Computational study of DNA in non-canonical environment

Annalisa Arcella

UNIVERSITAT DE BARCELONA

Computational study of DNA in non-canonical environment

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the
Faculty of Physics
Physics PhD program
Department of Fundamental Physics

Annalisa Arcella

August 2014

University of Barcelona

Physics PhD Program

Institute for Research in Biomedicine & Barcelona Supercomputing Center

Molecular Modelling and Bionformatics Group

INSTITUTE FOR RESEARCH IN BIOMEDICINE

Computational study of DNA in non-canonical environment

Author:
Annalisa Arcella

Supervisor:
Prof.Dr. Modesto Orozco

August 2014

"What get you into trouble
is not what we don't know,
It's what we know for sure
that just ain't so."

Mark Twain

Abstract

During my PhD thesis I used theoretical techniques, in particular Molecular Dynamics (MD) to study the structural properties of nucleic acids in non-canonical environment, especially in the gas phase and apolar conditions.

The highly charged nature of the nucleic acids backbone clearly suggests that the environment plays a key role in the behavior of these molecules. The water is the natural environment for DNA and is excellent to stabilize its structure, but it is not the ideal solvent for favoring specific recognition, certain reactions or physical processes such as charge transfer. Interest exist then to explore the nature of nucleic acids, particularly DNA, in non-aqueous solvents, where the universe of nucleic acids applications will expand even more.

In the first part of this thesis I used atomistic molecular dynamics simulations to investigate the structural and thermodynamics changes of a DNA hairpin when transferred from an aqueous solution to a low dielectric media, carbon tetrachloride (CTC), under different DNA charge states. I simulated the pulling of a short DNA hairpin from a water compartment through a CTC slap and estimated the free energy related to the transfer of the DNA from water to CTC through atomistic Umbrella Sampling simulation.

The second part of my thesis is centered in the challenge of the most recent experimental techniques, such as Mass Spectrometry and X-Ray Free Electron Laser which use gas phase ions to provide structural information of macromolecules. They are fast and require low sample consumption but the question is to what extent does gas phase structural information reflect the most populated conformation in solution.

A series of both experimental and theoretical studies with proteins have demonstrated that gas phase ensembles can be used to accurately model the solution structure. The question is then, whether or not these findings also stand for a highly flexible and charged non-globular molecule as DNA, depends on the solvent environment.

I combined extended classical and quantum MD simulations, validating the results by means of mass spectroscopy experiments, to reveal a picture of unprecedented quality on the nature of DNA in the gas phase.

Acknowledgements

There are many people I would like to thank for accompanying me during these years.

At first I thank all people in the lab., especially my supervisor Modesto that gave me the opportunity to join the group.

Thanks to all the guys that received me especially Marco, Andreas, Alberto and Tim .

Jose for several coffee breaks and funny, interesting, unusual and friendly conversations, and for his prompt support every time there was any technical/apocalyptical problem with 'Pluto and his *cluster* of friends'.

Adam for always amaze me with his tricks with perl.

Nacho to be always professional both in the joke and in little helps I asked him for.

Thanks to my friend Laura who welcome me into her vegetables garden to start one of the most important friendships of my Catalan landmark, and more...

Finally I would like to thank some special friends that represented my 'spanish family' especially Ginevra, Laura, Chiara, Cristina and Mario, my fantastic flatmates Jacynthe, Magui, Meri and Putxa and some people no living more in Barcelona, Davide and Christian.

Contents

List of Figures

List of Tables

To my parents...

Chapter 1

Introduction

In the 1944, Avery, MacLeod, and McCarty discovered that a nucleic acid, named DNA, was the transforming agent responsible for heredity (Avery 1944), a theory which was later confirmed in 1952 by Alfred Hershey and Martha Chase who established the role of DNA as the carrier of hereditary information (Hershey 1952). Scientists like Erwin Chargaff, who showed correspondence in the content of adenine and thymine, and of guanine and cytosine in DNA, or Rosalind Franklin, who performed X-ray diffraction experiment of DNA crystals, in 1953 lay the foundation for James Watson and Francis Crick who published the structure of DNA (Watson 1953). Advances in the field allowed the sequencing of fragments of DNA (Sanger 1977), whose culmination is the current genome sequencing projects. The first sequenced genome of DNA was published in 1977 (Sanger 1977) and the human genome did not arrive until 2004 (Lander, Linton et al. 2001; International Human Genome Sequencing 2004).

1.1 Structural properties of nucleic acids

The building blocks of nucleic acid are constituted by monomeric units called nucleotides. A nucleotide consists of three chemically and structurally distinct components: a base, a sugar, and a phosphate group (Figure 1.1). The phosphate groups give nucleic acids their acidic properties, as they are fully ionized at the physiological pH. There are two natural sugars in nucleic acids. Both are cyclic pentose β-D-deoxyribose in DNA and the closely related β-D-ribose in RNA. The bases are nitrogenous aromatic rings with lipophilic flat faces, and feature several hydrogen bond donors and acceptors along their edges.

Structurally, in natural nucleic acids, the nucleotide building block is first composed of a nucleoside subunit, formed when a base and sugar are linked by a C-N

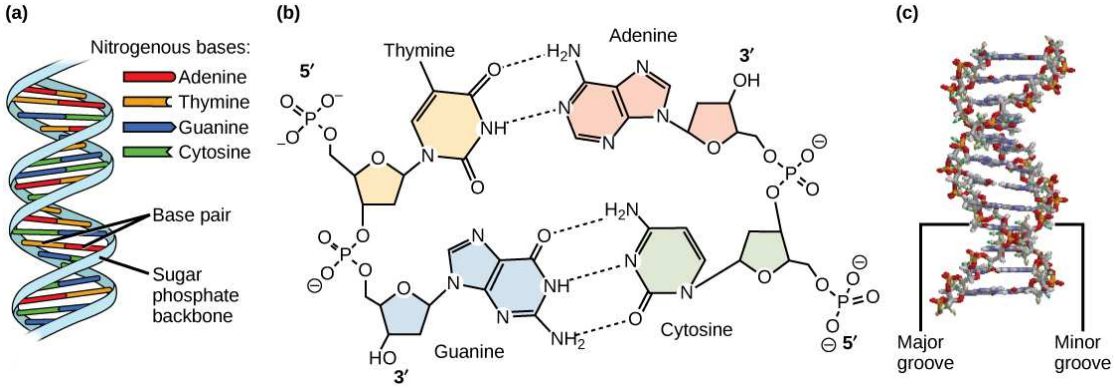

FIGURE 1.1: The structure of the DNA

glycosidic bond in the β stereochemistry. These nucleosides are then linked through phosphate groups that are attached to the 3 carbon of one nucleotide and the 5 carbon of the other, so that the full repeating unit in a nucleic acid is a 3,5-nucleotide.

Chemical analyses of cellular genetic material showed that there are four types of nucleotides, which only differ by the attached nitrogenous base. Two nucleotides, adenosine-5-phosphate (A) and guanosine-5-phosphate (G), contain fused-ring purines. Cytosine-5-phosphate (C) and thymine-5- phosphate (T) are single-ring pyrimidines. Thymine is replaced by its demethylated form, uracil, in RNA.

Early on, chemical analysis of DNA from genetic material showed that the molar ratio of guanine/cytosine and adenine/thymine are close to one (Zamenhof 1952). This information was critical for the determination of the DNA double helical structure, as Watson and Crick recognized that two pairs of bases could hydrogen bond and form an antiparallel double helix.

Nucleic acids are polar molecules and their polarity is derived from the oxygen and nitrogen atoms in the backbone. They are formed when nucleotides come together through phosphodiester linkages between the 5' and 3' carbon atoms (Anthony-Cahill 2012).

Nucleic acid sequence is the order of nucleotides within a DNA (GACT) or RNA (GACU) molecule that is determined by a series of letters. Sequences are presented from the 5' to 3' end and determine the covalent structure of the entire molecule. They can be complementary to another one in that the base on each position is complementary as well as in the reverse order. An example of a complementary sequence to AGCT is TCGA.

In contrast to DNA, RNA strands can perform a variety of functions. Besides carrying genetic information and regulation of gene activity, RNA molecules can even act as catalyzers, for example in the ribosome, where RNA catalyzes the formation of

peptide bonds (Cech 1993). The discovery of messenger and transfer RNAs came to clarify how DNA information is converted into functional proteins. These advances in molecular biology constituted the *Central Dogma* which shows the general pathway of information in three steps (Figure 1.2).

First, the replication of DNA assures an identical copy of DNA molecules in the next generation from parental DNA. Second, the transcription process by which some parts of the DNA are copied into RNA. And the third step which is called translation process in which messenger RNA molecules are translated into proteins in the ribosomes.

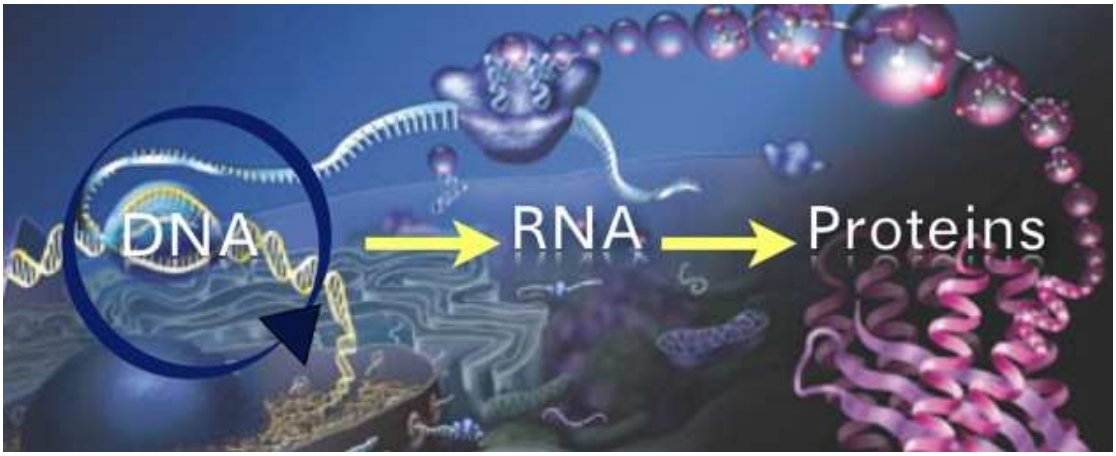

FIGURE 1.2: The Central Dogma of Molecular Biology.

1.1.1 Nucleic Acids as a nanomaterial

DNA (and related RNA) are not only a master pieces of chemical engineering that carry out the genetic information of an entire individual, but also molecules with incredible biotechnological applications related to their nanoscale geometry, biocompatibility, biodegradability, and molecular recognition properties; investigations into DNA and RNA structure and self-assembly (Yang 2010).

Since the eighties researchers from all over the world use DNA fragments to construct circuits, program instructions and assemble robots million times smaller than a grain of sand. Nature has however find ways to store even more efficiently information using a simple code of four letters, that is robust to external stress, self-replicative, and flexible enough to be adapted to cellular needs.

Scientists have used DNA in this way to construct various amazing nanostructures, such as ordered lattices, origami, supramolecular assemblies, and even three-dimensional objects. In addition, DNA has been utilized as a guide and template to direct the assembly of other nano-materials including nanowires, free standing membranes, and crystals.

Overall, these novel DNA materials have found applications not only in nanotechnology in general but also in various areas of biomedical field, nano-medicine in particular. Patterning DNA into one-, two-, and three-dimensional assemblies holds tremendous promise for biological applications, from encapsulation and release of molecules for drug delivery, to new substrates for tissue engineering, biological sensors, and selective therapeutics.

1.2 Canonical structure of the DNA

DNA is most commonly found as a B-form double helix (Figure 1.1B), having a 2 nm diameter and a 3.4 nm or 10.5 base pair pitch (height of a complete helix turn) (Watson 1953), with its complementary strand as a double helix; this structure is driven by supramolecular non-covalent interactions, such as hydrogen bonding and π-stacking between complementary base pairs (adenine (A) with thymine (T) and cytosine (C) with guanine (G)), hydrophobic effects, and electrostatic repulsion between the backbone phosphate units (Figure 1.1).

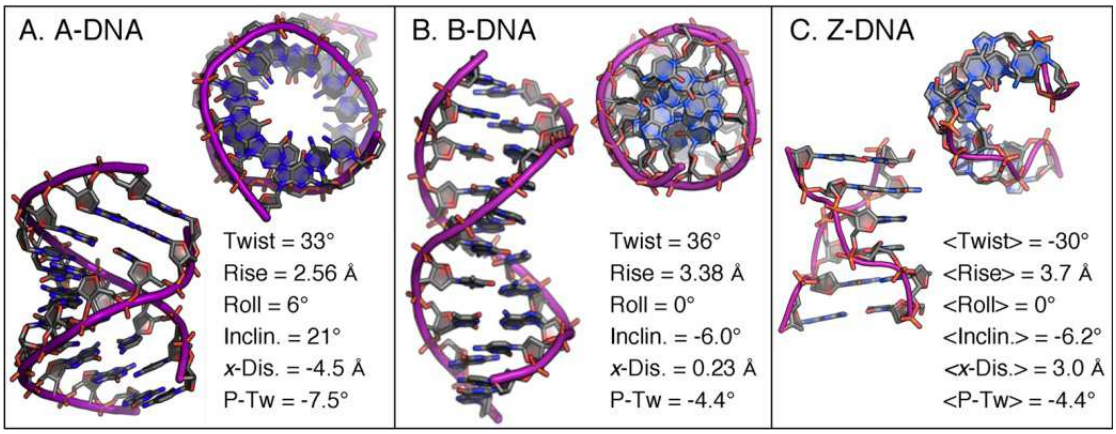

FIGURE 1.3: Canonical duplex DNA: A: A-DNA; B: B-DNA; C: Z-DNA.

Hydrogen bonds are crucial for DNA stability, for replication, and are also determinant in modulating biologically relevant interactions of DNA with proteins (Figure 1.4). One of the first features that were discovered from DNA sequence analysis was the identical quantities of A and T, and G and C nucleobases, something that Chargaff's experimental data (Chargaff 1950) pointed to be the key for mutual recognition and on which DNA replication and translation processes are based.

All nitrogen bases have several hydrogen bond acceptors and donors which allow several edges for interactions. While the Watson-Crick base pairing (WC) is the main interaction pattern between neighboring strands, there have been identified other kinds of interaction like wobble pairs (w) that allowed Francis Crick to explain the degeneracy

of the genetic code, and Hoogsteen pairs (Hoogsteen 1963) (H) that it is thought to play an important role in protein-DNA recognition (Nikolova, Gottardo et al. 2012).

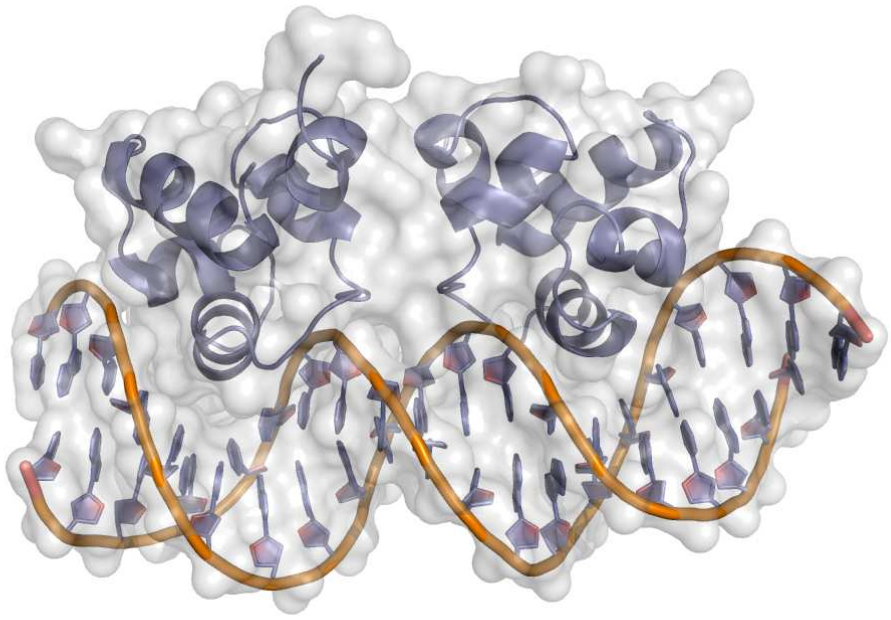

FIGURE 1.4: Example of DNA-protein docking interaction.

Double stranded helices are favored not only by hydrogen bonds but even more important, by **stacking interaction**. While hydrogen bonds have an electrostatic nature, in general, stacking interactions are mainly van der Waals interactions due to the π orbital overlap between aromatic rings (i.e., dispersive term), which are complemented by entropy-related hydrophobic effects.

Base stacking interactions have been shown to facilitate local conformational variability in DNA (Calladine 1982; Hunter 1993) probably through changes in the sugar-phosphate backbone which are known to be crucial in several protein-DNA interactions.

Different experimental techniques, such as osmometry, ultracentrifugation, calorimetry, denaturalization, NMR and others, have been widely used for thermodynamic description of this kind of interactions (Solie and Schellman 1968; POP 1974; Tribolet and Sigel 1987; Guckian, Schweitzer et al. 1996).

All these studies have shown that, stacking involved by guanine are the most stable one, while those involving thymine are the least stable ones. In general we can define the following order of stability for stacking interactions: purine-purine > purine-pyrimidine > pyrimidine-pyrimidine.

1.3 Non-canonical structure of nucleic acids

Other duplex shapes such as A- and Z-forms (Figures 1.3A, 1.3C) are observed depending on the surrounding environment. Nucleic acids are flexible and polymorphic molecules (Saenger 1984; Blackburn 1990; Sinden 1994; Bloomfield 2000) and unmodified DNA strands are also known to assemble into guanine quadruplexes (Figure 1.5C), cytosine-based, i-motifs (Figure 1.5D), and triple helices (Figure 1.7) through non standard hydrogen bonding or Hoogsteen base pairing (Bhatia 2011) (Figure 1.6). In this thesis I will consider triplex-stranded DNA and DNA hairpins that I am going to describe shortly.

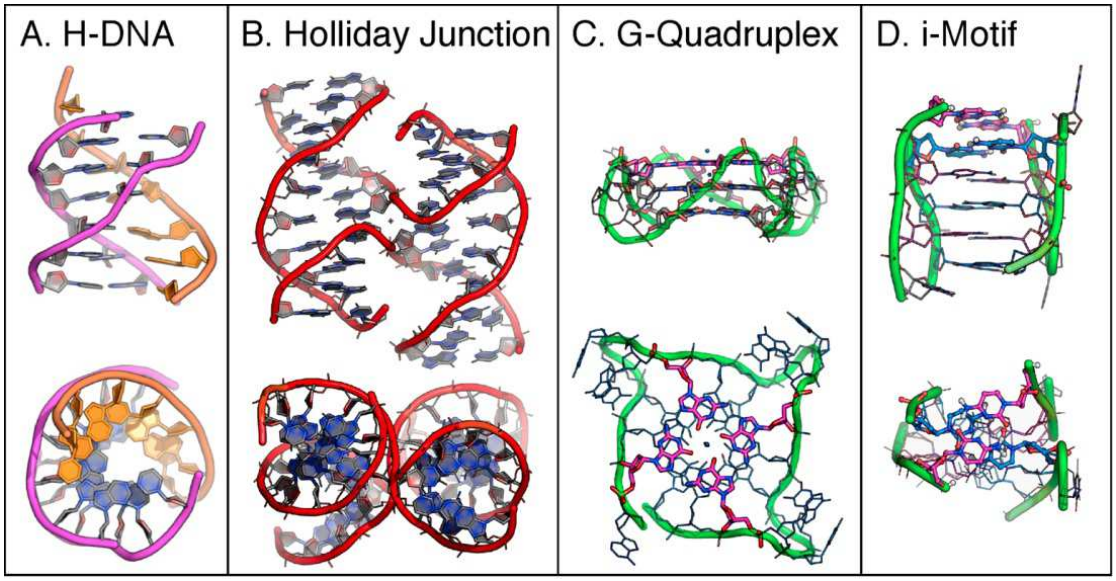

FIGURE 1.5: Non-canonical DNA: A: H-DNA; B: Holiday Junction; C: G-Quadruplex; D: i-motifs.

Triple-stranded DNA

Single-stranded triplex-forming oligonucleotide (TFO) can bind to the major groove of a double stranded DNA, generally to homopurine strands (Strobel and Dervan 1990). Their high affinity and specificity lead to the formation of DNA triplexes being involved in genomic events including inhibition of transcription and DNA replication, promotion of site-specific DNA damage, and enhancement of recombination (Davis 2010; McNeer, Schleifman et al. 2011). This unique ability of TFOs to recognize duplexes forming triplexes has been exploited in several biotechnological applications such as antigene strategies or for site-specific delivery of mutagenic agents (Chin, Schleifman et al. 2007; Vasquez 2010).

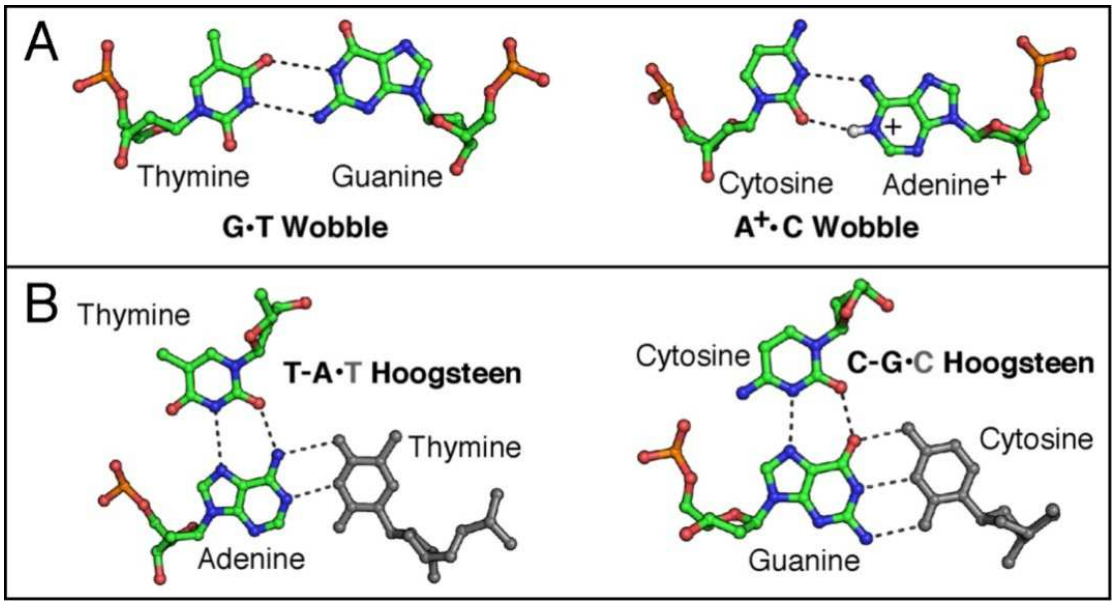

FIGURE 1.6: Non-canonical hydrogen bonds.

Triplexes can be formed by inter- or intra-molecular interaction with short oligonucleotides. Intra-molecular triple helix requires a single strand with appropriate sequence to fold back on itself driven by supercoiling (Htun and Dahlberg 1989). In general, triplexes are classified as either parallel or antiparallel depending on the orientation of the TFO relative to the interacting homopurine strand of the double helix.

Parallel triplexes present an omopyrimidine strand with either thymines or N3-protonated cytosines forming Hoogsteen base pairings (H) to bind to the dsDNA adopting $C^+ \cdot G\_C$ or $T \cdot A\_T$ pairings (where '_' denotes WC pairing and '·' represents Hoogsteen base pairing) (Figure 1.6). This requirement makes parallel triplexes extremely pH-dependent and rapidly unstable when pH rises above 5.0. In addition, antiparallel triplexes place a homopurine strand as TFO by reverse Hoogsteen (rH) base pairing between neutral nucleotides, avoiding the pH dependence from parallel triplexes and forming $G \cdot G\_C$ or $A \cdot A\_T$ pairings (where '·' stands for reverse Hoogsteen pairing) (Figure 1.6B).

Not only homopurine or homopyrimidine TFO sequences do not exclusively form triple helices but alternative GT-rich TFOs can also undergo both parallel and antiparallel orientations depending on their sequence, although these have been less studied and seems less stable (Thuong 1993).

From the thermodynamically point of view, triple helices are in general less stable than corresponding duplexes. However triplex stability can be enhanced by using modified nucleotides like 5-methylcytosine instead of cytosine or 5-bromouracil instead

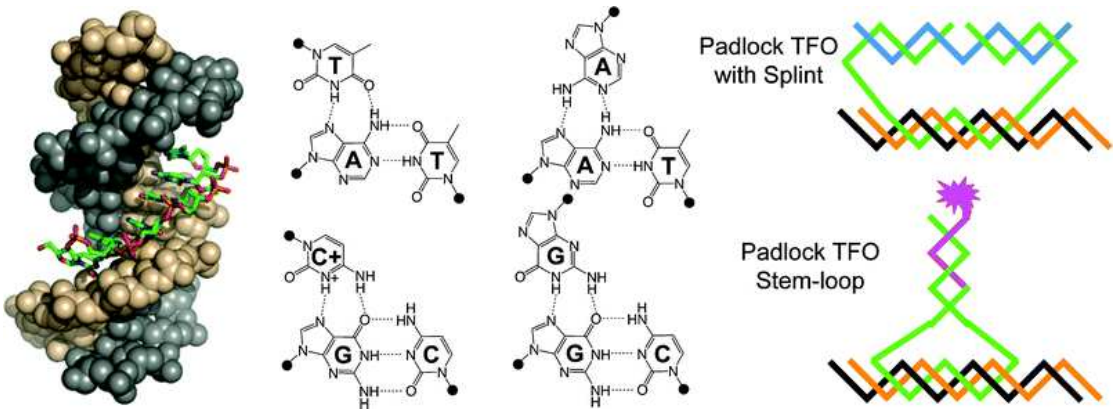

FIGURE 1.7: Triplex DNA major-groove binder.

of thymine and even more stabilizing is the use of ribonucleotides in the TFO (Roberts and Crothers 1992).

Peptide nucleic acids (PNA) sequences have been also shown to stabilize the complex with double stranded DNA molecules (Nielsen, Egholm et al. 1991; Betts, Josey et al. 1995)

DNA hairpin

Hairpin structures, an example shown in Figure 1.8, can be formed by sequences with inverted repeats (IRs), also termed palindromes, following two main mechanisms.

First, in several cellular processes, DNA is single stranded (single-stranded DNA [ssDNA]), for instance, during replication on the template for lagging-strand synthesis, during DNA repair, or, more importantly, during rolling-circle replication (RCR), bacterial conjugation, natural transformation, and infection by some viruses. ssDNA is not simply a transient inert state of DNA but can fold into secondary structures recognized by proteins, notably involved in site-specific recombination, transcription, and replication.

A second mechanism is the formation of hairpins from double-stranded DNA (dsDNA) as a cruciform, i.e., two opposite hairpins extruding through intrastrand base pairing from a palindromic sequence. The existence of cruciforms was already hypothesized soon after Watson and Crick's discovery (Platt 1955): the negative supercoiling of dsDNA could provide free energy to stabilize cruciform. Cruciforms then attracted much attention in the 1980s, when their existence was experimentally assessed in vitro under natural superhelical densities (Panayotatos and Wells 1981).

However, most studies at that time rejected their possible implication in cellular processes because of the slow kinetics of cruciform formation, which made them theoretically very unlikely to occur in vivo (Courey and Wang 1983; Sinden, Broyles et al. 1983). Nonetheless, this point of view was revised when techniques revealing cruciforms in vivo were developed and biological functions involving DNA secondary structures were discovered.

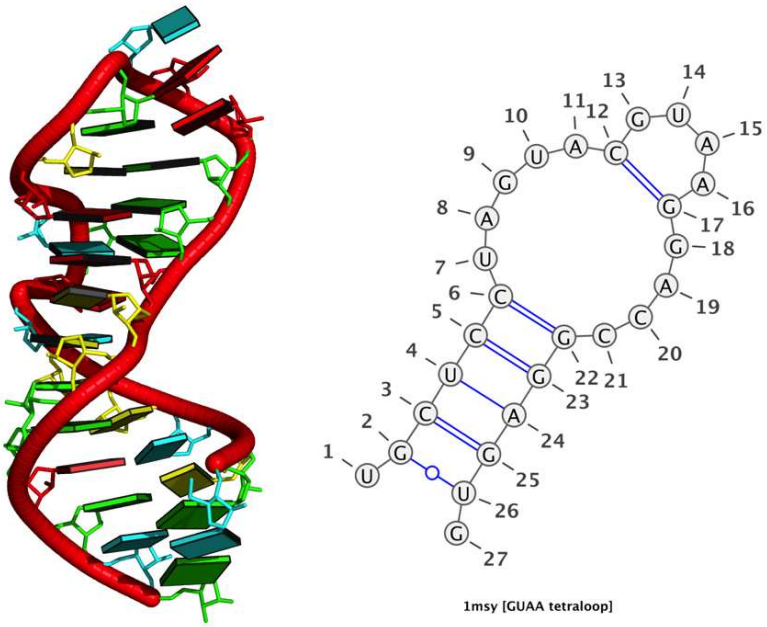

FIGURE 1.8: Hairpin 1msy [GUAA tetraloop].

DNA hairpins can interact with proteins in three different ways and impact cell physiology: (i) cruciform formation modifies the coiling state of DNA (Wang and Lynch 1993), which is known to affect the binding of regulatory proteins for transcription, recombination, and replication (Wing, Drew et al. 1980; Hatfield and Benham 2002); (ii) the DNA-protein interaction can be inhibited if a hairpin overlaps a protein recognition site (Horwitz and Loeb 1988); and (iii) proteins can directly recognize and bind DNA hairpins (Masai and Arai 1997; Val 2005 ; MacDonald, Demarre et al. 2006; Gonzalez-Perez, Lucas et al. 2007; Barabas, Ronning et al. 2008). It thus seems quite easy both for proteins to evolve hairpin-binding activity and for hairpins to evolve in such a way that they can exploit host proteins.

Hairpin recognition can be seen as a way for living systems to expand the repertoire of information storage in DNA beyond the primary base sequence.

1.4 The role of the solvent

Under physiological condition the DNA is found in aqueous solution. The solvation is directly related to the conformation of nucleic acids and plays a key role in the structure of nucleic acids. The most important function is the screening of the phosphate groups, whose electrostatic repulsion would lead to strand separation (Rueda, Kalko et al. 2003).

Water molecules are structured around the DNA forming two hydration shells. An average of around 11/20 molecules per nucleotides are situated in the inner shell that bind to the atoms of the sugar-phosphates through hydrogen bonds. The narrow minor groove usually contains the so-called spine of hydration, especially in AT base pares, present as constituent elements in different nucleic acids (Drew and Dickerson 1981).

The second shell is located around the first group of water (Tereshko 1999). It is less structured and many times is indiscernible from the rest of solvent with respect to the permeability of ions and mobility of water, even though is crucial for the stabilization of the DNA structure.

A series of systematic studies of hydration of DNA double helices has revealed that water molecules are well organized around both bases (Schneider 1993; Schneider and Berman 1995) and phosphates (Sharon 2007), and that the observed hydration patterns depend on the DNA conformational type as well as base identity. These hydration sites represent sites of preferred polar binding to DNA.

For instance, water and the polar groups of drugs and amino acids bind to similar sites in the B-DNA minor groove (Moravek, Neidle et al. 2002). Calculated hydration sites can also be used to predict polar binding of amino acids to DNA in protein/DNA complexes (Woda, Schneider et al. 1998).

A methodologically similar study has shown that metal cations such as Na^{+}, Mg^{++}, Ca^{++} or Zn^{++} also have preferred binding sites around phosphate groups, and that these sites are characteristic for each metal (Schneider 1998).

1.5 DNA in apolar environment

To understand the dependence of DNA structure and properties with solvent we need to analyze the nature of DNA interaction landscape. DNA is a heavily charged polyanion. Negative charge density is so large that Coulombic repulsion should unfold DNA duplex.

However, under physiological conditions, Coulombic repulsion is dramatically reduced by the charge screening effect of water, yielding to a destabilizing interaction

small enough as to be compensated by a myriad of interactions. Very simple back-of-the-envelope type calculations indicate that DNA duplex, as we know, should be very unstable in the absence of a highly polar environment as water.

The impact of solvent in DNA structure is known fifties. For example it is well established that changes in ionic strength or water activity leads to changes in nucleic-acid structure and stability (Saenger 1984; Blackburn 1990; Sinden 1994; Bloomfield 2000). Already in the original works of Wilkins, Franking, Watson and Crick was clear that even the addition of certain co-solvents, such as ethanol, leads to a dramatic conformational change between B and A forms in DNA duplexes (Huey 1981), high ionic strength is coupled with the $B \rightarrow Z$ transition in d($G\backslash C$) rich DNA duplexes (Pohl and Jovin 1972), some divalent ions are very important to stabilize triplexes (Bernues, Beltran et al. 1990), and G-quadruplexes display a dramatic loss of stability as a consequence of the reduction in the monovalent cationic concentration (Williamson, Raghuraman et al. 1989; Wang and Patel 1993) (see below for the description of these structures). Even more impressive: moderate concentrations of some co-solvents of reduced polarity, such as phenol, pyridine, propanol or 1,4 Dioxane, lead to the complete unfolding of duplex DNA (Turner 2000), while in some cases can stabilize other forms of DNA (Miyoshi 2009).

Recent work by Cui et al. (Cui S. 2007) showed that dsDNA splits apart when dragged from water to apolar solvent, such as octane, suggesting that the principle of spontaneous unwinding/splitting of dsDNA by providing a low-polarity (in other word, hydrophobic) micro-environment is exploited as one of the catalysis mechanisms of helicases. In other recent studies, Hud and coworkers have recently demonstrated that three DNA and RNA structures (i.e., duplex, triplex and G-quadruplex structures) can exist in a water-free solvent that is comprised of urea and choline chloride (a so-called deep eutectic solvent) (Lannan, Mamajanov et al. 2012).

The demonstration that nucleic acid structures can be maintained in a deep eutectic solvent and a water-free ionic liquid supports the feasibility of investigating and using DNA for nanotechnological applications.

Current and future applications of nucleic acids are however hampered by the common use of water as nucleic acid solvent. Water is a fantastic solvent, but it is not the ideal one for favoring specific recognition, certain reactions or physical processes such as charge transfer. Interest exist then to explore the nature of nucleic acids, particularly DNA, in non-aqueous solvents, where the universe of nucleic acids applications will expand even more (Lin J. 2008).

For example, a system such as the amphipatic DNA 6-helix bundle (Mathieu 2005) could function as a large pore and, with a hydrophobic or aromatic section on the inside, could bind to carbon nanotubes. In the same manner as helical peptides, DNA nanostructures with half-hydrophobic surfaces might be able to self-associate into larger well-defined units.

Furthermore DNA nanostructures with hydrophobic regions on the outside could be inserted into the hydrophobic patch of membrane cell in gene-delivery techniques. For example, non–viral vectors such as functionalized cationic lipids or polymers that can condense DNA in a non-aqueous media (Strobel and Dervan 1990; Feng Y. 1999; Pereira G.G. 2001; Montesi A. 2004), show great advantages over viruses in terms of safety and low immunogenicity (Brown 2001). However, the transfection efficiency of non-viral vectors is lower than achievable with viruses, which is likely to be related to the nature of the structure formed by the DNA interaction with cationic lipids or polymers (Koltover 1998; Ewert 2004; Ewert 2005). Knowledge of the dependence between solvent and structure will certainly favor our ability to desire more efficient nucleic acids carriers.

The first part of this thesis aims to elucidate the behavior of nucleic acids in apolar environment, in particular to estimate the free energy barrier to overcome polar/apolar boundary, i.e. water/lipid barrier useful for gene delivery application.

1.6 Nucleic Acids in the gas phase

The gas phase is an extreme case of non-aqueous solvent. Apparently, little interest exists in studying the behaviour of DNA or RNA in such exotic solvent. However, there is a rising interest in understanding how DNA is in the gas phase, due to extreme impact that experimental techniques working with samples in vacuum (such as mass spectroscopy) have in current research in biochemistry and molecular biology. The interest will increase even more if promises on X-Ray Free Electron Laser (XFEL) stands, and high resolution structures of nucleic acids in the gas phase are available.

1.6.1 X-ray Free Electron Laser

Current nucleic acids structure analysis techniques are limited to either crystallizable (X-ray Crystallography) or rather small soluble DNA (NMR) and one of the “grand challenges” of structural biology at the beginning of the 21th century is to successfully image a single, isolated biological molecule, without the need for crystalline sample. The realization of such an experiment is possible using the highly focussed and powerful

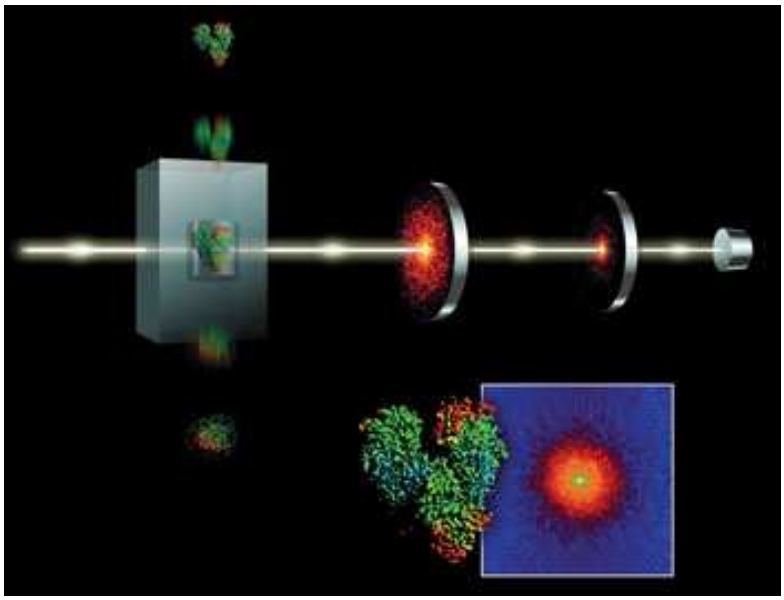

FIGURE 1.9: Experimental set-up for atomic resolution images based on single pulse diffraction patterns using the radiation of free-electron laser.

X-ray beam from an X-ray free electron laser (McCullough, Entwistle et al. 2009). The worlds first unprecedented XFEL super-microscope using hard x-ray will soon open to researchers worldwide at the Linac Coherent Light Source (LCLS) in Stanford (USA) and at the XFEL at Spring-8 (Japan). A European XFEL is under construction in Hamburg and will be operative in the 2015.

X-ray crystallography has long been used for molecular structure determination down to atomic resolution but requires the formation of well-ordered crystals. Unfortunately, not all interesting matter systems form crystals; in fact many of the most interesting biological systems are not amenable to crystallization, i.e. membrane proteins. For many classes of molecules, it is simply not possible to form regular crystals. Nor is a regular crystal lattice necessarily desirable for functional studies, because the reactions to be studied may be suppressed or altered by the crystalline lattice or may result in the breakdown of the crystal. Coherent X-ray imaging relaxes the requirement of crystalline periodicity, enabling the imaging of one single molecule at a time.

More recently, a new form of X-ray imaging based on diffraction from single isolated particles has been developed in which the concepts employed in X-ray crystallography are applied to noncrystalline objects. This technique bypasses the limitations of crystallization issues. Because of a combination of theoretical, experimental, and technological advances, it is now possible to have X-ray free-electron lasers.

They can revolutionize the way we study matter at the atomic and molecular level, allowing atomic resolution snapshots on the ultrafast timescale associated with the intrinsic atomic motions of atoms in matter. The new experiments need X-rays pulses

with a large number of photons focused on a sample as small as a molecule, squeezed into a time of a few femtoseconds, to study the dynamics of atomic and molecular processes.

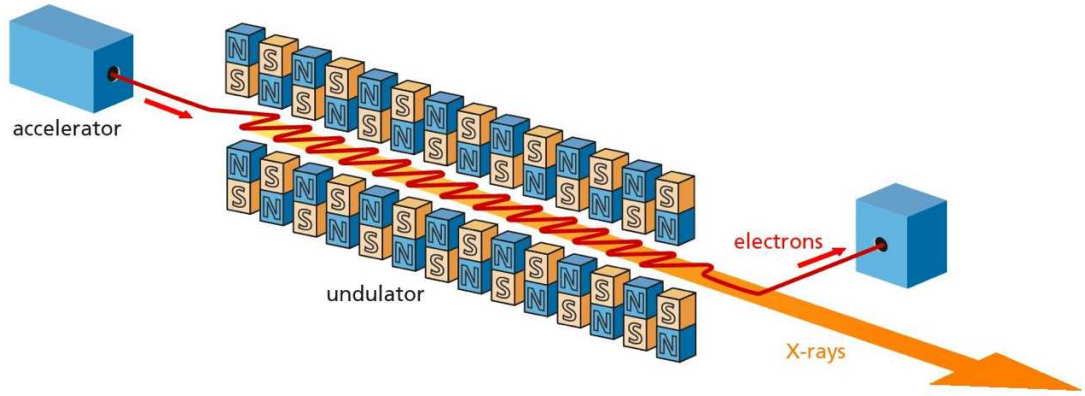

FIGURE 1.10: As at an undulator beamline of a 3rd generation synchrotron facility, the X-rays of an XFEL are generated by the oscillatory motion of relativistic electrons (typically 215 GeV) as they traverse the undulatora linear array of alternately-poled magnets. BOTTOM: Experimental set-up for atomic resolution images based on single pulse diffraction patterns using the radiation of free-electron laser.

Operative principle of the X-ray Free Electron Laser

A free-electron laser (FEL), is a type of laser that shares the same optical properties as conventional lasers such as emitting a beam consisting of coherent electromagnetic radiation that can reach high power, but that uses some very different operating principles to form the beam.

Unlike solid-state lasers in which electrons are excited in bound atomic or molecular states, free-electron lasers use a relativistic electron beam that moves freely through a magnetic structure that acts as an ondulator. As a result an electron beam is produced that emits a coherent electromagnetic wave-train (Figure 1.10). The free-electron laser has the widest frequency range of any laser type, and can be widely tunable, currently ranging in wavelength from microwaves, infrared, to the visible spectrum, ultraviolet, and X-ray (at a wevelenght of 1.5 Å).

The conceptual basis of coherent X-ray imaging is simple (Figure 1.9) (Anton Barty 2013). X-rays illuminate an object such as a molecule, and the diffracted (elastically scattered) photons are measured on a detector located some distance downstream. Instead of a crystal that produces Bragg spots on the detector, the single molecule produces a continuous diffraction pattern.

Gas phase ions as sample

Methods for characterization, selection and injection of single molecules into intense X-ray pulse are based on techniques used in electrospray ionization mass spectrometers (see below) (Neutze 2004). The sample is composed by gas phase ions produced using a soft ionization technique and immobilized on a membrane or in an ion-trap. In-vacuo techniques have the common advantage of having sensitivity which allows analysis of μg or even ng of sample as compared to mg required for crystallography or NMR.

Recent proof of principle experiments with a soft (25fs, 1013 W/cm2) (Chapman, Woods BW et al. 2006) XFEL flash at 60nm resolution have underlined the feasibility of the concept for proteins (Neutze 2004; Chapman, Woods BW et al. 2006; Barends, Foucar et al. 2014), and will be suitable for virus particles (Siuzdak, Bothner et al. 1996) and ribosomes (Rostom, Fucini et al. 2000) whose structures remain intact into the vacuum chamber of the mass spectrometer while maintaining the structural integrity of the samples at cryogenic temperatures.

1.6.2 Electrospray Ionization Mass Spectrometry

Most studies of DNA in the gas phase are derived from the use of Electrospray Ionization (ESI) mass spectroscopy (MS). ESI is a soft ionization technique extensively used for production of gas phase ions without fragmentation of thermally labile large supramolecules. The resulting ions bearing different net changes travel across the spectrometer tube until they arrive to a detector separated by their mass charge ratio, m/z values. Current MS-ESI protocols are mild enough to, not only, preserve covalent structures of macromolecules, but also non-covalent complexes between macromolecules. Recently drift tubes have been incorporated to mass spectrometers.

When travelling along a drift tube the macromolecule ions feel a flow of inert gas particles travelling in opposite directions. Collision of the macromolecular ions with the inert gas slow down the migration of the ion in a non-uniform manner: while compact structures will suffer few collisions, extended structures (even with the same mass) will suffer large number of collision and their migration along the tube will be much slower. Ions with the same mass/charge ratio can then be separated and drift times can be manipulated using simple physical models to gain information on the general shape of the macromolecule (in reality of all the macromolecular ions) in the gas phase. This information is condensed in the so named *Collision-Cross-Section* (CCS) (Wyttenbach 2013) (see below section 2.11.1)

A series of both experimental and theoretical studies with proteins (Fenn 1989; Gale 1995; Banks 1996; Benjamin 1998; Gabelica 1999; Hofstadler and Griffey 2001; Wyttenbach 2003; Heck 2004; Benesch and Robinson 2006; Sharon 2007; Hopper 2009) have demonstrated that if vaporization is done under mild conditions, the structure becomes stable during long periods of time, probably similar to the time-scale of the experiment, and gas phase ensembles can be used to accurately model the solution structure (Patriksson 2007; DAbramo 2009; Meyer 2009).

The question is whether these findings also stand for a highly flexible and charged non-globular molecule as DNA, whose conformation is known to be very dependent on the solvent environment. An analysis of the energetics of DNA (see above) seems to suggest that in the absence of solvent screening, DNA should unfolds.

However, mass spectroscopic studies using mild electrospray vaporization conditions (ESI-MS) (Gale 1995; Schnier, Klassen et al. 1998; Gabelica, De Pauw et al. 1999; Hofstadler and Griffey 2001; Reyzer, Brodbelt et al. 2001; Rosu, Gabelica et al. 2002; Vairamani and Gross 2003; A Prez 2005; Gabelica 2005; Gidden 2005; Gabelica, Rosu et al. 2007; Rosu, Nguyen et al. 2007; Rosu, De Pauw et al. 2008; Wan, Guo et al. 2008), has provided convincing experimental evidence demonstrating that nucleic acids maintain some of the solution-phase structural features when transferred into the gas phase.

It is clear, for example, that duplexes, triplexes and quadruplexes are in the gas phase in the same oligomeric state found in solution, and that even non-covalent complexes of different DNA structures with small ligands are detected intact in the gas phase (see references (Gale 1995; Schnier, Klassen et al. 1998; Gabelica, De Pauw et al. 1999; Wan, Gross et al. 2000; Hofstadler and Griffey 2001; Reyzer, Brodbelt et al. 2001; Rosu, Gabelica et al. 2002; Vairamani and Gross 2003; Gabelica 2005; Gidden 2005; Gabelica, Rosu et al. 2007; Rosu, Nguyen et al. 2007; Rosu, De Pauw et al. 2008).

These experimental findings suggest that some structural determinants (for example groove geometries, and hydrogen bonding pattern) modulating the specific ligand-DNA binding in aqueous solution are somehow conserved in the gas phase. Even more surprising, ion mobility electrospray ionization experiments (IM-ESI) demonstrated that the general shape of nucleic acids (experimentally measured by the collision cross section (Wyttenbach 2013)) in gas phase can be close to the expected values in solution, suggesting that contrary to chemical intuition nucleic acids in the gas phase remain quite compact. Clearly, basic concepts on the role of electrostatic repulsion on DNA structure need to be revisited.

During my doctoral thesis I did a computational study to investigate the behaviour of small pieces of DNA to gain insight into the structure and dynamics of DNA in the gas phase.

The nature of the vaporization process

In order to model gas-phase experiments with theoretical methods we need to understand the process followed to transfer molecules from solution to the gas phase (the vaporization process; see (Meyer 2012; López 2013) and references therein).

In a mild ESI-MS experiment, the solution containing the nucleic acids is injected in a capillary whose tip is placed in front of the mass spectrometer, see scheme inFigure 1.11. A potential difference between the capillary and the entrance of the mass spectrometer is then applied, generating an electric field in which the ions move towards the counter electrode (contrary to proteins, nucleic acids are typically sprayed in the negative mode).

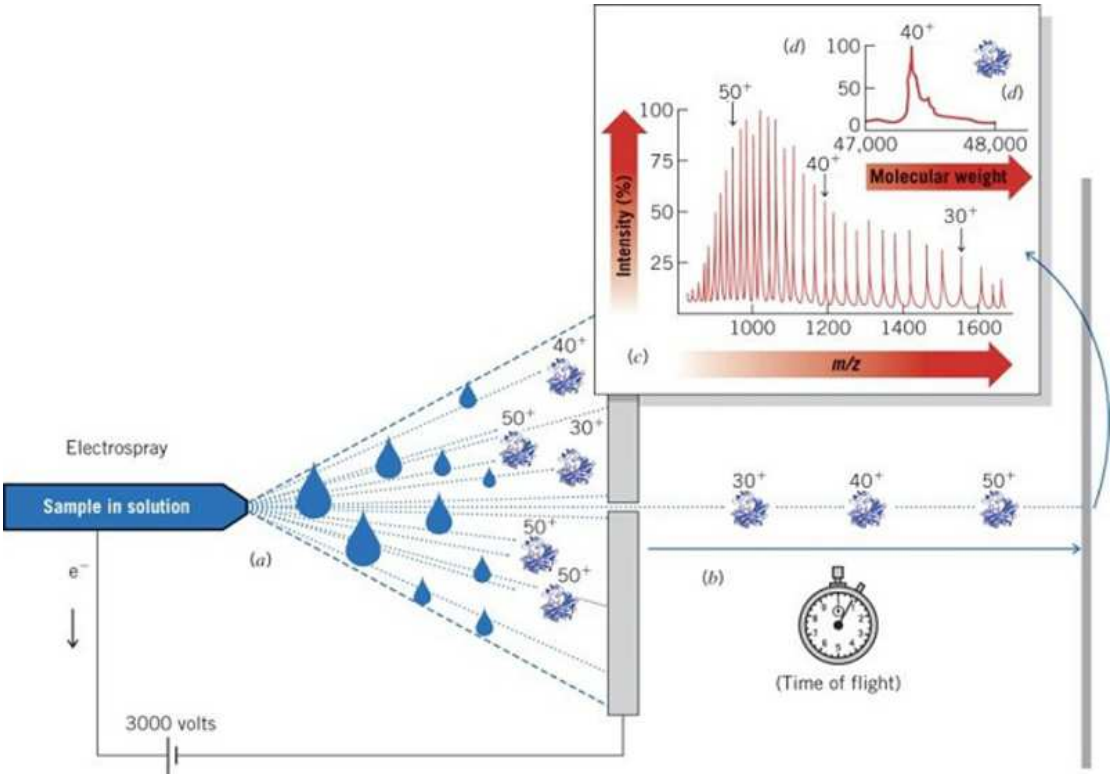

FIGURE 1.11: Electrospray ionization scheme.

This results in a jet formation and in the emission of charged droplets containing nucleic acids and solvent molecules. As the droplets move to the counter-electrode, water molecules are evaporated, which increases the charge density at the droplet surface up to the Rayleigh limit (De la Mora 2000; Meyer 2012; Rayleigh 1882). This charge density

increase induces droplet fission and the emission of a secondary jet or smaller droplets, whose subsequent fission proceeds in a similar way to that of the parent drop. Fully dehydrated nucleic acids are obtained after a few generations of fissions as shown in Figure 1.11.

The evaporation of solvent from the nucleic-acid containing droplets is supposed to occur following the charge residue mechanism (Fenn 1993; Hautreux M 2004; Meyer 2012). Accordingly, solvent migration is gradual and fast until the last solvent molecule leaves the droplet and the nucleic acids is in a pure gas-phase environment.

During this migration and evaporation process part of the original nucleic-acid charge moves to the solvent in form of OH- ions, i.e. the nucleic acids capture protons from the surrounding medium. The final total charge on the oligonucleotide should correlate then with the solvent accessible surface of the nucleic acid (Fenn 1993; Felitsyn, Kitova et al. 2002; Hautreux M 2004; Wan, Guo et al. 2008; Rayleigh 1882).

Analysis of nucleic acids ESI-MS spectra shows clear charge/mass signals that agree well with molecular surfaces similar to those found in solution (Felitsyn, Kitova et al. 2002; Rueda, Kalko et al. 2003; Rueda, Luque et al. 2005; Arcella, Portella et al. 2012), but as charge increases nucleic acid show shapes not longer consistent with the folded state.

Mild vaporization processes imply desolvation of the sample and a significant reduction on the total negative charge on the nucleic acid to a final preferred final charge (Gale 1995; Schnier, Klassen et al. 1998; Gabelica, De Pauw et al. 1999; Wan, Gross et al. 2000; Hofstadler and Griffey 2001; Reyzer, Brodbelt et al. 2001; Rosu, Gabelica et al. 2002; Vairamani and Gross 2003; Gabelica 2005; Gidden 2005; Gabelica, Rosu et al. 2007; Rosu, Nguyen et al. 2007; Rosu, De Pauw et al. 2008; Wan, Guo et al. 2008; Rosu, Gabelica et al. 2010). If the solution sample contains no significant amount of unfolded oligonucleotide, it is at a suitable ionic strength (typically from 100 mM to 150mM ammonium acetate), and the vaporization is mild enough, major signals detected experimentally generally correspond to the preferred charge $\pm$ 1. This observation implies that the experimental set-up and the structural characteristics of the nucleic acids fix within a narrow range the final charge state of the nucleic acid when it starts migrating along the drift tube.

Ion mobility spectrometry

In classical ion mobility spectrometry, dehydrated nucleic acids are forced to move in a chamber filled with inert gas at a given density. The drift velocity (ν), which determines

the time interval a given nucleic acid takes to arrive to the detector, depends linearly on the electric field and the intrinsic mobility of the oligonucleotide

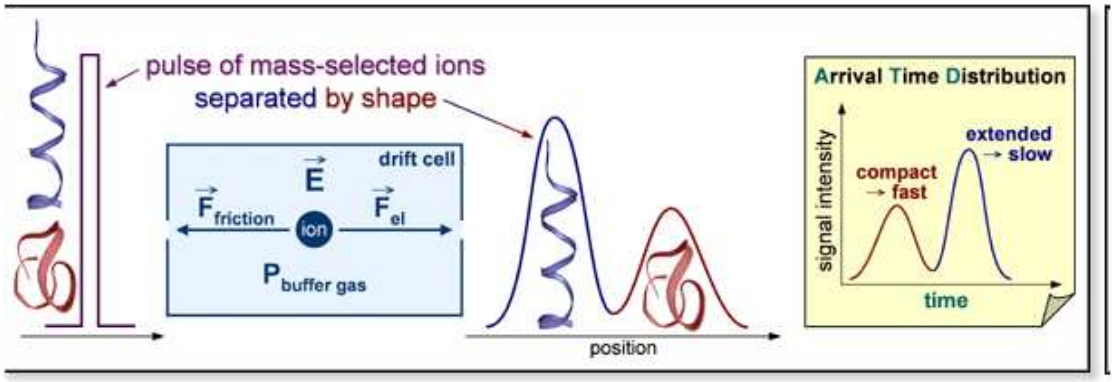

FIGURE 1.12: Drift chamber scheme for ion mobility and collision cross section measurement.

$$\nu = KE \tag{1.1}$$

with K being the mobility and E the external field. The oligonucleotide mobility depends on its mass and charge, but also its shape and size modulates the kinetic energy lost by collision against the inert gas particles in the drift tube (see Figure 1.12). In the low field limit, the mobility can be determined by the Mason-Schamp equation:

$$K = \frac{3}{16} N_0 \frac{q}{N^2} \left(\frac{2\pi}{\mu k T} \right)^{\frac{1}{2}} \frac{1}{\Omega} \tag{1.2}$$

where N is the gas number density, N_0 is the gas number density at standard temperature and pressure, q is the ion charge, μ is the reduced mass ($\mu = \frac{mM}{(m+M)}$, where M is the mass of the gas and m is the mass of the ion), k is the Boltzmann constant, T is the gas temperature, and Ω is a standard property related to the molecular surface that is exposed to collisions with the inert gas in the drift tube (Wyttenbach 2013)).

The exact thermodynamic conditions during the vaporization and the ion migration are not know exactly, hindering the comparison between experimental results and theoretical calculations (Meyer 2012). For example, the effective temperature of ions in the vaporization chamber is difficult to determine, since vaporization is an endothermic process that tends to reduce the effective temperature. All these considerations cannot be ignored when comparing the structural information obtained from simulations with observables derived from ESI-MS experiments, since the degree of uncertainties does not allow for a straightforward comparison of experiments and simulations.

1.7 Aims of the thesis

During my PhD thesis I used theoretical techniques, in particular Molecular Dynamics to study the structural properties of nucleic acids in non-canonical environment, especially in the gas phase and apolar conditions. Objectives of this study are:

1. Nucleic acids in apolar environment
   (a) Determination of structural properties of DNA in apolar solvent under different protonation state.
   (b) Estimate the free energy cost for transferring the DNA from polar to apolar environment for different charge state and the structural and topological consequences of phase transfer.
2. Structure of DNA Triplex in the gas phase
   (a) Determination of the structural properties of Triplex DNA in vacuum condition on atomistic detail.
3. Oligonucleotide structures in the gas phase
   (a) Explore the whole ensemble of DNA duplex in the vacuum conditions on millisecond timescale.
   (b) Study the impact of quantum effects in the structural ensemble of DNA in the gas phase.

Papers related to this thesis

1. **Arcella A.**, Portella G., Ruiz, M.L., Eritja R.,Vilaseca M., Gabelica V., Orozco M.,"The structure of Triplex DNA in the gas phase", *J.Am.Chem.Soc.*,2012, 134(15), 6596-6606
2. **Arcella A.**, Portella G., Orozco M.,2014, "Nucleic Acids in the gas phase", Chapter 3, Springer
3. **Arcella A.**, Dreyer J., Ippoliti E., Portella G., Gabelica V., Carloni P. and Orozco M."The structure of DNA in the gas phase", *Angew.Chemie.Int.Ed.*, *in revision*
4. **Arcella A.**, Portella G., Collepardo-Guevara R., Chakraborty D., Wales D.J. and Orozco M., "The structure and properties of DNA in apolar solvent", *J.Phys.Chem.B*, 2014, 118(29), 8540-8548

Other papers

5. Pujol R., **Arcella A.**, Guy M., Serra-Batiste M., Vilapriny S.P., Vilaseca M., Orozco M, Carulla N., "Early A-beta oligomers: new structural insight and well-defined synthetic mimics", *Nature Chemistry.*, *second revision*

6. **Arcella A.**, Portella G., Orozco M., "Free energy cost for DNA hairpin insertion into membrane",*in preparation*

Bibliography

A.Perez, J. S., P Jurecka, P Hobza, FJ Luque, M Orozco (2005). "Are the Hydrogen Bonds of RNA (A· U) Stronger Than those of DNA (A· T)? A Quantum Mechanics Study."Chemistry-A European Journal 11 (17): 5062-5066.

Anthony-Cahill, S. J., Mathews, C.K., van Holde, K.E., Appling, D.R. (2012). Biochemistry Englewood Cliffs, N.J. Anton Barty, J. K., and Henry N. Chapman (2013). "Molecular Imaging Using X-Ray Free-Electron Lasers." Annual Review of Physical Chemistry 64: 415-435.

Arcella, A., G. Portella, et al. (2012). "Structure of triplex DNA in the gas phase." J Am Chem Soc 134(15): 6596-6606.

Avery, O. T., MacLeod, C.M. and McCarthy, M. (1944). "Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus Type III." J.Exp.Med. 79: 137-158.

Banks, J. F., Jr., Whitehouse, C. M. (1996). "Electrospray ionization mass spectrometry." Methods Enzymol 270: 486-519.

Barabas, O., D. R. Ronning, et al. (2008). "Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection." Cell 132(2): 208-220.

Barends, T. R., L. Foucar, et al. (2014). "De novo protein crystal structure determination from X-ray free-electron laser data." Nature 505(7482): 244-247.

Benesch, J. L. and C. V. Robinson (2006). "Mass spectrometry of macromolecular assemblies: preservation and dissociation." Curr Opin Struct Biol 16(2): 245-251.

Benjamin, D. R., Robinson, C. V., Hendrick, J. P.,Hartl, F. U.,Dobson, C. M. (1998). "Mass spectrometry of ribosomes and ribosomal subunits." Proc Natl Acad Sci U S A 95(13): 7391-7395.

Bernues, J., R. Beltran, et al. (1990). "DNA-sequence and metal-ion specificity of the formation of *H-DNA." Nucleic Acids Res 18(14): 4067-4073.

Betts, L., J. A. Josey, et al. (1995). "A nucleic acid triple helix formed by a peptide nucleic acid-DNA complex." Science 270(5243): 1838-1841.

Bhatia, D., Sharma, S.,Krishnan, Y. (2011). "Synthetic, biofunctional nucleic acid-based molecular devices." Curr Opin Biotechnol 22(4): 475-484.

Blackburn, G. M. G., M. J. (1990). Nucleic Acids in Chemistry and Biology. Oxford, IRL Press.

Bloomfield, V. A. C., D. M.; Tinoco, I. (2000). Nucleic Acids. Structure, Properties and Functions. Sausalito, CA, University Science Books.

Brown, M. D., Schatzlein, A.G., Uchegbu I.F. (2001). "Gene delivery with synthetic (non viral) carriers." International Journal of Pharmaceutics. 229: 121.

Calladine, C. R. (1982). "Mechanics of sequence-dependent stacking of bases in B-DNA." J Mol Biol 161(2): 343-352.

Cech, T. R. (1993). "The efficiency and versatility of catalytic RNA: implications for an RNA world." Gene 135(1-2): 33-36.

Chapman, H., Barty A, Bogan MJ, Boutet S, Frank M, Hau-Riege SP, Marchesini S,, B. S. Woods BW, Benner H, London RA, Plonjes E, Kuhlmann M, Treusch R,, et al. (2006). "Femtosecond diffractive imaging with a soft-X-ray free-electron laser." Nature Phys. 2(12): 839-843.

Chargaff, E. (1950). "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation." Experientia 6(6): 201-209.

Chin, J. Y., E. B. Schleifman, et al. (2007). "Repair and recombination induced by triple helix DNA." Front Biosci 12: 4288-4297.

Courey, A. J. and J. C. Wang (1983). "Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions." Cell 33(3): 817-829.

Cui S., Y. J., Khner F., Schulten K., Gaub H.E.; (2007). "Double-Stranded DNA Dissociates into Single Strands when dragged into a Poor Solvent." J. Am. Chem. Soc. 129: 14710-14716.

DAbramo, M., Meyer, T., Bernad, P., Fernandez-Recio, J., Orozco, M. (2009). "On the use of low-resolution data to improve structure prediction of proteins and protein complexes." M. J. Chem. Theor. Comput. 5: 3129.

Davis, G. K., K.J. (2010). "Chromosomal Mutagenesis." Humana Press.

De la Mora, F. (2000). J. Anal Chim Acta (406): 93104.

Drew, H. R. and R. E. Dickerson (1981). "Structure of a B-DNA dodecamer. III. Geometry of hydration." J Mol Biol 151(3): 535-556. Ewert, K., Ahmad, A., Evans, H.M., Safinya, C.R. (2005). "Cationic lipid-DNA complexes for non-viral gene therapy:

relating supramolecular structures to cellular pathways." Expert Opinion on Biological Therapy 5: 3353.

Ewert, K., Slack, N.L., Ahmad, A., Evans, H.M., Lin, A.J.. (2004). "Cationic lipid-DNA complexes for gene therapy: Understanding the relationship between complex structure and gene delivery pathways at the molecular level." Current Medicinal Chemistry 11: 133149.

Felitsyn, N., E. N. Kitova, et al. (2002). "Thermal dissociation of the protein homodimer ecotin in the gas phase." J Am Soc Mass Spectrom 13(12): 1432-1442.

Feng Y., S. T. S., Hoh J.H.; (1999). "Ethanol-induced structural transitions of DNA on mica." Nucleic Acids Res. 27(8): 1943-1949.

Fenn, J. B. (1993). "Ion formation from charged droplets: roles of geometry, energy, and time." J Am Soc Mass Spectrom 4,: 524535.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science 246(4926): 6471.

Gabelica, V., E. De Pauw, et al. (1999). "Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry." J Mass Spectrom 34(12): 1328-1337.

Gabelica, V., De Pauw, E., Rosu, F. (1999). "Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry." Journal of Mass Spectrometry 34(12): 1328-1337.

Gabelica, V., F. Rosu, et al. (2007). "Base-dependent electron photodetachment from negatively charged DNA strands upon 260-nm laser irradiation." J Am Chem Soc 129(15): 4706-4713.

Gabelica, V. R., F.; Witt, M.; Baykut G.; De Pauw E.; (2005). "Fast gas-phase hydrogen/deuterium exchange observed for a DNA G-quadruplex." Rapid Commun Mass Spectrom 19(2): 201-208

Gale, D. C., Smith, R. D. (1995). "Characterization of noncovalent complexes formed between minor groove binding molecules and duplex DNA by electrospray ionization-mass spectrometry." J Am Soc Mass Spectrom 6(12): 1154-1164.

Gale, D. C. S., R. D. . (1995). "Characterization of noncovalent complexes formed between minor groove binding molecules and duplex DNA by electrospray ionization-mass spectrometry." J. Am. Soc. Mass. Spectrom 6(12): 1154.

Gidden, J. B., E. S.; Ferzoco, A. L.; Bowers, M. T. (2005). "Structural motifs of DNA complexes in the gas phase." Int. J. Mass Spectrom 240: 183-193.

Gonzalez-Perez, B., M. Lucas, et al. (2007). "Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides." EMBO J 26(16): 3847-3857.

Guckian, K. M., B. A. Schweitzer, et al. (1996). "Experimental Measurement of Aromatic Stacking Affinities in the Context of Duplex DNA." J Am Chem Soc 118(34): 8182-8183.

Hatfield, G. W. and C. J. Benham (2002). "DNA topology-mediated control of global gene expression in Escherichia coli." Annu Rev Genet 36: 175-203.

Hautreux M, H. N., Du Fou de Kerdaniel A, Zahir A, Malec V, Laprvote O. (2004). "Under nondenaturing solvent conditions, the mean charge state of a multiply charged protein ion formed by electrospray is linearly correlated with the macromolecular surface." Int J. Mass Spectrom. 231: 131137.

Heck, A. J., Van Den Heuvel, R. H. (2004). "Investigation of intact protein complexes by mass spectrometry." Mass Spectrom Rev 23(5): 368-389.

Hershey, A. D., Chase, M. (1952). "Independent functions of viral protein and nucleic acid in growth of bacteriophage." J Gen Physiol 36(1): 39-56.

Hofstadler, S. A. and R. H. Griffey (2001). "Analysis of noncovalent complexes of DNA and RNA by mass spectrometry." Chem Rev 101(2): 377-390.

Hoogsteen, K. (1963). "The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine." Acta Cryst. 16: 907-916.

Hopper, J. T., Oldham, N. J. (2009). "Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability." J Am Soc Mass Spectrom 20(10): 1851-1858.

Horwitz, M. S. and L. A. Loeb (1988). "An E. coli promoter that regulates transcription by DNA superhelix-induced cruciform extrusion." Science 241(4866): 703-705.

Htun, H. and J. E. Dahlberg (1989). "Topology and formation of triple-stranded H-DNA." Science 243(4898): 1571-1576.

Huey, R., Mohr, S. C. (1981). "Condensed states of nucleic acids. III. psi(+) and psi(-) conformational transitions of DNA induced by ethanol and salt." Biopolymers 20(12): 2533-2552.

Hunter, C. A. (1993). "Sequence-dependent DNA structure. The role of base stacking interactions." J Mol Biol 230(3): 1025-1054.

International Human Genome Sequencing, C. (2004). "Finishing the euchromatic sequence of the human genome." Nature 431(7011): 931-945.

Koltover, I., Salditt, T., Radler, J.O., Safinya, C.R. (1998). "An inverted hexagonal phase of cationic liposome-DNA complexes related to DNA release and delivery." Science 281: 7881.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature 409(6822): 860-921.

Lannan, F. M., I. Mamajanov, et al. (2012). "Human telomere sequence DNA in water-free and high-viscosity solvents: G-quadruplex folding governed by Kramers rate theory." J Am Chem Soc 134(37): 15324-15330.

Lin J., S. N. C., N. Vaidehi (2008). "Molecular-dynamics simulations of insertion of chemically modified DNA nanostructures into a water-chloroform interface." Biophys J. 95(3): 1099-1107.

Lpez, A. T., T.; Vilaseca,M.; Giralt,E. (2013). New. J. Chem. 37: 1283-1289.

MacDonald, D., G. Demarre, et al. (2006). "Structural basis for broad DNA-specificity in integron recombination." Nature 440(7088): 1157-1162.

Masai, H. and K. Arai (1997). "Frpo: a novel single-stranded DNA promoter for transcription and for primer RNA synthesis of DNA replication." Cell 89(6): 897-907.

Mathieu, F. S., Liao, C., Mao, J., Kopatsch, T., Wang, and N. C. Seeman. (2005). "Six-helix bundles designed from DNA." Nano Lett. 5(661665).

McCullough, B. J., A. Entwistle, et al. (2009). "Digital ion trap mass spectrometer for probing the structure of biological macromolecules by gas phase X-ray scattering." Anal Chem 81(9): 3392-3397.

McNeer, N. A., E. B. Schleifman, et al. (2011). "Polymer delivery systems for site-specific genome editing." J Control Release 155(2): 312-316.

Meyer, T., de la Cruz, X., Orozco, M. (2009). "An atomistic view to the gas phase proteome." Structure 17(1): 88-95.

Meyer, T. G., V.; Grubmller,H.; Orozco,M.; (2012). "Proteins in the gas phase." WIRES Computational Molecular Science.

Miyoshi, D. N., K.; Tateishi-Karimata, H.; Ohmichi, T.; Sugimoto, N., (2009,). "Hydration of Watson-Crick Base Pairs and Dehydration of Hoogsteen Base Pairs Inducing Structural Polymorphism under Molecular Crowding Conditions." Journal of the American Chemical Society 131: 3522-3531.

Montesi A., P. M., MacKintosh F. C.; (2004). "Collapse of a semiflexible polymer in poor solvent." PHYSICAL REVIEW E 69(021916).

Moravek, Z., S. Neidle, et al. (2002). "Protein and drug interactions in the minor groove of DNA." Nucleic Acids Res 30(5): 1182-1191.

Neutze, R., Huldt, G., Hajdu, J., and van der Spoel, D.; (2004). "Potential impact of an X-ray free electron laser on structural biology. ." Radiat. Phys. Chem 71: 905 916.

Nielsen, P. E., M. Egholm, et al. (1991). "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide." Science 254(5037): 1497-1500.

Nikolova, E. N., F. L. Gottardo, et al. (2012). "Probing transient Hoogsteen hydrogen bonds in canonical duplex DNA using NMR relaxation dispersion and single-atom substitution." J Am Chem Soc 134(8): 3667-3670.

Panayotatos, N. and R. D. Wells (1981). "Cruciform structures in supercoiled DNA." Nature 289(5797): 466-470.

Patriksson, A., Marklund, E., van der Spoel, D. (2007). "Protein structures under electrospray conditions." Biochemistry 46(4): 933-945.

Pereira G.G., W. D. R. M. (2001). "Toroidal Condensates of Semiflexible Polymers in Poor Solvents: Adsorption, Stretching, and Compression." Biophysical J. 80 161168.

Platt, J. R. (1955). "Possible Separation of Intertwined Nucleic Acid Chains by Transfer-Twist." Proc Natl Acad Sci U S A 41(3): 181-183.

Pohl, F. M. and T. M. Jovin (1972). "Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC)." J Mol Biol 67(3): 375-396.

POP, T. o. (1974). Basic Principles in Nucleic Acids Chemistry. New York, Academics.

Rayleigh, L. (1882). Phil. Mag.: 184-186.

Reyzer, M. L., J. S. Brodbelt, et al. (2001). "Evaluation of complexation of metal-mediated DNA-binding drugs to oligonucleotides via electrospray ionization mass spectrometry." Nucleic Acids Res 29(21): E103-103.

Roberts, R. W. and D. M. Crothers (1992). "Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition." Science 258(5087): 1463-1466.

Rostom, A. A., P. Fucini, et al. (2000). "Detection and selective dissociation of intact ribosomes in a mass spectrometer." Proc Natl Acad Sci U S A 97(10): 5185-5190.

Rosu, F., E. De Pauw, et al. (2008). "Electrospray mass spectrometry to study drug-nucleic acids interactions." Biochimie 90(7): 1074-1087.

Rosu, F., V. Gabelica, et al. (2002). "Triplex and quadruplex DNA structures studied by electrospray mass spectrometry." Rapid Commun Mass Spectrom 16(18): 1729-1736.

Rosu, F., V. Gabelica, et al. (2010). "Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry." Nucleic Acids Res 38(15): 5217-5225.

Rosu, F., C. H. Nguyen, et al. (2007). "Ligand binding mode to duplex and triplex DNA assessed by combining electrospray tandem mass spectrometry and molecular modeling." J Am Soc Mass Spectrom 18(6): 1052-1062.

Rueda, M., S. G. Kalko, et al. (2003). "The structure and dynamics of DNA in the gas phase." J Am Chem Soc 125(26): 8007-8014.

Rueda, M., F. J. Luque, et al. (2005). "Nature of minor-groove binders-DNA complexes in the gas phase." J Am Chem Soc 127(33): 11690-11698.

Saenger, W. (1984). Principles of Nucleic Acid Structure. New York, Springer-Verlag.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M. and Smith, M. (1977). "Nucleotide sequence of bacteriophage phi X174 DNA. ." Nature 265: 687-695.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A 74(12): 5463-5467.

Schneider, B. and H. M. Berman (1995). "Hydration of the DNA bases is local." Biophys J 69(6): 2661-2669.

Schneider, B., Cohen, D.M., Schleifer, L., Srinivasan, A.R., Olson, W.K. and Berman, H.M. (1993). "A systematic method for studying the spatial distribution of water molecules around nucleic acid bases." Biophys.J. 65: 2291-2303. Schneider, B. K., M. (1998). "Stereochemistry of binding of metal cations and water to a phosphate group. ." J. Am. Chem. Soc. 120: 161-165.

Schnier, P. D., J. S. Klassen, et al. (1998). "Activation energies for dissociation of double strand oligonucleotide anions: evidence for watson-crick base pairing in vacuo." J Am Chem Soc 120(37): 9605-9613.

Sharon, M., Robinson, C. V. (2007). "The role of mass spectrometry in structure elucidation of dynamic protein complexes." Annu Rev Biochem 76: 167-193.

Sinden, R. R., S. S. Broyles, et al. (1983). "Perfect palindromic lac operator DNA sequence exists as a stable cruciform structure in supercoiled DNA in vitro but not in vivo." Proc Natl Acad Sci U S A 80(7): 1797-1801.

Sinden, R. R. S., CA, (1994). DNA Structure and Function.

Siuzdak, G., B. Bothner, et al. (1996). "Mass spectrometry and viral analysis." Chem Biol 3(1): 45-48.

Solie, T. N. and J. A. Schellman (1968). "The interaction of nucleosides in aqueous solution." J Mol Biol 33(1): 61-77.

Strobel, S. A. and P. B. Dervan (1990). "Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation." Science 249(4964): 73-75.

Tereshko, V., Minasov, G. and Egli, M. (1999). "A "hydrat-ion" spine in a B-DNA minor groove." Journal of the American Chemical Society 121(15): 3590-3595.

Thuong, N. T. H. e. e., C. (1993). "Sequence-Specific Recognition and Modification of Double-Helical DNA by Oligonucleotides." Angewandte Chemie (International ed. in English) 32(5): 666-690.

Tribolet, R. and H. Sigel (1987). "Self-association of adenosine 5'-monophosphate (5'-AMP) as a function of pH and in comparison with adenosine, 2'-AMP and 3'-AMP." Biophys Chem 27(2): 119-130.

Turner, D. H. B., V. A., Crothers, D. M., Tinoco, I., Eds.; : (2000). Structure, Properties and Functions. Nucleic Acids. . Sausalito, CA, University Science Books: 308-310.

Vairamani, M. and M. L. Gross (2003). "G-quadruplex formation of thrombin-binding aptamer detected by electrospray ionization mass spectrometry." J Am Chem Soc 125(1): 42-43.

Val, M.-E., M. Bouvier, J. Campos, D. Sherratt, F. Cornet, D. Mazel, and F.-X. Barre. (2005). "The single-stranded genome of phage CTX is the form used for integration into the genome of Vibrio cholerae." Mol. Cell 19: 559-566.

Vasquez, K. M. (2010). "Targeting and processing of site-specific DNA interstrand crosslinks." Environ Mol Mutagen 51(6): 527-539.

Wan, C., X. Guo, et al. (2008). "Studies of the intermolecular DNA triplexes of C+.GC and T.AT triplets by electrospray ionization Fourier-transform ion cyclotron resonance mass spectrometry." J Mass Spectrom 43(2): 164-172.

Wan, K. X., M. L. Gross, et al. (2000). "Gas-phase stability of double-stranded oligodeoxynucleotides and their noncovalent complexes with DNA-binding drugs as revealed by collisional activation in an ion trap." J Am Soc Mass Spectrom 11(5): 450-457.

Wang, J. C. and A. S. Lynch (1993). "Transcription and DNA supercoiling." Curr Opin Genet Dev 3(5): 764-768.

Wang, Y. and D. J. Patel (1993). "Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex." Structure 1(4): 263-282.

Watson, J. D., Crick, F. H. (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." Nature 171(4356): 737-738.

Williamson, J. R., M. K. Raghuraman, et al. (1989). "Monovalent cation-induced structure of telomeric DNA: the G-quartet model." Cell 59(5): 871-880.

Wing, R., H. Drew, et al. (1980). "Crystal structure analysis of a complete turn of B-DNA." Nature 287(5784): 755-758. Woda, J., B. Schneider, et al. (1998). "An analysis of the relationship between hydration and protein-DNA interactions." Biophys J 75(5): 2170-2177.

Wyttenbach, T., Bleiholder, C.; Bowers, M. T. (2013). "Factors Contributing to the Collision Cross Section of Polyatomic Ions in the Kilodalton to Gigadalton Range: Application to Ion Mobility Measurements." Anal. Chem. 85: 2191-2199.

Wyttenbach, T., Bower, M. T. (2003). Top. Curr. Chem 255: 207.

Yang, D., Campolongo, M.J., Nhi Tran, T.N., Ruiz, R.C., Kahn, J.S., Luo D. (2010). "Novel DNA materials and their applications." Wiley Interdiscip Rev Nanomed Nanobiotechnol. 2(6): 648-669.

Zamenhof, S., Brawermann, G. and Chargaff, E. (1952). "On the Desoxypentose Nucleic Acids from Several Microorganisms." Biochim.Biophys.Acta 9: 402-405.

Chapter 2

Theory and Methods

Biomolecular modeling is a fertile and growing area, with exciting opportunities and potential applications. It is crucial for the biomolecular modeler to understand the issues of interest to biologists; but more important is to understand the complexity of biological systems and how to tackle them effectively by modeling. Vast amounts of data are being provided by large-scale research efforts in genomics, proteomics, glycomics and structural biology. The challenge for biomolecular modeling is to help in efforts to use these diverse data to develop new drugs, therapies, catalysts and biologically based nanotechnology. Increasingly, computer simulations of biological macromolecules are helping to meet this challenge (van der Kamp 2008). In particular, molecular simulations have been more and more being employed to investigate features not directly accessible to experiments.

Whereas it is indeed possible to take "still snapshots" of crystal structures and probe features of the motion of molecules through NMR, no current experimental technique allows access to all the time scales of motion with atomic resolution. Simulations based on fundamental physics offer the potential of filling-in these crucial 'gaps', modeling how proteins, nucleic acids and other biomolecules move, fluctuate, interact, react and function. Applications of theoretical models include studies of nucleic acids flexibility, nucleic acids and protein interaction, protein folding and conformational changes, association of proteins or nucleic acids with small molecules or other macromolecules, structure-based drug design, modeling the dynamics of ion channels and transport across membranes and modeling and analysis of enzyme mechanisms.

Improvements in computer hardware continue to deliver more computational power, which, when combined with theoretical and algorithmic developments, have led to an increasing range of applications of molecular modeling and molecular dynamic in biology. Molecular dynamics (MD) is a specialized computer-based discipline to simulate

time evolving atomistic or molecular systems. In such simulations, motion of atoms is determined within the framework of classical mechanics.

2.1 Nucelic Acids modeling

Molecular modelling of nucleic acids has unique characteristics mainly due to the difficulty of simultaneously represent both atomistic interactions at base pair scale, and on global phenomena scale such as at the chromosome scale. This variety in sizes, affects the choice of an approximation suitable for theoretical study.

The first step in the simulation of a nucleic acid is the choice of a functional connecting the three-dimensional coordinate system with its energy (Leach 2001). Ideally, quantum-mechanics techniques can be used to describe the systems. Unfortunately, these techniques require excessive computational cost and therefore its usefulness is limited to studies of reactivity or cases where a detailed description of the structure required some nucleotide bases (Cheatham and Young 2000).

Classical methods are an alternative to quantum methodologies to address the study of large chemical systems in which no changes occur in covalent bonds, or in the equilibrium electronic distribution, which allows to substitute the rigorous quantum Hamiltonian by a set of empirically parameterized equations, known as Force Field (McCammon 1987; Cheatham and Young 2000; Frenkel 2002). Nucleic acids can be modelled on three different size scale (Orozco, Perez et al. 2003): macroscopic, mesoscopic and microscopic scale.

2.1.1 Macroscopic (ideal elastic) models

In these models, it is assumed that the nucleic acid is like an *entropic spring*. Examples of macroscopic models include the *worm like chain* (WLC) model whereby the DNA is an isotropic flexible rod, and the *freely joined chain* (FJC) model according to which the DNA is modelled as a chain of links. Macroscopic methods are of low resolution, but computationally very efficient allowing the study of supramolecular properties such as chromosome packaging or the super-helicity of DNA strands. Typically, they are used to study properties of very long fragments, where properties are mostly extracted independent on the details of the sequence studied.

2.1.2 Mesoscopic model

It refers to the diffuse interface between microscopic and macroscopic descriptions of a system, i.e. the study of those systems too big to be treated at the microscopic level, but too complex to be represented at the macroscopic level. In the case of nucleic acids, mesoscopic studies refers to the representation of very large pieces of DNA, where sequence or environment effects makes unsuitable the use of the ideal-elastic rod models (Olson 1996; Bruant, Flatters et al. 1999; Olson and Zhurkin 2000; Matsumoto and Olson 2002).

These models divide the polynucleotide into small rod elements, often named beads. Each bead is flexible and allows the nucleic acid to adapt to external deformation forces.

The size of the beads depends on the type of problem to be analysed and the number of base pair that constitute the beads may range from one base pair to hundreds. Increasing the bead size introduces greater computational speed but may imply a reduction in resolution. To calculate the mesoscopic deformation energy the classic elastic potential is often used, which is parameterized on the base of observations of high-resolution structures, macroscopic properties or molecular dynamics simulations.

2.1.3 Microscopic Model

Within this category there are two groups: those based on collective variables and those considering atomistic descriptions. The first family of methods are based on prior knowledge of various parameters related to the structure of nucleic acids, which allows reducing their conformational degrees of freedom. An example is the Helicoidal model; wherein the structure is represented by a number of rotational and translational parameters at base pair or isolated bases scale.

Atomistic models use a detailed description of all degrees of freedom of the system, usually in Cartesian space. These methods have been the backbone of this thesis, and therefore will be explained in more detail below.

A huge part of this thesis benefits of Molecular Dynamics (MD) techniques that, actually is one of the principal tools in the theoretical study of biological molecules. This computational method calculates the time dependent behavior of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and

their complexes. They are also used in the determination of structures from x-ray crystallography and from NMR experiments. The principles and approximations on which MD simulations rest are briefly outlined in this chapter.

2.2 Molecular Dynamics

For a comprehensive description of MD I refer to (Berendsen 1990; Leach 2001; Hansson, Oostenbrink et al. 2002; Adcock 2006), and the reference manual of the GROMACS simulation package (Berk Hess 2010) used in this work.

2.2.1 Principles

The exact description of any physical system requires the solution of the time-dependent Schrdinger equation for the N-particle wave function $\Psi(\vec{r}, \vec{R})$ of the system, having $\vec{r}$ nuclear and $\vec{R}$ electronic degrees of freedom,

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{r}, \vec{R}) = H\Psi(\vec{r}, \vec{R}) \tag{2.1}$$

Here, H denotes the Hamiltonian and $\hbar = h/2\pi$ is the reduced Planck constant. Due to the large number of about 10^3 to 10^7 interacting particles for currently simulated biomolecular systems, any attempt at solving such systems via eq. 2.1 is prohibitive. Approximations are therefore needed to reduce computational demands on current available hardware. The most used one is the assumption of the Born-Oppenheimer (BO) approximation, which assume that for each nuclear configuration, electronic distribution has reached an equilibrium (this physically implies that nuclear are much slower than electronic movements). This means that within the BO approximation, following the dynamics of a system is equivalent to following the nuclei movements, which can be reasonably represented by Newtonian dynamics. *Ab initio* MD methods such as Car-Parrinello MD, shorly discribed in Appendix B, take into account the electronic energy by solving Schrödinger equation.

Empirical Force Field

As relevant biological processes usually involve large systems (thousands of atoms or more), and occur in relatively long timescales (from nano to microseconds or more), it is necessary to develop effective parameterized potentials (force-fields) that avoids the need to solve Newtons equations of motion. Empirical force-fields are faster to integrate,

albeit less accurate, in order to study this kind of systems. The term force field indicates a functional form for this approximation, which relates the configuration of the system (i = 1,...,N) (Adcock 2006) to its internal energy U, along with the set of parameters used in that function. The functional form can be written as:

$$\begin{aligned} U &= V_{bonds} + V_{angles} + V_{dihedr} + V_{Coul} + V_{LJ} \\ &= \sum_{bond} \frac{1}{2}(r - r_{eq})^2 + \sum_{angles} \frac{1}{2}(\theta - \theta_{eq})^2 \\ &\quad + \sum_{dihedr} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{dihedr} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \\ &\quad + \sum_{i<j} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{i<j} \left[4\epsilon_{ij} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \end{aligned} \tag{2.2}$$

Atom bond stretching, angle bending and improper dihedral angles are represented as harmonic terms, while dihedrals or torsional are described by a cosinusoidal term. Non-bonded interactions comprise two terms, the first is the Coulomb electrostatic, and the second term is a Lennard-Jones 6-12 that describes atom-atom repulsion and dispersion interactions. In eq 2.2, r and θ are respectively the bond length and valence angle; ϕ is the dihedral or torsion angle and r_{ij} is the distance between atoms i and j. Parameters include the bond force constant and equilibrium distance, K_n and r_{eq}, respectively; the valence angle force constant and equilibrium angle, K_θ, and θ_{eq}, respectively; the dihedral force constant, multiplicity and phase angle, V_n, n, and θ, respectively. The functional form used for out-of-plane distortions (e.g. in planar groups) is different in different force fields. For instance, in the AMBER (Weiner 1986) force field this term has the same form as that used for proper dihedrals, while in CHARMM (Brooks 1983) a harmonic term is used.

Non bonded parameters between atoms i and j include the partial atomic charges, qi, along with the LJ well-depth, ϵ_{ij}, and σ_{ij}, the (finite) distance at which the inter-particle potential is zero. These terms are also referred to as the interaction or external parameters. Typically, ϵ_{ij} and σ_{ij} are obtained for individual atom types and then combined to yield ϵ_{ij}, and σ_{ij} for the interacting atoms via combining rules. The dielectric constant ϵ is typically set to 1 (corresponding to the permittivity of vacuum) in calculations that incorporate explicit solvent representations.

Van der Waals and electrostatic interactions are calculated between atoms belonging to different molecules or for atoms in the same molecules separated by at least three

bonds. For the van der Waals potential, this truncation introduces only a small error in the energy. This is not the case for the electrostatic potential, because the Coulomb interaction between charges q_i and q_j decays slowly with distance. Hence it cannot be truncated, but when periodic boundary conditions are used, it is computed with efficient schemes such as Particle Mesh Ewald (see below) in conjunction with periodic boundary conditions, which aproximate the exact result to an acceptable error.

2.2.2 The force field for Nucleic Acids

Although the first molecular dynamics (MD) simulations of proteins were published in the late 1970s (McCammon 1976 ; McCammon, Gelin et al. 1977), the first restrained simulations of nucleic acids (NAs) did not appear until the mid-1980s (Levitt 1983), and reliable unrestrained simulations of these molecules were not possible until the mid-1990s, when new force fields were developed and methods for the proper representation of long-range electrostatic effects were incorporated into simulation codes (Darden 1993; Cornell 1995; MacKerell 1995; York 1995; Cheatham 1999; Foloppe 2000). This time lag illustrates the technical problems intrinsic to the MD simulation of very charged and flexible polymers, such as NAs, in aqueous solution.

With these difficulties addressed, the last decade has seen a wide use of MD to study a very large number of NAs (Beveridge 2000; Cheatham and Young 2000; Giudice 2002; Orozco 2003; Cheatham 2004) in water for simulation periods in the range 150 ns. Most of these simulations have used explicit models of solvent and the particle mesh Ewald method (PME) (Darden 1993) to account for long-range electrostatic effects. Although others are available, the force fields implemented in AMBER and CHARMM have been the most widely used (Cornell 1995; MacKerell 1995; Cheatham 1999; Foloppe 2000). In particular, MD simulations using AMBER force fields have been shown to accurately reproduce the structural and dynamic properties of a large variety of canonical and non-canonical NAs in water (Beveridge 2000; Cheatham 2000; Cheatham and Young 2000; Giudice 2002; Orozco, Perez et al. 2003; Cheatham 2004; Prez 2005; Perez 2005).

Moreover, they have satisfactorily described complex conformational changes such as the A→B transition in duplex and triplex DNAs (Cheatham 1996; Cheatham III 1997; Cheatham, Crowley et al. 1997; Shields 1997; Sprous 1998; Soliva, Luque et al. 1999; McConnell and Beveridge 2000) and have performed well in simulations of DNAs in extreme environments (Rueda 2003; Rueda, Luque et al. 2005; Rueda, Luque et al. 2006). Finally, several systematic studies have demonstrated the excellent ability of the standard AMBER force field to reproduce very high-level QM data for hydrogen bond and stacking interactions in the gas phase (Alhambra 1997; Hobza 1997; Sponer 2004;

Sponer 2005; Prez 2005.; Sponer 2006). Overall, these studies suggest that the AMBER force field is physically meaningful and retains a proper balance between intramolecular and intermolecular forces.

In particular the full re-parameterization of the α/γ torsional term to derive the most recent AMBER force field, the **parmbsc0** (Perez, Marchan et al. 2007) based on AMBER-parm99 and used in this thesis, has been widely demonstrated to model accurately the structural and dynamic properties of a large variety of NAs over μs MD simulation time scales.

2.3 Design a Molecular Dynamics Simulation in Biomolecular Field

A key decision in beginning a simulation of a biomolecular system is the choice of an appropriate method for that particular system and for the questions of interest. A modeling method should be capable of delivering a reliable result in a reasonable time (van der Kamp 2008). Studies involving multi-nanosecond dynamics simulations are now common. However, expert knowledge is still required, and care needs to be taken to ensure that the application of a biomolecular simulation method to a particular problem is meaningful and useful.

Concerning MD simulations the ingredients are basically three:

1. A model for the **interaction** between system constituents (atoms, molecules, surfaces, etc.) is needed.
2. An **intergrator** is needed, which propagates particle positions and velocities from time t to $t + \Delta t$. It is a finite difference scheme that moves trajectories discretely in time. The time step t has properly to be chosen to guarantee stability of the integrator, i.e. there should be no drift in the systems energy.
3. A **statistical ensemble** has to be chosen, where thermodynamic quantities like pressure, temperature or the number of particles are controlled.

These choices essentially define an MD simulation. Molecular dynamics simulations used in this thesis are performed with a classical description of the system under study; In this work, the AMBER (Weiner 1986; Perez, Marchan et al. 2007) force field has been used for the description atomic interactions required in point 1, while the GROMACS (Berk Hess 2010) package has been used to integrate the equation of motion. Leap-frog algorithm (Hockney 1973) is used as integrator (point 2) and systems

are simulated at the isobaric-isothermal NPT statistical ensemble (point 3). Details on the integrators and statistical ensemble are described in section 2.5.1 below.

2.4 Validity of MD

Although MD simulations have become an established tool in the study of biomolecules, the severe approximations used to obtain trajectories need to be kept in mind. For example, the description of the biomolecular system as point masses moving classically in an effective potential breaks down as soon as quantum effects such as electric reorganizations. In such case a combined quantum mechanical and classical mechanical (QM/MM) approach, originally proposed by Warshel and Levitt (Warshel 1976), may be taken that allows for an accurate description of electronic excitations, charge-fluctuations and transfer and the forming and breaking of chemical bonds.

In classical MD simulations, depending on the chosen force field and type of compound studied, each atom is assigned a partial charge that reflects the polarity and approximately models effective polarization. Throughout the simulation these charges are kept constant thereby excluding explicit polarization effects. Nowadays, several polarizable water model and force fields exists, see (Warshel 2007) for a recent review.

The approximation of the potential energy surface V($\mathbf{r}$) by some empirical force field naturally raises the question of how accurate these potential are, and how uncertainties in them affect reliability of the simulation results. Each force field has its own strengths in reproducing certain observations due to the data that were specifically used to parameterize it. Consequently, the choice of a particular force field will depend on the property and level of accuracy one is interested in.

2.5 Implementation details and simulation setup protocol

The above approximations lay the foundation for a practical realization of MD simulations of nucleic acids, as it is done in the GROMACS simulation suite which was used here and whose algorithms and methods will be sketched in the following.

2.5.1 Integrator

Newtons equations of motion are solved iteratively in discrete steps by means of the leap-frog algorithm (Hockney 1973), which has the advantage that the computationally intensive force calculations need to be done only once per integration step. The length of

one time step has to be chosen such that it is small in comparison to the fastest motions of the system. Bond vibrations involving hydrogen occur within several femtoseconds, restricting the time step to 0.5 fs. A number of algorithms to constrain covalent bond lengths have been developed that allow larger time steps. The SHAKE algorithm (Ryckaert 1977) is used to constrain all bond distances to their equilibrium value, allowing a 2 fs integration time step for solution conditions. Since forces are stronger in the absence of water, a more conservative 1 fs integration step (using also SHAKE) is used for gas phase calculations.

Besides interactions with membranes and other macromolecules, water is the natural environment for proteins. For a simulation of a model system that matches the in vivo system as close as possible, water molecules and sodium chloride in physiological concentration are added to the system in order to solvate the DNA. Transferable interaction potential 3 point (TIP3P) (Jorgensen 1983) parameters for water and Dangs parameters (Smith 1994) for ions are used.

2.5.2 Periodic Boundary Condition

Having a simulation box filled with solvent and solute, artifacts due to the boundaries of the system may arise, such as evaporation, high pressure due to surface tension and preferred orientations of solvent molecules on the surface. To avoid such artifacts, periodic boundary conditions are applied. In this way, the simulation system does not have any surface.

This, however, may lead to new artifacts if the molecules artificially interact with their periodic images due to e.g. long-range electrostatic interactions. These periodicity artifacts are minimized by enlarging the simulation box, but caution needs to be taken even for significantly large boxes in the study of large macromolecules. Different choices of unit cells, e.g., cubic, dodecahedral or truncated octahedral allow an improved fit to the shape of the DNA, and, therefore, permit a substantial reduction of the number of solvent molecules. Obviously, no boundary conditions are used for simulation in the gas phase.

2.5.3 Long range interaction

Long-range Coulomb interactions in periodic systems are treated by the Particle-Mesh-Ewald (PME) method (Darden 1993; Essmann 1995), which, in contrast to simple cut-off methods (Brooks III 1985; Steinbach 1994), allows their correct and computationally

efficient evaluation. Again, no cutoff is used for either electrostatic or Lennard-Jones interactions in the gas phase and accordingly no Ewald correction has to be considered.

2.5.4 Statistical Ensemble

A solution of Newtons equations of motion conserves the total energy of the system, resulting in a microcanonical NVE ensemble, in which the number of particles N, the volume V and system energy E are constant. However, real biological subsystems of the size studied in simulations constantly exchange volume and energy with their surrounding to keep constant pressure and temperature. To account for these features, algorithms are introduced which couple the system to a temperature and pressure bath. From the many proposed thermostats (Anderson 1980; Nose 1984; Berendsen 1990; Kast 1994; Kast 1996), the popular Berendsen thermostat is used which simply rescales the velocities in each step using

$$\nu' = \lambda\nu, \qquad \lambda = \left[1 + \frac{\Delta t}{\tau t}\left(\frac{T_0}{T} - 1\right)\right]^{1/2} \tag{2.3}$$

Here, T_0 denotes the reference temperature of the heat bath, τT the coupling constant, and Δt the integration time step. Pressure coupling in this work is done by the Berendsen barostat (Berendsen 1984), which rescales the coordinates at each n umber of steps using a similar formalism than that considered for temperature. Thus, isobaric-isothermal NPT ensembles are created.

2.5.5 Molecular Dynamics simulation in the Gas Phase

When using classical MD simulations we can sample time-dependent conformational transitions for different charge states and sub-states, obtaining structural information that can be directly compared with experimental one (see below). We cannot forget, however, the dramatic simplifications assumed in these calculations which hampers their reliability: i) we assume that force-fields calibrated to study nucleic acids in solution are valid also in the gas phase; ii) we explicitly assume that no intra-molecular proton-transfer occurs during the simulation, and iii) we need to assume an ensemble (constant energy or constant temperature) for our calculations, which might not fit exactly with the real nature of the experiments.

2.6 From Microscopic to Macroscopic: Simulations as a bridge between theory and experiments

All the results obtained in this thesis are constantly related with experimental data. Thus, it is crucial to establish relationships between experiments and molecular simulations, and between macroscopic and microscopic world. In the following paragraph We will review these issues.

An experiment is usually made on a macroscopic sample that contains an extremely large number of atoms or molecules sampling an enormous number of conformations. On the other hand, computer simulations generate information at the microscopic level, including atomic positions and velocities. The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires the use of the Statistical Mechanics formalism.

2.6.1 Statistical Mechanics

Statistical Mechanics is fundamental to the study of biological systems by molecular dynamics simulations because it provides the rigorous mathematical formalism that links measurable macroscopic properties to the distribution and motion of the atoms and the molecules of the N-body system.

Definitions: thermodynamics state and microscopic state

The thermodynamic state of a system is usually defined by a set of parameters (state variables) that are physical observables, like for example, the temperature, T, the pressure, P, and the number of particles, N.

The mechanical or microscopic state of a system is defined by the atomic positions, q, and momenta, p; these can also be considered as coordinates in a multidimensional space called phase space (Γ). A single point in phase space, denoted by G, describes the state of the system. Thus, the macroscopic or thermodynamic state (macrostate) described by thermodynamic quantities, is statistically interpreted in terms of all the accessible microscopic mechanical states (microstates) where each is characterized by all the N atoms positions,

$$(q_1, q_2, \ldots, q_N) \equiv q^N \tag{2.4}$$

and momenta

$$(p_1, p_2, \ldots, p_N) \equiv p^N \tag{2.5}$$

i.e. points in the multidimensional phase space Γ

An ensemble is a collection of points in phase space satisfying the conditions of a particular thermodynamic state. Or we can also say that it is a collection of all possible systems, which have different microscopic states but have an identical macroscopic or thermodynamic state.

Common techniques to sample the configurations assumed by a system at equilibrium are Molecular Dynamics (MD), Monte Carlo (MC) sampling and Stochastic/Brownian dynamics. MD simulation generates a sequence of points in phase space as a function of time; these points belong to the same ensemble, and they correspond to the different conformations (and associated momenta) of the system. Thermodynamic quantities are determined in computer simulations as ensemble averages over a large number of microscopic configurations assumed by the system under study. This is allowed due to the **ergodic hypothesis** assumption.

2.6.2 Calculating averages from a Molecular Dynamics simualtion: the Ergodic Hypotesis

In statistical mechanics, averages corresponding to experimental observables are defined in terms of ensemble averages. Indeed thermodynamic observables can be alternatively modeled by considering at once a collection of identical systems. Each system represents one of all the accessible microstates. If this collection (statistical ensemble) is allowed to evolve in time, its behavior can be characterized by a time dependent distribution function, (qN(t), pN(t)) for the microstates. The instantaneous average value of the observable, O, over the phase space is interpreted as:

$$\frac{\int_r O(q^N(t)p^N(t))\rho(q^N(t)p^N(t))dq^N dp^N}{\int_r \rho(q^N(t)p^N(t))dq^N dp^N} \tag{2.6}$$

If we assume equal probability for all microstates, then the distribution of points in phase space is frozen into one single shape, i.e., the distribution function is time invariant, and the condition:

$$\frac{\rho(q^N(t)p^N(t))dq^N dp^N}{dt} = 0 \tag{2.7}$$

describes the thermodynamic equilibrium and the so-called ensemble average is defined as:

$$\langle O \rangle_{ensemble} = \frac{\int_r O(q^N(t)p^N(t))\rho(q^N(t)p^N(t))dq^N dp^N}{\int_r \rho(q^N(t)p^N(t))dq^N dp^N} \tag{2.8}$$

In a molecular dynamics simulation, the points in the ensemble are calculated sequentially in time, thus, assuming that the equations of motion of the system are solved, each observable can be empirically associated with a function (O), of the instantaneous microstate, (qN (t); pN (t)), of the system. Nevertheless, the quantity O (qN (t); pN (t)) is not an observable since measurements are performed in a macroscopic time; thus any microscopic observable is assumed to be a time averaged value:

$$\langle O \rangle_{ensemble} = \lim_{t\to\infty} \int_{\tau_0}^{\tau} O(q^N(t)p^N(t))dt \tag{2.9}$$

The ergodic hypothesis asserts that at the thermodynamic equilibrium the time average and the ensemble average are equal. That is, the systems path in phase space on a surface of constant energy will, in long run, visit each region of that surface equally often, and the experimental measurement will coincide with the calculated time and ensemble averages. Assuming the validity of the hypothesis means to assume fully sample the accessible phase space (hyper) volume during the observation (i.e. simulation) time. Although the ergodic hypothesis is typically assumed, it might not be always (for some example see e.g.(Frenkel 2002). This issue will be treated in the following sections.

2.7 Enhanced sampling methods for nucleic acids simulation

Computer simulations have assisted in the development of structural biology, enhancing the results attained by experimental methods; in simulating biological molecules, molecular dynamics (MD) simulations have moved from re-creating experimental structures to predicting secondary and folded structures and displaying the pathways by which they occur though both experimental and theoretical methods still face many limitations (Liu, Spielmann et al. 1995).

When simulating transitions and pathways, structures are frequently trapped in local energy minima where the thermal energy of the system is insufficient to traverse the energy barriers in an affordable amount of simulation time. For this reason, different

biasing techniques, such as the locally Enhanced Sampling (LES) and the multiple variations of the Replica Exchange Method (REM) have been developed to make computer simulations more efficient in reducing barriers between local minima and native states, biasing then the trajectory to favor sampling of infrequent conformational changes.

Standard molecular dynamics (MD) simulations (on the multi-nanosecond timescale) with explicit solvent are able to model only a few significant conformational transitions. For example, Cheatham et al. reported that MD simulations in explicit solvent were able to model the A-DNA→B-DNA transition under conditions that stabilize the B-DNA form in solution (Cheatham 1996). Many other cases of successful transitions followed by unrestrained MD simulations have been reported (see Orozco 2014, in press), but there are many other cases where transitions are not detected due the kinetic trapping of the system in local minima during simulations on affordable timescales.

2.7.1 Conformational sampling

The aim of computer simulations of molecular systems is to calculate macroscopic behaviour from microscopic interactions. Thus, in order to describe the thermodynamics and kinetic of nucleic acids, a through sampling of the conformational space of the system is required. Following equilibrium statistical mechanics, any observable that can be connected to microscopic experiments is defined as an ensemble average $< O >$ over all possible realizations of the system.

Given current computer hardware, a fully converged sampling of all possible conformational states with the respective Boltzmann weight is attainable for simple systems comprising, for example, a small piece of DNA. For longer nucleic acids conventional MD simulations often do not fully converge, which generates doubts on the validity of observables derived from the simulation ensembles.

2.7.2 Energy landscape

The inefficiency in sampling is, as noted above, a direct consequence of the ruggedness of the systems energy landscape, a concept put by Frauenfelder (Frauenfelder, Sligar et al. 1991; Frauenfelder and Leeson 1998). The term energy landscape is ambiguously used within literature, defined either as the potential or the free energy of the system as a function of all structural degrees of freedom N_{df}. The global shape of the free energy landscape of macromolecules is supposed to be funnel-liked, with the native state populating the global energy minimum (Anfinsen 1973). Looking in more detail, the complex high-dimensional free energy landscape is characterized by a multitude of

almost iso-energetic minima, separated from each other by energy barriers of various heights. Each of these minima corresponds to one particular conformational sub state, with neighbouring minima corresponding to similar conformations. Within this picture, structural transitions are barrier crossings, with the transition rate depending on the height of the barrier.

A plethora of enhanced sampling methods have been developed to overcome barriers and tackle this multi-minima problem (Berendsen 1990; Tai 2004; Adcock and McCammon 2006). Among them, generalized ensemble algorithms have been widely used in recent years (Iba 2001; Mitsutake 2001).

2.8 Replica Exchange Molecular Dynamics

2.8.1 Generalized Ensemble Algorithms

Generalized ensemble algorithm sample an artificial ensemble that is either constructed from compositions or extensions of the original ensemble. The multi-canonical algorithm (Berg 1991) and its variants simulated tempering Monte-Carlo (MC) (Lyubartsev 1992; Marinari 1992) are examples of this second category.

In the multi-canonical algorithm, the bell-shaped canonical distribution of the potential energy $p(E)$ is modified by a so-called multi-canonical weight factor $w(E)$ making the resulting distributions uniform ($p(E)w(E) = const.$). In a single multi-canonical simulation MD or MC can then sample this flat distribution extensively because potential energy barriers are not longer present. For simulated tempering, the temperature is no longer fixed but becomes a dynamical variable, and both the configuration and the temperature are updated during a single MC simulation with a weight factor. The latter is chosen such that a probability distribution of the temperature is constant ($p(T) = const.$).

Hence, a random walk in temperature space is made, which in turn induces a random walk in potential energy space and allows the system to escape from local energy minima. In both algorithms, estimates for canonical ensemble average of physical quantities are obtained by reweighting techniques (Kumar 1992; Chodera 2007). The main problem with these algorithms, however, is the non-trivial determination of the different multi-canonical weight factor by an iterative process involving short trial simulations. For complex system this procedure can be very tedious and attempts have been made to accelerate convergence of the interactive process (Berg and Celik 1992; Kumar 1996; Smith 1996; Hansmann 1997; Bartels 1998).

The replica exchange (REx) algorithm, developed as an extension of simulated tempering, removes the problem of finding correct weight factor. It belongs to the first category of algorithms where a composition of the original ensemble is sampled. The standard temperature REx algorithm is reviewed in the following section.

2.8.2 Temperature Replica Exchange

Consider a simulation system of N atoms of mass m_k $(k = 1, \ldots, N)$ with their coordinate and velocity vectors denoted by $x := (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{3N}, \mathbf{x}_i \in \mathbb{R}^n$ and $v := (\mathbf{v}_1, \ldots, \mathbf{v}_N), \mathbf{v}_i \in \mathbb{R}^{3N}$, respectively. The Hamiltonian $H(\mathbf{x}, \mathbf{v}) = E(\mathbf{x}) + K(\mathbf{v})$ is given by the sum of the potential energy $E(\mathbf{x})$ and the kinetic energy $K(\mathbf{v}) = \sum_{k=1}^{N} m_k \nu_k^2/2$. In the canonical ensemble at temperature T, each state $s := (\mathbf{x}, \mathbf{v})$ with the Hamiltonian $H(s)$ has a probability given by its Boltzmann factor, with the inverse temperature and the Boltzmann constant. Via the equipartition theorem, the average kinetic energy is linked to the number of degrees of freedom of the system,

$$\langle K(v) \rangle_T = \frac{N_{df}}{2} k_B T \tag{2.10}$$

Usually $N_{df} \ll 3N$ since constraint algorithm (Miyamoto 1992; Hess 1997) considerably restrict the number of degrees of freedoms. When flexible bonds are simulated, $N_{df} = 3N$ and the standard textbook expression for a free N-particle system is recovered (Figure 2.1).

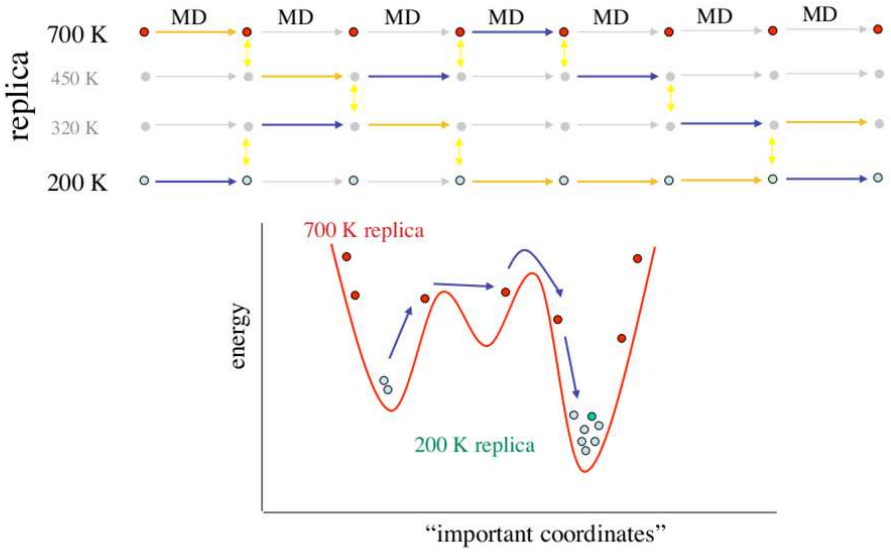

FIGURE 2.1: Replica Exchange scheme showing the interchange between *replica* 1 at temperature $T_1 = 200K$ and *replica* 2 at $T_2 = 700K$.

In REx a generalized ensemble is reconstructed, which consists of M+1 *non-interacting copies* (or *replicas*) of the original system in the canonical ensemble with temperatures $\{T_0, T_1, \ldots, T_M\}$ and $T_m = T_{m+1}\,(m = 0, \ldots, M)$. Equal temperatures for two or more replicas are possible, but seldomly used.

A state of this generalized ensemble is characterized by

$$S = \{\ldots, s_m^{[i]}, \ldots\} \tag{2.11}$$

Where the configuration $s_m^{[i]} = (x_m^{[i]}, v_m^{[i]})$ represents the coordinates $(x_m^{[i]})$ and velocities $(v_m^{[i]})$ of all atoms of the i_{th} replica at temperature T_m. Because the replicas are non-interacting, the product of Boltzmann factors for each replica gives the statistical weight of a state S of this generalized ensemble,

$$W(S) = \exp\big\{ - \sum_{m=0}^{M} \beta_m H(s_m^{[i]})\big\} \tag{2.12}$$

We now consider an exchange between a pair of replicas i and j,

$$S = \big\{\ldots, s_m^{[i]}, \ldots, s_n^{[j]}, \ldots\big\} \rightarrow S = \big\{\ldots, s_m^{[i]'}, \ldots, s_n^{[j]'}, \ldots\big\} \tag{2.13}$$

In more details, eq 2.13 reads

$$s_m^{[i]} = (x_m^{[i]}, v_m^{[i]}) \rightarrow s_n^{[i]'} = (x_n^{[i]'}, v_n^{[i]'}) \tag{2.14}$$

$$s_n^{[j]} = (x_n^{[j]}, v_n^{[j]}) \rightarrow s_m^{[j]'} = (x_m^{[j]'}, v_m^{[j]'}) \tag{2.15}$$

Unlike the original implementation of REx using Monte-Carlo updating steps (Swendsen 1986; Hukushima 1996; Tesi 1996), REx MD requires a rescaling of velocities, indicated by primes in 2.14 and 2.15. Velocity rescaling is done in such a way, that the equipartition theorem holds for each replica at all times. To this end, we look at the situation immediately after a temperature exchange $S \rightarrow S'$.

Starting with replica $s_m^{[i]} = (x_m^{[i]}, v_m^{[i]})$, the equipartition theorem reads

$$2\left\langle K(v_m^{[i]})\right\rangle_{T_m} = N_{df} k_B T_m$$

Upon exchange, replica receives the rescaled velocities of replica $s_m^{[i]}$, thus

$$2\left\langle K(v_n^{[i]'})\right\rangle_{T_n} = N_{df}k_BT_n$$

Combining both expression yields

$$2\left\langle K(v_n^{[i]'})\right\rangle_{T_n} = \frac{T_n}{T_m}\left\langle K(v_m^{[i]})\right\rangle_{T_m}. \tag{2.16}$$

Since $K(v) \sim v^2$, we arrive at the primed velocities

$$v_n^{[i]'} = \sqrt{\frac{T_n}{T_m}v_m^{[i]}}, \qquad v_m^{[j]'} = \sqrt{\frac{T_m}{T_n}v_n^{[j]}}, \tag{2.17}$$

where all atoms of the replicas are rescaled uniformly. For the exchange process to converge towards the equilibrium distribution (3.2), it is sufficient to impose the detailed-balance condition on the transition probability (exchange/acceptance probability) $S \to S'$,

$$W(S)P(S \to S') = W(S')P(S' \to S) \tag{2.18}$$

it is

$$\begin{aligned}\frac{(S \to S')}{(S' \to S)} &= \frac{\exp\{-\beta_m H(s_m^{[j]'})\}\exp\{-\beta_n H(s_n^{[i]'})\}}{\exp\{-\beta_m H(s_m^{[i]})\}\exp\{-\beta_n H(s_n^{[j]})\}} \\ &= \exp\{-\beta_m\big[H(s_m^{[j]'} - H(s_m^{[i]})\big] - \beta_n\big[H(s_n^{[i]'}) - H(s_n^{[j]})\big]\} \\ &= \exp\{\beta_m\big[E(x_m^{[i]} - E(x_m^{[j]})\big] - \beta_n\big[E(x_n^{[i]}) - E(x_n^{[j]})\big]\} \\ (\star) &= \exp\{(\beta_m - \beta_n)\big[E(x_m^{[i]} - E\big(x_n^{[j]}\big)\big]\}\end{aligned} \tag{2.19}$$

In the last step (*) we used the fact that the potential energy of the system immediately after the change solely depends on the respective conformation of the system and not on the temperature; thus, $x_m^{[i]} = x_n^{[j]}$ and $x_n^{[i]} = x_m^{[i]}$.

Detailed-balance can be satisfied by usual Metropolis Monte-Carlo criterion:

$$P(S \to S') = min\big\{1, exp\big\{\big(\beta_m - \beta_n\big) \mid E\big(x_m^{[i]}\big) - E\big(x_n^{[j]}\big) \mid \big\}\big\} \tag{2.20}$$

For simulations performed in the NPT-ensemble, above equation is modified by a pressure correction term (Okabe 2001). Putting together everything, a simulation using the REx algorithm is realized by alternately performing the following two steps:

- *Simultaneous* and *independent* simulation of each replica for a certain number of MD steps.
- Exchange of two replicas according to the Metropolis criterion 2.20.

In practice, only neighbouring replicas are exchanged since the acceptance probability 2.20 exponentially decreases with the temperature and potential energy difference. Within the generalized ensemble S a random walk in temperature space is performed, translating into a random walk in potential energy conformational sampling of the rugged free energy landscape of the system.

2.8.3 Algorithm performance

The appropriate choice of temperatures is crucial for an optimal performance of the REx algorithm. Depending on the problem under study, the properties of the system at the lowest temperature T_0 within the replica setup are usually of particular interest. Therefore, replica temperatures have to be chosen such that (a), the lowest temperature sufficiently samples energy states of interest; (b), the highest temperature is large enough to overcome energy barriers of the system; and (c), the acceptance probability is sufficiently high, requiring an adequate overlap of potential energy distributions for neighbouring replicas, see eq. 2.20. Protocols for this task can be found in (Okabe 2001; Fukunishi 2002; Sanbonmatsu 2002).

Once a temperature distribution is established, the frequency with which exchanges between replicas are attempted (exchange attempt frequency) needs to be fixed. There is some discussion within literature (Kofke 2002; Kofke 2004; Predescu 2005; Zuckerman 2006; Periole 2007; Zuckerman 2006) analysing the interplay between acceptance probability and sampling efficiency. Likewise, findings an appropriate criterion for judging REx efficiency is still a matter of controversy (Lyman 2006; Murdock 2006; Zuckerman 2006; Lyman 2007; Zuckerman 2006).

Besides its simplicity and ease of implementation, the REx algorithm has some advantages that are reflected in the widespread use of this method over the last few years. The main advantage of REx over other generalized ensemble method over the last few years. The main advantage of REx over other generalized ensemble method lies in the fact that the weight factor W(S) is known a priori and does not have to be determined

by a tedious and time-consuming procedure. Furthermore, each replica s_m pertains to a Boltzmann ensemble at replica temperature at T_m, that can be manipulated to obtain predicted observables (at this reference temperature).

REx is often used in the study of phase transition (Ferrenberg 1988)] such folding/unfolding simulations (Sanbonmatsu 2002; Zhou and Berne 2002; Garca 2003; Rao 2003; Zhou 2003; Kokubo 2004; Seibert 2005) or aggregation phenomena (Cecchini, Rao et al. 2004). In particular, the temperature dependence of calculated free energies $\Delta G = \Delta H - T\Delta S$ allows interferences about enthalpic and entropic contributions (Nguyen, Mu et al. 2005).

2.9 Free energy and Molecular Dynamics

The canonical thermodynamics ensemble describes a system, in which the temperature, the volume and the number of particles are kept fixed. Here we consider that the interactions between all N particles in the system are described by a classical Hamiltonian, consisting of the kinetic energy depending on the momenta $\mathbf{p}_i$ of the particles and the potential energy that depends on the positions of all particles $\mathbf{x}^N$,

$$H(\mathbf{p}^N, \mathbf{x}^N) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + Vx^N \tag{2.21}$$

The associated partition function is the (Mcquarrie June 2000)

$$Z_N(T, V) = \int d\mathbf{x}^N d^N\mathbf{p} e^{-\beta H(\mathbf{x}^N, \mathbf{p}^N)} \tag{2.22}$$

And the probability P to find the system in a particular configuration $\{\mathbf{x}^N, \mathbf{p}^N\}$ is

$$P(\{\mathbf{x}^N, \mathbf{p}^N\}) = \frac{e^{-\beta H(\mathbf{x}^N, \mathbf{p}^N)}}{Z_N} \tag{2.23}$$

The partition function of the system in connected to the Helmholtz free energy via

$$F(T, V, N) = -k_B T ln Z_N(T, N) \tag{2.24}$$

An important consideration is that when a system of N particles is in a state in which the temperature and the volume are constant, the system spontaneously evolves to minimize the (Helmholtz) free energy.

The free energy is the driving force of the system, it discriminates whether changes in the configuration of the system are spontaneous or require an external amount of energy to be realized. Although all the derivations here are carried out for the Helmholtz free energy, the calculations to obtain the Gibbs free energy in the case of an isobaric ensemble are formally equivalent.

2.9.1 Constrained free energy

As seen in eq. 2.22, the free energy depends on all degrees of freedom of the system, i.e., is a multivariate function. In many cases, however it is desirable to have a formulation that allows quantifying the impact of only a subset of parameters on the free energy.

A constrained free energy is defined exactly in this manner (Chaikin 2000), it expresses the value of the free energy of the whole system when a given observable of the system A(x,p) is constrained to a given generalized coordinate ξ. By interesting out the rest of degrees of freedom, the constrained partition function is

$$Z(\xi) = \int d\mathbf{x}^N d^N\mathbf{p}\delta(A(x,p) - \xi)e^{-\beta H\left(\mathbf{x}^N,\mathbf{p}^N\right)} \tag{2.25}$$

where δ is the Dirac δ-function. The corresponding constrained free energy F is then obtained by

$$\beta F(\xi) = -lnZ(\xi) \tag{2.26}$$

The probability to find the system in the state described by ξ is

$$P(\xi) = \frac{Z(\xi)}{Z_N} = \frac{1}{Z_N}\int d\mathbf{x}^N \delta(A(x,p) - \xi)e^{-\beta V(x)} \tag{2.27}$$

which can be rationalized as follow: the ratio between the phase space volume available for an observable constrained at ξ and the total available phase space is exactly the probability that the system is found at ξ, eqs. 2.26 and 2.27, provide a direct connection between the $F(\xi)$ and our molecular dynamics simulations,

$$\beta F(\xi) = -lnP(\xi) + \beta F \tag{2.28}$$

Which expresses the constrained free energy as the probability to find the system in ξ and a contribution from the total free energy, which is constant. Since we are interested in how the constrained free energy changes when the system evolves along ξ, the unknown term βF is not needed.

2.9.2 Potential of Mean Force is a constrained free energy

A useful constrained free energy is obtained by defining $A(\mathbf{x}, \mathbf{p}) = \mathbf{x}_1$, fixing a given particle to the position $\mathbf{x}_1$. The free energy required to move the particle from a reference state to the position $\mathbf{x}_1$ is known as the Potential of Mean Force (PMF) (Kirkwood 1935), and we denote it by $W(\mathbf{x}_1)$.

The PMF along a given pathway for x_1 generates a free energy profile, which is useful for a reduced representation of the dynamics of a system along that pathway. The PMF is, e.g., a basic ingredient for the Generalized Langevin Equation (Zwanzing 1954), which describes the dynamics evolution of a given subset of the total degrees of freedom. The PMF can be expressed as

$$e^{-\beta W(\mathbf{x}_1)} = C \int dx_2 \dots dx_N e^{-\beta V(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)} \tag{2.29}$$

Where the constant C denotes a reference state. The integration is done over all positions of the particles except the one of interest, x_1. The name of potential of mean force for this kind of constrained free energy is due to the fact that the derivative of $-W(x_1)$, with respect to x_1 is the mean average force exerted on the particle 1 located at position x_1, is the mean average force exerted on the particle 1 located at position x_1,

$$-\frac{\partial W(x_1)}{\partial x_1} = \langle F(x_1) \rangle \tag{2.30}$$

The equation provides another link between molecular dynamics simulations and constrained free energies, in this case $W(x_1)$. By integrating the average force exerted on particle x_1 from a reference state $x_1^{'}$ we obtain the free energy difference between these two states.

2.9.3 Methods to extract free energies from simulations

Extracting absolute free energies via eq 2.24 from simulations involving thousands of atoms is challenging since it requires the calculation of the partition function Z_N, something that is typically prohibitive in a high dimensional phase space. However, for most practical applications if suffices to determinate the relative free energy between two states, e.g., the free energy difference between an ion in the central channel position and the ion in the bulk. Relative constrained free energies can be readily computed from molecular dynamics simulations by employing eq. 2.27 for each state of interest and computing the difference,

$$\Delta F = -k_B T \ln \frac{P(\xi)}{P(\xi')} \tag{2.31}$$

Since we would like to obtain a position-dependent constrained free energy profile, we work with the potential of mean force, previously introduced. Also for convenience, we define our potential of mean force with respect to a reference point z_0 located in the bulk. In Cartesian coordinates, we express the potential of mean force as

$$e^{-\beta W(z_0)} = e^{-\beta W(z_0)} \frac{\int dx \int dy\, e^{-\beta V(x,y,z)}}{\int dx \int dy\, e^{-\beta V(x,y,z_0)}} \tag{2.32}$$

The integration carried out for x and y needs to be bound to be meaningful. At this aim, in chapter 3, to compute the free energy for transferring a DNA hairpin from water to carbon tetrachloride solvent, I consider a box composed by $water - CCl_4 - water$ and define the center of mass of CCl_4 as a reference group. I will assume that the hairpin moves along the z exe of a cylinder of radius r_1 and only atoms dynamically within the cylinder are considered. To ensure a well-defined reference state, in my calculations I restrict the integration outside the cylinder to a second cylinder defined with r_0, with a linear switch function that weights the contribution of atoms between r_0 and r_1 with distance. This smooths the effects of atoms moving in and out of the cylinder (which causes jumps in the pull forces applied to the center of mass of the hairpin to drag it from water to CCl_4). I finally define

$$\Delta W(z) = -k_B T ln \frac{P(z)}{P(z_0)} \tag{2.33}$$

where the notation $\Delta W(z)$ implicitly assumes that a reference state at z_0 is used to define the difference in (constrained) free energy. The relative probability to find a particle at position z with respect to the position z_0 can be extracted from simulations if

the system is able to sample all degrees of freedom. For PMF profiles with high barriers this is impractical, since probability to find the atom/molecule decreases exponentially with the barrier height.

There are several methods proposed to circumvent the lack of sampling in regions inaccessible due to its very low probability. Most of these methods overcome this limitation by guiding or confining the system along a given reaction coordinate. The external bias is designed such that its effect on the underlying free energy can be removed or accounted for.

2.9.4 Free energy from enforced equilibrium simulation

In transitions between states that occur infinitely slow, the work reversible W performed by (or against) the system is exactly the free energy difference ($W = \Delta F$). Since a free energy is a state function, several methods use conveniently chosen reaction coordinates to drive the system from state A to state B. By performing this transition infinitely slow, and recording either the mean force or a probability distribution, we can calculate ΔF. A potential of mean force W(x_1) can also be obtained if the transitions represent a succession of states located infinitely close along the reaction coordinate. Here I present a brief list of the most methods,

- **Free energy perturbation** (Zwanzig 1954; Straatsma 1992) (FEP), thermodynamic integration (Kirkwood 1935) (TI) and related methods (Meirovitch 2007) Compute the reversible work associated to a perturbation (geometrical or of any other type) in the Hamiltonian. These methods are typically used to compute free energies changes related to chemical modifications, which are then manipulated using thermodynamic cycles to obtain experimental observables. For example from the free energy associated to mutations in the sequence of DNA in both single strand and duplex states estimates of the stabilizing effect of mutation can be derived. As relative free energies are computed along pathways without chemical sense, this type of calculations are often named as computational alchemy).

- **Particle insertion method** (Widom 1963). This method uses ghost particle to randomly probe the interaction energy a real particle would have in selected configurations. It can be used, e.g., to compute the free energy cost of inserting a water molecule in specific locations of a lipid membrane.

- **Constrained force simulations** (McDonald 1975; Carter 1989). The system is simulated at discrete positions along the reaction coordinate, and displacements from the reference position are avoided by continuously resetting the particle to

its predefined positions. The average forces exerted on the particle are recorded; integration along the reaction coordinate yields a discrete PMF.

- **Umbrella sampling** (Torrie 1974). The position of the particle along the reaction coordinate is confined around predefined positions by an external potential. The position along the reaction coordinates and the strength of external potential are chosen to ensure that the system completely samples the reaction coordinate. Since this method is extensively used in this thesis, it will be introduced in more detail below.

2.10 Free Energy from non-equilibrium simulations

The work performed when the transition between different states of the system is not carried out under conditions of equilibrium is due to the free energy difference and a dissipative term, $W = \Delta F + W_{diss}$. To compute the free energy from the work distribution, the Jarzynski equality (Jarzynski 1997), a special case of the Crooks theorem (Crooks 2000) is typically used. The Jarzynski equation,

$$e^{-\beta \Delta F} = \left\langle e^{-\beta \Delta F} \right\rangle \tag{2.34}$$

where the $\langle \dots \rangle$ denotes the ensemble average of the work under non-equilibrium conditions, allows to compute free energies from multiple non-equilibrium simulations, each contributing a work of W to the ensemble average.

Direct application of eq. 2.34 is however impractical, since it requires the sampling of regions of the probability distribution P(W) with negative dissipative work, which are very rare. To alleviate this problem, a common approach to obtain the free energy is to use a cumulative expansion of $\left\langle e^{-\beta \Delta F} \right\rangle$ up to second order (Zwanzig 1954; Jarzynski 1997; Gore, Ritort et al. 2003), such that the free energy reads

$$\begin{aligned} \Delta F &= \langle W \rangle - \frac{1}{2} \beta \left\langle (W - (\langle W \rangle)^2 \right\rangle \\ &= \langle W \rangle - \frac{1}{2} \beta \sigma_W^2 \end{aligned} \tag{2.35}$$

Where σ_W^2 is the variance of the work distributions P(W) and $\langle W \rangle$ is the average work.

Another possibility is the use of the transient fluctuation theorem (Evans and Searles 1994; Evans and Searles 1996; Crooks 2000), which connects the work distribution of forward and backward transitions between state A and state B to the dissipative work (note that $W - \Delta F = W_{diss}$),

$$\frac{P(W)\{A \rightarrow B\}}{P(W)\{A \leftarrow B\}} = e^{-\beta(W-\Delta F)} \tag{2.36}$$

From the eq. 2.36 follows that the value of W for which the two probability distributions, $P(W)\{A \rightarrow B\}$ and $P(W)\{A \leftarrow B\}$ are equal, corresponds to the free energy difference ΔF between these two states.

Applications that makes use of non-equilibrium trajectories are, e.g., force probe molecular dynamics (FPMD) simulations (Grubmuller, Heymann et al. 1996), where the system of interest is guided via a harmonic potential. During the simulation, the harmonic potential is moved along the reaction coordinate, and the mechanical work is recorded. By performing multiple FPMD simulations a work distributions $P(W)$ is collected, which can be translated into a free energy by means of eq. 2.34 or eq. 2.35. Alternatively, several FPMD simulations can be performed from $A \rightarrow B$ and $B \leftarrow A$, and use eq. 2.36 to obtain ΔF.

Another method that requires non-equilibrium equalities is fast-growth Termodynamics Integration (Hukushima 1996; Hendrix 2001), since the transition between states are enforced without allowing sufficient relaxation times for the system, as opposite to the traditional (slow-growth) TI.

2.10.1 Umbrella Sampling

A widely used approach to deal with rare events, such as the permeation of a DNA hairpin from water to carbon tetrachloride bulk discussed in chapter 3, is to enforce the system to sample all the position along a previously defined reaction coordinate ξ, which is discretized in a number of bins N_w with reference position at ξ_i.

The method known as umbrella sampling (US), proposed by Torrie and Valleu (Torrie 1974), uses an external biasing potential w_i, to confine the atom/molecule around an established position ξ_i. A common form of biasing potential w_i is an harmonic potential,

$$w_i = \frac{1}{2}(\xi - \xi_i)^2 \tag{2.37}$$

which restrains the atom/molecule of interest in the vicinity of ξ_i. Therefore, the potential energy acting on a particle restrained at ξ_i is due to $V(x) + w_i(\xi)$. The probability to find the particle at position ξ_i is

$$P_i^{bias}(\xi) = \frac{\int d^N x\, \delta(\xi(x) - \xi) e^{-\beta(V(x)+w_i(\xi))}}{\int d^N x\, e^{-\beta(V(x)+w_i(\xi))}} \tag{2.38}$$

which can be rewritten as

$$P_i^{bias}(\xi) = e^{-\beta w_i(\xi)} \frac{\int d^N x \delta(\xi(x) - \xi) e^{-\beta V(x)}}{\int d^N x e^{w_i(\xi)} e^{-\beta V(x)}} \tag{2.39}$$

or

$$P_i^{bias}(\xi) = e^{-\beta w_i(\xi)} \frac{P(\xi)}{\left\langle e^{-\beta w_i(\xi)} \right\rangle} \tag{2.40}$$

where $\langle \dots \rangle$ is a short hand notation for the ensemble average of $e^{-\beta w_i(\xi)}$, and $P(\xi)$ is a constrained probability of the system in the absence of bias. By obtaining $P(\xi)$ we can extract the potential of mean force by means of eq. 2.33 up to an additive constant due to a reference state $P(\xi_0)$. More explicitly,

$$W(\xi) = -k_B T ln \frac{P(\xi)}{P(\xi_0)} - w_i(\xi) + F_i + W(\xi_0) \tag{2.41}$$

where we have defined $F_i = \left\langle e^{-\beta w_i(\xi_i)} \right\rangle$. Apart from the arbitrary reference $W(\xi_0)$ and F_i, we can extract all the ingredients needed from the biased simulations. The constant Fi reflects the free energy required to introduce the bias into the system

Using WHAM to reconstruct the unbiased probability density

Although several methods have been proposed to determine Fi, the method that is currently considered the most suitable (Roux 1995) is the so-called *weighted histogram analysis method* (WHAM) (Kumar 1992). From N_w umbrella sampling simulations along the reaction coordinate ξ, the WHAM method expresses the overall bias-free profile of probability distributions $P_i(\xi)$ by minimizing the statistical error in recombining all $P_i(\xi)$. The overall probability is expressed as,

$$P(\xi) = \sum_{i=1}^{N_w} P_i(\xi) \frac{n_i e^{-\beta(w(\xi)-F_i)}}{\sum_{j=1}^{N_w} n_j e^{-\beta(w_j(\xi)-F_j)}} \tag{2.42}$$

where n_i is the number of points in the histogram used to construct $P_i(\xi)$. Using eq. 2.41 and $\left\langle e^{-\beta w_i(\xi_i)} \right\rangle$ we obtain

$$P(\xi) = \sum_{i=1}^{N_w} e^{-\beta w_i(\xi)} P_i(\xi) e^{-\beta F_i} \frac{n_i e^{-\beta(w(\xi)-F_i)}}{\sum_{j=1}^{N_w} n_j e^{-\beta(w_j(\xi)-F_j)}} \tag{2.43}$$

The WHAM equations are

$$P(\xi) = \sum_{i=1}^{N_w} P_i(\xi) \frac{n_i e^{-\beta(w(\xi)-F_i)}}{\sum_{j=1}^{N_w} n_j e^{-\beta(w_j(\xi)-F_j)}} \tag{2.44}$$

$$e^{-\beta F_i} = \int d\xi \, e^{-\beta w_i(\xi)} P(\xi) \tag{2.45}$$

These two equations are solved in a self-consistent manner for F_i. In practice, this is achieved by formulating an initial guess for F_i and iterating the WHAM equation until the difference between $\mid F_i - F_i^{last} \mid$ is below a given threshold.

2.11 Analysis of the results: structural analysis

Root Mean Square deviation

A MD trajectory contains detailed information of the temporal fluctuations of the system. This information can be processed globally or locally depending on the level of study. For example, one can calculate a global descriptor such as the root mean square deviation (RMSD) with respect to a reference structure, usually the original one or in other cases the average along a part or the entire simulation or only a part of it:

$$RMSD = \left[\frac{1}{M} \sum_{l=1}^{3N} m_l (x_{kl} - x_l)^2 \right]^{1/2} \tag{2.46}$$

where M corresponds to the total mass, m_l is the atomic mass associated to l, x_{kl} represents the l-coordinate in the structure k, x_l the reference value in the reference structure and N is the number of atoms.

DNA parameters

In nucleic acids, the structural analysis requires the study of geometric parameters which usually include base pair relative positions (3 rotations and 3 translations) backbone torsions and groove volume analysis among others. These physical parameters are computed according to the Tsukuba convention (Olson, Bansal et al. 2001) with a variety of software programs, e.g., *Curves+* and 3DNA (Lu and Olson 2003; Lavery, Moakher et al. 2009) that have been used in this PhD.

Hydrogen bonds

Hydrogen bonds are computed with the program $g_h bonds$ from GROMACS.4.5. package, that analyzes the hydrogen bonds (H-bond) between all possible donors and acceptors. To determine if an H-bond exists, a geometrical criterion is used:

$$r \leq r_{HB} = 0.35nm$$

$$\alpha \leq \alpha_{HB} = 30^{\circ}$$

The value of $r_{HB} = 0.35$ nm corresponds to the first minimum of the EDF or SPC water. In the gas phase H-bonds are computed fulfilling only the first condition between canonical acceptors and donors.

Stacking interactions are usually classified in intramolecular and intermolecular interactions while hydrogen bonding interactions involve opposing bases.

2.11.1 Geometrical parameters

The Radius of gyration and the Collision Cross Section of nucleic acids have been estimated in this thesis as important parameters to compare the shape and the geometry of the structure in the gas phase with the structure in solution.

The Radius of Gyration

The *Radius of Gyration* gives a measure for the compactness of the structure, it is calculated with the program $g_g yrate$ from GROMACS.4.5 package as follows:

$$R_g = \Big(\frac{\sum_i \mid r_i \mid^2 m_i}{\sum_i m_i}\Big)^{1/2}$$

where m_i is the mass of atom i and r_i the position of atom i with respect the center of mass of the molecule.

The Collision Cross Section

Of the various methods used to study gas-phase biomolecules, measurements of collision cross sections give the most direct insights to folding (Hoaglund-Hyzer, Counterman et al. 1999).

The collision cross section is a measure of ion size, which can provide direct evidence for changes of higher order structure of ions of gas-phase bio-molecule like protein and DNA. Collision theory (Mason 1988) predicts a direct dependence of the collision frequency on the cross section of an ion. Thus, ions of two different sizes (i.e. different conformers) are expected to exhibit different behaviour in a collision cell. Larger unfolded ions (higher charge states) will have higher collision frequencies and correspondingly will lose more axial kinetic energy than smaller folded ions (Cox, Julian et al. 1994). Therefore, measurements that depend on the number of collisions that are experienced by different conformers can be used to discriminate between conformer shapes.

Ion mobility methods

Cross section can be determined by different techniques (Covey and Douglas 1993; Reimann 1998), The ion mobility, used in this thesis, is the highest resolution methods able to separate ions of a biopolymer that have the same mass to charge ratio but different conformations. The mobility of a gas phase ion is a measure of how fast it moves through a buffer gas under the influence of an electric field and the ccs is the probability surface of interaction between the gas phase ion and the atoms that constitute the buffer gas.

For calculation of ion mobility a variety of schemes have been developed ranging from a simple projection to methods based on trajectory calculations. The method used in this thesis are the **Projection Approximation** (PA) method and the **Exact Hard-Sphere Scattering** (EHSS) and I will describe them briefly in following sections.

In the **Projection Approximation** (von Helden 1993; Wyttenbach 1997; Wyttenbach 2000) the molecule is projected onto a randomly chosen plane in space, and a circle with the corresponding collision radius is drawn in that plane at the position of each projected atom. Then points in the plane are randomly picked within a square of area A that encloses the projected molecule. If a selected point is inside one or more

circles, it is counted as a hit. The ratio of hits to the number of tries multiplied by A is the collision cross section of that particular projection. The procedure is repeated for many different randomly selected projection

The **Exact Hard-Sphere Scattering** (Shvartsburg 1996) developed by Shvartsburg and Jarrold include the contribution of the scattering angle to cross section. In this method the averaged collision integral is solved numerically.

$$\Omega_{avg} = \frac{1}{4\pi^2} \int_0^{2\pi} \mathrm{d}\theta \int_0^{\pi} \mathrm{d}\phi sin\phi \int_0^{2\pi} \mathrm{d}\gamma X \int \mathrm{d}b 2b(1 - cos\chi(\theta, \phi, \gamma, b))$$

The scattering angle, $\chi(\theta, \phi, \gamma, b)$ is determined by following a trajectory through any all of its collision with the polyatomic ion. This model considers the details of the collision process, however, the effect of long-range potential has not been taken into account in this approach.

Bibliography

Adcock, S. A., McCammon, J. A. (2006). "Molecular dynamics: survey of methods for simulating the activity of proteins." Chem Rev 106(5): 1589-1615.

Alhambra, C., Luque, F. J., Gago, F., and Orozco, M. (1997). "Ab initio study of stacking interactions in A- and B-DNA." J. Phys. Chem. B. 101: 38463853.

Anderson, H. C. (1980). "Molecular dynamics simulations at constant pressure and/or temperature." J. Chem. Phys. 72: 23842393.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science 181: 223230.

Bartels, C., and, Karplus, M. (1998). "Probability distributions for complex systems: Adaptive umbrella sampling of the potential energy." J. Phys. Chem. B. 102: 865880.

Berendsen, H. J. C., Postma, J. P. M., Di Nola, A., and, Haak, J. R. (1984). "Molecular dynamics with coupling to an external bath." J. Chem. Phys. 81: 36843690.

Berendsen, W. F. v. G. a. H. J. C. (1990). "Computer-simulation of molecular-dynamics - methodology, applications, and perspectives in chemistry." Angew. Chem. Int. Edit. Engl. 29: 9921023.

Berg, B. A., and, Neuhaus, T. (1991). "Multicanonical algorithms for first-order phase transitions." Phys. Lett. 267: 249253.

Berg, B. A. and T. Celik (1992). "New approach to spin-glass simulations." Phys Rev Lett 69(15): 2292-2295.

Berk Hess, v. d. S., D., and Erik Lindahl. (2010). "Gromacs user manual version 4.5.6." website: http://www.gromacs.org.

Beveridge, D. L., and, McConnell, K. J. (2000). "Nucleic acids: theory and computer simulation." Y2K. Curr. Opin. Struct. Biol. 10: 182196.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and, Karplus, M. (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations." J. Comp. Chem. 4: 187217.

Brooks III, C. L., Pettitt, B. M., and, Karplus, M. (1985). "Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids." J. Chem. Phys 83: 58975908.

Bruant, N., D. Flatters, et al. (1999). "From atomic to mesoscopic descriptions of the internal dynamics of DNA." Biophys J 77(5): 2366-2376.

Carter, E. A., and, Ciccotti, G. Hynes, J.T., and, Kapral, R. (1989). "Constrained reaction coordinate dynamics for the simulation of rare events. ." Chem. Phys. Lett. 156: 472477.

Cecchini, M., F. Rao, et al. (2004). "Replica exchange molecular dynamics simulations of amyloid peptide aggregation." J Chem Phys 121(21): 10748-10756.

Chaikin, P. M., Lupensky, T.C. (2000). Principles of Consensed Matter Physics. Canbridge.

Cheatham III, T. E., and, Kollman, P. A. (1997). "Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]2 in the presence of hexaamminecobalt(III)." Structure (London) 5: 12971311.

Cheatham, T. E., 3rd, M. F. Crowley, et al. (1997). "A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions." Proc Natl Acad Sci U S A 94(18): 9626-9630.

Cheatham, T. E., 3rd and M. A. Young (2000). "Molecular dynamics simulation of nucleic acids: successes, limitations, and promise." Biopolymers 56(4): 232-256.

Cheatham, T. E., III (2004). "Simulation and modeling of nucleic acid structure, dynamics and interactions." Curr. Opin. Struct. Biol 14: 360367.

Cheatham, T. E., III, and Kollman, P. A. (2000). "Molecular dynamics simulation of nucleic acids." Annu. Rev. Struct. Dyn. 51(435471).

Cheatham, T. E., III, Cieplak, P., and, Kollman, P. A. (1999). "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat." J. Biomol. Struct. Dyn. 16: 845862.

Cheatham, T. E. K., P. A. (1996). "Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. ." J Mol Biol 259: 434-444.

Chodera, J. D., Swope, W. C., Pitera, J. W., Seok, C., and, Dill, K. A. (2007). "Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations." J. Chem. Theory Comput. 3: 2641.

Choen, M. J., Karasek, F.W. (1970). "Plasma Chromatography-a new dimension for Gas Chromatography and Mass Spectrometry." J. Chromatogr. Sci. 8: 330-337.

Clemmer, D. E., and, Jarrold, M.F. (1997). "Ion Mobility Measurements and their Applications to Clusters and Biomolecules." J. Mass Spectrom. 32: 577-592.

Clemmer, D. E., Hudgins, R.R., and, Jarrold, M.F. (1995). "Naked Protein Conformations: Cytochrome C in the Gas Phase." J. Am. Chem. Soc. 117: 10141-10142.

Cornell, W. D., P. Cieplak, C. I. Baily, I. R. Gould, K. M. Merz, Jr., D. C. Ferguson, T. Fox, J. W. Caldwell, and P. A. Kollman. (1995). "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules." J. Am. Chem. Soc. 117: 51795197.

Covey, T. and D. J. Douglas (1993). "Collision cross sections for protein ions." J Am Soc Mass Spectrom 4(8): 616-623.

Cox, K. A., R. K. Julian, et al. (1994). "Conformer selection of protein ions by ion mobility in a triple quadrupole mass spectrometer." J Am Soc Mass Spectrom 5(3): 127-136.

Crooks, G. E. (2000). "Path-ensemble averages in systems driven far from equilibrium." Phys. Rev. E 61(3): 23612366.

Darden, T., York, D., and, Pedersen, L. (1993). "Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems." J. Chem. Phys. 98: 1008910092.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and, Pedersen, L. G. (1995). "A smooth particle mesh Ewald potential." J. Chem. Phys. 103: 85778592.

Evans, D. J. and D. J. Searles (1994). "Equilibrium microstates which generate second law violating steady states." Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 50(2): 1645-1648.

Evans, D. J. and D. J. Searles (1996). "Causality, response theory, and the second law of thermodynamics." Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 53(6): 5808-5815.

Ferrenberg, A. M., and, Swendsen, R.H. (1988). "New Monte Carlo technique for studying phase transitions." Phys. Rev. Lett. 61: 26352638.

Foloppe, N., and, Mackerell, A. D. (2000). "All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data." J. Comput. Chem. 21: 86104.

Frauenfelder, H. and D. T. Leeson (1998). "The energy landscape in non-biological and biological molecules." Nat Struct Biol 5(9): 757-759.

Frauenfelder, H., S. G. Sligar, et al. (1991). "The energy landscapes and motions of proteins." Science 254(5038): 1598-1603.

Frenkel, D., and, Smit, B. (2002). Understanding molecular simulation: from algorithms to applications.

Frenkel, D., Smit, B. (2002). "Understanding Molecular Simulation." Academic Press.

Fukunishi, H., Watanabe, O.and,Takada, S. (2002). "On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction." J. Chem. Phys. 116: 90589067.

Garca, A. E., and, Onuchic, J. N. (2003). "Folding a protein in a computer: An atomic description of the folding/unfolding of protein." Proc. Natl. Acad. Sci. 100: 1389813903.

Giudice, E., and, Lavery, R. (2002). "Simulations of nucleic acids and their complexes." Acc. Chem. Res. 35: 350357.

Gore, J., F. Ritort, et al. (2003). "Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements." Proc Natl Acad Sci U S A 100(22): 12564-12569.

Grubmuller, H., B. Heymann, et al. (1996). "Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force." Science 271(5251): 997-999.

Hagen, D. F. (1979). "Characterization of isomeric compounds by gas plasma chromatography." Anal. Chem. 51(7): 870-874.

Hansmann, U. H. E. (1997). "Effective way for determination of multicanonical weights." Phys. Rev. E 56: 62006203.

Hansson, T., C. Oostenbrink, et al. (2002). "Molecular dynamics simulations." Curr Opin Struct Biol 12(2): 190-196.

Hendrix, D. A., and, Jarzynski, C. (2001). "A "fast growth" method of computing free energy differences." J. Chem. Phys. 114(14): 59745981.

Hess, B., Bekker, H., Berendsen, H. J. C., and, Fraaije, J. G. E. M. (1997). "LINCS: A linear constraint solver for molecular simulations." J. Comp. Chem. 18: 14631472.

Hoaglund-Hyzer, C. S., A. E. Counterman, et al. (1999). "Anhydrous protein ions." Chem Rev 99(10): 3037-3080.

Hobza, P., Kabelac, M., Sponer, J., Mejzlik, P., and, Vondrasek, J. (1997). "Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO / M, PM3), and ab initio Hartree-Fock method for interaction of DNA bases: comparison with nonempirical beyond Hartree-Fock results. ." J. Comp. Chem. 18: 11361150.

Hockney, R. W., Goel, S. P., and, Eastwood, J. W. (1973). "A 10000 particle molecular dynamics model with long-range forces." Chem. Phys. Lett. 21: 589591.

Hukushima, K., and, Nemoto, K. (1996). "Exchange Monte Carlo method and application to spin glass simulations." J. Phys. Soc. (Jap.) 65: 16041608.

Iba, Y. (2001). "Extended ensemble Monte Carlo." Int. J. Mod. Phys. C. 12: 623656.

Jarrold, M. F. (1995). "Drift Tube Studies of Atomic Clusters." J. Phys. Chem. B. 99: 11-21.

Jarzynski, C. (1997). "Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach." Phys. Rev. E 56: 50185035.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and, Klein, M.L. (1983). "Comparison of Simple Potential Functions for Simulating Liquid Water. ." J. Chem. Phys. 79: 926.

Kast, S. M., and, Brickmann, J. (1996). "Constant temperature molecular dynamics simulations by means of a stochastic collision model. 2. the harmonic oscillator." J. Chem. Phys. 104: 37323741.

Kast, S. M., Nicklas, K., Br, H.-J., and, J. Brickmann (1994). "Constant temperature molecular dynamics simulations by means of a stochastic collision model. 1. noninteracting particles." J. Chem. Phys. 100:: 566576.

Kirkwood, J. G. (1935). "Statistical Mechanics of fluid mixture." J. Chem. Phys. 3(5): 300-313.

Kofke, D. A. (2002). "On the acceptance probability of replica-exchange Monte Carlo trials." J. Chem. Phys. 117: 69116914.

Kofke, D. A. (2004). "Erratum: On the acceptance probability of replica-exchange Monte Carlo trials." J. Chem. Phys. 120: 10852.

Kokubo, H., and, Okamoto, Y. (2004). "Prediction of membrane protein structures by replica-exchange Monte Carlo simulations: Case of two helices. ." J. Chem. Phys. 120: 1083710847.

Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and, Rosenberg, J. M. (1992). "The weighted histogram analysis method for free-energy calculations on

biomolecules. I. the method." J. Comp. Chem. 13: 10111021.

Kumar, S., Payne, P.W., and, Vsquez, M. (1996). "Method for free-energy calculations using iterative techniques." J. Comput. Chem. 17: 12691275.

Lavery, R., M. Moakher, et al. (2009). "Conformational analysis of nucleic acids revisited: Curves+." Nucleic Acids Res 37(17): 5917-5929.

Leach, A. R. (2001). Molecular Modelling: Principles and Applications. . England, Essex, second edition.

Levitt, M. (1983). "Computer simulation of DNA double-helix dynamics." Cold Spring Harb Symp Quant Biol 47 Pt 1: 251-262.

Liu, H., H. P. Spielmann, et al. (1995). "Interproton distance bounds from 2D NOE intensities: effect of experimental noise and peak integration errors." J Biomol NMR 6(4): 390-402.

Liu, Y., Valentine, S. J., Counterman, A. E., Hoaglund, C. S., Clemmer, D. E. (1997). "Injected-ion Mobility Analysis of Biomolecules." Analytical Chemistry 69: 728A.

Lu, X. J. and W. K. Olson (2003). "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures." Nucleic Acids Res 31(17): 5108-5121.

Lyman, E. a., Zuckerman, D.M. (2006). "Ensemble-based convergence analysis of biomolecular trajectories." Biophys. J. 91: 164172.

Lyman, E. a., Zuckerman, D.M. (2007). "The structural de-correlation time: A robust statistical measure of convergence of biomolecular simulations." J Phys Chem B 111(44): 12876-12882.

Lyubartsev, A. P., Martinovski, A. A., Shevkunov, S. V., and, Vorontsov-Velyaminov, P. N. (1992). "New approach to Monte Carlo calculations of the free energy: Method of expanded ensembles." J. Chem. Phys. 96: 17761783.

MacKerell, A. D., Wiorkiewicz-Kuczera, J., and, Karplus, M. (1995). "An all-atom empirical energy function for the simulation of nucleic acids." J. Am. Chem. Soc. 117: 1194611975.

Marinari, E., and, Parisi, G. (1992). "Simulated tempering: A new Monte Carlo scheme." Europhys. Lett. 19: 451458.

Mason, E. A., McDaniel, E.W. (1988). Transport Properties of Ions in Gases. New York.

Matsumoto, A. and W. K. Olson (2002). "Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level." Biophys J 83(1): 22-41.

McCammon, J. A. (1976). Models for Protein Dynamics Orsay (France).

McCammon, J. A., B. R. Gelin, et al. (1977). "Dynamics of folded proteins." Nature 267(5612): 585-590.

McCammon, J. A., Harvey, S. C. (1987). Dynamics of Proteins and Nucleic Acids. Cambridge.

McConnell, K. J. and D. L. Beveridge (2000). "DNA structure: what's in charge?" J Mol Biol 304(5): 803-820.

McDonald, I. R., and, Rasaiaha, J.C. (1975). "Monte carlo simulation of the average force between two ions in a stockmayer solvent." Chem. Phys. Lett. 34: 382386.

Mcquarrie, D. A. (June 2000). Statistical Mechanics.

Meirovitch, H. (2007). "Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation." Curr Opin Struct Biol 17(2): 181-186.

Mitsutake, A., Sugita, Y., and, Okamoto, Y. (2001). "Generalized-ensemble algorithms for molecular simulations of biopolymers." Biopolymers 60: 96123.

Miyamoto, S., and, Kollman, P. A. (1992). "SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models." J. Comp. Chem. 13: 952962.

Murdock, S. E., Tai, K., Ng, M. H., Johnston, S., Wu, B., Fangohr, H., Laughton, C. A., Essex, J. W., and, Sansom (2006). "Sansom. Quality assurance for biomolecular simulations." J. Chem. Theory Comput. 2: 14771481.

Nguyen, P. H., Y. Mu, et al. (2005). "Structure and energy landscape of a photoswitchable peptide: a replica exchange molecular dynamics study." Proteins 60(3): 485-494.

Nose, S. (1984). "A unified formulation of the constant temperature molecular dynamics method." J. Chem. Phys. 81: 511519.

Okabe, T., Kawata, M. Okamoto, Y., and, Mikami, M. (2001). "Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble." Chem. Phys. Lett. 335:: 435439.

Olson, W. K. (1996). "Simulating DNA at low resolution." Curr Opin Struct Biol 6(2): 242-256.

Olson, W. K., M. Bansal, et al. (2001). "A standard reference frame for the description of nucleic acid base-pair geometry." J Mol Biol 313(1): 229-237.

Olson, W. K. and V. B. Zhurkin (2000). "Modeling DNA deformations." Curr Opin Struct Biol 10(3): 286-297.

Orozco, M., A. Perez, et al. (2003). "Theoretical methods for the simulation of nucleic acids." Chem Soc Rev 32(6): 350-364.

Orozco, M., Prez, A., Noy, A., and, Luque, F. J. (2003). "Theoretical methods for the simulation of nucleic acids." Chem. Soc. Rev. 32: 350364.

Orozco,M. Chem.Soc.Rev., 2014 In Press

Prez, A., Blas, J. R., Rueda, M., Lpez-Bes, J. M., de la Cruz, X., and, Orozco, M. (2005). "Exploring the essential dynamics of B-DNA." J. Chem. Theor. Comput. 1: 790800.

Perez, A., I. Marchan, et al. (2007). "Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers." Biophys J 92(11): 3817-3829.

Prez, A., Sponer, J., Jurecka, P., Hobza, P., Luque, F. J., and, Orozco, M. (2005.). "Are the RNA(AU) hydrogen bonds stronger than the DNA(AT) ones?" Chem. Eur. J. 11: 50625066.

Periole, X., and, Mark, A. E. (2007). "Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent." J. Chem. Phys. 126: 014903.

Predescu, C., Predescu, M.and, Ciobanu, C.V. (2005). "On the efficiency of exchange in parallel tempering Monte Carlo simulations." J. Phys. Chem. B 109: 41894196.

Radi, P. P., von Helden, G., Hsu, M. T., Kemper, P. R., and, Bowers, M. T. (1991). "Thermal Bimolecular Reactions of Size Selected Transition Metal Cluster Ions: Nbn+ + O2, n = 1-6." Int. J. Mass Spectrom. Ion Proc. 109: 49-73.

Rao, F., and, Caflisch., A. (2003). "Replica exchange molecular dynamics simulations of reversible folding." J. Chem. Phys. 119: 40354042.

Reimann, C. T., Sullivan, P. A., Axelsson, J., Quist, A. P., Altmann, S., Roepstorff, P., Velzquez, I., and, Tapia, O. (1998). "Conformation of Highly-Charged Gas-Phase Lysozyme Revealed by Energetic Surface Imprinting." J. Am. Chem. Soc. 120(30): 7608-7616.

Roux, B. (1995). "The calculation of the potential of mean force using computer simulations." Computer Physics Communications 91: 275282.

Rueda, M., Kalko, S. G., Luque, F. J., and, Orozco, M. (2003). "The structure and dynamics of DNA in the gas phase." J. Am. Chem. Soc. 125: 80078014.

Rueda, M., F. J. Luque, et al. (2005). "Nature of minor-groove binders-DNA complexes in the gas phase." J Am Chem Soc 127(33): 11690-11698.

Rueda, M., F. J. Luque, et al. (2006). "G-quadruplexes can maintain their structure in the gas phase." J Am Chem Soc 128(11): 3608-3619.

Ryckaert, J. P., Ciccotti, G., Berendsen, H. J. C. (1977). "Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes " Journal of Computational Physics 23(3): 327-341.

Sanbonmatsu, K. Y., and, Garca, A.E. (2002). "Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics." Proteins 46: 225234.

Seibert, M. M., Patriksson, A., Hess, B., and, van der Spoel., D. (2005). "Reproducible polypeptide folding and structure prediction using molecular dynamics simulations." J. Mol. Biol. 354: 173183.

Shelimov, K. B., and, Jarrold, M.F. (1996). ""Denaturation" and Refolding of Cytochrome C in Vacuo." J. Am. Chem. Soc. 118: 10313-10314.

Shelimov, K. B., and, Jarrold, M.F. (1997). "Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas Phase Protein Folding." J. Am. Chem. Soc. 119: 2987-2994.

Shelimov, K. B., Clemmer, D.E., Hudgins, R.R. and, Jarrold, M.F. (1997). "Protein Structure in Vacuo: The Gas Phase Conformations of BPTI and Cytochrome C." J. Am. Chem. Soc. 119: 2240-2248.

Shields, G. C., Laughton, C. A., and, Orozco, M. (1997). "Molecular dynamics simulations of the d(TAT) triple helix." J. Am. Chem. Soc. 119: 74637469.

Shvartsburg, A. A., and, Jarrold, M.F. (1996). "An Exact Hard Spheres Scattering Model for the Mobilities of Polyatomic Ions." Chem. Phys. Lett. 261: 86-91.

Smith, D. E., Dang, Liem, X. (1994). "Computer simulations of NaCl association in polarizable water." Journal of Chemical Physics 100 (5): 3757-3766.

Smith, G. R., and, Bruce, A.D. (1996). "Multicanonical Monte Carlo study of solid-solid phase coexistence in a model colloid." Phys. Rev. E 53: 65306543.

Smith, R. D., Barinaga, C.J. (1990). "Internal energy effects in the collision-induced dissociation of large biopolymer molecular ions produced by electrospray ionization tandem mass spectrometry of cytochrome c." Rapid Communications in Mass Spectrometry 4(2): 54-57.

Soliva, R., F. J. Luque, et al. (1999). "Role of sugar re-puckering in the transition of A and B forms of DNA in solution. A molecular dynamics study." J Biomol Struct Dyn 17(1): 89-99.

Sponer, J., Jurecka, P., Marchan, I., Luque, F. J., Orozco, M., and, Hobza, P. (2006). "Nature of base stacking. Reference quantum chemical stacking energies in ten unique B-DNA base pair steps." Chem. Eur. J. 12: 28542865.

Sponer, J., Jurecka, P., and, Hobza, P. (2004). "Accurate interaction energies of hydrogen-bonded nucleic acid base pairs." J. Am. Chem. Soc. 126: 1014210151.

Sponer, J. E., Spackova, N., Leszczynski, J., and, Sponer, J. (2005). "Principles of RNA base pairing: structures and energies of the trans Watson-Crick/sugar edge base pairs." J. Phys. Chem. B. 109: 1139911410.

Sprous, D., Young, M. A., and, Beveridge, D. L. (1998). "Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: Theoretical considerations of the A double left right arrow B transition." J. Phys. Chem. B. 102: 46584667.

Steinbach, P. J., and, Brooks, B.R. (1994). "New spherical-cutoff methods for long-range forces in macromolecular simulation." J. Comp. Chem. 15: 667683.

Straatsma, T., and, McCammon, J.A. (1992). "Computational alchemy." Ann. Rev. Phys. Chem Biol 43:: 407435.

Swendsen, R. H., and, Wang, J.S. (1986). "Replica Monte Carlo simulation of spin-glasses." Phys. Rev. Lett. 57: 26072609.

Tai, K. (2004). "Conformational sampling for the impatient." Biophys Chem 107(3): 213-220.

Tesi, M. C., van Rensburg, E. J. J., Orlandini, E., and, Whittington, S. G. (1996). "Monte Carlo study of the interacting self-avoiding walk model in three dimensions." J. Stat. Phys. 82((1-2)): 155181.

Torrie, G. M., and, Valleau, J.P. (1974). "Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid." Chem. Phys. Lett. 28: 578581.

Valentine, S. J., Anderson, J. G., Ellington, A. D., Clemmer, D. E. (1997). "DisulfideIntact and Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding,." J. Phys. Chem. B. 101: 38913900.

Valentine, S. J., Counterman, A. E., Clemmer, D. E. (1997). "Conformer-Dependent Proton-Transfer Reactions of Ubiquitin Ions." J. Am. Soc. Mass Spectrom. 8: 954961.

van der Kamp, M. W., Shaw, K.E., Woods, C.J., and, Mulholland, A.J. (2008). "Biomolecular simulation and modelling: status, progress and prospects." J R Soc Interface 5:: S173-S190.

van der Kamp, M. W., Shaw, K.E., Woods, C.J., and, Mulholland, A.J. (2008). "Biomolecular simulation and modelling: status, progress and prospects." J. R. Soc. Interface 5: S173-S190.

von Helden, G., Hsu, M.-T., Gotts, N., Bowers, M. T. (1993). "Carbon Cluster Cations with up to 84 Atoms: Structures, Formation Mechanism, and Reactivity." J. Phys. Chem. B. 97: 8182-8192.

Warshel, A., Kato, M., and, Pisliakov, A. V. (2007). "Polarizable force fields: history, test cases and prospects." J. Chem. Theory Comput.

Warshel, A., Levitt, M. (1976). "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme." J Mol Biol 103(2): 227-249.

Weiner, S. J., Kollman, P.A., Nguyen, D.T., Case, D.A. (1986). "An all atom force field for simulations of proteins and nucleic acids." 7(2): 230252.

Widom, W. (1963). "Some topics in theory of fluids." J. Chem. Phys. 39: 2808.

Wyttenbach, T., von Helden, G, and, Bowers, M.T. (1996). "Gas-Phase Conformation of Biological Molecules: Bradykinin." J. Am. Chem. Soc. 118: 8355-8364.

Wyttenbach, T., von Helden, J., Batka, J.J., Jr., Carlat, D., and, Bowers, M.T. (1997). "Effect of the Long-Range Potential on Ion Mobility Measurements." J. Am. Soc. Mass Spectrom. 8: 275-282.

Wyttenbach, T., Witt, M., and, Bowers, M.T. (2000). "On the Stability of Amino Acid Zwitterions in the Gas Phase: The Influence of Derivatization, Proton Affinity, and Alkali Ion Addition." J. Am. Chem. Soc. 122: 3458-3464.

York, D. M., Yang, W., Lee, H., Darden, T., and, Pedersen, L. G. (1995). "Toward the accurate modeling of DNA: the importance of long-range electrostatics." J. Am. Chem. Soc. 117: 50015002.

Zhou, R. (2003). "Free energy landscape of protein folding in water: explicit vs. implicit solvent." Proteins 53(2): 148-161.

Zhou, R. and B. J. Berne (2002). "Can a continuum solvent model reproduce the free energy landscape of a beta -hairpin folding in water?" Proc Natl Acad Sci U S A 99(20): 12777-12782.

Zuckerman, D. M., and, Lyman, E. (2006). "Erratum: A second look at canonical sampling of biomolecules using replica exchange simulation." J. Chem. Theory Comput. 2: 1693.

Zuckerman, D. M., and, Lyman, E. (2006). "A second look at canonical sampling of biomolecules using replica exchange simulation. ." J. Chem. Theory Comput. 2: 12001202.

Zwanzig, R. W. (1954). "High-temperature equation of state by a perturbation method. 1. Nonpolar gases." J. Chem. Phys. 22: 14201426.

Zwanzing, R. W. (1954). "Time-correlation functions and transport coefficient in statistical mechanics." Annu. Rev. Phys. Chem. 22: 1420-1426.

Part I

The DNA in apolar environment

Chapter 3

The Structure and Properties of the DNA in apolar solvents

Annalisa Arcella, Guillem Portella, Rosana Collepardo-Guevara Collepardo, Debayan Chakraborty, David J.Wales and Modesto Orozco, 2014, *J.Phys.Chem.B*, 118(29), 8540-8548

Abstract

The study of nucleic acids in low polarity environments paves the way for novel biotechnological applications of DNA. Here, I use a repertoire of atomistic molecular simulation tools to study the nature of DNA when placed in a highly apolar environment, and when forced to pass from aqueous to apolar solvent. Our results show that DNA becomes stiffer in apolar solvents and suggest that highly negatively charged states, which are the most prevalent in water, are strongly disfavored in apolar solvents, and neutral states with conformations not far from the aqueous ones are the dominant forms. Transfer from water to an apolar solvent is unlikely to occur, but results suggest that if forced, the DNA would migrate surrounded by a small shell of waters (the higher the DNA charge, the larger the number of water molecules in this shell). Even the neutral form (predicted to be the dominant one in apolar solvents) would surround itself by a small number of highly stable water molecules when moved from water to a highly apolar environment. Neutralization of DNA charges seems a crucial requirement for transfer of DNA to apolar media, and the best mechanism to achieve good transfer properties.

3.1 Introduction

DNA is a highly polar molecule, evolved to be stable in high-dielectric environments such as aqueous solution, which is the main environment considered in most experimental and theoretical studies. However, understanding the properties of DNA in low polarity environments is crucial for many nanotechnological applications of DNA (Lin 2008; Mann 2009.), and is instrumental in understanding how nucleic acids acting as antigene or antisense drugs pass the highly hydrophobic cellular membrane, or how DNA can be transported in the body by means of liposome carriers (Bailey 2000; Maurer 2001; Jeffs 2005; Hayes 2006; Skjorringe 2009).

Water and counter-ions are crucial to screen the electrostatic repulsion among charged phosphates, as well as to favor the apolar stacking of the bases. Accordingly, the high impact of solvent modification in the properties of DNA is not surprising. For example, a subtle change in the type of the neutralizing cation can lead to a dramatic conformational change (Gao 1995; Varnai 2004; Rueda, Luque et al. 2006) or even to a complete alteration in the sequence-stability rules of the duplex (Portella, Germann et al. 2014). Addition of ethanol to an aqueous solution induces drastic changes in the duplex structure, forcing a B→A transition (Watson 1953; Wilkins 1953; Franklin 1953;1; Klug 2004; Noy, Perez et al. 2007).

Some organic solvents such as methanol, formamide, pyridine or DMSO (Herskovits 1972; Bonner 2000; Ke 2010), induce unfolding and strand separation, or formation of toroid-like conformations (Feng 1999; Pereira 2001; Montesi 2004), while others, such as glycerol (Bonner 2000) or room-temperature ionic liquids (Mamajanov, Engelhart et al. 2010; Portella, Germann et al. 2014) maintain the duplex structure. More surprisingly, a DNA duplex does not loose its structure completely, and the two strands remain bound when DNA is transferred from aqueous solution to the gas phase (Herskovits 1972; Rueda 2003; A Arcella 2012). It is clear that changes in the balance of stabilizing and destabilizing terms controlling DNA structure due to alterations in the solvent is complex, and still not well understood.

Experimental studies in apolar solvents are challenging because of the difficulty of transfering DNA from water to these solvents. This problem has motivated several theoretical works, most of them where molecular dynamics (MD) simulations are used. For example, Cui et al. (Cui 2007) have studied the structure and stability of DNA in octane, correlating the observed changes to those required for helicase activity. Lin and Vaidehi (Lin 2008) explored the stability of decorated DNAs in interfaces between DNA and chloroform, and our group explored the reaction of DNA to high concentrations of pyridine (Perez and Orozco 2010).

Finally, Khalid et al (Khalid 2008) used coarse grained models to study the insertion of duplex DNA into a lipid bilayer, finding a significant associated free energy barrier, even for small duplexes.

Here I use massive atomistic MD simulations to investigate the nature of a short DNA duplex in carbon tetrachloride CCl_4, which is considered a simplified model of a highly apolar environment. I explored also the (free) energy and structural changes associated to the change of phase of a duplex DNA from water to CCl_4, and back to water. Finally I analyzed for the first time the possibility that DNA acts as a proton donor/acceptor, when it crosses from a polar to an apolar medium, such as from water to a biological membrane.

Results suggest that DNA, at least a short fragment, can maintain its duplex structure for significant periods of time in an apolar solvent, and surprisingly appears more rigid in such apolar solvents, than when immersed in aqueous solution. Transfer from aqueous to apolar solvents is, obviously, strongly disfavored, especially if the DNA remains fully charged, and typically the transferred DNA is not fully dehydrated but maintains (irrespective of its formal charge) a significant amount of water around. Calculations suggest that a neutral DNA state with a compact structure (not too different to that in aqueous solution) is the most prevalent form of DNA in apolar solvent.

3.2 Methods

I used a small oligonucleotide (d(GCGAAGC)) as a model system for all calculations. In aqueous solution, this small DNA forms a stable hairpin with a GAA triloop and a two-d(CG)-pairs stem, organized in a canonical B-type duplex as noted by NMR experiments (Padrta 2002). Experiments and theory have shown that this oligonucleotide folds in the (multi)microsecond (Ansari 2001; Ma 2007; Portella 2010), when immersed in water and unfolds in a similar time scale in a denaturant solvent such as pyridine (Perez and Orozco 2010). The small size of the system accelerates calculations, favors sampling and reduces memory effects, which might be very important for longer oligos. Previous studies (Orozco, Perez et al. 2003; Portella 2010; Perez, Luque et al. 2012) have demonstrated the ability of current force-field and MD simulation protocols to simulate this structure in a variety of environments (Orozco, Perez et al. 2003; Portella 2010; Perez, Luque et al. 2012).

3.2.1 Equilibrium molecular dynamics simulations

The 7-mer DNA hairpin was first simulated in aqueous solution to have a reference for the native trajectory. The NMR structure (Padrta 2002) was neutralized by adding 6 Na+ (placed at optimal CMIP positions (Gelpi, Kalko et al. 2001)), and solvated by 7200 water molecules, defining an octahedral box. The system was optimized, thermalized and pre-equilibrated using standard procedures with triplicated simulation windows (Shields 1997; Portella 2010), and equilibrated for 100 ns prior to 1 μs of production trajectory at constant temperature and pressure (T=300K; P=1 atm). Long range electrostatic corrections were represented by means of periodic boundary conditions (PBC) and the particle mesh Ewald (PME) correction with default parameters (Darden T. 1993).

Starting from the aqueous solution equilibrium structure, I performed unbiased MD simulations in CCl_4 considering the canonical DNA protonation state (Q=-6) and a fully neutralized system in which all the phosphates were protonated (Q=0). The simulation systems contained one DNA molecule and around 4288 CCl_4 molecules. Optimization, thermalization and equilibration procedures were the same as those used in the aqueous simulations.

Trajectories were also collected for 1 μs at constant temperature and pressure (T=300K; P=1 atm) using PBC-PME to account for long-range effects. Note that the CCl_4-simulation systems do not contain counterions, which means that Ewalds plasma is the responsible for charge valance in these simulations. To discard potential Ewald artifacts in CCl_4-simulation, I repeated calculations for the Q=-6 system using a reaction field correction with a 1.1 nm cutoff to define the boundary between atomistic system and the exterior dielectric continuum.

3.2.2 Replica exchange molecular dynamics (RExMD) simulations

To analyse the conformational conformational space of the hairpin in CCl_4 I performed replica exchange RExMD (Sugita 1999) simulations for the two charge states (Q=-6 and Q=0) starting from: i) a fully extended conformation and ii) the aqueous-equilibrated structure. As an additional test, I extended the RExMD simulations to a minimum-hydration hairpin (around 50 water molecules) obtained from simulations of the transfer of the hairpin from water to CCl_4 (see below).

A total of 25 temperatures were selected to span the gap between 300 and 440 K. The distribution of temperatures was chosen to guarantee an average exchange probability of 25%. Each of the temperature replicas was explored for 150 ns (for a total

simulation time of $2 \times 2 \times 35 \times 0.150 = 15\mu s$), with exchanges being attempted every 1000 MD steps.

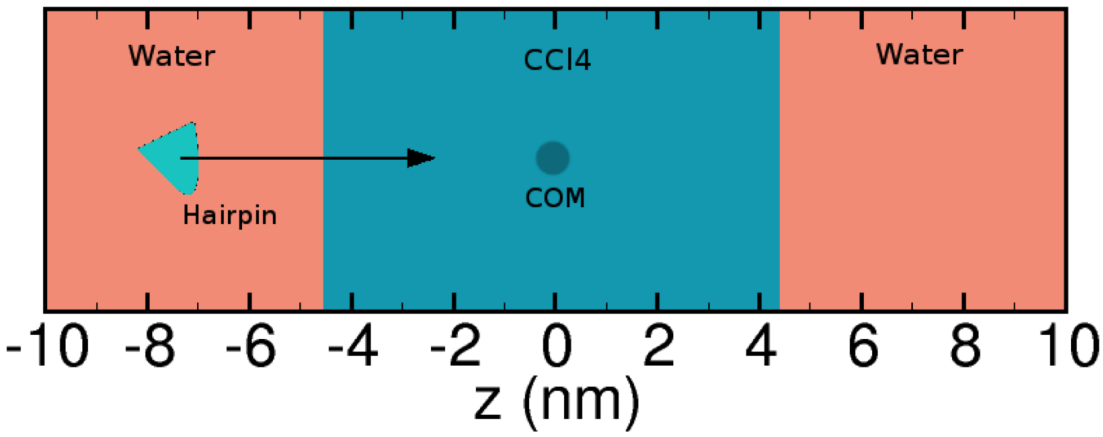

FIGURE 3.1: Schematic representation of the set-up of the steering MD simulation protocol used to study the transfer of the DNA hairpin from polar →← apolar solvents.

3.2.3 Steered molecular dynamics and umbrella sampling calculations

To obtain an atomistic representation of the transfer of the hairpin from aqueous solution to CCl_4, I performed an steered MD simulation (Hess 2010), where we moved the hairpin across the major axis of an orthorhombic box (size $(5.85 \times 5.85 \times 19.4)nm^3$) containing 9460 CCl_4 molecules (placed in the center of the long axis) and 34935 water molecules (placed at both sides of the CCl_4 phase; see Figure 3.1). Simulations for the Q=-6 DNA included 6 Na^+ counterions (see above for neutralization protocol), while no explicit counterions were included for the Q=0 simulations. The steered MD simulations either move the hairpin from one water box to the passing through the CCl_4 phase, or from the center of the CCl_4 phase to one of the water boxes. A slow steering velocity (0.05 nm/ns; with force 1000 $kJ/(mol \cdot nm^2)$) was used to reduce hysteresis effects.

The steered MD trajectory was then used to extract 75 starting configurations (spaced 0.2 nm along the major axis of the simulation system; see Figure 3.1) from which umbrella sampling (US) calculations were performed. Prior to the US calculations all configurations selected from the stereed MD simulations were relaxed for 20 ns. For the US procedure, I used the position of the DNA hairpin in the transfer direction (z) (see Figure 3.1) as the reaction coordinate, and a harmonic umbrella potential ($k = 1000kJ/(mol \cdot nm^2)$) with 75 overlapping windows, centered at consecutive values of the reaction coordinate that span the complete water→ CCl_4 →water transfer path.

I define the central window as that centered at the center of mass (COM) of the CCl_4 box (z = 0 nm), and center the first and last windows at points located at z=-7.3

and z=+7.3 nm with respect this COM. Individual trajectories for each of the 75 US windows extend for 100 ns (for a total simulation time of 9 microseconds per system).

The potential of mean force (PMF) was calculated using the cyclic implementation of the Weighted Histogram Analysis Method (WHAM) in GROMACS 4.5.5 (Kumar 1992; Hub 2010). WHAM incorporates the Integrated Autocorrelation Times (IACT) of umbrella windows. Because IACT is subject to uncertainties due to limited sampling, I have smoothed the IACT along the reaction coordinate using a Gaussian filter of width of 0.5 nm (Hub 2010). The statistical uncertainty of the PMFs was estimated using bootstrap analysis (Hub 2010).

3.2.4 Poisson-Boltzmann calculations of the phase transfer

I use finite difference resolution of Poisson-Boltzmann (PB) equation as implemented in our CMIP code (Gelpi, Kalko et al. 2001) to determine the change in solvation free energy when the hairpin moves from regions of low (CCl_4) and high (H_2O) dielectrics. Dielectric boundaries were defined from the placement of the interfaces between phases in the US calculations. For each of the 75 windows considered in the US calculations I selected 50 individual hairpin conformations (for Q=-6 and Q=0 states) performing then a total of 15300 PB calculations from which an average free energy of solvation for each point of the transfer process was determined. To solve PB equation I used a regular grid (spacing 0.5 nm) with three dielectric regions (DNA $\epsilon = 1$, CCl_4 $\epsilon = 2$, and H_2O $\epsilon = 80$).

3.3 Results and Discussions

3.3.1 The DNA hairpin in water

The d(GCGAAGC) oligonucleotide forms a very stable DNA hairpin in aqueous solution, with a well-organized and stable tri-loop and a short, but stable, B-DNA stem. The native NMR conformation is stabilized by eight base-pairing hydrogen bonds (H-bonds) (six $d(GC)$ bonds plus two $d(GA)$ bonds) (Padrta 2002), which I term 'canonical H-bonds'. Despite backbone fluctuations, around 98% of these canonical H-bonds are preserved in the simulation. Long MD simulations in aqueous solvent suggest that the DNA native hairpin conformation is stable at room temperature (Figure 3.2), in agreement with previous experimental (RMSD = (0.08 ± 0.01) nm with respect to the NMR average structure) and MD studies (Ansari 2001; Ma 2007; Portella 2010).

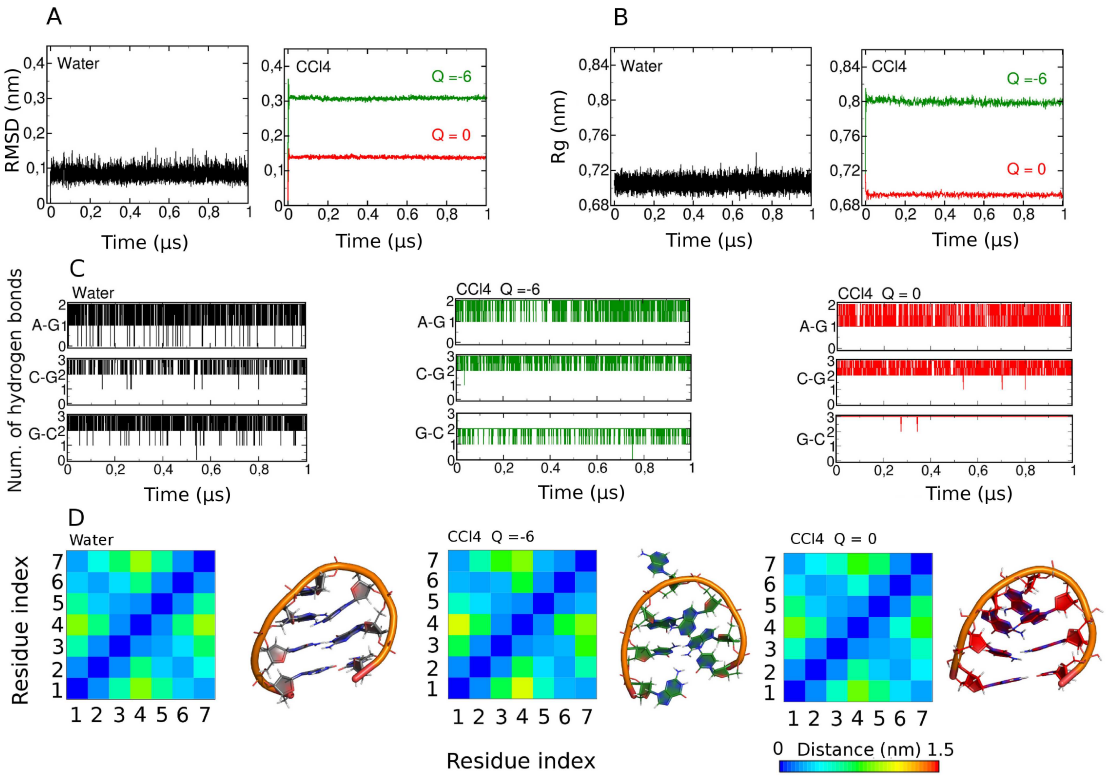

FIGURE 3.2: Summary of structural descriptors of unbiased MD simulations of the DNA hairpin in aqueous and CCl4 solution (Q=-6 and Q=0 states; starting always from the equilibrated structure in water). Panel A: RMSd with respect to NMR structure. Panel B: Radii of gyration, Panels C: hydrogen bonds. Panel D represent average contact maps obtained from the simulations and the representative structures.

3.3.2 The DNA hairpin in dry CCl_4

A transfer without any topological alteration of the DNA hairpin from water to an apolar media implies that DNA should have a Q=-6 net charge in the apolar phase, as it happens in water. However, the acidity of phosphate groups is expected to change dramatically in an apolar solvent, and it is thus possible that phosphates will instead become protonated upon transfer to CCl_4, generating a neutral hairpin. In order to cover both possibilities, I first performed unbiased MD simulations for the hairpin in the charge states Q=-6 and Q=0 in a box of CCl_4.

MD simulations of the Q=-6 state sample (in the microsecond regime) conformations that are less compact than those found in aqueous solution (increase around 0.1 nm in the Rg, Figure 3.2). The pattern of H-bonds is distorted, with a net loss of around 20% with respect to the aqueous solvent (Figure 3.2), and the loop geometry of Q=-6 hairpin is corrupted in CCl_4, with a significant loss of canonical G-A contacts (see contact maps in Figure 3.2).

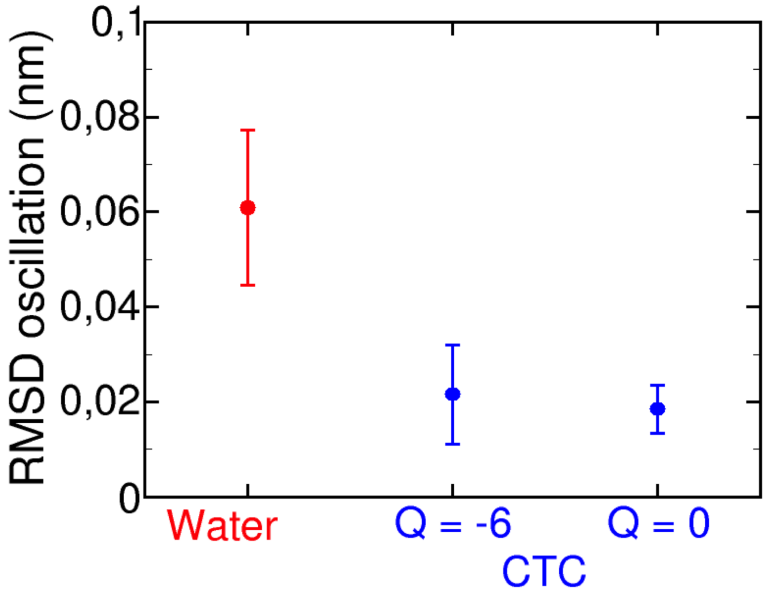

FIGURE 3.3: Averaged RMSd oscillations, calculated with rispect the averaged strcuture in the last 20 ns of simulation (and associated standard deviations) obtained from MD simulations of the DNA hairpin in water and in CCl4 solution (Q=-6 and Q=0 states).

However, despite all these changes the samples collected in microsecond long MD simulations of the Q=-6 hairpin DNA in CCl_4 (at room temperature) are not dramatically different to those obtained in simulations performed considering room temperature water as solvent. Deviations from the aqueous solution structure are even smaller in unbiased microsecond-long simulations of the Q=0 hairpin (RMSd from NMR structure around 0.14 nm; see Figure 3.2).

There is again a loss of canonical H-bonds (in average 16%, see Figure 3.2), but the loop geometry is well preserved, and the changes of the radius of gyration are small with respect to the aqueous simulations (just a small compression of the structure due to the lack of phosphate-repulsions in a neutralized DNA).

In summary, unbiased MD simulations suggest that either DNA hairpin has a strong memory of aqueous structure when placed in apolar environment (longer than microsecond), or no dramatic structural changes happen upon phase transfer (see below).

When transferring the DNA from a highly structured solvent like water to a poorly structured apolar solvent such as CCl_4, I would expect movements in the DNA to be enlarged, because of the larger fluidity of the apolar solvent, and because of the lack of strong DNA-solvent contacts. However, MD simulations show that RMSD fluctuations with respect to the average structure are much smaller in the simulations in CCl_4 than in equivalent trajectories (for the same temperature) in water (see Figure 3.3).

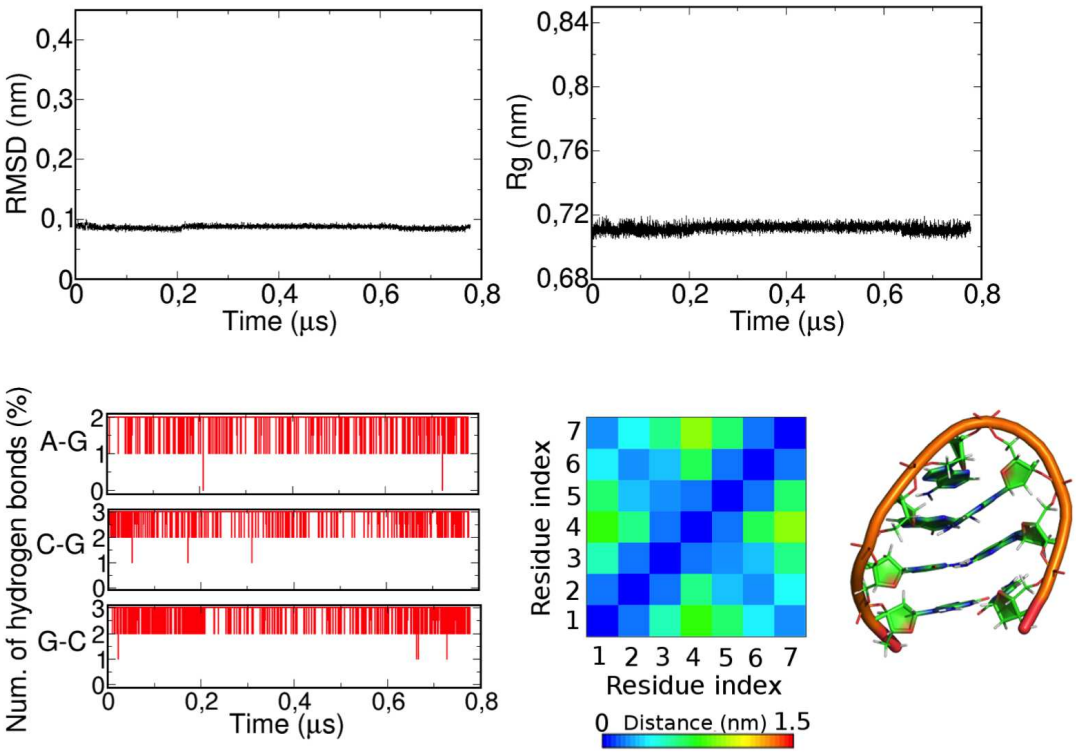

FIGURE 3.4: Summary of structural descriptors of unbiased MD simulations of the DNA hairpin in CCl4 solution (Q=-6 state) performed using a reaction field correction instead of the PME protocol. Compare plots with those presented in Figure 3.2

Clearly, the ability of water to establish H-bond interactions with the DNA favors DNA flexibility, since in water temporarily lost intra-DNA interactions (due to thermal fluctuations) are compensated with DNA-water contacts. In contrast, in an apolar solvent, lost of intra-DNA interactions cannot compensated. The final result is that, as previously found "in vacuum" (Rueda 2003; A Arcella 2012), polar structures such as DNA become stiffer in anhydrous environments.

I were concerned in all the analysis reported to date that the stability of the Q=-6 state could be related to PME-artifacts, since PME plasma neutralizes the hairpin inside the simulation box, which might not reflect real composition of a solution of DNA in dry CCl_4. To verify that PME was not a source of artifacts in the CCl_4 simulations, I repeated the Q=-6 hairpin simulation using a reaction field to correct distant interactions instead of the default PBC-PME procedure. As shown in Figure 3.4 and Figure 3.2 reaction field and Ewald simulations give nearly equivalent results, confirming previous claims by others (Hub, de Groot et al. 2014) that Ewald corrections are not a major source of errors in this type of simulations.

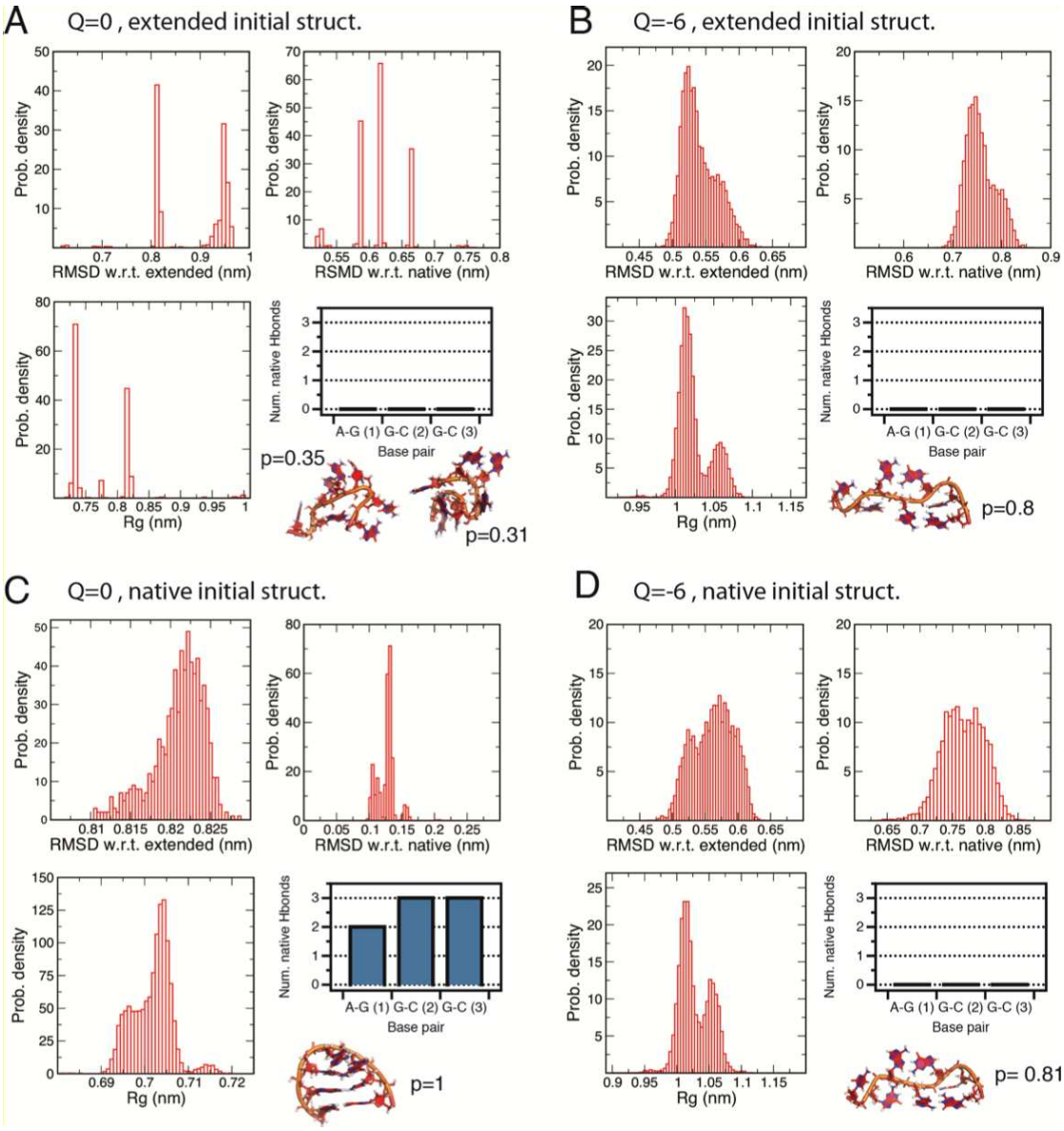

FIGURE 3.5: Summary of RexMD results for the dry hairpin in CCl4 solution (Q=-6 and Q=0) started from the native conformation (the folded state in water) and a fully extended structure. Results shown correspond to the last 50 ns of the replicas at room temperature (extracted from aggregated simulation time of 3.75 μs (see Methods).

As noted above, it is important to realize that the stability of the hairpin in one microsecond of unbiased MD simulation in CCl_4 does not mean that the real structure of the hairpin in CCl_4 will be close to that found in aqueous solution, since equilibration times in apolar solvent (when DNA is stiffer) can be much longer than the time scale sampled here. To explore possible memory effects I performed large-scale RExMD simulations with a wide range of temperatures and an aggregated simulation time of more than 15 microseconds (see Methods 5.2).

As describe above, I started simulations from both an "extended" and a "native-like" conformation for the two charged states of the hairpin. For an infinite simulation

time, both sets of simulations should converge to the same structural ensemble and divergence should signal the existence of a very rough conformational landscape, which can yield very strong memory effects in unbiased MD simulations at room temperature.

RExMD simulations of the Q=-6 hairpin leads to quite extended structures at the end of the simulation (Figure 3.5). The structures sampled have a significant amount of H-bonds, but few of them are native (Figure 3.5). The loop and general hairpin structure is also destroyed, and it is difficult to find any signal of the original hairpin structure in the sampled structures at room temperature (Figure 3.5). From these RExMD and unbiased MD simulations, I can conclude that the native structure is a metastable (in the microsecond time scale) conformation for a hypothetical fully charged hairpin, but it is not the most prevalent conformation in CCl_4.

The results of the RExMD simulations are more difficult to interpret for the Q=0 state, since convergence is not achieved. Simulations starting from the native state remain very close to the aqueous structure, maintaining the pattern of native H-bonds (Figure 3.5), while simulations starting from the extended state collapses to a variety of different compact, but non-native structures. It is clear then, that the conformational landscape of the Q=0 hairpin in CCl_4 is dominated by compact conformations, and that it is too frustrated for RExMD or unbiased MD simulations alone to characterize properly and determine then the most dominant conformer.

3.3.3 The transfer of the DNA hairpin between polar and apolar phases

DNA is naturally found in water, and large barriers are expected when transferred to an apolar media, such as a biological membrane or a liposome. Modeling this energetically costly process requires biasing schemes to force transfer from water to CCl_4. As described in Methods (5.2) section, I first performed a very slow velocity steered MD simulation to obtain starting conformations for umbrella sampling free energy calculations. I first pull the hairpin from one water box to the other passing through the CCl_4 phase (see schematic representation in Figure 3.1).

As seen in Figure 3.7, even for very small pulling velocities, the hairpin crosses the hydrophobic box hydrated, and keeping well the general hairpin structure (see Figures 3.6, 3.8). The number of water molecules in the micro-hydration drop depends on the charge state of the hairpin (see Figure 3.6), but even for the neutral hairpin there is a small, but non-negligible number of water molecules protecting the hairpin from the apolar solvent (see Figure 3.8). Despite displacements of the hairpin within the box were very slow, I cannot rule out potential problems in the relaxation of solvent environment around the DNA.

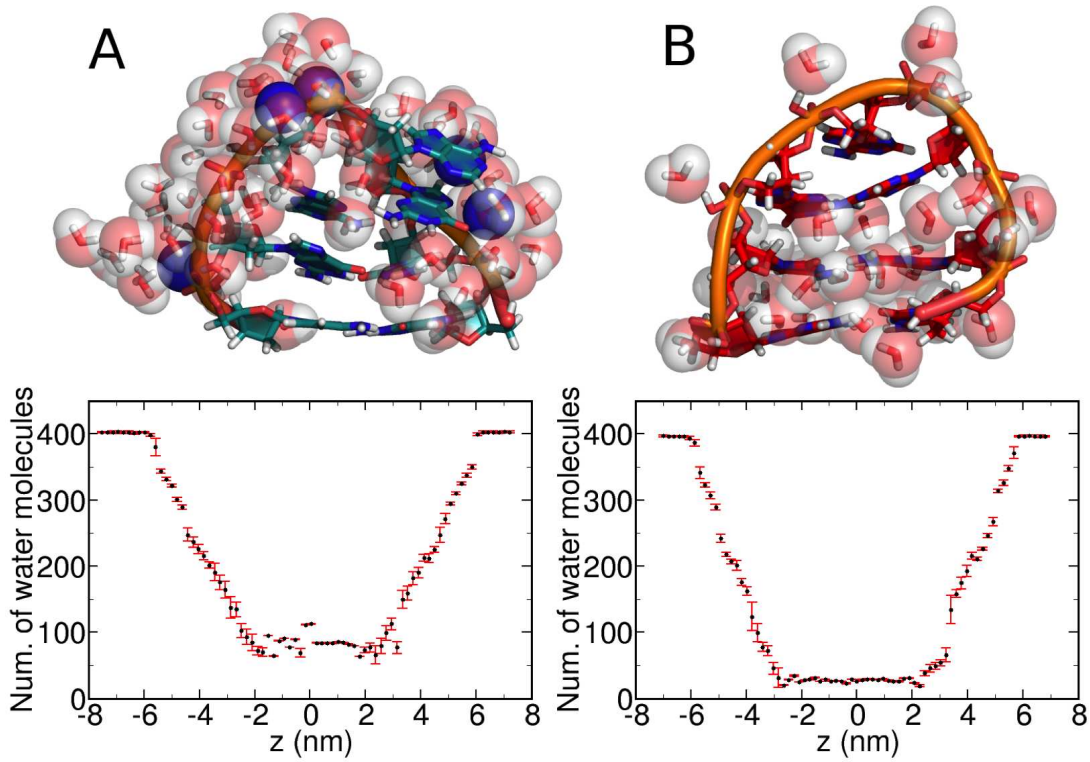

FIGURE 3.6: Evolution of the number o waters bound to the DNA hairpin during the phase-transfer water$\rightarrow CCl_4 \leftarrow$ water. (A) For Q=-6 hairpin; (B) for Q=0 hairpin.

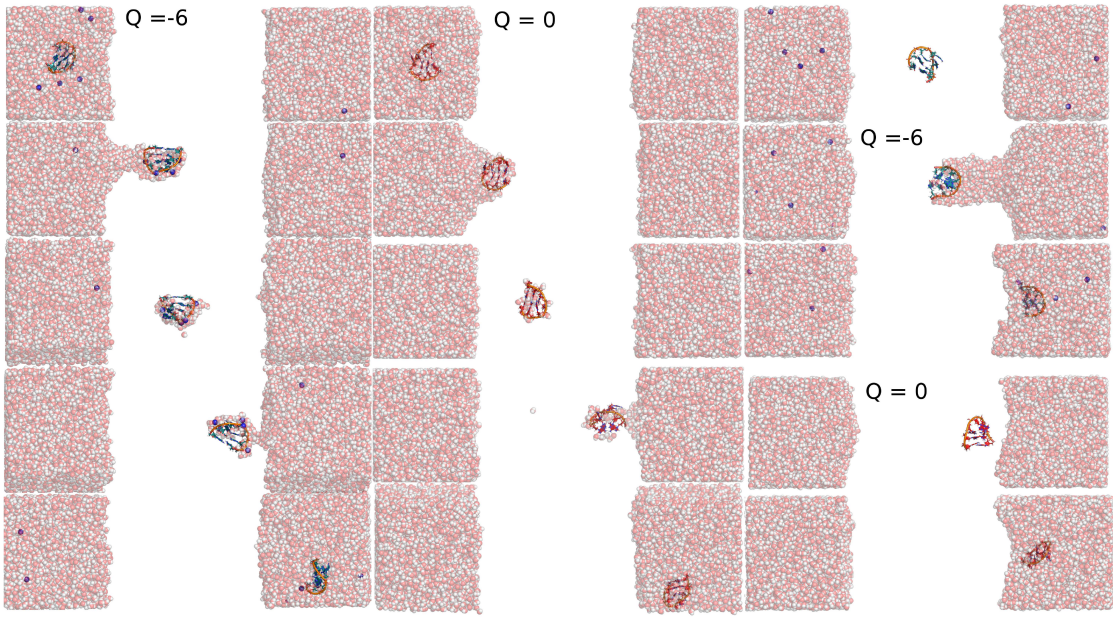

FIGURE 3.7: Representation of different snapshots obtained in steered MD simulations forcing the phase transfer of DNA hairpin. (A) From water to CCl_4 and back to water for Q=-6 DNA hairpin. (B) From water to CCl4 and back to water for Q=0 DNA hairpin. (C) From CCl4 to water for the Q=-6 DNA hairpin. (D) From CCl4 to water for the Q=0 DNA hairpin.

To discard these potential relaxation artifacts, and to confirm that even a neutral hairpin is hydrated in CCl_4, I repeated the pulling procedure, but starting with a dry hairpin in the middle of CCl_4 phase. These complementary simulations confirm that as soon as the hairpin moves from the center of the apolar phase, a stream of water molecules flows in, and spontaneously generate a micro-hydration environment, which is especially noticeable for the fully charged hairpin (Figure 3.6). This means that the invasion of water molecules in the apolar phase is not a relaxation artifact, but a real physical effect. Hence I predict that under normal conditions a hairpin would cross a hydrophobic phase, not in the dry state, but surrounded by a small number of water molecules.

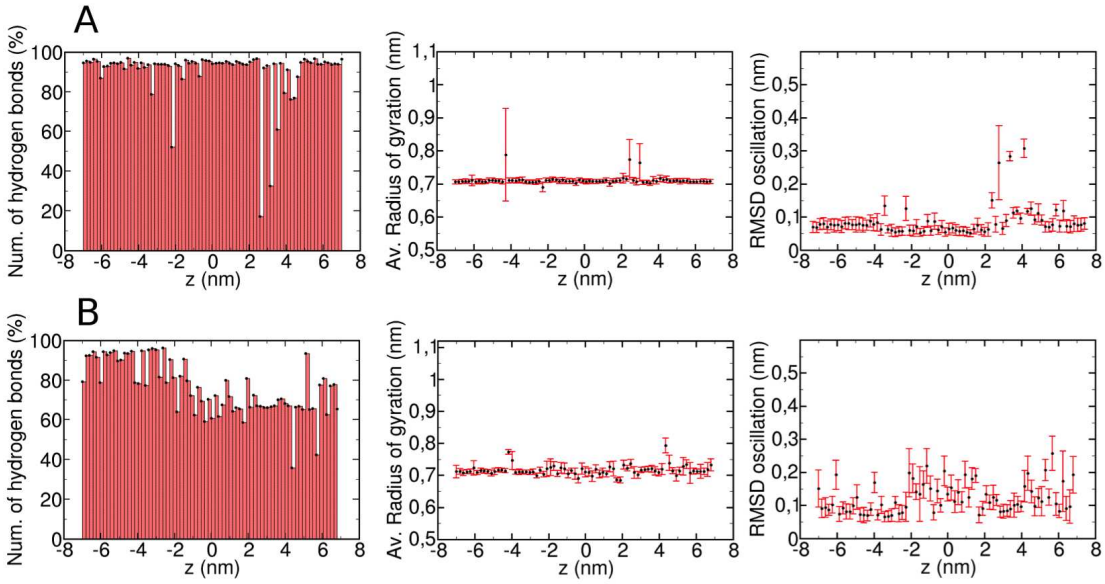

FIGURE 3.8: Evolution of several structural descriptors of the DNA hairpin during the phase-transfer water→ CCl_4 ←water. (A) For Q=-6 hairpin; (B) for Q=0 hairpin. Data shown here correspond to ensembles obtained at the end of the umbrella sampling windows (see below), not just the steered-MD collected snapshots.

The fact that a DNA hairpin in an apolar environment appears as a micro-hydrated species rises questions on the validity of previous RExMD simulations of the hairpin that were performed in dry of CCl_4.

In order to check this, I repeated RexMD calculations, but using micro-hydrated species (Q=0 and Q=-6) obtained from the pulling experiments when the center of mass of the hairpin is placed at the middle of the apolar phase.

I found that the impact of the micro-hydration environment helps stabilizing the folded state (see Figure 3.9), shielding the electrostatic repulsion between the charged phosphate groups.

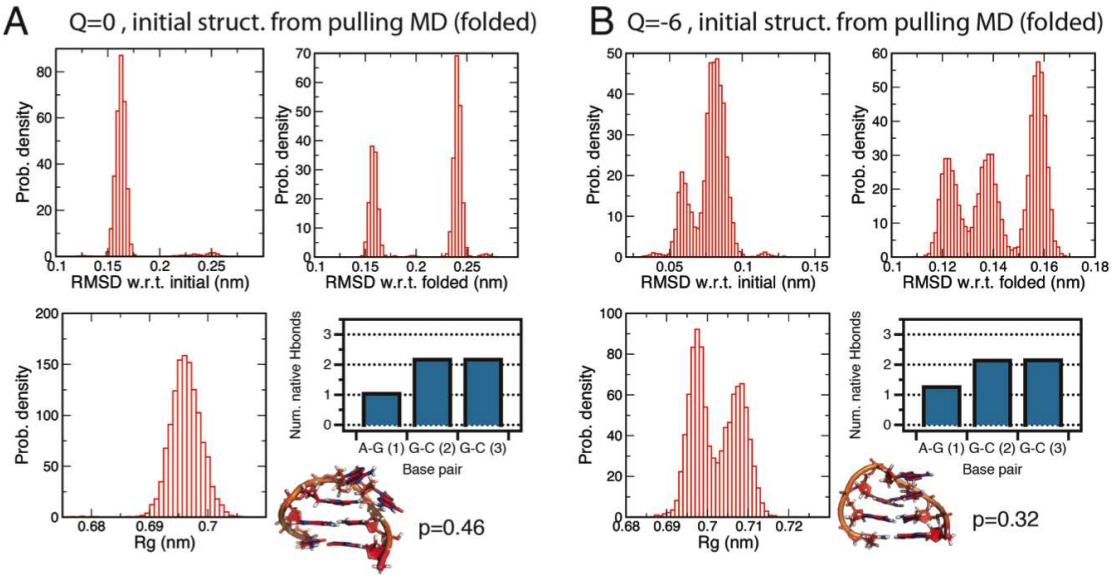

FIGURE 3.9: Summary of RexMD results for the micro-hydrated hairpin in CCl4 solution (Q=-6 and Q=0) started from the geometry of the hairpin when placed in the middle of the CCl4 phase in a pulling experiment of phase transfer. Results shown correspond to the last 50 ns of the replicas at room temperature (extracted from aggregated simulation time of 3.75 microseconds (see Methods 5.2).

The Free Energy to the transfer from polar to apolar phases

As described in Methods 5.2, I used WHAM and umbrella sampling (from steered-MD collected snapshots) to compute the free energy change associated to the transfer of the hairpin across polar and apolar phases. Results displayed in Figure 3.8, 3.10 and 3.12 clearly illustrate the large free barrier associated with crossing a hydrophobic phase (around 650 kJ/mol) for a fully charged hairpin and 150-200 kJ/mol for the neutral one.

These numbers are qualitatively close to those expected from Poisson-Boltzmann calculation (see Figure 3.12), supporting their validity, which demonstrated that transfer of a hairpin from water to an apolar solvent, even if micro-hydrated, is largely disfavored. These results reinforce the notion that decoration of the DNA with apolar moieties (Bailey 2000; Maurer 2001; Jeffs 2005; Hayes 2006; Skjorringe 2009) is desirable to improve permeation.

In order to validate the free energy estimates obtained from umbrella sampling simulation, I performed a large-scale Poisson-Boltzmann study (see Methods 5.2) of the change in solvation free energy associated to the transfer of Q=0 and Q=-6 hairpin from a high dielectric (water) to a low dielectric (CCl_4) back to high dielectric (water). Results are graphically summarized in Figure 3.12. The agreement between discrete atomistic MD-umbrella sampling calculations and macroscopic Poisson-Boltzmann calculation is remarkable and provides support for the results presented in Figure 3.10.

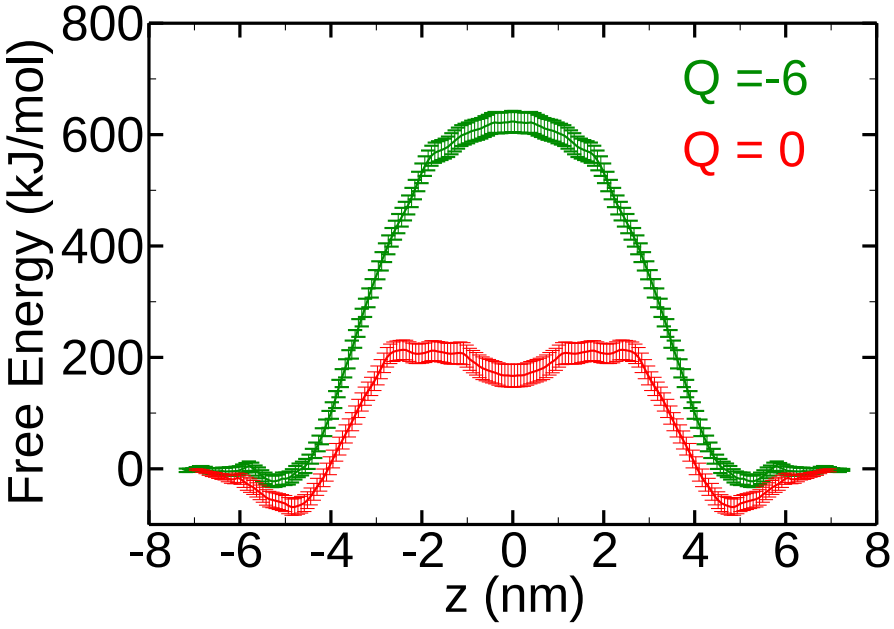

FIGURE 3.10: Potential of mean force associated to the transfer of hairpin (Q=-6 and Q=0) across a polar/apolar/polar simulation box (see Methods and Figure 3.10).

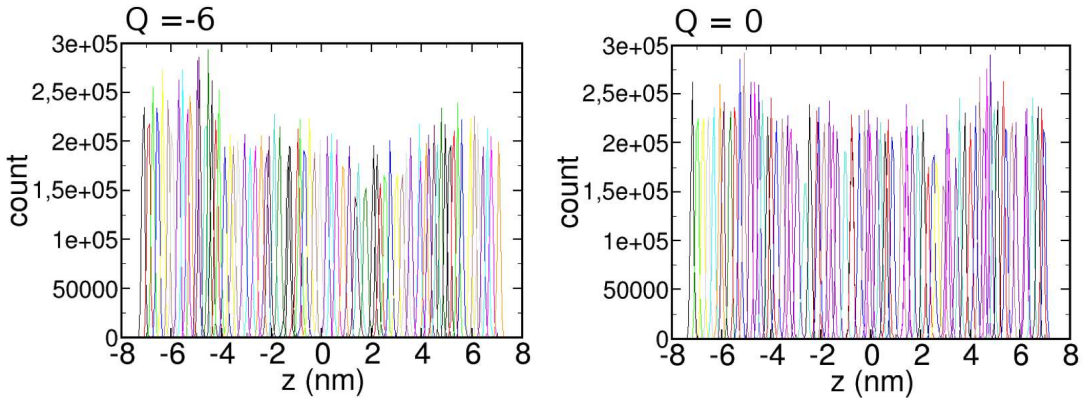

FIGURE 3.11: Histogram overlap obtained from umbrella sampling simulations of the transfer of hairpin from polar to apolar to polar phases.

As a final point to our descriptive study, it is worth to analyze the relative weight of neutral and charged states in a hairpin in CCl_4. This can be done using a standard thermodynamic cycle like that in Figure 3.13, where I compute the free energy cost of transforming a neutral to a charged state in CCl_4, the cost of removing the hairpin charge in water, and the cost associated to the phase transfer of the neutral and charged hairpins.

The free energy cost associated to the annihilation of the hairpin charge can be easily computed at neutral pH from the experimental pKa in water (around 2) of a DNA phosphate (pKa $\sim$ 2; $\Delta G \sim$ 160 kJ/mol assuming that all phosphate groups protonate independently of each other), from which I conclude that charge annihilation in water

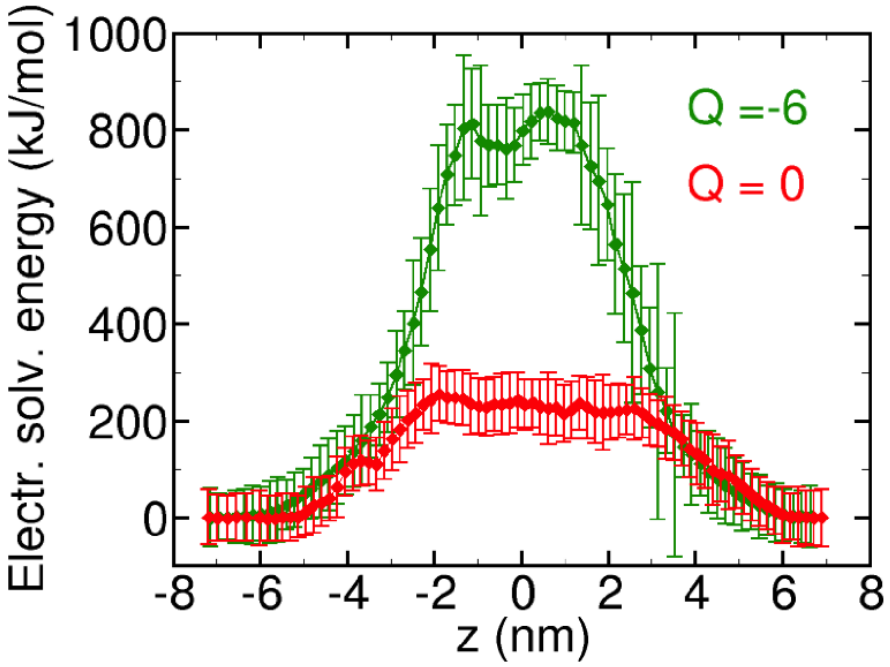

FIGURE 3.12: Change in solvation free energy associated to the transfer water→ CCl_4 ←water of a fully charged (Q=-6) and neutral (Q=0) hairpin.

is disfavored by c.a. 350 kJ/mol. The free energy associated to the transfer of a fully charged hairpin from water to CCl_4 is suggested to be around 450-500 kJ/mol higher than the equivalent process for the neutral state. In other words, I suggest that DNA in apolar media should have a reduced charge, and can be even completely neutral. Under these conditions DNA seems able to maintain its native structure in solution and is expected to show similar properties, despite a larger rigidity.

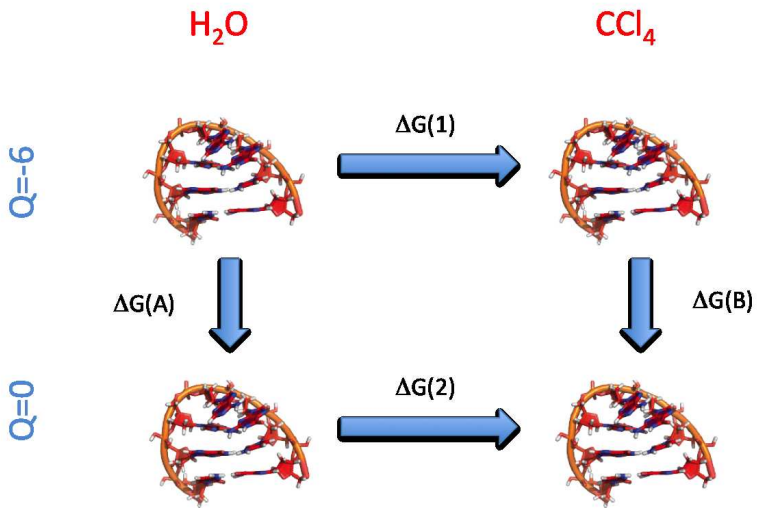

FIGURE 3.13: Thermodynamic cycle and associated free energy values to determine the charge state of a short DNA hairpin in solution (see text).

3.4 Conclusions

DNA is a highly apolar molecule which evolved during millions of years to be stable and functional in aqueous solution. DNA will not spontaneously transfer to an apolar phase such as a biological membrane, which represents a clear advantage in terms for confining DNA in a given cellular compartment.

However, biotechnological and biomedical applications often require the DNA to cross biological membranes, being encapsulated into a hydrophobic, environment, or just being a chemical reactant in a non-aqueous solvent. It is then necessary to study the behavior of nucleic acids in non-polar solvents as a preliminary step to design modifications that can favor its transfer from water. For this purpose, theoretical approaches have no competition, since they are the only approach that allows us to study DNA in such hostile environments.

The massive simulation effort presented here, illustrates the significant changes occurring in the conformational landscape of a small model system of DNA when transferred from water to CCl_4. The DNA becomes stiffer and the landscape increases its roughness in the apolar phases, which might generate folding frustration. Compact states, probably not far from the aqueous structure, are sampled if the DNA is neutralized, while more extended conformations are expected if the DNA maintains its ionic state in water.

The transfer of DNA from aqueous solution to CCl_4 is largely disfavored, especially if the DNA maintains its aqueous ionic state. The transfer always happens with the DNA placed in a micro-drop of water, not for a naked DNA. The size of the water drop increases with the net charge of the DNA, but even for a neutral DNA a significant number of water molecules are detected in the apolar phase.

Combination of umbrella sampling potential of mean force simulations (supported by Poisson-Boltzmann calculations) with experimental pKa estimates demonstrates that neutral state is the favored ionic state for the DNA, suggesting that entrance of DNA can generate a net proton flux across the membrane.

Bibliography

A. Arcella, G. Portella, M. L. Ruiz, R. Eritja, M. Vilaseca, V. Gabelica, and Modesto Orozco (2012). "Structure of Triplex DNA in the Gas Phase." J. Am. Chem. Soc. 134: 6596-6606.

A. Ansari, S. V. K., Y. Shen (2001). PNAS 98: 7771. Ansari, A., Kuznetsov, S.V., Shen, Y. (2001). "Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins,." PNAS 98(14): 7771-7776.

Bailey, A. L., Sullivan, S.M. (2000). "Efficient encapsulation of DNA plasmids in small neutral liposomes induced by ethanol and calcium. ." Biochimica Et Biophysica Acta-biomembranes 1468: 239252.

Becker, O. M. and M. Karplus (1997). "The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics." Journal of Chemical Physics 106(4): 1495-1517.

Bonner, G., Klibanov, A.M. (2000). "Structural stability of DNA in nonaqueous solvents." Biotechnol Bioeng 68(3): 339-344.

Carr, J. M., S. A. Trygubenko, et al. (2005). "Finding pathways between distant local minima." Journal of Chemical Physics 122(23).

Case, D. A., T. A. Darden, et al. (2006). "AMBER 9." University of California, San Francisco.

Cheatham, T. E., III, Cieplak, P., and, Kollman, P.A. (1999). "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat." J. Biomol. Struct. Dyn. 16: 845-862.

Cui, S., Yu, J., Khner, F., Schulten, K., Gaub, H.E.; (2007). "Double-Stranded DNA Dissociates into Single Strands when dragged into a Poor Solvent." J. Am. Chem. Soc. 129: 14710-14716.

Darden T., Y., D., Pedersen, L. (1993). "Particle Mesh Ewald-an N.Log(N) method for Ewald sums in large systems." J. Chem. Phys. 98: 10089-10092.

Evans, D. A. and D. J. Wales (2003). "Free energy landscapes of model peptides and proteins." Journal of Chemical Physics 118(8): 3891-3897.

Feng, Y., Spisz, T.S., Hoh, J.H. (1999). "Ethanol-induced structural transitions of DNA on mica." Nucleic Acids Res. 27(8): 1943-1949.

Franklin, R. E., Gosling, R.G. (1953;1). "Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate." Nature 72: 156-157.

Gao, Y. G., Robinson, H., van Boom, J.H., Wang A.H. (1995). "Influence of counter-ions on the crystal structures of DNA decamers: binding of [Co(NH3)6]3+ and Ba2+ to A-DNA." Biophys J 69(2): 559-568.

Gelpi, J. L., S. G. Kalko, et al. (2001). "Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins." Proteins 45(4): 428-437.

Hayes, M. E., Drummond, D.C., Kirpotin, D.B., Zheng, W.W., Noble, C.O. (2006). "Genospheres: self-assembling nucleic acid-lipid nanoparticles suitable for targeted gene delivery." Gene Therapy 13: 646651.

Henkelman, G. and H. Jonsson (1999). "A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives." Journal of Chemical Physics 111(15): 7010-7022.

Henkelman, G. and H. Jonsson (2000). "Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points." Journal of Chemical Physics 113(22): 9978-9985.

Henkelman, G., B. P. Uberuaga, et al. (2000). "A climbing image nudged elastic band method for finding saddle points and minimum energy paths." Journal of Chemical Physics 113(22): 9901-9904.

Herskovits, T. T., Harrington J.P. (1972). "Solution studies of the nucleic acid bases and related model compounds. Solubility in aqueous alcohol and glycol solutions." Biochemistry 11 (25): 4800-4811.

Hess, B., van der Spoel, D., Lindhal, E. (2010). Pull-code.

Hub, J. S., de Groot, B.L., van der Spoel, D. (2010). "g-wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates." Journal of Chemical Theory and Computation 6: 3713-3720.

Hub, J. S., B. L. de Groot, et al. (2014). "Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge." Journal of Chemical Theory and Computation 10(1): 381-390.

Jeffs, L. B., Palmer, L.R., Ambegia, E.G., Giesbrecht, C., Ewanick, S. (2005). "A scalable, extrusion-free method for efficient liposomal encapsulation of plasmid DNA." Pharmaceutical Research 22: 362372.

Ke, F., Luu, Y.K., Hadjiargyrou, M., Liang., D. (2010). "Characterizing DNA condensation and conformational changes in organic solvents." PLoS One. 11(5(10)): e13308.

Khalid, S., Bond, J.P., Holyoake, J., Hawtin, W.R., and, Sansom, M (2008). "DNA and lipid bilayers: self-assembly and insertion." J. R. Soc. Interface 5(241-250). Klug, A. (2004). "The discovery of the DNA double helix." J. Mol. Biol. : 335:333-326.

Kumar, S., Rosemberger, J.M., Bouzida, D., Swendsen R.H. and Kollman, P.A. (1992). "The weighted histogram analysis method for free-energy calculations on biomolecules." J. Comput. Chem. 13: 1011-1021.

Lin, J., Seeman, N.C., Vaidehi, N. (2008). "Molecular-dynamics simulations of insertion of chemically modified DNA nanostructures into a water-chloroform interface." Biophys J. 95(3): 1099-1107.

Liu, D. C. and J. Nocedal (1989). "On the Limited Memory Bfgs Method for Large-Scale Optimization." Mathematical Programming 45(3): 503-528.

Ma, H., Wan, C., Wu, A., Zewail, A. H. (2007). PNAS 104(712): 14.

Ma, H., Wan, C., Wu, A., Zewail, A. H. (2007). "DNA folding and melting observed in real time redefine the energy landscape." Proc Natl Acad Sci U S A 104(3): 712-716.

Mann, S. (2009.). "Self-assembly and transformation of hybrid nano-objects and nanostructures under equilibrium and non-equilibrium conditions." Nat. Mater. 8(10): 781-792.

Maurer, N., Wong, K.F., Stark, H, Louie, L., McIntosh, D. (2001). "Spontaneous entrapment of polynucleotides upon electrostatic interaction with ethanoldestabilized cationic liposomes." Biophysical Journal 80: 23102326.

Montesi, A., Pasquali, M., MacKintosh, F. C. (2004). "Collapse of a semiflexible polymer in poor solvent." PHYSICAL REVIEW E 69(021916).

Munro, L. J. and D. J. Wales (1999). "Defect migration in crystalline silicon." Physical Review B 59(6): 3969-3980.

Nocedal, J. (1980). "Updating Quasi-Newton Matrices with Limited Storage." Mathematics of Computation 35(151): 773-782.

Noy, A., A. Perez, et al. (2007). "Theoretical study of large conformational transitions in DNA: the B¡–¿A conformational change in water and ethanol/water." Nucleic Acids Res 35(10): 3330-3338.

Orozco, M., A. Perez, et al. (2003). "Theoretical methods for the simulation of nucleic acids." Chem Soc Rev 32(6): 350-364.

Padrta, P., Stefl, R., Krlk, L., Zdek, L., Sklenr, V. (2002). "Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings." Journal of Biomolecular NMR 24: 1-4.

Pereira, G. G., Williams, D.R.M.; (2001). "Toroidal Condensates of Semiflexible Polymers in Poor Solvents: Adsorption, Stretching, and Compression." Biophysical J. 80 161168.

Perez, A., F. J. Luque, et al. (2012). "Frontiers in molecular dynamics simulations of DNA." Acc Chem Res 45(2): 196-205.

Perez, A., I. Marchan, et al. (2007). "Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers." Biophys J 92(11): 3817-3829.

Perez, A. and M. Orozco (2010). "Real-time atomistic description of DNA unfolding." Angew Chem Int Ed Engl 49(28): 4805-4808.

Portella, G., M. W. Germann, et al. (2014). "MD and NMR Analyses of Choline and TMA Binding to Duplex DNA: On the Origins of Aberrant Sequence-Dependent Stability by Alkyl Cations in Aqueous and Water-Free Solvents." J Am Chem Soc.

Portella, G., Orozco, M. (2010). "Multiple Routes to Characterize the Folding of a Small DNA Hairpin." Angew. Chem. Int. Ed. 49: 7673 7676.

Rueda, M., F. J. Luque, et al. (2006). "G-quadruplexes can maintain their structure in the gas phase." Journal of the American Chemical Society 128(11): 3608-3619.

Rueda, M. K., S. G.; Luque, F. J.; Orozco, M (2003). "The structure and dynamics of DNA in gas phase." J. Am. Chem.Soc 125: 80078014.

Shields, G. C. L., C. A.; Orozco, M. (1997). "Molecular Dynamics Simulations of the d(TAT) Triple Helix." J. Am. Chem. Soc. 119: 7463.

Skjorringe, T., Gjetting, T., Jensen, T.G. (2009). "A modified protocol for efficient DNA encapsulation into pegylated immunoliposomes (PILs)." Journal of Controlled Release 139: : 140145.

Sugita, Y., Okamoto Y. (1999). "Replica-exchange molecular dynamics method for protein folding." Chemical Physics Letters 314(1-2): 141151.

Trygubenko, S. A. and D. J. Wales (2004). "A doubly nudged elastic band method for finding transition states (vol 120, pg 2082, 2004)." Journal of Chemical Physics 120(16): 7820-7820.

Varnai, P., and, Zakrzewska, K. (2004). "DNA and its counterions: a molecular dynamics study." Nucleic Acids Res 32(14): 4269-4280.

Wales, D. J. "OPTIM: A program for optimising geometries and calculating pathways. http://www-wales.ch.cam.ac.uk/software.html.

Wales, D. J. "PATHSAMPLE: A program for generating connected stationary point databases and extracting global kinetics. http://www-wales.ch.cam.ac.uk/software.html."
Wales, D. J. (2002). "Discrete path sampling." Molecular Physics 100(20): 3285-3305.

Wales, D. J., M. A. Miller, et al. (1998). "Archetypal energy landscapes." Nature 394(6695): 758-760.

Watson, J. D., Crick, F.H. (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." Nature Rev. Mol. Cell Biol. 171: 737-738.

Wilkins, M., Stokes, A.R., Wilson, H.R. (1953). "Molecular structure of deoxypentose nucleic acids." Nature Rev. Mol. Cell Biol. 171: 738-740.

Part II

The DNA in the gas phase

Chapter 4

Structure of Triplex DNA in the Gas Phase

Annalisa Arcella, Guillem Portella, Maria Luz Ruiz, Ramon Eritja, Marta Vilaseca, Valérie Gabelica, Modesto Orozco, 2012, *J.Am.Chem.Soc.*, 134(15), 6596-6606

Abstract

Extensive (more than 90 microseconds) molecular dynamics simulations complemented with ion-mobility mass spectrometry experiments, in collaboration with Dr.Gabelica of Bordeaux University, have been used to characterize the conformational ensemble of DNA triplexes in the gas phase. Our results suggest that the ensemble of DNA triplex structures in the gas phase is well-defined over the experimental time scale, with the three strands tightly bound, and for the most abundant charge states it samples conformations only slightly more compact than the solution structure. The degree of structural alteration is however very significant, mimicking that found in duplex and much larger than that suggested for G-quadruplexes. Our data strongly supports that the gas phase triplex maintains an excellent memory of the solution structure, well-preserved helicity, and a significant number of native contacts. Once again, a linear, flexible, and charged polymer as DNA surprises us for its ability to retain three-dimensional structure in the absence of solvent. Results argue against the generally assumed roles of the different physical interactions (solvent screening of phosphate repulsion, hydrophobic effect, and solvation of accessible polar groups) in modulating the stability of DNA structures.

4.1 Introduction

Most of the available structural information on biomacromolecules (among them nucleic acids) has been obtained by high-resolution experimental techniques such as X-ray crystallography and NMR spectroscopy. In favorable cases, these methods solve the structure of biomacromolecules with uncertainties below 0.1 nm. However, despite all the power of X-ray crystallography and NMR spectroscopy, we cannot ignore that the practical use of these techniques is limited by many technical problems, such as the difficulty to obtain crystals or well-defined nuclear overhauser effect (NOE) maps, often making necessary the use of low-resolution methods as sources of structural data.

Mass spectrometry (MS) experiments have become one of the sources of such low-resolution structural information, both through electrospray-MS (ESI-MS) and through its combination with ion mobility spectrometry (IMS-MS) (Fenn 1993; Gale 1995; Banks and Whitehouse 1996; Benjamin, Robinson et al. 1998; Hofstadler and Griffey 2001; Heck and Van Den Heuvel 2004; Benesch and Robinson 2006; Koeniger, Merenbloom et al. 2006; Hopper and Oldham 2009; Wyttenbach 2013).

MS is fast and requires low sample consumption, but in contrast to other techniques, the information is recorded from samples in the gas phase, which raises the question of to what extent does gas phase structural information reflect the most populated conformation in solution. This question will become even more crucial in the near future when X-ray free electron lasers (Neutze 2004) will provide high-resolution structural information of molecules in the gas phase.

A series of studies with proteins (Fenn 1993; Gale 1995; Banks and Whitehouse 1996; Benjamin, Robinson et al. 1998; Hofstadler and Griffey 2001; Heck and Van Den Heuvel 2004; Benesch and Robinson 2006; Koeniger, Merenbloom et al. 2006; Hopper and Oldham 2009; Wyttenbach 2013) have demonstrated that if vaporization is done under mild conditions, the structure becomes stable during long periods of time, and gas phase ensembles can be used to accurately model the solution structure (Patriksson, Marklund et al. 2007; Meyer, de la Cruz et al. 2009). In other words, despite the absence of the hydrophobic effect, protein structure remains quite stable in the gas phase.

The question is whether these findings also stand for a highly flexible and charged nonglobular molecule as DNA, whose conformation is known to be very dependent on the solvent environment (Perez 2010). In this respect, early ESI-MS experiments by different authors (Gale 1995; Gabelica, De Pauw et al. 1999; Wan, Gross et al. 2000; Hofstadler and Griffey 2001; Reyzer, Brodbelt et al. 2001; Rosu, Gabelica et al. 2002; Gidden, Ferzoco et al. 2004; Patriksson, Marklund et al. 2007; Meyer, de la Cruz et al. 2009) provided convincing evidence that the two strands of a DNA duplex remain

bound in the gas phase and that even noncovalent drug-DNA complexes remain stable in the complete absence of water.

Molecular dynamics (MD) simulations suggested that duplexes are severely distorted when transferred to the gas phase (Rueda 2003; Rueda, Luque et al. 2005) but confirmed that the two strands remain bound by several native contacts, maintaining the general helical shape at least in the sub-microsecond time scale.

Very surprisingly, similar calculations with G-quadruplex DNA (four-stranded DNA helices) suggested a full maintenance of native helical structure in the gas phase, at least in the microsecond time scale, (Rueda 2006) a striking finding that is fully consistent with ESI-MS and ESI-IMS-MS experiments (Vairamani and Gross 2003; Gabelica 2005; Galeva, Esch et al. 2005; Gabelica, Rosu et al. 2007; Rosu, Gabelica et al. 2010) The behavior of the other canonical DNA structure, the DNA triplex, in the gas phase remains unknown.

DNA triplexes are formed when a single strand of DNA (named as the triplex forming oligonucleotide; TFO) recognizes a polypurine segment of duplex DNA (the triplex target sequence; TTS) through specific major-groove-mediated hydrogen bonds. The third strand can be either polypurine (antiparallel triplexes) or polypyrimidine (parallel triplexes).

Under normal laboratory conditions, the parallel triplexes are much more stable than the antiparallel ones, (Washbrook and Fox 1994; Chandler and Fox 1996; Scaria and Shafer 1996), and accordingly, they are better characterized both experimentally (Chandler and Fox 1996; Scaria and Shafer 1996; Bartley, Brown et al. 1997; Koshlap 1997; Asensio, Brown et al. 1998; Phipps, Tarkoy et al. 1998; Tarkoy, Phipps et al. 1998; Lane 2001; Patriksson, Marklund et al. 2007; Schultze 1997) and theoretically (Shields 1997; Soliva, Luque et al. 1999; Cubero, Guimil-Garcia et al. 2001; Spackova 2004).

Several NMR and MD studies (Bartley, Brown et al. 1997; Koshlap 1997; Shields 1997; Asensio, Brown et al. 1998; Phipps, Tarkoy et al. 1998; Tarkoy, Phipps et al. 1998; Soliva, Luque et al. 1999; Cubero, Guimil-Garcia et al. 2001; Lane 2001; Orozco, Perez et al. 2003; Schultze 1997) have shown that triplex DNA forms a right-handed helix with general features close to B-form of DNA, including bases perpendicular to the helical axis, a narrow minor groove (mG), and south sugar puckerings.

The most obvious effect of the third strand is the partition of the duplex major groove (MG) in two new grooves: the major part of the major groove (MMG), similar in width to the major groove of DNA, but with quite different recognition properties, and

the minor part of the major groove (mMG), a very narrow groove that can coordinate small polar groups (Shields 1997; Cubero, Guimil-Garcia et al. 2001).

The presence of the third strands has very important consequences in the functionality of the DNA. The most important one is the dramatic modification of its ability to be recognized by specific major-groove binder proteins (Shields 1997; Jimnez 1998). For example, DNA repairing systems do not recognize triplex DNAs well, explaining the use of triplexes to generate random mutations in DNA (Havre, Gunther et al. 1993).

Furthermore, proteins involved in gene regulation also have large difficulties recognizing DNA sequences when placed on triplexes, and in fact triplexes generated in promoter regions can knock-down or even knock-out the expression of the corresponding genes (Curiel 1992; Soyfer 1995; Majumdar, Khorlin et al. 1998; Majumdar 1998; Shen, Buck et al. 2003). This behavior, combined with the high density of triplex target sequences in promoters (Goi 2006) explains the interest of triplexes for antigene therapy (Curiel 1992; Havre, Gunther et al. 1993; Soyfer 1995; Chan and Glazer 1997; Majumdar, Khorlin et al. 1998; Majumdar 1998; Rosu, Gabelica et al. 2002; Shen, Buck et al. 2003; Goi 2006).

To our knowledge, first ESI-MS studies of the DNA triplex structures in the gas phase are due to Rosu et al.,(Rosu, Gabelica et al. 2002) who found that the triplex remains stable in the gas phase under mild vaporization conditions. More detailed studies confirmed early results and suggested that triplex maintains noncovalent binding properties in the gas phase (Rosu, Nguyen et al. 2007; Rosu, De Pauw et al. 2008; Wan 2008) which points toward a certain maintenance of its original structure in solution.

Unfortunately, MS experiments alone cannot determine the magnitude of the structural degradation induced by vaporization, whether they are large as in duplex DNA or very small as found for G-quadruplex DNA.

In this chapter, I present a very extensive atomistic MD study on two stable parallel triplexes (12-mer and 18-mer) based on the (T,C) motif, i.e., triplexes based on the repetition of $d(TC^+)d(GA) \cdot d(TC)$ triads (the dot indicates WatsonCrick base pairing, while the dash stands for Hoogsteen base pairing).

The computational results presented here, supported by experimental measurements, demonstrate that DNA triplex is not unfolded in the gas phase, but on the contrary, retains approximate helical structure and a non-negligible amount of native contacts.

However, the magnitude of structural distortion found for triplexes is very important, similar to that in duplexes (Rueda 2003; Vairamani and Gross 2003; Rueda, Luque

et al. 2005; Rueda 2006) and larger than that detected for G-quadruplexes (Rueda 2006).

The present study completes the atlas of gas phase structures of DNA and confirms that basic concepts on the role of electrostatic repulsion on DNA structure need to be revisited.

4.2 Methods

4.2.1 Molecular Dynamics Simulations in Solution

The 12-mer and 18-mer systems were simulated in aqueous solution to establish a reference state for the structural study of their behavior in the gas phase. Initial structures were built using a previously optimized template for triplexes (Shields 1997; Soliva, Luque et al. 1999) and were subsequently solvated in a truncated octahedron box and neutralized with 33 sodium ions for triplex 12-mer, and 51 for the 18-mer, distributed around the phosphate groups at optimum positions according to Poisson-Boltzmann calculations.

The resulting 12-mer system contains roughly 5300 water molecules, and the 18-mer system encloses approximately 9300 water molecules. The 12-mer system was simulated at 300 K, whereas the 18-mer was simulated both at 300 K and at 372 K. The systems were equilibrated using our multistep protocol (Jorgensen 1983; Shields 1997) and were simulated for 100 ns. Transferable interaction potential 3 point (TIP3P) parameters for water and Dangs parameters for ions were used (Jorgensen 1983; Smith 1994); previous studies (Noy 2009) demonstrated that no changes in the simulations can be expected related to the addition of 50200 mM extra salt or to the use of alternative water or ion models.

4.2.2 Gas-Phase Simulations

I have used snapshots collected after 5/10 ns of MD simulations in water of the corresponding triplex as starting guess structures in the gas phase (typical heavy atoms root mean square deviation (RMSd) between snapshots around 0.05 nm). I made the reasonable assumption that all cytosines in the Hoogsteen strand remained protonated after the triplexes have been transferred to the gas phase and that extra protons would migrate to the phosphate groups, without cytosine deprotonation.

At this point two crucial decisions had to be taken: (i) the total charge at DNA and (ii) the protonation state. For the present simulations, I selected the expected most populated charge ions that showed complete desolvation and declustering (loss of all ammonium adducts as ammonia): -6 for the 12-mer, -7, -8 for the 18-mer.

The determination of the protonation state (i.e., which phosphates are protonated) is more complex, since the total number of different states for a given charge is equal to $\frac{m!}{[n!(m-n)!]}$, where n is the number of protons that are required for a given total charge and m is the number of phosphates (all Hoogsteen cytosines are considered protonated).

The simplest way to solve the problem is to assume a distributed neutralization, i.e., to assign equal fractional charge to all phosphates; this leads to the distributed charged scheme, (Rueda 2003) which was found to be reasonable for other forms of DNA.

However, for triplexes, I found it is necessary to consider phosphate-phosphate hydrogen bond across the minor part of the major groove (mMG), which requires individual titration of phosphates (the localized charge scheme (Rueda 2003).

This was done using a Monte Carlo procedure (in the order of 500 million Metropolis tests were done), from which I retained the 10 lowest energy configurations, assuming that they are representative of the ensemble of quasi-degenerated charge states detected experimentally for a given total charge.

A second important decision in defining the simulation conditions is the temperature of the system. Because of the loss of kinetic energy during vaporization and the unclear partition of thermal energy among internal and external degrees of freedom, there is no clear connection between simulation and experimental temperatures.

Thus, following our previously used strategy (Vairamani and Gross 2003; Gabelica 2005), I considered two working temperatures: T = 300 K, which simulates ideally mild vaporization conditions, and a higher temperature (T = 372 K), which was used to simulate the heating of ions that could have occurred in the electrospray source region, (Gabelica 2005) as well as ion heating inside the traveling wave IMS cell (Morsa, Gabelica et al. 2011).

Contrary to drift tube ion mobility spectrometry, where the ion temperature is equal to the bath gas temperature, significant field heating of the ions can occur in traveling wave ion mobility spectrometers as used here, leading to an effective temperature equal to the following eq 4.1 (Shvartsburg 2008)

$$T_{eff} = T_{gas} + \frac{3\pi}{128k^2T}(1+M)\left(\frac{zeE}{\rho\Omega}^2\right) \approx T_{gas} + C\frac{z^2}{\Omega^2} \tag{4.1}$$

where T_{gas} is the bath gas temperature (298K), k is the Boltzmann constant, M is the mass of the gas, z is the ion nominal charge, e is the charge of an electron, E is the effective electric field, ρ is the gas number density, m is the ion mass, and Ω is the ion collision cross section.

The effective temperature data of Morsa et al. (Morsa, Gabelica et al. 2011) were used to extrapolate the value of the constant C at the working conditions considered in this work. This extrapolation leads, for the ions of the highest mobility (or z/Ω ratio), to an effective temperature (T_{eff}) around 372K, which was enforced in my simulations.

After 3 ns of thermalization in gas phase, the systems were simulated for 1 μs. In order to investigate equilibration issues, I extended the simulation time of one of the systems up to 30 μs. Finally, to check for dependence on microscopic initial conditions, five replicas obtained using different starting coordinates and velocities were followed for a selected charge state of the 12-mer.

Simulation Details

PARMBSC0 revision of amber parm99 force field (Cheatham 1999; Perez, Marchan et al. 2007) was used to describe DNA, taking parameters for protonated cytosine (Soliva, Luque et al. 1999) and neutralized phosphates from previous works (Rosu, Gabelica et al. 2002; Rueda 2003; Gidden, Ferzoco et al. 2004; Rueda, Luque et al. 2005; Rueda 2006). As found in other studies,(Rueda 2003) preliminary tests showed that scaling down the charges at neutral nucleosides by a factor of 0.8 (to better simulate gas phase charge distribution) does not significantly change the results.

Simulations in water were carried out using periodic boundary conditions and the particle mesh Ewald method (Darden T. 1993) for long-range electrostatic treatment (0.12 nm grid size), combined with a cutoff radius of 1 nm for LennardJones interactions. No cutoff was used for either electrostatic or Lennard-Jones interactions in the gas phase.

All aqueous simulations were carried out at the isothermalisobaric ensemble (T = 300 K and P = 1 atm) using Berendsen thermostat and barostat (Berendsen 1984). The SHAKE algorithm (Ryckaert 1977) was used to constrain all bond distances to their equilibrium value, allowing a 2 fs integration time step for solution conditions.

Since forces are stronger in the absence of water, a more conservative 1 fs integration step (using also SHAKE) was used for gas phase calculations. All calculations in solution were carried out using AMBER9.0 computer program (Case 2005), while gas phase calculations were performed using the most efficient GROMACS 4.5 program (Hess 2008).

Analysis of Data

A wide variety of tools have been used to process more than 90 μs of data. This includes standard geometrical analysis, stacking and hydrogen bond calculations performed using in house programs, clustering analysis by AMBER9/GROMACS 4.5.3 tools, and theoretical determination of collision cross sections (CCS; the experimental observable), which was carried out for the different MD ensembles using the Sigma program, which uses a projection approximation procedure (Wyttenbach 1997; Wyttenbach 2000; Shvartsburg 2008).

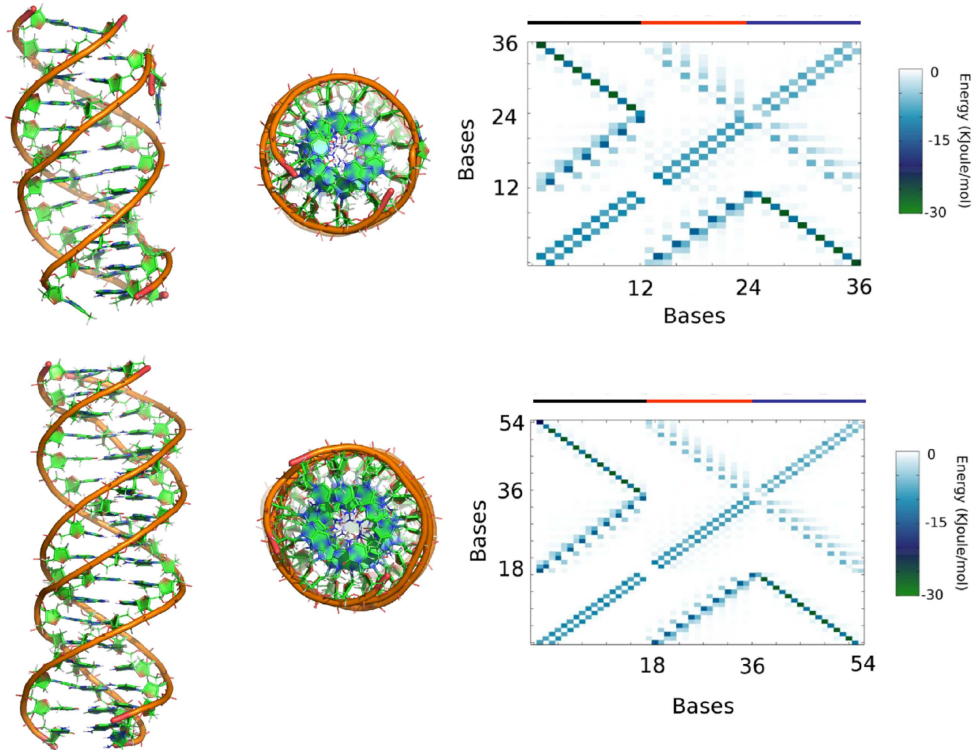

FIGURE 4.1: Triplex structure in aqueous solution at room temperature. Left: orthogonal views of the MD averaged (last 5ns) structures obtained for d(TC+)6d(GA)6d(TC)6 (12-mer at the top) and d(TC+)9d(GA)9d(TC)9 (18-mer at the bottom). Right: energy-based contact maps for 12- and 18-mer triplexes (energy values in kcal/mol). Reference lines are added to label the three strands: black for the WatsonCrick (WC), red for the WC pyrimidine, and blue for the Hoogsteen (H) one.

4.3 Control of Simulations in Aqueous Solvent

MD simulations performed using a fully hydrated environment yield stable trajectories in the 0.1 μs time scale for both the 12- and 18-mer triplexes. Helicity is fully preserved (see Figure 4.1), defining a structure with a general B-shape form (see Figures 4.1 and 4.2, Table 4.1), as previously suggested by shorter MD simulations and NMR experiments

(see the Introduction) and in clear disagreement with early models derived from X-ray fiber diffraction data (see ref (Shields 1997) for discussion).

Sugar puckerings are in general in the south to southeast region, but pyrimidine nucleotides reach around 9-11% of the north region, indicating that sugar puckering is quite flexible in triplexes. The largest structural distortions found are located at the ends of the helices, where fraying of a protonated cytosine is detected in both 12- and 18-mer simulations (see Figures 4.1 and 4.2).

		12-mer	18-mer
RMSD (nm)	all	2.3 ± 0.4	2.23 ± 0.33
	central	1.5 ± 0.2	2.05 ± 0.29
% H-bonds (all)	WC	96	96
	Hoogsteen	91	89
% H-bonds (all)	WC	99	100
	Hoogsteen	100	99
Av.Phase Ang.	$d(TC^+)_6$	100 ± 26	100 ± 44
	$d(GA)_6$	139 ± 28	139 ± 28
	$d(TC)_6$	108 ± 26	116 ± 45
Groove width (nm)	minor	1.2 ± 0.1	1.2 ± 0.2
	minor-Major	1.1 ± 0.06	1.1 ± 0.05
	Major-Major	2.05 ± 0.06	2.6 ± 0.06
Twist	GA	31.34 ± 3.38	31.5 ± 3.6
	AG	30.53 ± 4.15	29.6 ± 3.6
Roll	GA	1.74 ± 5.05	1.8 ± 4.9
	AG	3.73 ± 5.32	2.77 ± 5.4
Slide	GA	-1.2 ± 0.3	-1.2 ± 0.4
	AG	1.1 ± 0.5	1.2 ± 0.5

TABLE 4.1: Key structural descriptors of the 12-mer and 18-mer triplexes in aqueous solution obtained by averaging the last 50 ns of 100 ns trajectories. For groove nomenclature see Shields et al. (Shields,G.C., Laughton,C.A.,& Orozco,M. (1997) J. Am. Chem. Soc. 119, 7463). Distances are in nm and angles in degrees. Standard deviations in the average are also displayed.

At 372 K, the triplex shows indications of unfolding (data not shown), but the slow kinetics of the process preclude a direct visualization of these events in the accessible simulated time scale. Very interestingly, end-fraying effects involving WatsonCrick interactions are largely reduced compared to the situation in duplexes, affecting only marginally d(TAT) triads.

In summary, extended MD simulations suggest that the aqueous triplex is very stable at room temperature, is quite rigid, and displays a general B-like conformation that fits well with the experimental and previous MD data (Bartley, Brown et al. 1997; Koshlap 1997; Shields 1997; Asensio, Brown et al. 1998; Gotfredsen 1998; Phipps, Tarkoy et al. 1998; Tarkoy, Phipps et al. 1998; Soliva, Luque et al. 1999; Schultze 1997)

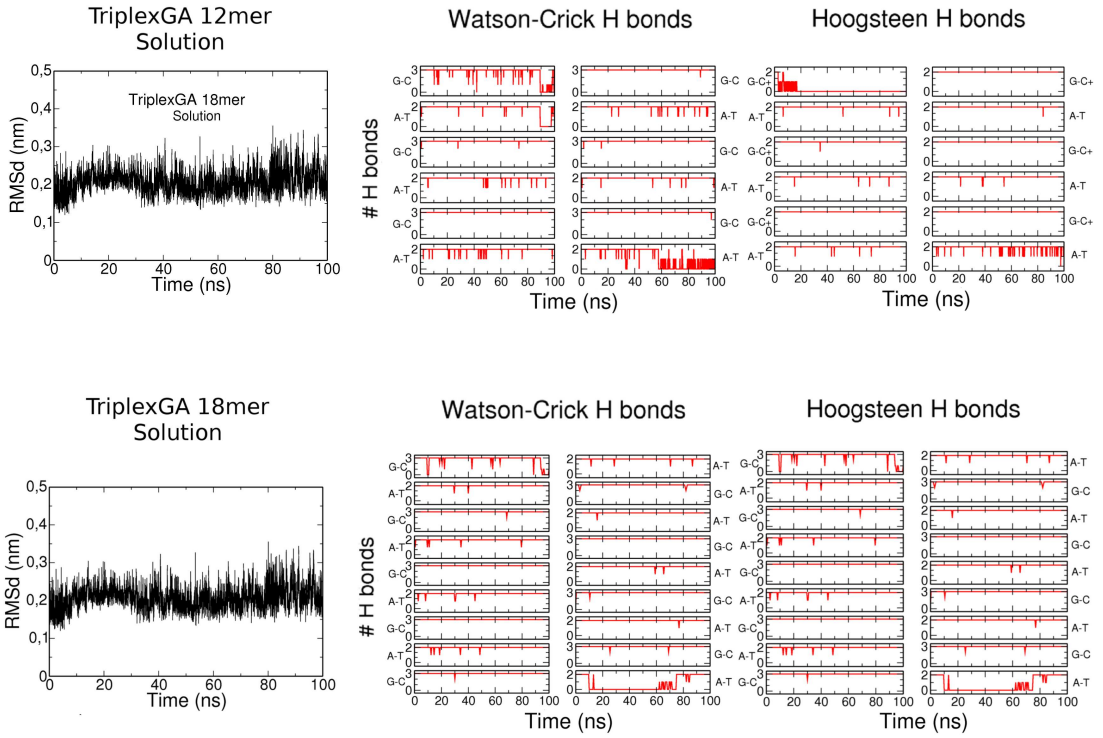

FIGURE 4.2: Time evolution of selected structural descriptors of the 12-mer (TOP) and 18-mer (BOTTOM) triplexes in aqueous solution. Left panels, root mean square deviation curves (RMSds) from canonical triplex structure. Middle and Right panels correspond to numbers the Watson Crick and Hoogsteen hydrogen bonds for the different base pairs, respectively

4.4 Gas Phase Simulations

As discussed above, I planned to explore the diversity of the gas phase conformational space by selecting a series of different reasonable simulation conditions, which can define the boundaries of "real experimental conditions" (372 K) and of "ideally mild gas phase conditions" (300 K); see the Methods 4.2. Additionally, I explored the impact of simulation details and of the length of the triplex, with the ultimate goal to characterize the gas phase conformational space of triplexes as accurately and universally as possible.

4.4.1 Time Dependence

The time dependence was explored first, since current computers do not allow us to reach experimental time scales, forcing us to perform multiple shorter simulations (in this paper, 1.0 μs long), which might ignore slow conformational rearrangements.

To check the expected impact of these very slow conformational arrangements in the gas phase, I perform a 30 μs simulation for one charge state (-6) of the 12-mer

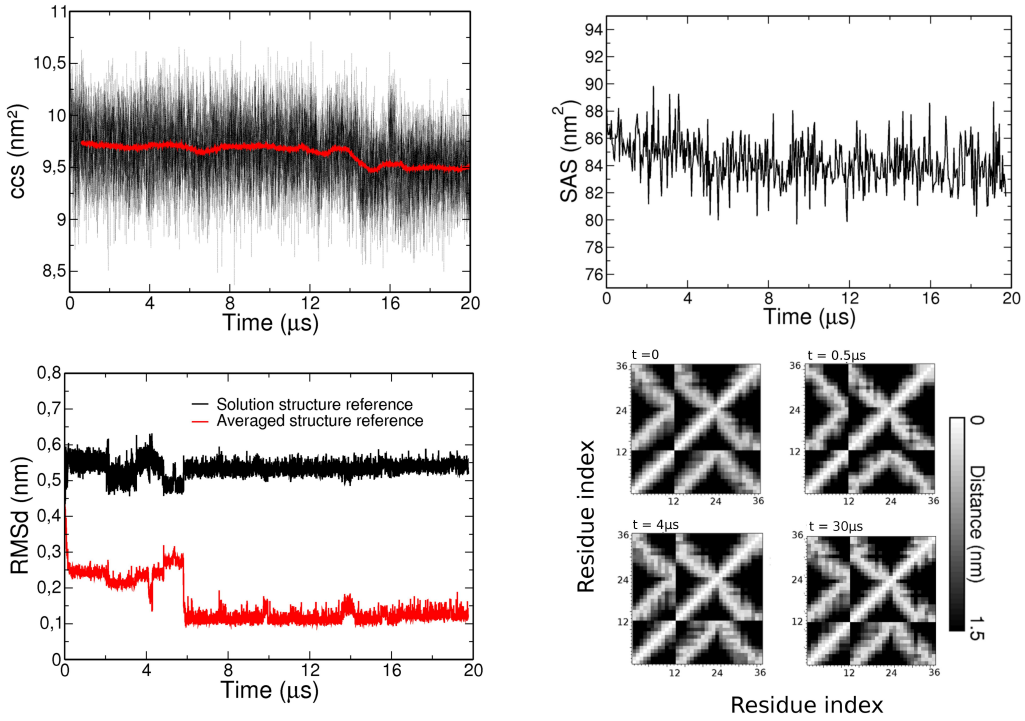

FIGURE 4.3: Key simulation results obtained in the 30 μs trajectory of [12-mer triplex]6- in the gas phase. Left-top: evolution with time of the RMSd (in nm) from the solution structure (black) and the MD-averaged one (red). Right-top: variation of the solvent accessible surface (in nm^2) along time. Left-bottom: evolution with time of the collision cross section. Right-bottom: distance-based contact maps (distances are color coded) obtained from different 5 μs windows along simulation compared to that found in the original solution structure.

triplex (T = 300K). Results (Figures 4.3 and 4.4, 4.6) clearly demonstrate that no strand dissociation happens even after 10^{10} integration of Newtons equations of motion. The helix remains quite well-defined for the entire period, and structural rearrangements from the beginning to the end of the simulation are minuscule.

This is noted (see Figures 4.3 and 4.5, 4.6)) in structural representations, RMSd plots, macroscopic descriptors, or even in the hydrogen bond pattern: 47% of Watson-Crick (WC) and 24% of Hoogsteen (H) hydrogen bonds are maintained after the first μs of MD, values that remain unaltered in the last microsecond of the trajectory (40% of WC and 30% of H hydrogen bonds are maintained).

Very interestingly, I do not observe a loss of structure or an increase in internal mobility as time advances (Figures 4.3 and 4.5, 4.6), and in fact the compactness (as measured by radii of gyration, solvent accessible surface, or CCS) even increases slightly from 1 to 30 μs, suggesting that no unfolding is expected to happen in the millisecond time scale periods either, i.e., in the typical experimental time region.

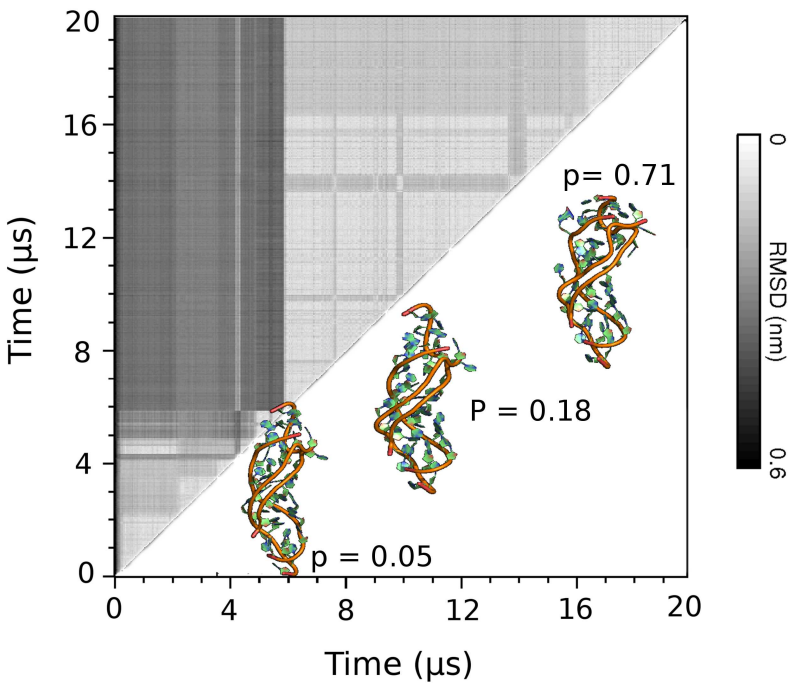

FIGURE 4.4: Bidimensional RMSd map for the 30 μs trajectory of the 12-mer triplex in the gas phase. Representative structure of cluster obtained at different simulation frames are displayed, noting the weight of each cluster in the total ensemble.

In summary, vaporization leads to a very fast transition from the canonical triplex to compact structures, which remain quite stable in the sub-millisecond time scales (looking at the extreme stiffness of the structure, probably even at the millisecond scale). Furthermore, the lack of significant differences between 30 and 1 μs structures strongly suggest that useful information can be derived from shorter simulations as those performed in the rest of this work.

4.4.2 Trajectory Heterogeneity

Trajectory heterogeneity was another source of concern in my simulations. MD is stochastic in nature, and it is a priori unclear whether different trajectories starting from slightly different starting conditions converge to the same ensemble.

To investigate this point, we performed 5×1.0 μs trajectories for the 12-mer duplex using different starting configurations (taken always from the solution ensemble)

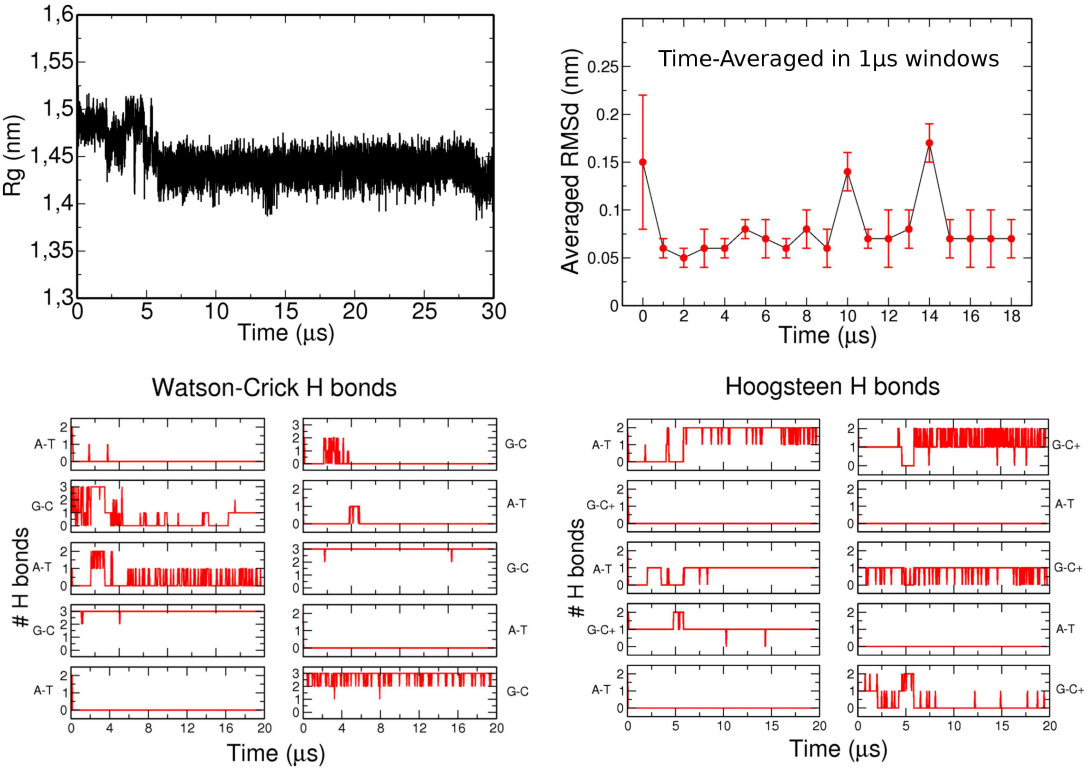

FIGURE 4.5: Time variation of several structural descriptors in the 30μs trajectory of the 12-mer triplex in the gas phase. TOP-LEFT: Radii of Gyration. TOP-RIGHT: average RMSd (with respect to MD-averaged structures) in one microsecond windows. BOTTOM: number of Watson-Crick and Hoogsteen hydrogen bonds.

and velocities (see the 5.2) and a common charge state (for completeness, I selected one different to that studied in the 30 μs simulations).

Results summarized in Figure 4.7 clearly show that trajectories separate in all cases quickly from the solution structure (typical RMSd around 0.7 nm), adopting different stable folds, which are very distant in terms of RMSd but share many common structural characteristics (Figure 4.7). As found in the very long trajectory, the helical structure is well maintained in all replicas, even though the minor grooves are strongly reduced, mostly as a consequence of the formation of hydrogen bonds between charged and neutral phosphates (Figure 4.8).

Additional helical distortions are due to the formation of saline bridges between Hoogsteen cytosines and anionic phosphates. These bridges induce changes in the placement of the nucleobases, which in some cases are extruded outside the pseudohelix, while in others they remain stacked inside.

This arrangement allows a non-negligible amount of canonical hydrogen bonding to be kept, especially the WatsonCrick ones (Figure 4.7), while establishing many other noncanonical interactions.

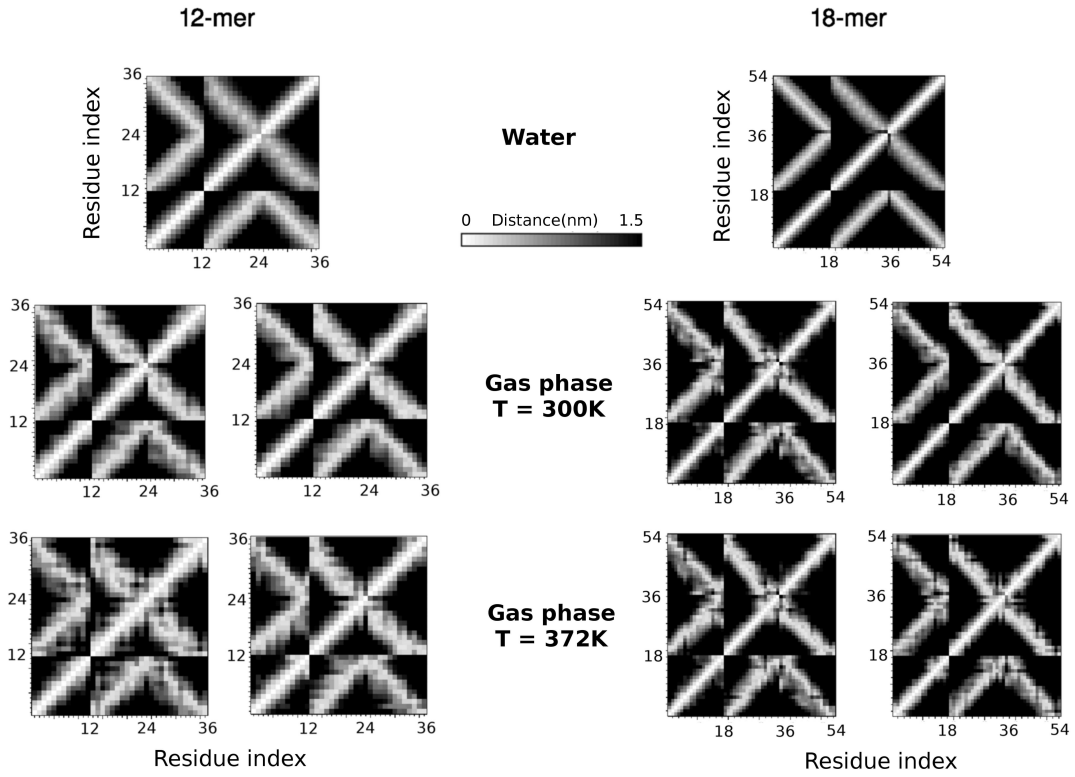

FIGURE 4.6: Examples of contact maps found in gas phase simulations (300K and 372K) for 12-mer and 18-mer triplexes. Charge states displayed were randomly selected. The reference maps obtained in aqueous solution are shown as reference.

Overall, replica simulations illustrate that conformational transitions induced by vaporization have a quite stochastic nature, leading then to quite diverse conformers (in terms of RMSd). However, most of the RMSds are due to terminal movements (for example base flipping) or different packing of nucleobases around the helix axis.

Very good convergence is achieved in the gas phase sampled structures in terms of global descriptors, like overall shape and type of contacts, including those properties experimentally measurable.

4.4.3 Effect of Charge Partitioning at Constant Total Charge

Vaporization is a fast process, and there are many phosphate groups in DNA triplex than can capture protons from the solvent or from the ammonium counterions during the transition to gas phase.

Accordingly, many quasi-degenerated charge distributions can be populated for each charge state, which might lead to different structures. That is, I cannot expect structural diversity due only to the different total charge of the different triplex ions

Replica	Averaged rms (nm)	Averaged ccs (nm²)	Hbond%	
			Watson-Crick	Hoogsteen
1	0.54 ± 0.01	9.6 ± 0.3	42	15
2	0.74 ± 0.02	9.5 ± 0.3	18	6
3	0.70 ± 0.01	9.4 ± 0.2	17	16
4	0.82 ± 0.01	9.7 ± 0.2	22	6
5	0.78 ± 0.01	9.8 ± 0.3	18	16
Av.	**0.6 ± 0.1**	**9.6 ± 0.3**	**23**	**12**

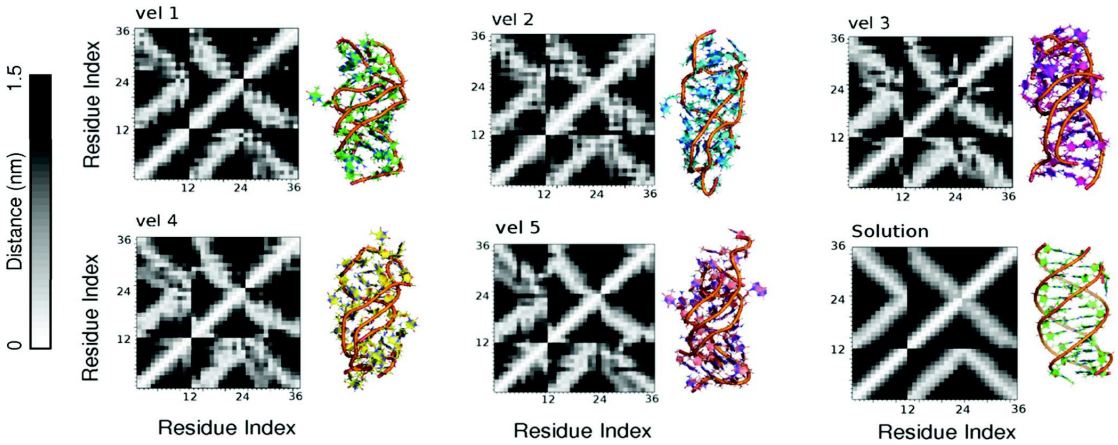

FIGURE 4.7: Key simulation results obtained from 5 replicas of a single charge state of the 12-mer triplex in gas phase. Top: MD-averaged values of RMSd (from solution structure), collision cross-section, and percentage of WatsonCrick and Hoogsteen hydrogen bonds. Bottom: averaged structures obtained in the last 50 ns of the replicas and distance contact maps. As a reference, I have included the water structure and its associated contact map.

(see experimental results below), but also to the existence of different ways to distribute a given total charge over the structure.

In order to cover this source of variability, I selected 10 alternative charge partitionings among the most stable found in our Monte Carlo titration procedure (see the Methods sec 4.2) and performed 1μs MD simulations with all of them.

I detected fast and major structural rearrangements upon vaporization of the 12-mer triplex at room temperature (see RMSd data and Figures 4.8 and 4.9) but not strand detachment or unfolding for any of the studied charge partitioning possibilities.

Gas phase structures are more rigid than the ones obtained from solution (collected at the same temperature), as noted in the mean RMSd with respect to their respective averages (Figure 4.9). The global helicity is maintained, and all the structures collected for the 10 charge partitioning possibilities display a clear memory of the solution conformation (see Figures 4.10 and 4.6).

As described above, the tendency of negatively charged phosphates to interact with neutral phosphates and protonated cytosines led to a dramatic narrowing of the minor and major groove.

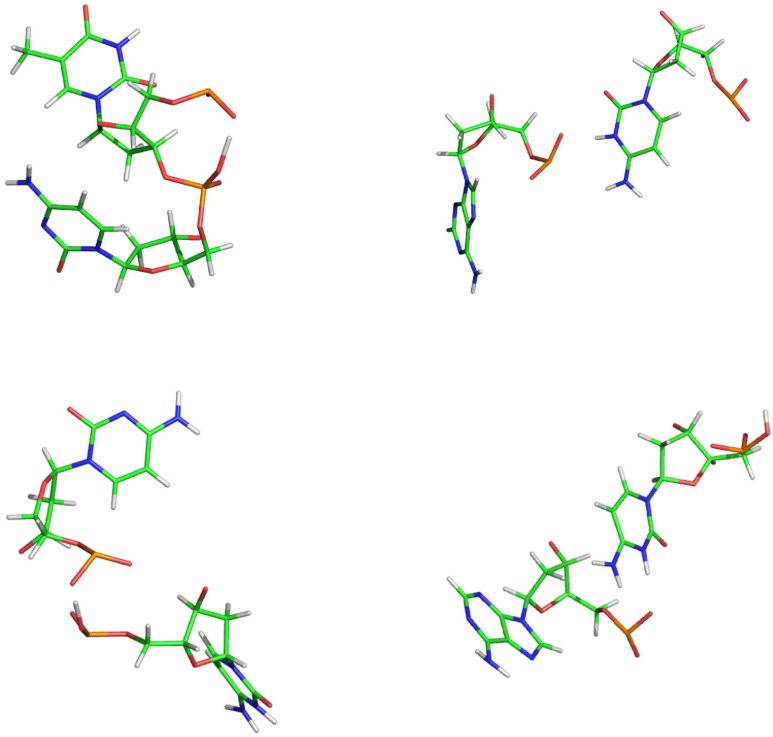

FIGURE 4.8: Schematic representation of typical molecular environments around charged phosphates and protonated cytosines. On the left side the images illustrate hydrogen bond mediated interaction between phosphate groups, whereas on the right side they depict the interaction between protonated cytosines and phosphate groups.

This narrowing causes the disruption of several nucleobase arrangements, some of which become extra-helical, justifying the slight increase in radii of gyration upon vaporization. Around 32% of canonical Watson-Crick and 14% of Hoogsteen hydrogen bonds are preserved in these simulations, values in close agreement with those extracted from multiple replicas and from the very long trajectory. Furthermore, around 50% native contacts are preserved, and many non-native contacts emerge.

This leads to a small but consistent compression (around 5% in the CCS) of the triplex with respect to the solution situation, matching again the results found in the replicas of a single charge state and in the very extended MD simulation. Despite all the heterogeneity and the uncertainties in defining simulation setup, the reproducibility of the overall MD results is impressive.

4.4.4 Validation with a Larger Triplex and Different Total Charges

The overall results obtained for the 18-mer triplex agree with those obtained for the 12-mer (Figures 4.7, 4.10 and 4.6), and they are qualitatively equivalent for the charge states simulated that correspond to mild vaporization conditions (Q = -7 and Q = -8).

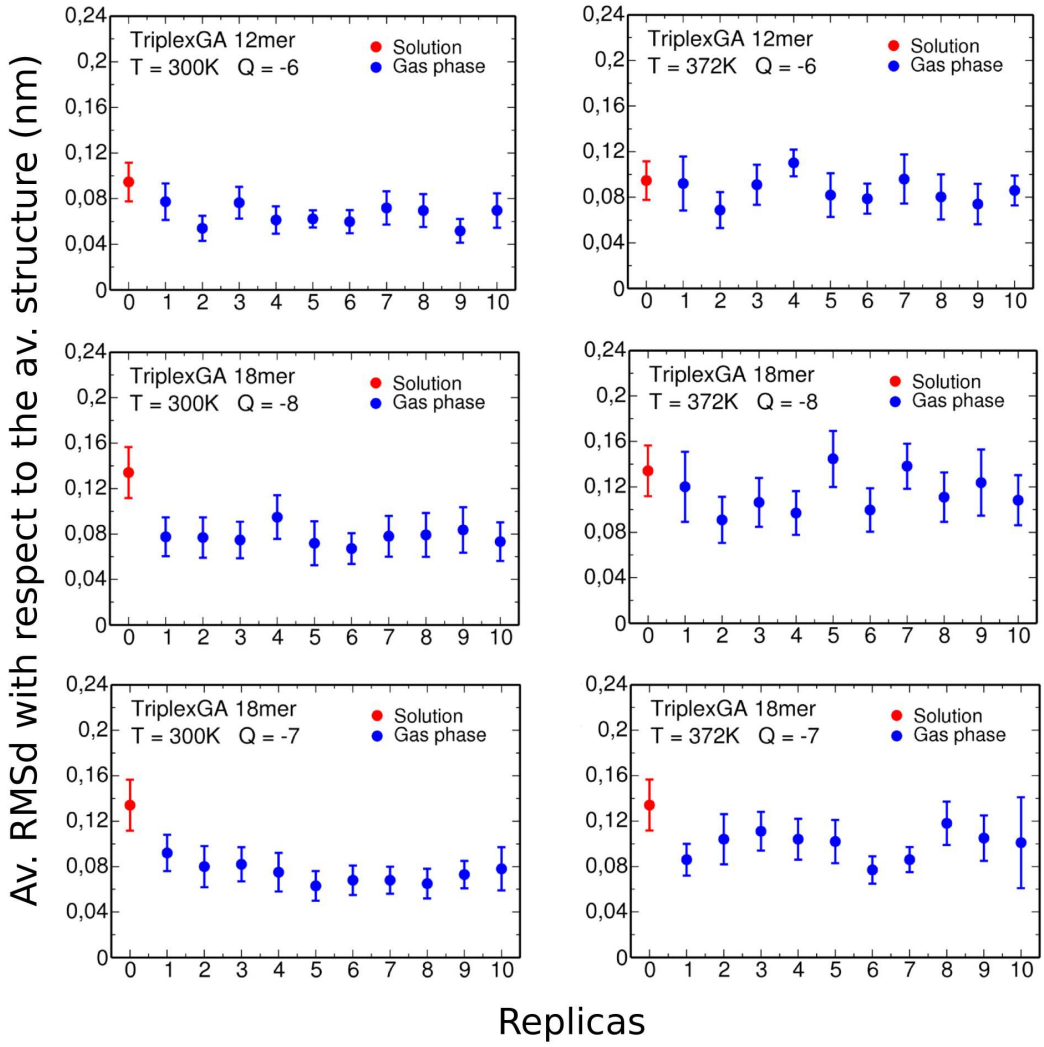

FIGURE 4.9: RMSd (bars correspond to associated standard derivations) with respect to the respective MD-averaged structures obtained for the 10 different charge states of the 12-mer and 18-mer triplexes in the gas phase at two different temperatures and charged states. The values obtained for the 100 ns MD simulations in water are indicated in all cases in red.

Vaporization leads to major changes in the structure (RMSd around 0.7 nm from solution conformation), but the general helical conformation is well maintained. As in the 12-mer simulations, the largest source of distortion is the narrowing of the minor grooves as a consequence of phosphate-phosphate and cytosine+-phosphate contacts (examples in Figure 4.8), which leads again to the disruption of many triads.

A very significant number of canonical hydrogen bonds (around 31 (WC) and 24% (H)) and native contacts (around 43 and 21%, respectively) are maintained, which combined with a myriad of non-native interactions justify the compactness of the structure

(average CCS in vacuum and solution are almost the same), as found for the 12-mer triplex.

In summary, the results seem well converged with respect to the triplex length, and I cannot expect dramatic changes in the behavior of triplexes in the gas phase related to their different sizes, at least in a range of triplex lengths below the persistence length of DNA, which are typically used in mass spectroscopy.

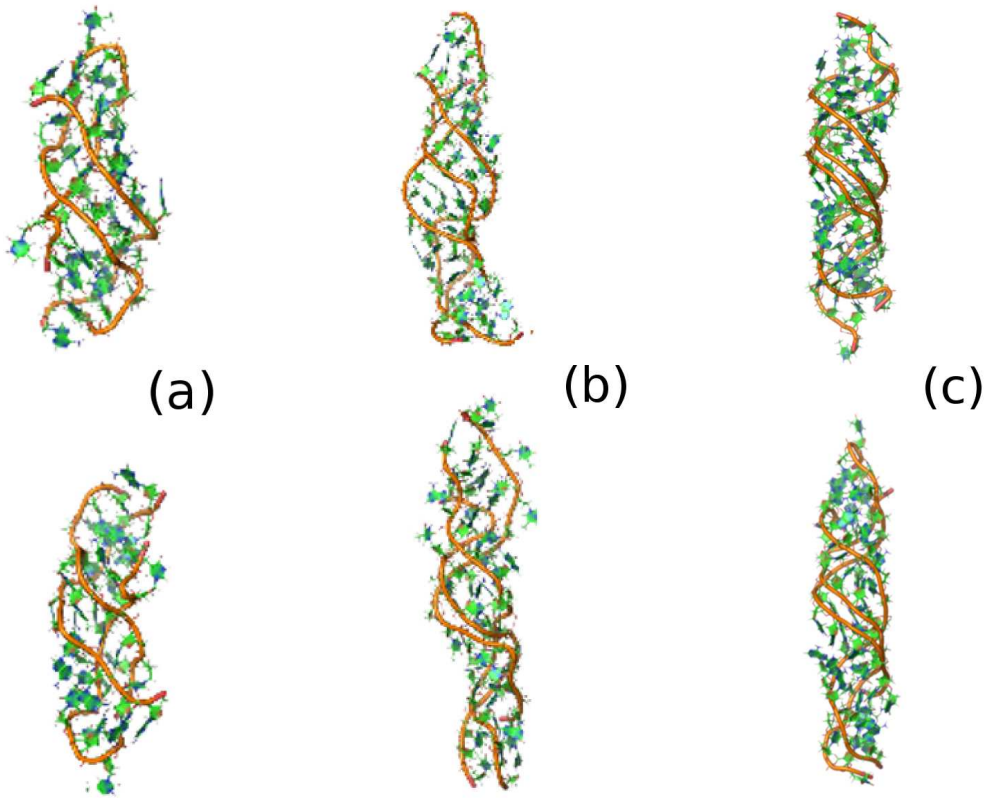

FIGURE 4.10: Representative averaged structures obtained from the last 50 ns of the trajectories of the 10 different charge distributions of the 12-mer and 18-mer triplexes in the gas phase at two different temperatures (above at 300 K, below at 372 K): (a) 12-mer, (b) 18-mer and Q = -7, (c) 18-mer and Q = -8. The RMSd between the shown structures and the rest are below 0.8 nm for each state and temperature.

4.4.5 Effect of Temperature

As commented above, it is difficult to map the spectrometer temperature into a simulation temperature, as the first one is a macroscopic concept, whereas the second is a microscopic magnitude derived exclusively from the internal velocities of the atoms in the macromolecule.

However, I can define a range of temperatures (spanning from T=300K to 372K, see the Methods 4.2) that will cover from the "ideally" mild (low field drift tube ion mobility) to more disruptive (traveling wave ion mobility) conditions.

Thus, to make our conclusions extensible to a reasonable range of temperatures (see the Methods 4.2 and refs (Rueda 2003; Rueda, Luque et al. 2005; Rueda 2006), I repeated the simulation of the 10 charge distributions for 12- and 18-mer triplexes using a higher effective temperature (T = 372 K).

Temperature has a dramatic effect in the conformation of DNA in solution, visible already in the multinanosecond scale (Perez and Orozco 2010; Portella and Orozco 2010), but quite interestingly, it has little impact in the gas phase. All macroscopic descriptors (Rg, or CCS) are quite robust to changes in temperature, and in fact it is difficult to find any clear and systematic difference in the global shapes of the triplexes at low and high temperature (see Figure 4.10).

Contact plots confirm that the global helical structure is not altered (Figure 4.2) when gas phase simulations are run at high temperature and the only difference is related to the reduction of directional hydrogen bond contacts due to thermal vibrations.

In summary, gas phase structures seem to be quite resistant to temperature variations within reasonable margins, and mass spectrometric experiments are probably not far from following ideally mild vaporization conditions, supporting the reliability of simulation conditions followed in the paper.

4.5 Experimental Validation

ESI-MS data clearly confirm previous results (Chandler and Fox 1996; Rosu, Nguyen et al. 2007; Wan 2008) suggesting that vaporization does not lead to a disruption of the two triple helices considered here and that despite phosphatephosphate repulsion, the three strands remain tightly bound. Figure 4.11 shows that the triplexes sprayed from acidic (pH = 3) conditions (see the Method 4.2) were ionized at charge states -5, -6, and -7 for the 12-mer (Figure 4.11A) and from -6 to -9 for the 18-mer (Figure 4.11B).

When samples at neutral pH were sprayed (not shown), the main charge states were -6, -7, and -8 for the 12-mer triplex and -7, -8, and -9 for the 18-mer triplex. The collision cross sections of the triplexes were measured as a function of the IMS cell bias voltage (Figure 4.11C,D).

Increasing this voltage causes the ion population to heat up transiently, because of collisions with the nitrogen gas leaking out of the IMS cell, before being thermalized

at around 372 K in the IMS cell itself. We see that the CCS decreases slightly when the bias voltage is increased. The CCS were also measured both for the naked triplex (no ammonium adduct, filled symbols) and for the whole ammonium adduct distribution when observable (open symbols).

The CCS was found to increase very slightly and steadily with ammonium adducts, suggesting no significant conformational changes. The values of the CCS obtained at bias = 20 V are summarized in Table 4.2. Results obtained with the different experimental protocols (different sample preparation, instrument tuning, and calibration curve) agree.

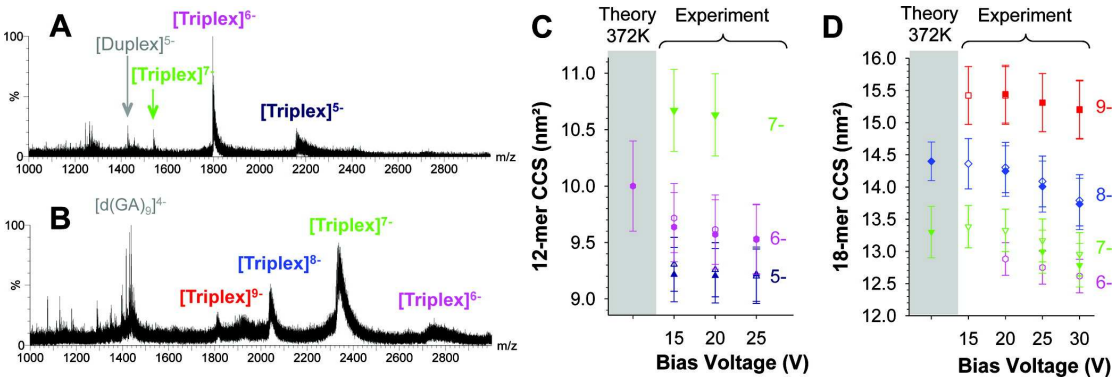

FIGURE 4.11: Comparison between simulations and experiments. (A,B) ESI-MS spectra of the 12-mer triplex (A) and 18-mer triplex (B) obtained from acidic conditions at a bias voltage of 20 V. (C,D) Experimental and theoretical collision cross sections obtained for the different charge states of the 12-mer triplex (C) and 18-mer triplex (D), as a function of the bias voltage. Dark blue, 5-; pink, 6-; green, 7-; navy blue, 8-; red, 9-; filled symbols, triplex with no ammonium adducts; open symbols, triplexes including the whole adduct distribution visible on the respective mass spectra.

Comparison between experimental CCS and values derived from the MD-derived structural samples can help to detect significant errors in our simulations. Nevertheless, also in case of agreement these results should be taken with caution. The sources of error in the experimental determination of the CCS with traveling wave IMS (errors in the published CCS and errors in traveling wave IMS measurements) are taken into account by using the 95% prediction interval, as described in Figure **??**, which are reported in Figure 4.11C,D and Table 4.2.

Also, there are expected errors in atomistic MD simulations, and probably more importantly there are intrinsic errors in the algorithms used to transform coordinates into CCS, which would probably require recalibration for nucleic acids as was already done for peptides (Siu 2009).

The analysis of the 55 μs of MD ensembles of the 12-mer triplex gives a theoretical CCS of $(10.0 \pm 0.4) nm^2$ (total charge Q = -6), and the analysis of the cumulated 60 μs of MD trajectories for the 18-mer triplex leads to a theoretical CCS of $(13.3 \pm 0.4) nm^2$ for Q = -7 and $(14.4 \pm 0.3) nm^2$ for Q = -8. These values are all close to the experimental

estimates at bias voltages '$20V$: $(9.6 \pm 0.3)nm^2$ for the 12-mer, $(13.3 \pm 0.4)nm^2$ (Q = -7) and $(14.4 \pm 0.5)nm^2$ (Q = -8) for the 18-mer (see Table 4.2).

For the most abundant charge states, the results reproduce the collision cross section increase when the total charge of the ion increases. Keeping in mind all the sources of noise in theoretical or experimental CCS determination, the agreement between MD (using the projection approximation, see Method section 2.11.1) and IMS-MS data is probably partially fortuitous but clearly supports the validity of our MD-derived conclusions and in general the use of last generation MD methods to obtain structural information under conditions where no high resolution structural data is available.

Caution is however necessary to avoid overestimation of the capability of simulating different ions. Thus, for very charged ions, I cannot expect that classical MD simulations with simple force-fields and fixed topology can provide accurate results. In fact, test calculations of Q = -9 (for the 18-mer oligo) yield too compact structures in the microsecond range compared to experimental observable.

Triplex	Charge	Theor.CCS (372K)	Exp.CCS (pH=7)	Exp.CCS (pH=3)
12-mer	in solution	10.2 ± 0.2		
	-5			9.2 ± 0.2
	-6	10.0 ± 0.3	9.7 ± 1.1	9.6 ± 0.3
	-7		10.2 ± 1.3	10.6 ± 0.4
	-8		11.8 ± 1.5	
18-mer	in solution	14.3 ± 0.5		
	-6			12.9 ± 0.3
	-7	13.2 ± 0.1	13.4 ± 1.3	13.3 ± 0.3
	-8	14.3 ± 0.1	14.2 ± 1.5	14.3 ± 0.4
	-9		15.0 ± 1.8	15.3 ± 0.5
	-10			16.6 ± 0.5

TABLE 4.2: Theoretical (MD) and experimental estimates of CCS (nm^2) for the triplexes, at bias = 20 V using neutral and acidic conditions. The experimental errors were calculated from the 95% prediction interval of the respective calibration curves, meaning that the true value has 95% chance of lying in this interval. The theoretical error bars were determined by propagation of standard errors in the different estimates obtained from individual trajectories, charge states and temperatures.

4.6 Conclusions

Simulations presented here provide an extended description of the nature of DNA triplexes in the gas phase and complete the atlas of canonical structures of DNA in the gas phase. Our simulations do not support the hypothesis that the triplex adopts a well-defined structure in the gas phase, as found for the DNA G-quadruplex. On the

contrary, I suggest that it adopts a large number of stable but distorted structures, more similar to the situation found for duplex DNA.

It is likely that different triplexes bearing the same total charge might display different distributions of charges in the molecule, which can evolve in the gas phase to different structures, and that even different molecules with the same distribution of charges might evolve to adopt different conformations.

However, all these structures, which can be quite distinct in terms of RMSd, are close in terms of general conformational characteristics and macroscopic structural descriptors such as the collision cross section (CCS).

The different metastable structures defining the gas phase ensemble are very rigid and do not interconvert even in the sub-millisecond time scale explored here. Interestingly, the results from the unprecedented amount of MD simulations performed for this work demonstrate that despite major changes in conformation, the triplexes maintain a clear memory of the helical structure in solution when transferred to the gas phase.

Irrespective of the temperature and length of the oligonucleotide, a significant amount of native interactions are maintained upon vaporization, including many canonical hydrogen bonds, which combined with a myriad of non-native nucleobase and backbone interactions (especially important are those involving negatively charged phosphates) avoid the elongation of the triplex and leads to a small reduction of CCS with respect to solution state.

The good agreement between experimental and calculated CCS suggests that these structural models adequately describe the triplexes produced by ESI-MS at their most abundant charge states.

Very interestingly, this work shows evidence that the key physical interactions that stabilize the secondary structure of nucleic acids, namely, screened Coulombic repulsion and the nucleobase-nucleobase hydrophobic terms, does not lead to the complete unfolding of a highly charged and nonglobular polymer such as triplex DNA when these interactions are tuned by the lack of solvent.

On the contrary, when transferred to the gas phase, the DNA triplex maintains very clear signatures of the conformation in solution. This finding opens interesting possibilities for using gas phase structural data to infer solution structure of nucleic acids

Bibliography

Asensio, J. L., T. Brown, et al. (1998). "Comparison of the solution structures of intramolecular DNA triple helices containing adjacent and non-adjacent CG.C+ triplets." Nucleic Acids Res 26(16): 3677-3686.

Banks, J. F., Jr. and C. M. Whitehouse (1996). "Electrospray ionization mass spectrometry." Methods Enzymol 270: 486-519.

Bartley, J. P., T. Brown, et al. (1997). "Solution conformation of an intramolecular DNA triplex containing a nonnucleotide linker: comparison with the DNA duplex." Biochemistry 36(47): 14502-14511.

Benesch, J. L. and C. V. Robinson (2006). "Mass spectrometry of macromolecular assemblies: preservation and dissociation." Curr Opin Struct Biol 16(2): 245-251.

Benjamin, D. R., C. V. Robinson, et al. (1998). "Mass spectrometry of ribosomes and ribosomal subunits." Proc Natl Acad Sci U S A 95(13): 7391-7395.

Berendsen, H. J. C. P., J. P. M.; van Gunsteren, W. F.; Dinola, A.; Hook, J. R. (1984). "Molecular dynamics with coupling to an external bath." J. Chem. Phys. 81.

Case, D. A. D., T.; Gohlke, H.; Luo, R.; Merz, J. R.; K., M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. (2005). "The Amber biomolecular simulation programs." J. Comput. Chem. 26: 1668 1688.

Chan, P. P. and P. M. Glazer (1997). "Triplex DNA: fundamentals, advances, and potential applications for gene therapy." J Mol Med (Berl) 75(4): 267-282.

Chandler, S. P. and K. R. Fox (1996). "Specificity of antiparallel DNA triple helix formation." Biochemistry 35(47): 15038-15048.

Cheatham, T. E., III, Cieplak, P., and, Kollman, P.A. (1999). "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat." J. Biomol. Struct. Dyn. 16: 845-862.

Cubero, E., R. Guimil-Garcia, et al. (2001). "The effect of amino groups on the stability of DNA duplexes and triplexes based on purines derived from inosine." Nucleic Acids Res 29(12): 2522-2534.

Curiel, D. T. W., E.; Cotten, M.; Birnstiel, M. L.; Agarwal, S.; Li, C. M.; Loechel, S.; Hu, P. C. (1992). "High-Efficiency Gene Transfer Mediated by Adenovirus Coupled to DNAPolylysine Complexes." Hum. Gene Ther. 147.

Darden T., Y., D., Pedersen, L. (1993). "Particle Mesh Ewald-an N.Log(N) method for Ewald sums in large systems." J. Chem. Phys. 98: 10089-10092. Fenn, J. B. (1993). "Ion formation from charged droplets: roles of geometry, energy, and time." J Am Soc Mass Spectrom 4,: 524535.

Gabelica, V., E. De Pauw, et al. (1999). "Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry." J Mass Spectrom 34(12): 1328-1337.

Gabelica, V., F. Rosu, et al. (2007). "Base-dependent electron photodetachment from negatively charged DNA strands upon 260-nm laser irradiation." J Am Chem Soc 129(15): 4706-4713.

Gabelica, V. R., F.; Witt, M.; Baykut G.; De Pauw E.; (2005). "Fast gas-phase hydrogen/deuterium exchange observed for a DNA G-quadruplex." Rapid Commun Mass Spectrom 19(2): 201-208

Gale, D. C. S., R. D. . (1995). "Characterization of noncovalent complexes formed between minor groove binding molecules and duplex DNA by electrospray ionization-mass spectrometry." J. Am. Soc. Mass. Spectrom 6(12): 1154.

Galeva, N. A., S. W. Esch, et al. (2005). "Rapid method for quantifying the extent of methionine oxidation in intact calmodulin." J Am Soc Mass Spectrom 16(9): 1470-1480.

Gidden, J., A. Ferzoco, et al. (2004). "Duplex formation and the onset of helicity in poly d(CG)n oligonucleotides in a solvent-free environment." J Am Chem Soc 126(46): 15132-15140.

Goi, J. R. V., J. M.; Dopazo, J.; Orozco, M (2006). "Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions." BMC Genomics 7(63).

Gotfredsen, C. H. S., P.; Feigon, J. (1998). "Solution Structure of an Intramolecular Pyrimidine-Purine-Pyrimidine Triplex Containing an RNA Third Strand." J. Am. Chem. Soc. 120: 4281.

Havre, P. A., E. J. Gunther, et al. (1993). "Targeted mutagenesis of DNA using triple helix-forming oligonucleotides linked to psoralen." Proc Natl Acad Sci U S A 90(16): 7879-7883.

Heck, A. J. and R. H. Van Den Heuvel (2004). "Investigation of intact protein complexes by mass spectrometry." Mass Spectrom Rev 23(5): 368-389.

Hess, B. K., C.; van der Spoel, D.; Lindahl, E. (2008). "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." J. Chem. Theory Comput. 4.

Hofstadler, S. A. and R. H. Griffey (2001). "Analysis of noncovalent complexes of DNA and RNA by mass spectrometry." Chem Rev 101(2): 377-390.

Hopper, J. T. and N. J. Oldham (2009). "Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability." J Am Soc Mass Spectrom 20(10): 1851-1858.

Jimnez, E. V., A.; Espins, M. L.; Soliva, R.; Orozco, M.; Bernus, J.; Azorn, F. (1998). "The GAGA factor of Drosophila binds triple-stranded DNA." J. Biol. Chem. 273: 24640.

Jorgensen, W. L. C., J.; Madura, J. D.; Impey, R. W.; Klein, M. L. (1983). "Comparison of simple potential functions for simulating liquid water." J. Chem. Phys. 79(926).

Koeniger, S. L., S. I. Merenbloom, et al. (2006). "Evidence for many resolvable structures within conformation types of electrosprayed ubiquitin ions." J Phys Chem B 110(13): 7017-7021.

Koshlap, K. M., Schultze, P., Brunar, H., Dervan, P. D., Feigon, J. (1997). "Solution Structure of an Intramolecular DNA Triplex Containing an N7-Glycosylated Guanine Which Mimics a Protonated Cytosine." Biochemistry 26: 2659.

Lane, A. N. J., T. C. Curr. Org. Chem. (2001). "Structures and Properties of Multi-stranded Nucleic Acids." Curr. Org. Chem. 5: 845.

Majumdar, A., A. Khorlin, et al. (1998). "Targeted gene knockout mediated by triple helix forming oligonucleotides." Nat Genet 20(2): 212-214.

Majumdar, A. K., A.; Dyatkina, N.; Lin, F. L.; Powell, J.; Liu, J.; Fei, Z.; Khirpine, Y.; Watanable, K. A.; George, J.; Glazer, P. M.; Seidman, M. M. (1998). "Triplex-forming oligodeoxynucleotides targeting survivin inhibit proliferation and induce apoptosis of human lung carcinoma cells." Nat. Genet. 20(212).

Meyer, T., X. de la Cruz, et al. (2009). "An atomistic view to the gas phase proteome." Structure 17(1): 88-95.

Morsa, D., V. Gabelica, et al. (2011). "Effective temperature of ions in traveling wave ion mobility spectrometry." Anal Chem 83(14): 5775-5782.

Noy, A. S., I.; Luque, F. J.; Orozco, M. (2009). "The impact of monovalent ion force field model in nucleic acids simulations." Phys. Chem. Chem. Phys. 11: 10596.

Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. Nature 2000, 406, 752

Neutze, R.; Hudt, G.; van der Spoel, D. Radiat. Phys. Chem. (2004), "Potential impact of an X-ray free electron laser on structural biology", 71, 905

Orozco, M., A. Perez, et al. (2003). "Theoretical methods for the simulation of nucleic acids." Chem Soc Rev 32(6): 350-364. Patriksson, A., E. Marklund, et al. (2007). "Protein structures under electrospray conditions." Biochemistry 46(4): 933-945. Perez, A., I. Marchan, et al. (2007). "Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers." Biophys J 92(11): 3817-3829.

Perez, A. and M. Orozco (2010). "Real-time atomistic description of DNA unfolding." Angew Chem Int Ed Engl 49(28): 4805-4808.

Phipps, A. K., M. Tarkoy, et al. (1998). "Solution structure of an intramolecular DNA triplex containing 5-(1-propynyl)-2'-deoxyuridine residues in the third strand." Biochemistry 37(17): 5820-5830.

Portella, G. and M. Orozco (2010). "Multiple routes to characterize the folding of a small DNA hairpin." Angew Chem Int Ed Engl 49(42): 7673-7676.

Reyzer, M. L., J. S. Brodbelt, et al. (2001). "Evaluation of complexation of metal-mediated DNA-binding drugs to oligonucleotides via electrospray ionization mass spectrometry." Nucleic Acids Res 29(21): E103-103.

Rosu, F., E. De Pauw, et al. (2008). "Electrospray mass spectrometry to study drug-nucleic acids interactions." Biochimie 90(7): 1074-1087.

Rosu, F., V. Gabelica, et al. (2002). "Triplex and quadruplex DNA structures studied by electrospray mass spectrometry." Rapid Commun Mass Spectrom 16(18): 1729-1736.

Rosu, F., V. Gabelica, et al. (2010). "Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry." Nucleic Acids Res 38(15): 5217-5225.

Rosu, F., C. H. Nguyen, et al. (2007). "Ligand binding mode to duplex and triplex DNA assessed by combining electrospray tandem mass spectrometry and molecular modeling." J Am Soc Mass Spectrom 18(6): 1052-1062.

Rueda, M., Kalko, S. G., Luque, F. J., Orozco, M. (2003). "The structure and dynamics of DNA in the gas phase." J Am Chem Soc 125(26): 8007-8014.

Rueda, M., F. J. Luque, et al. (2005). "Nature of minor-groove binders-DNA complexes in the gas phase." J Am Chem Soc 127(33): 11690-11698.

Rueda, M., Luque, F. J., Orozco, M. (2006). "G-quadruplexes can maintain their structure in the gas phase." J Am Chem Soc 128(11): 3608-3619.

Ryckaert, J. P. C., G.; Berendsen, H. J. C. (1977). "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes." J. Comput. Phys. 23: 327.

Scaria, P. V. and R. H. Shafer (1996). "Calorimetric analysis of triple helices targeted to the d(G3A4G3).d(C3T4C3) duplex." Biochemistry 35(33): 10985-10994.

Schultze, P. K., K. M.; Feigon, J. (1997). "Solution Structure of an Intramolecular DNA Triplex Containing an N7-Glycosylated Guanine Which Mimics a Protonated Cytosine." Biochemistry 36: 2659.

Shen, C., A. Buck, et al. (2003). "Triplex-forming oligodeoxynucleotides targeting survivin inhibit proliferation and induce apoptosis of human lung carcinoma cells." Cancer Gene Ther 10(5): 403-410.

Shields, G. C. L., C. A.; Orozco, M. (1997). "Molecular Dynamics Simulations of the d(TAT) Triple Helix." J. Am. Chem. Soc. 119: 7463.

Shvartsburg, A. A. S., R. D. (2008). "Fundamentals of Traveling Wave Ion Mobility Spectrometry." Anal. Chem. 80: 9689 9699.

Siu, C.-K. G., Y.; Saminathan, I. S.; Hopkinson, A. C.; Siu, K. W. M. (2009). "Optimization of Parameters Used in Algorithms of Ion-Mobility Calculation for Conformational Analyses." J. Phys. Chem. B 114: 1204 1212.

Smith, D. E. D., L. X. (1994). "Computer simulations of NaCl association in polarizable water." J. Chem. Phys. 100: 3757 3766.

Soliva, R., F. J. Luque, et al. (1999). "Can G-C Hoogsteen-wobble pairs contribute to the stability of d(G. C-C) triplexes?" Nucleic Acids Res 27(11): 2248-2255.

Soyfer, V. N. P., V. N. (1995). Triple Helical Structures. New York, Springer.

Spackova, N. C., E.; Sponer, J.; Orozco, M. (2004). "Theoretical Study of the Guanine → 6-Thioguanine Substitution in Duplexes, Triplexes, and Tetraplexes." J. Am. Chem. Soc. 126: 1464214650.

Tarkoy, M., A. K. Phipps, et al. (1998). "Solution structure of an intramolecular DNA triplex linked by hexakis(ethylene glycol) units: d(AGAGAGAA-(EG)6-TTCTCTCT-(EG)6-TCTCTCTT)." Biochemistry 37(17): 5810-5819.

Turner, K. B., N. A. Hagan, et al. (2006). "Inhibitory effects of archetypical nucleic acid ligands on the interactions of HIV-1 nucleocapsid protein with elements of Psi-RNA." Nucleic Acids Res 34(5): 1305-1316.

Vairamani, M. and M. L. Gross (2003). "G-quadruplex formation of thrombin-binding aptamer detected by electrospray ionization mass spectrometry." J Am Chem Soc 125(1): 42-43.

Wan, C. G., X.; Liu, Z.; Liu, S. (2008). "Studies of the intermolecular DNA triplexes of C+GC and TAT triplets by electrospray ionization Fourier-transform ion cyclotron resonance mass spectrometry." J. Mass Spectrom 43(164).

Wan, K. X., M. L. Gross, et al. (2000). "Gas-phase stability of double-stranded oligodeoxynucleotides and their noncovalent complexes with DNA-binding drugs as revealed by collisional activation in an ion trap." J Am Soc Mass Spectrom 11(5): 450-457.

Washbrook, E. and K. R. Fox (1994). "Alternate-strand DNA triple-helix formation using short acridine-linked oligonucleotides." Biochem J 301 (Pt 2): 569-575.

Wyttenbach, T., Bleiholder, C.; Bowers, M. T. (2013). "Factors Contributing to the Collision Cross Section of Polyatomic Ions in the Kilodalton to Gigadalton Range: Application to Ion Mobility Measurements." Anal. Chem. 85: 2191-2199.

Wyttenbach, T. v. H., G.; Batka, J.; Carlat, D.; Bowers, M. (1997). "Effect of the long-range potential on ion mobility measurements." J. Am. Soc. Mass Spectrom. 8.

Wyttenbach, T. W., M.; Bowers, M. T. (2000). "On the Stability of Amino Acid Zwitterions in the Gas Phase: The Influence of Derivatization, Proton Affinity, and Alkali Ion Addition." J. Am. Chem. Soc. 122: 3458.

Chapter 5

The structure of DNA in the gas phase

Annalisa Arcella, Jens Dreyer, Emiliano Ippoliti, Guillem Portella, Valérie Gabelica, Paolo Carloni and Modesto Orozco, *Angew.Chemie.Int.Ed.*, *in revision*.

In this chapter I will describe an exhaustive gas phase structural analysis of a small DNA model that forms a stable hairpin in solution, by combining ion mobility mass spectroscopy, in collaboration with Dr. Valérie Gabelica in Bordeaux, sub-millisecond classical molecular dynamics and *ab initio* molecular dynamics. The latter, in collaboration with the Paolo Carloni's group in Jülich.

Abstract

The gas phase structure of the DNA duplexes and the nature of the dynamics changes accurring upon vaporization remains unclear. Our results show that the free energy landscape of DNA in vacuum is dominated by compact conformations for low-charged states and extended conformations for high-charged state. Three dynamics ranges are identified: a sub-nanosecond time scale related to electronic rearrangements producing topological changes in the DNA, a nano-to-microsecond time scale where small conformational changes between conformers occur, and a micro-to-millisecond time scale (detected only for highly charged ions) leading to significant rearrangements of the structure. *Ab initio* molecular dynamics simulations show that mobile protons exist not only localized on phosphate groups,but also between phosphates and nucleobases.

5.1 Introduction

DNA duplex structure and dynamics result from a subtle balance between stabilizing interactions (mostly nucleobases hydrogen bonding and stacking) and destabilizing terms (mostly dimerization-induced desolvation and phosphate-phosphate coulombic repulsion). Under physiological conditions the stabilizing terms dominate mostly due to the ability of the solvent to screen phosphate-phosphate repulsions. Coulomb repulsion in the absence of water generates strong tensions in the duplex, which should lead, in principle, to the dissociation of the two strands (Saenger 1984; Blackburn 1990; Sinden 1994; Bloomfield 2000).

However, twenty years ago mass spectroscopy (MS) experiments reveal that under mild electrospray conditions (ESI), DNA sequences that are double-stranded in solution are indeed detected as dimers in the gas phase (Ganem, Li et al. 1993; Light-Wahl, Springer et al. 1993; Doktycz, Habibi-Goudarzi et al. 1994). In ESI-MS from native conditions (sufficiently high ionic strength so that the duplex is formed in the starting solution), partial neutralization of the DNA phosphates by protons or by contaminating sodium ions happens during vaporization, reducing the total charge in the DNA, and accordingly the Coulomb repulsion.

Besides dimer detection, correlations between the gas-phase kinetic stability of these dimers and the base composition and solution stability suggested that a fraction of hydrogen bonding and base stacking interactions were conserved in these dimers (Schnier, Klassen et al. 1998; Gabelica and Pauw 2001; Gabelica and De Pauw 2002).

Ion mobility spectrometry measurements, in which ion shape is probed by friction with a gas under the drag of an electric field, suggested that A-form helices could be preserved in gas-phase d(CG)n duplexes starting with 8 to 10 base pairs (Gidden, Ferzoco et al. 2004), and that B-form helices could be preserved from 22 base pairs and longer (Baker and Bowers 2007).

However, the actual picture must be more complex, because the duplex DNA ion mobility is highly dependent on sequence (Gidden, Ferzoco et al. 2004; Burmistrova, Gabelica et al. 2013) and on internal energy (Burmistrova, Gabelica et al. 2013), and because gas phase coarse-grain observables are not univocally translated to atomic-resolution structures. Deciphering the rules governing the conformational changes induced by full dehydration is still a puzzle that needs to be solved to take advantage of current ESI-MS infrastructure, and specially of the XFEL technology (Neutze 2004).

Recently, atomistic classical molecular dynamics (MD) simulations have been used to describe the nature of the conformational ensemble of different nucleic acids in the gas

phase under conditions resembling those in an electrospray ionization mass spectrometry (ESI-MS) experiment. DNA duplexes and triplexes are maintained in the gas phase, showing memory of the solution conformation, but also severe distortions and structural diversity (Gabelica, De Pauw et al. 1999; Rueda 2003; Arcella 2012). On the contrary, G-quadruplexes structural changes induced by vaporization are minor (Rueda 2006).

The general picture obtained from MD simulations suggest that a fast vaporization freezes the DNA in a long-lived meta-stable conformation, not far away from the native state, where a myriad of native and non-native hydrogen bond and stacking, and ion-dipole interactions alleviate the Coulombic repulsion (Rueda 2003; Breuker and McLafferty 2008; Arcella 2012; Meyer, Gabelica et al. 2013; Avi, Portella et al. 2014).

The previous classical MD simulations in the microsecond regime raise fundamental questions that are crucial for the understanding of vaporization-induced conformational transitions and possibly fragmentation: i) does the covalent structure of the oligonucleotide change in the gas phase due to intra-molecular proton transfer? ii) if intra-molecular proton transfer occurs, what is its impact in the conformational ensemble?, iii) do long time-scale rearrangements (sub-millisecond scale) occur in the gas phase?

I present here a massive effort to fully characterize the conformational space of a small model duplex DNA (a hairpin with two CG pairs joined by a tri-loop), extensively characterized in solution (Ansari, Kuznetsov et al. 2001; Ma 2007; Portella 2010). For this model oligonucleotide We first studied the experimental behaviour in the gas phase by state-of-the-art ion-mobility ESI-MS, obtaining shape descriptors consistent with compact states for the most stable low-charge states, and with more extended structures for high-charge states.

I studied then the dynamics of all relevant charge sub-states of the most prevalent charge states by using massive atomistic MD simulations in (sub)millisecond time scale, finding long-time scale structural rearrangements only for the highly charged states. Finally, in collaboration with Paolo Carloni group in Jülich, our extremely rich ensemble will be used to perform, for the first time, full quantum dynamics on these systems (accumulated time above 0.5 ns), showing that despite the lack of the facilitating role of water, protons are quite mobile in the structure, and many proton-transfer events (mostly centered at the phosphates) occur spontaneously.

Overall, the dynamics of the 'rigid' DNA in the gas phase is more complex than anticipated, and previously accepted concepts need to be revisited.

5.2 Methods

I used a short oligonucleotide (d(GCGAAGC)) as a minimum model of a duplex DNA. Thanks to its size, this 7-mer can indeed be completely characterized both theoretically and experimentally. This oligonucleotide folds under physiological conditions very quickly (micro to millisecond time scale (Ansari, Kuznetsov et al. 2001; Ma 2007; Portella 2010) to form a stable hairpin with a duplex B-form stem (d(GC· GC)), and a d(GAA) triloop (Ansari, Kuznetsov et al. 2001; Padrta 2002; Ma 2007; Portella 2010). Previous studies (Portella 2010; Perez, Luque et al. 2012) have demonstrated the ability of MD simulations with parm99+parmbsc0 force-field (Cheatham 1999; Perez, Marchan et al. 2007) to fully characterize the structure and dynamics of this hairpin in aqueous solution.

5.2.1 Classical Molecular Dynamics Simulations

The folded hairpin from NMR model 1PQT (Padrta 2002) was equilibrated for 100 ns in aqueous solution following the protocol described in Portella et al. (Portella 2010).The trajectory was stable (see Figure 5.1) and provided starting conformations for gas phase simulations below (snapshots taken randomly from the 5-100 ns simulation window).

Based on ESI-MS measures three charge states (Q=-2, Q=-3 (the most populated one) and Q=-4 (the least populated one) were considered in simulations. For each charge state I explored all the possible charge sub-states resulting from phosphate protonation using the localized charge protocol (Rueda 2003; Rueda, Luque et al. 2005; Rueda 2006; Arcella 2012), and the standard force-field parameters (Rueda 2003) for gas phase DNA. I did not include those charge distribution that lead to extreme repulsion resulting from over-concentration of charge in certain regions.

This analysis resulted in a total of five charges sub-states for charge state Q=-2, four sub-states for charge state Q=-3 and four for charge state Q=-4 (these sub-states are labeled as LC#; where LC0 corresponds to the charge distribution with the lowest Coulomb repulsion, the smaller the number # is, the lower the predicted energy for that sub-state). These charge distribution are also supported by previous MS-ESI experiments, which have shown that for 9-nt oligos it is very likely that Q=-2 ion bears charges on each end, and the Q=-3 ion has one additional charge in the center of the sequence (Pan, Verhoeven et al. 2005).

These considerations produced a total of 13 systems to be simulated in the gas phase. To improve sampling for each of them I generated 9 replicas taking solution snapshots as seeds for replica exchange simulations (RExMD; (Sugita 1999)) covering

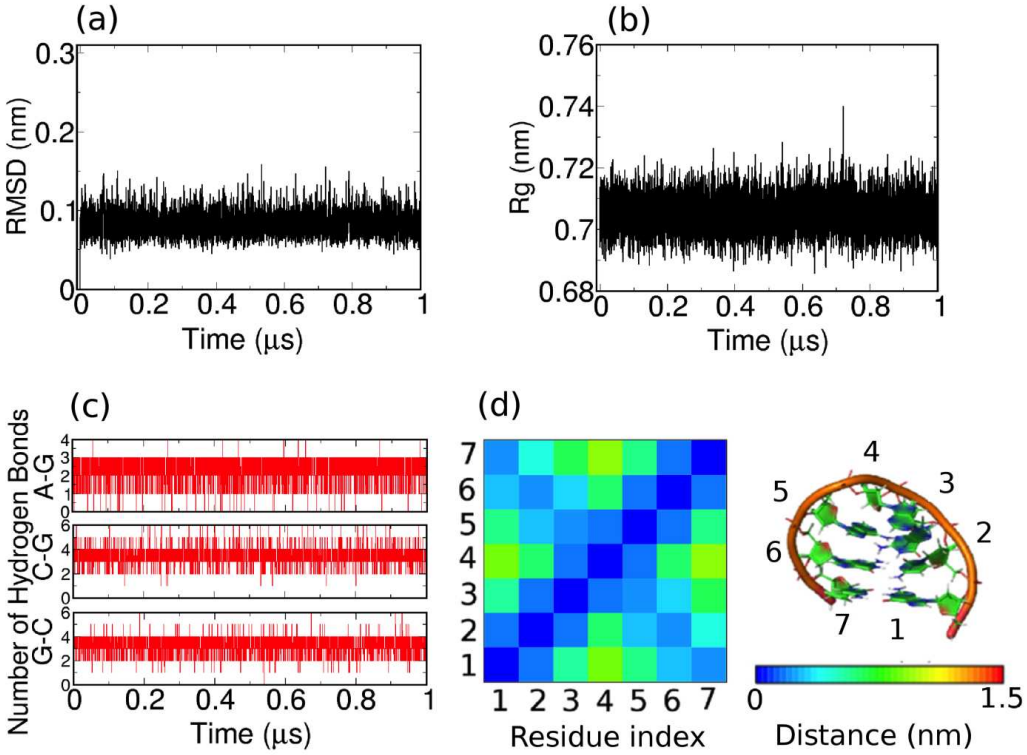

FIGURE 5.1: Panels from (a) to (c) represent variation of structural descriptors for the hairpin in control simulation in water. Panel (d) shows the contact map (note the X shape characteristic of an antiparallel duplex and the density of contacts at the central core) and the averagestructure.

the temperature range from 300 to 800 K (Morsa, Gabelica et al. 2011). Individual replicas were extended to 2.5 μs for a total accumulated time of 22.5 μs per substate, 90-112 μs per state, and 0.3 ms of aggregated simulation time.

In order to explore slow time-dependent conformational transitions for Q=-4 state, I performed extended unbiased MD simulations for the lowest energy sub-states (LC0) for a simulation time of 0.25 ms (Q=-4) using constant temperature (at the predicted effective temperature for ESI-MS T=400 K). All simulations were done using 1 fs integration step and standard simulation details explained in Appendix B

5.2.2 Collission Cross Section calculations from atomistic models

From each trajectoty, I extracted 250 euqally spaced structures and computed the predicted collision cross section (CCS) using the exact hard-sphere scattering (EHSS) appoximation (Shvartsburg 2007) and parameters by Siu et al. (Siu 2010), which were then averaged to obtain the ensemble CCS. Previous studies have shown that this approach provides good representation of MD CCS for nucleic acids (Avino 2014).

5.3 Results

5.3.1 Experimental collision cross sections

The mass spectra of the oligonucleotide is dominated by three peaks corresponding in order of abundance to charge states -3 ≫ -2 > -4 (see Figure 5.2A). The average collision cross sections (CCS) values are very dependent on charge state CCS(Q=-4) ≫ CCS(Q=-3) > CCS(Q=-2), but they are quite independent on the bias voltage, at least in the 10-30 V range (see Figure 5.2B for typical arrival time distributions, and Table A.2 in Appendix A for the full list of values).

Also, despite changes of stability are predicted when ammonium buffer is substituted by pure water, the collision cross sections were identical whether starting from the aqueous solution or from the 150 mM ammonium acetate solution (see Table A.2 in Appendix A).

The invariance of the CCS with regard to solution phase composition or bias voltage already suggests that the structure or distribution of structures inside the IMS cell have somehow lost the memory of the solution phase structure, and that they have done so before entering into the ion mobility cell. This supports that this short hairpin can explore the conformational landscape in the gas phase with little memory bias.

The theoretical CCS estimate obtained using the MD-equilibrated structure at room temperature in aqueous solution is CCS=3.60 nm^2. Transfer to the gas phase therefore collapses the structure for charge state Q=-2, maintains the compactness virtually unmodified for the dominant charge state Q=-3, and clearly enlarges the structure for the highest charge state Q=-4. Similar transitions to very extended structures have been noted before for the oligonucleotide dT10 (Hoaglund, Liu et al. 1997), which on the contrary to our system is incapable of forming Watson-Crick base pairs.

5.3.2 RExMD simulations

RExMD suggests that vaporization produces very significant structural changes (RMSD in the range 0.5-0.7 nm) in the structure of the DNA, with a complete loss of structural memory of the solution conformation (see Figures 5.3 and 5.6, and compare them with Figure 5.1).

The stiff stabilizing interactions within the different sampled conformations requires high temperatures in the RExMD simulations to induce exchange between the conformational states. Despite the diversity of conformers sampled all of them show

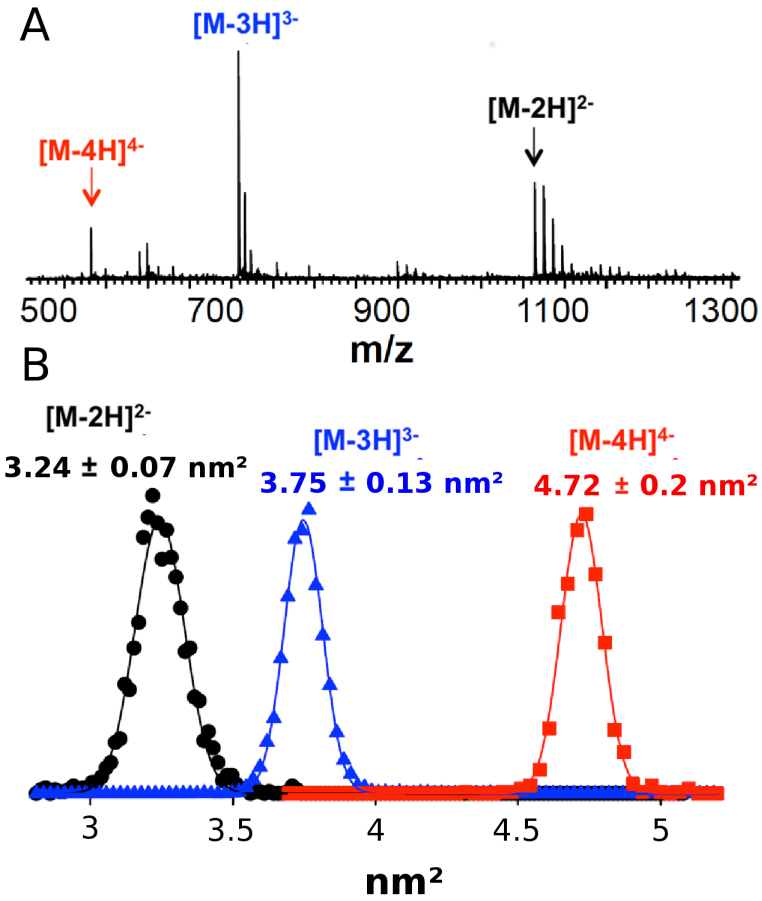

FIGURE 5.2: MS-ESI results a short DNA hairpin. (A) Mass spectrum of 10 μM hairpin electrosprayed from 150 mM $NH_4 4OAc$ solution, with a bias voltage of 15 V. (B) Collision cross section distributions for the three main charge states. The error mentioned next to the CCS is the 95% prediction error obtained from the calibration curve. The full width at half maximum (FWHM) of the distribution is also indicated. For the 3− and 4− charge states, the FWHM is limited by instrumental resolution. For the 2− charge state, the slightly broader CCS distribution indicates conformational heterogeneity.

similar compactness and overall shape for a given charge state, independently of the charge distribution considered (see Figure 5.3 and 5.6).

A remarkable loss of helicity is evident in all cases except for some structures collected in the Q=-3 charge state, the most abundant in ESI-MS mass/charge profile (Figure 5.2), in which a vague memory of helical arrangement appears in contact maps (Figure 5.6). As the temperature increases the structures sampled become more extended, but the effect is rather small, except for the highest charge state Q=-4, for which rising the temperature from 300 to 600 K leads to an increase of more than 0.05 nm^2 in the CCS (Figure 5.3 right panels).

Taking as reference the room temperature or the IMS effective temperatures (Morsa, Gabelica et al. 2011; Arcella 2012), I can conclude that according to MD

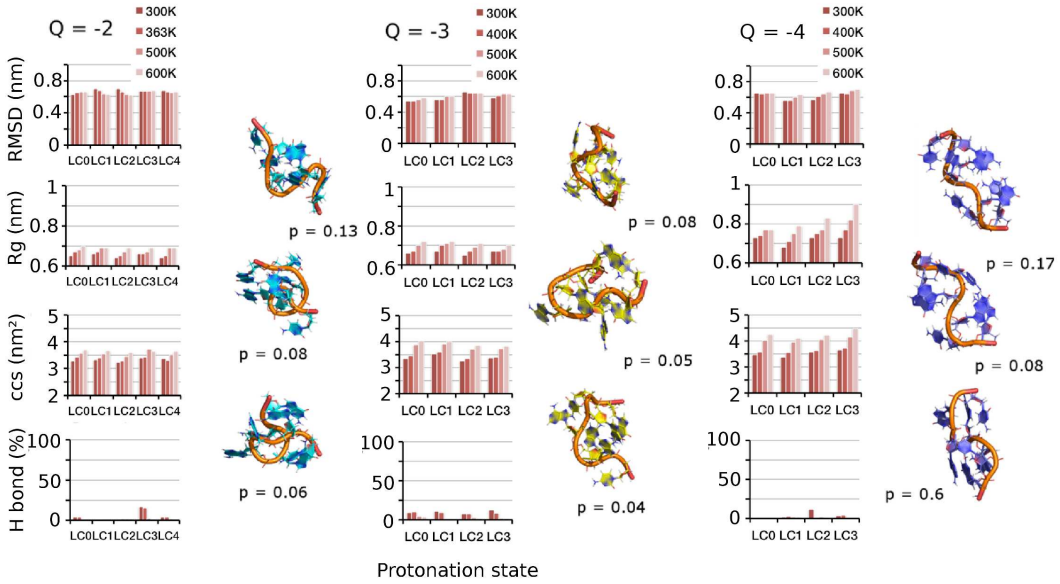

FIGURE 5.3: Summary of RExMD results in the gas phase for charge states (left to right) Q=-2,-3 and -4 for different temperatures (values for different charge substates, from LC0 to LC3 or LC4 are displayed in all cases). At the right of each panel I displayed representative structures of the most populated clusters for all charges states (determined by pooling snapshots from all charge substates) at the MS-ESI effective temperature.The populations (relative to 1) of the cluster are shown.

simulations, vaporization leads to a clear compression of the structure for the lowest charge state (Q=-2), in excellent agreement with experimental measures (predicted CCS (RExMD)=$(3.3 \pm 0.1) nm^2$; experimental value $(3.24 \pm 0.07) nm^2$).

For this charge state, the gas phase structure is severely distorted, with a clear loss (in average around 96%) of canonical hydrogen bonds, which are substituted by stacking interactions, and by a multitude of non-canonical hydrogen bonds; the most prevalent of which are the phosphate(anion)-phosphate(neutral) contacts. These contacts are present in more than 90% of collected snapshots for this charge state (see one typical structure showing phosphate-phosphate hydrogen bonds in Figure 5.4).

For the major charge state Q=-3 CCS derived from RExMD simulations suggest little changes from solution compactness (in qualitative agreement again with experimental measures in Figure 5.2 and Table A.2; predicted CCS (RExMD)= (3.50±0.02) nm^2; experimental value (3.75±13) nm^2), even though the sampled structures appear severely distorted with respect to the solution structure (Figure 5.3 and 5.5), suggesting that caution is needed before assuming that little change in CCS from solution conformation (CCS=3.60 nm^2) necessarily mean structural conservation.

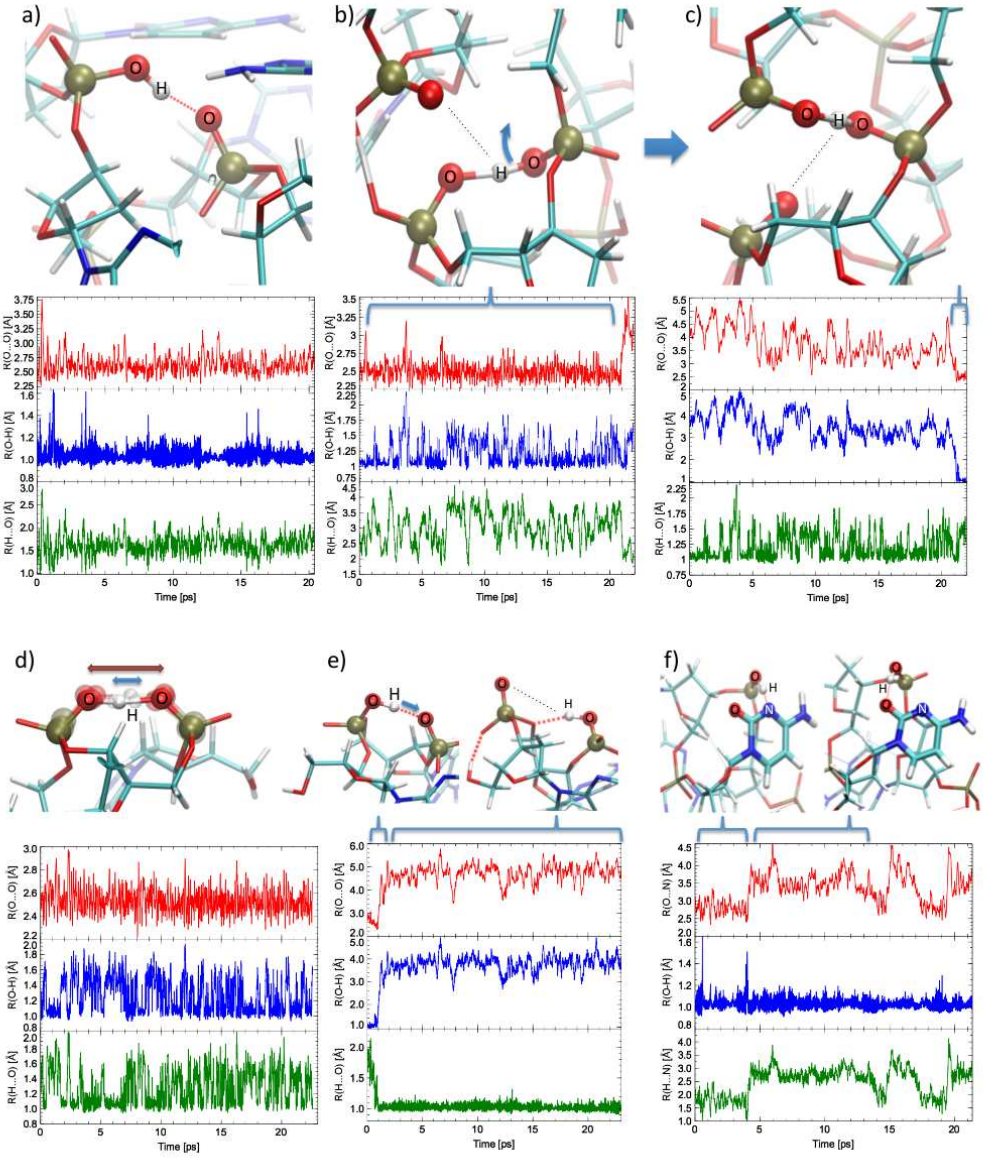

FIGURE 5.4: Examples of possible hydrogen bond configurations between phosphate groups or between phosphate groups and a nucleobase, respectively, with corresponding distance profiles describing potential proton transfer processes.a) no proton transfer;b,c) change of the hydrogen bonding scheme between different phosphate groups without proton transfer; d) the proton is shared between donor and acceptor;e) proton transfer between phosphate groups with subsequent disruption of the hydrogen bond; f) change of hydrogen bonding scheme between phosphate and cytosine (between N3 and C=O).

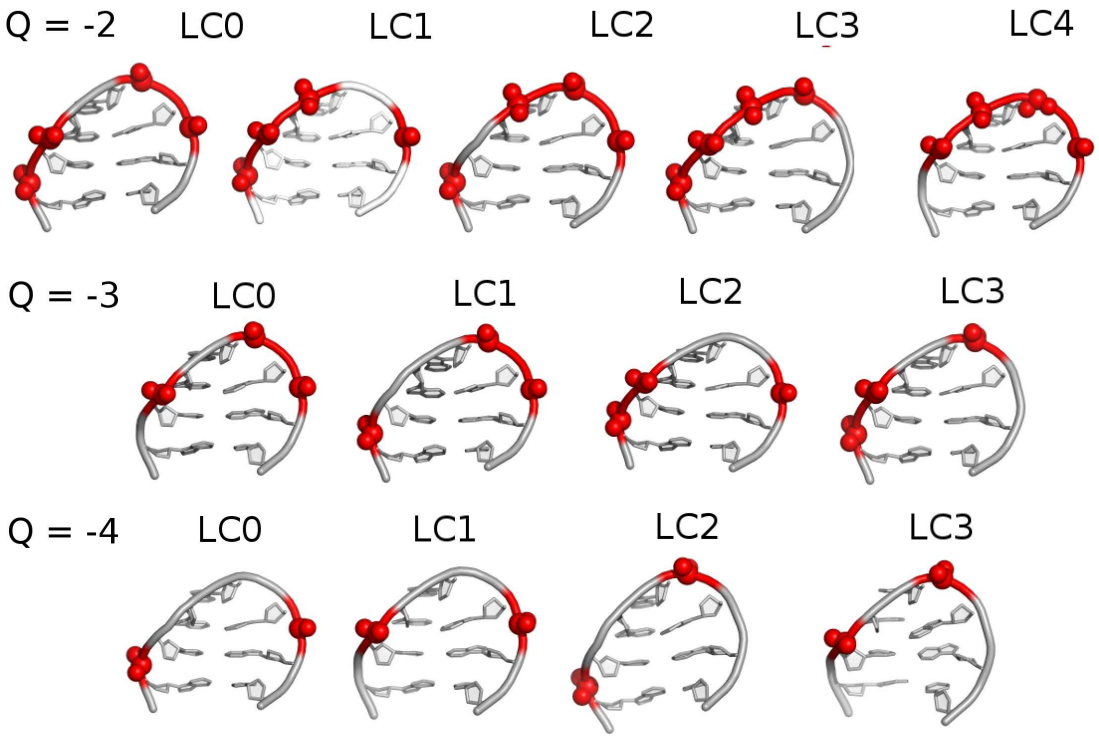

FIGURE 5.5: Schematic representations of the different charge subs-states considered (LC#) for each charge state (Q=-#).I represent as red balls the position of the neutralized phosphates.

The number of preserved canonical hydrogen bonds (around 10% at room, or effective temperatures) is slightly larger than that found for other charge states, showing a residual population of canonical G-C hydrogen bonds, but is still far from those characteristics of a lengthier DNA duplex. Finally, for the least populated high charge state (Q=-4) the type of intramolecular interactions detected by RExMD ensembles are similar to those found in the other charge states, but the structures appeared more elongated, with larger CCS and radius of gyration (Rg), and slightly less densely packed (see Figure 5.3 and 5.5).

It is worth noting that the loss of compactness upon vaporization found for the Q=-4 state agrees qualitatively, but not quantitatively with experimental measures, since ESI-MS experiments suggest a larger extension of the structure at the same effective temperature than that found in simulation (predicted $CCS(RExMD) = (3.6 \pm 0.2) nm^2$; experimental value $(4.72 \pm 0.19)\ nm^2$).

This disagreement between theory and experiment, which contrasts with the similarity found for the other charge states, strongly suggests the existence of slow conformational rearrangements, which are not sampled in our RExMD simulations.

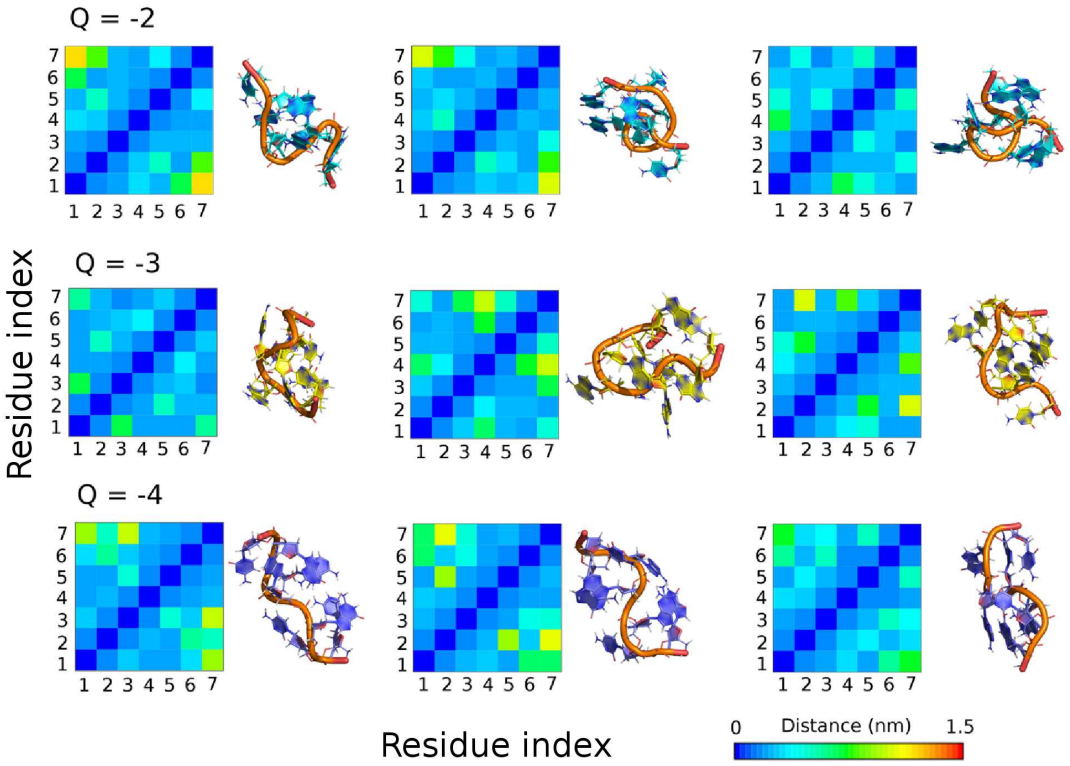

FIGURE 5.6: Contact maps corresponding to representative of major clusters (at right from each contact map) for a given charge state obtained from RExMD simulations at effective MS-ESI temperature (see also Figure 5.3)

5.3.3 Extended MD simulations

To explore the possibility of slow conformational changes in Q=-4 oligonucleotide in the gas phase I performed very extended simulations (see Methods 5.2) reaching time scales not far from the experimental ones (never explored in previous simulations).

Results, summarized in Figure 5.7 demonstrate the existence of very slow cyclic conformational changes happening in the 25-75 μs range, that implies the interchange between compact conformers (CCS clearly below (3.8) nm^2 , as in RExMD simulations) to more extended structures (CCS above (4.4±0.2) nm^2, in much better agreement with the experiment).

Analysis of the trajectory shows that the large change in CCS (and in the radii of gyration; see Figure 5.7) is due to a surprisingly modest structural change related to the loss of residual stem interactions (see structural models and RMSD plot in Figure 5.7, and contact maps in Figure 5.8).

Putting together very extended unbiased MD, RExMD simulations and ESI-IMS-MS data, I can conclude that the gas phase ensemble for the Q=-3 and Q=-2 states

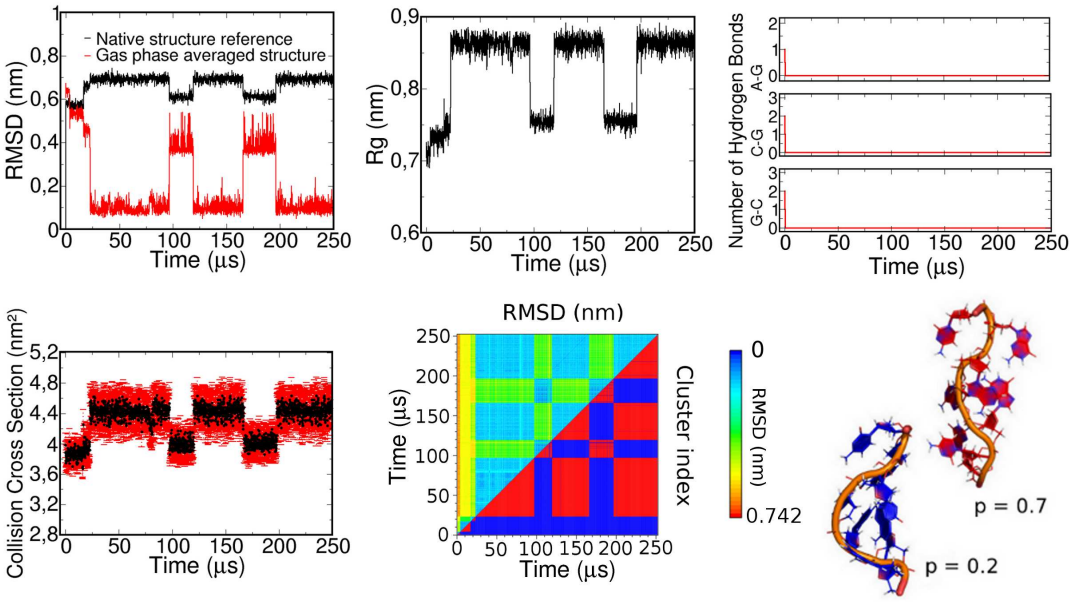

FIGURE 5.7: Variation of structural descriptors (root mean square deviation, radii of gyration, number of hydrogen bonds and collision cross section) for the hairpin at charge state R4 along very extended classical MD simulation (the most stable LC0 charge substate was considered here). Representative structures of the two conformational states (low CCS in blue font and blue nucleotides and high CCS in red font and red nucleotides) are shown with their associated probability (relative to 1,following the same color code).

is dominated by many macroscopically similar, yet microscopically different, compact conformations with little memory of the solution structure.

Such conformations are quite stiff due to the strong stabilization generated by a variety of non-native contacts including extensive nucleobase stacking and hydrogen bonds, especially those involving anionic and neutral phosphate groups. The situation for Q=-4 is more diverse than anticipated, since it appears to be dominated by two major conformational families, one compact characterized by abundant nucleobase contacts in the core of the structure, and another (the most populated one) more extended and with less packing in the core, which reproduce well the major CSS signal observed ESI-IMS-MS experiments.

5.3.4 Proton transfer and topological changes

Classical studies above (as well as all previous MD studies correlated with IMS data) ignore the possible existence of changes in the chemical topology of the DNA in the gas phase on the course of the experiment. If such changes occur, they are likely to involve acidic protons, which might migrate to acceptors. To study this possibility and its impact in the conformational ensemble, we performed a vast CPMD study from a

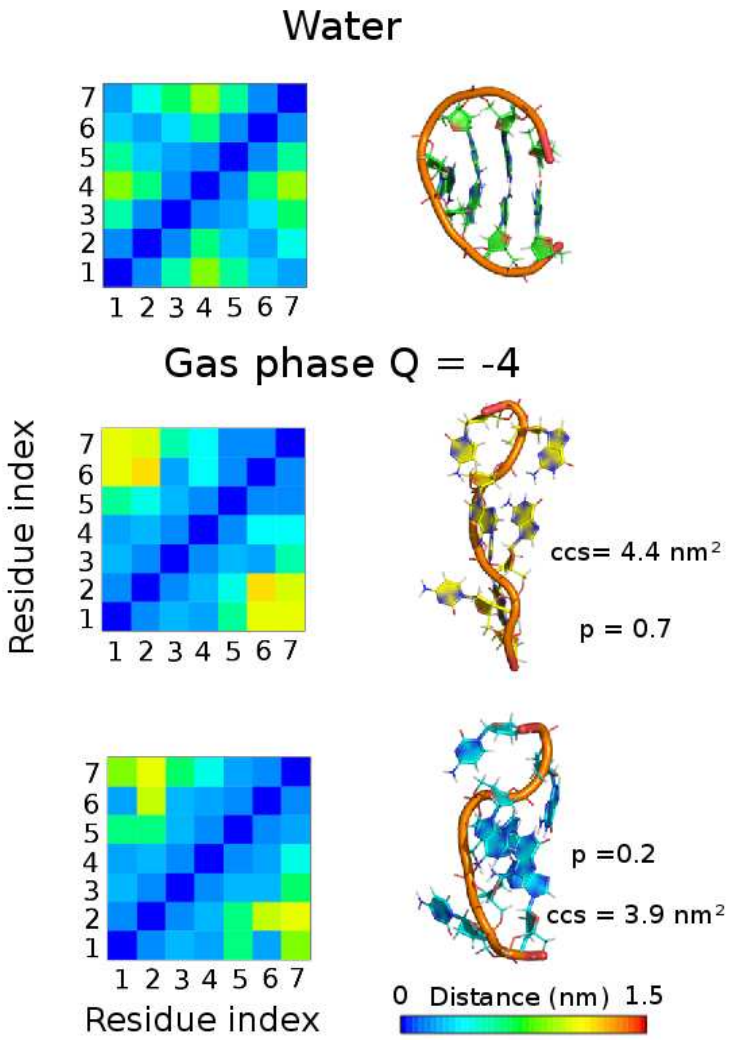

FIGURE 5.8: Contact maps corresponding to representative of the low and high CCS states sampled during the very extended MD simulation of the Q=-4 state (see Figure 5.7)

variety of snapshots collected from the MD simulations for the different charge states and sub-states at IMS effective temperatures. The seed snapshots were selected from distant regions of the trajectories (see Methods 5.2) with the requisite to display at least a couple of hydrogen bonds between acidic hydrogen atoms and potential acceptors.

We then analyzed around 70 individual interactions for an accumulated *ab initio* molecular dynamics simulation time above 0.5 ns (more than 4.2 million of QM-derived snapshots). Details of Car-Parrinello MD are discribed in Appendix B.

The analysis of this huge amount of data shows the existence of six possible scenarios for the different hydrogen bonds: i) no proton transfer, i.e. the proton remains bound to the donor atom for the entire QM trajectory, ii) even though no proton transfer occurs, the original hydrogen bonding scheme is lost and/or a new scheme is formed

along the CPMD simulation, iii) during the simulation time proton transfer occurs, but it is reversible and the proton remains most of the time bound to the donor site, iv) proton transfer occurs many times and the proton is shared for significant periods of time by the donor and acceptor groups (the fingerprint of a canonical low-barrier hydrogen bond), v) full proton transfer occurs, with the proton being most of the time bound to the acceptor group, and vi) as a consequence of donor→acceptor proton transfer there is a conformational change that implies non-negligible changes in the geometry of the donor-acceptor groups.

Contrary to the classical view of macromolecules in the gas phase provided by MD, with a fixed chemical topology, there is a very significant amount of proton transfer occurring spontaneously along the CPMD trajectories (see Figure 5.4 for schematic examples of the process in a few cases, Figure 5.9 for average values). For the most stable charge state (Q=-3) around 15% of the hydrogen bonds analyzed in our study experienced proton transfer processes, the percentage increases to 37% for the Q=-2 state and to 79% in the least stable Q=-4 state.

Proton transfer is extremely fast (sub-picosecond time scale) and for the active hydrogen bonds (those where at least one proton transfer event is detected; see label to Figure 5.9 for definition) many proton transfer events occur during individual CPMD trajectories (in average from 5 for the most stable Q=-3 state, to nearly 12 for the least stable highly charged Q=-4 state). In a few cases we were able to detect up to 40 proton transfer transitions in a 20-25 ps trajectory for a single active hydrogen bond (see Table 5.9), and in average for all the active hydrogen bonds 4-9% of time the proton can be considered "shared" by donor and acceptor, a fingerprint of low barrier hydrogen bonds.

Most of the proton transfers detected here are reversible (see Table 5.9 and Figure 5.8) and full transfer (i.e. transfers where the proton remains tightly bound to the acceptor) happens only in 6% (Q=-2) to 14% (Q=-4) of the "active hydrogen bonds".

In most, but not all cases (see Table 5.9 and Figure 5.8) proton transfer has little impact on the conformation, since the donor-acceptor distances remain essentially unchanged after the transfer. Contact promiscuity (i.e. hydrogen bond contacts lost or created during the CPMD trajectories) is not detected for the most stable charge state (Q=-3), but is not negligible (17/17% (Q=-2) or 4/7% Q=-4) for the less experimentally abundant charge states, even changes of hydrogen bonded partners do not lead to overall dramatic conformational changes.

Most of the proton transfer events detected correspond to proton interchange between phosphates (neutral and anionic), a result that is somehow expected considering that phosphate-phosphate contacts are the most prevalent ones, and that acidity of

Observables	Q=-2	Q=-3	Q=-4
% contacts with PT	33 (8/24)	53 (8/15)	79 (22/29)
% full transfer	4 (1/24)	7 (1/15)	24 (7/29)
% broken contacts	21 (5/24)	-	5 (1/22)
% new contacts	17 (4/24)	-	5 (1/22)
% conf. changes	-	-	9 (2/22)
Av. transf./cont.	6.7 (54/8)	6.0 (48/8)	12.0 264/22)
Max. transf./cont.	26	12	40
% time in acceptor	22 (178.7/8)	26 (205/8)	42 (927.6/22)
% time in share zone	9 (69.5/8)	4 (28.4/8)	7 (158/22)

FIGURE 5.9: Summary of the CPMD analysis of selected H-bond interactions and associated proton transfer (PT) for ensembles of structures collected for different charge states. Values reported here correspond to hydrogen bond contacts involving as donor the acidic proton of phosphate. In the vast majority of cases acceptor group is a deprotonated phosphate. The percentage of contacts that are lost leading to conformation changes are those where at the end of the CPMD trajectory donor-acceptor heteroatoms have significantly separated from original seed conformation. For all the contacts where PT happens we computed the number of PT events, showing in the Table the average number of events and the maximum number of events among all contacts. We display the percentage of simulation time that on average protons (in the case of contacts with PT events) stay close to the acceptor and in the share zone between donor and acceptor in the CPMD trajectory. Further details on the quantification of PT events are described in Appendix B material.

donor and acceptor groups is expected to be very similar. No proton transfer from phosphate to the ribose was detected in any of the simulations (even in those trajectories starting from well defined phosphate-ribose hydrogen bonds). However, proton transfers from phosphates to nucleobases, as well as changes of hydrogen bond patterns between nucleobases were observed.

No proton transfers with nucleobases acting as donors were detected. Of the phosphate-nucleobase proton transfer processes analyzed, only in two cases the proton transferred from the phosphate group remained tightly bound to the acceptor N3 group of guanine (Q=-4) or cytosine (Q=-3), respectively. In two other cases, phosphate-nucleobase hydrogen bonds changed to a different acceptor group (guanine C=O to N7 (Q=-2) and cytosine N3 to C=O, Q=-2 in both cases). In addition, we observed a persisting (Q=-3) and cleaved (Q=-2) hydrogen bond to adenine N7 atoms. Proton transfer to nucleobase (transient or permanent in the time frame of the simulation) did not result in major structural rearrangements in any of the studied cases.

5.4 Discussion

The present study with the smallest possible model of a DNA duplex (a 2-bp stem plus a 3-nt loop) provides a picture of unprecedented quality of the conformational space sampled in the gas phase by a short oligonucleotide capable of folding intramolecularly. The small size of the system reduces memory effects in the experiments, allows the extension of the classical trajectories to time scales close to the experimental ones, and allows for the first time, the representation of quantum effects in the study of the structure and dynamics of nucleic acids in the gas phase.

The oligonucleotide studied gives a major IMS-MS signal for charge state -3, followed by a minor one for charge -2 and an even smaller signal for charge -4. Vaporization of the most prevalent Q=-3 state leads to a conformational ensemble very diverse at the microscopic level, with individual structures showing slow inter-conversion kinetics (test unbiased trajectories starting from one structure representative of a cluster tends to sample conformations close to the seed-one in the multi-microsecond time scale), but defining a quite homogeneous ensemble in terms of macroscopic observables. MD simulations suggest a hairpin with a degree of compactness similar to that of the solution structure, in excellent agreement with ESI-MS experiments performed here.

The strength of a myriad of van der Waals contacts and non-canonical hydrogen bond interactions (those between phosphates are the most prevalent ones) explain the rigidity and stability of the individual structures sampled. However, as expected from

its small size, there is little memory of the solution structure, the finger print of contacts defining the helix are mostly lost, and only around 10% of the canonical hydrogen bonds are maintained in RExMD experiments.

The low charged (Q=-2) ion is sampling conformations that are more compact than those sampled at Q=-3, but qualitatively similar, and again without any clear structural signals of the original helix originally present in solution. Clearly the small size of the system precludes the existence of memory effects that affect longer structures, where clear resemblances between solution and gas phase structures were found (Rueda 2003; Rueda 2006; Arcella 2012). Both for the Q=-2, and Q=-3 states, agreement between theoretical and experimental observables is excellent, supporting the reliability of the derived model.

For the highly charged Q=-4 ion, the situation is completely different: we experimentally determined CCS values that are far from theoretical estimates obtained from microsecond timescale simulations. Experiments demonstrate that the oligo is sampling more extended conformations than those predicted by calculations, illustrating the problems of current RExMD protocols (or any other MD biasing scheme) to represent very roughly conformational landscape, were local minima can be separated by very large energy barriers.

To gain insight into the conformational ensemble of the Q=-4 charge state I performed extended MD simulations, approaching to the millisecond regime. I confirmed that this ion was subjected to infrequent transitions between two major states, structurally very similar by visual inspection, but leading to quite distinct CCS signals. The major conformation corresponds to a quite extended family of structures characterized by CCS very similar to those reported experimentally, while the minor agrees with the sampling detected in RExMD simulations.

These results suggest that despite its intrinsic rigidity, the DNA structure can undergo non-negligible conformational transitions, especially when it is heavily charged and therefore displays a strong tendency to minimize Coulomb repulsion. The idea that gas phase structure is a frozen picture of the solution structure needs to be revisited, at least for small systems. It would also be interesting to characterize the influence of system size on memory effects, taking into account the effects of temperature and the time spent by the desolvated ions in the mass spectrometer.

The hypothesis that covalent structure (the chemical topology) remains unaltered upon vaporization needs also to be revisited. It is commonly assumed that during the fast vaporization process the required protonation should occur at solvent accessible basic sites. For oligonucleotides, it seems logical to assume that most of the protonation

should happen at phosphates, but there is no guarantee that protons will remain in the same places during the entire time scale of the experiment.

The lack of water makes distant solvent mediated proton transfer impossible, but direct hydrogen bonds can justify short-range proton transfer. Proton transfer between phosphates could be at the origin of asymmetric charge partitioning upon dissociation of some DNA duplexes (Madsen and Brodbelt 2010). Besides, proton transfer events from phosphates to riboses (Cerny, Gross et al. 1986; Barry, Vouros et al. 1995; Bartlett, McCloskey et al. 1996) or from phosphates to nucleobases (Wang, Wan et al. 1998; Wan and Gross 2001; Wan, Gross et al. 2001) have been proposed as key steps in fragmentation mechanisms of oligonucleotides (reviewed in Wu et al. (Wu and McLuckey 2004)).

A massive CPMD study (to our knowledge, the first of this type in these systems) reveals the surprisingly large amount of proton transfer occurring for all charged states. Even for the most stable Q=-3 state, proton transfer is detected in the 15% of the hydrogen bond analyzed, the figures being significantly increased for the less stable Q=-2 and Q=-4 states. Proton transfer is very fast and surprisingly frequent (on average 8-12 times in a 20 ps window). Proton transfer is in most cases reversible, and the proton returns to the original donor group after a few picoseconds. Most of proton transfers occur between phosphates, but we were able to detect a few cases where nucleobases were protonated, and two cases in which the proton remained stably bound to the acceptor base.

This validates the mechanism initially proposed by Gross and co-workers for DNA (Wang, Wan et al. 1998; Wan and Gross 2001), and later also proposed by McLuckey and co-workers for RNA (Huang, Kharlamova et al. 2008), in which proton transfer from phosphate to nucleobase and thereby zwitterion formation would be the first step in the oligonucleotide fragmentation. The frequencies of these hetero-atomic proton transfers in our simulations is low, but if we detect them in 0.5 ns of CPMD sampling (generated from seeding derived from very extended classical trajectories), they may be not negligible at higher effective temperatures necessary to cause significant fragmentation.

In contrast with a previous study on a trinucleotide where no evidence for zwitterion formation was found based on collision cross sections and short-timescale MD simulations (Gidden and Bowers 2003), our simulations suggest that zwitterion formation is entirely possible for our 7-mer oligonucleotide while migrating along the drift tube. Besides, it is worth to note that most proton transfer are not necessarily resulting in significant conformational changes and CCS changes (in agreement with the fact that classical calculations show little differences between the different charge substates), but even with our limited quantum sampling, we have found a few cases where transfer of

the proton to the acceptor group leads to separation of donor and acceptor groups and accordingly to potential structural changes.

This suggests that for a millisecond-scale process structural rearrangements catalyzed by proton transfer may not be negligible and might contribute to structural variability in the ensemble.

5.5 Conclusions

The present study, which combines very extended classical MD simulations, quantum mechanical CPMD calculations and IMS-MS experiments, reveals a picture of unprecedented quality on the nature of DNA in the gas phase. The classical picture of DNA in the gas phase as a frozen structure with rigid topology must be revisited.

Different dynamic time scales are detected, one in the femto-to-picosecond that involves proton transfer between neighboring groups, another in the nano-to-microsecond time scale that implies movements among close conformations with similar macroscopic descriptors, and finally one very slow, in the micro-to-millisecond time scale that implies conformational transitions leading to change in macroscopic structural observables.

In summary, the conformational life of DNA in the gas phase seems richer than previously anticipated, and a mobile proton model such as now widely accepted to explain peptide fragmentation (Wysocki, Tsaprailis et al. 2000; Boyd and Somogyi 2010) also entirely make sense for oligonucleotides.

Bibliography

Ansari, A., S. V. Kuznetsov, et al. (2001). "Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins." Proc Natl Acad Sci U S A 98(14): 7771-7776.

Arcella, A., Portella, G., Ruiz, M. L., Eritja, R., Vilaseca, M., Gabelica, V., Orozco, M. (2012). "Structure of triplex DNA in the gas phase." J Am Chem Soc 134(15): 6596-6606.

Avi, A., G. Portella, et al. (2014). "Specific loop modifications of the thrombin-binding aptamer trigger the formation of parallel structures." FEBS Journal 281(4): 1085-1099.

Baker, E. and M. Bowers (2007). "B-DNA Helix Stability in a Solvent-Free Environment." Journal of the American Society for Mass Spectrometry 18(7): 1188-1195.

Barry, J. P., P. Vouros, et al. (1995). "Mass and sequence verification of modified oligonucleotides using electrospray tandem mass spectrometry." Journal of Mass Spectrometry 30(7): 993-1006.

Bartlett, M. G., J. A. McCloskey, et al. (1996). "The Effect of Backbone Charge on the Collision-Induced Dissociation of Oligonucleotides." Journal of Mass Spectrometry 31(11): 1277-1283.

Becke, A. D. (1988). "Density-functional exchange-energy approximation with correct asymptotic behavior." Physical Review A 38(6): 3098-3100.

Blackburn, G. M. G., M. J. (1990). Nucleic Acids in Chemistry and Biology. Oxford, IRL Press.

Bloomfield, V. A. C., D. M.; Tinoco, I. (2000). Nucleic Acids. Structure, Properties and Functions. Sausalito, CA, University Science Books.

Boyd, R. and . Somogyi (2010). "The Mobile Proton Hypothesis in Fragmentation of Protonated Peptides: A Perspective." Journal of the American Society for Mass Spectrometry 21(8): 1275-1278.

Breuker, K. and F. W. McLafferty (2008). "Stepwise evolution of protein native structure with electrospray into the gas phase, 10-12 to 102 s." Proceedings of the National Academy of Sciences 105(47): 18145-18152.

Burmistrova, A., V. Gabelica, et al. (2013). "Ion Mobility Spectrometry Reveals Duplex DNA Dissociation Intermediates." Journal of The American Society for Mass Spectrometry 24(11): 1777-1786.

Car, R. and M. Parrinello (1985). "Unified Approach for Molecular Dynamics and Density-Functional Theory." Physical Review Letters 55(22): 2471-2474.

Cerny, R. L., M. L. Gross, et al. (1986). "Fast atom bombardment combined with tandem mass spectrometry for the study of dinucleotides." Analytical Biochemistry 156(2): 424-435.

Cheatham, T. E., III, Cieplak, P., and, Kollman, P.A. (1999). "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat." J. Biomol. Struct. Dyn. 16: 845-862.

Doktycz, M. J., S. Habibi-Goudarzi, et al. (1994). "Accumulation and Storage of Ionized Duplex DNA Molecules in a Quadrupole Ion Trap." Analytical Chemistry 66(20): 3416-3422.

Gabelica, V. and E. De Pauw (2002). "Collision-induced dissociation of 16-mer DNA duplexes with various sequences: evidence for conservation of the double helix conformation in the gas phase." International Journal of Mass Spectrometry 219(1): 151-159.

Gabelica, V., E. De Pauw, et al. (1999). "Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry." Journal of Mass Spectrometry 34(12): 1328-1337.

Gabelica, V. and E. D. Pauw (2001). "Comparison between solution-phase stability and gas-phase kinetic stability of oligodeoxynucleotide duplexes." Journal of Mass Spectrometry 36(4): 397-402.

Ganem, B., Y.-T. Li, et al. (1993). "Detection of oligonucleotide duplex forms by ion-spray mass spectrometry." Tetrahedron Letters 34(9): 1445-1448.

Gidden, J. and M. Bowers (2003). "Gas-phase conformations of deprotonated trinucleotides (dGTT-, dTGT-, and dTTG-): the question of zwitterion formation." Journal of the American Society for Mass Spectrometry 14(2): 161-170.

Gidden, J., A. Ferzoco, et al. (2004). "Duplex Formation and the Onset of Helicity in Poly d(CG)n Oligonucleotides in a Solvent-Free Environment." Journal of the American Chemical Society 126(46): 15132-15140.

Grimme, S. (2004). "Accurate description of van der Waals complexes by density functional theory including empirical corrections." Journal of Computational Chemistry 25(12): 1463-1473.

Hoaglund, C. S., Y. Liu, et al. (1997). "Gas-Phase DNA: Oligothymidine Ion Conformers." Journal of the American Chemical Society 119(38): 9051-9052.

Huang, T.-y., A. Kharlamova, et al. (2008). "Ion Trap Collision-Induced Dissociation of Multiply Deprotonated RNA: c/y-Ions versus (a-B)/w-Ions." Journal of the American Society for Mass Spectrometry 19(12): 1832-1840.

Lee, C. T., W. T. Yang, et al. (1988). "Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density." Physical Review B 37(2): 785-789.

Light-Wahl, K. J., D. L. Springer, et al. (1993). "Observation of a small oligonucleotide duplex by electrospray ionization mass spectrometry." Journal of the American Chemical Society 115(2): 803-804.

Ma, H., Wan, C., Wu, A., Zewail, A. H. (2007). "DNA folding and melting observed in real time redefine the energy landscape." Proc Natl Acad Sci U S A 104(3): 712-716.

Madsen, J. and J. Brodbelt (2010). "Asymmetric charge partitioning upon dissociation of DNA duplexes." Journal of the American Society for Mass Spectrometry 21(7): 1144-1150.

Meyer, T., V. Gabelica, et al. (2013). "Proteins in the gas phase." Wiley Interdisciplinary Reviews: Computational Molecular Science 3(4): 408-425.

Morsa, D., V. Gabelica, et al. (2011). "Effective temperature of ions in traveling wave ion mobility spectrometry." Anal Chem 83(14): 5775-5782. Neutze, R., Huldt, G., Hajdu, J., and van der Spoel, D.; (2004). "Potential impact of an X-ray free electron laser on structural biology." Radiat. Phys. Chem 71: 905 916.

Padrta, P., Stefl, R., Krlk, L., Zdek, L., Sklenr, V. (2002). "Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings." Journal of Biomolecular NMR 24: 1-4.

Pan, S., K. Verhoeven, et al. (2005). "Investigation of the initial fragmentation of oligodeoxynucleotides in a quadrupole ion trap: Charge level-related base loss." Journal of the American Society for Mass Spectrometry 16(11): 1853-1865.

Perez, A., F. J. Luque, et al. (2012). "Frontiers in molecular dynamics simulations of DNA." Acc Chem Res 45(2): 196-205.

Perez, A., I. Marchan, et al. (2007). "Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers." Biophys J 92(11): 3817-3829.

Portella, G., and, Orozco M. (2010). "Multiple Routes to Characterize the Folding of a Small DNA Hairpin." Angew. Chem. Int. Ed. 49: 7673 7676.

Rueda, M., Kalko, S. G., Luque, F. J., Orozco, M. (2003). "The structure and dynamics of DNA in the gas phase." J Am Chem Soc 125(26): 8007-8014.

Rueda, M., F. J. Luque, et al. (2005). "Nature of minor-groove binders-DNA complexes in the gas phase." J Am Chem Soc 127(33): 11690-11698.

Rueda, M., Luque, F. J., Orozco, M. (2006). "G-quadruplexes can maintain their structure in the gas phase." J Am Chem Soc 128(11): 3608-3619.

Saenger, W. (1984). Principles of Nucleic Acid Structure. New York, Springer-Verlag.

Schnier, P. D., J. S. Klassen, et al. (1998). "Activation Energies for Dissociation of Double Strand Oligonucleotide Anions: Evidence for WatsonCrick Base Pairing in Vacuo." Journal of the American Chemical Society 120(37): 9605-9613.

Shvartsburg, A. A., S. V. Mashkevich, et al. (2007). "Optimization of algorithms for ion mobility calculations." The journal of physical chemistry. A 111(10): 2002-2010.

Sinden, R. R. S., CA, (1994). DNA Structure and Function.

Siu, C.K., et al., (2010) "Optimization of Parameters Used in Algorithms of Ion-Mobility Calculation for Conformational Analyses".Journal of Physical Chemistry B, 114(2):p.1204-1212.

Sugita, Y., and, Okamoto, Y. (1999). "Replica-exchange molecular dynamics method for protein folding." Chem. Phys. Lett. 314: 141-151.

Troullier, N. and J. L. Martins (1991). "Efficient pseudopotentials for plane-wave calculations." Physical Review B 43(3): 1993-2006.

Wan, K. and M. Gross (2001). "Fragmentation mechanisms of oligodeoxynucleotides: Effects of replacing phosphates with methylphosphonates and thymines with other bases in T-rich sequences." Journal of the American Society for Mass Spectrometry 12(5): 580-589.

Wan, K. X., J. Gross, et al. (2001). "Fragmentation mechanisms of oligodeoxynucleotides studied by H/D exchange and electrospray ionization tandem mass spectrometry." Journal of the American Society for Mass Spectrometry 12(2): 193-205.

Wang, Z., K. Wan, et al. (1998). "Structure and fragmentation mechanisms of isomeric T-rich oligodeoxynucleotides: A comparison of four tandem mass spectrometric methods." Journal of the American Society for Mass Spectrometry 9(7): 683-691.

Wu, J. and S. A. McLuckey (2004). "Gas-phase fragmentation of oligonucleotide ions." International Journal of Mass Spectrometry 237(23): 197-241.

Wysocki, V. H., G. Tsaprailis, et al. (2000). "Mobile and localized protons: a framework for understanding peptide dissociation." Journal of Mass Spectrometry 35(12): 1399-1406.

Chapter 6

Summary and Conclusions

The highly charged nature of the nucleic acids backbone clearly suggests that the environment plays a key role in the behavior of these molecules (Clausen-Schaumann 2000; Cui 2007). As the DNA is found in aqueous solution under physiological condition, the vast majority of published works on nucleic acids naturally investigates its behavior under these conditions. It is well known that water and counterionsplay a fundamental role in defining the structure and function of biomolecules, e.g. by screening of the phosphate-phosphate repulsion, allowing then a stable double helical structure (Gao YG 1995; Varnai and Zakrzewska 2004). Thus, it is kwnon since the fifties that changes in solvent composition leads to dramatic changes in nucleic-acid structure and stability (Herskovits T.T. 1972; Bonner G. 2000; Ke F. 2010).

Water is excellent to stabilize DNA, but it is not the ideal solvent for favoring specific recognition, certain reactions or physical processes such as charge transfer. Interest exist then to explore the nature of nucleic acids, particularly DNA, in non-aqueous solvents, where the universe of nucleic acids applications will expand even more (Lin 2008) (Mathieu 2005). For example DNA nanostructures with hydrophobic regions on the outside could be inserted into the hydrophobic patch of membrane cell in gene-delivery techniques (Koltover 1998; Ewert 2004; Ewert 2005), and a wide range of reactivity banned in polar solvents will emerge for DNA if it can be transferred to non-aqueous solvents. Knowledge of the dependence between solvent and structure will certainly favor our ability to desire more efficient nucleic acids carriers, and to expand the universe of DNA chemistry.

DNA in apolar environment

In the chapter 3 "Nucleic acids in apolar environment" I present a study on the behavior of nucleic acids in apolar environment. In particular, I used atomistic molecular dynamics simulations to investigate the structural and thermodynamics changes of a DNA hairpin when transferred from an aqueous solution to a low dielectric media, carbon tetrachloride (CTC), under different DNA charge states. I simulated the pulling of a short DNA hairpin from a water compartment through a CTC slap and estimated the free energy related to the transfer of the DNA from water to CTC through atomistic Umbrella Sampling simulation.

Since charge desolvation establishes a large free energy barrier for the transfer of DNA from polar to apolar media, it is very likely that nucleic acids will prefer to capture a proton from the water environment to neutralize the charge on the phosphates. In this work I used atomistic molecular dynamics simulations to investigate the thermodynamics changes of a DNA hairpin when transferred from aqueous solution to carbon tetrachloride, under different DNA charge states. Contrary to previous studies conducted on DNA in organic solvent, I found that the DNA hairpin does not denature in a pure apolar environment and maintains well its native conformation for low-charged states. Steered MD coupled to umbrella sampling simulations allowed us to determine the free energy cost associated to the phase transfer water→CTC, which range from 650 kJ/mol for the fully charged oligo (a small DNA hairpin with two CG steps) to 200 kJ/mol for the netural species. A natural extension of the project will be to change CTC by a real biological membrane, a project that I have already started.

Nucleic acids in the gas phase

The second part of my thesis is centered in the challenge of the most recent experimental techniques, such as Mass Spectrometry (Fenn 1993) and X-Ray Free Electron Laser (XFEL) (Neutze 2004) which use gas phase ions to provide structural information of macromolecules. They are fast and require low sample consumption but the question is to what extent does gas phase structural information reflect the most populated conformation in solution.

A series of both experimental and theoretical studies with proteins (Fenn 1993; Gale 1995; Banks and Whitehouse 1996; Benjamin, Robinson et al. 1998; Gabelica, De Pauw et al. 1999; Hofstadler and Griffey 2001; Heck and Van Den Heuvel 2004; Benesch and Robinson 2006; Koeniger, Merenbloom et al. 2006; Hopper and Oldham 2009; Hu and Easley 2011; Wyttenbach 2013) have demonstrated that if vaporization

is done under mild conditions, the structure of the protein becomes stable during long periods of time, probably similar to the time-scale of the experiment, and gas phase ensembles can be used to accurately model the solution structure (Parkinson, Lee et al. 2002; Meyer, de la Cruz et al. 2009; Meyer 2012).

The question is then, whether or not these findings also stand for a highly flexible and charged non-globular molecule as DNA, whose conformation is known to be very dependent on the solvent environment. An analysis of the energetics of DNA suggest that in the absence of solvent screening DNA should unfolds, but experimental studies clearly points in the opposite. In the chapter 4 "The structure of Triplex DNA in the gas phase" (Arcella 2012).

Using massive MD simulations we characterized the conformational ensemble of DNA triplexes in the gas phase, validating the results by means of state of the art mass spectroscopy experiments. Our results suggest that the ensemble of DNA triplex structures in the gas phase is well-defined over the experimental time scale, with the three strands tightly bound, and for the most abundant charge states it samples conformations only slightly more compact than the solution structure.The degree of structural alteration in the triplex is however very significant, mimicking that found in duplex and much larger than that suggested for G-quadruplexes (Rueda 2003; Rueda, Luque et al. 2005).

The data strongly supports that the gas phase triplex maintains memory of the solution structure, well-preserved helicity, and a significant number of native contacts.

In Chapter 5 we considered a very small model of a DNA hairpin containing 2 duplex steps. We characterize the mass spectra of molecular structure of DNA hairpin through mass spectroscopy experiment and performed very extensive MD simulations (both biased and unbiased). In collaboration with Paolo Carloni group in Jülich, Car-Parrinello Molecular Dynamics computations on representative MD snapshots were performed to investigate on proton-transfer energy, crucial to understand structural distortions of nucleic acids in vacuum conditions (Arcella A. 2014).

The study, which combines very extended classical MD simulations, quantum mechanical CPMD calculations and IMS-MS experiments, reveals a picture of unprecedented quality on the nature of DNA in the gas phase. The classical picture of DNA in the gas phase as a frozen structure with rigid topology must be revisited. Different dynamics time scales are detected, one in the femto-to-picosecond that involves proton transfer between neighboring groups, another in the nano-to-microsecond time scale that implies movements among close conformations with similar macroscopic descriptors, and

finally one very slow, in the micro-to-millisecond time scale that implies conformational transitions leading to change in macroscopic structural observables.

In summary, the conformational life of DNA in the gas phase seems richer than previously anticipated, and a mobile proton model such as now widely accepted to explain peptide fragmentation (Wysocki, Tsaprailis et al. 2000; Boyd and Somogyi 2010) also entirely make sense for oligonucleotides.

Bibliography

Banks, J. F., Jr. and C. M. Whitehouse (1996). "Electrospray ionization mass spectrometry." Methods Enzymol 270: 486-519.

Benesch, J. L. and C. V. Robinson (2006). "Mass spectrometry of macromolecular assemblies: preservation and dissociation." Curr Opin Struct Biol 16(2): 245-251.

Benjamin, D. R., C. V. Robinson, et al. (1998). "Mass spectrometry of ribosomes and ribosomal subunits." Proc Natl Acad Sci U S A 95(13): 7391-7395.

Bonner G., K. A. M. (2000). "Structural stability of DNA in nonaqueous solvents." Biotechnol Bioeng 68(3): 339-344.

Boyd, R. and . Somogyi (2010). "The Mobile Proton Hypothesis in Fragmentation of Protonated Peptides: A Perspective." Journal of the American Society for Mass Spectrometry 21(8): 1275-1278.

Brown, M. D., Schatzlein, A.G., Uchegbu I.F. (2001). "Gene delivery with synthetic (non viral) carriers." International Journal of Pharmaceutics. 229: 121.

Clausen-Schaumann, H., Rief, M., Tolksdorf, C., Gaub, H. E. (2000). "Mechanical stability of single DNA molecules." Biophys J 78(4): 1997-2007.

Cui, S., Yu, J., Khner, F., Schulten, K., Gaub, H.E.; (2007). "Double-Stranded DNA Dissociates into Single Strands when dragged into a Poor Solvent." J. Am. Chem. Soc. 129: 14710-14716.

Ewert, K., Ahmad, A., Evans, H.M., Safinya, C.R. (2005). "Cationic lipid-DNA complexes for non-viral gene therapy: relating supramolecular structures to cellular pathways." Expert Opinion on Biological Therapy 5: 3353.

Ewert, K., Slack, N.L., Ahmad, A., Evans, H.M., Lin, A.J.. (2004). "Cationic lipid-DNA complexes for gene therapy: Understanding the relationship between complex structure and gene delivery pathways at the molecular level." Current Medicinal Chemistry 11: 133149.

Fenn, J. B. (1993). "Ion formation from charged droplets: roles of geometry, energy, and time." J Am Soc Mass Spectrom 4,: 524535.

Gabelica, V., E. De Pauw, et al. (1999). "Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry." J Mass Spectrom 34(12): 1328-1337.

Gabelica, V., F. Rosu, et al. (2005). "Fast gas-phase hydrogen/deuterium exchange observed for a DNA G-quadruplex." Rapid Commun Mass Spectrom 19(2): 201-208.

Gale, D. C. S., R. D. . (1995). "Characterization of noncovalent complexes formed between minor groove binding molecules and duplex DNA by electrospray ionization-mass spectrometry." J. Am. Soc. Mass. Spectrom 6(12): 1154.

Gao YG, R. H., van Boom JH, Wang AH (1995). "Influence of counter-ions on the crystal structures of DNA decamers: binding of [Co(NH3)6]3+ and Ba2+ to A-DNA." Biophys J 69(2): 559-568.

Heck, A. J. and R. H. Van Den Heuvel (2004). "Investigation of intact protein complexes by mass spectrometry." Mass Spectrom Rev 23(5): 368-389. Herskovits T.T., H. J. P. (1972). "Solution studies of the nucleic acid bases and related model compounds. Solubility in aqueous alcohol and glycol solutions." Biochemistry 11 (25): 4800-4811.

Hofstadler, S. A. and R. H. Griffey (2001). "Analysis of noncovalent complexes of DNA and RNA by mass spectrometry." Chem Rev 101(2): 377-390. Hopper, J. T. and N. J. Oldham (2009). "Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability." J Am Soc Mass Spectrom 20(10): 1851-1858.

Hu, J. and C. J. Easley (2011). "A simple and rapid approach for measurement of dissociation constants of DNA aptamers against proteins and small molecules via automated microchip electrophoresis." Analyst 136(17): 3461-3468.

Huang, T.-y., A. Kharlamova, et al. (2008). "Ion Trap Collision-Induced Dissociation of Multiply Deprotonated RNA: c/y-Ions versus (a-B)/w-Ions." Journal of the American Society for Mass Spectrometry 19(12): 1832-1840.

Ke F., L. Y. K., Hadjiargyrou M., Liang. D. (2010). "Characterizing DNA condensation and conformational changes in organic solvents." PLoS One. 11(5(10)): e13308.

Koeniger, S. L., S. I. Merenbloom, et al. (2006). "Evidence for many resolvable structures within conformation types of electrosprayed ubiquitin ions." J Phys Chem B 110(13): 7017-7021.

Koltover, I., Salditt, T., Radler, J.O., Safinya, C.R. (1998). "An inverted hexagonal phase of cationic liposome-DNA complexes related to DNA release and delivery." Science 281: 7881.

Lin, J., Seeman, N.C., Vaidehi, N. (2008). "Molecular-dynamics simulations of insertion of chemically modified DNA nanostructures into a water-chloroform interface." Biophys J. 95(3): 1099-1107.

Mathieu, F. S., Liao, C., Mao, J., Kopatsch, T., Wang, and N. C. Seeman. (2005). "Six-helix bundles designed from DNA." Nano Lett. 5(661665).

Meyer, T., X. de la Cruz, et al. (2009). "An atomistic view to the gas phase proteome." Structure 17(1): 88-95.

Meyer, T. G., V.; Grubmller,H.; Orozco,M.; (2012). "Proteins in the gas phase." WIRES Computational Molecular Science.

Neutze, R., Huldt, G., Hajdu, J., and van der Spoel, D.; (2004). "Potential impact of an X-ray free electron laser on structural biology." Radiat. Phys. Chem 71: 905 916.

Parkinson, G. N., M. P. Lee, et al. (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA." Nature 417(6891): 876-880.

Rueda, M., F. J. Luque, et al. (2005). "Nature of minor-groove binders-DNA complexes in the gas phase." J Am Chem Soc 127(33): 11690-11698.

Rueda, M., Luque, F. J., Orozco, M. (2006). "G-quadruplexes can maintain their structure in the gas phase." J Am Chem Soc 128(11): 3608-3619.

Rueda, M. K., S. G.; Luque, F. J.; Orozco, M (2003). "The structure and dynamics of DNA in gas phase." J. Am. Chem.Soc 125: 80078014.

Sugita, Y., and, Okamoto, Y. (1999). "Replica-exchange molecular dynamics method for protein folding." Chem. Phys. Lett. 314: 141-151.

Vairamani, M. and M. L. Gross (2003). "G-quadruplex formation of thrombin-binding aptamer detected by electrospray ionization mass spectrometry." J Am Chem Soc 125(1): 42-43.

Varnai, P. and K. Zakrzewska (2004). "DNA and its counterions: a molecular dynamics study." Nucleic Acids Res 32(14): 4269-4280.

Wan, K. and M. Gross (2001). "Fragmentation mechanisms of oligodeoxynucleotides: Effects of replacing phosphates with methylphosphonates and thymines with other bases in T-rich sequences." Journal of the American Society for Mass Spectrometry 12(5): 580-589.

Wan, K. X., J. Gross, et al. (2001). "Fragmentation mechanisms of oligodeoxynucleotides studied by H/D exchange and electrospray ionization tandem mass spectrometry." Journal of the American Society for Mass Spectrometry 12(2): 193-205.

Wang, Z., K. Wan, et al. (1998). "Structure and fragmentation mechanisms of isomeric T-rich oligodeoxynucleotides: A comparison of four tandem mass spectrometric methods." Journal of the American Society for Mass Spectrometry 9(7): 683-691.

Wysocki, V. H., G. Tsaprailis, et al. (2000). "Mobile and localized protons: a framework for understanding peptide dissociation." Journal of Mass Spectrometry 35(12): 1399-1406.

Wyttenbach, T., Bleiholder, C.; Bowers, M. T. (2013). "Factors Contributing to the Collision Cross Section of Polyatomic Ions in the Kilodalton to Gigadalton Range: Application to Ion Mobility Measurements." Anal. Chem. 85: 2191-2199.

Chapter 7

Resumen

Estudio computacional del ADN en entorno no canónico

La naturaleza altamente cargada del esqueleto de los ácidos nucleicos sugiere claramente que el entorno juega un papel clave en el comportamiento de estas moléculas (Clausen - Schaumann 2000 ; Cui 2007). El ADN se encuentra en disolución acuosa en condiciones fisiológicas y, por eso, la gran mayoría de los trabajos publicados sobre ácidos nucleicos investiga, naturalmente, su comportamiento en estas condiciones.

Es bien conocido que tanto el agua como los contraiones juegan un papel fundamental en la definici'on de la estructura y función de las biomoléculas, como por ejemplo, por el apantallamiento de la repulsión fosfato-fosfato, permitiendo entonces una estructura de doble hélice estable (Gao YG 1995; Varnai y Zakrzewska 2004). Por lo tanto, es noto desde los aos cincuenta que los cambios en la composición del disolvente conduce a cambios dramáticos en la estructura de los ácido nucleico y de su estabilidad (Herskovits TT 1972 ;Bonner G. 2000 ;Ke F. 2010).

El agua es excelente para la estabilidad de la estructura del ADN, pero no es el disolvente ideal para favorecer algunos reconocimientos específicos, ciertas reacciones o procesos físicos, tales como la transferencia de carga. Existe un fuerte interés en la exploraci'on de la naturaleza de los ácidos nucleicos, especialmente del ADN, en disolventes no acuosos, donde el universo de las aplicaciones de los ácidos nucleicos se expandirá aún más (Lin 2008)(Mathieu 2005).

Por ejemplo las regiones hidrofóbicas en la parte externa de las nanoestructuras de ADN podrían insertarse en el parche hidrofóbico de una membrana celular en las técnicas de transfección celular (Koltover 1998;Ewert 2004;Ewert 2005), y una amplia gama de reacciones que son energéticamente prohibitivas en disolventes polares se pondrán de

manifiesto para el ADN si se pudiera transferir a disolventes no acuosos. El conocimiento de la dependencia la estructura de los ácidos nucleicos del disolvente, sin duda aumentará nuestra capacidad de diseñar portadores de ácidos nucleicos más eficientes, y ampliando así el universo de la química del ADN.

El ADN en el entorno apolar

En el capítulo 3 describo un estudio sobre el comportamiento de los ácidos nucleicos en el entorno apolar. He usado simulaciones atomísticas de Dinámica Molecular para investigar los cambios estructurales y termodinámicas de una horquilla de ADN cuando se transfiere de una solución acuosa a un medio de constante dieléctrica más baja, tetracloruro de carbono (CTC), en diferentes estados de carga de ADN. He simulado la tracción de una horquilla corta del ADN a partir de un compartimiento de agua a través de un CTC y he calculado la energ'ia libre relacionada con la transferencia del ADN del agua a CTC.

Ya que la desolvatación de carga establece una barrera de energía libre muy grande para la transferencia de ADN de medios polares a medios apolares, es muy probable que los ácidos nucleicos prefieren capturar un protón desde el entorno de agua para neutralizar la carga de los fosfatos. En este trabajo he utilizado simulaciones atomísticas de dinámica molecular para investigar los cambios termodinámicos de una horquilla de ADN cuando se transfieren de una solución acuosa a un medio de menor constante dieléctrica, tal como tetracloruro de carbono (CTC) , en diferentes estados de carga de ADN.

Contrario a estudios anteriores realizados sobre el ADN en un didisolvente orgánico, mis resultados demuestran que una horquilla de ADN no se desnaturaliza en un entorno apolar puro y mantiene así su conformación nativa de estados de baja cargada. La combinación de dinámica molecular juntos con el método de Umbrella Sampling nos permitió determinar el coste de la energía libre asociada a la transferencia de agua a CTC, con valores que van desde 650 kJ/mol para el oligo completamente cargado (una pequeã horquilla de ADN con dos pasos CG) y a 200 kJ/mol para el estado neutralizado. El siguiente paso será el estudio de la transferencia de una orquilla de ADN a través de una membrana biológica real, un proyecto que ya he empezado.

La estructura del ADN de triple hélice en la fase gaseosa

La segunda parte de la tesis se centra en el desafío de las técnicas experimentales más recientes, tales como la espectrometría de masas (Fenn 1993) y X- Ray Láser de Electrones

Libres (XFEL) (Neutze 2004) que utilizan iones en fase gaseosa para proporcionar información estructural de macromoléculas. Éstas técnicas son r'apidas y requieren un pequea muestra, pero la pregunta es hasta qué punto la información estructural en la fase gas refleja la conformaci'on más poblada en disolución

Estudios experimentales y teóricos sobre proteínas (Fenn 1993; Gale 1995; Bancos y Whitehouse 1996; Benjamin,Robinson et al , 1998; Gabelica,De Pauw et al 1999; Hofstadler y Griffey 2001;Heck y Van Den Heuvel 2004; Benesch y Robinson 2006; Koeniger, Merenbloom et al 2006; Hopper y Oldham 2009; Hu y Easley 2011; Wyttenbach 2013) han demostrado que si la vaporización se lleva a cabo en condiciones suaves, la estructura de las proteínas es estable durante largos períodos de tiempo, probablemente similar a la escala de tiempo del experimento, y conjuntos de fase gaseosa se puede utilizar para modelar con precisi'on la estructura en solución (Parkinson, Lee et al 2002; Meyer, de la Cruz y otros,2009; Meyer 2012). La pregunta es entonces, si estos hallazgos valen también para moléculas no globulares, muy flexibles y cargadas como el ADN, la cuya conformación depende mucho del entorno.

Un análisis de la energética del ADN sugieren que en ausencia del apantallamento del disolvente de ADN se despliega, pero estudios experimentales sealan claramente lo contrario. En el capítulo 4 "La estructura de triple hélice de ADN en la fase de gas" (Arcella 2012) mediante el uso masivo de simulaciones de Dinámica Molecular y la validación de los resultados por medio de experimentos de espectroscopia pudimos caracterizar el colectivo conformacional del tr'iplex de ADN en vacío. Nuestros resultados sugieren que el colectivo de las estructuras de triple hélice de ADN en la fase gas está bien definido. En la escala de tiempo experimental, las tres hebras se quedan fuertemente unidos, y para el estado de carga más abundante, las conformaciones se muestran ligeramente m'as compactas que las estructuras en solución.

Sin embargo, el grado de distorsi'on estructural en el triplex es muy significativo, imitando la que se encuentra en d'uplex y mucho más grande que el observado en el G-quadruplex (Rueda 2003; Rueda, Luque et al 2005). Los datos apoyan firmemente que el triplex en fase gas mantiene la memoria de la estructura en solución, la helicidad y un n'umero importante de contactos nativos.

La estructura de un dúplex minimal de ADN en la fase gaseosa

En el capítulo 5 se considera un pequeo modelo de una horquilla de ADN que contiene 2 pasos de dúplex. Se caracteriza los espectros de masas de la molécula obtenidos mediante experimentos de espectroscopía de masas y he realizado simulaciones de Dinámica Molecular. En colaboración con el grupo de Paolo Carloni en Jülich, se realizaron

cálculos cuánticos utilizando el método de Dinámica Molecular de Car-Parrinello para investigar la energía de transferencia de protones que ocurren alo largo del ADN durante el proceso de evaporización, crucial para entender las distorsiones estructurales de los ácidos nucleicos en condiciones de vacío (Arcella A. 2014).

El estudio, que combina simulaciones de dinámica molecular clásicas y cuánticas y de experimentos de espectroscopía de masas, revela una imagen de una calidad sin precedentes la estructura y la naturaleza del ADN en la fase gaseosa. La imagen clásica de ADN en la fase gaseosa como si fuera una estructura congelada con topología rígida debe ser revisada. Se detectan diferentes escalas de tiempos dinámica, uno de femto- a picosegundo, que implica la transferencia de protones entre los grupos vecinos, otra en la escala de tiempo de nano- a microsegundo que implica movimientos entre conformaciones cercanas con descriptores macroscópicas similares, y, finalmente, uno muy lento, en la escala de tiempo del orden del micro- y milisegundos que implica transiciones conformacionales que conducen a cambios en los observables estructurales macrosc'opicos. En resumen, la vida conformacional del ADN en la fase de gas parece más rica de lo que se había previsto anteriormente, y un modelo de protones móviles, que como actualmente es aceptado para explicar la fragmentación de péptidos (Wysocki, Tsaprailis et al 2000; Boyd y Somogyi 2010) tiene enteramente sentido también para los oligonucleótidos.

Bibliography

Banks, J. F., Jr. and C. M. Whitehouse (1996). "Electrospray ionization mass spectrometry." Methods Enzymol 270: 486-519.

Benesch, J. L. and C. V. Robinson (2006). "Mass spectrometry of macromolecular assemblies: preservation and dissociation." Curr Opin Struct Biol 16(2): 245-251.

Benjamin, D. R., C. V. Robinson, et al. (1998). "Mass spectrometry of ribosomes and ribosomal subunits." Proc Natl Acad Sci U S A 95(13): 7391-7395.

Bonner G., K. A. M. (2000). "Structural stability of DNA in nonaqueous solvents." Biotechnol Bioeng 68(3): 339-344.

Boyd, R. and . Somogyi (2010). "The Mobile Proton Hypothesis in Fragmentation of Protonated Peptides: A Perspective." Journal of the American Society for Mass Spectrometry 21(8): 1275-1278.

Brown, M. D., Schatzlein, A.G., Uchegbu I.F. (2001). "Gene delivery with synthetic (non viral) carriers." International Journal of Pharmaceutics. 229: 121.

Clausen-Schaumann, H., Rief, M., Tolksdorf, C., Gaub, H. E. (2000). "Mechanical stability of single DNA molecules." Biophys J 78(4): 1997-2007.

Cui, S., Yu, J., Khner, F., Schulten, K., Gaub, H.E.; (2007). "Double-Stranded DNA Dissociates into Single Strands when dragged into a Poor Solvent." J. Am. Chem. Soc. 129: 14710-14716.

Ewert, K., Ahmad, A., Evans, H.M., Safinya, C.R. (2005). "Cationic lipid-DNA complexes for non-viral gene therapy: relating supramolecular structures to cellular pathways." Expert Opinion on Biological Therapy 5: 3353.

Ewert, K., Slack, N.L., Ahmad, A., Evans, H.M., Lin, A.J.. (2004). "Cationic lipid-DNA complexes for gene therapy: Understanding the relationship between complex structure and gene delivery pathways at the molecular level." Current Medicinal Chemistry 11: 133149.

Fenn, J. B. (1993). "Ion formation from charged droplets: roles of geometry, energy, and time." J Am Soc Mass Spectrom 4,: 524535.

Gabelica, V., E. De Pauw, et al. (1999). "Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry." J Mass Spectrom 34(12): 1328-1337.

Gabelica, V., F. Rosu, et al. (2005). "Fast gas-phase hydrogen/deuterium exchange observed for a DNA G-quadruplex." Rapid Commun Mass Spectrom 19(2): 201-208.

Gale, D. C. S., R. D. . (1995). "Characterization of noncovalent complexes formed between minor groove binding molecules and duplex DNA by electrospray ionization-mass spectrometry." J. Am. Soc. Mass. Spectrom 6(12): 1154.

Gao YG, R. H., van Boom JH, Wang AH (1995). "Influence of counter-ions on the crystal structures of DNA decamers: binding of [Co(NH3)6]3+ and Ba2+ to A-DNA." Biophys J 69(2): 559-568.

Heck, A. J. and R. H. Van Den Heuvel (2004). "Investigation of intact protein complexes by mass spectrometry." Mass Spectrom Rev 23(5): 368-389. Herskovits T.T., H. J. P. (1972). "Solution studies of the nucleic acid bases and related model compounds. Solubility in aqueous alcohol and glycol solutions." Biochemistry 11 (25): 4800-4811.

Hofstadler, S. A. and R. H. Griffey (2001). "Analysis of noncovalent complexes of DNA and RNA by mass spectrometry." Chem Rev 101(2): 377-390. Hopper, J. T. and N. J. Oldham (2009). "Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability." J Am Soc Mass Spectrom 20(10): 1851-1858.

Hu, J. and C. J. Easley (2011). "A simple and rapid approach for measurement of dissociation constants of DNA aptamers against proteins and small molecules via automated microchip electrophoresis." Analyst 136(17): 3461-3468.

Huang, T.-y., A. Kharlamova, et al. (2008). "Ion Trap Collision-Induced Dissociation of Multiply Deprotonated RNA: c/y-Ions versus (a-B)/w-Ions." Journal of the American Society for Mass Spectrometry 19(12): 1832-1840.

Ke F., L. Y. K., Hadjiargyrou M., Liang. D. (2010). "Characterizing DNA condensation and conformational changes in organic solvents." PLoS One. 11(5(10)): e13308.

Koeniger, S. L., S. I. Merenbloom, et al. (2006). "Evidence for many resolvable structures within conformation types of electrosprayed ubiquitin ions." J Phys Chem B 110(13): 7017-7021.

Koltover, I., Salditt, T., Radler, J.O., Safinya, C.R. (1998). "An inverted hexagonal phase of cationic liposome-DNA complexes related to DNA release and delivery." Science 281: 7881.

Lin, J., Seeman, N.C., Vaidehi, N. (2008). "Molecular-dynamics simulations of insertion of chemically modified DNA nanostructures into a water-chloroform interface." Biophys J. 95(3): 1099-1107.

Mathieu, F. S., Liao, C., Mao, J., Kopatsch, T., Wang, and N. C. Seeman. (2005). "Six-helix bundles designed from DNA." Nano Lett. 5(661665).

Meyer, T., X. de la Cruz, et al. (2009). "An atomistic view to the gas phase proteome." Structure 17(1): 88-95.

Meyer, T. G., V.; Grubmller,H.; Orozco,M.; (2012). "Proteins in the gas phase." WIRES Computational Molecular Science.

Neutze, R., Huldt, G., Hajdu, J., and van der Spoel, D.; (2004). "Potential impact of an X-ray free electron laser on structural biology." Radiat. Phys. Chem 71: 905 916.

Parkinson, G. N., M. P. Lee, et al. (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA." Nature 417(6891): 876-880.

Rueda, M., F. J. Luque, et al. (2005). "Nature of minor-groove binders-DNA complexes in the gas phase." J Am Chem Soc 127(33): 11690-11698.

Rueda, M., Luque, F. J., Orozco, M. (2006). "G-quadruplexes can maintain their structure in the gas phase." J Am Chem Soc 128(11): 3608-3619.

Rueda, M. K., S. G.; Luque, F. J.; Orozco, M (2003). "The structure and dynamics of DNA in gas phase." J. Am. Chem.Soc 125: 80078014.

Sugita, Y., and, Okamoto, Y. (1999). "Replica-exchange molecular dynamics method for protein folding." Chem. Phys. Lett. 314: 141-151.

Vairamani, M. and M. L. Gross (2003). "G-quadruplex formation of thrombin-binding aptamer detected by electrospray ionization mass spectrometry." J Am Chem Soc 125(1): 42-43.

Varnai, P. and K. Zakrzewska (2004). "DNA and its counterions: a molecular dynamics study." Nucleic Acids Res 32(14): 4269-4280.

Wan, K. and M. Gross (2001). "Fragmentation mechanisms of oligodeoxynucleotides: Effects of replacing phosphates with methylphosphonates and thymines with other bases in T-rich sequences." Journal of the American Society for Mass Spectrometry 12(5): 580-589.

Wan, K. X., J. Gross, et al. (2001). "Fragmentation mechanisms of oligodeoxynucleotides studied by H/D exchange and electrospray ionization tandem mass spectrometry." Journal of the American Society for Mass Spectrometry 12(2): 193-205.

Wang, Z., K. Wan, et al. (1998). "Structure and fragmentation mechanisms of isomeric T-rich oligodeoxynucleotides: A comparison of four tandem mass spectrometric methods." Journal of the American Society for Mass Spectrometry 9(7): 683-691.

Wysocki, V. H., G. Tsaprailis, et al. (2000). "Mobile and localized protons: a framework for understanding peptide dissociation." Journal of Mass Spectrometry 35(12): 1399-1406.

Wyttenbach, T., Bleiholder, C.; Bowers, M. T. (2013). "Factors Contributing to the Collision Cross Section of Polyatomic Ions in the Kilodalton to Gigadalton Range: Application to Ion Mobility Measurements." Anal. Chem. 85: 2191-2199.

Appendix A

Description of Ion-Moblity Mass Spectrometry

A.1 Details of the experiment performed in Chapter 4

A.1.1 Sample Preparation for Experimental Validation

MS experiments cannot determine alone the structure of the triplex in the gas phase. However, low-resolution structural data derived from IMS-MS experiments can help us validate our theoretical models. The triplexes were prepared in neutral or acidic conditions. Single stranded d(GA)6, d(GA)9, d(TC)6, and d(TC)9 were bought from either Sigma-Aldrich or Eurogentec. For the neutral conditions, the strands were mixed in suitable stoichiometry and resuspended in 150 mM ammonium acetate (pH = 5.5) to a final 2 μsM concentration. Samples were heated at 90 degrees and annealed overnight to form 12- and 18-mer d(TC)d(GA)d(TC) triplexes, which were easily detected in 20% polyacrylamide gels. Melting curves revealed clear triplex signatures with melting temperatures around 50 (12-mer) and 65 (18-mer) degrees in UV-melting experiments (data not shown). Triplex samples were lyophilized before use and resuspended for MS analysis in water and mixed 1:1 (v/v) with solution A to a final concentration of 150 μM (solution A: 100 mM NH4OAc neutralized with triethylamine (pH 7.11)/isopropanol (9:1)) immediately prior to the IMS-MS analysis. Neutral pH and 5% isopropanol favored ionization in the negative mode and significantly improved signal-to-noise ratio with chip-based nanoelectrospray in the negative mode (Turner, Hagan et al. 2006). For acidic conditions, strands were dissolved in water, and triplexes were prepared by annealing in solution B to a final concentration of 100 μM (solution B: 150 mM NH4OAc

supplemented with acetic acid to pH = 3). Samples under neutral conditions were injected at a concentration of 150 μM, while those at acidic pH (where triplex is more stable) were diluted before injection to a concentration of 10 μM (12-mer) or 15 μM (18-mer).

A.1.2 Ion-Mobility Mass Spectrometry

Two sets of experiments were performed in Barcelona and Liege, with the triplexes in neutral (Barcelona) and acidic (pH = 3; Liege) conditions. Both sets of experiments were performed in the negative ion mode on Synapt G1 HDMS traveling wave ion mobility mass spectrometers (Waters, Manchester, U. K.). The main difference is the ionization source used. Experiments on neutral triplexes were performed using a 384-well plate refrigerated at 15 C and introduced by automated chip-based nanoelectrospray using a Triversa NanoMate (Advion BioSciences), whereas experiments at pH = 3 were performed using the standard electrospray source (flow rate of 4 μL/min) at room temperature. Each instrument was tuned so as to obtain a proper ion signal in the softest possible conditions. The details of voltages, pressures, and temperatures used for the ion transfer to the IMS cell are given in table S2 (SI) ref (Arcella A. 2012) . On both instruments, the traveling wave IMS cell was operated at a wave velocity s = 300 m/s and a wave height WH = 8 V. The instruments were mass-calibrated using sodium iodide or cesium iodide and mobility-calibrated using oligonucleotides of known collision cross sections. Different calibrants lists and calibration procedures were used for each instrument (details in Figure **??**, Tables S2 and S3 in (SI) in ref (Arcella A. 2012)), and therefore, the two sets of results are those of totally independent traveling wave IMS experiments.

A.2 Details of the experiment performed in Chapter 5

A.2.1 Ion mobility mass spectrometry experiments

Ion mobility data were recorded on a Synapt G1 HDMS mass spectrometer (Waters, Manchester, UK) with electrospray ionization in the negative mode. The samples were prepared at 10 μM strand concentration, either in pure H2O or in 150 mM aqueous NH4OAc, and infused at 4 μL/min. The source capillary voltage was -2.2 kV, the sampling cone voltage was 20 V, the extraction cone voltage was 4 V, the source backing pressure was 3.10 mbar. All these settings ensured the softest possible conditions in the source region in order to minimize internal energy uptake in that region. The IMS cell was filled with nitrogen at 0.532 mbar, the wave height was 8 V and the wave velocity

was 300 m/s. The IMS was calibrated the same day as data acquisition, using DNA samples of known helium collision cross section. The centroid positions of the arrival time distributions of each calibrant were used to construct the calibration curve, Figure .

The calibration curve was then used to obtain the equation that converts experimental arrival time distributions into experimental collision cross section (CCS) distributions. The values reported for each charge state of the hairpin are the centroid positions of the corresponding distributions. The error on the absolute value of the CCS reported here is obtained from the 95% prediction band of the calibration curve. The full width at half maximum (FWHM) of the CCS distributions are provided separately.

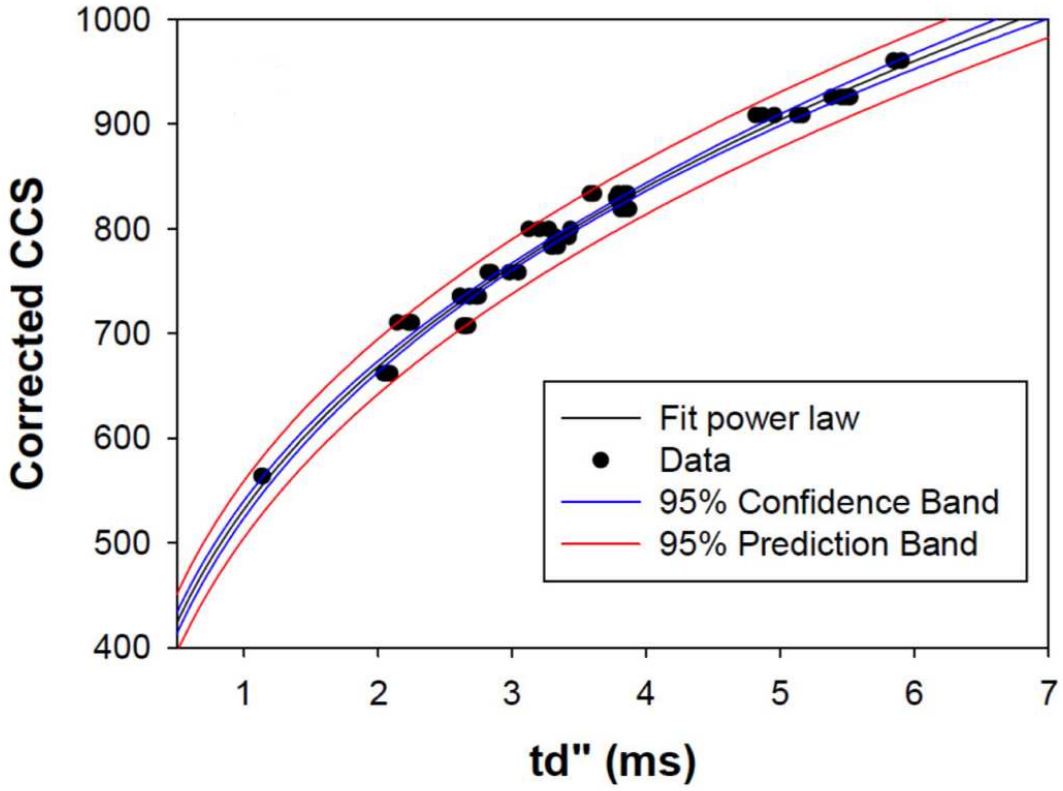

FIGURE A.1: Calibration of the Synapt HDMS arrival times with nucleic acid anions of known collision cross sections. Data collected at bias voltages 10, 15, 20, 25 and 30 V were all merged in a single calibration curve. The calibrants lists are given in the table. The arrival times were corrected for the traveling time outside the IMS and the traveling time of one wave in the IMS and transfer, yielding td values that are plotted against the corrected collision cross sections (corrected CCS), following the protocol of Ruotolo et al (Ruotolo,B.T.; Benesch,J.L.; Sandercock,A.M.; Hyung,S.J.; Robinson,C.V.Nat.Protoc. 2008,3,1139-1152). A power law was used to fit the data with Sigmaplot 11.0. The 95% prediction interval of the power calibration was used to estimate the error on the experimental CCS. This is therefore the CCS interval in which one has 95% chances of finding the true value using the prediction of the calibration curve..

Hairpin	Teff (8V@300m/s)	10V	15V	20V	25V	30V
$[M-2H]^{2-}$ from NH_4OAc solution	358 K	325 ± 7 (FWHM: 23)	**324 ± 7 (FWHM: 20)**	324 ± 7 (FWHM: 21)	325 ± 7 (FWHM: 21)	324 ± 7 (FWHM: 21)
$[M-2H]^{2-}$ from H_2O solution	358 K	325 ± 7 (FWHM: 23)	327 ± 7 (FWHM: 25)	326 ± 7 (FWHM: 22)	326 ± 7 (FWHM: 22)	325 ± 7 (FWHM: 21)
$[M-3H]^{3-}$ from NH_4OAc solution	400 K	375 ± 13 (FWHM: 17)	**375 ± 13 (FWHM: 16)**	374.5 ± 13 (FWHM: 16)	374 ± 13 (FWHM: 16)	373.5 ± 13 (FWHM: 15.5)
$[M-3H]^{3-}$ from H_2O solution	400 K	375 ± 13 (FWHM: 17)	376 ± 13 (FWHM: 17)	375 ± 13 (FWHM: 16)	375 ± 13 (FWHM: 16)	375 ± 13 (FWHM: 16)
$[M-4H]^{4-}$ from NH_4OAc solution	412 K	473 ± 19 (FWHM: 17)	**472 ± 19 (FWHM: 17)**	472 ± 19 (FWHM: 17)	471.5 ± 19 (FWHM: 17)	--

FIGURE A.2: Average experimental CCS (in $Å^2$) for the charge states of the hairpin in the gas phase, obtained using different voltages from samples in ammonium acetate buffer and in water (see experimental section in main text). The corresponding effective temperature in each case is noted in the second column. FWHM is the full width at half maximum of the reconstructed collision cross section distributions. The error mentioned next to the CCS is the 95text Figure 5.2B. The table shows that the CCS values and peak shapes do not change depending on the bias voltage or depending on the starting solution conditions.

Bibliography

Arcella A. Portella G., Ruiz, M.L., Eritja R.,Vilaseca M., Gabelica V., Orozco M.,"The structure of Triplex DNA in the gas phase", J.Am.Chem.Soc., 2012, 134(15), 6596-6606

Turner,K.B.,N.A.Hagan, et al. (2006). Inhibitory effects of archetypicalnucleic acid ligands on the interactions of HIV-1 nucleocapsid protein with elements of Psi-RNA. Nucleic Acids Res 34(5): 1305-1316.

Appendix B

Details of classical and quantum mechanics calculations peformed in Chapter 4

B.1 Details of classical calculations

We carried out all classical MD simulations using the Gromacs-4.5 software (Hess, Kutzner et al. 2008), without periodic boundary conditions and all-to-all calculation for electrostatics and for the short-range repulsive and attractive dispersion interactions, which we modelled via a Lennard-Jones potential. We used the Settle algorithm (Miyamoto and Kollman 1992) to constrain bond lengths and angles of water molecules, and we used P-Lincs (Hess 2008) for all other bond lengths, allowing an integration time step of 1fs. We kept a constant temperature in the simulation box by using the thermostat method of Bussi et al. (Bussi, Donadio et al. 2007). We took the parameters describing the interactions within the system from the amber99sb+parmbsc0 (Hornak, Abel et al. 2006; Perez, Marchan et al. 2007) force field, with parameters for protonated phosphate groups taken from previous works (Rueda 2003; Rueda 2006). Water molecules were represented by the SPC/E potential (Berendsen, Grigera et al. 1987), and we used the Na+ parameters from Smith et al. (Smith and Dang 1994)

We prepared the simulations performed in water by first immersing the hairpin structure 1PQT (Padrta 2002) in a water box of octahedron box containing approximately 3000 water molecules. We used minimum neutralizing conditions, which result in six Na+ ions in solution. The systems were minimized and thermalized for 100 ps before starting the production simulation of 100 ns. We performed 13 sets of RExMD (Sugita 1999) simulations in the gas phase of 2.5 microseconds each, one for each charge

distribution (LC#) and total charge (all the details are described in the manuscript). A total of 13 temperatures were selected to span the gap between 300 K and 863.8 K for Q = -2, and 9 temperatures ranging from 300 K to 621 K for Q=-3 and Q=-4. The exchange frequency between different temperatures was 0.25 on average, and we attempted a temperature exchange every 1000 steps.

We computed root mean squared deviations (RMSD), radius of gyration (Rg) and contact maps using the tools available in the Gromacs-4.5 software suite. We used the exact hard-spheres scattering (EHSS) approximation (Shvartsburg, Mashkevich et al. 2007) to calculate the collision cross section (CCS) from the structures from our simulations. For the CCS calculations, we obtained around 9000 structures for each charge distribution, with a time separation equally distributed along the simulation.

B.2 Car-Parrinello ab initio Molecular Dynamics

B.2.1 The idea of Car-Parrinello ab initio Molecular Dynamics (CPMD) method

The term ab-initio molecular dynamics is used to refer to a class of methods for studying the dynamical motion of atoms, where a huge amount of computational work is spent in solving, as exactly as is required, the entire quantum mechanical electronic structure problem. When the electronic wavefunctions are reliably known, it will be possible to derive the forces on the atomic nuclei using the Hellmann-Feynman theorem (Hellman-Feynman). The forces may then be used to move the atoms, as in standard molecular dynamics.

CarParrinello Molecular Dynamics (CPMD) (Car and Parrinello 1985) is related to the more common Born-Oppenheimer molecular dynamics (BOMD) method in that the quantum mechanical effect of the electrons is included in the calculation of energy and forces for the classical motion of the nuclei, but whereas BOMD treats the electronic structure problem within the time-independent Schrödinger equation, CPMD explicitly includes the electrons as active degrees of freedom, via (fictitious) dynamical variables (Car, Parrinello 1985).

B.2.2 CPMD calculations performed in chapter 5

Full ab initio molecular dynamics simulations were carried out to investigate the existence and structural impact of covalent changes in the structure of the oligo. Calculations

were done using Car-Parrinello Molecular Dynamics (Car and Parrinello 1985) with a dispersion-corrected density functional BLYP-Grimme (Becke 1988; Lee, Yang et al. 1988; Grimme 2004) with norm-conservative pseudopotentials of the MartinsTroullier type (Troullier and Martins 1991) for core electrons along with the Kleinman-Bylander approach for the non-local part. A step of 5 a.u. (0.12 fs) was used for integrating the equations of motion (additional CPMD simulation details are explained in Suppl. Methods). Initially, CPMD calculations were carried out from randomly selected MD snapshots collected at ESI-MS effective temperatures (one for each 4-5 charge sub-states of the three prevalent charge states) and were extended for at least 20 ps. For the most interesting (even less prevalent) charge state Q=-4, we then performed 10 additional (20 ps-long) CPMD simulations taking now snapshots collected from the entire 0.25 ms range and the different charge sub-states (see below). Selected snapshots contain in most cases the prevalent phosphate-phosphate- hydrogen bonds, which are likely to favor phosphate-phosphate proton transfers. To explore other potential proton transfer events, we selected other (less frequent) snapshots with geometries that can favor phosphate-nucleobase proton transfers and analyzed them through additional CPMD simulations. A total of more than 0.5 ns-long ab initio QM/MD trajectories have been collected.

B.2.3 Details of Quantum calculations

We carried out the ab initio MD simulations using the CPMD software package 3.15.1 (CPMD http://www.cpmd.org/, Copyright IBM Corp 1990-2008, Copyright MPI fu r Festk orperforschung Stuttgart 1997-2001). The wavefunction was expanded through a plane-wave basis set up to an energy cutoff of 70 Ry. Periodic boundary conditions were imposed with isolated system conditions using the Martyna-Tuckerman Poisson solver scheme (Martyna 1999).

The geometries of the starting classical MD structures were preliminarily optimized and a subsequent equilibration procedure were performed in order to bring the temperature of the systems at 400 K. Production run simulations were performed in NVT ensemble using a standard Nos-Hoover chain thermostat (Nos 1984; Hoover 1985; Martyna 1992) with a reference temperatura of 400 K.

To determine the number of contact showing proton transfer (PT) we consider that a PT has occurred if the proton jumps from the donor region (1.2$\mathring{A}$ from the donor nucleus) to the acceptor region (1.2$\mathring{A}$ from the acceptor nucleus) and stays bonded to it for at least more than 150 fs (20 times the vibrational period of OH stretching). We consider full transfer has occurred when for the entire CPMD trajectory (typically 20

ps) the proton remains more time bound to the acceptor than to the donor site (as seen not a common situation). Broken contacts are those where the original hydrogen bond contact is lost, in most cases the lost contacts are replaced by new ones.

Bibliography

Berendsen, H. J. C., J. R. Grigera, et al. (1987). "The missing term in effective pair potentials." Journal of Physical Chemistry 91(24): 6269-6271.

Bussi, G., D. Donadio, et al. (2007). "Canonical sampling through velocity rescaling." J Chem Phys 126(1): 014101.

Car, R. and M. Parrinello (1985). "Unified approach for molecular dynamics and density-functional theory." Phys Rev Lett 55(22): 2471-2474.

Hellman-Feynman The so-called theorem of quantum mechanical forces was originally proven by P. Ehrenfest, Z. Phys. 45, 455 (1927), and later discussed by Hellman (1937) and independently rediscovered by Feynman (1939).

Hess, B. (2008). "P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation." Journal of Chemical Theory and Computation 4(1): 116-122.

Hess, B., C. Kutzner, et al. (2008). "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." Journal of Chemical Theory and Computation 4(3): 435-447.

Hoover, W. (1985). "Canonical dynamics: Equilibrium phase-space distributions." Phys. Rev. A 31: 1695-1697

Hornak, V., R. Abel, et al. (2006). "Comparison of multiple Amber force fields and development of improved protein backbone parameters." Proteins 65(3): 712-725.

Martyna, G. J., Tuckerman, M.E. , Klein, M.L. (1992). "Nose-Hoover chains: The canonical ensemble via continuous dynamics. ." J. Chem. Phys. 97 . Martyna, G. J., Tuckerman, M.E. (1999). "A reciprocal-space based method for treating long range forces in force-field based and ab initio calculations of clusters . ." J. Chem. Phys. 110: 2810

Miyamoto, S. and P. A. Kollman (1992). "SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models." Journal of Computational Chemistry 13: 952-962.

Nos, S. (1984). "A unified formulation of the constant temperature molecular dynamics methods." J. Chem. Phys. 81: 511-519.

Padrta, P., Stefl, R., Krlk, L., Zdek, L., Sklenr, V. (2002). "Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings." Journal of Biomolecular NMR 24: 1-4.

Perez, A., I. Marchan, et al. (2007). "Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers." Biophys J 92(11): 3817-3829.

Rueda, M., Kalko, S. G., Luque, F. J., Orozco, M. (2003). "The structure and dynamics of DNA in the gas phase." J Am Chem Soc 125(26): 8007-8014.

Rueda, M., Luque, F. J., Orozco, M. (2006). "G-quadruplexes can maintain their structure in the gas phase." J Am Chem Soc 128(11): 3608-3619.

Shvartsburg, A. A., S. V. Mashkevich, et al. (2007). "Optimization of algorithms for ion mobility calculations." The journal of physical chemistry. A 111(10): 2002-2010.

Smith, D. E. and L. X. Dang (1994). "Computer simulations of NaCl association in polarizable water." Journal of Chemical Physics 100(5): 3757-3766.

Sugita, Y., and, Okamoto, Y. (1999). "Replica-exchange molecular dynamics method for protein folding." Chem. Phys. Lett. 314: 141-151.

Troullier, N. and J. L. Martins (1991). "Efficient pseudopotentials for plane-wave calculations." Physical Review B 43(3): 1993-2006.

Annalisa Arcella

PERSONAL DATA

Name	Annalisa Arcella
Address	C/ Doctor Modrego 14,
	08196 Sant Cugat del Vallès,
	Barcelona-Spain
E-mail address	annalisa.arcella'at'gmail.com
Nationality	Italian

EDUCATION

Sep. 2007-June 2014	Institute for Research in Biomedicine (IRB), Barcelona-Spain. PhD Thesis in Computational Biophysics. Title: Computational study of DNA in non-canonical environment. Supervisor: Prof.Dr. Modesto Orozco.
May 2006-Feb. 2007	International School of Advanced Studies (SISSA), Trieste-Italy. Post-graduate fellowship. Supervisor: Prof.Dr. Christian Micheletti.
March 2006	Università Federico II, Napoli-Italy. M.sc in Physics *with honor*. Master Thesis in Statistical Mechanics, Title: Scaling of the elasticity at the sol-gel transition. Supervisor: Emeritus Prof.Dr. Antonio Coniglio.