UNIVERSITÀ DEGLI STUDI DI PALERMO

Dottorato di Ricerca in Scienze Chimiche
Dipartimento di Fisica e Chimica
S.S.D. – CHIM/03

The Interaction of Metal Complexes with G-quadruplex DNA

IL DOTTORE
Angelo Spinello

IL COORDINATORE
Prof. Paolo Lo Meo

IL TUTOR
Prof. Giampaolo Barone

IL CO-TUTOR
Prof. Antonino Lauria

CICLO XXVI
2013-2015

Università degli Studi di Palermo

Abstract

Doctor of Philosophy

The Interaction of Metal Complexes with G-quadruplex DNA

by Angelo Spinello

A computational approach was proposed to study the binding and the stability of metal complex-nucleic acid supramolecular systems. In particular, the interaction of transition metal complexes with DNA structures named "G-quadruplexes" was considered. G-quadruplex conformations are present in telomeres and several oncogenes and they are involved in the inhibition of telomerase, a protein responsible for immortalization of cancer cells. The main purpose of the project was then to provide a computational tool to design chemical compounds able to selectively stabilize G-quadruplex structures.

Università degli Studi di Palermo

Abstract

Doctor of Philosophy

The Interaction of Metal Complexes with G-quadruplex DNA

by Angelo Spinello

Un approccio computazionale è stato proposto per lo studio dell'interazione di complessi metallici di basi di Schiff con DNA. Nel capitolo 2, è stato investigato il meccanismo di azione di complessi di Nichel(II), Rame(II) e Zinco(II) con B e G-quadruplex DNA. Il G-quadruplex è una conformazione non canonica adottata da particolari sequenze ricche in guanina. Recentemente, è stata dimostrata la sua esistenza in cellule umane, in regioni telomeriche e non telomeriche, ed è stato proposto come un possibile target per una nuova categoria di agenti antineoplastici. I capitoli successivi sono basati su dati raccolti durante due periodi di ricerca all'estero.

Nel capitolo 3, basato sugli studi eseguiti presso l'Università tecnica di Braunschweig, verrà mostrato come i campi di forza, oggigiorno di uso comune in Chimica/Biofisica Computazionale, siano in grado di riprodurre correttamente la stabilità relativa di G-quadruplex modello. Inoltre, è stata studiata in dettaglio l'interazione di una classe di leganti organici, noti G-quadruplex binders, con un modello di quadruplex parallelo. L'approccio computazionale ha messo in evidenza l'importanza del considerare esplicitamente la protonazione dei leganti.

Nel capitolo 4, risultato di una COST Short-Term Scientific Mission in Francia presso l'Université de Lorraine, è descritta la procedura usata per riprodurre gli spettri di Dicroismo Circolare delle principali conformazioni dei G-quadruplex.

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka' but 'That's funny...'"

Isaac Asimov

Contents

Chapter 1

Introduction

Contents

1.1 Double helical DNA

1.1.1 DNA structure

DNA, deoxyribonucleic acid, carries all the genetic information in cells. It is a linear polymer composed by nucleotides, each monomer is constituted by a phosphate group, a cyclic furanoside sugar (β-D-2-deoxyribose) and a nitrogenous base. The sugar is phosphorylated in the 5' position and substituted at C1' by one of four different heterocycles attached by a β-glycosil C1'-N linkage. Nucleotides contain either a purine (adenine, guanine) or a pyrimidine (thymine, cytosine, uracil in RNA) base (see Figure 1.1).

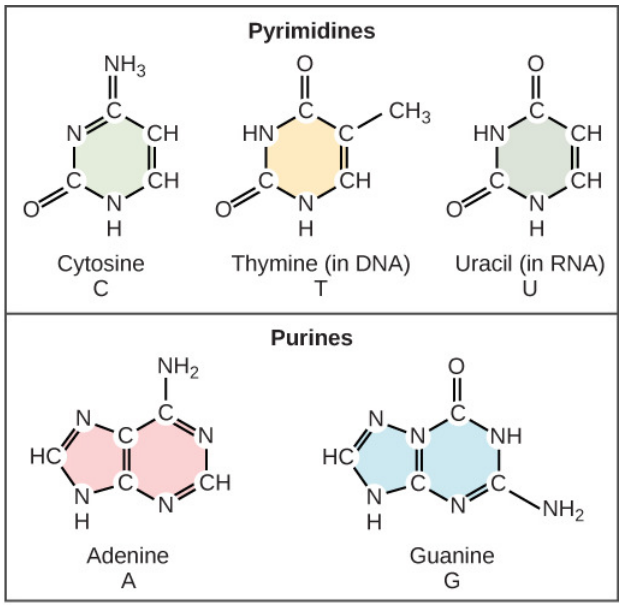

FIGURE 1.1: The structure of the four nitrogenous bases [1]

The most common DNA structure (duplex) was proposed by Watson and Crick in 1953 [2]. It is composed by two strands in a double helix conformation. Each strand has a backbone made up of deoxyribose molecules linked together by phosphate groups. The 3' C of a sugar molecule is connected through a phosphate group to the 5' C of the next sugar. This linkage is also called 3'-5' phosphodiester linkage. All DNA strands are read from the 5' to the 3' end where the 5' end terminates in a phosphate group and the 3' end terminates in a sugar molecule. Bases fit in the double helical model if a pyrimidine is always paired with a purine.

According to Chargaff's rules, the two strands will pair A with T and G with C. Two H-bonds can form between A and T, and three can form between G and C. This third H-bond in the G-C base pair is between the additional exocyclic amino group of guanine and the C2 keto group of cytosine. The pyrimidine C2 keto group is not involved in hydrogen bonding in the A-T base pair. Because glycosidic bonds branch off from one side of the base-pairs and because base-pairs are displaced from the helix axes, the outer envelope of the double helix is not cylindrically smooth but can display two grooves of different width and depth [3]. The major groove is wider than the minor groove in DNA, and many sequence specific proteins are able to interact with it . The N7 and C6 groups of purines and the C4 and C5 groups of pyrimidines face into the major groove, thus they can make specific contacts with amino acids in DNA-binding proteins. Specific amino acids serve as H-bond donors and acceptors to form H-bonds with specific nucleotides in the DNA. H-bond donors and acceptors are also in the minor groove, and indeed some proteins bind specifically in the minor groove [4].

1.1.2 Mechanism of DNA duplex - ligand binding

DNA-binders can be classified according to the type of association with DNA: covalent binding, major or minor groove binding and intercalation. As an example, it is well known that one of the most successful anticancer drugs, cisplatin [cis-diamminedichloroplatinum(II)], acts through the formation of coordination bonds with the N7 nitrogen atoms on two neighbouring guanine DNA bases (see Figure 1.2) [5]. In this form, cellular proteins cannot process or correctly repair DNA, the cell is therefore unable to replicate and dies. Nevertheless, cisplatin is effective in only a limited range of cancers, has lot of severe side-effects and for some other tumors often induces "resistance" during the treatment [7]. During the past years a number of platinum-based drugs with less side-effects and improved pharmacological properties were synthesized [8]. DNA-intercalation involves the insertion of a planar aromatic system between the base pairs of DNA, leading to significant π stacking and for this reason is favored by the presence of extended planar aromatic

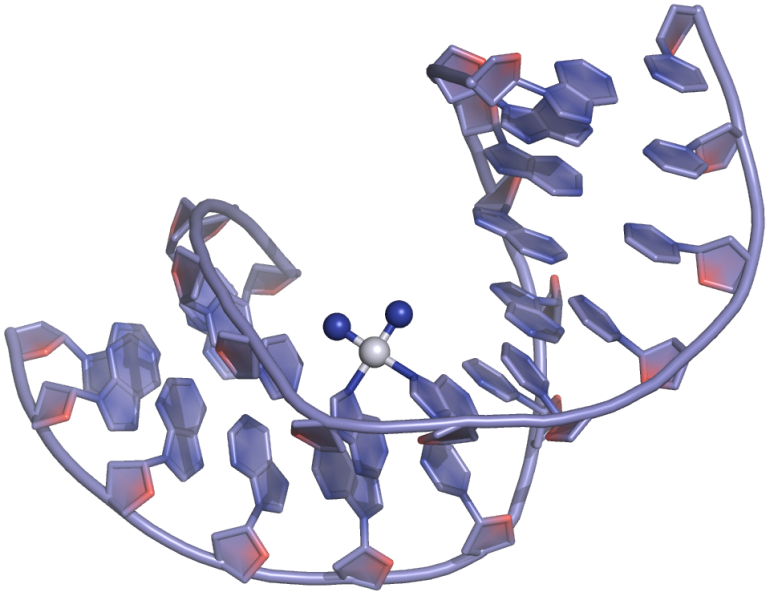

FIGURE 1.2: Schematic view of a cisplatin intra-strand adduct. The platinum atom is shown as a white sphere; the NH_3 ligands are shown as blue spheres [6].

ligand. Many different compounds have been proven to be DNA-intercalators, for example actinomycin D (see Figure 1.3) proflavine and ethidium bromide. The first, as an example, is a cyclic polypeptide-containing antibiotic that binds to DNA and inhibits RNA synthesis. It is known, by X-ray crystallography [9], that the phenoxazone ring system on actinomycin intercalates between adjacent base pairs, while pentapeptide chains lie in the narrow groove of the B helix and form hydrogen bonds with guanine residues on opposite chains. Generally, intercalators distort the canonical structure of the DNA double helix. There is a decrease in the twist angle between the base pairs around the intercalation site, and the DNA is elongated. All of these effects are usually reversible upon removal of the intercalator [10]. The biological effects of this class of DNA binders are exerted mostly interfering with the recognition and function of proteins that recognize the polynucleotide, such as polymerases and topoisomerases [11]. Although intercalation was proposed many years ago, it is still one of the most important mechanism of interaction with nucleic acids.

Canonical double helical B-DNA has two grooves, called major and minor. Proteins and large biomolecules, e.g. oligonucleotides, bind to B-DNA preferentially in the major groove, because it is a good receptor in terms of size, flexibility

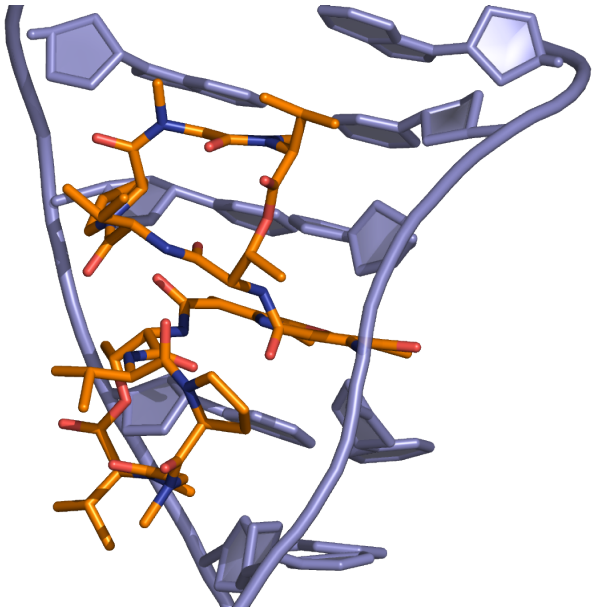

FIGURE 1.3: Schematic view that show the intercalation of actinomicyn D (orange) in double-stranded DNA [6].

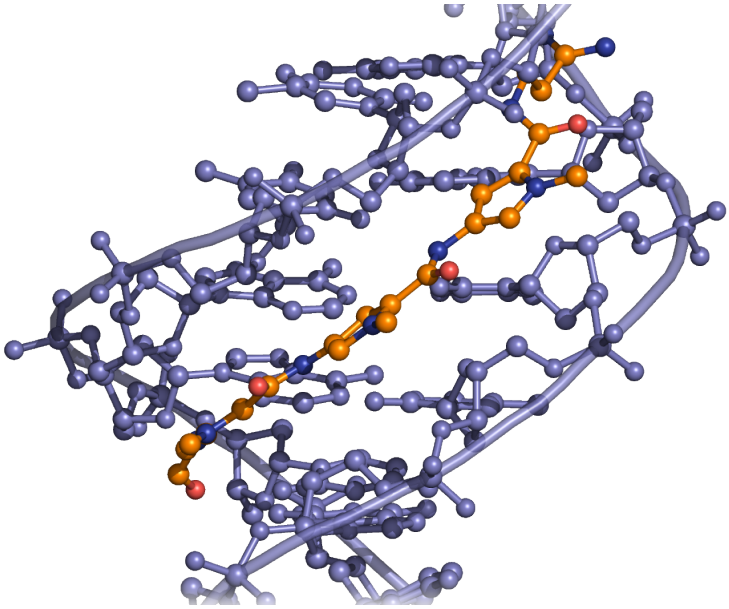

FIGURE 1.4: View from the three-dimensional structure of a complex between distamycin (orange) and a DNA duplex, showing the binding of distamycin in the minor groove [6].

and possibility to form H-bonding [12]. Small DNA binders prefer minor groove, where sequence recognition is limited compared to the major groove. Usually they possess a high degree of sequence specificity. The natural molecule distamycin A (see Figure 1.4), synthetic diarylamidines and bis-enzimidazoles (such as Hoechst 33258) are well-known groove binders. They are the most studied DNA-binding agents and they have shown several biological activities. Some have found clinical application in treating several diseases, like cancers, or as anti-viral and anti-bacterial agents [13]. Moreover, some metal complexes proves to be also DNA-groove binders. As an example, Pt(II) complexes of substituted 1,2,4-oxadiazole derivatives, have shown anticancer activity towards human ovarian cancer cell lines [14].

1.2 G-quadruplex DNA

1.2.1 G-quadruplex structure

G-quadruplexes (G4) formed by human telomeric DNA have become a focus of attention in recent years because of their role in important biological processes, such as aging and cancer [15], and potential as a therapeutic target for cancer [16, 17]. In quadruplexes, four guanine bases align in a pseudo-plane through hydrogen-bond alignments involving the Watson–Crick edge of a guanine and the Hoogsteen edge of its partner; resulting in a (G-G-G-G) tetrad (see Figure 1.5a). The quadruplex stem is composed of stacked tetrads with phosphodiester backbones delimiting cavities denominated grooves. The tetrads are held together by cations and interactions of p orbitals of stacked aromatic bases.

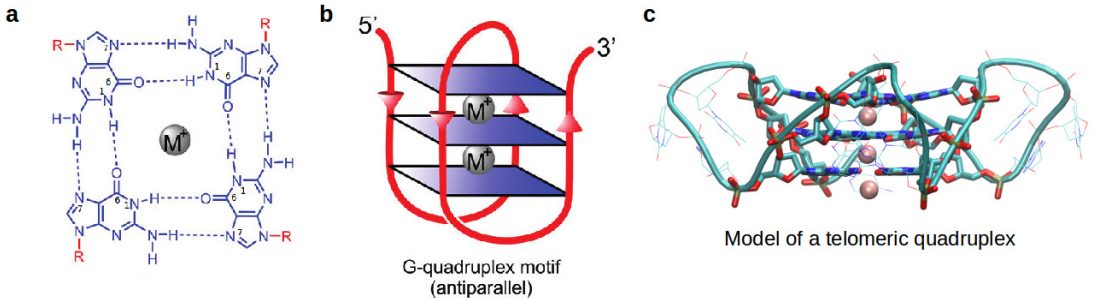

FIGURE 1.5: A quartet with Hoogsteen hydrogen bonds (a). A scheme that shows an example of G-quadruplex folding (b). A picture of a parallel telomeric G-quadruplex DNA (c).

The concept of targeting G-quadruplexes as a therapeutic strategy was first developed for telomeric DNA and telomerase inhibition, where the single stranded 3' G-rich ends of chromosomes are available to be stabilized as four stranded structures. The single stranded G-rich DNA is free of its complimentary C-rich strand which allows for novel folding topologies involving G-tetrad formation. In humans the single stranded 3' telomere ends are heterogeneous in length but typically extend up to 200 nucleotides (7-33 hexanucleotide repeats), allowing for many self-associations as well as binding to specific single stranded binding proteins. One protein in particular, telomerase, a reverse transcriptase that uses its own RNA template to hybridize to the DNA 3' end and synthesize additional d(TTAGGG)

repeats, maintains the 3' overhang. Telomerase was identified to be up regulated in over 85% of cancers but not in somatic cells [18]. This link to cancer biology propelled researchers to develop new strategies to interfere with telomere maintenance and alter cell growth through telomerase inhibition [19]. Recently, the existence of G-quadruplex structure in human cells was proved through the generation and application of an engineered, structure-specific antibody employed to quantitatively visualize DNA G-quadruplex [20]. Interestingly, their presence was also observed in non-telomeric regions. In fact, there are also regions of the human genome beyond the telomere that have potential quadruplex-forming ability. Suitable G-rich sequences were identified in a number of genes and in breakpoint regions before the human genome was fully sequenced. The first to be systematically studied was the promoter region of the c-myc oncogenes, and it is still of great interest because 1) c-myc is a ubiquitous control oncogens that plays an important role in several human cancers; 2) the MYC gene product is virtually not a proper target due to its instability; 3) targeting the c-myc quadruplex with small ligand has become a paradigm for the rapidly expanding field of promoter quadruplex targeting [21, 22].

1.2.2 G-quadruplex binders

An innovative strategy was developed to target the DNA substrate of telomerase by stabilizing G-quadruplex formation and inhibit hybridization. Several selective drug-like small molecule ligands were developed to target the quadruplex forming 3' telomeric ends that showed strong binding and an ability to stabilize these motifs and inhibit telomerase, however it was shown that they lacked selectivity towards telomere quadruplex motifs. Thus began structural investigations to understand the folded topologies of quadruplexes as therapeutic targets using a traditional structure based drug discovery approach. To stabilize G-quadruplex structures the binders should posses an aromatic system, able to do stacking interactions with the quartet, and positively charged side chains to improve both the water solubility and the electrostatic interactions with the negatively charged

phosphates. Moreover, a positive central charge can be useful to improve the affinity. It is possible that the positive charge will fall in line with the potassium channel, leading to a further stabilization of the adduct. Few years ago, it was proposed that metal complexes coordinated to aromatic ligands could have the potential to be potent quadruplex DNA binders, and this was subsequently demonstrated to be the case with several examples [23–25]. Recently, the first X-ray crystal structures of nickel(II) and copper(II) salphen metal complexes bound to a quadruplex DNA were presented [26]. Two structures were reported and show that these salphen-metal complexes bind to human telomeric quadruplexes by end-stacking, with the metal in each case almost in line with the potassium ion channel (see Figure 1.6).

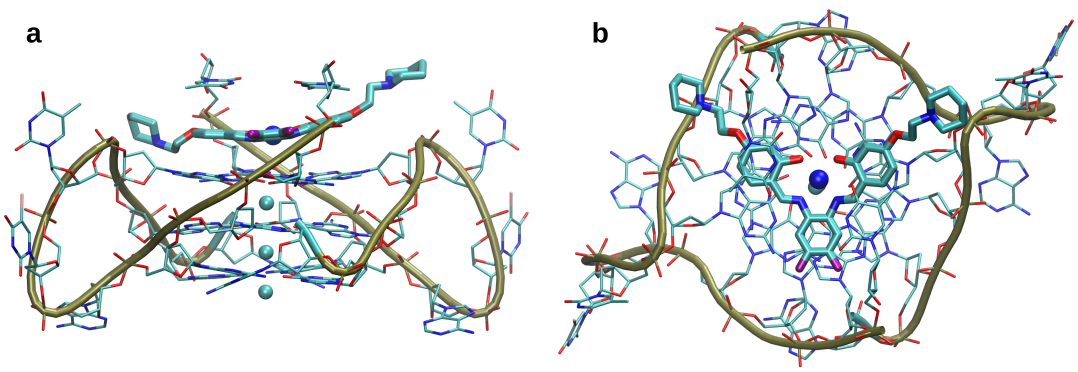

FIGURE 1.6: Side (a) and top (b) views of the crystal structure of a copper-salphene complex. Potassium atoms (cyan) form a canal in the middle of the structure, copper (blue) is in line with them [26].

In this concept, the metal center plays key structural, electronic, and electrostatic roles [27]. Structurally, it was envisaged that the metal can "organize" ligands into specific geometrical conformations ideally suited to $\pi - \pi$ stack onto a G-quartet. In addition, the metal can also have an important electronic role of "pulling" electrons from coordinated aromatic ligands, making them more electron-deficient and therefore more likely to be involved in substantial $\pi - \pi$ interactions with G-quartets or nitrogenous bases. Finally, the metal itself can also play an electrostatic role, being positively charged, and because it may be positioned at one end of the central ion channel in a quadruplex complex, it would thus occupy the

place that a stabilizing potassium ion would normally occupy at the top of the G-quartet stack at the quadruplex core [26].

1.3 Methods

In the present thesis work, the host-guest interactions described above have been investigated by the computational approaches described in this section.

Experimental techniques such as UV–vis spectroscopy or viscosity measurements, provide indirect evidence of the interaction between metal complexes and DNA. Unfortunately, they hardly are able to provide atomic-level structural details, such as the binding site or the kind of interaction, e.g. whether non-covalent or covalent, occurring between small molecules and DNA. Challenging experiments are necessary to detect such structural features, for example involving the application of X-ray crystallography or NMR spectroscopy techniques and the use of small synthetic oligonucleotides [28, 29]. On the other side, the knowledge of the structural and energetic details of the interaction between small molecules and nucleic acids is of crucial importance, because it can provide an atomistic model for the interpretation of the macroscopic properties related to such phenomenon. For this reason, several computational methods have been proposed as complementary tools to investigate the molecule-DNA interaction. Usually, classic molecular dynamics (MD) simulations in the presence of explicit solvent molecules have been performed to study the properties of DNA models and the non-covalent interactions of DNA with small molecules or with proteins [30]. In fact, empirical force fields are nowadays able to correctly reproduce the structural and dynamic features of DNA models as well as of their interaction complexes (force fields will be discussed in section 3.1.2) [31]. As an example, with this method, the groove binding of metal complexes with DNA dodecamers was successfully described [32]. Moreover, the geometry of $[Ru(phen)_2(dppz)]^{2+}$ (phen = 1,10-phenanthroline) intercalated into an adenine-thymine tetramer, $d(ATAT)_2$, including Na^+ counterions and only eight water molecules, was optimized by Car-Parrinello MD [33].

MD simulations have been also used to sample the DNA conformational space and select the most representative geometries, whose relative energy was successively calculated by density functional theory (DFT) methods [34]. In this regard, it should be pointed out that the accuracy obtained by empirical force field methods in the structural description of non-covalent interactions between small molecules and biomolecular systems has been checked and confirmed by quantum chemical calculations [35].

Recently, hybrid quantum mechanics (QM) and molecular mechanics (MM) methods are being increasingly considered [36]. It has been shown that these QM/MM methods provide structural and energetic results concerning the high level layer more reliable than those obtained by MM methods only. For example, the QM/MM methodology, with DFT used to describe the higher-level QM layer, allowed to successfully reproduce the coordination geometry and spectroscopic properties of metalloprotein active sites [37, 38]. Very often DFT and DFT/MM methods have been used for the study of the interaction between small molecules and oligonucleotides, e.g. by covalent [39, 40] and/or non covalent binding, with particular emphasis on hydrogen bonding or the $\pi-\pi$ stacking interactions [41, 42]. The combination of MD and DFT/MM approaches resulted successfully in the description of both non-covalent and covalent interactions of DNA models with main group metal ions [43] or ruthenium(II)-arene complex [44]. In this context, very few computational approaches have up to now been employed for describing the formation of both intercalation and covalent binding between metal complexes and DNA oligonucleotides. In particular, it is known that the most common DFT functionals fail to correctly describing the $\pi - \pi$ stacking interactions among planar aromatic molecules and DNA nitrogen bases [45]. In Chapter 2 it will be shown that using the M06-2X functional [46] as the DFT method of a DFT/MM calculation, it is possible to reliably describe both covalent and non-covalent binding between a metal complex and a DNA model. Recent studies showed that this functional correctly describes both hydrogen bonding and $\pi - \pi$ stacking interactions among DNA nucleobases in the presence of implicit solvent [47].

The methods that will be described below were successfully applied to the investigation of other biological macromolecules and their interaction with small ligands.

- The binding of a copper complex to B-DNA [48]
- Evaluating the mechanism of action of a small ligand, PTC124 (currently under clinical trial), with a small messanger RNA [49, 50]
- The determination of the quaternary structure of the chaperones GroEL and Hsp60 [51] and the interaction of a known binder, epolactaene, with Hsp60 [submitted]
- The inhibition of water and glycerol permeation in aquaporin 3 induced by mercury [52]
- Folding of small antimicrobial peptides [53] [submitted] and their interaction with biological membranes [in preparation]

1.3.1 Molecular Dynamics

The first molecular dynamics simulation of a macromolecule was published almost 40 years ago [54]. This simulation regarded the bovine pancreatic trypsin inhibitor (BPTI), which was considered the 'hydrogen molecule' of protein dynamics because of its small size, high stability and accurate X-ray structure. Although this simulation was performed in vacuum with a simple molecular mechanics potential and lasted only 9.2 ps, the trajectory clearly have shown that proteins are not rigid structures but dynamic systems, and their internal motions can play a important functional roles. MD simulations provide details concerning individual motions of atoms as a function of time [55]. In particular, they can be useful to answer specific questions about the properties of a system, more easily than performing experiments on the same system. Molecular dynamics studies the temporal evolution of the coordinates and the momenta of a given macromolecule. Such an evolution is

called a trajectory. A typical trajectory is obtained by solving Newton's equations of motion

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \tag{1.1}$$

where $r_i(t)$ is the position vector of ith particle and F_i is the force acting upon ith particle at time t and m_i is the mass of the particle. The trajectory is important in assessing numerous time-dependent observables [56] such as the accessibility of a given molecular surface [57], the interaction between a small molecule (e.g. dasatinib) and the Src kinase [58], amongst others. To integrate the differential equations it is necessary to specify the instantaneous forces that act on the particles and their initial positions and velocities. The MD trajectories are defined by both position and velocity vectors and they describe the dynamic evolution of the system during the time. Moreover, positions and velocities are propagated with a finite small time interval using numerical integrators, e.g. the Verlet algorithm. During the simulation the trajectories may be displayed and analyzed, providing important structural details. The dynamic events that may influence the functional properties of the system, for example conformational changes in proteins, can be observed with an atomistic detail. The forces acting on each atoms have to be recalculated at every step to update the positions and velocities in the numerical integration procedure. Force fields (FF will be discussed in Section 3.1.2) calculate the long-range electrostatic and dispersion interactions and in order to account for all the nonbonded pairs a summation of order N^2 has to be performed.

Thus, the overall MD algorithm is composed by the repeated calculation of the forces and many efforts were made to develop techniques to deal with the problem of the electrostatic interactions. Moreover, an aqueous solution and the presence of ions represent the typical environment for biological macromolecules and it has to be accounted for in reliable simulations. In order to speed up the calculations, only a finite sample of an extended and theoretically infinite system can be represented inside the model. The treatment of long-range interactions is strictly connected to the choice of boundary conditions (see Figure 1.7) that are imposed on the

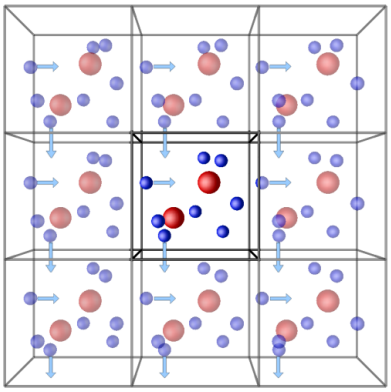

FIGURE 1.7: Schematic representation of the idea of periodic boundary conditions.

system during the simulation. Periodic boundary conditions (PBC) are used in molecular dynamics simulations to avoid problems with boundary effects caused by finite size. The existence of PBC means that any atom that leaves the simulation box, for example from the right-hand face, then enters the box by the left-hand face. Regarding the electrostatic treatment, recent algorithms, for example the particle mesh Ewald (PME) method, allows efficient computation of the long-range interactions without recurring to cutoff approximation. In this approximation, contributions from sites separated by distance larger than a certain cut-off are neglected [59].

1.3.2 QM/MM

Nowadays, it is still an extremely time-consuming task to calculate in an accurate way the structure and properties of biomolecular systems using QM methods. During the last decades many efforts have been made to make these kind of calculations practicable. One of the results of such efforts is the hybrid method [60]. In a hybrid method, a large biomolecule is divided into many part. Different methods are then applied to the fragments, going from very expensive and accurate QM methods to less expensive and accurate molecular mechanics (MM) methods. The hybrid method was initially developed by Honig and Karplus in 1971 [61]. This was further improved and extended by Warshel and Karplus [62] and was formalized in a more general QM/MM scheme by Warshel and Levitt [63]. Karplus,

Levitt, and Warshel were awarded the 2013 Nobel Prize in Chemistry, for the "Development of Multiscale Models for Complex Chemical Systems" [64–66]. In the QM/MM formalism the whole molecular system is divided into two sections: the model and the environment. The small model part, the more interesting layer of the system, is treated with an accurate QM method. The environment part is treated with a more efficient but less accurate MM method (which cannot describe bond breaking/formation or electronic state). In this scheme, as shown in equation:

$$E_{QM/MM} = E_{QM} + E_{MM} + E_{QM-MM} \tag{1.2}$$

the total energy of the entire system, $E_{QM/MM}$, is the sum of the energy of the model system by the QM method (E_{QM}), the energy of the environment system by the MM method (E_{MM}), and the interactions (E_{QM-MM}) between the QM model system and the MM environment system. For this reason, the former scheme is called an additive [67], in which the energies of the two systems and the interactions between the two systems are added to obtain the total energy of the whole system. The ONIOM method instead is a subtractive scheme [68]. ONIOM can be considered as a hybrid method based on a different concept from the one mentioned above. Although ONIOM can be used as a two-layer QM/MM method, it can also combine different QM and QM methods, and can easily be extended to more than two layers. In a two layer ONIOM calculation, the total energy of the system is obtained from:

$$E^{ONIOM(QM/MM)} = E^{QM}_{model} + E^{MM}_{real} - E^{MM}_{model} = E^{high}_{model} + E^{low}_{real} - E^{low}_{model} \tag{1.3}$$

the real system contains all the atoms, and is calculated only at the low level. The model system contains the more interesting part of the system and is treated at the QM level, along with the atoms that are used to cap dangling bonds resulting from the cut of covalent bonds between the QM and the MM regions. To evaluate

the ONIOM energy, both QM and MM calculations need to be performed for the model system. Because the positions of the link atoms are defined in terms of the atoms in the real system, the potential energy surface (PES), and therefore geometry optimization, is well defined [69].

1.3.3 Conformational searches

Conformational search techniques can be classified broadly into two categories: stochastic and deterministic methods. Stochastic methods (e.g. Monte Carlo, simulated annealing, genetic algorithm, etc.) rely on a probabilistic descriptions to aid the location of the global minimum, and there is no natural endpoint to the procedure. On the other hand, deterministic methods (e.g. systematic search method, etc.) provide a certain level of assurance in locating the position of the global minimum and there is a defined endpoint to the procedure. Monte Carlo methods are stochastic techniques, in which the energy E_0 is calculated starting for an arbitrary conformation. New conformations are generated by randomly changing the dihedral angles, and the energy E is calculated. These conformations are accepted or rejected depending on the Boltzmann Factor:

$$BF = e^{-(E-E_0)/RT)} \tag{1.4}$$

This factor is compared with a random number between 0 and 1. If the Boltzmann Factor is greater than this random number, the new conformation is accepted; else, it is rejected. The iterative process is continued until a set of low energy conformers has been generated. Several modified Monte Carlo methods have been applied for peptides and proteins [70]. In contrast to Molecular Dynamics, Monte Carlo simulations do not provide information about time evolution. Nevertheless, they provide an ensemble of representative configurations and therefore conformations from which probabilities and relevant thermodynamic observables (e.g. free energy) may be calculated. Monte Carlo simulations also play a fundamental role when designing complex and hybrid molecular dynamic (MD) algorithms [71].

Chapter 2

Nickel(II), Copper(II), Zinc(II) Schiff base complexes: G-quadruplex vs. duplex DNA-binding

Contents

2.1 Introduction

The condensation of an amine with an aldehyde to form a Schiff base is an old and well known reaction [72]. The condensation of a salicylaldehyde with 1,2-diaminoethane, 1,2-phenyldiamine or 2,3-diaminonaphtalene gives respectively molecules known with the acronyms of "Salen", "Salphen" and "Salnaphen" [73, 74].

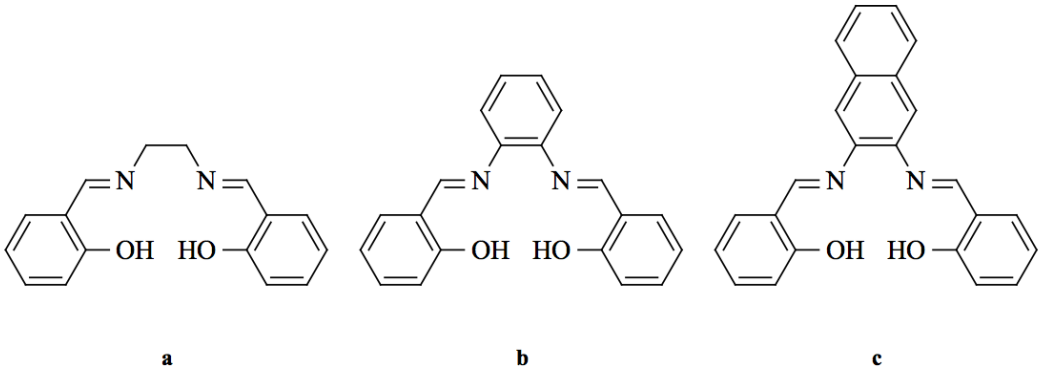

FIGURE 2.1: Chemical structures of H_2Salen (N,N'-bis(salicylidene)ethylene diamine) (a), H_2Salphen (N,N'-bis(salicylidene)-1,2-phenylenediamine) (b), H_2Salnaphen (N,N'-bis(salicylidene)-2,3-naphtalenediamine) (c).

Schiff bases are a class of well-known ligands in coordination chemistry. Because of its Lewis basicity, the imino nitrogen form coordination bonds with metal ions [75]. The ligands in Figure 2.1 posses four coordination sites, two oxygens and two nitrogens, in a square planar conformation. Thus, they are able to coordinate transition metals in the equatiorial position, leaving two axial sites free for the potential coordination of other ancillary ligands. Salen metal complexes have been extensively used in transition-metal chemistry for a large variety of applications, for example in catalysis [76]. Besides, Salen metal complexes have found several biomedical applications, for example as anti-viral drugs, as models for superoxide dismutase and as marker for several pathological conditions including ovarian cancer [77]. The presence of a planar aromatic moiety confers to these complexes a remarkable affinity towards DNA structures. In fact, there are several reported studies of the interaction of Salphen and Salnaphen metal complexes with duplex-DNA [78]. Moreover, it is well-known that Schiff base metal

complexes are a potent class of a G-quadruplex stabilizers. As an example Vilar et al. recently have reported the synthesis of two Nickel and Copper square planar Salphen metal complexes able to bind to human telomeric G-quadruplexes by end-stacking. Moreover, the complexes show significant antiproliferative activity in a panel of cancer cell lines [26].

The subject of this chapter will be the interaction of novel Nickel(II) (**1**), Copper(II) (**2**) and Zinc(II) (**3**) complexes of a Salnaphen-like ligand with duplex and G-quadruplex DNA (see Figure 2.2. Such studies have been recently published [79, 80]. Electrostatic interactions between the positively charged and the negatively charged phosphate backbone should strengthen the affinity with both DNA conformations.

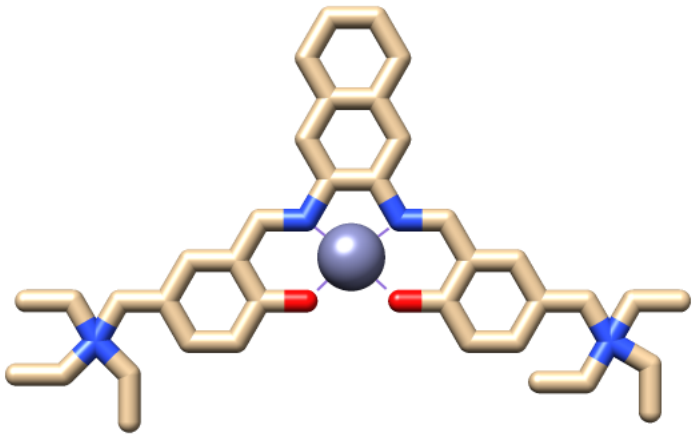

FIGURE 2.2: Nickel(II), copper(II) and zinc (II) complexes, 1-3, of N,N'-bis-5-(triethyl ammonium methyl)-salicylidene-2,3-naphtalendiiminato).

In the last paragraph the photophysical and DNA-binding properties of the cationic zinc(II) complex of 5-triethyl ammonium methyl salicylidene orthophenylendiiminato are reported, which have been investigated by a combined experimental and computational approach and which have been recently published [81]. Fine structural details of the molecule-DNA binding mode in solution cannot be obtained by solution techniques such as UV-vis absorption, CD and viscometry. Nonetheless, above-mentioned techniques are interesting because they can be routinely applied to biological samples in physiological conditions. On the other hand,

computational chemistry may support the interpretation of experimental data with atomistic models. In this context, quantum mechanics/molecular mechanics (QM/MM) calculations can provide reliable structural information on the local region of biomolecular systems. It was recently shown that Molecular Dynamics (MD) simulations, followed by QM/MM calculations, with the QM part described by density functional theory (DFT), provided a detailed structure of the intercalation complexes of dodecanucleotide double-helices with a copper(II) metallointercalator, and formation energy data in a good agreement with the experimental DNA-binding constant [48].

2.2 Interaction of complexes 1-3 with duplex DNA

2.2.1 Experimental Data

UV-vis absorption, circular dichroism (CD) and viscometry titrations provided clear evidence of the intercalative mechanism of the three square-planar metal complexes, allowing to determine the intrinsic DNA-binding constants (K_b), equal to 1.3×10^7, 2.9×10^6, $6.2\times10^5 M^{-1}$ for **1**, **2** and **3**, respectively. Preferential affinity, of one order of magnitude toward AT compared to GC base pair sequences was detected by UV-vis absorption titrations of **1** with $[poly(dG\text{-}dC)]_2$ and $[poly(dA\text{-}dT)]_2$.

2.2.2 MD simulations and DFT/MM calculations

Due to the higher affinity towards polyAT DNA, MD simulations have been performed on the complex **3**/$[dodeca(dA\text{-}dT)]_2$. The root mean square deviation (RMSD) along the simulation for all non-hydrogen atoms is shown in Figure 2.2, as can be seen the dodecamer is stable during the simulation time. The metal

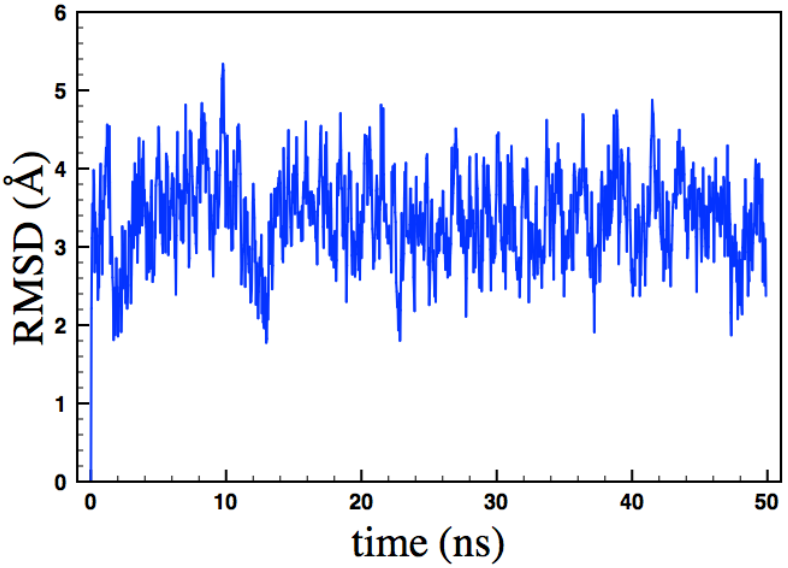

FIGURE 2.3: RMSD of all non-hydrogen atoms of $[dodeca(dA\text{-}dT)]_2$.

TABLE 2.1: Calculated coordination distances (Å) of the metal ion of **1**, **2**, **3** (M=Ni, Cu, Zn) with the oxygen atom of thymine O4 and with amine nitrogen atom of adenine N6, in the intercalation pocket of $[dodeca(dA\text{-}dT)]_2$.

	1	**2**	**3**
O4	3.02	2.57	2.94
N6	3.46	3.05	2.41

complex remains inside the binding pocket for the whole simulation time. The equilibrium geometry, after about 50 ns, has been used as starting point for further geometry optimizations, by hybrid two-layer QM/MM calculations, using DFT as QM method and the Amber99 force field as MM method, of the intercalation complexes of the three metal complexes **1-3** with $[dodeca(dA\text{-}dT)]_2$ (Figure 2.4). The higher level of the DFT/MM structures shown in Figure 2.4 involves the four nitrogen bases of the intercalation pocket and the metal complex. The selected DFT functional, M06-2X, allows to reliably describe weak interactions in implicit solvent, notably H-bond and π - π stacking [47] that are fundamental in the binding between the metal complex and the DNA biomolecule. Important structural details can be obtained by the analysis of the optimized structures reported in Figure 2.4, which provide atomic level models explaining the strong DNA-binding experimentally detected. In fact, three are the complementary contributions ruling the DNA-metal complex interaction: 1) the electrostatic attraction between the two positively charged residues of the metal complexes, the triethylammonium-methyl groups and the two positively charged phosphate groups; 2) intercalation of the naphthalene moiety between the stacked and H-bonded nitrogen bases, as summarized in Table 2.1.

Concerning the first contribution, electrostatic attraction, it is interesting to see that the triethylammoniummethyl groups are close to the nearest-neighbors phosphate groups of the DNA backbone, with minimum P-N distances (of the phosphate and of the triethylammoniummethyl groups, respectively) in the range 5-11 Å for the three metal complex-dodecanucleotide systems. With regards to the intercalation of the naphthalene residues between the stacked and H-bonded DNA bases, the pictures in Figure 2.4 show that such a process is more pronounced

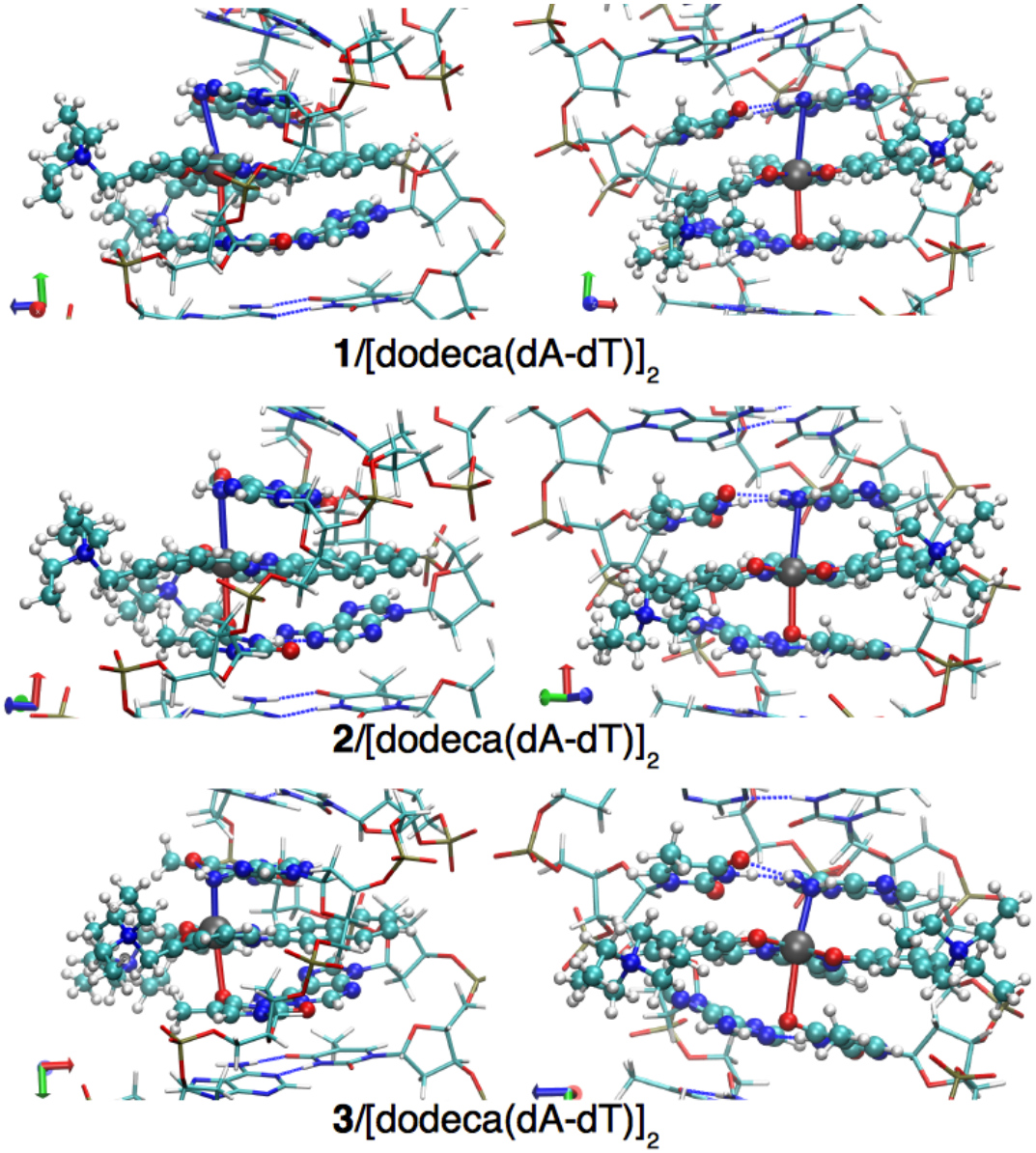

FIGURE 2.4: Two different views of the intercalation site of the supramolecular complexes between the three metal complexes [**1** - **3** with $[dodeca(dA\text{-}dT)]_2$, highlighting the interatomic distances of the metal ion with the O4 keto atom of thymine [in red, 3.02 Å(**1**), 2.57 Å(**2**) and 2.94 Å(**3**)] and with N6 amine atom of adenine [in blue, 3.46 Å(**1**), 3.05 Å(**2**) and 2.41 Å(**3**)]. DFT and MM regions are represented by "ball and stick" and "sticks" styles, respectively.

TABLE 2.2: Formation energy (kJ/mol) in terms of standard enthalpy (ΔH°) and Gibbs free energy (ΔG°), calculated at 298.15 K for the complexes of **1**, **2** and **3** with [dodeca(dA-dT)]$_2$.

Model system	ΔH° (vacuo)	ΔG° (vacuo)	ΔG° (water)
1/[dodeca(dA-dT)]$_2$	-141.1	-96.5	-58.2
2/[dodeca(dA-dT)]$_2$	-153.3	-89.5	-49.0
3/[dodeca(dA-dT)]$_2$	-192.8	-171.2	-32.7

for the nickel complex **1** and, in decreasing order, for the copper and zinc ones **2** and **3**. This trend is highlighted by the distortion from linearity of the N-M-O angle between the metal ion and axial exocyclic atoms N and O of the amine and keto groups interatomic distaces, shown in blue and in red, respectively. In details, this angle assumes the following values: 172 °, 175 ° and 168 ° for **3**, **2** and **1**, respectively. Concerning metal coordination, it has been reported that N,O coordination distances of 2.5-3.0 Å, revealed by X-ray crystallography for copper and zinc in active sites of biomolecules, are considered weak coordination bonds [82]. Moreover, the DFT/MM calculations support the hypothesis that, in the intercalation complexes with DNA, while the copper and zinc ions are axially coordinated in an octahedral environment, the Nickel ion is not and preserves its square planar coordination geometry (Table 2.1). Standard enthalpy and Gibbs free energy values, calculated at 298.15 K, were used to evaluate, *in vacuo* and in solution, the formation energy of the supramolecular complexes between **1**, **2** and **3** with [dodeca(dA-dT)]$_2$ (Table 2.2).

The analysis of Table 2.2 allows to make interesting considerations on the energetic contributions involved in the DNA binding of the three metal complexes. First, the formation of the intercalation complexes is always exothermic, *in vacuo* and in solution. Although, both entropy and solvation seem to destabilize the DNA-binding energy. The structure and electronic properties of the intercalation complexes may provide an explanation of the observed trends. The role of the solvent can be rationalized taking into account that is a considerable electrostatic character in the solvation energy that is screened going from the gas phase to water solution. However, the solvent destabilization of the zinc(II) complex **3** is much

larger than those of the nickel(II) and copper(II) complexes **1** and **2**. This result finds a nice support in the calculated APT charge on the zinc atom, that is about 1.4, also in the intercalation complexes, while those on the nickel and copper atoms are about 1.0 and 1.2, respectively. The formation free energy is always smaller than the formation enthalpy, both *in vacuo* and in solution, indicating that the entropic contribution, in the equation $\Delta G^\circ = \Delta H^\circ$ - T ΔS°, is always negative. However, such entropic destabilization is larger for the intercalation complexes of $[\text{dodeca(dA-dT)}]_2$ with **1**, **2** and smaller for that with the zinc complex **3**. This result also is in agreement with the DNA-binding affinity, that is greater for **1** and **2** and smaller for **3**.

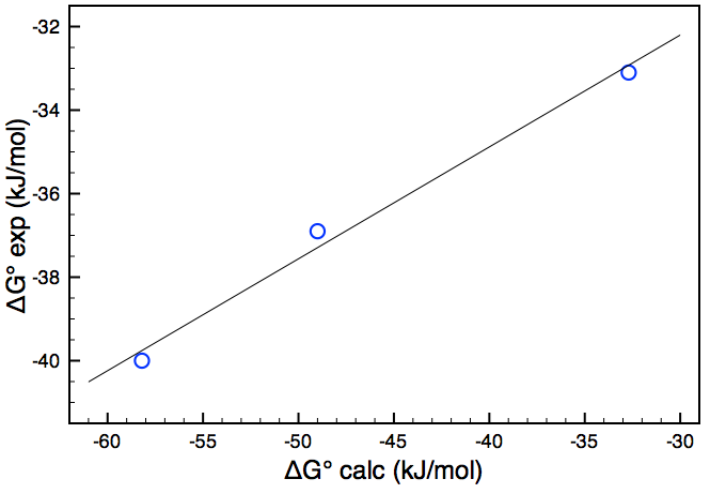

FIGURE 2.5: Linear correlation plot between the experimental and calculated formation free energies for **1**, **2** and **3** with native ct-DNA and with $[\text{dodeca(dA-dT)}]_2$, respectively.

Remarkably, an excellent agreement exists between the calculated formation free energy values of the intercalation complexes with $[\text{dodeca(dA-dT)}]_2$ and the analogous formation free energy values experimentally obtained from the DNA-binding constants reported in Table 2.1, as ΔG° = - $RTln(K_b)$. In fact, the latter are -40.0, -36.9 and -33.1 kJ/mol, for **1**, **2**, **3**, respectively. Moreover, the experimental DNA-affinity trend **1** > **2** > **3** is reproduced by the calculation. In fact, a good linear correlation (R=0.995) was found between the experimental and calculated formation Gibbs free energies, as shown in Figure 2.5. In conclusion, although it

is known that there are ten possible ways in which four nitrogen bases can produce intercalation sites [83], the DNA model selected for our calculations correctly describes the DNA-binding interaction mechanism of the three metal complexes.

2.2.3 Conclusions

Detailed information on the DNA-binding of three square planar nickel(II) (**1**), copper(II) (**2**) and zinc(II) (**3**) Schiff-base complexes was obtained through the application of MD simulations and quantum mechanics/molecular mechanics calculations. Spectroscopic and viscometry measurements confirmed that **1**, **2** and **3** are DNA-intercalators with DNA-affinity decreasing in the order Ni > Cu > Zn and with selective affinity of the three metal complexes toward AT with respect to GC sequences. DFT/MM calculations provided detailed local information on the DNA-binding site, which consists of 1) electrostatic attraction between the lateral cationic groups of the metal complexes and the anionic phosphate groups of the DNA backbone, 2) the DNA-intercalation of the planar aromatic fragment of the chelating Schiff-base ligand, 3) the metal ion coordination by the exocyclic keto-oxygen and amine-nitrogen bases. The values of the DNA-binding constants and their decreasing trend in the order **1** > **2** > **3** are correctly reproduced by the calculated formation Gibbs free energy values of the supramolecular complexes of **1**, **2** and **3** with $[\text{dodeca(dA-dT)}]_2$ in solution.

2.3 Interaction of complexes 1-3 with G-quadruplex DNA

2.3.1 Experimental Data

The affinity of the complexes **1-3** towards wild-type human telomeric (*h-Telo*) and protooncogene *c-myc* G-quadruplex (G4) DNA was also investigated. Binding constants (K_b) were determined by spectrophotometric titrations for G4-DNA and duplex-DNA. The results obtained point out that the three metal complexes selectively bind G4-DNA with higher affinity, up to two orders of magnitude, with respect to duplex-DNA. The nickel(II) complex **1** was found to be the most effective G4-DNA stabilizer and the K_b values decrease in the order **1** > **2** ∼ **3**. interestingly, **1** is able to induce G4-DNA formation of *h-Telo* sequences, also in the absence of K^+ cations.

2.3.2 MD simulations and DFT/MM calculations

MD simulations have been performed on the two complexes between **3** and both *c-myc* and *h-Telo* sequences in G4-DNA. The RMSD along the simulation, for all non hydrogen atoms and for guanine bases, are shown in Figure 2.6. The results of the MD simulations show that the interaction between the zinc(II) complex **3** and both G4-DNA models is driven by the strong electrostatic attraction between the positively charged triethylammoniummethyl groups of the ligand and the negatively charged DNA phosphate groups. This long range interaction allows the metal complex to easily approach the biomolecule. Moreover, a strong $\pi - \pi$ stacking interaction between the naphthalene moiety of the ligand and the terminal G-tetrad is present at the equilibrium. The three snapshots reported in the RMSD plot in Figure 2.6 describe the interaction of the zinc(II) complex toward the *c-myc* G4 structure. It is worth noting that the guanine basis lying above the terminal G-tetrad, colored in green, performs a rotation at about 100

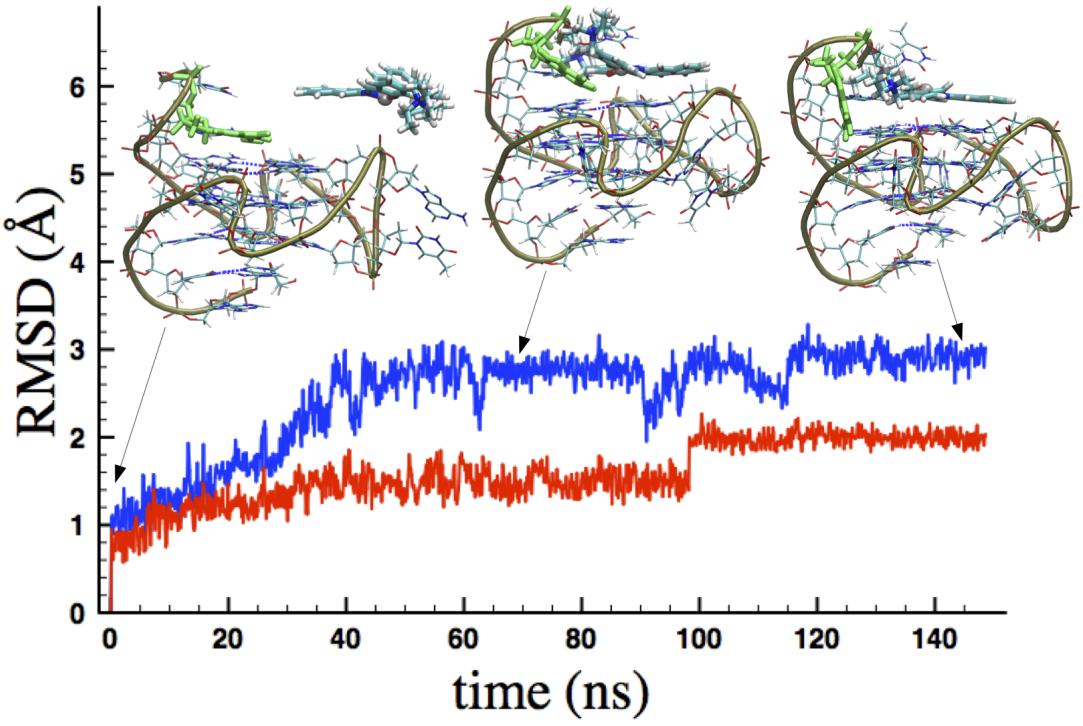

FIGURE 2.6: Plot of the RMSD obtained for **3**/*c-myc* up to 150 ns of MD simulations. Representative snapshots are also reported.

ns, as highlighted by the step in the RMSD of the guanine bases (red line in Figure 2.6). This rotation allows a better stacking interaction between the aromatic moiety of the metal complex and the tetrad of *c-myc* quadruplex. Concerning the interaction with the telomeric model, the RMSD plot in Figure 2.7 shows that the equilibrium is quickly reached after about 5 ns and the stacked metal complex remains tightly bound to the biomolecule up to the end of the MD simulation. The equilibrium geometry, after about 50 ns, has been used as starting point for further geometry optimizations, by hybrid two-layer QM/MM calculations, using DFT as QM method and the Amber99 force field as MM method, of the intercalation complexes of the three metal complexes **1-3** with *h-Telo* in G4 conformation (Figure 2.8). The DFT/MM structures shown in Figure 2.8 provide interesting atomistic details of the binding complexes, explaining the strong DNA-binding experimentally detected. In particular, the metal ion of the three complexes is in line with the two potassium cations in the central channel formed by the three stacked G-tetrads. Moreover, a weak metal coordination occurs in **2**/*h-Telo* and

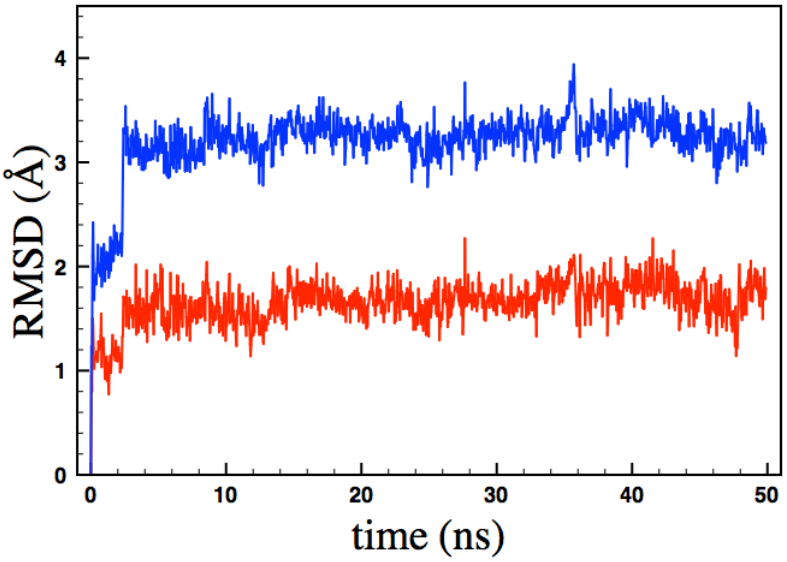

FIGURE 2.7: All non-hydrogen atom (blue) and guanine residues (red) plot of the RMSD obtained for **3**/*h-Telo* up to 50 ns of MD simulations.

3/*h-Telo*, by one of the four O6 keto oxygen atoms of the guanine bases, as reported in Figure 2.8. Such metal coordination and the concomitant distortion of the square planar geometry of the complexes, that decreases in the order Zn > Cu > Ni. Together with consideration on the solvent and thermodynamic contribution, provide an explanation of the decreasing affinity order, Ni > Cu ∼ Zn, experimentally detected, between the three complexes and G4-DNA. Standard enthalpy and Gibbs free energy values, calculated at 298.15 K, were used to evaluate, *in vacuo* and in solution, the formation energy of the supramolecular complexes between **1**, **2** and **3** with *h-Telo* G4-DNA (Table 2.3). The tabulated data allow to make interesting considerations of the energetic contribution involved in the G4-DNA binding of the thee metal complexes with duplex-DNA (compare with Table 2.2). First, the binding with the biomolecule is always accompanied by a

TABLE 2.3: Formation energy (kJ/mol) in terms of standard enthalpy (ΔH°) and Gibbs free energy (ΔG°), calculated at 298.15 K for the complexes of **1**, **2** and **3** with *h-Telo* G4-DNA.

Model system	ΔH° (vacuo)	ΔG° (vacuo)	ΔG° (water)
1/*h-Telo*	-233.9	-122.6	-34.6
2/*h-Telo*	-242.5	-107.6	-14.4
3/*h-Telo*	-290.9	-154.9	-20.9

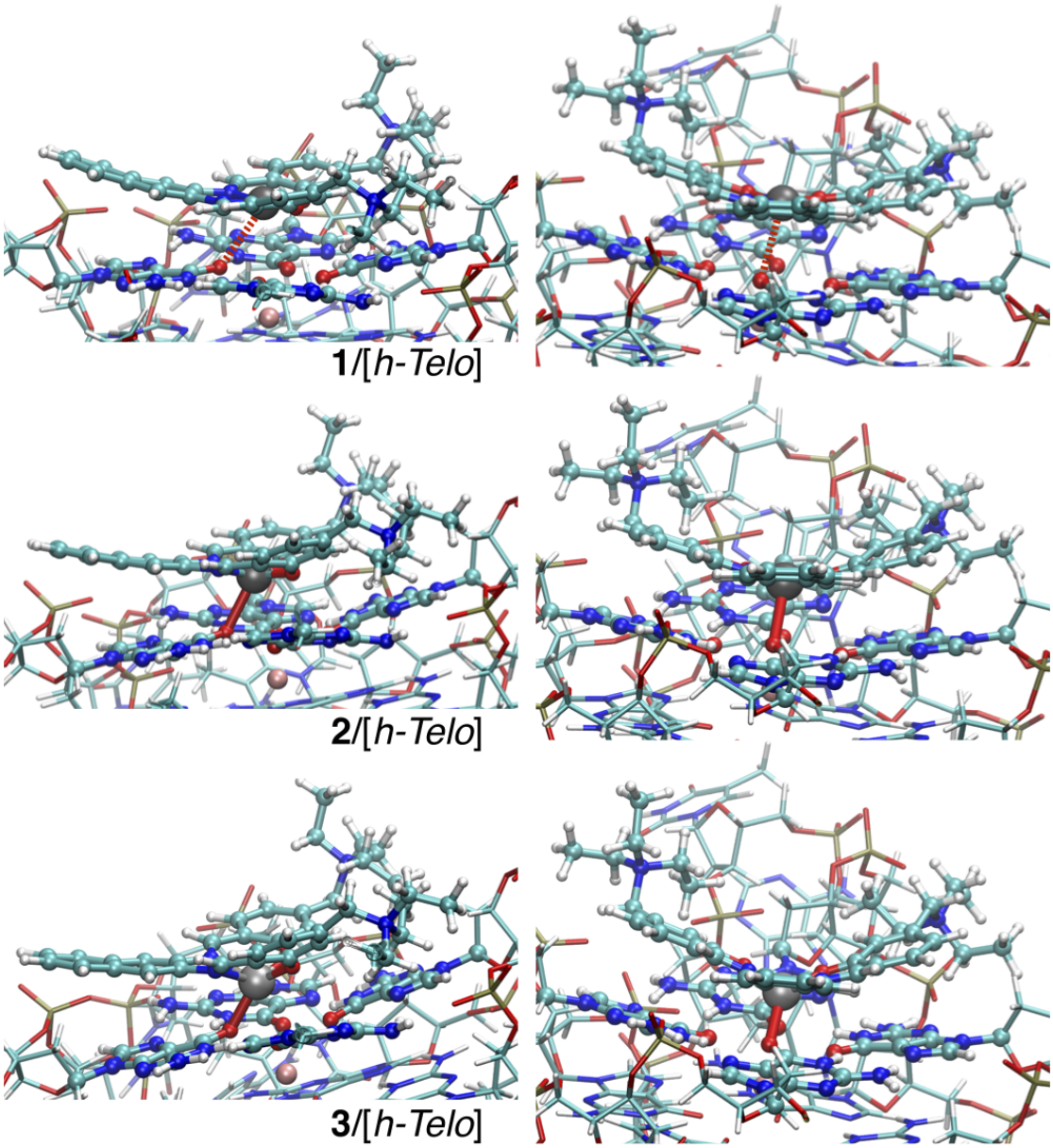

FIGURE 2.8: Two different views of the binding site of the supramolecular complexes between the three metal complexes **1** - **3** with *h-Telo* G4, highlighting the interatomic distances of the metal ion with one of the four O6 keto atoms of guanine (in red, 3.12 Å(**1**), 2.46 Å(**2**) and 2.12 Å(**3**)). DFT and MM regions are represented by "ball and sticks" and "sticks", respectively.

strong exothermic contribution, both *in vacuo* and in solution. However, both entropy and solvation play a destabilizing effect on the DNA-binding energy. The role of the polar solvent can be rationalized taking into account that there is a considerable electrostatic character in the interaction energy, which is screened going from the gas phase to water solution. The solvent destabilization decreases in the order Zn > Cu > Ni, in parallel with the decrease of the calculated charges, *in vacuo*, of the three ions in the binding complexes shown in Figure 2.8, 1.61, 1.34 and 1.04, for Zn, Cu, Ni, respectively. The formation free energy is always smaller than the formation enthalpy, both *in vacuo* and in solution, indicating that the entropic contribution, in the equation $\Delta G^\circ = \Delta H^\circ$ - T ΔS°, is always negative. Moreover, such entropic destabilization is lower for the complexes of *h-Telo* with **1** and higher for **2** and **3**. This result is probably related to the existence of a chemical bond between the exocyclic keto oxygen O6 and both the copper and zinc ions in **2**/*h-Telo* and **3**/*h-Telo*, while this coordination bond does not form with the nickel ion in **1**/*h-Telo* (see Figure 2.8). Finally, there is a good agreement between the calculated formation free energy values in solution, -34.6, -14.4 and -20.9 kJ/mol, and the experimental values, -36.2, -30.3 and -30.2 kJ/mol. The latter were obtained by the equation ΔG° = - $RTln(K_b)$ and using the K_b values reported in Table 2.3 for the interaction of **1**, **2** and **3** with *h-Telo* G4-DNA.

2.3.3 Conclusions

Detailed information on the binding of three square planar Nickel(II) (**1**), Copper(II) (**2**) and Zinc(II) (**3**) Schiff-base complexes with two quadruplex models was obtained through the complementary application of molecular dynamics simulations (MD) and quantum mechanics/molecular mechanics (DFT/MM) calculations. The results confirmed that (**1**), (**2**) and (**3**) are strong G4-binders, with affinity decreasing in the order Ni > Cu $\sim$ Zn and with selective affinity of the three metal complexes toward G4-DNA compared to duplex-DNA. In particular, the nickel compound (**1**) binds the telomeric quadruplex 100 times stronger that the duplex and this value is among the highest reported in the literature. MD

simulations provide a possible interaction mechanism between the zinc complex (**3**) with both *c-myc* and *h-Telo* G4-DNA, while DFT/MM calculations provided detailed local information on the DNA-binding site and an explanation of the solvent and thermodynamics contributions in the binding with the biomolecules. In particular, the higher entropic destabilization following the formation of both (**2**)/*h-Telo* and (**3**)/*h-Telo*, compared to the (**1**)/*h-Telo* complex, follows the coordination of the apical empty site of the copper and zinc ions by exocyclic keto-oxygen of a guanine base in the terminal G-tetrad, while the nickel ion maintains its square planar coordination geometry of the isolated Schiff-base complex. The values of the DNA-binding constants and their decreasing trend in the order (**1**) > (**2**) $\sim$ (**3**), are correctly reproduced by the calculated formation Gibbs free energy values of the supramolecular complexes of (**1**), (**2**) and (**3**) with *h-Telo* G4-DNA in solution.

2.4 An investigation of the spectroscopic properties of a DNA-intercalator Zn Salphen-type complex

As no experimental intercalated structures are available for Zn-Salphen-like complex, $ZnSalphen^{2+}$, two possible pockets of the intercalated system were obtained in alternating AT and GC dodecanucleotides from MD simulations. The metal complex remains tightly intercalated between the 5^{th} and 6^{th} bases for further geometry optimizations, by hybrid two-layer QM/MM calculations, using DFT as QM method and the Amber99 force field as MM method. The double stranded DNA pockets were modeled using the $d(ApT)_2$ and $d(GpC)_2$ simplified models, which limit the interaction between $ZnSalphen^{2+}$ and the DNA to the first neighborhoods. The inclusion of at least the effects of the lateral sugar/phosphate chains in addition to the stacking bases, is crucial in order to obtain a reliable description of the short range interactions between the DNA-bond molecule and the pocket; the selected model systems are reported in Figure 2.9.

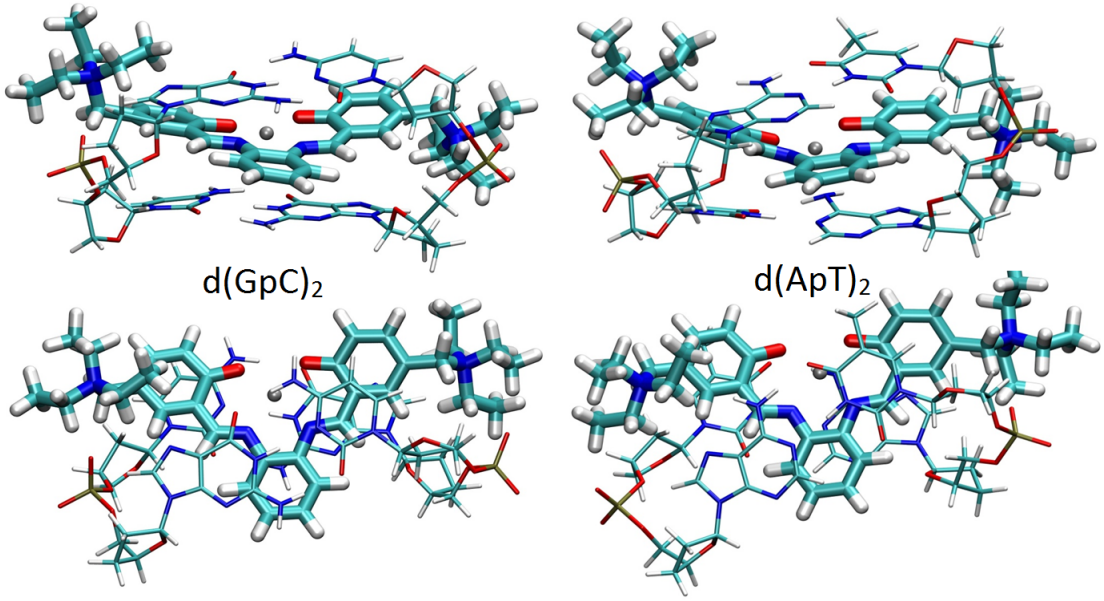

FIGURE 2.9: Optimized molecular structure of $ZnSalphen^{2+}$ intercalated in $d(GpC)_2$ (left) and $d(ApT)_2$ (right).

A straightforward method for the calculation of the transition energies relies on a fully QM strategy for both ZnSalphen^{2+} and the intercalation pocket. However, another approach involves the use of a QM/polarizable molecular mechanics (QM/Mmpol) [84, 85] method describing the DNA pocket using a polarizable classical force field, given by point charges and induced dipoles. The comparison between the fully QM and the QM/MMpol method will be useful in order to disentangle the different DNA pocket effects. The solvent can be taken into account at the PCM level using a QM/PCM or a QM/Mmpol/PCM [86] approach. The optimized ground state structures of ZnSalphen^{2+} intercalated in d(GpC)$_2$ and d(ApT)$_2$ are reported in Figure 2.9. In Figure 2.10 the experimental absorption spectrum is compared to the calculated ones in the case of ZnSalphen^{2+} intercalated in the d(GpC)$_2$ and d(ApT)$_2$ pockets. This comparison shows that the intercalation causes significant changes in the absorption spectrum. Note that since a fully QM study is performed for both ZnSalphen^{2+} and the DNA pocket our calculated spectra explicitly describe the effect of the QM intercalation pocket that is crucial in order to model the spectrum at wavelengths larger than 300 nm that are characterized by strong mixing between ZnSalphen^{2+} and the DNA-pocket. There is a general good qualitative agreement between the calculated and the experimental spectra. In fact, intercalation would cause a decrease of the intensity of the dye bands and their shift to higher wavelength that is fully reproduced in the calculated spectra. The inclusion of the solvent using a continuum PCM causes only a slight modification of the spectra.

The most relevant transitions in the spectra were analyzed using a NTO analysis (shown in Figure 2.11 for AT). The latter shows that all the transitions (even the ones at lower energies) are characterized by non-negligible charge transfer contributions between the base pairs and the DNA. The relevance of this mixing has been previously suggested [87], considering a different intercalated molecule. The comparison between the isodensity surfaces of the free and DNA-bound ZnSalphen^{2+} shows that the intercalation leads to the breaking of the symmetry of the virtual orbitals involved in the different transitions hows that the first two excitations

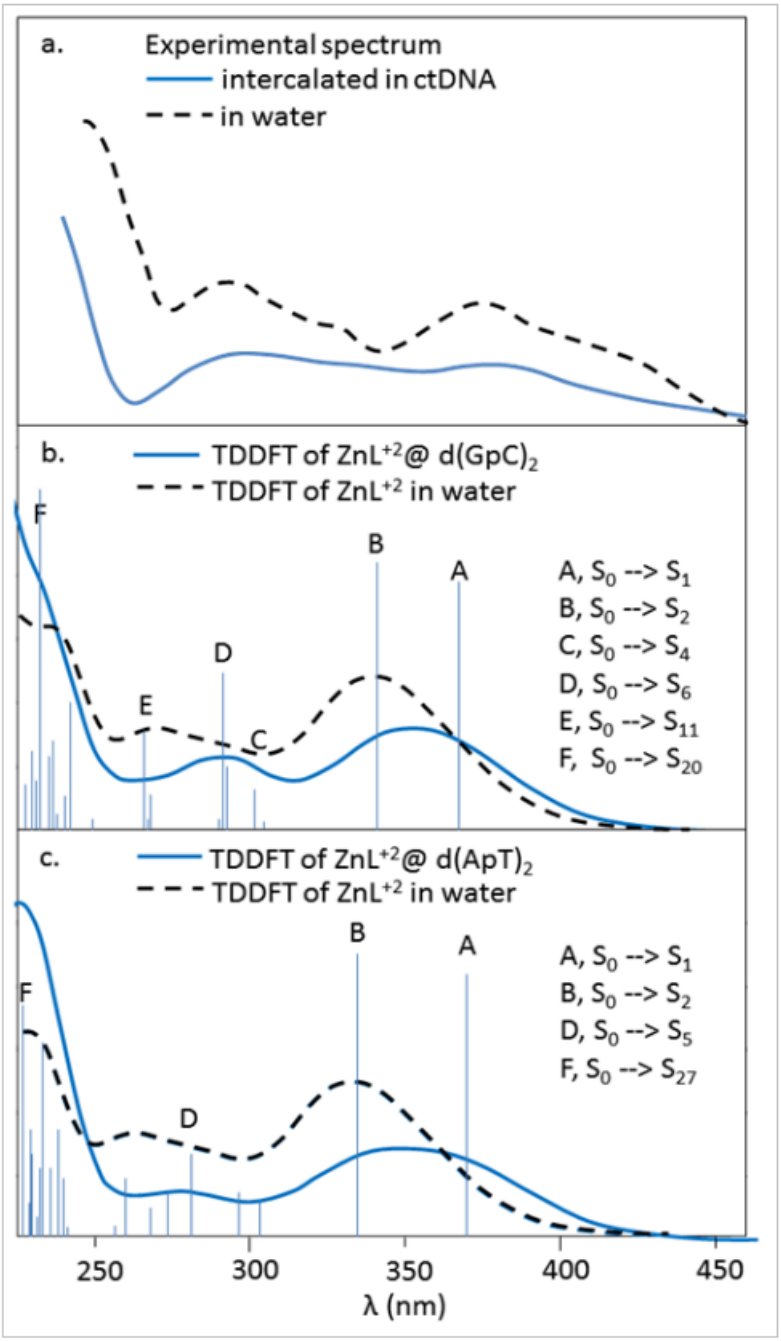

FIGURE 2.10: Comparison between the experimental absorption spectrum (a) of ZnSalphen^{2+} intercalated in ctDNA and those calculated in water (PCM) for ZnSalphen^{2+} intercalated in d(GpC)$_2$ (b) and d(ApT)$_2$ (c). The comparison with the corresponding spectra of free ZnSalphen^{2+} is also shown for each case (dashed line). For the experimental spectrum of ZnSalphen^{2+}/DNA complex (25 C) low ionic strength (1×10^{-3} M buffer) and polymer excess conditions were used to ensure quantitative dye reaction; in the shown spectrum, the DNA contribution to absorbance was subtracted.

are mainly located on the ligand, whereas higher energy transitions also involve increasing contributions from both the zinc atom and the adjacent bases.

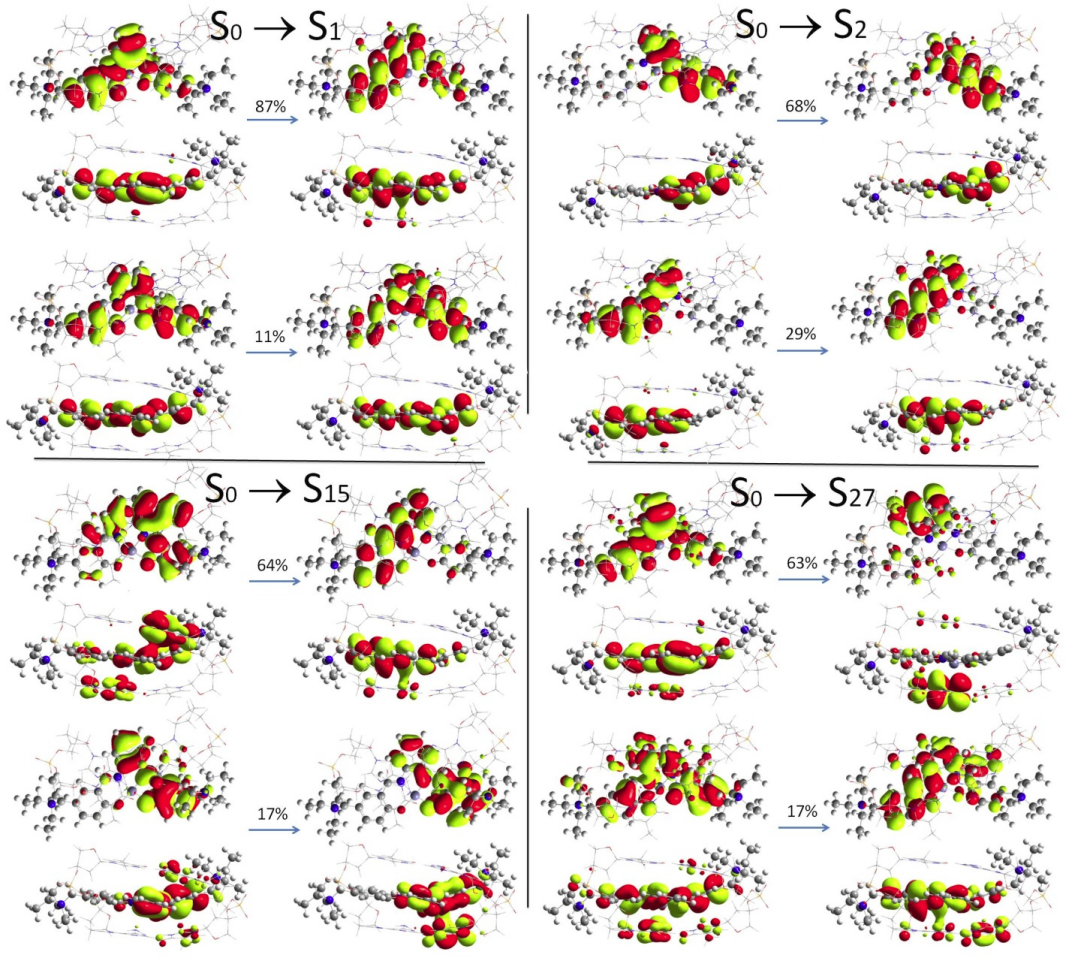

FIGURE 2.11: Dominant NTOs pairs for the most important transitions of ZnSalphen^{2+} bound to d(ApT)$_2$. The NTOs transition amplitudes are also shown.

2.4.1 Conclusions

The calculated spectra of ZnSalphen^{2+} intercalated in both d(ApT)$_2$ and d(GpC)$_2$, correctly reproduce the hypochromic effect experimentally observed for ZnSalphen^{2+} intercalated in natural DNA. In details, the infra-ligand band of the zinc(II) complex, which consists of two transitions to the first and second excited electronic

states, is red shifted of about 0.12 eV going from the free (in water) to the intercalated system and the calculation indicate 0.11 eV when the shift is averaged on the two pockets. The values of the experimental DNA-binding constants obtained at different temperatures provided the enthalpic and entropic contributions in the DNA binding and show that two interaction mechanism are involved: 1) intercalation and 2) external interaction, at high and low temperature, respectively, and that the two mechanisms coexist at room temperature in physiological conditions. Remarkably, such result is supported by the analysis of the structures of intercalation complexes, optimized by DFT. In fact, the latter are characterized by two main complementary binding interactions: 1) intercalation in the stacked AT-AT or GC-GC base pairs; 2) external electrostatic interaction between the negatively charged phosphate groups and the triethylammonium cationic groups of the Schiff base ligand.

2.5 Computational Details

2.5.1 Molecular Dynamics simulations

An alternating adenine-thymine sequence of 12 base pairs, [dodeca(dA-dT)]$_2$, was constructed in the double-helix B-DNA conformation by the NUCLEIC routine of the TINKER program package [88]. To create the intercalation pocket in the starting DNA structure, the torsion angles α - ζ and χ of the sugar phosphate backbone were opportunely modified [3]. The naphthalene moiety of the zinc(II) complex **3** was intercalated from the major groove side between the 6^{th} and 7^{th} base pairs using a home-made routine. The G-quadruplex models used are: human telomeric DNA (pdb id 1KF1) and the human c-myc promoter (pdb id 1XAV). We have performed four 50 ns MD simulations (150 ns in the case of c-myc/complex), two for the quadruplexes and two for the quadruplex/complex systems. The zinc(II) complex was positioned about 7-8 Å far over 3' G-quartet in order to simulate the recognition process.

MD simulations on the **3**/[dodeca(dA-dT)]$_2$ complex were performed by the GROMACS 4.5.3 software package [89, 90] using the Amber99 force field [91] with Parmbsc0 nucleic acid torsions [92, 93]. The starting geometry and the partial atomic charges of the metal complex were obtained by DFT calculations (see below), while other intramolecular parameters were generated with ACPYPE software [94]. A triclinic box of TIP3P water molecules were added around the oligonucleotides to a depth of 1.5 nm on each side of the solutes, to obtain a solution density of about 1.02 g/mL. Na^+ counterions were added to neutralize the negative charges of the phosphate groups, while other Na^+ and Cl^- ions were added to set the solution ionic strenght to about 0.15 M. Van der Waals parameters for zinc (σ = 0.195998 nm, ε = 0.0115897 kJ/mol), potassium (σ = 0.473602 nm, ε = 0.001372 kJ/mol) and chlorine (σ = 0.440104 nm, ε = 0.418400 kJ/mol), were taken from the Amber99 force fields implemented in GROMACS. Explicit solvent simulations were performed in the isothermal-isobaric NPT ensemble, at a temperature of 300 K, under control of a velocity rescaling thermostat [95]. The particle

mesh Ewald method [96] was used to describe long-range electrostatic interactions. The timestep for integration was 2 fs and all covalent bonds, including the four bonds between the metal ion and the tetracoordinate Schiff base ligand, constrained with the LINCS algorithm. There were two temperature coupling groups in this simulations, the first for the DNA and, if present, for the metal complex, the second for water and ions. Preliminary MD simulations showed that the structure of the isolated metal complex is maintained in solution. Preliminary energy minimizations were run for 5000 steps with the steepest descend algorithm. During the equilibration, the oligonucleotide and the metal complex/oligonucleotide system were harmonically restrained with a force constant of 1000 kJ mol^{-1} nm^{-2}, gradually relaxed into five consecutive steps of 100 ps each, to 500, 200, 100, 50 kJ mol^{-1} nm^{-2}.

2.5.2 DFT/MM calculations

The relaxed geometries of the complexes between DNAs and the metal complexes, were optimized by two-layer quantum mechanics/molecular mechanics (QM/MM) hybrid calculations, as implemented in the ONIOM method [97, 98], with the aim to perform a high-level calculation on the intercalation pocket and to take account of the constraining effects of the double-helical structure at lower levels of theory. The M06-2X [46] DFT functional and the dzvp basis set [99] were used in the higher QM layer, to suitably model the hydrogen bonding and $\pi - \pi$ stacking interactions between the sixth and seventh Watson-Crick base pairs. The Amber99 force field was used in the lower MM layer of the DFT/MM calculations. The highest layer of the model includes, in the case of duplex-DNA, the sixth and seventh base pairs and the cationic complexes **1-3**, with charge set to +2 and spin multiplicity 1 or 2, the latter in the presence of the copper ion. In the case of G-quadruplex DNA the model includes the four guanine bases of the 3' G-quartet and the cationic complex, with charge set to +2 and spin multiplicity 1. Default atomic partial charges were used for the DNA atoms, implicitly included in the force field

parameters. Vibration frequency calculations, within the harmonic approximation, where performed on the optimized geometries by using the same DFT/MM method. Solvent effects were evaluated by performing M06-2X/dzvp single point calculations on the high layer model extracted by the DFT/MM optimized geometry, with the implicit water solvent reproduced by the polarizable continuum model (PCM) [100, 101]. The Integral Equation Formalism version of the Polarizable Continuum Model (PCM) [100] has been used to describe the effects of the solvent both on the ground and the excited states. PCM cavities have been built as a series of interlocking spheres centered on atoms with the UFF radii. Standard enthalpy and Gibbs free energy values, at 298.15 K, of each energy minimum structure, both in vacuo and in solution, were calculated by adding the thermal correction obtained by vibration frequency analysis of the DFT/MM systems to the DFT energy calculated for the high layers both in vacuo and in water solution. The PCM energy data contain also non-electrostatic effects. The ground and excited state geometry optimizations, and the absorption and emission calculations were performed at (TD)DFT level of QM theory using the M05-2X functional [102] and the 6-311+G(d,p) basis set. This functional was chosen for its ability to describe both non-covalent interactions [103] and zinc complexes [104]. Open shell DFT/TDDFT calculations for triplet and doublet states were carried out using the unrestricted self-consistent field formalism. All the calculated absorption spectra have been produced by the convolution of the vertical transitions energies using Gaussian functions of fixed half-length-width of 0.24 eV. All the geometry optimizations were performed without imposing any constraint, whereas, in the case of the intercalated systems, the DNA pocket structures were kept frozen and described using the smaller 6-31G basis set analogously to previous works [105]. In order to evaluate the importance of a QM description of the intercalation pockets, constituted by the $d(ApT)_2$ and $d(GpC)_2$ dinucleoside-monophosphate duplexes, in the determination of absorption and fluorescence energies of $ZnSalphen^{2+}$, two models were compared. In the first model, indicated as QM/QM/PCM, a full QM description was used for both the complex and the DNA fragments constituting the pockets. Exactly as done for the geometry optimizations, the DNA fragments

were described with a smaller basis set (6-31G). The second model, indicated as QM/MMPol/PCM, [106] instead used a QM description only for ZnSalphen^{2+} whereas the pockets were described with fixed charges and induced dipoles. The fixed charges were obtained from a fit of the electrostatic potential of fragment, using the Merz and Kollman method [107] using the same functional and the same basis set used in the QM calculations. The induced dipoles were obtained in terms of isotropic polarizabilities placed on each MM atom. We adopted the Thole model, which avoids intramolecular overpolarization problems by using a smeared dipole-dipole interaction tensor [87]. Atomic isotropic polarizability values were taken from the fit of experimental molecular polarizabilities performed by van Duijnen and Swart using the linear version of Thole dipole-dipole tensor [108]. In both the QM/QM and QM/MMPol descriptions the effects of the rest of DNA and of the solvent were simulated using a PCM description. In the case of the QM/MMPol/PCM the two classical parts (MMPol and PCM) are allowed to mutually polarize. The vertical ionization potential (VIP) was evaluated as the difference between the energy of the optimized structure before ionization and that of the ionized system at the same geometry. In order to effectively model the solvent effect, it is necessary to use the non-equilibrium version of the PCM. In fact, the ionization process can be seen as a vertical process in which the fast change in the molecular charge density is coupled with a fast (mainly electronic) and a slow (mainly orientational) response of the solvent molecules, similarly to the cases of electronic excitation and electron transfer. All calculations were performed using the Gaussian 09 package [109].

Chapter 3

Empirical force fields are able to describe the relative energies of G-quadruplex conformations and their interactions with naphthalene diimide ligands

Contents

3.1 Introduction

The results reported in the present chapter have been obtained after a period of about nine months spent as visiting student at the Technical University of Braunschweig (Germany), working in the laboratory of Prof. Jörg Grunenberg, which has led to the publication of two papers [110, 111].

3.1.1 Glycosidic bond angle

Concerning the topology of G4-DNA arrangements, the base of each nucleotide is attached via a glycosidic bond from the N9 nitrogen of the purine to the C1' carbon of the deoxyribose sugar. The rotation about the glycosidic bond, the torsional χ-angle, defines two general conformational classes: the *anti* conformation with the base extended away from the sugar, and the *syn* conformation with the base essentially lying on the top of the sugar ring (Figure 3.1). There are evidences that an "*anti-anti*" stem orientation (AA) of the nucleobases is preferred against an "*syn-anti*" (SA) arrangement [112].

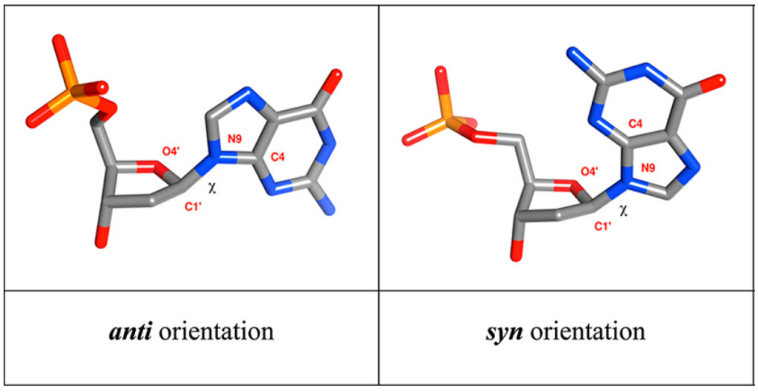

FIGURE 3.1: Two orientations of the glycosidic torsion angle χ: anti and syn. χ torsion angle is defined as O4'-C1'-N9-C4 and O4'-C1'-N1-C2 for purines and pyrimidines, respectively [112].

G-quadruplex structures are in some cases highly polymorphic due to the variability in the number of stacked G-quartets, the G-strand (or G-tract) orientations, the orientation patterns of glycosidic bonds of the guanines along a G-tract

and the various different loop types that can connect the G-strands. Parallel G-quadruplexes are structurally consistent and do not display significant structural polymorphisms. This occurs because all the guanine residues adopt anti conformations for the glycosidic bond orientation and all the four strands have the same orientation. In contrast, anti-parallel G-quadruplexes display both *anti* and *syn* guanines and at least one of the four strands must be oriented anti-parallel to the others [113]. G-quadruplex pholymorphism will be further discussed in section 4.1.1.

3.1.2 Force Fields

An empirical force field (FF) is a mathematical expression describing the dependence of the energy of a system on the coordinates of its particles [114]. It consists of an analytical form of the potential energy, and a set of parameters entering into this form. The parameters are typically obtained either from *ab initio* or semi-empirical quantum mechanical calculations or by fitting to experimental data such as neutron, X-ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy. Molecules are defined as a set of atoms that is held together by elastic forces and the force field replaces the true potential with a simplified model valid in the region being simulated. Ideally it should be simple enough to be evaluated quickly, but sufficiently detailed to reproduce the properties of interest of the system studied. There are many force fields available in the literature, designed to study different kinds of systems. However this is a typical expression for a FF:

$$U = \sum_{bonds} \frac{1}{2} k_b (d - d_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \frac{1}{2} k_\varphi [1 + cos(n\varphi - \varphi_0)] + \sum_{impropers} \frac{1}{2} k_\psi (\psi - \psi_0)^2 + \sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_i j} \right)^{12} - \left(\frac{\sigma_{ij}}{r_i j} \right)^{6} \right] + \sum_{i,j} \frac{q_i q_j}{\varepsilon_D r_{ij}} \tag{3.1}$$

Where the first four terms refer to intramolecular contributions to the total energy (bond stretching, angle bending, dihedral and improper torsions), and the last two terms describe the non-covalent Van der Waals interactions and Coulomb interactions. A harmonic potential is frequently used to represent the bond stretching term. k_b is the force constant of the bond, which is generally high, indicating that a large amount of energy is needed to stretch or compress a chemical bond. d_0 is the reference bond length, that is the value of the bond length when all other terms in the potential energy function are equal to zero. If other energy terms are different from zero, as it normally happens in large molecules, they will also influence the equilibrium bond length. The harmonic potential is usually a good approximation to describe the energy of a system only for small deviations from the reference bond length. The second term in equation 3.1

$$\sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{3.2}$$

represents the difference in the potential energy due to the deformation of angles. Force constants are typically smaller compared to bond stretching, this indicate that less energy is needed for a bond angle to deviate from its reference value. Similarly to the bond term, the accuracy of the force field can be increased by incorporating higher order contributions in the bond angle definition. The third term in equation 3.1

$$\sum_{dihedrals} \frac{1}{2} k_\varphi [1 + cos(n\varphi - \varphi_0)] \tag{3.3}$$

is called torsional term, and represents the potential energy of the molecule as a function of the rotation about each dihedral angle. The energies involved here are lower than for bond and angle. Torsional terms include atoms that are separated by 3 bonds. k_φ represents the barrier height, n the number of minima in the energy function (usually called multiplicity) and φ_0 is the phase factor, determining the position of the minima. The last bonded term:

$$\sum_{impropers} \frac{1}{2} k_\psi (\psi - \psi_0)^2 \tag{3.4}$$

is the improper torsional term. It is used in order to maintain the planarity of groups with flat geometry, and sometimes the correct chirality. As an example for planar groups, this term provides a penalty function for bending out-of-plane. Good examples of such groups are the peptide bond in proteins and the aromatic rings in side chains. The fifth and sixth term in equation 3.1

$$\sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_i j} \right)^{12} - \left(\frac{\sigma_{ij}}{r_i j} \right)^{6} \right] + \sum_{i,j} \frac{q_i q_j}{\varepsilon_D r_{ij}} \tag{3.5}$$

are the non-bonded terms, which are calculated pairwise for atoms separated by 3 or more bonds and for non bonded atoms. The first is a Lennard-Jones 12-6 potential and contains an attractive and a repulsive term. The attractive term is generated by the dispersion forces between instantaneous dipoles as explained by London [115]. The repulsive term is due to the observation that, below about 0.3 nm, atoms start to repel each other. Some electrostatic properties can be directly calculated using quantum chemical methods: for example, the electrostatic potential can be calculated directly from the electron density, which is defined by the wave function. One way to represent the electrostatic potential of a molecule is to place partial charges on atomic nuclei. In this representation, the electrostatic potential can be calculated as the last term of equation 3.1, where r_{ij} is the distance between nuclei i and j. In the last decades, a lot of efforts has been made into developing methods for the determination of atomic partial charges that were able to reproduce electrostatic properties of molecules, and in particular the electrostatic potential obtained from quantum mechanics calculations [116–118]. In recent years, intensive efforts have been directed to the development of computational models able to account also for induced polarization explicitly for the purpose of generating MD trajectories [119].

3.1.3 Compliance Constants

The force constants measure the intrinsic bond strength. However, the Cartesian force constants have to be transformed to a complete and non-redundant set of internal coordinates in order to gain any chemical implication. This transformation can be settled by several approaches [120]. Nevertheless, as it was pointed out, even force constants expressed in a set of internal coordinates, are not invariant to the choice of the individual coordinates [121, 122]. Thus, different choices of internal coordinates lead to different values of a specific bond strength. This numerical ambiguity of the internal force constants has always hampered their application [123]. To eliminate this ambiguity, it was suggested to use an inverted matrix of force constants, which was introduced as early as 1947 by Taylor and Pitzer [124]. Decius in 1962 proved that the inverse matrix of force constants (called compliance matrix) is indeed invariant to the internal coordinates using the complete set of experimentally determined force constants in molecules such as NO_2 and $GeCl_4$. Together with the possibility of modern computer systems and programs to efficiently calculate the complete matrix of energy second derivatives for large molecular systems, using correlated wave functions, the use of the inverted Hessian matrix has recently made the unequivocal calculation of bond strengths in polyatomic molecules a straightforward task [35, 125].

3.1.4 Naphthalene Diimide ligands

One promising novel anti-cancer strategy is to use small molecules able to bind and stabilize G-quadruplex structures by means of telomerase inhibition. Consequently, a huge number of ligands which are able to bind selectively to the G-quadruplex motif were studied in the recent past [126–128]. In this context, piperazine substituted naphthalene diimide (ND) ligands, pioneered by the Neidle group, attracted special attention. These molecules are characterized by a common naphthalene diimide aromatic core, which is able to stack effectively to the G4 3' quartet, while the piperazine substituents strengthen the recognition

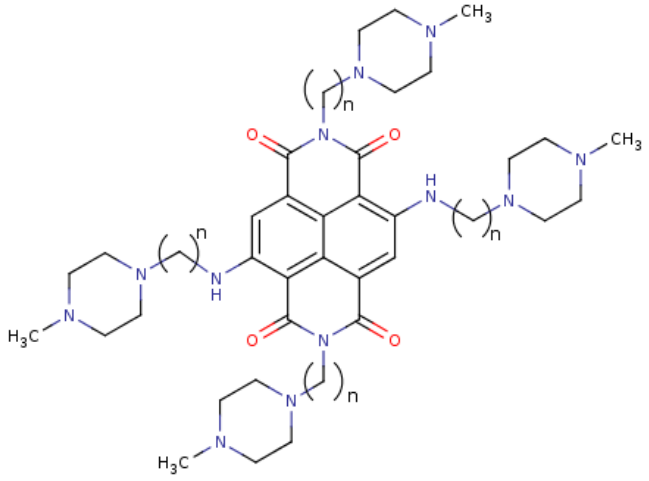

FIGURE 3.2: Naphthalene diimide ligands **1-4**, where n = 2-5 is the number of CH_2 groups in the chain.

process via hydrogen bonds. Under physiological pH-conditions, the piperazine nitrogens are protonated, leading to four positively charged ND arms. Due to this strong, non-covalent interactions, experimental [129–131] and computational [132] biophysical studies could demonstrate that such ligands are effective G4-binders that inhibit the growth of several cancer cell lines. In this context, the protonation state of a ligand plays a key role in drug design and in molecular recognition [133, 134]. Adding a proton to a neutral ligand leads to different binding modes, enhancing the hydrogen bonding aptitude with the negatively charged phosphate groups. It is clear, that in the case of ND/G4 complexes, the strength of the hydrogen bonds, and by this the stability of the telomerase inhibition, depends on the pH-conditions as well. In order to optimize the process of recognition, it is therefore of crucial interest to understand and control ligand protonation state. A computational study is reported in section 3.3, concerning the binding of a series of naphthalene diimide (ND) ligands **1**–**4** (Figure 3.2) with a well defined telomeric quadruplex model. The methyl-piperazione moiety can be considered as a diprotic base. For 1,4- dimethylpiperazione the reported values pK_{a1} and pK_{a2} at 298 K are respectively 3.81 and 8.38 [135]. Two limiting protonation states were therefore designed (see Figure 3.3): 1) The internal nitrogen atoms of the four methyl-piperazine rings are protonated (termed *intra*), and 2) the terminal piperazine nitrogen atoms (termed *ter*) are protonated. In both cases the total

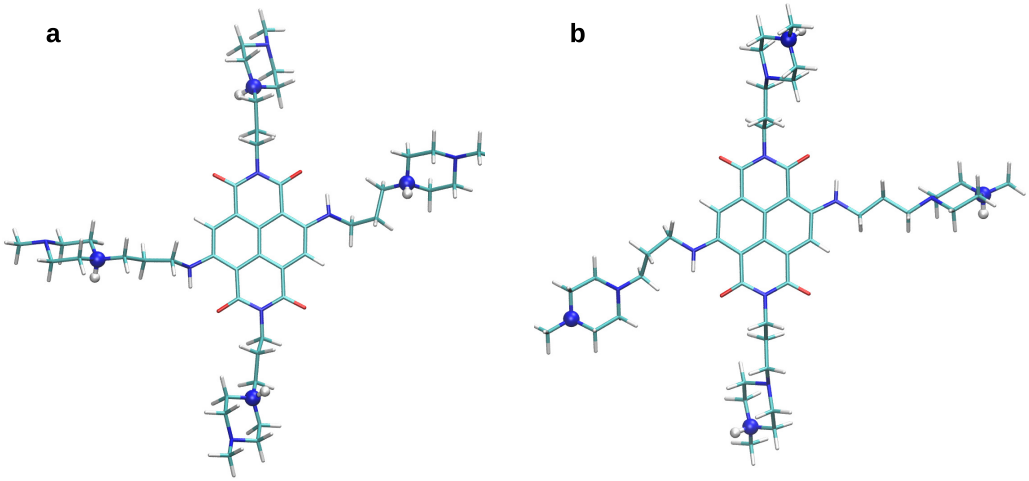

FIGURE 3.3: The proposed protonation states (a) *intra* and (b) *ter*.

charge of the protonated molecules is +4. The principal interactions of the ND ligands with G-quadruplex DNA are shown in Figure 3.4.

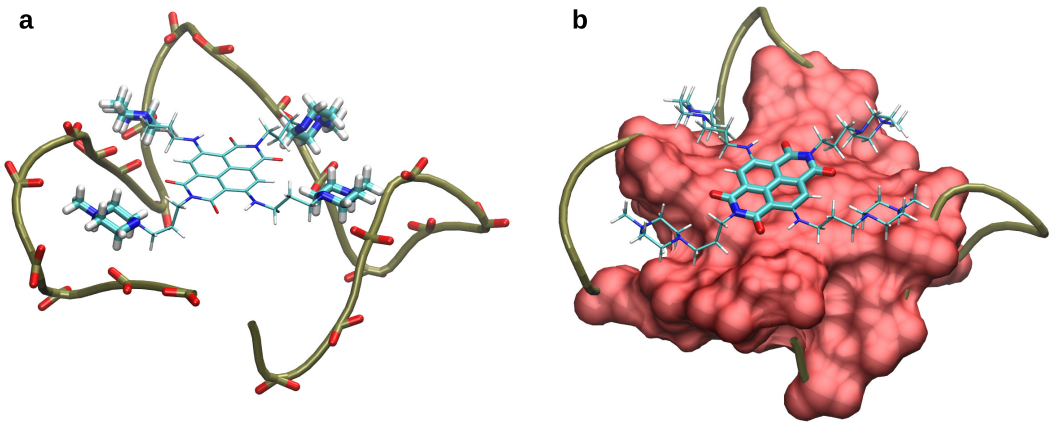

FIGURE 3.4: The most important interactions for the molecular recognition process. (a) Electrostatic (phosphate oxygen atoms in red) and hydrogen bond interactions. (b) $\pi - \pi$ stacking interactions between the naphthalene diimide core and the telomeric quadruplex.

These well-known G4 binders will be used as model systems and the information obtained from the following calculation will be used to guide a rational design of new metal complex with an improved selectivity towards G-quadruplex structures.

3.2 The relative energies of G-quadruplex stems

A recent publication by Sponer and Grimme [112] raises doubts about the reliability of MD simulations of non-canonical DNA structures. Their conclusion, based on the observed reversed stem stability comparing force field and DFT results, goes well beyond the known weak points of force fields (for example an over-stabilization of base stacking), stating that empirical force fields are not able to describe the G-quadruplex energy surface because of their pairwise-additive philosophy. In simple MM calculations, the SA-abab stem (in which the strands are in alternating direction) was described as the lowest minimum while approximate density functional theory puts the AA stem (in which the four strand point in the same direction) 5 kcal/mol below the SA-abab stem, in line with the experimental evidence. Nevertheless, the robustness of this conclusion was not tested against different force field parameters or implementations [136]. In this regard, several force fields (e.g. AMBER, OPLS-AA, MMFF) were compared with DFT results in order to review the general capabilities of empirical force fields in describing the conformational space of G-quadruplex structures. Since the description of biomolecular systems in condensed phase relies on an authoritative reproduction of the conformer space *in vacuo*, in a first step, nevertheless the simulations were conducted in gas phase. Solvent effects were included in a second step. Further, because different rotamers are considered (i.e. species consisting of the same number of atoms and with the same bond connectivity) it is possible to directly compare the different topologies by means of their computed steric force field energy directly. The starting conformers of the present study, each containing tetrads of hydrogen-bonded guanine bases and different glycosidic torsions defining the nucleobase orientation relative to the sugar, were taken from the literature [112]. All the computations started with the quantum chemically geometry optimized *in vacuo*, called "QMopt" in the original paper. The superposition of the optimized quantum chemical and force field geometries shows only small dihedral and angle deviations. Comparing the energetics of quantum chemical and force field results, considering the same energy wells, is a prerequisite of any meaningful discussion

TABLE 3.1: Calculated relative energy (kJ/mol) of the AA and the SA-abab stems, respectively. A positive sign is connected with a destabilized SA-abab rotamer.

	Vacuum	**Implicit Solvent**
WB97XD	+123.4	+78.7
Hartree-Fock	+117.3	+68.3
OPLS-2005	+47.4	+24.1
MMFF	+89.1	+10.9
AMBER 94	-72.3	-69.6
AMBER bsc0	-70.9	-65.9
AMBER bsc0 (Qeq)	+127.1	+106.1

in the following. Table 3.1 compiles the energetic results of the optimizations: 1) In line with the results by Sponer and Grimme, the AMBER force field (both, the AMBER94 and the AMBER-bsc0 implementations) sees the all-*anti* conformer AA higher in energy in comparison with the SA-abab (*syn-anti*) conformer. 2) Both, the OPLS-AA and the MMFF force field reproduce the experimental (and DFT) trend: a higher stability for the AA stem in comparison with the SA-abab stem. The overall picture is the same after the inclusion of implicit solvent effects (using the Poisson-Boltzmann model, PB). It is interesting to note that all three force field implementations (AMBER, OPLS-AA and MMFF) are based on pair-wise additive interactions. Nevertheless, different methodologies were used during their fitting process, respectively. The OPLS force field, for example, was trained using experimental properties of liquids, such as density and heat of vaporization, while the MMFF force field parameters were fitted against high level *ab initio* gas phase data. Remarkably, the molecular training set for all three evaluated force fields (OPLS; AMBER and MMFF) did not include any of the non-canonical DNA structures. One obvious reason for the wrong description of the conformer space could be the bogus account of dispersion effects. Remarkably, both simple Hartree-Fock/6-31G(d) theory (without dispersion) and DFT (including dispersion, wb97xd/6-31G(d)) reproduce the correct experimental trend (see Table 3.1). Dispersion is obviously not the reason for the energy differences. A second reason for the erratic behavior of the AMBER force field could be the deficient description of the relevant hydrogen bonds. In a second step, the relaxed force constants (inverse compliance constants) of the most important non-covalent interactions

TABLE 3.2: Comparison of non-covalent interaction in the AA and the SA-abab conformers of the quartet G-DNA computed at the DFT (WB97XD) and Force Field (AMBER) level of theory in N/cm.

Vacuo	**NH—N**	**NH—O**	**K^+—O**	**mean**
Guanine dimer G_2	0.11	0.24	-	**0.17**
AA (DFT)	0.34	0.34	0.19	**0.29**
SA-abab (DFT)	0.31	0.36	0.19	**0.28**
AA (AMBER)	0.41	0.29	0.19	**0.29**
SA-abab (AMBER)	0.30	0.31	0.20	**0.27**

were analyzed. In fact, the description in terms of compliance constants allows a direct comparison of individual covalent and non-covalent interactions for polyvalent supramolecular complexes. In the last years the accumulation of studies, which proof the utility of a deliberate application of compliance constants for strong and soft interactions is nevertheless significant [137, 138]. Here, the inverse of the computed compliance constants (relaxed force constants in N/cm) is presented in order to facilitate the comparison with older literature values. Each G-DNA stem contains two quartets comprising four Hoogsteen hydrogen bonds of the NH—N and the NH—O type, as well as four dipole-ion interactions of the guanine—K^+ type, respectively. The relaxed force constants computed by the wb97xd/6-31G(d) DFT method, as a measure of the elasticity for the AA and the SA-abab stem, are also listed in Table 3.2. The analysis of the individual Hoogsteen hydrogen bonds confirms the recently reported H-bond cooperativity in G-tetrads [139]. While for an isolated guanine dimer (G_2) the calculated relaxed force constants of the hydrogen bonds point to quite soft interactions (NH—O: 0.25 N/cm; NH—N: 0.10 N/cm), the G-tetrads are characterized by strong NH—O (0.35 N/cm) and NH—N hydrogen bonds (0.31-0.34 N/cm). Comparing the sum (0.29 versus 0.28 N/cm) of all inter-residue hydrogen bonds as well as the eight contacts between the guanine oxygen and the channel cation K^+, kept in place by eight soft interactions (0.18-0.20 N/cm), there seems to be no difference between the AA and SA stems (in this case, the SA-abab rotamer) at all. This is true for both the DFT and the AMBER description. Thus, if dispersion and hydrogen bonding, are not responsible, what could be the reason for the erratic behavior of the AMBER force field? In order to describe the total energy, an

accurate representation of the charge distribution throughout the entire molecule is essential for all atomic resolution modeling techniques and especially for highly charged systems such as nucleic acids. Since the guanine quadruplex structures are non-canonical DNA structures, which were not included in the original force field training set, the standard assignment of the atomic charges might therefore be the source of small individual electrostatic errors, even if the atoms in question are not part of the hydrogen bond network. The accumulation of these small errors for the nearly 20000 subtle, mutual non-bonded interactions might nevertheless lead to an erratic description of the total steric energy in force fields. Therefore, a way to verify this hypothesis is the calculation of the adapted charges, which a) depend upon the environment and b) are allowed to fit during the geometry optimization. The last row of Table 3.1 also includes the result of a modified AMBER force field using quantum adapted charges applying Reppe's and Goddard's charge equilibrium (QEq) formalism [140]. The resulting adapted charges are based on (1) experimental atomic ionization potentials, (2) electron affinities, and (3) atomic radii. Using these experimental parameters, an atomic chemical potential is constructed leading to charges which are in excellent agreement with experimental dipole moments. The AMBER-QEq force field indeed reproduced the experimental and DFT trend: the AA rotamer is now calculated to be the most stable conformer using the AMBER force field. In comparison with the SA-abab stem, the stabilization is quite pronounced (127.1 kJ/mol) in vacuo and slightly shielded (106.1 kJ/mol) applying the implicit solvent correction (see Table 3.1). Therefore, the AMBER-QEq trend is in line with the experimental results: a higher stability for the AA-stem. A comparison of the relevant atomic charge before and after the application of the QEq formalism reveals that the following atom types are involved, in decreasing order: the phosphorus atom, whose charge decreases by 0.6 units, and the N*, CA, and O2 atom types, whose charge values change by 0.4 units (see Table 3.3). However, it is important to refrain from any overinterpretation of the atomic charges, the more so as any partial adaption of the force field parameters without adaption of all other nonbonded interaction or torsional parameters impedes their interpretation as physical observables.

TABLE 3.3: Relevant AMBER bsc0 charges (a.u.) before and after the Qeq formalism was applied.

Atom type	**bsc0**	**bsc0 Qeq**	**OPLS**
P	1.2	0.6	1.8
O2	-0.8	-0.4	-1.0
N*	-0.1	-0.4	-0.5
NB	-0.6	-0.4	-0.5
CA	0.7	0.3	0.5
N2	-0.9	-0.6	-0.8
NC	-0.7	-0.3	-0.5

3.2.1 Conclusions

In conclusion, provided that the empirical potential function is chosen correctly and flexible enough to adapt to the very special electrostatics, force fields are indeed able to reproduce, and hence predict, the correct relative energetic ordering of different arrangements of guanine quadruplex (G-DNA) stems. Since the simulation of molecular recognition processes, where DNA G-quadruplexes function as a receptor for small organic or inorganic molecules, relies on a reliable and fast mapping of the system's conformational space, a combination of force field methods, controlled by DFT checkpoints could lead to a robust protocol for the development of effective and potent G-DNA binders.

3.2.2 Computational Details

All quantum chemical calculations were performed with the Gaussian 09 suite of programs. The structures were fully optimized using a long-range corrected (LC) hybrid density functional with which employs 100% Hartree–Fock (HF) exchange for long-range electron–electron interactions, while include dispersion corrections are included via empirical atom–atom terms introduced by Head-Gordon coworkers the (wb97xd). A polarized medium sized all-electron Gaussian atomic orbitals (AO) basis set (6-311G(d,p)) was used in the optimizations and compliance constants calculations. Continuum solvation effects were represented by the Polarizable Continuum Model (PCM) assuming the bulk dielectric constant of 78.35

for water. The force field calculations were carried out with Macromodel software (AMBER94, MMFF and OPLS-2005) and with the Gaussian 09 set of programs (AMBER bsco, AMBER bsco/Qeq). A generalized-Born (GB) algorithm was applied for including solvent effects.

3.3 The interaction of Naphthalene Diimide (ND) ligands with G-quadruplex DNA

The relative enthalpic interaction strengths of the ND guests (ΔE) were calculated as the difference between the global minimum energy of the complex (E_{G4-ND}), and the sum of the global minima of the G4 receptor (E_{G4}) and of the individual isolated (liquid phase) ND (**1-4**) guests (E_{ND1}, E_{ND2}, E_{ND3}, E_{ND4}). Entropic contributions were ignored in this part. In Table 3.4, the computed ΔE values for the binding of ligands **1-4** are compared with experimental melting data published by Neidle et al. [130]: a higher ΔT_m value is connected with a higher stability of the complex.

Besides the general reproduction of the trend in terms of binding strength, the following conclusions can be done: 1) Terminal protonation of the piperazine substituents leads to a better recognition (stronger bond) in comparison with the inner protonation. 2) Changing the protonation state dramatically changes the stability of the ligand-G4 complex, destabilizing the inner protonation. 3) Only the ΔE values for the terminal protonation state reproduce the experimental stability trend: $2 > 3 > 4 > 1$. In order to test the robustness of the results in terms of the relative proton affinity of ligands **1-4**, the energy of the two protonation states was calculated for the most effective binder, ND **2**, with electronic structure methods (DFT; B3LYP-D3 functional and a standard 6-31g(d) basis set), starting a DFT minimization from the OPLS-2005 global minimum conformer geometry. The DFT gas phase calculation seems to reproduce the force field simulation of

TABLE 3.4: Comparison of the computed trend for the ND ligands **1-4** in terms of the calculated (OPLS-2005) enthalpic binding energy (ΔE in kJ/mol) with experimental melting data (in °C). Both protonation states, *intra* and *ter*, were considered (see Figure 3.3).

	1	**2**	**3**	**4**
$\Delta \mathbf{T}_m$	19.0	28.3	24.7	23.8
Δ**E *intra***	-274.7	-252.6	-314.9	-314.0
Δ**E *ter***	-294.5	-420.1	-405.8	-388.2

the liquid phase. In fact, the terminal protonation of the guest ND **2** is preferred by -146.7 kJ/mol.

In order to compare the binding mode in the solid state and in solution, the crystal structure 3SC8 was compared with the global minimum found by the conformational Monte Carlo scan in the liquid phase (*ter* protonation) for guest ND **2** (Figure 3.5). Besides the overall analogy, there are some differences in the stacking position of the ligand, in particular the position of the ND guest is closer to the 3' quartet and shifted horizontally in the solid state. It was speculated that this difference is due to crystal packing effects. In the solid state, the ND **2** ligand interacts with two quadruplex receptors at the same time, shifting the minimum position of the diimide core. The same seems to be true for the second protonation state (*intra* in Figure 3.6).

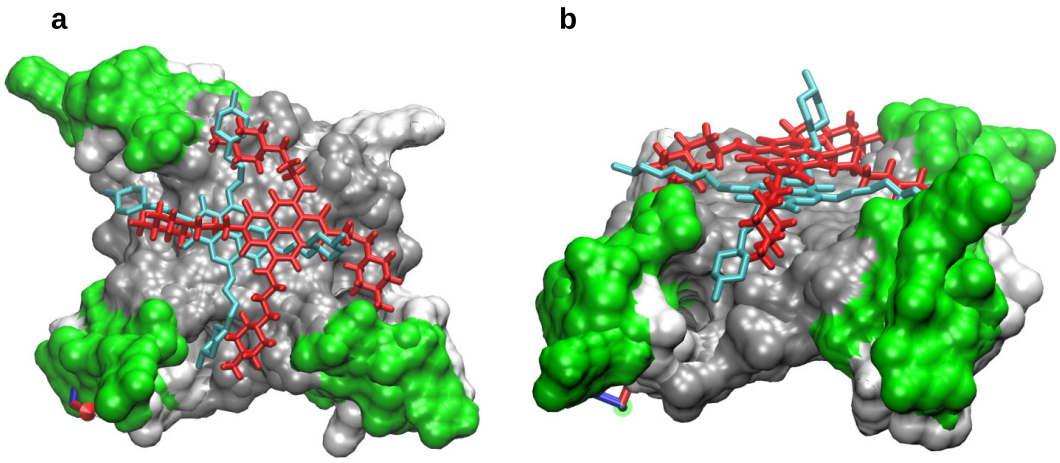

FIGURE 3.5: Superposition of the crystal structure (light blue) with the global minimum found for ligand **2** *ter* (red).

In order to analyze the individual contributions of the individual non-covalent interactions to the overall recognition process, the strength of the most important non-covalent interactions between the receptor and ligand **2** was computed by means of their relaxed force constants (inverse compliance constants). The values obtained are reported in Table 3.5.

In the case of the *ter* protonation state, the global minimum obtained from the conformational search shows two direct and two water mediated hydrogen bonds

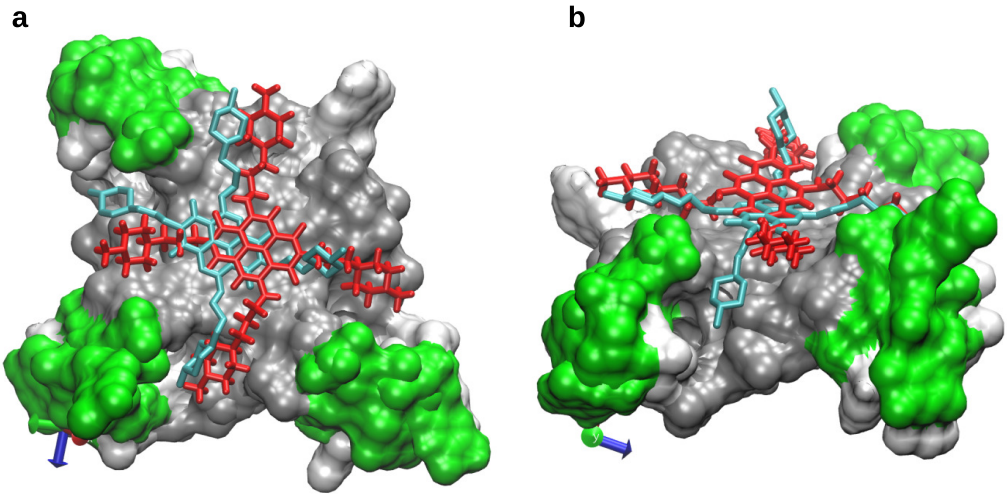

FIGURE 3.6: Superposition of the crystal structure (light blue) with the global minimum found for ligand **2** *intra* (red).

with the phosphate groups, while the *intra* protonation state is characterized by water mediated hydrogen bonds exclusively. Figure 3.7 depicts the two kinds of hydrogen bonds observed in the global minimum of the **2** *ter* complex. Looking at Table 3.5, in terms of their stiffness, the direct piperazine/phosphate hydrogen bonds (N-H—O-phosphate) seem to be more or less of the same quality then the mediated bonds (N-H—O-water). The calculated relaxed force constants for this kind of non-covalent interactions ranges between approximately 0.3 and 0.6 N/cm. In contrast, the calculated softness of the hydrogen bonds between the mediating water and accepting phosphate oxygen is significantly lower: the values between 0.1 and 0.2 N/cm point to very short hydrogen bond lifetimes. The results

TABLE 3.5: Relaxed force constants (OPLS-2005 force field in N/cm) of the hydrogen bonds between ligands and the telomeric receptor.

ND2 intra complex		**ND2 ter complex**	
NH—O (Water)	0.50	NH—O (Phosphate)	0.29
NH—O (Water)	0.38	NH—O (Phosphate)	0.45
NH—O (Water)	0.59	NH—O (Water)	0.42
NH—O (Water)	0.34	NH—O (Water)	0.53
Water—O (Phosphate)	0.09	Water—O (Phosphate)	0.19
Water—O (Phosphate)	0.03	Water—O (Phosphate)	0.14
Water—O (Phosphate)	0.12		
Water—O (Phosphate)	0.15		

obtained point out that the static view of hydrogen bonding alone is not enough to explain the energy difference between the two protonation states.

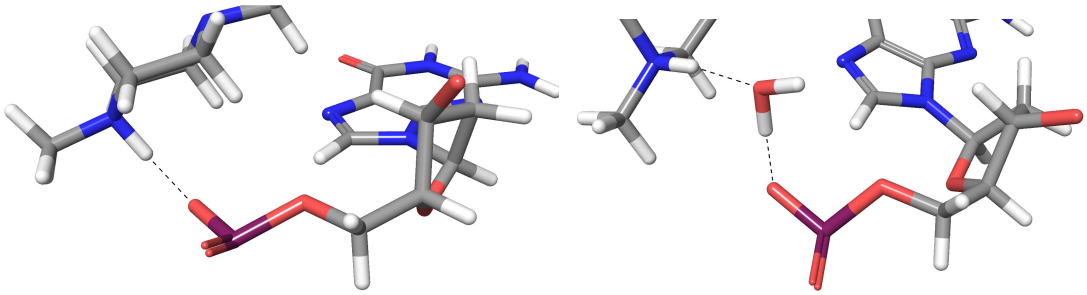

FIGURE 3.7: Direct (left) and water mediated (right) hydrogen bonds observed for the computed (OPLS-2005) lowest energy conformer of the complex between the telomeric G-quadruplex receptor and the terminal protonated (*ter*) naphthalene diimide ligand ND **2**.

In its *ter* protonation state, ND **2** guest is characterized by short range interactions, showing two direct hydrogen bonds. The main difference is visible in the long range interactions: the average positions of the terminal rings are closer to phosphate groups. The methyl-piperazine rings are locked deep inside the quadruplex grooves, leading to a stronger electrostatic interaction with the receptor. The

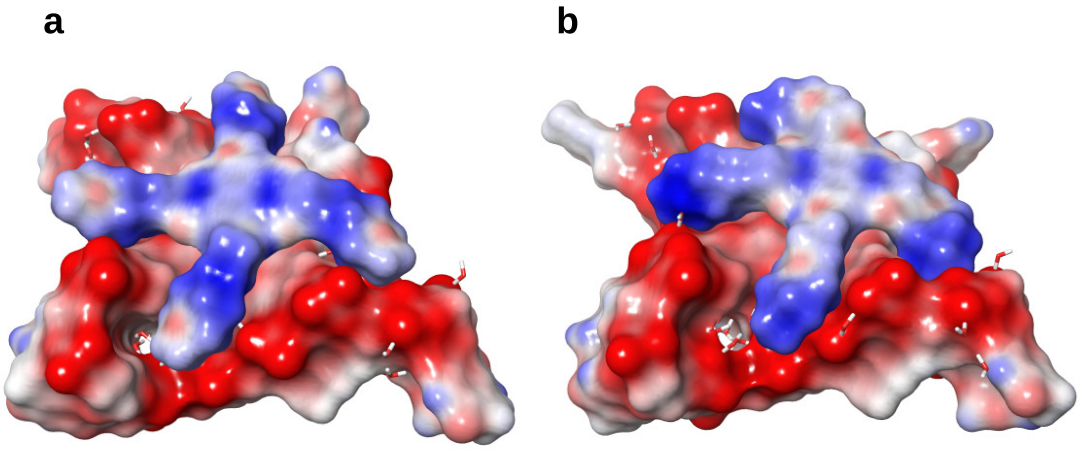

FIGURE 3.8: The electrostatic potential for the *intra* (a) and *ter* (b) isomers of the G4 complex with ND **2**.

different position of the charged arms is clearly shown by the electrostatic potential surfaces for the global minimum (Figure 3.8), where the negatively charged phosphates is plotted in red and the positively charged nitrogens in blue, respectively. Molecular Dynamics simulations were performed in order to further test

the hypothesis concerning the differences of the hydrogen bond lifetimes in both protonation states and the results are summarized in Figure 3.9. As it can be seen, in strong contrast to the internal protonation, the terminal protonation state is stabilized by a larger number of direct NH—OP hydrogen bonds during the simulation time. The 50 ns trajectory of the terminal protonation state is characterized by at least one strong hydrogen bond, while the second and third hydrogen bonds are auxiliary by nature. Interestingly, a single time step, characterized by four simultaneous hydrogen bonds, is never observed during the whole 50 ns seconds of MD simulation.

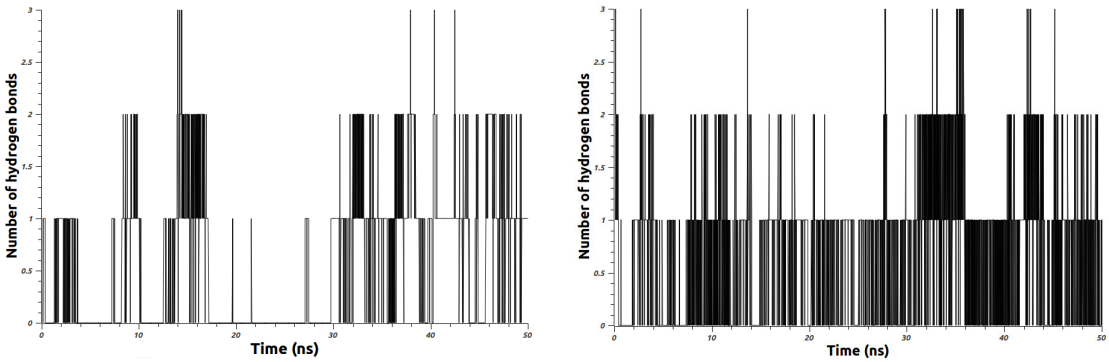

FIGURE 3.9: Here the number of NH—OP hydrogen bonds versus the simulation time was plotted for the *intra* (left) and the *ter* (right) isomers of ligand **2**.

3.3.1 Conclusions

Due to the combination of experimental data, thorough Monte Carlo conformational scans, MD simulations, DFT calculations and simulations of the mechanical binding strengths, 1) the different protonation states of tetrasubstituted naphthalene diimide (ND) guests in complex with G-quadruplex (G4) DNA receptors, and 2) the importance of water mediated vs. direct hydrogen bonding have been analyzed. Based on the reproduction of the relative stabilities of a series of ND ligands published by the Neidle group, the following conclusions have been reached: 1) the terminal protonation of the naphthalene diimide is the dominant state, leading to persistent, direct NH—OP bonds; 2) the calculation of the relaxed force constants in combination with MD simulations point out that the direct bonding is by far stronger and longer lasting than their water mediated counterparts. In this case the static picture of the X-ray structure is misleading; 3) in the liquid phase, besides hydrogen bonding, the longer range electrostatic interactions are maximized in the terminal protonation, leading to a perfect fit of the methyl-piperazine rings which are locked deep inside the quadruplex grooves in this case.

3.3.2 Computational Details

Two pdb entries of telomeric G-quadruplexes targeted by naphthalene diimide ligands were used as starting points for the construction of the models. The two structures, published by S. Neidle, G. N. Parkinson et al., were downloaded from the protein data bank, hydrogen atoms were added and basic amines were protonated according to the definition of the *intra* and *ter* protonation states, by hand. The crystallographic water molecules and the two internal potassium ions were retained in the liquid state simulation. The OPLS-2005 force field [141, 142] and the generalized Born/solvent accessible surface (GB/SA) continuum solvation model [143] were used applying the Schrödinger software package [144]. The conformational space was scanned applying a Monte Carlo Multiple Minimum (MCMM) algorithm as implemented in MacroModel [145], while the Polak-Ribiere conjugate

gradient (PRCG) algorithm was used for the minimizations. After a preliminary conformational Monte Carlo search of 90,000 steps, the resulting liquid phase global minimum of the G4 receptor together with the computed OPLS-2005 energy (E_{G4}) represented the model for the subsequent ND binding study. Each telomeric ND (**1-4**) complex was generated using the following protocol: 1) the individual ND guests were posed 8–9 Å away from the receptor, keeping the diimide core parallel to the 3' quartet of the telomere. 2) 20,000 minimization steps (steepest descent) were conducted. 3) Finally, 120,000 Monte Carlo steps were performed in order to scan the low energy region. In this final step, the internal potassium ions were constrained with a force constant of 100 kJ/mol. All other degrees of freedom were included as search parameters. The Gaussian09 software [146] was used for quantum mechanical and AMBER molecular mechanics (MM) calculations. Geometry optimizations were conducted with the B3LYP DFT functional and the 6-31g(d) basis set. The Amber99 force field was used [91]. The COMPLIANCE software and its graphical interface [147, 148] were used to evaluate the strength of the most important hydrogen bonds between ligands and quadruplex receptor. Molecular Dynamics (MD) simulations on both isomers of ligand **2** were performed by the GROMACS 4.5.3 software package as described in section 2.5.1. 15 K^+ counterions were added to neutralize the negative charges of the phosphate groups, while other 22 K^+ and Cl^- ions were added to set the solution ionic strength to about 0.15 M. The two internal potassium ions were kept from the Monte Carlo simulations. The hydrogen bonds networks were analyzed using the GROMACS tool g_hbond.

Chapter 4

Calculation of the Circular Dichroism of G-quadruplex DNA

Contents

4.1 Introduction

This final chapter is the result of the investigations carried out at the University of Nancy (France) during a short period (1.5 months) spent in the laboratory of Prof. Antonio Monari, in the ambit of a "short term scientific mission" organized by the COST action 1201: Biomimetic Radical Chemistry.

4.1.1 G-quadruplex isomorphism

G-quadruplexes are highly polymorphic structures. As an example, quadruplexes formed by human telomeres can adopt a variety of different folding topologies according to the relative orientations of the four strands held together by G-quartets (Figure 4.1). Rigorous information about the exact conformation that human telomeric quadruplex forms under physiological conditions is important, not only for understanding its biological function but also for the development of selective G-quadruplex binders. G-quadruplex motifs depends on the enviromental conditions. Animal cells maintain high intracellular concentration of K^+ (150 mM) and a lower concentration of Na^+ (5-15 mM) [149].

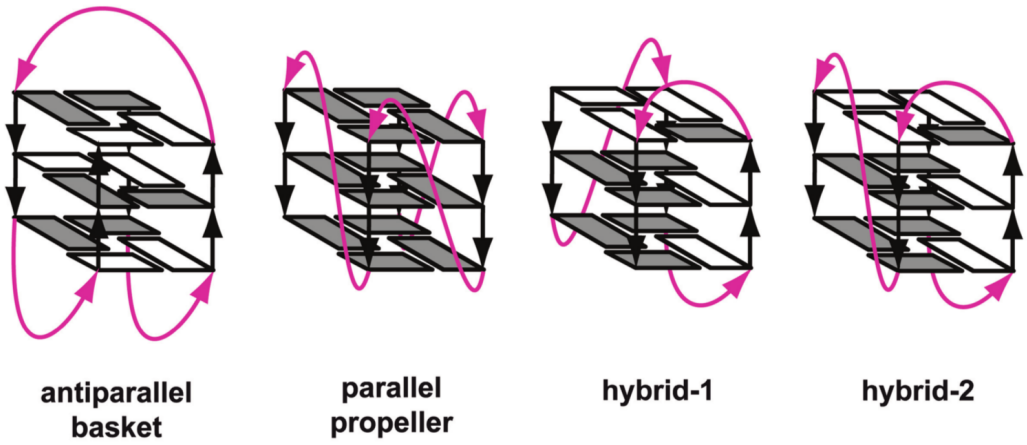

FIGURE 4.1: Schematic structures of the folding topologies adopted by telomeric DNA. Loops are colored in magenta; anti and syn guanines are colored grey and white, respectively [150].

The quadruplex formed in Na^+ solutions is well-characterized by NMR and adopts an antiparallel conformation (see Figure 4.1, antiparallel basket) [17]. During the last years, investigations on the structures formed in a physiological K^+ solution raised some controversy. In fact, the telomeric structure in K^+ solution has not been identified as well as the one formed in Na^+ solution and originally it was suggested to contain both parallel and antiparallel conformations [151]. In 2002, a novel propeller-like telomeric parallel structure, in the presence of K^+, was solved by X-ray crystallography [152]. In 2006, three research groups described a new hybrid-type structure in K^+ solution, with mixed parallel/antiparallel strands (see Figure 4.1, hybrid 1), which was proposed to be the favored conformation under physiological conditions [153–155]. Therefore, it was suggested that the parallel conformation, originally observed in the crystal structure, might have been selected under the particular condition concerning the crystallization procedure and thus might not be the biologically relevant conformation [156]. One year later a novel hybrid-type mixed parallel/antiparallel G-quadruplex, was reported [157]. This structure (see Figure 4.1, hybrid-2) and the previously reported hybrid-1 structure differ in their loop arrangements, strand orientations and capping structures. While the hybrid-type G-quadruplexes appear to be the major conformations in K^+ solution, human telomeric sequences are always in equilibrium between hybrid-1 and hybrid-2 structures, which is largely determined by the 3'-flanking sequence. Subsequently, it was shown that the intramolecular human telomeric quadruplex adopts the same parallel conformation observed in the crystal structure in K^+ solution under molecular crowding conditions. Because the intracellular environment is crowded with macromolecules up to 40% (w/w) in concentration [158], it was proposed that the parallel conformation could well represent the most relevant structure that human telomeric G-quadruplex can adopt under physiological conditions. Moreover, this structure was also found to have a dramatically enhanced stability [159]. Recently, it was shown that the conformational behaviour of the telomeric repeats in *X. Laevis egg* extract or in Ficoll is notably different from that observed in the presence of PEG. In fact, while the high-order parallel G-quadruplex folding topology is dominant in a PEG containing environment, the

equilibrium is dramatically shifted in favour of monomeric antiparallel or hybrid conformations in Ficoll or cellular extracts [150].

4.1.2 Frenkel Exciton Theory

The exciton model is a way to study the excited states of non-covalent stacked chromophores [160, 161]. Considering for simplicity a dimer composed of the monomers ϕ_A and ϕ_B, the global excited states may be constructed as the linear combination of individual monomer's excitations:

$$|\psi_{\pm}\rangle = c_{\pm}^{A}\left|\phi_A^{\dagger}\phi_B^{0}\right\rangle \pm c_{\pm}^{B}\left|\phi_A^{0}\phi_B^{\dagger}\right\rangle \tag{4.1}$$

where $\left|\phi_A^{\dagger}\phi_B^{0}\right\rangle$ represent an excitation on the monomer A. To obtain the global excitation energies and the eigenstates that define the excitation wavefunction, is it possible to write a secular equation and diagonalize an effective Hamiltonian matrix H whose elements will be:

$$H_{ii} = \varepsilon_i, H_{ij} = \left\langle \phi_i \middle| \hat{H} \middle| \phi_j \right\rangle \tag{4.2}$$

where the diagonal elements ε_i are the excitation energies on the monomer i, while the off-diagonal elements represent the coupling between the different excited states. The coupling between monomers breaks the degeneracy between the excited states giving rise to two transitions. Since the coupling is usually weak, the two transitions will normally be embedded in the vibronic band of the absorption spectra. However, in the case of chiral aggregates, such as duplex and G-quadruplex DNA structures, the two transitions may have different signs of the rotatory strength and therefore they can be easily differentiated in the spectrum. Although some more complicated schemes using the coupling between the monomers' excited state density matrices exist [162–164], the simplest formalism consists in considering the coupling between the monomeric transition dipoles as

well as their distances and orientations. In this case, the Hamiltonian matrix elements become:

$$H_{ij} = \frac{\mu_i \cdot \mu_j}{|R_{ij}|} - 3\frac{(\mu_i \cdot R_{ij})(\mu_j \cdot R_{ij})}{|R_{ij}|^5} \tag{4.3}$$

In Eq.4.3, μ_i is the transition dipole moment of the monomer i, R_{ij} is the distance vector relying the center of charges of the two chromophores i and j, and $|R_{ij}|$ is its module. After diagonalization of the H matrix, the eigenvalues ε are obtained, i.e., the excitation energies and the matrix U whose columns store different eigenvectors. The absorption oscillator strengths for an exciton state k will therefore be proportional to:

$$f_k = \sum_{i,j} (\mu_i \cdot \mu_j) U_{ik} U_{jk} \tag{4.4}$$

and the rotatory strength vector to:

$$r_k = \sum_{i,j} \varepsilon_k [R_{ij}(\mu_i \cdot \mu_j)] U_{ik} U_{jk} \tag{4.5}$$

For electronic circular dichroism in the case of homogeneous solution, the final spectroscopic observable intensity will be the average over the three components of the r_k vector.

This approach was successfully used to model the electronic circular dichroism spectra of two DNA model sequences, poly[d(GC)] and poly[d(AT)] [165].

4.2 Reproduction of the Circular Dichroism spectra

Circular dichroism (CD) refers to the differential absorption of the two circularly polarized components, one rotating counter-clockwise (left handed) and the other clockwise (right handed) of a plane polarized light. The idea behind CD is that the interaction between left and right handed photons and a chiral molecule will be different, thus only chiral molecules are able to produce a CD signal [166]. There are three common quadruplex topologies based on the directionality of neightbouring strands, namely parallel, antiparallel and hybrid (two possible types, called 1 and 2), each having a characteristic CD spectrum (see Figure 4.2).

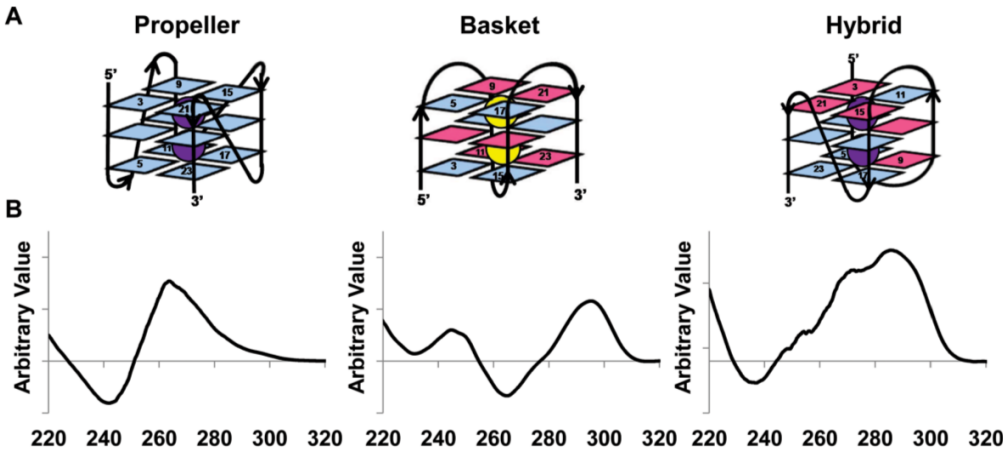

FIGURE 4.2: Folding topology (A) and experimental CD spectra (B) of the three quadruplex conformations. Propeller (parallel) on the left, Basket (antiparallel) center, hybryd on the right. *anti* and *syn* glycosidic angle are shown in cyan and pink, respectively [167].

Parallel quadruplex is a right-handed helix with four-fould symmetry and with all guanines having a *anti* conformation of the glycosidic bonds. Parallel structure is characterized by a negative peak near 240 nm and a positive peak near 265 nm. The structure of an antiparallel quadruplex is also a right-handed helix but guanines have alternating *syn* and *anti* glycosidic bond conformations (for a discussion about DNA glycosidic angle χ see section 3.1.1), antiparallel CD spectrum displays a negative peak near 260 nm and a positive peak 295 nm. In K^+

solution the CD spectrum of the telomeric sequence $G_3(T_2AG_3)_3$ shows a small negative peak near 240 nm and a positive peak near 290 nm with a shoulder at 270 nm, it is known that it adopts in these condition an hybrid conformation. It was hypothesized that the position of the typical bands in G-quadruplex spectra are related to the stacking orientation of G-quartets. Specifically, it was proposed

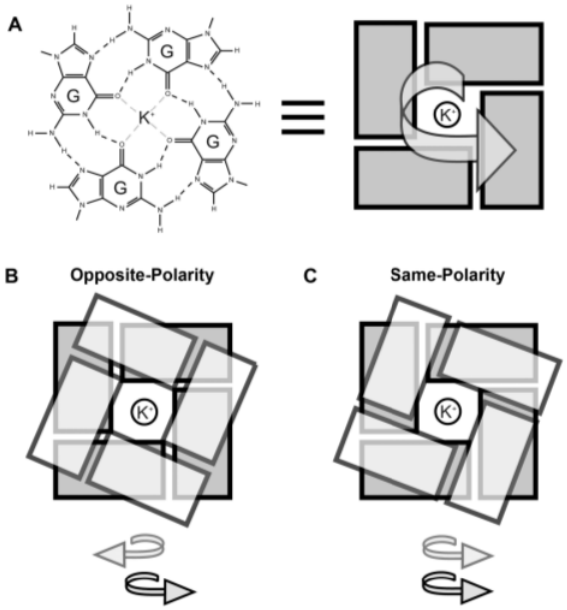

FIGURE 4.3: G-quartets stacking: (A) The polarity is defined by the direction of the Hoogsteen hydrogen bond donor-to-acceptor pattern. The two stacking possibility are (B) opposite-polarity or (C) same-polarity stacking [168].

that the band near 260 nm arises from the stacking of the stacking of the same polarity while the one observed at longer wavelenght arises from the stacking of quartets with alternating polarities [169] (see Figure 4.3).

Four crystal structures were considered, representing each one a particular quadruplex folding topology: 1KF1 (parallel), 143D (antiparallel), 2HY9 (hybrid-1), 2JPZ (hybrid-2). A 80 ns molecular dynamics simulation were performed on each structure to explore the DNA conformational space, a snapshot is saved at every ns and it is used as the starting geometry for the calculation of the excited states. In the case of G-quadruplex DNA the chromophores absorbing in the UV-vis region are represented by guanines while other contributions are negligible. For this reason, the three stacked guanine quartets were considered as part of the QM region

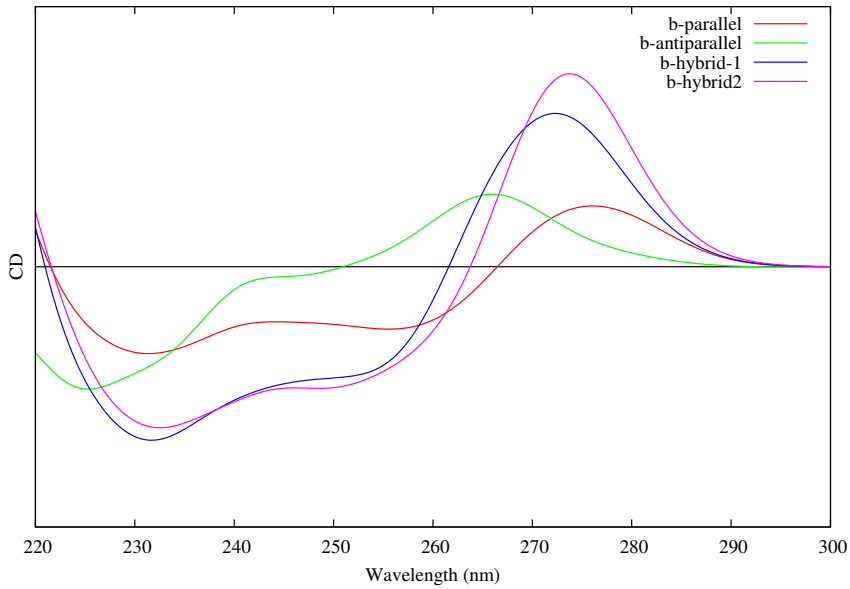

FIGURE 4.4: CD spectra obtained with the single base QM/MM calculations.

while leaving loops and flanking base in the MM region. First, a single guanine base was included in the QM region of each calculation and the Frenkel excitation theory was used to obtain the CD spectra of the overall structure (Figure 4.4). As it can be seen the position of the positive peak is similar for the four conformations. Moreover, the hybrid structures does not show the typical shoulder on the positive peak and the position of the parallel and antiparallel positive band is inverted. Both hybrids have a similar spectrum, so only the hybrid-1 conformation will be further discussed. Considering the pour accord with the experimental data obtained with the quartet approach, a complete row (three stacked guanines) was included in the QM level. If the shape of CD spectra depends on the stacking orientation of the guanine bases having a different glycosidic angle, it could be possible to obtain a better agreement with experimental data including the complete row in the QM level. The results are shown in Figure 4.5. With the row-approach, the band shapes for both systems appear to be quite correctly reproduced. The parallel spectrum shows a positive band around 250 nm, the experimental one is around 260 nm. The antiparallel shows a negative band at about 220 nm and a positive peak around 250 nm, they result shifted of about 40 nm. In the case of the hybrid a negative band at 225 nm and a positive band at 255 nm with a shoulder at about 240 nm were found. Also in this case there is a shift of about

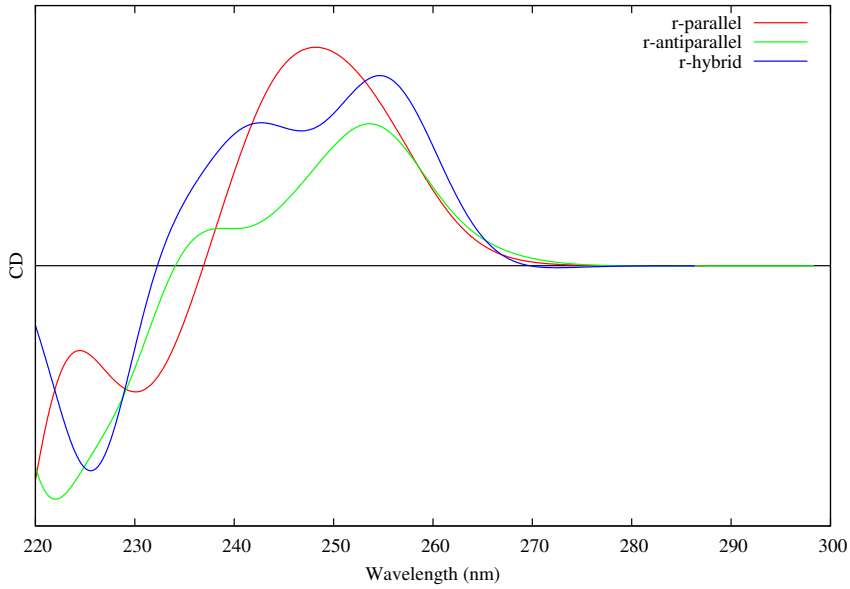

FIGURE 4.5: CD spectra of the three quadruplex conformations using the row approach.

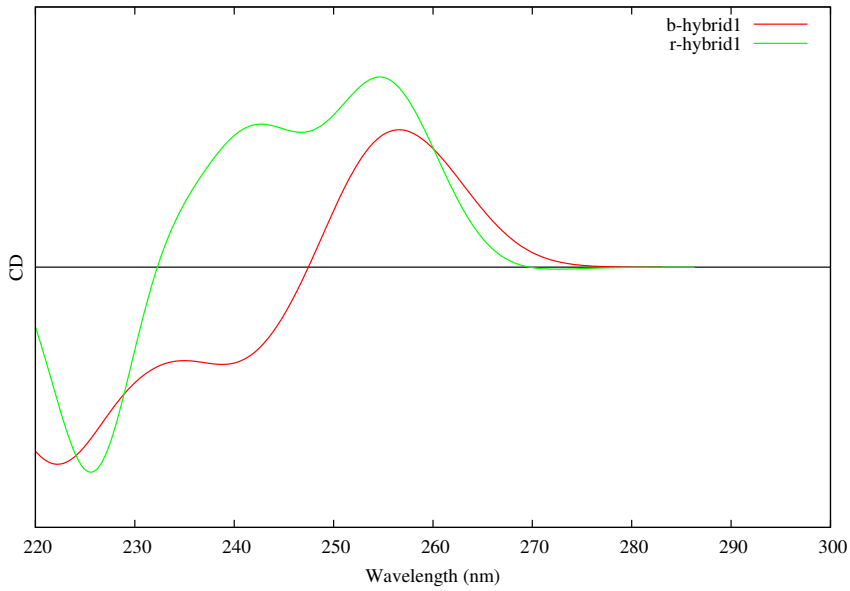

FIGURE 4.6: A comparison of the hybrid CD spectra calculated with single-base and the row selection.

40 nm comparing with the experimental data.

Finally, in Figure 4.6 G-quadruplex spectra for the hybrid-1 conformation are reported using the two different considered approaches. Remarkably, in the hybrid spectrum only when the rows are explicitly included inside the QM level the typical shoulder (around 240 nm) appears. This is due to the fact that in the hybrid conformation the first and the second quartets are stacked with an opposite-polarities orientation (like in antiparallel structure), while the second and the third adopt a same-polarity orientation (like parallel structure). Thus, the hybrid conformation would have a CD spectrum that is composite of contributions of the two types of G-quartet stacks (see Figure 4.2).

4.2.1 Conclusions

Summarizing, the electronic circular dichroism spectra were calculated for the most relevant G-quadruplex conformations. The proposed protocol consisted in molecular dynamics simulations, to sample the DNA conformational space, and QM/MM calculations to compute the properties of quadruplex excited states. Frenkel Hamiltonian approach proved to be a valid method also for the prediction of G-quadruplex ECD spectra. An accurate selection of the bases to be included in the high level of theory of the QM/MM calculations seems to be fundamental. In fact, only selecting the rows (three stacked guanines) a good agreement with the experimental data was achieved.

4.3 Computational Details

Molecular Dynamics simulations were performed with the AMBER software [170], details about MD simulations are discussed in section 2.5.1. The QM/MM border was placed at the junction between the sugar and the guanine base, and the dangling bond was treated with the link atom scheme [171]. The quantum level of the computations is performed at time-dependent density functional theory level, using the M06-2X [46] functional and 6-31G basis set [172]. The energy of the six first exited states of each base were calculation as well as their electronic transition dipole moments. Excitation energies and transition dipole moments were used to build the Frenkel Hamiltonian and simulate the electronic circular dichroism spectra of the different quadruplex structures. All the QM/MM calculations were performed using a modified version of Gaussian 09, Revision B01 [109], coupled with Tinker [88]. Spectra band shapes have been obtained convoluting each snapshot with a Gaussian function of full width and half-length of 0.2 eV. Frenkel Hamiltonian ECD spectra have been obtained using a code developed at the University of Lorraine (Nancy) to post-process Gaussian 09 outputs.

Angelo Spinello

Curriculum Vitae

Personal Data

citizenship: Italian
languages: Italian (native)
English (B2)
German (A2)
cell phone: +39 328 5674705
email:
angelo.spinello@gmail.com
angelo.spinello@unipa.it

Professional Experience

Université de Lorraine, Structure and Reactivity of Complex Molecular Systems (SRSMC). COST Short-Term Scientific Mission (tutor Dr. Antonio Monari).

06/2015 – 07/2015

Prediction of G-quadruplex circular dichroism spectra. Binding free energy of G-quadruplex stabilizers.

University of Groningen, Research Institute of Pharmacy. COST Short-Term Scientific Mission (tutor Prof. Angela Casini).

03/2015 – 05/2013

Molecular Dynamics simulations of aquaporins, inhibition of water and glycerol permeation in AQP3 induced by mercury.

TU Braunschweig, Institut für Organische Chemie. Visiting PhD student (tutor Prof. Jörg Grunenberg).

03/2013 – 12/2013

Study of the interaction of small organic binders with G-quadruplex DNA. A comparison of different force fields to describe the relative stability of G-quadruplex stems.

Education

University of Palermo, PhD in Chemistry - applied for Doctor Europaeus - (tutor Dr. Giampaolo Barone). Title: Interaction of metal complexes with G-quadruplex DNA.

01/2013 – 12/2015

University of Palermo, Master degree in Chemistry, summa cum laude.

01/2010 – 03/2012

Knowledge and Skills

- **Molecular Dynamics (GROMACS, Amber)**
- **Quantum Chemistry: DFT, QM/MM (Gaussian)**
- **Conformational Analysis (Macromodel)**
- **Molecular visualization and modeling (VMD, molden, maestro, MOE)**
- **Basic knowledge of Python, Bash, Fortran and LaTeX**

Proceedings

2013

- ***A. Spinello, G. Barone, J. Grunenberg*: A computational study of the molecular recognition process of small molecules by telomeric quadruplex-DNA (poster), Conference on Scientific Computing - CSC 2013, 02-05 December 2013, Paphos, Cyprus.**
  Awarded PRACE fellowship for the participation to the conference.

2014

- ***A. Terenzi, R. Bonsignore, A. Spinello, A.M. Almerico, A. Lauria, G. Barone*: Quadruplex-DNA vs. B-DNA binding of Schiff base transition metal complexes, Gordon Research Conference on Metals in Medicine, 22-27 June 2014, Andover, NH, USA.**

- ***A. Spinello, A. Terenzi, R. Bonsignore, A. Marrone, A. Martorana, A.M. Almerico, A. Lauria, G. Barone*: The Interaction of Small Molecules with Biomolecules, Congresso "Ricerca di base, interdisciplinare e traslazionale in ambito Biologico e Biotecnologico", 26-27 Giugno 2014, Palermo.**

- ***A. Marrone, R. Bonsignore, A. Terenzi, A. Spinello, C. Gentile, A. Martorana, A.M. Almerico, A. Lauria, G. Barone*: Ni^{II}, and Zn^{II} Schiff Base Complexes: Telomeric G-quadruplex Stabilizers, Congresso "Ricerca di base, interdisciplinare e traslazionale in ambito Biologico e Biotecnologico", 26-27 Giugno 2014, Palermo.**

- ***A. Terenzi, R. Bonsignore, A. Spinello, A.M. Almerico, A. Lauria, G. Barone*: DNA-Binding of Ni^{II}, Cu^{II} and Zn^{II} Complexes of Salen Derivatives, Third Whole Action Meeting of the COST Action CM1105, 22 August 2014, Zurich.**

- ***G. Barone, A. Terenzi, R. Bonsignore, A. Spinello, A.M. Almerico, A. Lauria*: G4-DNA vs. B-DNA binding of Schiff base transition metal complexes, 12th European Biological Inorganic Chemistry Conference, 24-28 August 2014, Zurich.**

- ***A. Spinello, A. Terenzi, R. Bonsignore, A.M. Almerico, A. Lauria, G. Barone*: G4-DNA vs. B-DNA binding of Schiff base transition metal complexes, COST Training School: Chemistry of Metals in Biological Systems, 07-14 September 2014, Université Catholique De Louvain, Lovain-La-Neuve, Belgium.**

- ***A. Spinello, G. Barone, M. G. Ortore, F. Spinozzi, S. Vilasi, A. Palumbo Piccionello*: Quaternary structure of GroEL and Hsp60 chaperones (oral communication), XXII meeting of the "Italian Biophysical Society", 21-24 September 2014, Palermo, Italy.**

- ***Giampaolo Barone, Alessio Terenzi, Riccardo Bonsignore, Angelo Spinello, Anna Maria Almerico, Antonino Lauria*: DNA-Binding of Ni^{II}, Cu^{II} and Zn^{II} Schiff Base Complexes, Convegno**

Congiunto delle Sezioni Calabria e Sicilia 2014, 1-2 December 2014, Palermo.

2015

- ***A. Spinello, G. Barone, A. Pace, F. Cappello, S. Buscemi, A.P. Piccionello*: Binding of Epolactaene to Hsp60: a Computational Study, XIX School of Pure and Applied Biophysics "Theoretical and Computational Approaches to Biophysics", 26-30 January 2015, Venice.**

- ***A. Spinello, M. G. Ortore, F. Spinozzi, C. Ricci, G. Barone, A. M. Gammazza, A. Palumbo Piccionello*: Quaternary Structure of GroEL and Hsp60 Chaperones, 10th European Biophysics Congress, 18 July2015, Dresden, Germany.**
  Awarded EBSA fellowship for the participation to the conference.

- ***A. Spinello*:The interaction of Nickel(II) metal complexes with B and G-quadruplex DNA, European Summer School in Quantum Chemistry, ESQC 2015, Torre Normanna, 6-19 September 2015, Palermo.**

- ***G. Barone, R. Bonsignore, A. Spinello, A. Terenzi*: Schiff Base Metal Complexes as Selective G-quadruplex Binders, Meeting of the WG2 of the COST Action CM1105, 5 October 2015, Lisbon.**

- ***R. Bonsignore, A. Terenzi, A. Spinello, C. Gentile, A. Martorana, A. Lauria, G. Barone, A.M. Almerico*: Complessi di Ni^{II} e Zn^{II} di basi di Schiff: sintesi e studio dell'interazione con B-DNA e quadruplex telomerico, Convegno Congiunto SCI Calabria-Sicilia 2015, 3-4 December 2015, Catanzaro.**

Workshops and Schools

- **Classical Molecular Dynamics for the Simulation of Biological Systems, 22-24 February 2012, CASPUR, Rome.**

- **COST Training School: Chemistry of Metals in Biological Systems, 07-14 September 2014, Université Catholique De Louvain, Lovain-La-Neuve, Belgium. *A. Spinello, A. Terenzi, R. Bonsignore,***

A.M. Almerico, A. Lauria, G. Barone: G4-DNA vs. B-DNA binding of Schiff base transition metal complexes (poster).
Awarded COST fellowship for the participation to the School.

- Efficient use of Molecular Dynamics simulation application in an HPC environment, 17-19 November 2014, CINECA (PRACE Advanced Training Centres course), Bologna.
- XIX School of Pure and Applied Biophysics on "Theoretical and Computational Approaches to Biophysics", 26 - 30 January 2015, Venice, Campo Santo Stefano. ***A. Spinello, G. Barone, A. Pace, F. Cappello, S. Buscemi, A. P. Piccionello***: Binding of epolactaene to Hsp60: a computational study (poster).
- European Summer School in Quantum Chemistry, ESQC 2015, 6-19 September 2015, Torre Normanna, Palermo.
- DNA damages: modeling and rationalize structure and reactivity, to be attended, 3-6 November 2015, Lyon.
- Python for computational science, to be attended, 1-3 December, CINECA (PRACE Advanced Training Centres course), Rome.

HPC Projects

- Standard HPC Grant 2012: The interaction of small molecules with oligonucleotides, approved on MATRIX machine of CASPUR, Rome.
- ISCRA B 2012: Structure, dynamics and reactivity of DNA intercalating transition metal complexes, approved on PLX machine of CINECA, Bologna.
- ISCRA B 2013: The interaction of transition metal complexes with G-quadruplex DNA, approved on FERMI machine of CINECA, Bologna.

Publications

Submitted

- **A. Spinello, G. Barone, F. Cappello, A. Pace, S. Buscemi, A.P. Piccionello: The binding mechanism of epoactaene to Hsp60 unveiled by in silico modelling.**
- **D. Schillaci, G. Barone, M.G. Cusimano, S. Cascioferro, A. Spinello, M. Vitale, V. Arizza: A human β thymosin and the analogue fragment from sea urchin as a platform for the development of new anti-biofilms agents for Staphylococcus spp. and Pseudomonas aeruginosa.**

2016

- **A. Spinello, A. de Almeida, A. Casini, G. Barone: The inhibition of glycerol permeation through Aquaporin 3 induced by mercury(II): a molecular dynamics study, *J. Inorg. Biochem.*, DOI: 10.1016/j.jinorgbio.2015.11.027.**
- **A. Spinello, G. Barone, J. Grunenberg: Molecular Recognition of Naphthalene Diimide Ligands by Telomeric Quadruplex-DNA: The Importance of the Protonation State and non-Mediated Hydrogen Bonds, *Phys. Chem. Chem. Phys.*, DOI: 10.1039/C5CP05576H.**

2015

- **I. Pibiri, L. Lentini, R. Melfi, G. Gallucci, A. Pace, A. Spinello, G. Barone, A. Di Leonardo: Enhancement of Premature Stop Codon Readthrough in the CFTR Gene by Ataluren (PTC124) Derivatives, *Eur. J. Med. Chem.*, 101 (2015) 236.**
- **A. Spinello, M.G. Ortore, F. Spinozzi, C. Ricci, G. Barone, A. Marino Gammazza, A.P. Piccionello: Quaternary structure of GroEL and naïve-Hsp60 chaperones: a combined SAXS-MD study, *RSC Adv.*, 5 (2015) 49871.**

2014

- **L. Lentini, R. Melfi, A. Di Leonardo, A. Spinello, G. Barone, A. Pace, A.P. Piccionello, I. Pibiri: Towards a rationale for the**

PTC124 (Ataluren) promoted readthrough of premature stop codons: a computational approach and GFP-reporter cell-based assay, ***Mol. Pharmaceutics***, 11 (2014) 653.

- A. Lauria, R. Bonsignore, A. Terenzi, A. Spinello, F. Giannici, A. Longo, A.M. Almerico, G. Barone: Nickel(II), copper(II) and zinc(II) metallo-intercalators: structural details of the DNA-binding by a combined experimental and computational investigation, ***Dalton Trans.***, 43 (2014) 6108.

- A. Biancardi, A. Burgalassi, A. Terenzi, A. Spinello, G. Barone, T. Biver, B. Mennucci: A combined theoretical-experimental investigation of the spectroscopic properties of a DNA-intercalator ZnII Salphen-type complex, ***Chem. Eur. J.***, 20 (2014) 7439.

- A. Terenzi, R. Bonsignore, A. Spinello, C. Gentile, A. Martorana, C. Ducani, B. Högberg, A.M. Almerico, A. Lauria, G. Barone: Selective G-Quadruplex Stabilizers: Schiff-base Metal Complexes with Anticancer Activity, ***RSC Adv.***, 4 (2014) 33245.

- J. Grunenberg, G. Barone, A. Spinello: The right answer for the right electrostatics: Force field methods are able to describe relative energies of DNA guanine quadruplexes, ***J. Chem. Theory Comp.***, 10 (2014) 2901.

- D. Schillaci, M.G. Cusimano, A. Spinello, G. Barone, M. Vitale, D. Russo, D. Parrinello, V. Arizza: Paracentrin 1, a synthetic antimicrobial peptide from the sea-urchin Paracentrotus lividus, interferes with staphylococcal and Pseudomonas aeruginosa biofilm formation, ***AMB express***, 4 (2014) 78.

2013

- A. Spinello, A. Terenzi, G. Barone: Metal Complex–DNA Binding: Insights from Molecular Dynamics and DFT/MM Calculations, ***J. Inorg. Biochem.***, 124 (2013) 63.

2012

- A. Terenzi, L. Tomasello, A. Spinello, G. Bruno, C. Giordano, G. Barone: (Dipyrido[3,2-a:2',3'c]phenazine) (glycinato) copper(II) perchlorate: A novel DNA-intercalator with anti-proliferative activity against thyroid cancer cell lines, ***J. Inorg. Biochem.***, 117 (2012) 103.

Bibliography

[1] http://philschatz.com/biology-book/contents/m44486.html, .

[2] James D Watson and Francis HC Crick. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738, 1953.

[3] Martin Egli and Wolfram Saenger. *Principles of nucleic acid structure*. Springer Science & Business Media, 1984.

[4] Carole A Bewley, Angela M Gronenborn, and G Marius Clore. Minor groove-binding architectural proteins: structure, function, and dna recognition 1. *Annual review of biophysics and biomolecular structure*, 27(1):105–131, 1998.

[5] Viktor Brabec and Jana Kasparkova. Modifications of dna by platinum complexes: relation to resistance of tumors to platinum antitumor drugs. *Drug Resistance Updates*, 8(3):131–146, 2005.

[6] http://www.atdbio.com/content/16/nucleic-acid-drug-interactions, .

[7] Bernhard Lippert. Impact of cisplatin on the recent development of pt coordination chemistry: a case study. *Coordination chemistry reviews*, 182(1): 263–295, 1999.

[8] X Chen, Y Wu, H Dong, C Y Zhang, and Y Zhang. Platinum-based agents for individualized cancer treatmen. *Current molecular medicine*, 13(10):1603–1612, 2013.

[9] Shigehiro Kamitori and Fusao Takusagawa. Multiple binding modes of anticancer drug actinomycin d: X-ray, molecular modeling, and spectroscopic studies of d (gaagcttc) 2-actinomycin d complexes and its host dna. *Journal of the American Chemical Society*, 116(10):4154–4165, 1994.

[10] LS Lerman. The structure of the dna-acridine complex. *Proceedings of the National Academy of Sciences of the United States of America*, 49(1):94, 1963.

[11] Federico Gago. Stacking interactions and intercalative dna binding. *Methods*, 14(3):277–292, 1998.

[12] Nguyen T Thuong and Claude Hélène. Sequence-specific recognition and modification of double-helical dna by oligonucleotides. *Angewandte Chemie International Edition in English*, 32(5):666–690, 1993.

[13] Pier Giovanni Baraldi, Andrea Bovero, Francesca Fruttarolo, Delia Preti, Mojgan Aghazadeh Tabrizi, Maria Giovanna Pavani, and Romeo Romagnoli. Dna minor groove binders as potential antitumor and antimicrobial agents. *Medicinal research reviews*, 24(4):475–528, 2004.

[14] Helen M Coley, Julien Sarju, and Gabriele Wagner. Synthesis and characterization of platinum (ii) oxadiazoline complexes and their in vitro antitumor activity in platinum-sensitive and-resistant cancer cell lines. *Journal of medicinal chemistry*, 51(1):135–141, 2007.

[15] David Gilley, Hiromi Tanaka, and Brittney-Shea Herbert. Telomere dysfunction in aging and cancer. *The international journal of biochemistry & cell biology*, 37(5):1000–1013, 2005.

[16] Haiyong Han and Laurence H Hurley. G-quadruplex dna: a potential target for anti-cancer drug design. *Trends in pharmacological sciences*, 21(4):136–142, 2000.

[17] Naijie Jing, Wei Sha, Yidong Li, Weijun Xiong, and David J Tweardy. Rational drug design of g-quartet dna as anti-cancer agents. *Current pharmaceutical design*, 11(22):2841–2854, 2005.

[18] Nam W Kim, Mieczyslaw A Piatyszek, Karen R Prowse, Calvin B Harley, Michael D West, P d L Ho, Gina M Coviello, Woodring E Wright, Scott L Weinrich, and Jerry W Shay. Specific association of human telomerase activity with immortal cells and cancer. *Science*, 266(5193):2011–2015, 1994.

[19] Daekyu Sun, Brian Thompson, Brian E Cathers, Miguel Salazar, Sean M Kerwin, John O Trent, Terence C Jenkins, Stephen Neidle, and Laurence H Hurley. Inhibition of human telomerase by a g-quadruplex-interactive compound. *Journal of medicinal chemistry*, 40(14):2113–2116, 1997.

[20] Giulia Biffi, David Tannahill, John McCafferty, and Shankar Balasubramanian. Quantitative visualization of dna g-quadruplex structures in human cells. *Nature chemistry*, 5(3):182–186, 2013.

[21] Tracy A Brooks and Laurence H Hurley. Targeting myc expression through g-quadruplexes. *Genes & cancer*, 1(6):641–649, 2010.

[22] Stephen Neidle. *Therapeutic applications of quadruplex nucleic acids*. Academic Press, 2011.

[23] Julie E Reed, Anna Arola Arnal, Stephen Neidle, and Ramón Vilar. Stabilization of g-quadruplex dna and inhibition of telomerase activity by square-planar nickel (ii) complexes. *Journal of the American Chemical Society*, 128 (18):5992–5993, 2006.

[24] Julie E Reed, Stephen Neidle, and Ramon Vilar. Stabilisation of human telomeric quadruplex dna and inhibition of telomerase by a platinum–phenanthroline complex. *Chem. Commun.*, (42):4366–4368, 2007.

[25] Hélène Bertrand, David Monchaud, Anne De Cian, Régis Guillot, Jean-Louis Mergny, and Marie-Paule Teulade-Fichou. The importance of metal geometry in the recognition of g-quadruplex-dna by metal-terpyridine complexes. *Organic & biomolecular chemistry*, 5(16):2555–2559, 2007.

[26] Nancy H Campbell, Nurul H Abd Karim, Gary N Parkinson, Mekala Gunaratnam, Vanessa Petrucci, Alan K Todd, Ramon Vilar, and Stephen Neidle. Molecular basis of structure–activity relationships between salphen metal complexes and human telomeric dna quadruplexes. *Journal of medicinal chemistry*, 55(1):209–222, 2011.

[27] Savvas N Georgiades, Nurul H Abd Karim, Kogularamanan Suntharalingam, and Ramon Vilar. Interaction of metal complexes with g-quadruplex dna. *Angewandte Chemie International Edition*, 49(24):4020–4034, 2010.

[28] Aneta Oleksi, Alexandre G Blanco, Roeland Boer, Isabel Usón, Joan Aymamí, Alison Rodger, Michael J Hannon, and Miquel Coll. Molecular recognition of a three-way dna junction by a metallosupramolecular helicate. *Angewandte Chemie*, 118(8):1249–1253, 2006.

[29] Mary Elizabeth Peek and Loren Dean Williams. X-ray crystallography of dna-drug complexes. *Methods in enzymology*, 340:282–290, 2001.

[30] Charles A Laughton and Sarah A Harris. The atomistic simulation of dna. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(4):590–600, 2011.

[31] Alberto Pérez, F Javier Luque, and Modesto Orozco. Frontiers in molecular dynamics simulations of dna. *Accounts of chemical research*, 45(2):196–205, 2011.

[32] Syma Khalid, Michael J Hannon, Alison Rodger, and P Mark Rodger. Simulations of dna coiling around a synthetic supramolecular cylinder that binds in the dna major groove. *Chemistry-a European Journal*, 12(13):3493–3506, 2006.

[33] Simona Fantacci, Filippo De Angelis, Antonio Sgamellotti, Alessandro Marrone, and Nazzareno Re. Photophysical properties of [ru (phen) 2 (dppz)] 2+ intercalated into dna: An integrated car-parrinello and tddft study. *Journal of the American Chemical Society*, 127(41):14144–14145, 2005.

[34] Daniel Svozil, Pavel Hobza, and Jiří Šponer. Comparison of intrinsic stacking energies of ten unique dinucleotide steps in a-rna and b-dna duplexes. can we determine correct order of stability by quantum-chemical calculations? *The Journal of Physical Chemistry B*, 114(2):1191–1203, 2009.

[35] Jörg Grunenberg. Complexity in molecular recognition. *Physical Chemistry Chemical Physics*, 13(21):10136–10146, 2011.

[36] Frank H Wallrapp and Victor Guallar. Mixed quantum mechanics and molecular mechanics methods: looking inside proteins. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(2):315–322, 2011.

[37] Lung Wa Chung, Hajime Hirao, Xin Li, and Keiji Morokuma. The oniom method: its foundation and applications to metalloenzymes and photobiology. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(2):327–350, 2012.

[38] Hauke Paulsen, Alfred X Trautwein, Patrick Wegner, Christian Schmidt, Aleksandr I Chumakov, and Volker Schünemann. Interpretation of nuclear resonant vibrational spectra of rubredoxin using a combined quantum mechanics and molecular mechanics approach. *ChemPhysChem*, 12(17):3434–3441, 2011.

[39] Qianqian Hou, Likai Du, Jun Gao, Yongjun Liu, and Chengbu Liu. Qm/mm study on the reaction mechanism of o6-alkylguanine- dna alkyltransferase. *The Journal of Physical Chemistry B*, 114(46):15296–15300, 2010.

[40] Konstantinos Gkionis and James A Platts. Qm/mm investigation into binding of square-planar platinum complexes to dna fragments. *JBIC Journal of Biological Inorganic Chemistry*, 14(8):1165–1174, 2009.

[41] Célia Fonseca Guerra, Tushar van der Wijst, Jordi Poater, Marcel Swart, and F Matthias Bickelhaupt. Adenine versus guanine quartets in aqueous solution: dispersion-corrected dft study on the differences in π-stacking and hydrogen-bonding behavior. *Theoretical Chemistry Accounts*, 125(3-6):245–252, 2010.

[42] Rebeca Ruiz, Begoña García, Giuseppe Ruisi, Arturo Silvestri, and Giampaolo Barone. Computational study of the interaction of proflavine with d (atatatatat) 2 and d (gcgcgcgcgc) 2. *Journal of Molecular Structure: THEOCHEM*, 915(1):86–92, 2009.

[43] Roman V Reshetnikov, Jiri Sponer, Olga I Rassokhina, Alexei M Kopylov, Philipp O Tsvetkov, Alexander A Makarov, and Andrey V Golovin. Cation binding to 15-tba quadruplex dna is a multiple-pathway cation-dependent process. *Nucleic acids research*, 39(22):9789–9802, 2011.

[44] Christian Gossens, Ivano Tavernelli, and Ursula Rothlisberger. Dna structural distortions induced by ruthenium- arene anticancer compounds. *Journal of the American Chemical Society*, 130(33):10921–10928, 2008.

[45] Stefan Grimme. Density functional theory with london dispersion corrections. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(2):211–228, 2011.

[46] Yan Zhao and Donald G Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, 120(1-3):215–241, 2008.

[47] Raphael F Ribeiro, Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. The solvation, partitioning, hydrogen bonding, and dimerization of nucleotide bases: a multifaceted challenge for quantum chemistry. *Physical Chemistry Chemical Physics*, 13(23):10908–10922, 2011.

[48] Angelo Spinello, Alessio Terenzi, and Giampaolo Barone. Metal complex–dna binding: Insights from molecular dynamics and dft/mm calculations. *Journal of inorganic biochemistry*, 124:63–69, 2013.

[49] Laura Lentini, Raffaella Melfi, Aldo Di Leonardo, Angelo Spinello, Giampaolo Barone, Andrea Pace, Antonio Palumbo Piccionello, and Ivana Pibiri. Toward

a rationale for the ptc124 (ataluren) promoted readthrough of premature stop codons: a computational approach and gfp-reporter cell-based assay. *Molecular pharmaceutics*, 11(3):653–664, 2014.

[50] Ivana Pibiri, Laura Lentini, Raffaella Melfi, Giulia Gallucci, Andrea Pace, Angelo Spinello, Giampaolo Barone, and Aldo Di Leonardo. Enhancement of premature stop codon readthrough in the cftr gene by ataluren (ptc124) derivatives. *European journal of medicinal chemistry*, 101:236–244, 2015.

[51] Angelo Spinello, Francesco Spinozzi, Caterina Ricci, Giampaolo Barone, Maria Grazia Ortore, Antonella Marino Gammazza, and Antonio Palumbo Piccionello. Quaternary structures of groel and naïve-hsp60 chaperones in solution: a combined saxs-md study. *RSC Advances*, 2015.

[52] Angelo Spinello, Andreia de Almeida, Angela Casini, and Giampaolo Barone. The inhibition of glycerol permeation through aquaglyceroporin-3 induced by mercury(ii): A molecular dynamics study. *Journal of Inorganic Biochemistry*, 2015. doi: http://dx.doi.org/10.1016/j.jinorgbio.2015.11.027.

[53] Domenico Schillaci, Maria Grazia Cusimano, Angelo Spinello, Giampaolo Barone, Debora Russo, Maria Vitale, Daniela Parrinello, and Vincenzo Arizza. Paracentrin 1, a synthetic antimicrobial peptide from the sea-urchin paracentrotus lividus, interferes with staphylococcal and pseudomonas aeruginosa biofilm formation. *AMB Express*, 4(1):78, 2014.

[54] J Andrew McCammon, Bruce R Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977.

[55] Jarosaw Meller. *Molecular Dynamics*. John Wiley and Sons, Ltd, 2001. ISBN 9780470015902. doi: 10.1038/npg.els.0003048. URL `http://dx.doi.org/10.1038/npg.els.0003048`.

[56] Kresten Lindorff-Larsen, Paul Maragakis, Stefano Piana, Michael P Eastwood, Ron O Dror, and David E Shaw. Systematic validation of protein force fields against experimental data. *PloS one*, 7(2):e32131, 2012.

[57] Audray K Harris, Joel R Meyerson, Yumiko Matsuoka, Oleg Kuybeda, Amy Moran, Donald Bliss, Suman R Das, Jonathan W Yewdell, Guillermo Sapiro, Kanta Subbarao, et al. Structure and accessibility of ha trimers on intact 2009 h1n1 pandemic influenza virus to stem region-specific neutralizing antibodies. *Proceedings of the National Academy of Sciences*, 110(12):4592–4597, 2013.

[58] Yibing Shan, Eric T Kim, Michael P Eastwood, Ron O Dror, Markus A Seeliger, and David E Shaw. How does a drug molecule find its target binding site? *Journal of the American Chemical Society*, 133(24):9181–9183, 2011.

[59] Celeste Sagui and Thomas A Darden. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure*, 28(1):155–179, 1999.

[60] Lung Wa Chung, W. M. C. Sameera, Romain Ramozzi, Alister J. Page, Miho Hatanaka, Galina P. Petrova, Travis V. Harris, Xin Li, Zhuofeng Ke, Fengyi Liu, Hai-Bei Li, Lina Ding, and Keiji Morokuma. The oniom method and its applications. *Chemical Reviews*, 115(12):5678–5796, 2015.

[61] Barry Honig and Martin Karplus. Implications of torsional potential of retinal isomers for visual excitation. 1971.

[62] A Warshel and M Karplus. Calculation of ground and excited state potential surfaces of conjugated molecules. i. formulation and parametrization. *Journal of the American Chemical Society*, 94(16):5612–5625, 1972.

[63] Michael Levitt and Jonathan Greer. Automatic identification of secondary structure in globular proteins. *Journal of molecular biology*, 114(2):181–239, 1977.

[64] Martin Karplus. Development of multiscale models for complex chemical systems: from h+ h2 to biomolecules (nobel lecture). *Angewandte Chemie International Edition*, 53(38):9992–10005, 2014.

[65] Michael Levitt. Birth and future of multiscale modeling for macromolecular systems (nobel lecture). *Angewandte Chemie International Edition*, 53(38): 10006–10018, 2014.

[66] Arieh Warshel. Multiscale modeling of biological functions: from enzymes to molecular machines (nobel lecture). *Angewandte Chemie International Edition*, 53(38):10020–10031, 2014.

[67] Hans Martin Senn and Walter Thiel. Qm/mm methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.

[68] Thom Vreven and Keiji Morokuma. Hybrid methods: Oniom (qm: mm) and qm/mm. *Annual reports in computational chemistry*, 2:35–51, 2006.

[69] Stéphane Humbel, Stefan Sieber, and Keiji Morokuma. The imomo method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and sn2 reaction: Rcl+ cl-. *The Journal of chemical physics*, 105(5):1959–1967, 1996.

[70] Zhenqin Li and Harold A Scheraga. Monte carlo-minimization approach to the multiple-minima problem in protein folding. *Proceedings of the National Academy of Sciences*, 84(19):6611–6615, 1987.

[71] Kristen A Fichthorn and W Hh Weinberg. Theoretical foundations of dynamical monte carlo simulations. *The Journal of chemical physics*, 95(2): 1090–1096, 1991.

[72] David A Atwood and Melanie J Harvey. Group 13 compounds incorporating salen ligands. *Chemical reviews*, 101(1):37–52, 2001.

[73] Giampaolo Barone, Noemi Gambino, Angela Ruggirello, Arturo Silvestri, Alessio Terenzi, and Vincenzo Turco Liveri. Spectroscopic study of the interaction of ni ii-5-triethyl ammonium methyl salicylidene ortho-phenylendiiminate with native dna. *Journal of inorganic biochemistry*, 103(5):731–737, 2009.

[74] Arturo Silvestri, Giampaolo Barone, Giuseppe Ruisi, Daniele Anselmo, Serena Riela, and Vincenzo Turco Liveri. The interaction of native dna with zn (ii) and cu (ii) complexes of 5-triethyl ammonium methyl salicylidene ortophenylendiimine. *Journal of inorganic biochemistry*, 101(5):841–848, 2007.

[75] Carlos Baleizao and Hermenegildo Garcia. Chiral salen complexes: an overview to recoverable and reusable homogeneous and heterogeneous catalysts. *Chemical reviews*, 106(9):3987–4043, 2006.

[76] Laetitia Canali and David C Sherrington. Utilisation of homogeneous and supported chiral metal (salen) complexes in asymmetric catalysis. *Chemical Society Reviews*, 28(2):85–93, 1999.

[77] Antonella Dalla Cort, Paolo De Bernardin, Gianpiero Forte, and Francesco Yafteh Mihan. Metal–salophen-based receptors for anions. *Chemical Society Reviews*, 39(10):3863–3874, 2010.

[78] Dennis J Gravert and John H Griffin. Steric and electronic effects, enantiospecificity, and reactive orientation in dna binding/cleaving by substituted derivatives of [salenmniii]+. *Inorganic chemistry*, 35(17):4837–4847, 1996.

[79] Antonino Lauria, Riccardo Bonsignore, Alessio Terenzi, Angelo Spinello, Francesco Giannici, Alessandro Longo, Anna Maria Almerico, and Giampaolo Barone. Nickel (ii), copper (ii) and zinc (ii) metallo-intercalators: structural details of the dna-binding by a combined experimental and computational investigation. *Dalton Transactions*, 43(16):6108–6119, 2014.

[80] Alessio Terenzi, Riccardo Bonsignore, Angelo Spinello, Carla Gentile, Annamaria Martorana, Cosimo Ducani, Björn Högberg, Anna Maria Almerico, Antonino Lauria, and Giampaolo Barone. Selective g-quadruplex stabilizers: Schiff-base metal complexes with anticancer activity. *RSC Advances*, 4(63): 33245–33256, 2014.

[81] Alessandro Biancardi, Azzurra Burgalassi, Alessio Terenzi, Angelo Spinello, Giampaolo Barone, Tarita Biver, and Benedetta Mennucci. A theoretical and experimental investigation of the spectroscopic properties of a dna-intercalator salphen-type znii complex. *Chemistry-A European Journal*, 20 (24):7439–7447, 2014.

[82] Marjorie M Harding. The geometry of metal-ligand interactions relevant to proteins. ii. angles at the metal atom, additional weak metal-donor interactions. *Acta Crystallographica Section D: Biological Crystallography*, 56(7): 857–867, 2000.

[83] Giampaolo Barone, Célia Fonseca Guerra, and F Matthias Bickelhaupt. B-dna structure and stability as function of nucleic acid composition: Dispersion-corrected dft study of dinucleoside monophosphate single and double strands. *ChemistryOpen*, 2(5-6):186–193, 2013.

[84] UC Singh and PA Kollman. Anapproach to computing electro static charges for molecules (j. *Comput Chem*, 5:129–145, 1984.

[85] Brent H Besler, Kenneth M Merz, and Peter A Kollman. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry*, 11(4):431–439, 1990.

[86] Tarita Biver, Fernando Secco, Maria Rosaria Tinè, Marcella Venturini, Andrea Bencini, Antonio Bianchi, and Claudia Giorgi. Intercalation of zn (ii) and cu (ii) complexes of the cyclic polyamine neotrien into dna: equilibria and kinetics. *Journal of inorganic biochemistry*, 98(9):1531–1538, 2004.

[87] B Th Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics*, 59(3):341–350, 1981.

[88] http://dasher.wustl.edu/tinker/.

[89] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.

[90] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.

[91] WD Cornell, P Cieplak, CI Bayly, IR Gould, KM Merz Jr, DM Ferguson, DC Spellmeyer, T Fox, JW Caldwell, and PA Kollman. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc*, 117:5179–5197, 1995.

[92] Alberto Pérez, Iván Marchán, Daniel Svozil, Jiri Sponer, Thomas E Cheatham, Charles A Laughton, and Modesto Orozco. Refinement of the amber force field for nucleic acids: improving the description of α/γ conformers. *Biophysical journal*, 92(11):3817–3829, 2007.

[93] Andrew T Guy, Thomas J Piggot, and Syma Khalid. Single-stranded dna within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study. *Biophysical journal*, 103(5):1028–1036, 2012.

[94] Alan W Sousa da Silva and Wim F Vranken. Acpype-antechamber python parser interface. *BMC research notes*, 5(1):367, 2012.

[95] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[96] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[97] Mats Svensson, Stephane Humbel, Robert DJ Froese, Toshiaki Matsubara, Stefan Sieber, and Keiji Morokuma. Oniom: A multilayered integrated mo+ mm method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and pt (p (t-bu) 3) 2+ h2 oxidative addition. *The Journal of Physical Chemistry*, 100(50):19357–19363, 1996.

[98] Thom Vreven and Keiji Morokuma. On the application of the imomo (integrated molecular orbital+ molecular orbital) method. *Journal of Computational Chemistry*, 21(16):1419–1432, 2000.

[99] Nathalie Godbout, Dennis R Salahub, Jan Andzelm, and Erich Wimmer. Optimization of gaussian-type basis sets for local spin density functional calculations. part i. boron through neon, optimization technique and validation. *Canadian Journal of Chemistry*, 70(2):560–571, 1992.

[100] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum mechanical continuum solvation models. *Chemical reviews*, 105(8):2999–3094, 2005.

[101] Giovanni Scalmani and Michael J Frisch. Continuous surface charge polarizable continuum models of solvation. i. general formalism. *The Journal of chemical physics*, 132(11):114110, 2010.

[102] Yan Zhao, Nathan E Schultz, and Donald G Truhlar. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *Journal of Chemical Theory and Computation*, 2(2):364–382, 2006.

[103] Edward G Hohenstein, Samuel T Chill, and C David Sherrill. Assessment of the performance of the m05- 2x and m06- 2x exchange-correlation functionals for noncovalent interactions in biomolecules. *Journal of Chemical Theory and Computation*, 4(12):1996–2000, 2008.

[104] Sérgio F Sousa, Emanuela S Carvalho, Diana M Ferreira, Isabel S Tavares, Pedro A Fernandes, Maria João Ramos, and JosÉ ANF Gomes. Comparative analysis of the performance of commonly available density functionals in the determination of geometrical parameters for zinc complexes. *Journal of computational chemistry*, 30(16):2752–2763, 2009.

[105] Alessandro Biancardi, Tarita Biver, Fernando Secco, and Benedetta Mennucci. An investigation of the photophysical properties of minor groove bound and intercalated dapi through quantum-mechanical and spectroscopic tools. *Physical Chemistry Chemical Physics*, 15(13):4596–4603, 2013.

[106] Stefano Caprasecca, Carles Curutchet, and Benedetta Mennucci. Toward a unified modeling of environment and bridge-mediated contributions to electronic energy transfer: A fully polarizable qm/mm/pcm approach. *Journal of Chemical Theory and Computation*, 8(11):4462–4473, 2012.

[107] BH Brent, KM Merz Jr, and PA Kollman. Atomic charges derived from semiempirical methods. *J. Comput. Chem. Vol*, 11:431–439, 1990.

[108] Piet Th Van Duijnen and Marcel Swart. Molecular and atomic polarizabilities: Thole's model revisited. *The Journal of Physical Chemistry A*, 102(14):2399–2407, 1998.

[109] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G Scalmani, V Barone, B Mennucci, GA Petersson, et al. Gaussian 09, revis. b01, 2010.

[110] Jörg Grunenberg, Giampaolo Barone, and Angelo Spinello. The right answer for the right electrostatics: Force field methods are able to describe relative energies of dna guanine quadruplexes. *Journal of Chemical Theory and Computation*, 10(8):2901–2905, 2014.

[111] A. Spinello, G. Barone, and J. Grunenberg. Molecular recognition of naphthalene diimide ligands by telomeric quadruplex-dna: the importance of the protonation state and mediated hydrogen bonds. *Phys. Chem. Chem. Phys.*, 2016. doi: 10.1039/C5CP05576H.

[112] Jiří Šponer, Arnošt Mládek, Naďa Špačková, Xiaohui Cang, Thomas E Cheatham III, and Stefan Grimme. Relative stability of different dna guanine quadruplex stem topologies derived using large-scale quantum-chemical computations. *Journal of the American Chemical Society*, 135(26):9785–9796, 2013.

[113] Sarah Burge, Gary N Parkinson, Pascale Hazel, Alan K Todd, and Stephen Neidle. Quadruplex dna: sequence, topology and structure. *Nucleic acids research*, 34(19):5402–5415, 2006.

[114] Luca Monticelli and D.Peter Tieleman. Force fields for classical molecular dynamics. In Luca Monticelli and Emppu Salonen, editors, *Biomolecular Simulations*, volume 924 of *Methods in Molecular Biology*, pages 197–213. Humana Press, 2013.

[115] Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.

[116] Robert S Mulliken. Electronic population analysis on lcao–mo molecular wave functions. i. *The Journal of Chemical Physics*, 23(10):1833–1840, 1955.

[117] U Chandra Singh and Peter A Kollman. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, 5(2): 129–145, 1984.

[118] François-Yves Dupradeau, Christine Cézard, Rodolphe Lelong, Elodie Stanislawiak, Julien Pêcher, Jean Charles Delepine, and Piotr Cieplak. Re dd. b.: a database for resp and esp atomic charges, and force field libraries. *Nucleic acids research*, 36(suppl 1):D360–D367, 2008.

[119] Wei Jiang, David J Hardy, James C Phillips, Alexander D MacKerell Jr, Klaus Schulten, and Benoît Roux. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical drude oscillators in namd. *The journal of physical chemistry letters*, 2(2):87–92, 2010.

[120] Peter Pulay, Geza Fogarasi, Frank Pang, and James E Boggs. Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives. *Journal of the American Chemical Society*, 101 (10):2550–2560, 1979.

[121] Jerzy Cioslowski and Stacey T Mixon. Covalent bond orders in the topological theory of atoms in molecules. *Journal of the American Chemical Society*, 113 (11):4142–4145, 1991.

[122] Pawel M Kozlowski, Andrzej A Jarzecki, and Peter Pulay. Vibrational assignment and definite harmonic force field for porphine. 1. scaled quantum mechanical results and comparison with empirical force field. *The Journal of Physical Chemistry*, 100(17):7007–7013, 1996.

[123] Peter Politzer and Shoba Ranganathan. Bond-order-bond-energy correlations. *Chemical physics letters*, 124(6):527–530, 1986.

[124] William J Taylor and Kenneth S Pitzer. Vibrational frequencies of semirigid molecules: A general method and values for ethylbenzene. *J. Res. Nat. Bur. Standards*, 38(1), 1947.

[125] Jörg Grunenberg and Norman Goldberg. How strong is the gallium - gallium triple bond? theoretical compliance matrices as a probe for intrinsic bond strengths. *Journal of the American Chemical Society*, 122(25):6045–6047, 2000.

[126] Mu-Yong Kim, Hariprasad Vankayalapati, Kazuo Shin-ya, Konstanty Wierzba, and Laurence H Hurley. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *Journal of the American Chemical Society*, 124(10):2098–2099, 2002.

[127] Erin Shammel Baker, Jeong Tae Lee, Jonathan L Sessler, and Michael T Bowers. Cyclo [n] pyrroles: size and site-specific binding to g-quadruplexes. *Journal of the American Chemical Society*, 128(8):2641–2648, 2006.

[128] David Monchaud and Marie-Paule Teulade-Fichou. A hitchhiker's guide to g-quadruplex ligands. *Organic & biomolecular chemistry*, 6(4):627–636, 2008.

[129] Marialuisa Micco, Gavin W Collie, Aaron G Dale, Stephan A Ohnmacht, Ingrida Pazitna, Mekala Gunaratnam, Anthony P Reszka, and Stephen Neidle. Structure-based design and evaluation of naphthalene diimide g-quadruplex ligands as telomere targeting agents in pancreatic cancer cells. *Journal of medicinal chemistry*, 56(7):2959–2974, 2013.

[130] Gavin W Collie, Rossella Promontorio, Sonja M Hampel, Marialuisa Micco, Stephen Neidle, and Gary N Parkinson. Structural basis for telomeric g-quadruplex targeting by naphthalene diimide ligands. *Journal of the American Chemical Society*, 134(5):2723–2731, 2012.

[131] Sonja M Hampel, Assitan Sidibe, Mekala Gunaratnam, Jean-François Riou, and Stephen Neidle. Tetrasubstituted naphthalene diimide ligands with selectivity for telomeric g-quadruplexes and cancer cells. *Bioorganic & medicinal chemistry letters*, 20(22):6459–6463, 2010.

[132] Gary Prato, Samantha Silvent, Sammy Saka, Massimiliano Lamberto, and Dmytro Kosenkov. Thermodynamics of binding of di-and tetrasubstituted naphthalene diimide ligands to dna g-quadruplex. *The Journal of Physical Chemistry B*, 119(8):3335–3347, 2015.

[133] Shengshuang Zhu, Michael F Brown, and Scott E Feller. Retinal conformation governs p k a of protonated schiff base in rhodopsin activation. *Journal of the American Chemical Society*, 135(25):9391–9398, 2013.

[134] Daniel P Oehme, Robert TC Brownlee, and David JD Wilson. Effect of atomic charge, solvation, entropy, and ligand protonation state on mm-pb (gb) sa binding energies of hiv protease. *Journal of computational chemistry*, 33(32):2566–2580, 2012.

[135] Farhad Khalili, Amr Henni, and Allan LL East. p k a values of some piperazines at (298, 303, 313, and 323) k. *Journal of Chemical & Engineering Data*, 54(10):2914–2917, 2009.

[136] Stefano Piana, Kresten Lindorff-Larsen, and David E Shaw. How robust are protein folding simulations with respect to force field parameterization? *Biophysical journal*, 100(9):L47–L49, 2011.

[137] Yevgen P Yurenko, Jan Novotnỳ, Vladimir Sklenář, and Radek Marek. Exploring non-covalent interactions in guanine-and xanthine-based model dna quadruplex structures: a comprehensive quantum chemical approach. *Physical Chemistry Chemical Physics*, 16(5):2072–2084, 2014.

[138] Georgios Markopoulos and Jörg Grunenberg. Predicting kinetically unstable c c bonds from the ground-state properties of a molecule. *Angewandte Chemie International Edition*, 52(40):10648–10651, 2013.

[139] Célia Fonseca Guerra, Hester Zijlstra, Gábor Paragi, and F Matthias Bickelhaupt. Telomere structure and stability: Covalency in hydrogen bonds, not resonance assistance, causes cooperativity in guanine quartets. *Chemistry-A European Journal*, 17(45):12612–12622, 2011.

[140] Anthony K Rappe and William A Goddard III. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry*, 95(8): 3358–3363, 1991.

[141] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[142] Jay L Banks, Hege S Beard, Yixiang Cao, Art E Cho, Wolfgang Damm, Ramy Farid, Anthony K Felts, Thomas A Halgren, Daniel T Mainz, Jon R Maple, et al. Integrated modeling program, applied chemical theory (impact). *Journal of computational chemistry*, 26(16):1752–1780, 2005.

[143] Winnfried Hasel, Thomas F Hendrickson, and W Clark Still. A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Computer Methodology*, 1(2):103–116, 1988.

[144] Maestro, version 9.2, schrödinger, llc, new york, ny, 2011.

[145] Macromodel, version 9.9, schrödinger, llc, new york, ny, 2011.

[146] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G Scalmani, V Barone, B Mennucci, GA Petersson, et al. Gaussian software, version 09 revision d01. *Gaussian Inc.: Wallingford, CT, USA*, 2009.

[147] Kai Brandhorst and Jörg Grunenberg. How strong is it? the interpretation of force and compliance constants as bond strength descriptors. *Chemical Society Reviews*, 37(8):1558–1567, 2008.

[148] Kai Brandhorst and Jörg Grunenberg. Efficient computation of compliance matrices in redundant internal coordinates from cartesian hessians for nonstationary points. *The Journal of Chemical Physics*, 132(18):184101, 2010.

[149] B Johnson Alberts and A Johnson. Lewi, s j.; raff, m.; roberts, k.; walter, p. *Molecular biology of the cell (Garland Science, Taylor and Francis Group, New York, 2002)*, 2002.

[150] Robert Hänsel, Frank Löhr, Silvie Foldynová-Trantírková, Ernst Bamberg, Lukáš Trantírek, and Volker Dötsch. The parallel g-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic acids research*, page gkr174, 2011.

[151] Wei Li, Peng Wu, Tastuo Ohmichi, and Naoki Sugimoto. Characterization and thermodynamic properties of quadruplex/duplex competition. *FEBS letters*, 526(1):77–81, 2002.

[152] Gary N Parkinson, Michael PH Lee, and Stephen Neidle. Crystal structure of parallel quadruplexes from human telomeric dna. *Nature*, 417(6891):876–880, 2002.

[153] Attila Ambrus, Ding Chen, Jixun Dai, Tiffanie Bialis, Roger A Jones, and Danzhou Yang. Human telomeric sequence forms a hybrid-type intramolecular g-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic acids research*, 34(9):2723–2735, 2006.

[154] Akimasa Matsugami, Yan Xu, Yuuki Noguchi, Hiroshi Sugiyama, and Masato Katahira. Structure of a human telomeric dna sequence stabilized by 8-bromoguanosine substitutions, as determined by nmr in a k+ solution. *FEBS Journal*, 274(14):3545–3556, 2007.

[155] Kim Ngoc Luu, Anh Tuân Phan, Vitaly Kuryavyi, Laurent Lacroix, and Dinshaw J Patel. Structure of the human telomere in k+ solution: an intramolecular (3+ 1) g-quadruplex scaffold. *Journal of the American Chemical Society*, 128(30):9963–9970, 2006.

[156] Jing Li, John J Correia, Lei Wang, John O Trent, and Jonathan B Chaires. Not so crystal clear: the structure of the human telomere g-quadruplex in solution differs from that present in a crystal. *Nucleic acids research*, 33(14): 4649–4659, 2005.

[157] Jixun Dai, Megan Carver, Chandanamali Punchihewa, Roger A Jones, and Danzhou Yang. Structure of the hybrid-2 type intramolecular human telomeric g-quadruplex in k+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic acids research*, 35(15):4927–4940, 2007.

[158] Steven B Zimmerman and Stefan O Trach. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of escherichia coli. *Journal of molecular biology*, 222(3):599–620, 1991.

[159] Yong Xue, Zhong-yuan Kan, Quan Wang, Yuan Yao, Jiang Liu, Yu-hua Hao, and Zheng Tan. Human telomeric dna forms parallel-stranded intramolecular g-quadruplex in k+ solution under molecular crowding condition. *Journal of the American Chemical Society*, 129(36):11185–11191, 2007.

[160] VI Prokhorenko, DB Steensgaard, and AR Holzwarth. Exciton theory for supramolecular chlorosomal aggregates: 1. aggregate size dependence of the linear spectra. *Biophysical journal*, 85(5):3173–3186, 2003.

[161] Frank C Spano, Stefan CJ Meskers, Emanuelle Hennebicq, and David Beljonne. Probing excitation delocalization in supramolecular chiral stacks by means of circularly polarized light: experiment and modeling. *Journal of the American Chemical Society*, 129(22):7044–7054, 2007.

[162] Sandro Jurinovich, Gennaro Pescitelli, Lorenzo Di Bari, and Benedetta Mennucci. A tddft/mmpol/pcm model for the simulation of exciton-coupled circular dichroism spectra. *Physical Chemistry Chemical Physics*, 16(31):16407–16418, 2014.

[163] Lucas Viani, Marina Corbella, Carles Curutchet, Edward J O'Reilly, Alexandra Olaya-Castro, and Benedetta Mennucci. Molecular basis of the exciton–phonon interactions in the pe545 light-harvesting complex. *Physical Chemistry Chemical Physics*, 16(30):16302–16311, 2014.

[164] Aurora Munoz-Losa, Sinisa Vukovic, Stefano Corni, and Benedetta Mennucci. Nonplasmonic metal particles as excitation energy transfer acceptors: an unexpected efficiency revealed by quantum mechanics. *The Journal of Physical Chemistry C*, 113(37):16364–16370, 2009.

[165] Hugo Gattuso, Xavier Assfeld, and Antonio Monari. Modeling dna electronic circular dichroism by qm/mm methods and frenkel hamiltonian. *Theoretical Chemistry Accounts*, 134(3):1–8, 2015.

[166] Bengt Nordén. *Circular dichroism and linear dichroism*, volume 1. Oxford University Press, 1997.

[167] Jia Zhou, Aaron M Fleming, April M Averill, Cynthia J Burrows, and Susan S Wallace. The neil glycosylases remove oxidized guanine lesions from telomeric and promoter quadruplex dna structures. *Nucleic acids research*, page gkv252, 2015.

[168] Christopher Jacques Lech, Brahim Heddi, and Anh Tuân Phan. Guanine base stacking in g-quadruplex nucleic acids. *Nucleic acids research*, 41(3): 2034–2046, 2013.

[169] Donald M Gray, Jin-Der Wen, Carla W Gray, Rudolf Repges, Charlotte Repges, Gerhard Raabe, and Jörg Fleischhauer. Measured and calculated cd spectra of g-quartets stacked with the same or opposite polarities. *Chirality*, 20(3-4):431–440, 2008.

[170] DA Case, V Babin, Josh Berryman, RM Betz, Q Cai, DS Cerutti, TE Cheatham Iii, TA Darden, RE Duke, H Gohlke, et al. Amber 14. 2014.

[171] Antonio Monari, Jean-Louis Rivail, and Xavier Assfeld. Theoretical modeling of large molecular systems. advances in the local self consistent field method for mixed quantum mechanics/molecular mechanics calculations. *Accounts of chemical research*, 46(2):596–603, 2012.

[172] RBJS Krishnan, J Stephen Binkley, Rolf Seeger, and John A Pople. Self-consistent molecular orbital methods. xx. a basis set for correlated wave functions. *The Journal of Chemical Physics*, 72(1):650–654, 1980.