r'^2} \right] \\
&= \sum_j q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} e^{-k^2/4\alpha} = \tilde{\varrho}_P(\mathbf{k})\tilde{\gamma}(\mathbf{k})
\end{aligned}
\tag{C.8}
$$

where $\tilde{\varrho}_P(\mathbf{k}) = \sum_j q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j}$ and $\tilde{\gamma}(\mathbf{k}) = e^{-k^2/4\alpha}$ are the Fourier transform of a point charges set and of the smearing Gaussian, respectively [1]

Inserting in Poisson's equation the potential takes the form:

$$
\tilde{\phi}_+^G(\mathbf{k}) = \tilde{g}(\mathbf{k})\tilde{\varrho}_P(\mathbf{k})\tilde{\gamma}(\mathbf{k})
\tag{C.9}
$$

where $\tilde{g}(\mathbf{k}) = \frac{4\pi}{k^2}$ is the Green's function. Note that $\tilde{\phi}_+^G(\mathbf{k})$ is defined only for $\mathbf{k} \neq 0$. This arises from the conditional convergenge of the series and is related to the assumption that the surrounding of our periodic system is an ideal conductor

---

[1] This is a property of the Fourier transform of a convolution ($\star$). If $a(x) = b(x) \star c(x) \equiv \int dx' b(x')c(x-x')$, the Fourier transform $\tilde{a}$ is simply the product of the Fourier transform $\tilde{b}$ and $\tilde{c}$. The charge density $\tilde{\varrho}_+^G(\mathbf{k})$ is the Fourier transform of

$$
\tilde{\varrho}_+^G(\mathbf{r}) = \sum_j q_j e^{-\alpha|\mathbf{r}-\mathbf{r}_j|^2} = \int d\mathbf{r}' \overbrace{e^{-\alpha r'^2}}^{\gamma(\mathbf{r}')} \overbrace{\sum_j q_j \delta\left(\mathbf{r}' - (\mathbf{r} - \mathbf{r}_j)\right)}^{\varrho_P(\mathbf{r}-\mathbf{r}')} = \gamma(\mathbf{r}) \star \varrho_P(\mathbf{r})
$$

Thus

$$
\tilde{\varrho}_+^G(\mathbf{k}) = \tilde{\varrho}_P(\mathbf{k})\tilde{\gamma}(\mathbf{k}) \,.
$$

$(\varepsilon = \infty)^2$. Antitrasforming the potential gives the electrostatic energy:

$$
\begin{aligned}
U^{rec} &= \frac{1}{2}\sum_i q_i \phi_+^G(\mathbf{r}_i) = \frac{1}{2}\sum_i q_i \left[\frac{1}{V}\sum_{\mathbf{k}\neq 0}\tilde{\phi}_+^G(\mathbf{k})\right] \\
&= \frac{2\pi}{V}\sum_{\mathbf{k}\neq 0}\sum_{ij}\frac{e^{-k^2/4\alpha}}{k^2}q_i q_j e^{-i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)} = \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\tilde{g}(\mathbf{k})\left|\tilde{\varrho}_+^G(\mathbf{k})\right|^2 e^{-k^2/4\alpha}
\end{aligned}
\tag{C.10}
$$

As mentioned above, this term contains the self-interaction of each charge q_i with a Gaussian distribution $\varrho_+^{self} = q_i(\alpha/\pi)^2 e^{-\alpha|\mathbf{r}-\mathbf{r}_i|^2}$. The potential due to this cloud is obtained solving the Poisson's equation in spherical coordinates ($x \equiv |\mathbf{r} - \mathbf{r}_i|$):

$$
\frac{1}{x} = \frac{\partial x \phi_+^{self}}{\partial x^2} = 4\pi\varrho_+^{self} \rightarrow \phi_+^{self} = \frac{q_i}{x}\left(\frac{2}{\sqrt{\pi}}\right)\int_0^{\sqrt{\alpha}x}e^{-x'^2}dx' = \frac{q_i}{x}\mathrm{erf}(\sqrt{\alpha}x)
\tag{C.11}
$$

where $\mathrm{erf}(x) = (2/\sqrt{(\pi)})\int_0^x \exp(-u^2)du$. Finally, we get:

$$
U^{self} = \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}}\sum_{i=1}^N q_i^2
\tag{C.12}
$$

The advantage of this scheme are now clear: the self-contribution energy is simply a constant, independent on system configuration, and does not enter in the calculation of the forces. Using the result for ϕ^{self} it is easy to calculate the short-range interaction energy U^{dir}. Indeed, this energy arises from a potential generated by charge densities $q_i - \varrho_-^G(\mathbf{r})$ centered at $\mathbf{r}_i$, which thus is given at any point $\mathbf{r}$ by:

$$
\phi^{dir}(\mathbf{r}) = \sum_{\mathbf{n}}\sum_j \frac{q_j}{|\mathbf{r}-(\mathbf{r}_j+\mathbf{n}L)|}\mathrm{erfc}\left(\sqrt{\alpha}\,|\mathbf{r}-(\mathbf{r}_j+\mathbf{n}L)|\right)
\tag{C.13}
$$

---

[2]In the case where the system is surrounded by a medium of finite dielectric constant ε_s dipoles will be present on its surface. A formula has been derived relating the energy U^{Coul} we are calculating here to that corresponding to a finite ε_s:

$$
U(\infty) = U(\varepsilon_s) - U^{pol} = U(\varepsilon_s) - \frac{2\pi}{(2\varepsilon_s+1)V}\left|\sum_i q_i\mathbf{r}_i\right|
$$

The term U^{pol} corresponds to the neglected $k \neq 0$ contribution. For further details on this point see [119] and Refs. therein.

The total electrostatic interaction energy can be finally expressed as:

$$
\begin{aligned}
U_{E}^{Coul}w &= \frac{1}{2} \sum_{i} q_i \phi^{tot}(\mathbf{r}_i) \\
&= U^{dir} - U^{self} + U^{rec} = \frac{1}{2} \sum_{i} q_i \left[\phi^{dir}(\mathbf{r}_i) + \phi^{rec}(\mathbf{r}_i) - \phi^{self}(\mathbf{r}_i) \right] \\
&= \frac{1}{2} \sum_{\mathbf{n}}{}' \sum_{ij}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|} \mathrm{erfc}\left(\sqrt{\alpha}\, |\mathbf{r}_{ij} + \mathbf{n}L| \right) \\
&+ \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \sum_{ij} \frac{e^{-k^2/4\alpha}}{k^2} q_i q_j e^{-i\mathbf{k}\cdot(\mathbf{r} - \mathbf{r}_j)} \\
&- \left(\frac{\alpha}{\pi} \right)^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2
\end{aligned}
\tag{C.14}
$$

Convergence and Accuracy. It is important to remind that, whereas ϕ depends linearly on the charge density ϱ, the electrostatic energy does not. Thus, U^{dir} and $U^{rec} - U^{self}$ are not the electrostatic energies due to $\varrho + \varrho_-^G$ and ϱ_+^G, as the potential these charge distribution generate is evaluated on the original ϱ. Consistently, they have been called real and Fourier space contributions to U^{Coul}, and their sum is would not give the energy of the original charge density ϱ. The accuracy of the method, defined as the difference $\epsilon = U_{Ew}^{Coul} - U^{Coul}$, is related to the optimal choice of the parameters α, width of the Gaussian distribution, and of the two cutoffs r_c and k_c in real and reciprocal space, respectively. It can be shown that ϵ for both the real and reciprocal space terms is proportional to $\exp(-s^2)/s^2$ [119], where $s \equiv r_c \alpha$ and $s \equiv \pi n_c / \alpha L$, respectively. Here n_c is a positive integer used to define the cutoff in the reciprocal space vectors as $k_c = 2\pi/Ln_c$, so that the number of Fourier components within this cutoff is $(4\pi/3)n_c^3$. If one requires the same accuracy on the evaluation of U^{dir} and U^{rec} the value of s is fixed, and the choice of α determines both the number of Fourier components required to calculate U^{rec} and the cutoff r_c. The optimal choice of α is obtained minimizing the total computational time

$$
\tau = \tau^{dir} N^{dir} + \tau^{rec} N^{rec}
\tag{C.15}
$$

where τ^{dir} is the time needed to evaluate the interaction between a couple of particles, τ^{rec} that needed to evaluate the interaction per k vector and per particle, N^{dir} and N^{rec} are the number of cycles per MD step needed for the evaluation of U^{dir} and U^{rec}, respectively. For a homogeneous system of density $D = N/L^3$ we have

$$N^{dir} = \frac{4}{3}\pi\left(\frac{r_c}{L}\right)^3 N^2 = \frac{4}{3}\pi\left(\frac{s}{\alpha L}\right)^3 N^2 \tag{C.16}$$

$$N^{rec} = \frac{4}{3}\pi n_c^3 N = \frac{4}{3}\pi\left(\frac{s\alpha L}{\pi}\right)^3 N \tag{C.17}$$

which gives for α and τ the values:

$$\alpha = \left(\frac{\tau^{dir}\pi^3 N}{\tau^{rec}L^6}\right)^{1/6} \sim N^{-1/6} \tag{C.18}$$

$$\tau = \frac{8\sqrt{\tau^{dir}\tau^{rec}}N^{3/2}s^3}{3\sqrt{\pi}} = \mathcal{O}(N^{3/2}) \tag{C.19}$$

C.1   Particle-Mesh Approaches

The computational cost for solving eq. C.14 thus scales as $\mathcal{O}(N^{3/2})$. In general, r_c can be safely set to a value $< L/2$, restricting the calculation of U^{dir} to the original box only ($|\mathbf{n}| = 0$). However, for convenience r_c is sometimes set equal to the cut-off for short-range interactions, and in this case α does not need to have its optimum value. This implies that, although the real part of the Ewald sum goes as $\mathcal{O}(N)$, the Fourier part goes as $\mathcal{O}(N^2)$. To see this, simply substitute $s = \alpha r_c$ in eqs. C.16, C.17 and recall that $N = DL^{-3}$ and $N_c \sim Dr_c^{-3}$, where N_c is the number of particles within the sphere of radius r_c. Either or noth α takes its optimum value, the computational cost of the Ewald sum becomes prohibitive for large ($N > 10^4$). To cope with this problem the so-called Particle-Mesh Approaches have been developed [119], in which the point charges are spread on a space grid. This leads to a Poisson equation expressed through Discrete Fourier Transform (DFT), which is solved using the Fast Fourier Transforms (FFT) [139], scaling as $\mathcal{O}(M \log_2 M)$, where M is the number of grid points or Discrete Fourier Components (in practice $\sim N$?). Since the FFT is a grid transformation, there are obvious discretization problems to be solved (*in primis* those of charge assignement and calculation of forces from grid points back to atoms) and corresponding discretization errors to be minimized. In this sense an improvement in the accuracy has been made by further splitting interactions in the reciprocal space into a short and long range term, and evaluating the first directly as particle-particle and the second through the use of a mesh (Particle-Particle Particle-Mesh, PPPM or P³M [138]). For excellent discussions and comparisons among various Particle-Mesh approaches the reader can refer to the book by Frenkel and Smit [119] and the two articles by Deserno and Holm [396,397]. What should be kept in mind is that Particle-Mesh approaches have been introduced as the allow to implement Ewald summation

using the FFT, not for reasons of accuracy. Without their advent, however, computer simulations would have never found spread use in biological problems.

Bibliography

[1] W. W. Gibbs. Untangling the roots of cancer. *Sci. Am. Magazine*, July, 2003.

[2] W. C. Hahn and R. A. Weinberg. Rules for making human tumor cells. *N. Engl. J. Med.*, 347:1593–1603, 2002.

[3] Prasad V. Jallepalli and Christoph Lengauer. Chromosome segregation and cancer: cutting through the mystery. *Nat Rev Cancer*, 1(2):109–117, November 2001.

[4] L. A. Loeb, K. R. Loeb, and J. P. Anderson. Multiple mutations and cancer. *Proc. Natl. Acad. Sci. USA*, 100(3):776–781, 2003.

[5] M. Esteller. Cancer epigenomics: Dna methylomes and histone-modification maps. *Nat. Rev. Genet.*, 8:286–298, 2007.

[6] C. A. M. La Porta. Cellular targets for anticancer strategies. *Curr. Drug Targets*, 5(4):347–355, May 2004.

[7] S. Neidle and D. E. Thurston. Chemical approaches to the discovery and development of cancer therapies. *Nature Rev. Cancer*, 5(4):285–296, 2005.

[8] B. A. Chabner and T. G. Roberts. Timeline: Chemotherapy and the war on cancer. *Nat Rev Cancer*, 5(1):65–72, Jan 2005.

[9] S. Neidle. Dna minor-groove recognition by small molecules. *Nat. Prod. Rep.*, 18(3):291–309, 2001.

[10] Enrique Espinosa, Pilar Zamora, Jaime Feliu, and Manuel González Barón. Classification of anticancer drugs–a new system based on therapeutic targets. *Cancer Treat Rev*, 29(6):515–523, Dec 2003.

[11] C. J. Suckling. Minor groove binders 1998-2004. *Exp. Opin. Ther. Patents*, 14(12):1693–1724, 2004.

[12] M. Gniazdowski, W. A. Denny, S. M. Nelson, and M. Czyz. Effects of anti-cancer drugs on transcription factor-dna interactions. *Exp. Op. Ther. Targets*, 9(3):471–489, June 2005.

[13] X. G. Xi. Helicases as antiviral and anticancer drug targets. *Curr. Med. Chem.*, 14:883–915, 2007.

[14] U. Pindur, M. Jansen, and T. Lemster. Advances in dna-ligands with groove binding, intercalating and/or alkylating activity: chemistry, dna-binding and biology. *Curr Med Chem*, 12(24):2805–2847, 2005.

[15] P. G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M. A. Tabrizi, M. G. Pavani, and R. Romagnoli. Dna minor groove binders as potential antitumor and antimicrobial agents. *Med. Res. Rev.*, 24:475–528, 2004.

[16] E. C. Friedberg. How nucleotide excision repair protects against cancer. *Nat Rev Cancer*, 1(1):22–33, Oct 2001.

[17] W. P. Roos and B. Kaina. Dna damage-induced cell death by apoptosis. *Trends Mol. Med.*, 12(9):440–450, September 2006.

[18] S. Neidle. Crystallographic insights into dna minor groove recognition by drug. *Biopolymers*, 44:105–121, 1997.

[19] S. Neidle. *Nucleic acid structure and recognition*. Oxford University Press, New York, 2002.

[20] A. D. Mackerell. Empirical force fields for biological macromolecules: Overview and issues. *J. Comp. Chem.*, 25:1584–1604, 2004.

[21] A. N. Lane and M. J. Forster. Determination of internal dynamics of deoxyri-boses in the dna hexamer d(cgtacg)2 by 1h nmr. *Eur Biophys J*, 17(4):221–232, 1989.

[22] D. G. Gorenstein. Conformation and dynamics of dna and protein-dna complexes by [31]p nmr. *Chem. Rev.*, 94:1315–1338, 1994.

[23] R. J. Isaacs and H. P. Spielmann. Nmr evidence for mechanical coupling of phosphate b(i)-b(ii) transitions with deoxyribose conformational exchange in dna. *J. Mol. Biol.*, 311(1):149–160, Aug 2001.

[24] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Jr. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[25] T. E. Cheatham III, P. Cieplak, and P. A. Kollman. A modified version of the cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, 16:845–862, 1999.

[26] A. D. MacKerell, J. Wiorkiewicz-Kuczera, and M. Karplus. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.*, 117(48):11946–11975, 1995.

[27] N. Foloppe and A. D. Jr. MacKerell. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, 21(2):86–104, 2000.

[28] A. D. MacKerell and N. K. Banavali. All-atom empirical force field for nucleic acids: Ii. application to molecular dynamics simulations of dna and rna in solution. *J. Comp. Chem.*, 21(2):105–120, 2000.

[29] D. R. Langley. Molecular dynamic simulations of environment and sequence dependent dna conformations: the development of the bms nucleic acid force field and comparison with experimental results. *J. Biomol. Struct. Dyn.*, 16(3):487–509, Dec 1998.

[30] T. A. Soares, P. H. Hünenberger, M. A. Kastenholz, V. Kräutler, T. Lenz, R. D. Lins, C. Oostenbrink, and W. F. van Gunsteren. An improved nucleic acid parameter set for the gromos force field. *J. Comp. Chem.*, 26(7):725–737, May 2005.

[31] A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham III, C. A. Laughton, and M. Orozco. Refinement of the amber force field for nucleic acids: Improving the description of α/γ conformers. *Bioph. J.*, 92(11):3817–3829, 2007.

[32] A. Lauria, A. Montalbano, P. Barraja, G. Dattolo, and A. M. Almerico. Dna minor groove binders: an overview on molecular modeling and qsar approaches. *Curr. Med. Chem*, 14(20):2136–2160, 2007.

[33] R. Lavery. Recognizing dna. *Quart. Rev. Biophys.*, 38(4):339–344, November 2005.

[34] T. E. Cheatham III. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.*, 14:360–367, 2004.

[35] M. Orozco, A. Noy Perez, and F. J. Luque. Theorethical methods for the simulation of nucleic acids. *Chem. Soc. Rev.*, 32:350–364, 2003.

[36] D. L. Beveridge and K. J. McConnell. Nucleic acids: theory and computer simulation, y2k. *Curr. Opin. Struct. Biol.*, 10:182–196, 2000.

[37] S. A. Shaikh, S. R. Ahmed, and B. Jayaram. A molecular thermodynamic view of dna-drug interactions: a case study of 25 minor-groove binders. *Arch. Bioch. Bioph.*, 429(1):81–99, September 2004.

[38] E. Giudice and R. Lavery. Simulation of nucleic acids and their complexes. *Acc. Chem. Res.*, 5:350–357, 2002.

[39] J. Srinivasan, T. E. Cheatham III, P. Cieplak, P. A. Kollman, and D. A. Case. Continuum solvent studies of the stability of dna, rna and phosphoramidate-dna helices. *J. Am. Chem. Soc.*, 120:9401–9409, 1998.

[40] T. E. Cheatham III, J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. 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.*, 117(14):4193–4194, April 1995.

[41] T. E. Cheatham and B. R. Brooks. Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution. *Theor. Chem. Acc.*, 99(5):279–288, September 1998.

[42] A. Noy, A. Pérez, F. Lankas, F. J. Luque, and M. Orozco. Relative flexibility of dna and rna: a molecular dynamics study. *J. Mol. Biol.*, 343(3):627–638, Oct 2004.

[43] M. A. Kastenholz, T. U. Schwartz, and P. H. Hunenberger. The transition between the b and z conformations of dna investigated by targeted molecular dynamics simulations with explicit solvation. *Biophys. J.*, 91(8):2976–2990, October 2006.

[44] P. Cieplak, T. E. Cheatham, and P. A. Kollman. Molecular dynamics simulations find that 3' phosphoramidate modified dna duplexes undergo a b to a transition and normal dna duplexes an a to b transition. *J. Am. Chem. Soc.*, 119(29):6722–6730, July 1997.

[45] T. E. Cheatham III and P. A. Kollman. Observation of the a-dna to b-dna transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.*, 259(3):434–444, June 1996.

[46] F. Lankas, J. Sponer, J. Langowski, and T. E. Cheatham. Dna basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.*, 85(5):2872–2883, November 2003.

[47] F. Lankas, J. Sponer, J. Langowski, and T. E. Cheatham. Dna deformability at the base pair level. *J. Am. Chem. Soc.*, 126(13):4124–4125, April 2004.

[48] S. B. Dixit, D. L. Beveridge, D. A. Case, T. E. Cheatham, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, K. M. Thayer, and P. Varnai. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of dna oligonucleotides. ii: Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.*, 89(6):3721–3740, December 2005.

[49] B. Hartmann and R. Lavery. Dna structural forms. *Quart. Rev. Biophys.*, 29(4):309–368, December 1996.

[50] B. Hartmann, D. Piazzola, and R. Lavery. Bi-bii transitions in b-dna. *Nucleic Acids Res.*, 21:561–568, 1993.

[51] M. Poncin, B. Hartmann, and R. Lavery. Conformational sub-states in b-dna. *J. Mol. Biol.*, 226(3):775–794, August 1992.

[52] P. Varnai, D. Djuranovic, R. Lavery, and B. Hartmann. α/γ transitions in the b-dna backbone. *Nucleic Acids Res.*, 30:5398–5406, 2002.

[53] F. Briki, J. Ramstein, R. Lavery, and D. Genest. Evidence for the stochastic nature of base pair opening in dna - a brownian dynamics simulation. *J. Am. Chem. Soc.*, 113(7):2490–2493, March 1991.

[54] J. Bernet, K. Zakrzewska, and R. Lavery. Modelling base pair opening: the role of helical twist. *J. Mol. Struct. (Theochem)*, 398:473–482, June 1997.

[55] P. Varnai and R. Lavery. Base flipping in dna: Pathways and energetics studied with molecular dynamic simulations. *J. Am. Chem. Soc.*, 124(25):7272–7273, June 2002.

[56] E. Giudice, P. Varnai, and R. Lavery. Base pair opening within b-dna: free energy pathways for gc and at pairs from umbrella sampling simulations. *Nucleic Acids Res.*, 31(5):1434–1443, March 2003.

[57] U. D. Priyakumar and A. D. MacKerell. Computational approaches for investigating base flipping in oligonucleotides. *Chem. Rev.*, 106(2):489–505, February 2006.

[58] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette. Stretched and overwound dna forms a pauling-like structure with exposed bases. *Proc. Natl. Acad. Sci. USA*, 95(24):14152–14157, November 1998.

[59] T. Lionnet, S. Joubaud, R. Lavery, D. Bensimon, and V. Croquette. Wringing out dna. *Phys. Rev. Lett.*, 96(17):178102, May 2006.

[60] M. Zacharias. Minor groove deformability of dna: A molecular dynamics free energy simulation study. *Bioph. J.*, 91:882–991, 2006.

[61] R. Lavery and B. Pullman. The molecular electrostatic potential, steric accessibility and hydration of dickerson b-dna dodecamer d(cpgpcpgpapaptptpcpgpcpg). *Nucleic Acids Res.*, 9(15):3765–3777, 1981.

[62] S. Y. Ponomarev, K. M. Thayer, and D. L. Beveridge. Ion motions in molecular dynamics simulations on dna. *Proc. Natl. Acad. Sci. USA*, 101:14771–14775, 2004.

[63] R. Lavery, A. Pullman, and B. Pullman. The electrostatic-field of b-dna. *Theor. Chim. Acta*, 62(2):93–106, 1982.

[64] R. Lavery and B. Pullman. The dependence of the surface electrostatic potential of b-dna on environmental-factors. *J. Biomol. Struct. Dyn.*, 2(5):1021–1032, 1985.

[65] K. Zakrzewska, A. Madami, and R. Lavery. Poisson-boltzmann calculations for nucleic acids and nucleic acids complexes. *Chem. Phys.*, 204(2-3):263–269, April 1996.

[66] R. A. O'Flanagan, G. Paillard, R. Lavery, and A. M. Sengupta. Non-additivity in protein-dna binding. *Bioinformatics*, 21(10):2254–2263, May 2005.

[67] B. Jayaram and T. Jain. The role of water in protein-dna recognition. *Annu. Rev. Biophys. Biomol. Structure*, 33:343–361, 2004.

[68] A. Sarai and H. Kono. Protein-dna recognition patterns and predictions. *Annu. Rev. Biophys. Biomol. Structure*, 34:379–398, 2005.

[69] G. Paillard and R. Lavery. Analyzing protein-dna recognition mechanisms. *Structure*, 12(1):113–122, January 2004.

[70] SB Singh, Ajay , DE Wemmer, and PA Kollman. Relative binding affinities of distamycin and its analog to d(cgcaagttggc) middle dotd(gccaacttgcg): Comparison of simulation results with experiment. *Proc. Natl. Acad. Sci. USA*, 91(16):7673–7677, 1994.

[71] D. R. Langley, J. Golik, B. Krishnan, T. W. Doyle, and D. L. Beveridge. The dna-esperamicin-a(1) complex - a model-based on solvated molecular-dynamics simulations. *J. Am. Chem. Soc.*, 116(1):15–29, January 1994.

[72] M. R. Reddy and M. D. Erion, editors. *Free Energy Calculations in Rational Drug Design*. Springer-Verlag, 2001.

[73] P. B Dervan. Molecular recognition of dna by small molecules. *Bioorg. Med. Chem.*, 9:2215–2235, 2001.

[74] S. A. Harris, E. Gavathiotis, M. S. Searle, M. Orozco, and C. A. Laughton. Cooperativity in drug-dna recognition: a molecular dynamics study. *J. Am. Chem. Soc.*, 123:12658–12663, 2001.

[75] N. Spackova, T. E. Cheatham III, F. Ryjacek, F. Lankas, van L. Meervelt, P. Hobza, and J. Sponer. Molecular dynamics simulations and thermodynamic analysis of dna-drug complexes. minor groove binding between 4',6-diamidino-2-phenylindole (dapi) and dna duplexes in solution. *J. Am. Chem. Soc.*, 125:1759–1769, 2003.

[76] J. Dolenc, C. Oostenbrink, J. Koller, and W. F. van Gunsteren. Molecular dynamics simulations and free energy calculations of netropsin and distamycin binding to an aaaaa dna binding site. *Nucleic Acids Res.*, 33:725–733, 2005.

[77] John A. Hartley, Victoria J. Spanswick, Natalie Brooks, Peter H. Clingen, Peter J. McHugh, Daniel Hochhauser, R. Barbara Pedley, Lloyd R. Kelland, Michael C. Alley, Robert Schultz, Melinda G. Hollingshead, Karen M. Schweikart, Joseph E. Tomaszewski, Edward A. Sausville, Stephen J. Gregson, Philip W. Howard, and David E. Thurston. Sjg-136 (nsc 694501), a novel rationally designed dna minor groove interstrand cross-linking agent with potent and broad spectrum antitumor activity: Part 1: Cellular pharmacology, in vitro and initial in vivo antitumor activity. *Cancer Res*, 64(18):6693–6699, 2004.

[78] A. Javanthan, A. Incornato, J. Fowler, R. Anderson, J. Whitlock, and A. Narendran. Pre-clinical studies on the novel chemotherapeutic agent (sjg-136) for the treatment of refractory acute lymphoblastic leukemia in children. *Pediat. Blood Cancer*, 48(6):636–636, 2007.

[79] A. Le Cesne, J. Y. Blay, I. Judson, A. Van Oosterom, J. Verweij, J. Radford, P. Lorigan, S. Rodenhuis, I. Ray-Coquard, S. Bonvalot, F. Collin, J. Jimeno, E. Di Paola, M. Van Glabbeke, and O. S. Nielsen. Phase ii study of et-743 in advanced soft tissue sarcomas: a european organisation for the research and treatment of cancer (eortc) soft tissue and bone sarcoma group trial. *J Clin Oncol*, 23(3):576–584, Jan 2005.

[80] A. M. Burger, P. M. Loadman, D. E. Thurston, R. Schultz, H. Fiebig, and M. C. Bibby. Preclinical pharmacology of the pyrrolobenzodiazepine (pbd) monomer drh-417 (nsc 709119). *J. Chemother.*, 19(1):66–78, 2007.

[81] SR Alberts, C Erlichman, JM Reid, JA Sloan, MM Ames, RL Richardson, and RM Goldberg. Phase I study of the duocarmycin semisynthetic derivative KW-2189 given daily for five days every six weeks. *Clin Cancer Res*, 4(9):2111–2117, 1998.

[82] K. Spiegel and A. Magistrato. Modeling anticancer drug-dna interactions via mixed qm/mm molecular dynamics simulations. *Org. Biomol. Chem.*, 4(13):2507–2517, 2006.

[83] K. Spiegel, U. Rothlisberger, and P. Carloni. Duocarmycins binding to dna investigated by molecular simulation. *J. Phys. Chem. B*, 110(8):3647–3660, 2006.

[84] A.V. Vargiu, P. Ruggerone, A. Magistrato, and P. Carloni. Anthramycin-dna binding explored by molecular simulations. *J. Phys. Chem. B*, 110(48):24687–24695, 2006.

[85] R. Garcia-Nieto, I. Manzanares, C. Cuevas, and F. Gago. Bending of dna upon binding of ecteinascidin 743 and phthalascidin 650 studied by unrestrained molecular dynamics simulations. *J. Am. Chem. Soc.*, 122(30):7172–7182, August 2000.

[86] F. C. Seaman and L. H. Hurley. Molecular basis for the dna sequence selectivity of ecteinascidin 736 and 743: Evidence for the dominant role of direct readout via hydrogen bonding. *J. Am. Chem. Soc.*, 120(50):13028–13041, December 1998.

[87] J. Dolenc, R. Baron, C. Oostenbrink, J. Koller, and W. F. van Gunsteren. Configurational entropy change of netropsin and distamycin upon dna minor-groove binding. *Biophys. J.*, 91(4):1460–1470, 2006.

[88] M. Rueda, F. J. Luque, and M. Orozco. Nature of minor-groove binders-dna complexes in the gas phase. *J. Am. Chem. Soc.*, 127(33):11690–11698, 2005.

[89] B. Wellenzohn, M. J. Loferer, M. Trieb, C. Rauch, R. H. Winger, E. Mayer, and K. R. Liedl. Hydration of hydroxypyrrole influences binding of imhppypy-beta-dp polyamide to dna. *J. Am. Chem. Soc.*, 125(4):1088–1095, January 2003.

[90] B. Wellenzohn, W. Flader, R. H. Winger, A. Hallbrucker, E. Mayer, and K. R. Liedl. Complex of b-dna with polyamides freezes dna backbone flexibility. *J. Am. Chem. Soc.*, 123:5044–5049, 2001.

[91] B. Wellenzohn, W. Flader, R. H. Winger, A. Hallbrucker, E. Mayer, and K. R. Liedl. Influence of netropsin's charges on the minor groove width of d(cgcgaattcgcg)(2). *Biopol.*, 61(4):276–286, 2001.

[92] K. Bielawski, A. Bielawska, D. Bartulewicz, and A. Rozanski. Molecular modelling of the interaction of carbocyclic analogues of netropsin and distamycin with d(cgcgaattcgcg)(2). *Acta Bioch. Pol.*, 47(3):855–866, 2000.

[93] B. Wellenzohn, R. H. Winger, A. Hallbrucker, E. Mayer, and K. R. Liedl. Simulation of ecori dodecamer netropsin complex confirms class i complexation mode. *J. Am. Chem. Soc.*, 122(16):3927–3931, April 2000.

[94] S. B. Singh and P. A. Kollman. Calculating the absolute free energy of association of netropsin and dna. *J. Am. Chem. Soc.*, 121(14):3267–3271, 1999.

[95] K. Boehncke, M. Nonella, K. Schulten, and A. H. J. Wang. Molecular-dynamics investigation of the interaction between dna and distamycin. *Biochemistry*, 30(22):5465–5475, June 1991.

[96] K. Zakrzewska, R. Lavery, and B. Pullman. A theoretical-study of the sequence specificity in binding of lexitropsins to b-dna. *J. Biomol. Struct. Dyn.*, 4(5):833–843, April 1987.

[97] K. Zakrewska, R. Lavery, and B. Pullman. Theoretical-studies of the selective binding to dna of 2 non-intercalating ligands - netropsin and sn-18071. *Nucleic Acids Res.*, 11(24):8825–8839, 1983.

[98] L. H. Hurley. Pyrrolo(1,4)benzodiazepine antitumor antibiotics: comparative aspects of anthramycin, tomaymycin and sibiromycin. *J. Antibiotic*, 30:349–370, 1977.

[99] M. L. Kopka, D. S. Goodsell, I. Bailakov, K. Grzeskowiak, D. Cascio, and R. E. Dickerson. Crystal structure of a covalent dna-drug adduct: Anthramycin bound to c-c-a-a-c-g-t-t-g-g and a molecular explanation of specificity. *Biochemistry*, 33:13593–13610, 1994.

[100] J. A. Gottlieb and B. Drewinko. Review of the current clinical status of platinum coordination complexes in cancer chemotherapy. *Cancer Chemother Rep*, 59(3):621–628, 1975.

[101] R. M. Wing, P. Pjura, H. R. Drew, and R. E. Dickerson. The primary mode of binding of cisplatin to a b-dna dodecamer: C-g-c-g-a-a-t-t-c-g-c-g. *EMBO J*, 3(5):1201–1206, May 1984.

[102] P. M. Takahara, A. C. Rosenzweig, C. A. Frederick, and S. J. Lippard. Crystal structure of double-stranded dna containing the major adduct of the anticancer drug cisplatin. *Nature*, 377(6550):649–652, Oct 1995.

[103] 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. *Biochemistry*, 37(26):9230–9239, Jun 1998.

[104] K. Spiegel, U. Rothlisberger, and P. Carloni. Cisplatin binding to dna oligomers from hybrid car-parrinello/molecular dynamics simulations. *J. Phys. Chem. B*, 108:2699–2707, 2004.

[105] I. Takahashi, K. Takahashi, M. Ichimura, M. Morimoto, K. Asano, I. Kawamoto, F. Tomita, and H. Nakano. Duocarmycin-a, a new antitumor antibiotic from streptomyces. *J. Antibiot.*, 41(12):1915–1917, 1988.

[106] C. H. Lin and D. J. Patel. Solution structure of the covalent duocarmycin a-dna duplex complex. *J. Mol. Biol.*, 248(1):162–179, Apr 1995.

[107] P. S. Eis, J. A. Smith, J. M. Rydzewski, D. A. Case, D. L. Boger, and W. J. Chazin. High resolution solution structure of a dna duplex alkylated by the antitumor agent duocarmycin sa. *J Mol Biol*, 272(2):237–252, Sep 1997.

[108] A. Dimarco, M. Gaetani, P. Orezzi, T. Scotti, and F. Arcamone. Experimental studies on distamycin-a-a new antibiotic with cytotoxic activity. *Cancer Chem. Rep.*, (18):15–19, 1962.

[109] M. Coll, C. A. Frederick, A. H. J. Wang, and A Rich. A bifurcated hydrogen-bonded conformation in the d(amiddle dot t) base pairs of the dna dodecamer d(cgcaaatttgcg) and its complex with distamycin. *Proc. Natl. Acad. Sci. USA*, 84(23):8385–8389, 1987.

[110] X. Chen, B. Ramakrishnan, S. T. Rao, and M. Sundaralingam. Binding of 2 distamycin-a molecules in the minor-groove of an alternating b-dna duplex. *Nat. Struct. Biol.*, 1(3):169–175, 1994.

[111] K. Uytterhoeven, J. Sponer, and L. Van Meervelt. Two 1 : 1 binding modes for distamycin in the minor groove of d(ggccaattgg). *Eur. J. Biochem.*, 269(12):2868–2877, 2002.

[112] B. K. Keppler, W. Rupp, U. M. Juhl, H. Endres, R. Niebl, and W. Balzer. Synthesis, molecular structure, and tumor-inhibiting properties of imidazolium trans-bis(imidazole)tetrachlororuthenate(iii) and its methyl-substituted derivatives. *Inorg. Chem.*, 26(26):4366–4370, 1987.

[113] G. Sava, S. Pacor, G. Mestroni, and E. Alessio. Na[trans-rucl4(dmso)im], a metal complex of ruthenium with antimetastatic properties. *Clinical and Experimental Metastasis*, 10(4):273–280, 1992. Cited By (since 1996): 24 Export Date: 21 December 2006 Source: Scopus.

[114] G. Mestroni, E. Alessio, G. Sava, S. Pacor, and M. Coluccia. *Metal Complexes in Cancer Chemotherapy*. VCH Verlagsgesellschaft, Weinheim(Germany), 1993.

[115] A. Bergamo, V. Scarcia, A. Furlani, G. Mestroni, E. Alessio, and G. Sava. Cell cycle perturbations of the anti-metastasis agent imh[trans-rucl4(dmso)im] in comparison to cisplatin. *Cell Proliferation*, 30(10-12):460, 1997. Export Date: 21 December 2006 Source: Scopus.

[116] P. J. Dyson and G. Sava. Metal-based antitumour drugs in the post genomic era. *Dalton Trans.*, (16):1929–1933, 2006. Cited By (since 1996): 6 Export Date: 21 December 2006 Source: Scopus.

[117] A. Bergamo and G. Sava. Ruthenium complexes can target determinants of tumour malignancy. *Dalton Trans.*, (13):1267–1272, 2007.

[118] A. Stirling, M. Iannuzzi, A. Laio, and M. Parrinello. Azulene-to-naphthalene rearrangement: The car-parrinello metadynamics method explores various reaction mechanisms. *Chem. Phys. Chem.*, 5:1558–1568, 2004.

[119] D. Frenkel and B. Smit. *Understanding Molecular Simulations: From Algorithms to Applications*. Academic Press, San Diego, CA, 2 edition, 2002.

[120] A. M. Liapunov. *Stability of motion*. Academic Press, New York, 1966.

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

[122] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.

[123] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384–2393, 1980.

[124] S. Toxvaerd and O. H. Olsen. Canonical molecular-dynamics of molecules with internal degrees of freedom. *Ber. Bunsenges Phys. Chem.*, 94(3):274–278, 1990.

[125] G. J. Martyna, M. L. Klein, and M. Tuckerman. Nosé–hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97(4):2635–2643, 1992.

[126] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamicsmethod. *J. App. Phys.*, 52:7182–7190, 1981.

[127] B. H. Bransden and C. J. Joachain. *Physics of atoms and molecules*. Longman, London, 1983.

[128] D. Marx and J. Hütter. *Modern Methods and Algorithms of Quantum Chemistry*, volume 1, chapter "Ab initio molecular dynamics: Theory and implementation", pages 301–449. John von Neumann Institute for Computing, Jülich, NIC Series, 2000.

[129] L. D. Landau and E. M. Lifshits. *Quantum mechanics - non-relativistic theory*. Oxford, Pergamon Press, 3 edition, 1977.

[130] C. J. Cramer. *Essentials of Computational Chemistry*. Wiley, New York, 2 edition, 2002.

[131] J. Wang, P. Cieplak, and P. A. Kollmann. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.*, 21:1049–1074, 2000.

[132] M. Clark, R. D. Cramer, and N. Vanopdenbosch. Validation of the general-purpose tripos 5.2 force-field. *J. Comp. Chem.*, 10(8):982–1012, 1989.

[133] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4(2):187–217, 1983.

[134] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The resp model. *J. Phys. Chem.*, 97:10269–10280, 1993.

[135] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1989.

[136] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

[137] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[138] R. W. Hockney and J. W. Eastwood. *Computer Simulations Using Particles*. McGraw-Hill, New York, 1981.

[139] W. H. Press, B. P. Flannery S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1 edition, 1986.

[140] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev. B.*, 136:864–871, 1964.

[141] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press US, 1989.

[142] T. Koopmans. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Physica*, 1:104, 1934.

[143] C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev*, 46(7):618–622, 1934.

[144] F. Jensen. *Introduction to Computational Chemistry*. Wiley, 2 edition, 2006.

[145] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A*, 140:1133–1138, 1965.

[146] K. Capelle. A bird's-eye view of density-functional theory. *arxiv:cond-mat/0211443*, 2006.

[147] W. J. Carr Jr. Energy, specific heat, and magnetic properties of the low-density electron gas. *Phys. Rev*, 122:1437, 1961.

[148] W. J. Carr Jr and A. A. Maradudin. Ground-state energy of a high-density electron gas. *Phys. Rev*, 133:A371, 1964.

[149] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45(7):566–569, 1980.

[150] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58:1200–1211, 1980.

[151] J. P. Perdew and Y. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45(23):13244–13249, 1992.

[152] F. Sim, A. St. Amant, I. Papai, and D. R. Salahub. Gaussian density functional calculations on hydrogen-bonded systems. *J. Am. Chem. Soc.*, 114(11):4391–4400, 1992.

[153] C. Tuma, A. D. Boese, and N. C. Handy. Predicting the binding energies of h-bonded complexes: A comparative dft study. *Phys. Chem. Chem. Phys*, 1:3939–3947, 1999.

[154] Y. Andersson, D. C. Langreth, and B. I. Lundqvist. van der waals interactions in density-functional theory. *Phys. Rev. Lett.*, 76(1):102–105, 1996.

[155] E. R. Johnson and A. D. Becke. A post-hartree–fock model of intermolecular interactions. *J. Chem. Phys.*, 123(2):024101, 2005.

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

[157] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B*, 37:785–589, 1988.

[158] O. V. Gritsenko, B. Ensing, P. R. T. Schipper, and E. J. Baerends. Comparison of the accurate kohn-sham solution with the generalized gradient approximations (GGAs) for the SN_2 reaction $F-+CH3F->FCH3+F-$: A qualitative rule to predict success or failure of GGAs. *J. Phys. Chem. A*, 104(37):8558–8565, 2000.

[159] J. Harris. Adiabatic-connection approach to kohn-sham theory. *Phys. Rev. A*, 29(4):1648–1659, 1984.

[160] A. D. Becke. A new mixing of hartree–fock and local density-functional theories. *J. Chem. Phys.*, 98(2):1372–1377, 1993.

[161] A. D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652, 1993.

[162] A. D. Boese and N. C. Handy. New exchange-correlation density functionals: The role of the kinetic-energydensity. *J. Chem. Phys.*, 116(22):9559–9569, 2002.

[163] M. Ernzerhof and G. E. Scuseria. Assessment of the perdew–burke–ernzerhof exchange-correlation functional. *J. Chem. Phys.*, 110(11):5029–5036, 1999.

[164] J. C. Slater. Atomic shielding constants. *Phys. Rev*, 36(1):57–64, 1930.

[165] S. F. Boys. Electronic wave functions. i. a general method of calculation for the stationary states of any molecular system. *Proc. R. Soc. (London) A*, 200:542–554, 1950.

[166] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self–consistent molecular orbital methods. xii. further extensions of gaussian–type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56(5):2257–2261, 1972.

[167] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Harcourt, Orlando, 1976.

[168] G. J. Martyna and M. E. Tuckerman. 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.

[169] N. Troullier and J. L. Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43:1993–2006, 1991.

[170] L. Kleinman and D. M. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.*, 48:1425–1428, 1982.

[171] M. F. Kropman and H. J. Bakker. Dynamics of Water Molecules in Aqueous Solvation Shells. *Science*, 291(5511):2118–2120, 2001.

[172] M. Born R. Oppenheimer. Zur quantentheorie der molekeln (on the quantum theory of molecules). *Annalen der Physik*, 84:457–484, 1927.

[173] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J. Mol. Biol.*, 103(2):227–249, 1976.

[174] D. Bakowies and W. Thiel. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.*, 100(25):10580–10594, 1996.

[175] M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello. A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys.*, 110(21):10452–10467, 1999.

[176] A. Laio, J. VandeVondele, and U. Rothlisberger. A hamiltonian electrostatic coupling scheme for hybrid car-parrinello molecular dynamics simulations. *J. Chem. Phys.*, 116:6941–6948, 2002.

[177] J. Hutter, A. Alavi, T. Deutsch, P. Ballone, M. Bernasconi, P. Focher, S. Goedecker, M. Tuckerman, and M. Parrinello. *CPMD*. Max-Planck-Institut für Festkürperforschung, Stuttgart and IBM Research Laboratory Zürich, 1995-1999.

[178] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Phys. Rev. B*, 54(3):1703–1710, 1996.

[179] A. Laio, J. VandeVondele, and U. Rothlisberger. D-resp: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulation. *J. Phys. Chem. B*, 106:7300–7307, 2002.

[180] C. Chipot and A. Pohorille, editors. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, volume 86 of *Springer Series in Chemical Physics*. Springer, 2007.

[181] N. Foloppe and R. Hubbard. Towards predictive ligand design with free-energy based computational methods? *Curr Med Chem*, 13(29):3583–3608, 2006.

[182] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.*, 156(5):472–477, April 1989.

[183] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J. Chem. Phys.*, 23:187–199, 1977.

[184] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comp. Chem.*, 16(11):1339–1350, November 1995.

[185] B. Roux. The calculation of the potential of mean force using computer-simulations. *Comput. Phys. Comm.*, 91(1-3):275–282, September 1995.

[186] C. Jarzynsky. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[187] E. Darve and A. Pohorille. Calculating free energies using average force. *J. Chem. Phys.*, 115(20):9169–9183, 2001.

[188] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.*, 108(5):1964–1977, 1998.

[189] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99(20):12562–12566, 2002.

[190] B. Ensing, A. Laio, F.L. Gervasio, M. Parrinello, and M.L. Klein. A minimum free energy reaction path for the e2 reaction between fluoro ethane and a fluoride ion. *J. Am. Chem. Soc.*, 126(31):9492–9493, 2004.

[191] A. Warshel. Computer simulations of enzyme catalysis: methods, progress, and insights. *Annu. Rev. Biophys. Biomol. Struct.*, 32:425–443, 2003.

[192] M. Klähn, E. Rosta, and A. Warshel. On the mechanism of hydrolysis of phosphate monoesters dianions in solutions and proteins. *J. Am. Chem. Soc.*, 128(47):15310–15323, Nov 2006.

[193] F.L. Gervasio, A. Laio, and M. Parrinello. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.*, 127(8):2600–2607, 2005.

[194] H.-J. Woo and B. Roux. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA*, 102(19):6825–6830, May 2005.

[195] P. A. Sims, C. F. Wong, and J. A. McCammon. A computational model of binding thermodynamics: the design of cyclin-dependent kinase 2 inhibitors. *J. Med. Chem.*, 46(15):3314–3325, Jul 2003.

[196] T. Simonson, G. Archontis, and M. Karplus. Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.*, 35(6):430–437, Jun 2002.

[197] H. Gouda, I. D. Kuntz, D. A. Case, and P. A. Kollman. Free energy calculations for theophylline binding to an rna aptamer: Comparison of mm-pbsa and thermodynamic integration methods. *Biopolymers*, 68(1):16–34, Jan 2003.

[198] K. M. Masukawa, P. A. Kollman, and I. D. Kuntz. Investigation of neuraminidase-substrate recognition using molecular dynamics and free energy calculations. *J Med Chem*, 46(26):5628–5637, Dec 2003.

[199] G. Archontis, K. A. Watson, Q. Xie, G. Andreou, E. D. Chrysina, S. E. Zographos, N. G. Oikonomakos, and M. Karplus. Glycogen phosphorylase inhibitors: a free energy perturbation analysis of glucopyranose spirohydantoin analogues. *Proteins*, 61(4):984–998, Dec 2005.

[200] N. Schmid, B. Zagrovic, and W. F. van Gunsteren. Mechanism and thermodynamics of binding of the polypyrimidine tract binding protein to rna. *Biochemistry*, 46(22):6500–6512, Jun 2007.

[201] D. A. Pearlman and P. A. Kollman. The calculated free energy effects of 5-methyl cytosine on the b to z transition in dna. *Biopolymers*, 29(8-9):1193–1209, 1990.

[202] N. K. Banavali and B. Roux. Free energy landscape of a-dna to b-dna conversion in aqueous solution. *J. Am. Chem. Soc.*, 127(18):6866–6876, 2005.

[203] J. G. Kirkwood and F. P. Buff. The statistical mechanical theory of solutions. i. *J. Chem. Phys.*, 19(6):774–777, 1951.

[204] A. M. Ferrenberg and R. H. Swendsen. New monte carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 61(23):2635–2638, Dec 1988.

[205] A. M. Ferrenberg and R. H. Swendsen. New monte carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 63(15):1658, Oct 1989.

[206] F. Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86(10):2050–2053, Mar 2001.

[207] D. Cvijovic and J. Klinowski. Taboo search - an approach to the multiple minima problem. *Science*, 267(5198):664–666, February 1995.

[208] T. Huber, A. E. Torda, and W. F. van Gusteren. Local elevation - a method for improving the searching properties of molecular-dynamics simulation. *J. Comput. Aided Mol. Des.*, 8(6):695–708, December 1994.

[209] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello. Assessing the accuracy of metadynamics. *J. Phys. Chem. B*, 109(14):6714–6721, 2005.

[210] B. Ensing, A. Laio, M. Parrinello, and M. L. Klein. A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B*, 109(14):6676–6687, 2005.

[211] R. Martoňák, A. Laio, and M. Parrinello. Predicting crystal structures: The parrinello-rahman method revisited. *Phys. Rev. Lett.*, 90(7):075503, Feb 2003.

[212] M. Iannuzzi, A. Laio, and M. Parrinello. Efficient exploration of reactive potential energy surfaces using car-parrinello molecular dynamics. *Phys. Rev. Lett.*, 90(23):238302, Jun 2003.

[213] C. Micheletti, A. Laio, and M. Parrinello. Reconstructing the density of states by history-dependent metadynamics. *Phys. Rev. Lett.*, 92(17):170601, 2004.

[214] M. Sprik and G. Ciccotti. Free energy from constrained molecular dynamics. *J. Chem. Phys.*, 109(18):7737–7744, 1998.

[215] L. H. Hurley, C. Chandler, T. F. Garner, R. Petrusek, and S. G. Zimmer. Dna binding, induction of unscheduled dna synthetis, and excision of anthramycin from dna in normal and repair-deficient human fibroplasts. *J. Biol. Chem.*, 254:605–608, 1979.

[216] R. L. Petrusek, E. L. Uhlenhopp, N. Duteau, and L. H. Hurley. Reaction of anthramycin with dna. biological consequences of dna damage in normal and xeroderma pigmentosum cell. *J. Biol. Chem.*, 257:6207–6216, 1982.

[217] R. P. Hertzberg, S. M. Hecht, V. L. Reynolds, I. J. Molineux, and L. H. Hurley. Dna sequence specificity of the pyrrolo[1,4]benzodiazepine antitumor antibiotics. methidiumpropyl-edta-iron(ii) footprinting analysis of dna binding sites for anthramycin and related drugs. *Biochemistry*, 25:1249–1258, 1986.

[218] L. H. Hurley, T. Reck, D. E. Thurston, and D. R. Langley. Pyrrolo[1,4]benzodiazepine antitumor antibiotics: relationship of dna alkylation and sequence specificity to the biological activity of natural and synthetic compounds. *Chem. Res. Toxicol.*, 1:258–268, 1988.

[219] M. A. Warpehoski and L. H. Hurley. Sequence selectivity of dna covalent modification. *Chem. Res. Toxicol.*, 1:315–333, 1988.

[220] F. L. Boyd, S. F. Cheatham, W. Remers, G. C. Hill, and L. H. Hurley. Characterization of the structure of the anthramycin-d(atgcat)2 adduct by nmr and molecular modeling studies. determination of the stereochemistry of the covalent linkage site, orientation in the minor groove of dna, and effect on local dna structure. *J. Am. Chem. Soc.*, 112(9):3279–3289, 1990.

[221] R. Kizu, P. H. Draves, and L. H. Hurley. Correlation of dna sequence specificity of anthramycin and tomaymycin with reaction kinetics and bending of dna. *Biochemistry*, 32:8712–8722, 1993.

[222] W. A. Remers, M. Mabilia, and A. J. Hopfinger. Conformations of complexes between pyrrolo[1,4]benzodiazepines and dna segments. *J Med. Chem.*, 29:2492–2503, 1986.

[223] M. C. Alley, M. G. Hollingshead, C. M. Pacula-Cox, W. R. Waud, J. A. Hartley, P. W. Howard, S. J. Gregson, D. E. Thurston, and E. A. Sausville. Sjg-136 (nsc 694501), a novel rationally designed dna minor groove interstrand cross-linking agent with potent and broad spectrum antitumour activity: Part 2: Efficacy evaluations. *Cancer Res.*, 64:6700–6706, 2004.

[224] C. Martin, T. Ellis, C. J. McGurk, T. C. Jenkins, J. A. Hartley, M. J. Waring, and D. E. Thurston. Sequence-selective interaction of the minor-groove interstrand cross-linking agent SJG-136 with naked and cellular dna: Footprinting and enzymeinhibition studies. *Biochemistry*, 44:4135–4147, 2005.

[225] A. M. James, F. H. Blaikie, R. A. J. Smith, R. N. Lightowlers, P. M. Smith, and M. P. Murphy. Specific targeting of a dna-alkylating reagent to mitochondria. *Eur. J. Biochem.*, 270:2827–2836, 2003.

[226] S. N. Rao, U. C. Singh, and P. A. Kollman. Molecular mechanics simulations on covalent complexes between anthramycin and b dna. *J Med. Chem.*, 29:2484–2492, 1986.

[227] K. Zakrewska and B. Pullman. A theoretical investigation of the sequence specificity in the binding of the antitumor drug anthramycin to dna. *J. Biomol. Struct. Dyn.*, 4:127–136, 1986.

[228] L. J. Adams, T. C. Jenkins, L. Banting, and D. E. Thurston. Molecular modelling of a sequence-specific dna-binding agent based on the pyrrolo[2,1-c][1,4]benzodiazepines. *Pharm. Pharmacol. Commun.*, 5:555–560, 1999.

[229] M. Freccero, R. Gandolfi, M. Sarzi-Amade, and A. Rastelli. Peroxy acid epoxidation of acyclic allylic alcohols. competition between s-trans and s-cis peroxy acid conformers. *J. Org. Chem.*, 70:9573–9583, 2005.

[230] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, and J. C. Burant. *Gaussian 98, Revision A.3*. Gaussian, Inc., Pittsburgh PA, 1998.

[231] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. R. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculationsto simulate the structural and energetic properties of molecules. *Comp. Phys. Comm.*, 91:1–41, 1995.

[232] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

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

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

[235] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.

[236] D. L. Beveridge, G. Barreiro K. S. Byun, D. A. Case, T. E. Cheatham III, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of dna oligonucleotides. i. research design and results on d(cpg) steps. *Bioph. J.*, 87:3799–3813, 2004.

[237] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91:43–56, 1995.

[238] E. Lindahl, B. Hess, and D. van der Spoel. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.*, 7:306–317, 2001.

[239] R. Lavery and H. Sklenar. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, 6:655–677, 1989.

[240] B. Lee and F. Richards. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.*, 55:379–400, 1971.

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

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

[243] A. D. Becke. Density-functional thermochemistry. ii. the effect of the perdew–wang generalized-gradient correlati correction. *J. Chem. Phys.*, 97(12):9173–9177, 1992.

[244] D. Sebastiani and U. Rothlisberger. Advances in density–functional–based modeling techniques - recent extensions of the car–parrinello approach. In P. Carloni and F. Alber, editors, *Quantum Medicinal Chemistry*, Weinheim, Germany, 2003. Wiley-VCH. pp. 5-36.

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

[246] N. Marzari and D. Vanderbilt. Maximally localized generalized wannier functions for composite energy bands. *Phys. Rev. B*, 56:12847–12865, 1997.

[247] F. Alber, G. Folkers, and P. Carloni. Dimethyl phosphate: stereoelectronic versus environmental effects. *J. Phys. Chem. B*, 103:6121–6126, 1999.

[248] C. Bailly, D. E. Graves, G. Ridge, and M. J. Waring. Use of a photoactive derivative of actinomycin to investigate shuffling between binding sites on dna. *Biochemistry*, 33:8736–8745, 1994.

[249] D. Gunz and H. Naegeli. A noncovalent binding-translocation mechanism for site-specific cc-1065-dna recognition. *Biochem. Pharmacol.*, 52:447–453, 1996.

[250] T. E. Cheatham III and M. A. Young. Molecular dynamics simulation of nucleic acids: Successes, limitations, and promise. *Biopol.*, 56:232–256, 2001.

[251] B. Wellenzohn, W. Flader, R. H. Winger, A. Hallbrucker, E. Mayer, and K. R. Liedl. Significance of ligand tails for interaction with the minor groove. *Bioph. J.*, 81:1588–1599, 2001.

[252] J. Schlitter. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.*, 256:617–622, 1993.

[253] C. Teijeiro, E. de la Red, and D. Marin. Electrochemical analysis of anthramycin: Hydrolysis, dna-interactions and quantitative determination. *Electroanal.*, 12:963–968, 2000.

[254] M. D. Barkley, S. Cheatham, D. E. Thurston, and L. H. Hurley. Pyrrolo[1,4]benzodiazepine antitumor antibiotics: evidence for two forms of tomaymycin bound to dna. *Biochemistry*, 25:3021–3031, 1986.

[255] J. D. Farmer Jr, G. R. Gustafson, A. Conti, M. B. Zimmt, and J. W. Suggs. Dna binding properties of a new class of linked anthramycin analogs. *Nucleic Acids Res.*, 19:899–903, 1991.

[256] F. A. Tanious, D. Hamelberg, C. Bailly, A. Czarny, D. W. Boykin, and W. D. Wilson. Dna sequence dependent monomer-dimer binding modulation of asymmetric benzimidazole derivatives. *J. Am. Chem. Soc.*, 126:143–153, 2004.

[257] M. Trieb, C. Rauch, B. Wellenzohn, F. R. Wibowo, T. Loerting, E. Mayer, and K. R. Liedl. Daunomycin intercalation stabilizes distinct backbone conformations of dna. *J. Biomol. Struct. Dyn.*, 21:713–724, 2004.

[258] R. M. Romero, P. Rojsitthisak, and I. S. Haworth. Dna interstrand crosslink formation by mechlorethamine at a cytosine-cytosine mismatch pair: Kinetics and sequence dependence. *Arch. Bioch. Bioph.*, 386:143–153, 2001.

[259] S. Ding, R. Shapiro, N. E. Geacintov, and S. Broyde. Equilenin-derived dna adducts to cytosine in dna duplexes: structures and thermodynamics. *Biochemistry*, 44:14565–14576, 2005.

[260] W. A. Denny. Dna minor groove alkylating agents. *Curr. Med. Chem*, 8:533–544, 2001.

[261] J. B. Chaires. Energetics of drug-dna interactiens. *Biopol.*, 44(3):201–215, 1997. 10.1002/(SICI)1097-0282(1997)44:3<201::AID-BIP2>3.0.CO;2-Z.

[262] J. B. Chaires. Drug–dna interactions. *Curr. Opin. Struct. Biol.*, 8(3):314–320, 1998.

[263] I. Haq. Thermodynamics of drug-dna interactions. *Arch. Bioch. Bioph.*, 403(1):1–15, 2002.

[264] L. H. Hurley. Dna and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer*, 2:188–200, 2002.

[265] K. L. Buchmueller, A. M. Staples, C. M. Howard, S. M. Horick, P. B. Uthe, N. M. Le, K. K. Cox, B. Nguyen, K. A. O. Pacheco, W. D. Wilson, and M. Lee. Extending the language of dna molecular recognition by polyamides: Unexpected influence of imidazole and pyrrole arrangement on binding affinity and specificity. *J. Am. Chem. Soc.*, 127(2):742–750, 2005.

[266] J. R. Schnell, R. R. Ketchem, D. L. Boger, and W. J. Chazin. Binding-induced activation of dna alkylation by duocarmycin sa: Insights from the structure of an indole derivative-dna adduct. *J. Am. Chem. Soc.*, 121(24):5645–5652, 1999.

[267] Eric J. Small, Robert Figlin, Daniel Petrylak, David J. Vaughn, Oliver Sartor, Ivan Horak, Rosemarie Pincus, Alton Kremer, and Chris Bowden. A phase ii pilot study of kw-2189 in patients with advanced renal cell carcinoma. *Invest. New Drugs*, V18(2):193–198, 2000. 10.1023/A:1006386115312.

[268] W. Leupin, W. J. Chazin, S. Hyberts, W. A. Denny, and K. Wüthrich. Nmr studies of the complex between the decadeoxynucleotide $d(GCATTAATGC)_2$ and a minor-groove-binding drug. *Biochemistry*, 25:5902–591, 1986.

[269] S. Y. Breusegem, R. M. Clegg, and F. G. Loontiens. Base-sequence specificity of hoechst 33258 and dapi binding to five (a/t)(4) dna sites with kinetic evidence for more than one high-affinity hoechst 33258-aatt complex. *J. Mol. Biol.*, 315(5):1049–1061, 2002.

[270] Kenneth J. Breslauer, David P. Remeta, Wan-Yin Chou, Robert Ferrante, James Curry, Denise Zaunczkowski, James G. Snyder, and Luis A. Marky. Enthalpy–entropy compensations in drug-dna binding studies. *Proc. Natl. Acad. Sci. USA*, 84(24):8922–8926, 1987.

[271] J. Lah and G. Vesnaver. Energetic diversity of dna minor-groove recognition by small molecules displayed through some model ligand-dna systems. *J. Mol. Biol.*, 342(1):73–89, 2004.

[272] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen. Gromacs: fast, flexible, and free. *J. Comp. Chem.*, 26(16):1701–1718, 2005.

[273] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, Jr. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A Pople. Gaussian 03, 2004.

[274] William L. Jorgensen. Quantum and statistical mechanical studies of liquids. 10. transferable intermolecular potential functions for water, alcohols, and ethers. application to liquid water. *J. Am. Chem. Soc.*, 103(2):335–340, 1981.

[275] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

[276] R. Lavery and H. Sklenar. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, 6:63–91, 1988.

[277] John G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3(5):300–313, 1935.

[278] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations

on biomolecules. i. the method. *J. Comp. Chem.*, 13(8):1011–1021, 1992. 10.1002/jcc.540130812.

[279] D. A. McQuarrie. *Statistical Mechanics*. Harper and Row, New York, 1976.

[280] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, and P. A. Kollman. Amber 9, 2006.

[281] B. Jayaram, D. Sprous, M. A. Young, and D. L. Beveridge. Free energy analysis of the conformational preferences of a and b forms of dna in solution. *J. Am. Chem. Soc.*, 120(41):10629–10633, 1998.

[282] S. Yan, M. Wu, D. J. Patel, N. E. Geacintov, and S. Broyde. Simulating structural and thermodynamic properties of carcinogen-damaged dna. *Biophys. J.*, 84(4):2137–2148, 2003.

[283] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E.Cheatam III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum model. *Acc. Chem. Res.*, 33:889–897, 2000.

[284] K. A. Sharp and B. Honig. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.*, 19(1):301–332, 1990.

[285] Ray Luo, Laurent David, and Michael K. Gilson. Accelerated poisson-boltzmann calculations for static and dynamic systems. *J. Comp. Chem.*, 23(13):1244–1253, 2002. 10.1002/jcc.10120.

[286] Doree Sitkoff, Kim A. Sharp, and Barry Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.*, 98(7):1978–1988, 1994.

[287] C. Tan, L. Yang, and R. Luo. How well does poisson-boltzmann implicit solvent agree with explicit solvent? a quantitative analysis. *J. Phys. Chem. B*, 110(37):18680–18687, 2006.

[288] Emilio Gallicchio, Linda Yu Zhang, and Ronald M. Levy. The sgb/np hydration free energy model based on the surface generalized born solvent reaction

field and novel nonpolar hydration free energy estimators. *J. Comp. Chem.*, 23(5):517–529, 2002. 10.1002/jcc.10045.

[289] Jonathan B. Chaires. A thermodynamic signature for drug-dna binding mode. *Arch. Bioch. Bioph.*, 453(1):26–31, 2006.

[290] Ken A. Dill. Additivity principles in biochemistry. *J. Biol. Chem.*, 272(2):701–704, 1997.

[291] F. Han, N. Taulier, and T. V. Chalikian. Association of the minor groove binding drug hoechst 33258 with d(cgcgaattcgcg)2: Volumetric, calorimetric, and spectroscopic characterizations. *Biochemistry*, 44(28):9785–9794, 2005.

[292] R. Baliga and D. M. Crothers. On the kinetics of distamycin binding to its target sites on duplex dna. *Proc. Natl. Acad. Sci. USA*, 97(14):7814–7818, 2000.

[293] L. Wang, A. Kumar, D. W. Boykin, C. Bailly, and W. D. Wilson. Comparative thermodynamics for monomer and dimer sequence-dependent binding of a heterocyclic dication in the dna minor groove. *J. Mol. Biol.*, 317(3):361–374, 2002.

[294] N. N. Degtyareva, B. D. Wallace, A. R. Bryant, K. M. Loo, and J. T. Petty. Hydration changes accompanying the binding of minor groove ligands with dna. *Biophys. J.*, 92(3):959–965, 2007.

[295] J. R. Kiser, R. W. Monk, R. L. Smalls, and J. T. Petty. Hydration changes in the association of hoechst 33258 with dna. *Biochemistry*, 44(51):16988–16997, 2005.

[296] M. Zewail-Foote and L. H. Hurley. Differential rates of reversibility of ecteinascidin 743-dna covalent adducts from different sequences lead to migration to favored bonding sites. *J. Am. Chem. Soc.*, 123(27):6485–6495, 2001.

[297] S. Y. Breusegem, F. G. Loontiens, P. Regenfuss, and R. M. Clegg. Kinetics of binding of hoechst dyes to dna studied by stopped-flow fluorescence techniques. In *Drug-Nucleic Acid Interactions*, volume 340 of *Methods in Enzymology*, pages 212–233. 2001.

[298] R. Baliga, E. E. Baird, D. M. Herman, C. Melander, P. B. Dervan, and D. M. Crothers. Kinetic consequences of covalent linkage of dna binding polyamides. *Biochemistry*, 40(1):3–8, 2001.

[299] A. Asai, S. Nagamura, H. Saito, I. Takahashi, and H. Nakano. The reversible dna-alkylating activity of duocarmycin and its analogues. *Nucl. Acids Res.*, 22(1):88–93, 1994.

[300] S. Y. Reddy, F. Leclerc, and M. Karplus. Dna polymorphism: A comparison of force fields for nucleic acids. *Biophys. J.*, 84(3):1421–1449, 2003.

[301] Benoît Roux. Ion conduction and selectivity in k(+) channels. *Annu Rev Biophys Biomol Struct*, 34:153–171, 2005.

[302] M. Sotomayor and K. Schulten. Single-molecule experiments in vitro and in silico. *Science*, 316(5828):1144–1148, May 2007.

[303] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu. Biomolecular modeling: Goals, problems, perspectives. *Angew Chem Int Ed Engl*, 45(25):4064–4092, Jun 2006.

[304] A. Barducci, R. Chelli, P. Procacci, V. Schettino, F. L. Gervasio, and M. Parrinello. Metadynamics simulation of prion protein: β-structure stability and the early stages of misfolding. *J. Am. Chem. Soc.*, 128(8):2705–2710, 2006.

[305] S. Piana and A. Laio. A bias-exchange approach to protein folding. *J. Phys. Chem. B*, 111(17):4553–4559, 2007.

[306] M. Ceccarelli, C. Danelon, A. Laio, and M. Parrinello. Microscopic mechanism of antibiotics translocation through a porin. *Biophys. J.*, 87(1):58–64, 2004.

[307] G. Fiorin, A. Pastore, P. Carloni, and M. Parrinello. Using metadynamics to understand the mechanism of calmodulin/target recognition at atomic detail. *Biophys. J.*, 91(8):2768–2777, 2006.

[308] S. Y. Breusegem, S. E. Sadat-Ebrahimi, K. T. Douglas, R. M. Clegg, and F. G. Loontiens. Increased stability and lifetime of the complex formed between dna and meta-phenyl-substituted hoechst dyes as studied by fluorescence titrations and stopped-flow kinetics. *J Mol Biol*, 308(4):649–663, May 2001.

[309] M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, and R. E. Dickerson. The molecular origin of dna-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. USA*, 82(5):1376–1380, 1985.

[310] C. Zimmer and U. Wahnert. Nonintercalating dna-binding ligands - specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Progr. Bioph. Mol. Biol.*, 47(1):31–112, 1986.

[311] P. B. Dervan and R. W. Burli. Sequence-specific dna recognition by polyamides. *Curr. Op. Chem. Biol.*, 3(6):688–693, 1999.

[312] C. Bailly. Topoisomerase i poisons and suppressors as anticancer drugs. *Curr. Med. Chem.*, 7(1):39–58, 2000.

[313] A. Taylor, K. A. Webster, T. A. Gustafson, and L. Kedes. The anti cancer agent distamycin a displaces essential transcription factors and selectively inhibits myogenic differentiation. *Mol. Cell. Biochem.*, 169(1-2):61–72, 1997.

[314] L. H. Hurley and D. E. Thurston. Pyrrolo(1,4)benzodiazepine antitumor antibiotics - chemistry, interaction with dna, and biological implications. *Pharm. Res.*, (2):52–59, 1984.

[315] D. E. Thurston and D. S. Bose. Synthesis of dna-interactive pyrrolo[2,1-c][1,4]benzodiazepines. *Chem. Rev.*, 94(2):433–465, 1994.

[316] I. Puzanov, W. Lee, J. D. Berlin, M. W. Calcutt, D. L. Hachey, W. Vermeulen, J. Low, and M. L. Rothenberg. Phase i and pharmacokinetic trial of sjg-136 administered on a daily x 5 schedule. *Clin. Cancer Res.*, 11(24):9061S–9061S, 2005. Part 2 Suppl. S.

[317] C. Geroni, S. Marchini, P. Cozzi, E. Galliera, E. Ragg, T. Colombo, R. Battaglia, M. Howard, M. D'Incalci, and M. Broggini. Brostallicin, a novel anticancer agent whose activity is enhanced upon binding to glutathione. *Cancer Res*, 62(8):2332–2336, Apr 2002.

[318] J. G. Pelton and D. E. Wemmer. Structural characterization of a 2:1 distamycin a.d(cgcaaattggc) complex by two-dimensional nmr. *Proc Natl Acad Sci U S A*, 86(15):5723–5727, Aug 1989.

[319] C. Bailly and J. B. Chaires. Sequence-specific dna minor groove binders. design and synthesis of netropsin and distamycin analogues. *Bioconj. Chem.*, 9(5):513–538, 1998.

[320] Jr. Brosh, R. M., J. K. Karow, E. J. White, N. D. Shaw, I. D. Hickson, and V. A. Bohr. Potent inhibition of werner and bloom helicases by dna minor groove binding drugs. *Nucl. Acids Res.*, 28(12):2420–2430, 2000.

[321] A. A. Vashisht and N. Tuteja. Cold stress-induced pea dna helicase 47 is homologous to eif4a and inhibited by dna-interacting ligands. *Arch. Bioch. Bioph.*, 440(1):79–90, 2005.

[322] N. R. Bachur, R. Johnson, F. Yu, R. Hickey, N. Applegren, and L. Malkas. Antihelicase action of dna-binding anticancer agents: relationship to guanosine-cytidine intercalator binding. *Mol Pharmacol*, 44(5):1064–1069, 1993.

[323] V. Barone and M. Cossi. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A*, 102(11):1995–2001, 1998.

[324] G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.*, 128(41):13435–13441, 2006.

[325] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. Peptide folding: When simulation meets experiment. *Angew Chem. Int. Ed.*, 38(1-2):236–240, 1999.

[326] R. S. Spolar and M. T. Record. Coupling of local folding to site-specific binding of proteins to dna. *Science*, 263(5148):777–784, 1994.

[327] P. B. Dervan. Design of sequence-specific dna-binding molecules. *Science*, 232(4749):464–471, 1986.

[328] A. V. Vargiu, P. Ruggerone, A. Magistrato, and P. Carloni. Sliding of alkylating anticancer drugs along the minor groove of dna: New insights on sequence selectivity. *Biophys. J.*, 94(2):550–561, 2008.

[329] G. Sava, S. Zorzet, C. Turrin, F. Vita, M. Soranzo, G. Zabucchi, M. Cocchietto, A. Bergamo, S. DiGiovine, G. Pezzoni, L. Sartor, and S. Garbisa. Dual action of nami-a in inhibition of solid tumor metastasis: Selective targeting of metastatic cells and binding to collagen. *Clin. Cancer Res.*, 9(5):1898–1905, 2003. Cited By (since 1996): 16 Export Date: 21 December 2006 Source: Scopus.

[330] J. M. Rademaker-Lakhai, D. van den Bongard, D. Pluim, J. H. Beijnen, and J. H. M. Schellens. A phase i and pharmacological study with imidazolium-trans-dmso-imidazole-tetrachlororuthenate, a novel ruthenium anticancer agent. *Clin. Cancer Res.*, 10(11):3717–3727, June 2004.

[331] E. Alessio, G. Mestroni, A. Bergamo, and G. Sava. Ruthenium antimetastatic agents. *Curr. Top. Med. Chem.*, 4:1525–1535, 2004.

[332] E. Alessio, G. Mestroni, A. Bergamo, and G. Sava. Ruthenium anticancer drugs. *Met. Ions Biol. Syst.*, 42:323–351, 2004. Cited By (since 1996): 12 Export Date: 21 December 2006 Source: Scopus.

[333] M. Bacac, A. C. G. Hotze, K. Van Der Schilden, J. G. Haasnoot, S. Pacor, E. Alessio, G. Sava, and J. Reedijk. The hydrolysis of the anti-cancer ruthenium complex nami-a affects its dna binding and antimetastatic activity: An nmr evaluation. *J. Inorg. Biochem.*, 98(2):402–412, 2004. Cited By (since 1996): 9 Export Date: 21 December 2006 Source: Scopus.

[334] G. Sava, A. Bergamo, S. Zorzet, B. Gava, C. Casarsa, M. Cocchietto, A. Furlani, V. Scarcia, B. Serli, E. Iengo, E. Alessio, and G. Mestroni. Influence of chemical stability on the activity of the antimetastasis ruthenium compound nami-a. *Eur. J. Cancer*, 38(3):427–435, 2002. Cited By (since 1996): 28 Export Date: 21 December 2006 Source: Scopus.

[335] M. Groessl, E. Reisner, C. G. Hartinger, R. Eichinger, O. Semenova, A. R. Timerbaev, M. A. Jakupec, V. B. Arion, and B. K. Keppler. Structure-activity relationships for nami-a-type complexes (hl)[trans-rucl4l(s-dmso)ruthenate(iii)] (l = imidazole, indazole, 1,2,4-triazole, 4-amino-1,2,4-triazole, and 1-methyl-1,2,4-triazole): Aquation, redox properties, protein binding, and antiproliferative activity. *J. Med. Chem.*, 50(9):2185–2193, 2007.

[336] M. Bouma, B. Nuijen, M. T. Jansen, G. Sava, A. Flaibani, A. Bult, and J. H. Beijnen. A kinetic study of the chemical stability of the antimetastatic ruthenium complex nami-a. *Int. J. Pharm.*, 248(1-2):239–246, 2002. Cited By (since 1996): 5 Export Date: 21 December 2006 Source: Scopus.

[337] L. Messori, P. Orioli, D. Vullo, E. Alessio, and E. Iengo. A spectroscopic study of the reaction of nami, a novel ruthenium(iii) anti-neoplastic complex, with bovine serum albumin. *Eur. J. Biochem.*, 267(4):1206–1213, 2000. Cited By (since 1996): 30 Export Date: 21 December 2006 Source: Scopus.

[338] A. Bergamo, L. Messori, F. Piccioli, M. Cocchietto, and G. Sava. Biological role of adduct formation of the ruthenium(iii) complex nami-a with serum albumin and serum transferrin. *Invest. New Drugs*, 21(4):401–411, 2003. Cited By (since 1996): 8 Export Date: 21 December 2006 Source: Scopus.

[339] F. Frausin, V. Scarcia, M. Cocchietto, A. Furlani, B. Serli, E. Alessio, and G. Sava. Free exchange across cells, and echistatin-sensitive membrane target for the metastasis inhibitor nami-a (imidazolium trans-imidazole dimethyl sulfoxide tetrachlororuthenate) on kb tumor cells. *J. Pharmacol. Exp. Ther.*,

313(1):227–233, 2005. Cited By (since 1996): 4 Export Date: 21 December 2006 Source: Scopus.

[340] M. Ravera, S. Baracco, C. Cassino, D. Colangelo, G. Bagni, G. Sava, and D. Osella. Electrochemical measurements confirm the preferential bonding of the antimetastatic complex [imh][rucl4(dmso)(im)] (nami-a) with proteins and the weak interaction with nucleobases. *J. Inorg. Biochem.*, 98(6):984–990, 2004. Cited By (since 1996): 9 Export Date: 21 December 2006 Source: Scopus.

[341] G. Sava, F. Frausin, M. Cocchietto, F. Vita, E. Podda, P. Spessotto, A. Furlani, V. Scarcia, and G. Zabucchi. Actin-dependent tumour cell adhesion after short-term exposure to the antimetastasis ruthenium complex nami-a. *Eur. J. Cancer*, 40(9):1383–1396, 2004. Cited By (since 1996): 9 Export Date: 21 December 2006 Source: Scopus.

[342] S. Kapitza, M. Pongratz, M. A. Jakupec, P. Heffeter, W. Berger, L. Lackinger, B. K. Keppler, and B. Marian. Heterocyclic complexes of ruthenium(iii) induce apoptosis in colorectal carcinoma cells. *J. Cancer Res. Clin. Oncol.*, 131(2):101–110, 2005. Cited By (since 1996): 11 Export Date: 21 December 2006 Source: Scopus.

[343] V. Brabec and O. Novakova. Dna binding mode of ruthenium complexes and relationship to tumor cell toxicity. *Drug Resist. Updates*, 9(3):111–122, June 2006.

[344] S. Zorzet, A. Bergamo, M. Cocchietto, A. Sorc, B. Gava, E. Alessio, E. Iengo, and G. Sava. Lack of in vitro cytotoxicity, associated to increased g2-m cell fraction and inhibition of matrigel invasion, may predict in vivo-selective antimetastasis activity of ruthenium complexes. *J. Pharmacol. Exp. Ther.*, 295(3):927–933, 2000. Cited By (since 1996): 33 Export Date: 21 December 2006 Source: Scopus.

[345] O. M. Ni Dhubhghaill, W. R. Hagen, B. K. Keppler, K-G. Lipponer, and P. J. Sadler. Aquation of the anticancer complex trans-[rucl4(him)2]- (him=imidazole). *J. Chem. Soc. Dalton Trans.*, 22:3305–3310, 1994.

[346] C. Anderson and A. L. Beauchamp. H-1-nmr study of the solvolysis of the paramagnetic tetrachloro-bis(imidazole)ruthenium(iii) anion in water, methanol, and dimethyl-sulfoxide. *Can. J. Chem.*, 73(4):471–482, April 1995.

[347] E. Holler, W. Schaller, and B. Keppler. Inhibition of escherichia coli dna polymerase i catalysed dna polymerization by trans-imidazolium-

bisimidazoletetrachlororuthenate(iii). *Arzneim-Forsch/Drug Res.*, 41(10):1065–8, 1991.

[348] M. Ravera, C. Cassino, S. Baracco, and D. Osella. New insights into the redox chemistry of ruthenium metallopharmaceuticals: The electrochemical behaviour of [lh][trans-ruiiicl4l 2] (l = imidazole or indazole) complexes. *Eur. J. Inorg. Chem.*, (4):740–746, 2006. Export Date: 21 December 2006 Source: Scopus.

[349] A. Küng, T. Pieper, R. Wissiack, Erwin Rosenberg, and Bernhard K. Keppler. Hydrolysis of the tumor-inhibiting ruthenium(iii) complexes him trans-[rucl4(im)2] and hind trans-[rucl4(ind)2] investigated by means of hpce and hplc-ms. *J. Biol. Inorg. Chem.*, 6(3):292–299, 2001. 10.1007/s007750000203.

[350] E. Alessio, G. Balducci, A. Lutman, G. Mestroni, M. Calligaris, and W. M. Attia. Synthesis and characterization of 2 new classes of ruthenium(iii)-sulfoxide complexes with nitrogen donor ligands (l) - na[trans-rucl4(r2so)(l)] and mer,cis-rucl3(r2so)(r2so)(l) - the crystal-structure of na[trans-rucl4(dmso)(nh3)].2dmso, na[trans-rucl4(dmso)(im)].h2o,me2co (im=imidazole) and mer,cis-rucl3(dmso)(dmso)(nh3). *Inorg. Chim. Acta*, 203:205–217, 1993.

[351] G. Sava, E. Alessio, A. Bergamo, and G. Mestroni. Non-toxic tools for the selective treatment of solid tumor metastases. In M. J. Clarke and P. J. Sadler, editors, *Top. Biol. Inorg. Chem.* Spkinger-Verlag, 1999. pp. 143-169.

[352] M. Ravera, S. Baracco, C. Cassino, P. Zanello, and D. Osella. Appraisal of the redox behaviour of the antimetastatic ruthenium(iii) complex [imh][rucl4(dmso)(im)], nami-a. *Dalton Trans.*, (15):2347–2351, 2004. Cited By (since 1996): 5 Export Date: 21 December 2006 Source: Scopus.

[353] G. Mestroni, E. Alessio, G. Sava, S. Pacor, M. Coluccia, and A Boccarelli. Water-soluble ruthenium(iii)-dimethyl sulfoxide complexes: Chemical behaviour and pharmaceutical properties. *Metal-Based Drugs*, 1(1):41–63, 1994.

[354] J. Chen, L. Chen, S. Liao, K. Zheng, and L. Ji. A theoretical study on the hydrolysis process of the antimetastatic ruthenium(iii) complex nami-a. *J. Phys. Chem. B*, 111(27):7862–7869, 2007.

[355] J. Chen, L. Chen, S. Liao, K. Zheng, and L. Ji. The hydrolysis process of the anticancer complex [imh][trans-rucl4(im)2]: a theoretical study. *Dalton Trans.*, pages 3507–3515, 2007.

[356] F. Q. Schafer and G. R. Buettner. Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Radical Biol. Med.*, 30(11):1191–1212, June 2001.

[357] Dirk V. Deubel and Justin Kai-Chi Lau. In silico evolution of substrate selectivity: comparison of organometallic ruthenium complexes with the anticancer drug cisplatin. *Chem. Comm.*, (23):2451–2453, 2006. 10.1039/b601590e.

[358] J. K. C. Lau and D. V. Deubel. Hydrolysis of the anticancer drug cisplatin: Pitfalls in the interpretation of quantum chemical calculations. *J. Chem. Theory Comput.*, 2(1):103–106, 2006.

[359] P. J. Hay and W. R. Wadt. Ab initio effective core potentials for molecular calculations. potentials for the transition metal atoms sc to hg. *J. Chem. Phys.*, 82(1):270–283, 1985.

[360] W. R. Wadt and P. J. Hay. Ab initio effective core potentials for molecular calculations. potentials for main group elements na to bi. *J. Chem. Phys.*, 82(1):284–298, 1985.

[361] P. J Hay and W. R. Wadt. Ab initio effective core potentials for molecular calculations. potentials for k to au including the outermost core orbitals. *J. Chem. Phys.*, 82(1):299–310, 1985.

[362] A. W. Ehlers, M. Bohme, S. Dapprich, A. Gobbi, A. Hollwarth, V. Jonas, K. F. Kohler, R. Stegmann, A. Veldkamp, and G. Frenking. A set of f-polarization functions for pseudo-potential basis sets of the transition metals sc—cu, y—ag and la—au. *Chem. Phys. Lett.*, 208(1-2):111–114, 1993.

[363] V. Barone, M. Cossi, and J. Tomasi. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.*, 107(8):3210–3221, August 1997.

[364] Michael H. Abraham. Relationship between solution entropies and gas phase entropies of nonelectrolytes. *J. Am. Chem. Soc.*, 103(22):6742–6744, 1981.

[365] David H. Wertz. Relationship between the gas-phase entropies of molecules and their entropies of solvation in water and 1-octanol. *J. Am. Chem. Soc.*, 102(16):5316–5322, 1980.

[366] J. Chatlas, R. van Eldik, and B. K. Keppler. Spontaneous aquation reactions of a promising tumor inhibitor trans-imidazolium-tetrachlorobis(imidazole)ruthenium(iii), trans-him[rucl4(im)2]. *Inorg. Chim. Acta*, 233(1-2):59–63, 1995.

[367] J. V. Burda, M. Zeizinger, and J. Leszczynski. Hydration process as an activation of trans- and cisplatin complexes in anticancer treatment. dft and ab initio computational study of thermodynamic and kinetic parameters. *J. Comp. Chem.*, 26(9):907–914, July 2005.

[368] E. R. Jamieson and S. J. Lippard. Structure, recognition, and processing of cisplatin-dna adducts. *Chem. Rev.*, 99(9):2467–2498, September 1999.

[369] J. Kozelka, R. Savinelli, G. Berthier, J. P. Flament, and R. Lavery. Force-field for platinum binding to adenine. *J. Comp. Chem.*, 14(1):45–53, January 1993.

[370] R. F. W. Bader. A quantum-theory of molecular-structure and its applications. *Chem. Rev.*, 91(5):893–928, July 1991.

[371] R. F. W. Bader. *Atoms in Molecules-A Quantum Theory.* Oxford University Press, 1990.

[372] S. T. Howard and O. Lamarche. Description of covalent bond orders using the charge density topology. *J. Phys. Org. Chem.*, 16(2):133–141, February 2003.

[373] A. Rybarczyk-Pirek, A. T. Dubis, S. J. Grabowski, and J. Nawrot-Modranka. Intramolecular hydrogen bonds in crystals of thiophosphorylbenzopyrane derivatives - x-ray and ft-ir studies. *Chem. Phys.*, 320:247–258, 2006.

[374] A. Robertazzi and J. A. Platts. Hydrogen bonding and covalent effects in binding of cisplatin to purine bases: Ab initio and atoms in molecules studies. *Inorg. Chem.*, 44(2):267–274, 2005.

[375] M. P. Waller, A. Robertazzi, J. A. Platts, D. E. Hibbs, and P. A. Williams. Hybrid density functional theory for pi-stacking interactions: Application to benzenes, pyridines, and dna bases. *J. Comp. Chem.*, 27(4):491–504, March 2006.

[376] F. Biegler-König and J. Schönbohm. Update of the aim2000-program for atoms in molecules. *J. Comp. Chem.*, 23(15):1489–1494, 2002.

[377] J. E. Carpenter and F. Weinhold. Analysis of the geometry of the hydroxymethyl radical by the "different hybrids for different spins" natural bond... *J. Mol. Struct. (Theochem)*, 169:41–62, 1988.

[378] S. Blasco, I. Demachy, Y. Jean, and A. Lledos. Dft study of the structural and redox properties of [cp2fe2s4](q) complexes (q=0,+2,+1 and-2). *New. J. Chem.*, 25:611–617, 2001.

[379] M. H. Baik and R. A. Friesner. Computing redox potentials in solution: Density functional theory as a tool for rational design of redox agents. *J. Phys. Chem. A*, 106(32):7407–7412, 2002. Cited By (since 1996): 31 Export Date: 21 December 2006 Source: Scopus.

[380] M. J. Clarke. Oncological implications of the chemistry of ruthenium. *Metal. Ions. Biol. Syst.*, 11:231, 1980.

[381] V. Exarchou, A. Troganis, I. P. Gerothanassis, M. Tsimidou, and D. Boskou. Do strong intramolecular hydrogen bonds persist in aqueous solution? variable temperature gradient h-1, h-1-c-13 ge-hsqc and ge-hmbc nmr studies of flavonols and flavones in organic and aqueous mixtures. *Tetrahedron*, 58:7423–7429, 2002.

[382] G. Costa, G. Balducci, E. Alessio, C. Tavagnacco, and G. Mestroni. Electrochemistry of sodium trans-bis(dimethyl sulphoxide) tetrachlororuthanate(iii) and mer-trichlorotris(dimethyl sulphoxide) ruthenium(iii). the first and complete electrochemical characterization of chlorodimethyl sulphoxide ruthenium(iii) complexes. *J. Electroanal. Chem.*, 296(1):57–76, 1990. Cited By (since 1996): 5 Export Date: 21 December 2006 Source: Scopus.

[383] A. Bergamo, G. Stocco, C. Casarsa, M. Cocchietto, E. Alessio, B. Serli, S. Zorzet, and G. Sava. Reduction of in vivo lung metastases by dinuclear ruthenium complexes is coupled to inhibition of in vitro tumour invasion. *Int. J. Oncol.*, 24(2):373–379, 2004. Cited By (since 1996): 4 Export Date: 21 December 2006 Source: Scopus.

[384] K. Büttner, S. Nehring, and K-P. Hopfner. Structural basis for dna duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol*, 14(7):647–652, Jul 2007.

[385] A. Bergamo, S. Zorzet, B. Gava, A. Sorc, E. Alessio, E. Iengo, and G. Sava. Effects of nami-a and some related ruthenium complexes on cell viability after short exposure of tumor cells. *Anti-Cancer Drugs*, 11(8):665–672, 2000. Cited By (since 1996): 15 Export Date: 21 December 2006 Source: Scopus.

[386] B. Gava, S. Zorzet, P. Spessotto, M. Cocchietto, and G. Sava. Inhibition of b16 melanoma metastases with the ruthenium complex imidazolium trans-imidazoledimethylsulfoxide-tetrachlororuthenate and down-regulation of tumor cell invasion. *J. Pharmacol. Exp. Ther.*, 317(1):284–291, 2006. Cited By (since 1996): 1 Export Date: 21 December 2006 Source: Scopus.

[387] J D. Watson and F. H. C. Crick. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171:737, 1953.

[388] Z. Vischer and Chargaff. Microbial nucleic acids: The desoxypentose nucleic acids of avian tubercle bacilli and yeast. *J.Biol.Chem.*, 177:429–438, 1949.

[389] W. K. Olson, A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin. Dna sequence-dependent deformability deduced from protein-dna crystal complexes. *Proc. Natl. Acad. Sci. USA*, 95(19):11163–11168, Sep 1998.

[390] Nomenclature committee of iub (nc-iub) and iupac-iub joint commission on biochemical nomenclature (jcbn). newsletter 1983. *Eur. J. Biochem.*, 131(1):1–3, 1983.

[391] K. Yanagi, G. G. Privé, and R. E. Dickerson. Analysis of local helix geometry in three b-dna decamers and eight dodecamers. *J. Mol. Biol.*, 217(1):201–214, Jan 1991.

[392] M. J. Gait G. M. Blackburn. *Nucleic acids in chemistry and biology*. Oxford University Press New York, 1996.

[393] C. J. Mundy M. E. Tuckerman and G. J. Martyna. On the classical statistical mechanics of non-hamiltonian systems. *Europh. Lett.*, 45(2):149–155, 1999.

[394] M. E. Tuckerman and G. J. Martyna. Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem. B*, 104(2):159–178, 2000.

[395] M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna. Non-hamiltonian molecular dynamics: Generalizing hamiltonian phase space principles to non-hamiltonian systems. *J. Chem. Phys.*, 115(4):1678–1702, 2001.

[396] M. Deserno and C. Holm. How to mesh up ewald sums. i. a theoretical and numerical comparison of various particle mesh routines. *J. Chem. Phys.*, 109(18):7678–7693, 1998.

[397] M. Deserno and C. Holm. How to mesh up ewald sums. ii. an accurate error estimate for the particle–particle–particle-mesh algorithm. *J. Chem. Phys.*, 109(18):7694–7701, 1998.