Modeling the influence of DNA lesion on the regulation of gene expression

Tom Miclot

Modeling the influence of DNA lesion on the regulation of gene expression

by **Tom Miclot**

PhD thesis publicly defended on December 12th, 2022 in order to obtain the PhD degree in Chemistry, for the *Université de Lorraine*, and the PhD degree in Molecular and Biomolecular Sciences, for the *Università degli studi di Palermo*

Jury members

Rapporteur	Dr. Matteo Lambrughi	Danish Cancer Society Research Center
Rapporteur	Prof. Iñaki Tuñón	Universidad de Valencia
Examiner	A. Prof. Adriana Pietropaolo	Università Magna Graecia di Catanzaro
Examiner	Dr. Antonio Santoro	Università degli Studi di Messina
President of the jury	Prof. Stéphanie Grandemange	Université de Lorraine
Director	Prof. Giampaolo Barone	Università degli Studi di Palermo
Co-Director	Prof. Antonio Monari	Université Paris Cité

Dottorati di Ricerca, UniPa and *Ecole doctorale C2MP*

Whatever your life is, make it useful. To do this, strive to make some progress, however modest it may seem to you, in the work you have chosen, and be certain that you will thus help the general welfare and increase the common heritage.

Gustave Eiffel

Personal translation

Acknowledgements

Child of the sun,
Your destiny is unparalleled,
Adventure is calling you,
Don't wait and run towards it.

Les Mystérieuses Cités d'or

With these few words, I would like to express my thanks to Giampaolo Barone, great baron of tiramissu, and Antonio Monari, the great connoisseur of the Chimay Red®. Thank you for being such good PhD directors. And I especially want to thank this beer they drank together and that gave birth to this PhD thesis project: *The Beerborn PhD*.

Also, I would like to thank all rapporteurs, Iñaki Tuñón, Matteo Lambrughi, for doing me the honor of accepting to examine my work.

Thanks to all my colleagues of both French and Italian laboratories, for their advice and their welcome during my internship and my thesis.

A big thank you to all my family, without whom I would not be here today. So it is important for me to thank you here for always supporting me and believing in me. I also thank my godfather and my godmother who were always present.

I also thank my friend Mathieu, known as Raven, who always has his head in the stars.

Contents

Curriculum vitae

Name	Tom Martin Manfred Miclot
Date of birth	18^{th} August 1994
Place of birth	Strasbourg, France
Nationality	French

Professional networks

Google Scholar *ResearchGate* *ORCID* *LinkedIn* *Scopus* *GitHub*

Diplomas

2020 Master Epistemology, History of Science and Technology
Université de Lorraine, Nancy, France

2019 Master in Biotechnologies
Université de Lorraine, Vandoeuvre-lès-Nancy, France

2017 Bachelor of Science in Life Science
Université de Lorraine, Vandoeuvre-lès-Nancy, France

2015 Advanced Technician's Certificate in Biotechnologies
Lycée Polyvalent Régional Stanislas, Villers-lès-Nancy, France

2013 Baccalaureate in Laboratory Science and Technology
Lycée Louis-Vincent, Metz, France

Laboratory experiences before PhD

2020 Master thesis - 6 month

Title : *Correlation between physical time and computer time.*

Supervisor : Dr. Baptiste Mélès

Place : Archives Henri Poincaré ; Nancy ; France

2019 Master thesis - 6 month

Title : *Cloning, expression, purification, crystallisation and structural resolution of armadillo repeat proteins fused to a target peptide and a chaperone protein for crystallisation.*

Supervisor : PD Dr. Peer Mittl

Place : Laboratory of Prof. Dr. Andreas Plückthun ; Biochemisches Institut ; Universität Zürich ; Switzerland

2018 Internship - 7 weeks

Title : *Study of the interaction between the damaged BRCA2 promoter (abasic site, thymine dimer and 64PP) and NF-kB by molecular dynamics.*

Supervisor : A.Pr Dr. Antonio Monari

Place : Laboratoire de Physique et Chimie Théoriques UMR7019 ; Université de Lorraine ; France

2017 Bachelor thesis - 7 weeks

Title : *Evaluation of the RNA content of pole-laying fish eggs of variable quality, by RT-qPCR.*

Supervisor : Dr. Bérénice Schaerlinger

Place : UR AFPA ; Université de Lorraine ; Nancy ;France

2015 Advanced technician certification thesis - 9 weeks

Title : *Extraction, purification and crystallization of the SCPb protease of* Streptococcus porcinus

Supervisor : Dr Jakki Cooney

Place : Laboratory of Dr Jakki Cooney ; Life Sciences Dept and MSSI ; University of Limerick ; Limerick ; Ireland

2014 Internship - 7 weeks

Title : *Microbiological control of fish aquarium filters.*

Supervisor : Yannick Ledorre

Place : UR AFPA ; Université de Lorraine ; Nancy ; France

Conferences and Schools

2022 European Summerschool in Quantum Chemistry
Università degli Studi di Palermo, Palermo, Italy

2021 Think Innovation week 2021 - Entrepreneurship and Innovation in a Global Environment
Université de Lorraine, Nancy, France

2021 Journées Théorie, Modélisation et Simulation (online)
Institut des Sciences Chimiques de Rennes, Rennes, France

2020 AEBIN Photochemistry School (online)
Donostia International Physics Center, Donostia-San Sebastian, Spain

Scientific Communication

2020 SHARPER - Note europea dei ricercatori
Università degli Studi di Palermo, Palermo, Italy

Awards and distinctions

2022 Premio Ricerc@STEBICEF 2022
Università degli Studi di Palermo, Palermo, Italy
This Call for Papers is aimed at awarding no. 5 prizes of 500.00 euro each. The Call is addressed to young PhD students, post-doctoral fellows, grant holders and contract holders who have distinguished themselves for their scientific merits by publishing, in the year 2021.

Supervision and Mentoring Activities

2022 Co-Supervisor of tow bachelor students in computer science, two months, working on the development of a VMD module facilitating the structural analysis of DNA. The *NucleicAnalysor* module is currently available on GitHub:
`https://github.com/Pitoucha/NucleicAnalysor`

Software development

2022 FindContacts

A Tcl script that adds functionalities to the VMD software. It especially allows to analyze and visualize interaction networks on a single molecule or in a complex.

https://github.com/TMiclot/FindContacts

List of published articles

(co-)First author articles

1. Tom Miclot et al. "G-Quadruplex Recognition by DARPIns through Epitope/Paratope Analogy**". In: *Chemistry A European Journal* (2022). issn: 0947-6539, 1521-3765. doi: 10.1002/chem.202201824.

2. Tom Miclot et al. "Never Cared for What They Do: High Structural Stability of Guanine-Quadruplexes in the Presence of Strand-Break Damage". In: *Molecules* 27.10 (2022), p. 3256. issn: 1420-3049. doi: 10.3390/molecules27103256.

3. Tom Miclot et al. "Structure and Dynamics of RNA Guanine Quadruplexes in SARS-CoV-2 Genome. Original Strategies against Emerging Viruses". In: *The Journal of Physical Chemistry Letters* 12.42 (2021), pp. 1027710283. issn: 1948-7185, 1948-7185. doi: 10.1021/acs.jpclett.1c03071.

4. Tom Miclot et al. "Forever Young: Structural Stability of Telomeric Guanine Quadruplexes in the Presence of Oxidative DNA Lesions**". In: *Chemistry A European Journal* 27.34 (June 2021), pp. 88658874. issn: 0947-6539, 1521-3765. doi: 10.1002/chem.202100993.

5. Cécilia Hognon et al. "Role of RNA Guanine Quadruplexes in Favoring the Dimerization of SARS Unique Domain in Coronaviruses". In: *The Journal of Physical Chemistry Letters* 11.14 (2020), pp. 56615667. issn: 1948-7185, 1948-7185. doi: 10.1021/acs.jpclett.0c01097.

Other articles

1. Emmanuelle Bignon et al. "Specific Recognition of the 5-Untranslated Region of West Nile Virus Genome by Human Innate Immune System". In: *Viruses* 14.6 (2022), p. 1282. issn: 1999-4915. doi: 10.3390/v14061282.

2. Emmanuelle Bignon et al. "Structure of the 5 untranslated region in SARS-CoV-2 genome and its specific recognition by innate immune system *via* the human oligoadenylate synthase 1". In: *Chemical Communications* 58.13 (2022), pp. 21762179. issn: 1359-7345, 1364-548X. doi: 10.1039/D1CC07006A.

3. Jeremy Morere et al. "How Fragile We Are: Influence of Stimulator of Interferon Genes (STING) Variants on Pathogen Recognition and Immune Response Efficiency". In: *Journal of Chemical Information and Modeling* (2022), acs.jcim.2c00315. issn: 1549-9596, 1549-960X. doi: 10.1021/acs.jcim.2c00315.

4. Antonio Francés-Monerris et al. "Microscopic interactions between ivermectin and key human and viral proteins involved in SARS-CoV-2 infection". In: *Physical Chemistry Chemical Physics* 23.40 (2021), pp. 2295722971. issn: 1463-9076, 1463-9084. doi: 10.1039/D1CP02967C.

5. Antonio Francés-Monerris et al. "Molecular Basis of SARS-CoV-2 Infection and Rational Design of Potential Antiviral Agents: Modeling and Simulation Approaches". In: *Journal of Proteome Research* 19.11 (2020), pp. 42914315. issn:1535-3893, 1535-3907. doi: 10.1021/acs.jproteome.0c00779.

6. Cristina García-Iriepa et al. "Thermodynamics of the Interaction between the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus-2 and the Receptor of Human Angiotensin-Converting Enzyme 2. Effects of Possible Ligands". In: *The Journal of Physical Chemistry Letters* 11.21 (2020), pp. 92729281. issn:1948-7185, 1948-7185. doi: 10.1021/acs.jpclett.0c02203.

1

Introduction

The artist not only carries humanity in himself, but he reproduces its history in the creation of his work: first disorder, a general view, aspirations, dazzle, everything is mixed (barbaric era); then analysis, doubt, method, arrangement of the parts (scientific era); finally he comes back to the first synthesis, more extended in the execution.

Flaubert, Carnet 2

The genetic information determines the formation of any living being. Its integrity favors the maintenance of the individual and the conservation of the species. For this reason, genetic alteration leads on a path strewn with light and shadow. On the one hand it is the theater of evolution; the one that has led us from a single-cell life form to the dominant species on the planet. However, the alteration of genetic information can also generate harmful or dangerous effects, with pernicious consequences. There is here the mark of an apparent dualism: genetic alteration manifests itself as the catalyst stimulating the progression and transformation of living organisms, or it presents itself as the fatal herald of their decline. This last point occurs when the genetic alteration is too important, so evolution has endowed every form of life with a means of ensuring the preservation and integrity of the genetic information. If we think of the gateway to life, then the information repair system is identified with Janus, the Roman god with two faces and guardian of the city. As the gatekeeper, the first face looks outward to the environment in which individuals live, and the other face looks inward to cellular life. The repair system strives to repel the harmful effects of genetic alteration and like Janus it closes the door to the evils of life. The Janus analogy takes on a special meaning since the causes of genetic alteration originate from metabolic activity within the cells themselves, such as replication errors and the production of reactive oxygen species [1], or they originate in the external environment, of which radiation such as ultraviolet light is a good example [1]. The universe centered around the mechanism of genetic information is governed by the laws of nature, but ... « What man calls the laws of nature is nothing more than a generalization of a problem that escapes him [1]. » This is the question of the scientist, whose job is to understand and explain these phenomena. It marks the way of elucidation of the knowledge of all the richness around the genetic information, its conservation and its alteration. And this is none other than the history of the nucleic acids; biological molecule that stores the genetic information.

[1]Quote from Dohko, Golden Knight of Libra, in the Saint Seiya manga (1986).

Subject of this thesis

If the bibliography is very rich with contributions explaining the damage of canonical DNA, i.e. in the form of a double helix, it remains poorer as concerns the study of the damage on the non-canonical form of this biological structure. In this respect, this PhD thesis proposes to understand the structural influence of DNA damages on non-canonical G-quadruplex folding. Then, the work seeks to elucidate the mechanisms of interactions between proteins and damaged or non-damaged G-quadruplexes in cellular and viral systems.

CHAPTER 1

Discovery and structures of nucleic acids

Chapter contents

1.1 A brief historical reminder

The discovery of nucleic acids dates back to 1871 with the work of Johannes Friedrich Miescher, a Swiss physician and biologist, which he called « nuclein » [2]. A whole series of works followed to characterize their biological and chemical properties. Moreover the term « nuclein » was replaced under the name of nucleic acid a few years later, 1889, after the proposition of Richard Altmann. Then, the different nucleic acids were identified and their biological role became clearer [3]. In the middle of the 20^{th} century, the work of Rosalind Franklin, including the famous Photo 51, elucidated the double helix structure of DNA. It was only afterwards that James Dewey Watson and Francis Crick published a paper based on her work [4]. The identification of the Watson-Crick-Franklin interactions between nucleotides: AT and GC, followed shortly [5]. Then different interactions were identified in the work conducted by Hoogsteen [6]. The continuity of the studies on nucleic acids have allowed to highlight their biological and structural complexity. Further on, these organic molecules present an important structural polymorphism: helix [4],

triplet [7], hairpin [8], i-motif [9], and G-quadruplex [10].

1.2 Structure of nucleic acids

1.2.1 Nucleic acid bases and basepairing

Nucleic acids are biopolymers, i.e. polymers of organic molecules. More precisely, nucleic acids are a chain of nucleotides covalently linked together by a 3',5'-phosphodiester bond (Figure 1.1). Nucleotides consist of three elements: a phosphate group and a five-carbon sugar: 2'-deoxyribose (for DNA) or ribose (for RNA) and a nitrogenous base. There are five natural nitrogenous bases, which are divided into two categories: those based on a purine base (adenine, A, and guanine, G) and those based on a base (thymine, T, cytosine, C, and uracil, U). But not all of them are present for each type of nucleic acid. Thus, A, T, G and C are found in DNA. But T is not present in RNA, where it is replaced by U[1].

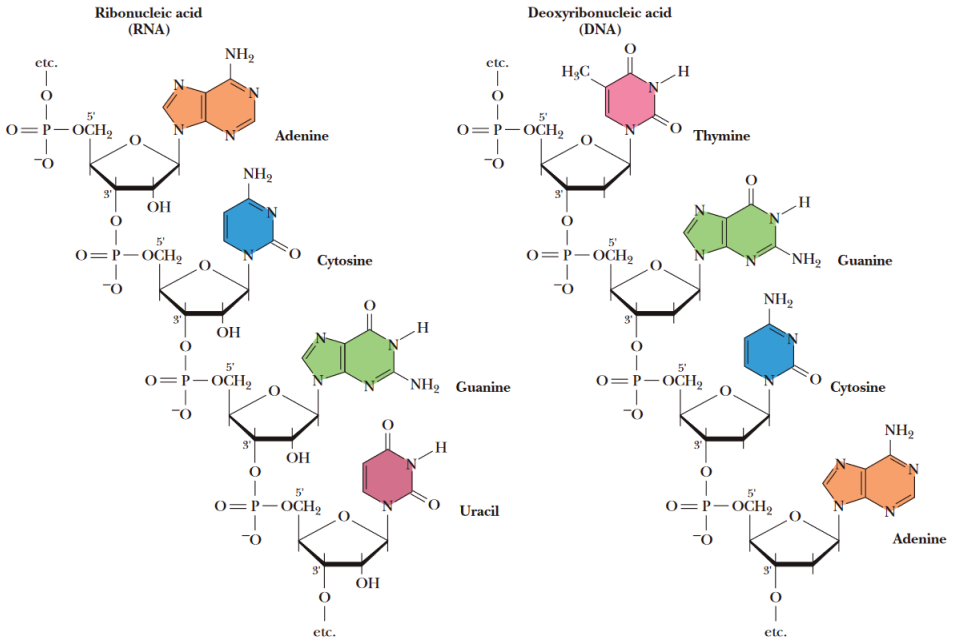


Figure 1.1: The four nucleotides that make up DNA and RNA *Reproduced from Garrett and Grisham [11].*

To fold into secondary structures, such as the double helix of DNA, nucleotides interact with each other through intermolecular hydrogen bonds (Figure 1.1). In the presence of

two strands of nucleic acid, a pyrimidine necessarily pairs with a purine base. Nucleotide make the selective pairs: G-C, which forms three hydrogen bonds, and A-T or A-U, which form only two [12, 1].

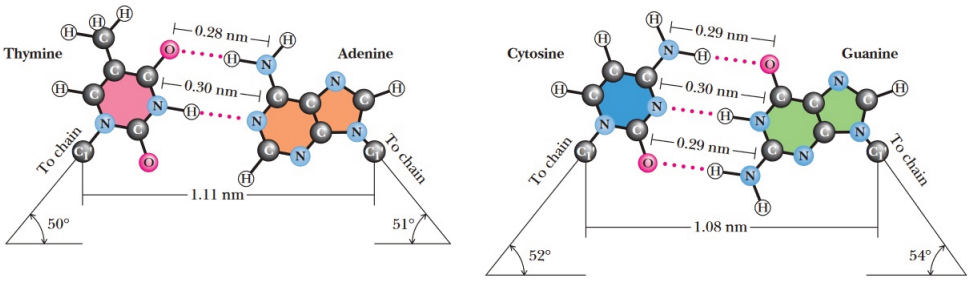


Figure 1.2: Characteristics of Watson-Crick-Franklin basepairing between nucleotides: A-T and G-C *Reproduced from Garrett and Grisham [11].*

1.2.2 Double helix: canonical structure of DNA

The best known nucleic acid structure is the DNA double helix [4]. It consists of two polynucleotide chains wrapped around each other and oriented in opposite directions. The double helix has two grooves: the major groove, which is located between the sugars of the backbones of the two chains, and the minor groove which is located between the paired bases. The presence of deprotonated phosphate groups induces a negative charge on the DNA. The base pairs globally form a hydrophobic core [13]. Also, they form π-stacking interactions that stabilize the the double helix [12].

Although seemingly simple, the double helix conformation of DNA has different variations. The most commonly encountered structure is B-DNA, but two other forms exist: A-DNA and Z-DNA. Their schematic representations are given in Figure 1.3 and their structural characteristics are given in Table 1.1.

Although the double helix structure of DNA is the most known, nucleic acids have a rich collection of secondary structures, as previously mentioned. In addition to the helical form, we may cite the triplet [7], the hairpin [8], the i-motif [9], the and G-quadruplex [10]. This last form is part of the main topic of this thesis and will be described in more detail in the following chapters.

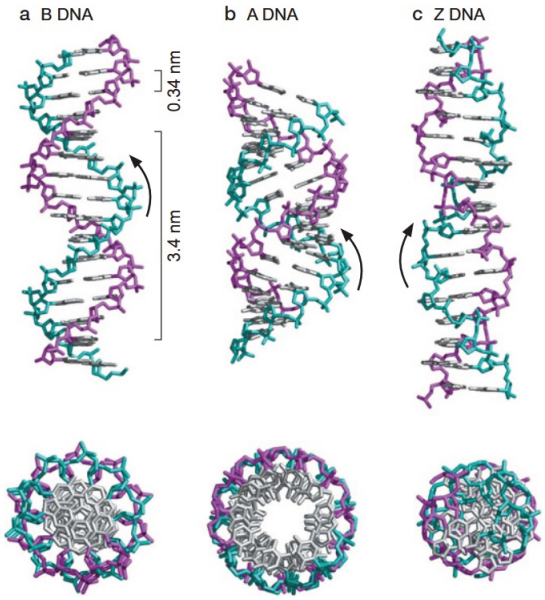


Figure 1.3: Illustration of the different conformations of the DNA double helix. *Reproduced from Watson [13].*

Structural Parameter		**A-DNA**	**B-DNA**	**Z-DNA**
Direction of helix rotation		Right handed	Right handed	Left handed
Residue per helical turn		11	10.5	12
Axial rise per residue		2.55 Å	3.4 V	3.7 Å
Pitch (length) of the helix		28.2 Å	34 Å	44.4 Å
Base pair tilt		20 °	-6 °	7 °
Rotation per residue		32.7 °	34.3 °	-30 °
Diameter of helix		23 Å	20 Å	18 Å
Configuration of glycosidic bond	dA, dT, dC	anti	anti	anti
	dG	anti	anti	syn
Sugar Pucker	dA, dT, dC	C3' endo	C2' endo	C2' endo
	dG	C3' endo	C2' endo	C3' endo

Table 1.1: Structural parameters for the different conformations of the DNA double helix. *Reproduced from Ohyama [14].*

CHAPTER 2

DNA modification: damage and mutation

Chapter contents

2.1 DNA mutation

A mutation is the modification of the genome of a living organism, it corresponds to a modification in the nucleotidic sequence of the DNA (gene mutation), or the arrangement of a chromosome (chromosomal mutation) [15], see Figure 2.1. Finally, mutations are changes that can have a positive, negative, or neutral biological impact on gene expression, transcription and phenotype [16].

2.1.1 Gene mutation

There are three types of mutations in the nucleotide sequence (point mutation) [16], namely:

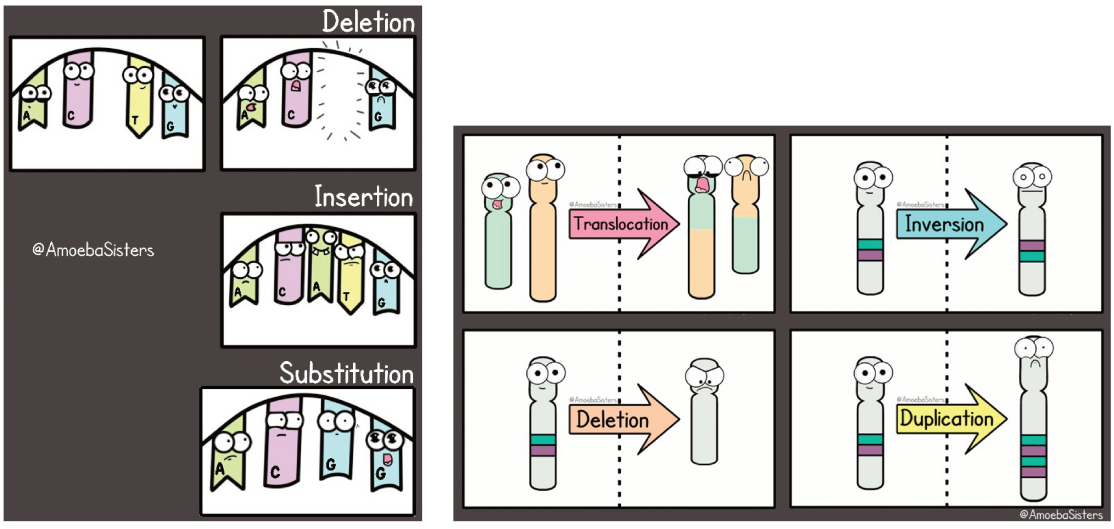
- Base deletion: Deletion of a nitrogenous base in the nucleotide sequence.

- Base insertion: Insertion of a nitrogenous base in the nucleotide sequence.
- Base substitution: One nitrogenous base is replaced by another in the nucleotide sequence. When a purine base is substituted by another purine base, or when a pyrimidine base is substituted by another pyrimidine base, an event called transition. Conversely, transversion is the substitution of a purine base by a pyrimidine base, or the substitution of a pyrimidine base by a purine base.

2.1.2 Chromosomal mutation

A chromosome mutation is a change in the arrangement of the chromosome. There are four types of rearrangements:

- Translocation: a segment of chromosome is moved on another chromosome.
- Deletion: a chromosome segment is lost.
- Inversion: the orientation of a chromosome segment is reversed.
- Duplication: a chromosome segment is doubled.



(a) (b)

Figure 2.1: Schematic representation of the different types of a) gene mutation and b) chromosome mutation. *Reproduced from the work of AmoebaSisters.*

2.2 DNA damage

DNA damage is defined as a change in the chemical structure of the nitrogenous base, the sugar or the phosphodiester bond of a nucleotide. It can be caused by the action of a mutagenic agent (chemical or physical), or be the result of a spontaneous reaction of the nucleotides. Depending on the conditions, damage may be present in a single spot, or it may form clusters of damage in the nucleotide sequence. Finally, damage is classified according to the chemical nature of the reaction product and the process that causes it.

2.2.1 Mutagenic agent induce

Base analogs It's a chemical compounds with a similarity close enough to that of the natural nitrogenous bases can be incorporated at a normal position in the nucleotide sequence. But their pairing properties are not always equivalent to those of nitrogenous bases [16]. The introduction of basic analogues can induce cell death [17, 18].

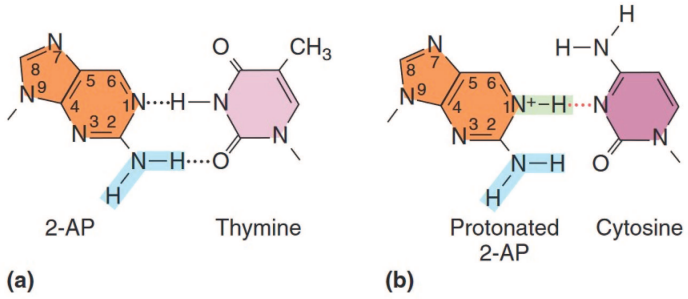


Figure 2.2: Illustration of the pairing of 2-aminopurine (2AP) with thymine and cytosine. *Reproduced from Griffiths et al. [16].*

Alkylating agents Alkylating agents are molecules that do not form permanent complexes with DNA, but modify the chemical structure of nucleotides. They add alkyl groups, sometimes ethyl or methyl, on the nucleotides. The modification causes an alteration of the pairing properties, and leads to mispairing [16, 19]. But the cells have systems that can counteract this type of damage [20, 21].

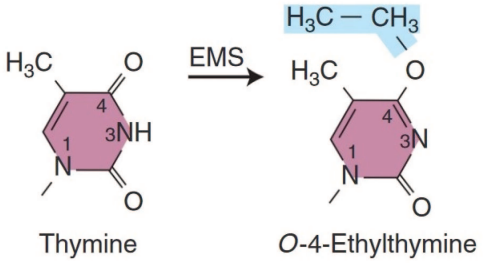


Figure 2.3: Thymine alkylation induced by ethylmethanesulfonate (EMS). *Reproduced from Griffiths et al. [16].*

Intercalating agents They are a class of planar molecules that mimic nitrogenous base pairs. They intercalate between the base pairs of the helical DNA, causing a distortion of its structure. This results in the deletion or insertion of a base pair in the next replication cycle [16, 22].

UV and ionizing radiation These are physical mutagenic agents constituted by photons of different energy. UV light mostly induces the formation of cyclobutane pyrimidine photodimer and the 6-4 photoproduct between two adjacent nucleotide bases. Here we can cite the thymine dimer whose presence correlates with the appearance of skin cancer, because these damages tend to block the replication [23, 24, 25].
Ionizing radiation causes the phosphodiester bond to break on one strand (single strand break) or on both strands (double strand breaks) of DNA [16, 26]. The product of the break in the DNA backbone gives two types of breaks. There are canonical strand break: $5' - PO_4^- / 3' - OH$, and non-canonical strand break: $5' - OH / 3' - PO_4^-$ [27, 28]. Double strand breaks are difficult to repair and hence are particularly toxic, causing high genetic instability.

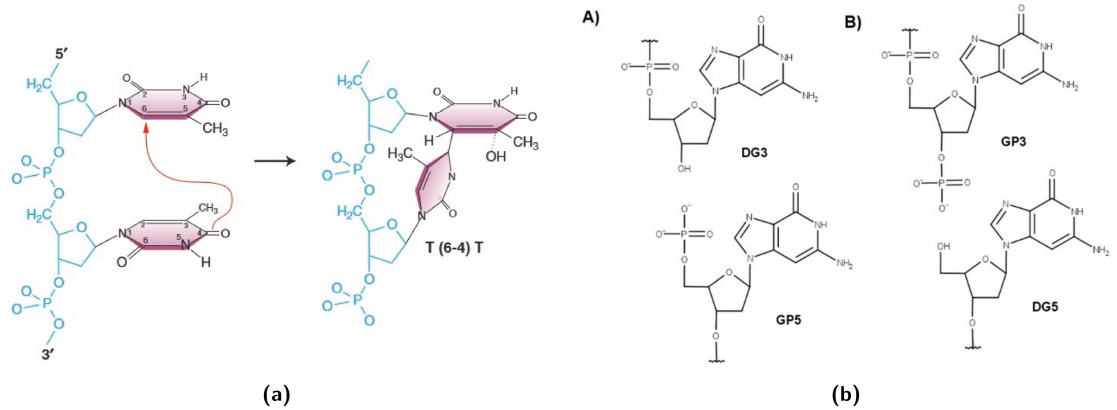


Figure 2.4: Illustration of a) the mechanism of formation of a 6-4 photoproduct, thymine dimer, induced by UV and b) products of ionizing irradiation on the backbone. *Reproduced and modified from Griffiths et al. [16] and Miclot et al. [29]*

2.2.2 DNA damage induced by the spontaneous reaction of nucleotides

Depurinatiation / Depyrimidination They correspond to the breaking of the glycosidic bond between the nitrogenous base and the sugar. If the damage occurs on a purine base (Guanine or Adenine), then it is a depurination. If the damage occurs on a pyrimidine base (thymine or cytosine), then it is a depyrimidation, the latter being

much less frequent. Both types of damage form sites where only the sugar remains in the sequence. [16, 30, 31]

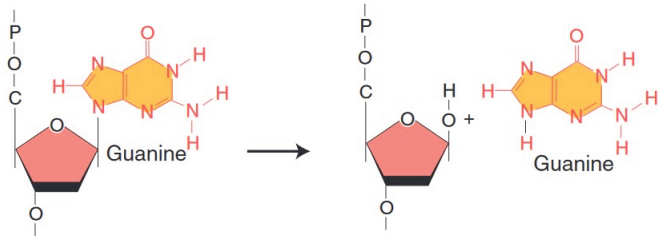


Figure 2.5: The depurination of a guanine induces the formation of an abasic site. *Reproduced from Griffiths et al. [16].*

Base deamination This is damage that results from the removal of an amine group ($-NH_3$) from a nitrogenous base. Consequently, only guanine, adenine and cytosine are sensitive to this type of reaction. For example, deamination converts a cytosine into a uracil, an RNA-specific nucleotide. [16, 32].

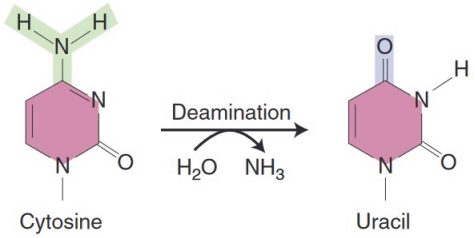


Figure 2.6: The deamination of a cytosine transforms the nucleotide into an uracil. *Reproduced from Griffiths et al. [16].*

Oxidative lesions and Reactive Oxygen Species (ROS) ROS are reactive species mainly including : superoxide radicals ($O_2^-\bullet$), hydrogen peroxide (H_2O_2) and hydroxyl radicals ($\bullet OH$). These molecules are very reactive and modify the chemical structure of nucleotides into various products [16] through oxidative reactions. Of all the nucleotides, guanine is the one with the highest reduction potential. In other words it is the most sensitive nucleotide to ROS, and gives a great diversity of reaction products which may be highly toxic [33, 34, 35].

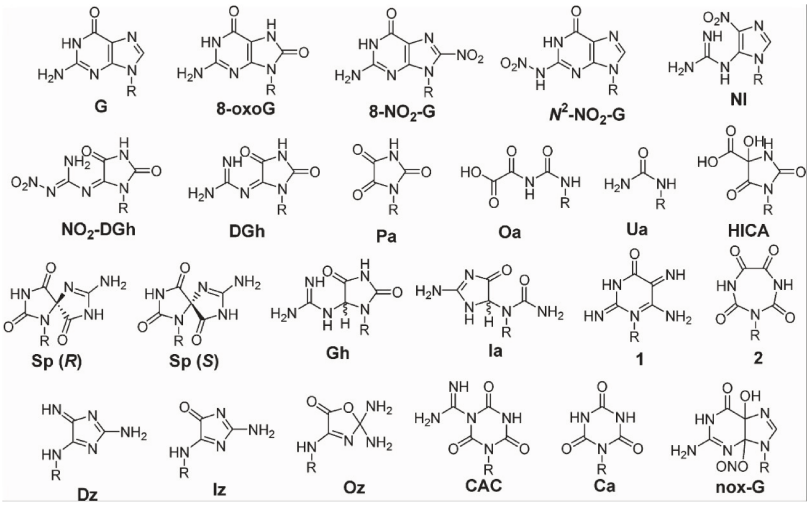


Figure 2.7: Example of some oxidation products of guanine. *Reproduced from Neeley and Essigmann [36].*

CHAPTER 3

DNA repair system

Chapter contents

To assure genetic stability, DNA is the only molecule in the cell that is not simply replaced, but is repaired when damaged. To carry out this operation cells uses several repair mechanisms [16].

3.1 Repair activity of DNA polymerase

DNA polymerases are a class of enzymes responsible for DNA replication. In addition to this function, they have the ability to identify and repair base mismatches. In reality, the replication/repair activity occurs simultaneously [37, 38, 39].

3.2 Base excision-repair (BER)

The BER repair system is one of the most important in the cell. It acts on spontaneous DNA damage: abnormal bases, apuric or apyrimidic sites (AP site), and single-strand

breaks, by involving a whole machinery of enzymes. First, DNA glycosylases, whose role is to recognize the damage and to cleave the glycosidic bond between the nitrogenous base and the sugar leading to an AP site which is recognized and removed from the nucleotide sequence by the AP-endonuclease, by cleavage of the phosphodiester bond. Then, phosphodiesterase reaches the environment of the backbone cut to remove a section of nucleotides. This "gap" is then repaired by DNA polymerase activity exploting sequence complementarity with the opposite strand [40]. BER activity is illustrated in Figure 3.1.

3.3 Nucleic excision-repair (NER)

The BER repair system is very efficient, but it is unable to intervene effectively in the presence of a large amount of damage in the nucleotide sequence or of bulky lesions. Moreover, it is unable to recognize pyrimidine dimers induced by UV light, nor damage that distorts the DNA double helix. Therefore, when one of these situations occurs, another repair system is put in place. This is the NER. In principle, it is very similar to BER, but involves a much larger protein machinery, see Figure 3.1. A deficiency of this repair system causes the appearance of two debilitating autosomal diseases: Cockayne syndrome and *Xeroderma pigmentosum* [40].

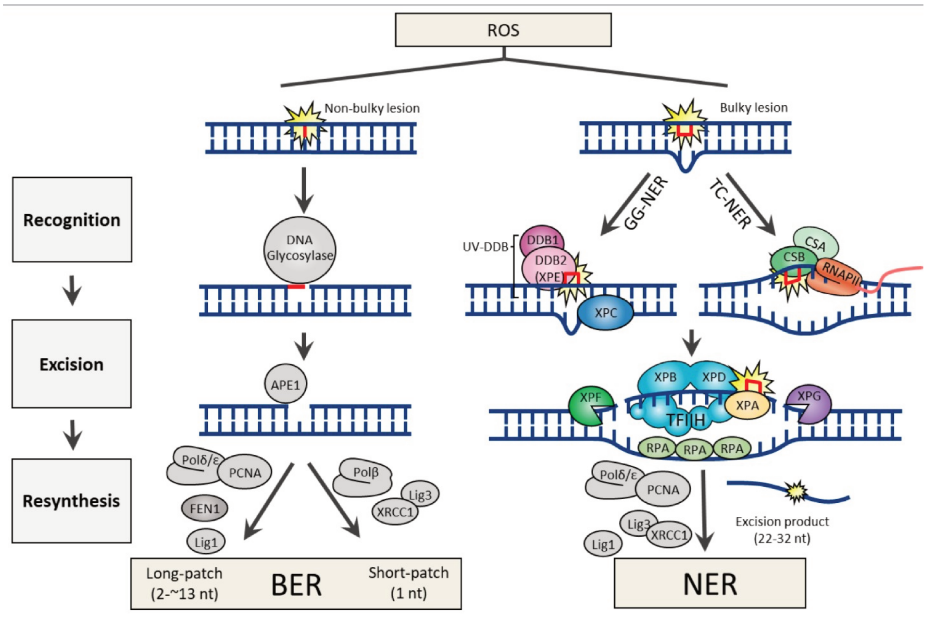


Figure 3.1: Mechanisms of the BER and NER repair systems, with their different protein contributors. *Reproduced from Lee and Kang [41].*

3.4 Double strand break repair

Double strand breaks are very toxic damage frequently leading to cell death. When this type of damage is present in the nucleotide sequence, two major repair systems can be mobilized: homologous recombination (HR) and nonhomologous DNA end joining (HNEJ). In the first case, the repair system degrades the DNA to obtain single-stranded sequences bordering the damaged area. Then it searches a healthy strand for a sequence homologous to the one present on the damaged strand. Then, the healthy complementary strand matches with the damaged strand and the activity of DNA polymerase allows the reconstruction of the nucleotide sequence. In the second repair system (HNEJ), proteins will degrade the DNA to obtain 5'-P and 3'-OH ends, which are recognized and repaired by DNA ligase IV [42, 43].

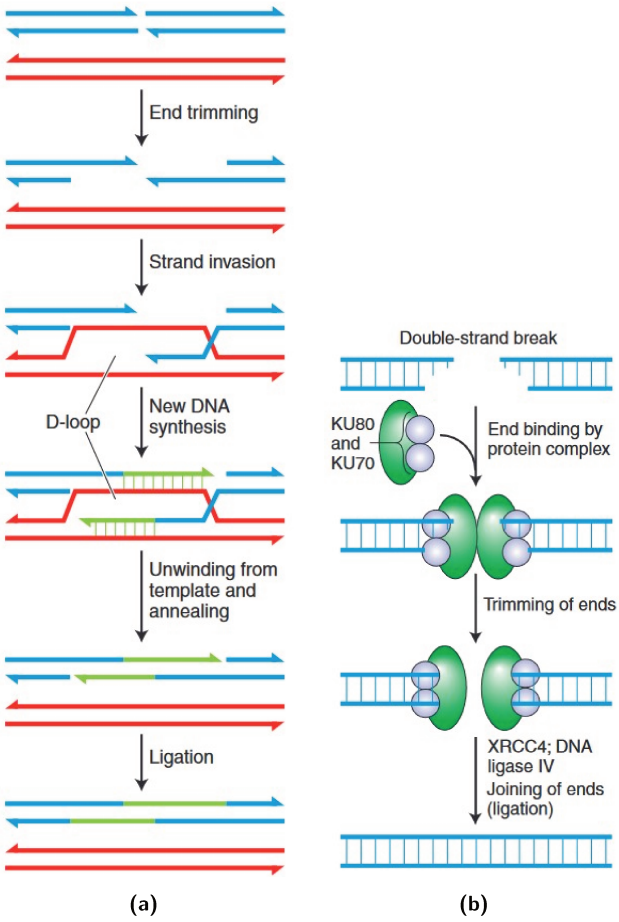


Figure 3.2: Mechanism of double strand break repair a) by homologous recombination and b) by the HNEJ system. *Reproduced from Griffiths et al. [16] and Miclot et al. [29]*

CHAPTER 4

G-quadruplex: non-canonical structure of nucleic acids

Chapter contents

4.1 G-quadruplex and bibliometric impact

Among non-canonic arrangements of nucleic acids G-quadruplex shows a growing interest. Indeed, an analysis of the literature in the *Web of Science* database shows an almost exponential increase in the proportion of publications presenting the keyword « G-quadruplex » , see Figure 4.1; the same is true for the number of citations to articles on this subject. More precisely, G-quadruplex research area produced few publications between the years 1990 and 2000. But, in the period between 2001 and 2006, the number of publications and citations has been steadily increasing. Consequently, starting in 2007 and continuing until now, G-quadruplex represent a research hot topic with a steep growth.

The importance of G-quadruplexes is not only revealed by the increase in the publications that mention them. In fact, a complex network is articulated around this research topic. For example, it is possible to shed light on the universe established around the research on G-quadruplexes, using the VOSviewer software [44]. This software creates bibliometric networks from the analysis of search results conducted in a database. The

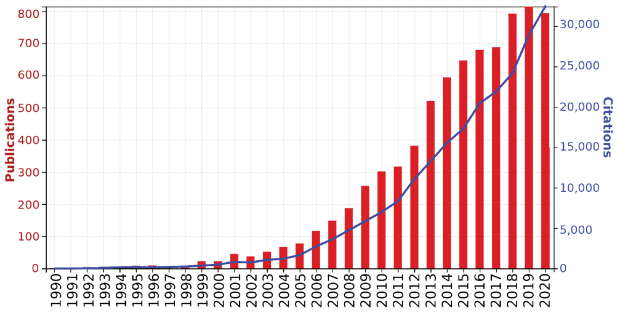


Figure 4.1: The analysis of the bibliography by the *Web of Science* database returns a growing trend in the number of publications and citations involving the keyword « G-quadruplex ». Citation Report graphic is derived from Clarivate *Web of Science*, Copyright Clarivate 2021. All rights reserved.

result of the analysis of the abstract of the articles from the previous research on the *Web of Science* database is represented in Figure 4.2. There is a visual representation of an important network involving G-quadruplex-related thermes. All of them are grouped into four distinct clusters: cell biology, genetics, G-quadruplex structure, and detection and biotechnology applications. Thus, the scientific area interested in the non-canonical G-quadruplex form of nucleic acids is growing and showing a large area of wealth.

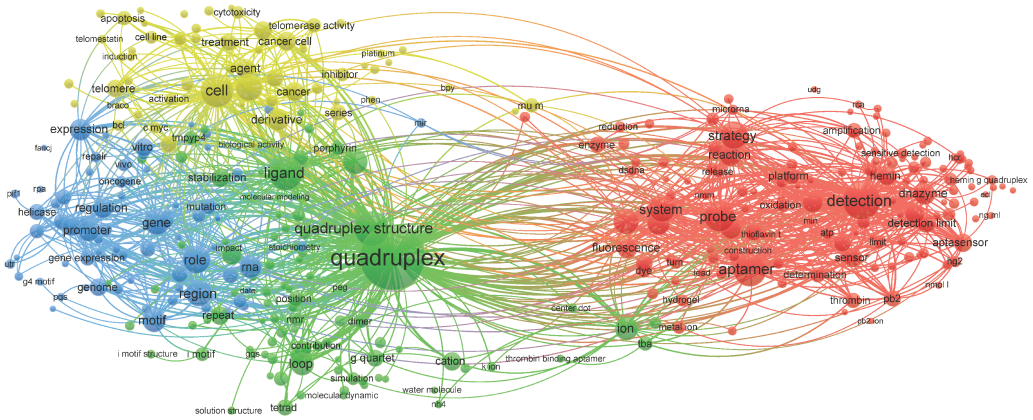


Figure 4.2: The universe built around the G-quadruplex is very rich and composed by four large clusters, represented here by different colors. Yellow : cell biology, Blue : genetic, Green : structural biology, Red : detection & biotechnologies.

4.2 Biological function of G-quadruplexes

To understand the biological importance of G-quadruplexes, it is necessary to recall that not all the nucleotide sequences are involved in the replication and transduction of the genetic information. Indeed, some sequences are involved in the regulation of genetic mechanisms [1]. The regulatory sequences can be presented in a canonical form, the so called double-helical or B-DNA, or in non-canonical forms such as G-quadruplexes. The biological role of the latter is, however,very subtle both considering their biological role

and the interactions with other non-canonical structures. On this point, we can mention the interdependence of the stability of G-quadruplexes and i-motifs: the stabilization of G-quadruplexes leads to the destabilization of i-motifs, conversely the stabilization of i-motifs leads to the destabilization of G-quadurplexes [45].

G-quadruplexes are involved in various cellular processes, such as transcription, recombination, replication and regulation of the cell life cycle [46, 47]. Telomeric G-quadruplexes (h-Telo) are found at the ends of chromosomes, in the telomeres, and their formation is controlled by telomere end-binding proteins [48]. h-Telo G-quadruplex are involved in the regulation of telomerase activity. Overexpression of telomerase is often linked to the immortality of malignant cells. Therefore, the understanding of the structure and dynamics of G-quadruplexes is fundamental to develop appropriate therapeutic strategies in cancer research [49]. From a more general medical point of view, the study of G-quadruplexes is also of interest, for exemple Asamitsu et al. [50] showed their role in the modification of nerves. Furthermore, they are responsible for the development of neurological diseases, including the X-linked intellectual disability syndrome [51].

In a virological context, G-quadruplexes also represent an interesting therapeutic target because they play a role in the viral replication cycle [52, 53, 54]. For example, the SARS unique domain (SUD) of the non-structural protein of SARS-CoV, and SARS-CoV2, can bind to guanine-rich RNA sequences, including G-quadruplex [55, 56], which allows the virus to evade the host immune system's response [57]. In the same context, G-quadruplex RNA structures were identified in the SARS-CoV-2 genome [58]. G-quadruplexes are present in the diversity of life. They are part of the resistance mechanism of the bacterium *Deinococcus radiodurans* towards ionizing radiations [59] and are involved in the mechanisms of plant development and growth [60, 61]. So, the interest of G-quadruplexes is so wide that bioinformatic algorithms and tools have been developed for the sole purpose of tentatively identifying and locating them in the genome [62, 63].

4.3 Biotechnological interests of G-quadruplex

G-quadruplexes represent an interesting topic in the development of nanotechnologies. Some G-quadruplexes are lipophilic, so they can be inserted into membranes and mimic ion channels [64]. They are also able to recruit hemes into cells, moreover the heme/G-quadruplex association forms a DNAzyme with peroxidase activity [65, 66]. They can also be used in the assembly of nanotubes [67]. Other biotechnological applications involve ionophores, hydrogel, enhanced fluorescence, assisted polymerization of crystalline nanorods, DNA Origamis or DNA logic gates [68, 69].

2

Structural parameters of G-quadruplexes

CHAPTER 5

Structures of G-quadruplexes

Chapter contents

5.1 Properties of the guanine and formation of G-quadruplexes

Guanine is an aromatic organic molecule with the chemical formula: $C_5H_5N_5O$. It is one of the nitrogenous bases found in nucleic acids (DNA or RNA). This molecule adopts a planar structure, which is derived from purine. Figure 5.1 shows the chemical structure of guanine. The proper nomenclature is as follows [70]:

- when the base only is referred then the name guanine should be used, because the R moiety is a hydrogen (H);

- when discussing the nucleoside then the name guanosine, or deoxy-guanosine, should be used, because the R moiety is a sugar (ribose or deoxyribose);
- when referring to the nucleotide then name guanylic acid, or deoxy-guanylic acid should be used, because the R moiety is now a sugar (ribose or deoxyribose, for RNA or DNA, respectively) covalently linked to a phosphate.

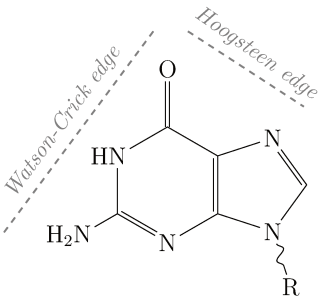
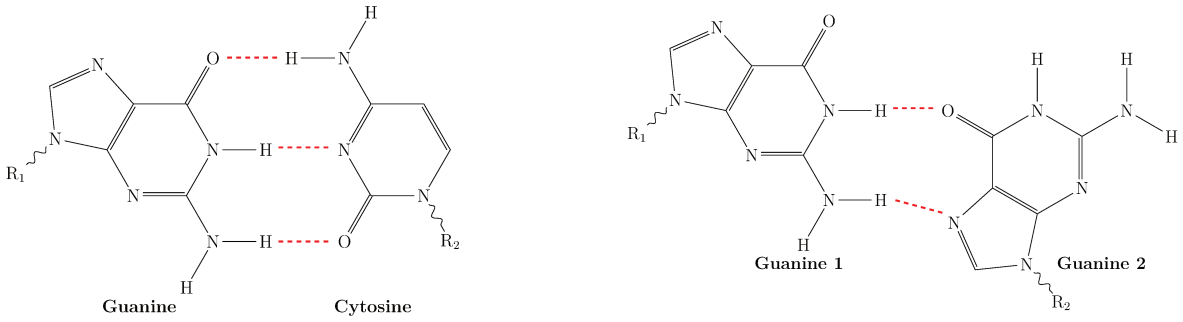


Figure 5.1: Chemical structure of the guanine, with the representation of the Watson-Crick edge and Hoogsteen edge involve in base pairing. *Adapted from Leontis [71]*

Guanine present two most prominent patterns leading to the formation of hydrogen bonds with other nucleotides. The first was described by Watson and Crick, while the other was described later by Hoogsteen [4, 6]. Each of the two interaction sites is shown in the Figure 5.1. Due to their different location, these two sites generate hydrogen matching of different organisation. The Watson-Crick area can make three H-bonds, and the Hoogsteen area can only make two H-bonds. Consequently, this arrangement geometrically limits the possibilities of matching with other nucleotides. The Watson-Crick matching of guanine is complementary to a cytosine, while the Hoogsteen matching changes this rule and grants guanine the ability to bind other nucleotides. Donohue and Trueblood [72] showed the possibility of an adenine-guanine pairing, which was then taken up and discussed further by Hoogsteen [6]. Nevertheless, a guanine is also able to form hydrogen bonds with another guanine, as shown in Figure 5.2. Although Watson-Crick type base pairing is the most commonly encountered, the Hoogsteen model also exists in double-stranded DNA, albeit in smaller proportions. Interestingly, Hoogsteen base pairs also have a biological role in damage and repair, in DNA/protein interactions and even in the replication machinery [73, 74].

Watson-Crick and Hoogsteen base pairs represent two different types of matching. They can also act in concert in the case of triple helix structure of nucleic acids [75]. Reverse Watson-Crick and Hoogsteen base pairing act in the formation of i-motifs, cytosine-rich four-stranded nucleic structures [76]. However, there are particular structures that



(a) Watson-Crick base pairing

(b) Hoogsteen base pairing

Figure 5.2: Hydrogen bonds of guanine with other nucleotides, according to the Watson-Crick or Hoogsteen models. H-bonds are shown as dotted red lines.

are formed by Hoogsteen type pairing only. It involves four guanines, in the same G-G configuration as shown in Figure 5.2, which establish an almost planar arrangement. This particular configuration was already observed *in vitro* in 1962 [77], and is called a guanine tetrad, or quartet. But these quartets are not found alone in a G-quaduplex. They are stacked on other quartets $\pi - \pi$ stacking interactions, forming a system of several parallel quartets. A guanine quartet arrangement and a schematic representation of a three-quadrtets G-quadruplex is shown in Figure 1.

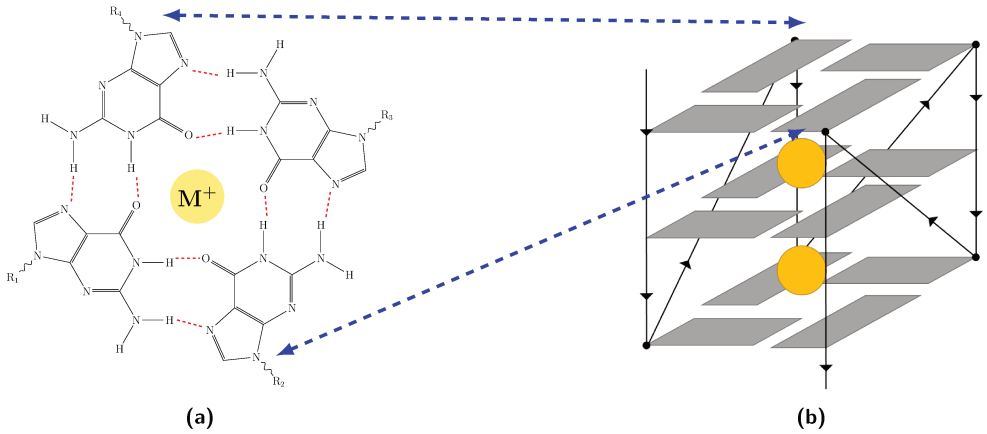
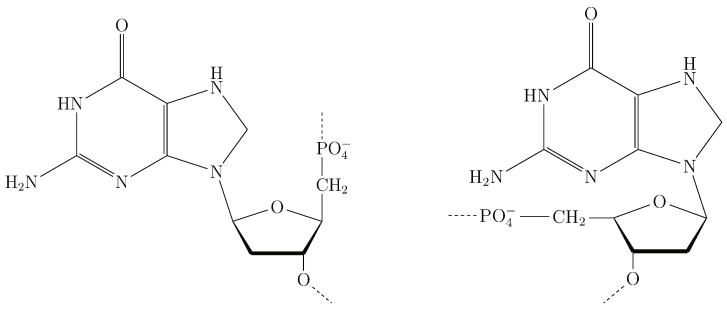


Figure 5.3: Guanine quartet **(a)** superimposed on two other quartets **(b)** in a G-quadruplex arrangement stabilised by the presence of alkali metal cations (M^+ represented by yellow spheres).

Hoogsteen pairings in guanine quartets can be formed in *anti* or *syn* arrangement with respect to the glycosidic bond (Figure 5.4) [78]. If a quartet undergoes alteration of the hydrogen bond pattern, such as a decrease of H-bonds, then the destabilisation of the quartet usually follows [79].

However, the stabilisation of a G-quadruplex structure does not rely solely on Hoogsteen base pairs. In fact, a G-quadruplex consists of at least two guanine quartets parallel to each other and including an alkali metal cation. The presence of the cation is necessary



(a) *Anti* **(b)** *Syn*

Figure 5.4: *Anti* and *Syn* conformations of a deoxy-guanylic acid fragment.

to counterbalance the negative charge of the phosphate groups in the sugar-phosphate backbone. The most common cation is potassium, K^+, but there are other types of cations within G-quadruplexes, such as alkali and alkali earths metals and sometimes non-metals, such as the ammonium ion, $NH_4{}^+$. Despite this rich variety, not all cations have the same stabilising strength [80, 81]. Largy, Mergny, and Gabelica [82] propose the following order: $Sr^{2+} > K^+ > Ca^{2+} > NH_4{}^+, Na^+, Rb^+ > Mg^{2+} > Li^+ \geq Cs^+$. However, the presence of a cation is crucial for the formation and stabilization of G-quadruplexes [78].

Two last important parameters for G-quadruplex stability and formation are: hydration and glycosidic bonds angles. Indeed, the hydration state controls the conformation, also called topology, during the folding of a nucleic acid into a G-quadruplex. More precisely, it is the dehydration that allows this specific folding. Depending on nucleotide sequence, length of loops [83], and environmental conditions, a G-quadruplex will fold in one of three possible topologies: parallel, antiparallel and hybrid (also call mixed parallel/antiparallel, or 3+1) [84, 85, 86]. Water molecules interact mostly with N2, N3 and O4' atoms of guanine [78]. Also, we know that the glycosidic bonds angles differ in the G-quadruplexes topologies [87]. Therefore, the environmental conditions capable of modifying these angles will also modify the global G-quadruplex topology.

5.2 Topology classification

G-quadruplexes are cation-stabilised stacks of guanine quartets, connected by sugar-phosphate loops. However, different DNA sequences produce different arrangements, i.e. conformations in which each nucleotide adopts different geometrical parameters. Factors

affecting the topology include the number of molecules forming the G-quadruplex, the organisation of the loops, and the arrangement of the quartets. Some experimental techniques, such as circular dichroism, are able to provide some information on the topology of G-quadruplexes. However, resolving at an atomistic scale the precise structures of G-quadruplexes is more delicate and time-consuming.

5.2.1 Bulge formation between quartets

The sequence of G-quadruplexes often consists of repeating motifs: for example `TTAGGG` for the human telomeric G-quadruplex. For this reason, G-quadruplex sequences are often represented with the guanines forming the quartets separated by linking sequences, such as in the consensus sequence: $G_{3+}N_{L1}G_{3+}N_{L2}G_{3+}N_{L3}G_{3+}$; where N_L is a nucleic sequence of arbitrary length [88, 89]. However, it has been seen that G-quadruplexes can be formed even if there are breaks in the sequence of three guanines, for example in `d(GGG TT GCGG A GGG T GGG CCT)`, where a cytosine interrupts the guanine sequence. The latter is however ejected from the G-quadruplex core, forming a small loop called a bulge (see Figure 5.5) [90, 89, 91].

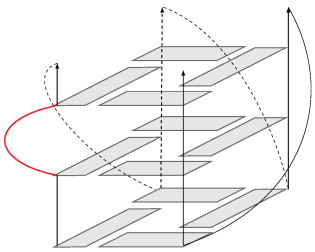
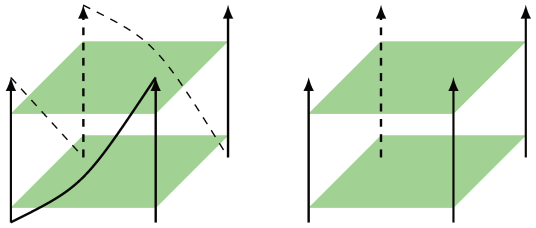


Figure 5.5: Intramolecular G-quadruplex containing a bulge, in red.

5.2.2 *Intra-* and *inter-* molecular folding of G-quadruplexes

G-quadruplexes form a four-stranded DNA or RNA superhelix, and thus four grooves. However, not all G-quadruplexes have the same topology. G-quadruplexes can be formed by the folding of a single DNA strand, in which case they are said to be intramolecular (or unimolecular) or of different DNA strands, in which case they are said intermolecular. For example bimolecular (dimeric) or tetramolecular (tetrameric), are formed by two or four different DNA strands, respectively [88], as shown in Figure 5.6.

Each G-quartet in a G-quadruplex can be constituted, theoretically, by 1 to 4 DNA strands, which we will name sequentially from A to D. In the example of a dimeric G-quadruplex: either two guanine bases in the tetrad belong to strand A and the other two



(a) Intra-molecular **(b)** Inter-molecular

Figure 5.6: Schematic representation of *intra-* and *inter*-molecular G-quadruplexes. Each green tetragon represents a quartet.

to strand B (A2-B2), or three guanine bases belong to strand A and only one to strand B (A3-B1). From this reasoning, it is possible to summarize in Table 5.1 all possibilities to form intramolecular and intermolecular G-quadruplexes.

Topology	A	B	C	D
Monomeric	4			
Dimeric	3	1		
	2	2		
Trimeric	2	1	1	
Tetrameric	1	1	1	1

Table 5.1: Possible foldind of a G-quadruplex with a maximum of four strands, from A to D. Each number represents the number of guanine bases in a G-tetrad belonging to a strand in the G-quadruplex.

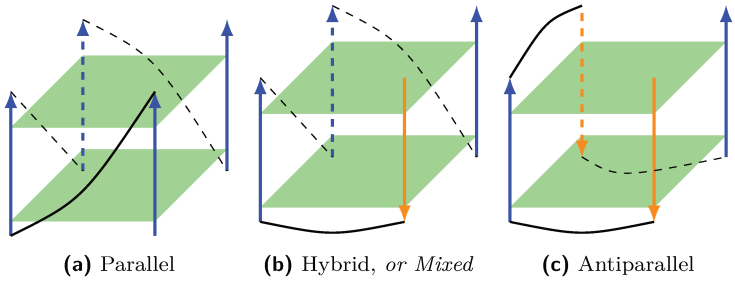
5.2.3 G-quadruplexes topologies definitions

Previously it was mentioned that three topologies exist for a G-quadruplex: parallel, hybrid (or mixed), and antiparallel. These nomenclature originate from the relative arrangement of the $(5' \rightarrow 3')$ directions of the four strands connecting the stacked G-tetrads [78], see Figure 5.7.

- If all strands forming the guanine tetrads are oriented following the same polarity with respect to each other, then the G-quadruplex is parallel. $(\uparrow\uparrow\uparrow\uparrow)$
- If only one strand has a reversed polarity orientation with respect to the others, then the G-quadruplex is hybrid. $(\uparrow\uparrow\uparrow\downarrow)$
- If two strands are oriented with reversed polarity with respect to the other two, then the G-quadruplex is antiparallel. $(\uparrow\uparrow\downarrow\downarrow)$

5.2.4 Linking loops types

In intramolecular G-quadruplexes the guanine bases in the G-tetrads are linked by loops made by a sequence of nucleotides that do not participate in the formation of

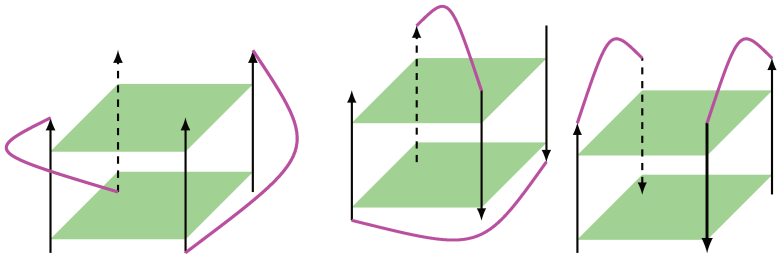


(a) Parallel **(b)** Hybrid, *or Mixed* **(c)** Antiparallel

Figure 5.7: Schematic representation of the three possible topologies in an intramolecular G-quadruplexes, where the arrows represent the ($5' \rightarrow 3'$) strand direction. Arrows with the same color represent parallel strands with the same direction.

the same guanidic core of the G-tetrad. There are three possible types of loops (see Figure 5.8): external, diagonal and lateral [78].

- External loops are established between two adjacent parallel strands.
- Diagonal loops connect two opposite anti-parallel strands.
- Lateral loops connect adjacent anti-parallel strands.

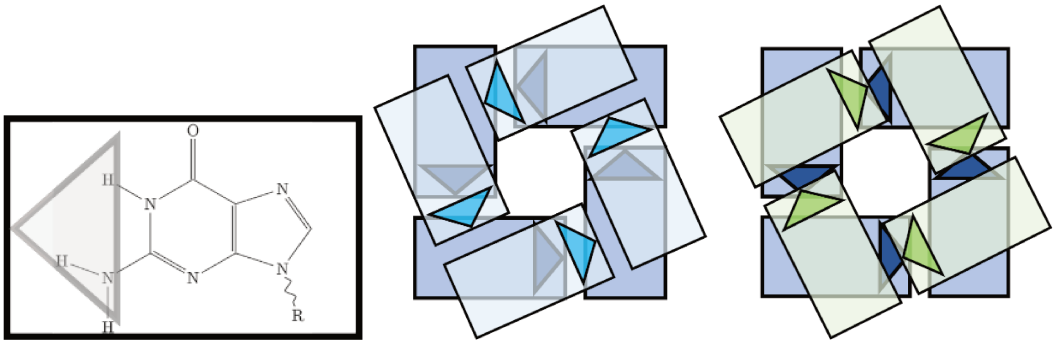


(a) External, or *Propeller* **(b)** Diagonal **(c)** Lateral

Figure 5.8: Schematic representation of the three loop arrangements (in magenta) in G-quadruplex. This examples are shown in inter-molecular G-quadruplexes.

5.2.5 Quartets stacking and classification

The parallel superposition of quartets is the core motif leading to the G-quadruplex formation, originating by a $\pi - \pi$ stacking interactions between the guanine bases. In each quartet, the guanines are arranged according to a direction of rotation, sometimes called quartet polarity, as we discussed in the previous section and in Figure 1: either the two superimposed quartets show homopolar stacking, i.e. the four guanines in the two tetrads rotate in the same direction; or they show heteropolar stacking, i.e. the guanines of the two tetrads rotate in opposite directions [85, 92], see figure Figure 5.9.



(a) Schematic guanine polarity **(b)** Homopolar stacking **(c)** Heteropolar stacking

Figure 5.9: Schematic representation of the stacking of two G-tetrads. The O6 oxygen is always near the center of the tetrad. *Adapted from Lech, Heddi, and Phan [92].*

CHAPTER 6

Structural Elucidation by Circular Dichroism

These days, every Tom, Dick and Harry thinks he knows what a photon is, but he is wrong.

Albert Einstein – 1951

Chapter contents

The use of circular dichroism (CD) properties of polarized light gives rise to a very powerful molecular analysis technique. By exploiting the optical activity of chiral molecules, or supramolecular aggregates such as proteins or nucleic acids, it is possible to characterize their secondary structures and monitor their conformational changes, for example following their interactions with ligands. This chapter proposes to recall some fundamental aspects of CD.

6.1 Wave properties of light

In the history of physics, light has been a subject that has attracted many intellectuals from all over the world. After the ancient and medieval theories, there was a debate lasting for some centuries opposing the corpuscular theory enunciated by Newton and the wave theory enunciated by Huygens [93]. Today, we know that light has both properties, i.e. it is both a wave and a particle.

The photon is the elementary particle of light, however, light can be described as an electromagnetic field. For more clarity, a field is what represents all the quantitative values (q) that a physical entity can take, at any point in a given space (x,y,z) and time (t) [94, 95]. There is a spatio-temporal dependence (x,y,z,t) of the constituent values of a field. Thus a field is a five-dimensional object, which is constituted by the set of all (x,y,z,t,q) possible coordinates.

More specifically, an electromagnetic field is the sum of an electric field ($\overrightarrow{E}$) and a magnetic field ($\overrightarrow{M}$), perpendicularly to each other. Both propagate perpendicularly to the axis of their direction of propagation ($\overrightarrow{d}$), as represented in Figure 6.1. In fact, this direction of propagation is none other than that of light. Electric and magnetic fields have a wave-like behaviour. As a result, an electromagnetic field behaves like a wave [96]. The beginnings of the formulation of the properties of electromagnetism were formalized by James Clerk Maxwell in the 19^{th} century [97]. Today the wave equations describing electromagnetic properties are presented in four differential equations [94, 98].

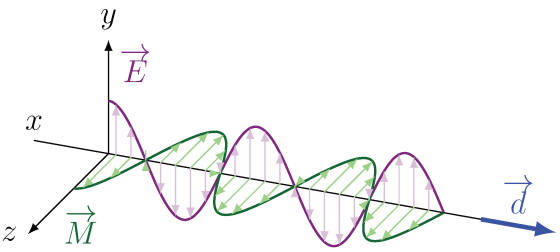


Figure 6.1: Perpendicular propagation in the direction $\overrightarrow{d}$ of an electromagnetic wave, composed of a magnetic wave $\overrightarrow{M}$ (in green) and an electric wave $\overrightarrow{E}$ (in purple), perpendicular to each other.

Here, we will focus only on the following properties of light waves:

- Light is an electromagnetic wave.
- Light can be polarized.

- The speed of light in vacuum c is exactly 299 792 458 $m.s^{-1}$ [99].

- The wavelength λ, which is the distance between two point (usually taken as the wave maximum amplitude) of the oscillation in phase. The energy of a photon is directly related to the wavelength of light.

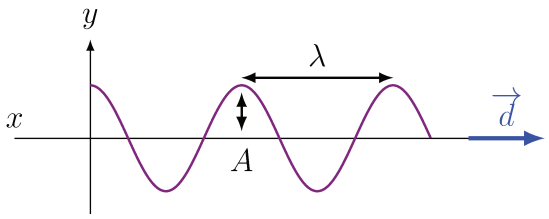


Figure 6.2: A wave is characterized by its wavelength λ and its amplitude A.

- The frequency f, which is dependent on the speed of the light in vacuum c and the wavelength of the light λ :

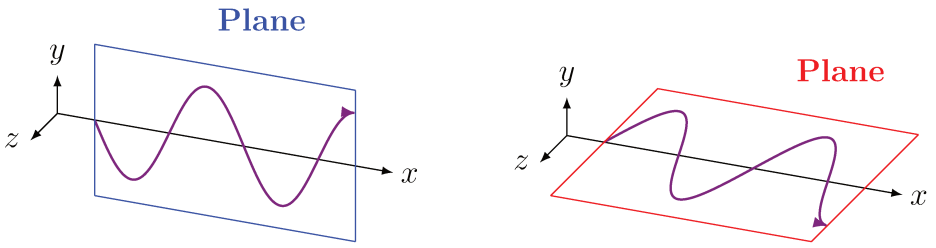
$$f = \frac{c}{\lambda} \tag{6.1}$$

6.2 Light polarisation

Light is a wave that it can be polarized, i.e. it can be oriented in space. We can explain this phenomenon if we consider that the oscillation of the electric wave in the direction d (here, we consider that the direction corresponds to the x-axis) is the result of two components able to oscillate separately from each other and at a given frequency. We called this two components : y component and z component. Finally, we notice that, this reasoning dismantle the simple vectorial picture of the light wave [100].

6.2.1 Linear polarisation

When the oscillations of the y and z components reach a maximum at the same time they are said to be "in phase". So when the vibrations of the y and z components are in phase, or in phase opposition, and have a certain proportionality to each other, then the result is a direction in the polarisation plane P_{xz} [100, 98]. There are several polarization planes generally ranging from $+90°$ to $-90°$. A $0°$ plane is a vertically polarized light and is perpendicular to a plane at $+90°$ or at $-90°$, corresponding to a horizontally polarized light. So there is an equivalence of orientation only between the $+90°$ and $-90°$ planes.

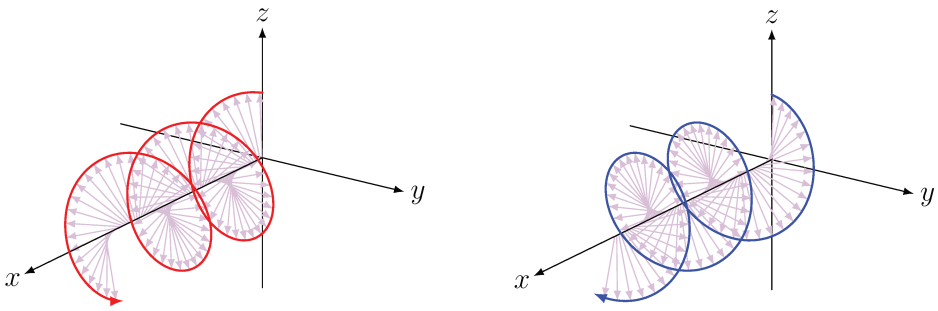


(a) Linear polarisation oriented at 0° **(b)** Linear polarisation oriented at −90°.

Figure 6.3: 0° polarization plane resulting to $y_{vibrations} = 1$ and $z_{vibrations} = 0$, and +90° polarisation plane resulting to $y_{vibrations} = 0$ and $z_{vibrations} = 1$ [100]. These two planes are perpendicular to each other.

6.2.2 Elliptical polarization and circular dichroism

The polarization of light does not always follow a plane, it can be elliptical if the oscillations of the y and z components are not in phase and present different amplitudes, leading to elliptically polarized light. But sometimes the phase shift of the y and z component vibrations is approximately 90° and happens with the same amplitude. In this case the polarization becomes circular, with two directions of polarization. If the electric wave describes a circle rotating clockwise with an angular velocity α, then it's named right-hand circular polarized light. Conversely, if the electric wave draws a counter-clockwise circle we are in presence of left-hand circular polarization [96, 98].



(a) Left-hand circular polarization **(b)** Right-hand circular polarization

Figure 6.4: Right-handed and left-handed circular polarization of light.

To conclude, the designation « Circular dichroism (CD) » refers to the circular polarisation of the light wave. Since there are only two possible oscillations, left or right, the designation is also based on the term « dichroism » which comes from the Greek word *dikhroos* which means « of two colors » [101]. A molecule is said « dichroic » when it absorbs left and right polarized light differently [102]. Finally, according to Fresnel's construction, we can consider that the resultant of two circularly polarized waves is a linearly polarized wave.[96]. There is an intrinsic link between linear and circular polar-

ization, since the sum of two left- and right-handed polarized lights, in phase and with the same amplitude, produces a linear polarized light.

6.2.3 « Natural » light

Light is typically non-polarized, which means that all overlapping waves sent by the source are polarized according to different and random planes [96].

6.3 Optical activity and chirality

The optical activity is the ability of a substance to change the plane of polarization of the incident polarized light. This property can be rationalized by the symmetry of the molecule (or macromolecule), indeed it is related to the absence of reflection symmetry [100]. Maurizot [96] goes into more details and lists the three geometrical conditions that a molecule must possess in order to be optically active. The molecule in question must have neither center of symmetry, nor plane of symmetry, nor improper axis of symmetry. When we pay attention to these details, we realize that these are the necessary conditions for a molecule to be chiral. As a reminder, two molecules with the same raw formula, but a different topology are called isomers. When two isomers are the mirror image of each other, and their structure is not superimposable, then they are chiral. The two mirror images of the molecule form a pair and are called enentiomers [103]. For example, the amino acid alanine, in Figure 6.5, is a chiral organic molecule.

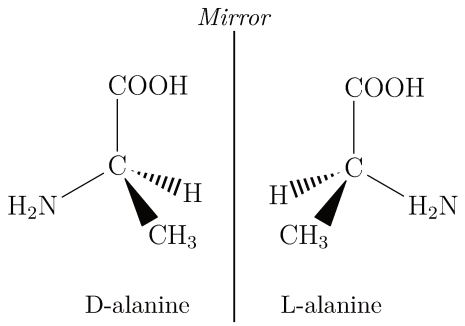


Figure 6.5: The amino acid alanine is a chiral organic molecule. It has two configurational isomers: D-alanine and L-alanine. One is the mirror image of the other, and are not superimposable. So D-alanine and L-alanine are enantiomers, with the internal C as chiral center.

Optical activity takes place only for one enantiomer of a chiral molecules. To explain this phenomenon, we must remember that linearly polarized light is a resultant of left and right circularly polarized light waves (respectively $\overrightarrow{W_l}$ and $\overrightarrow{W_r}$). If the optical path in

a solution is different for $\overrightarrow{W_l}$ and $\overrightarrow{W_r}$. Then the solution has a circular birefringence, i.e. different refractive indices for right and left circularly polarized light. Then a phase shift between $\overrightarrow{W_l}$ and $\overrightarrow{W_r}$ may be observed [96]. Concretely, the incidence of the polarized light on the molecule according to the angle α induces a down-up movement of the electrons in the direction α. Then, it will generate an electric field which comes out of the molecule. In other words, if the electric wave of the light enters in contact with the molecule at the point P, it will come out at the point $P + D$, with D the diameter of the molecule [100]. Consequently the phase shift between $\overrightarrow{W_l}$ and $\overrightarrow{W_r}$ depends on the diameter of the molecule and by its capacity to absorb preferentially $\overrightarrow{W_l}$ or $\overrightarrow{W_r}$. Now, by applying the Fresnel construction to $\overrightarrow{W_l}$ or $\overrightarrow{W_r}$ dephased, we have a polarized light in another plane [96]: $\alpha \pm \beta$, as shown in Figure 6.6 below.

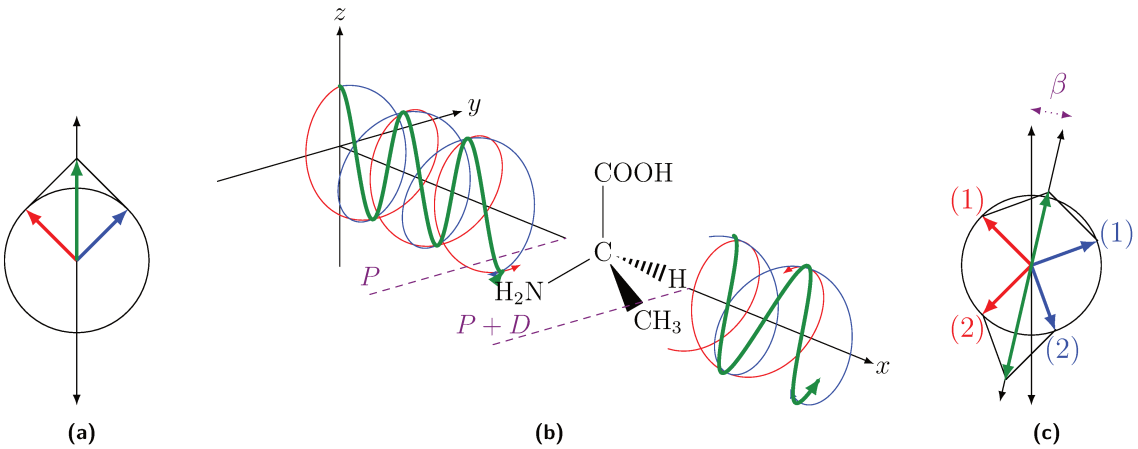


Figure 6.6: **(a)** Fresnel construction showing the superposition of $\overrightarrow{W_l}$ (red) and $\overrightarrow{W_r}$ (blue) forming a linear polarized light according to the angle *alpha* (here, equal to +90°). **(b)** This light beam is send to a D-alanine molecule. The ray $\overrightarrow{W_r}$ is absorbed by the molecule at point P and comes out at point $P+D$, while $\overrightarrow{W_l}$ is not absorbed. So $\overrightarrow{W_r}$ has a delay and is out of phase with $\overrightarrow{W_l}$. **(c)** This causes a shift of the polarization plane to the right, as shown by a second Fresnel construction. *Based on Maurizot [96] and Feynman, Sands, and Leighton [100]*

6.4 Circular dichroism

The phenomenon of circular dichroism is intimately linked to absorption and optical activity. Since it is only observed in the spectral range absorbed by an optically active molecule. But unlike absorbance which is always positive, circular dichroism can have positive and/or negative values. In fact, the circular dichroism is explained by developing what has been mentioned in section 6.3. An optically active molecule preferentially

absorbs left or right polarized light. This generates a phase shift of the two components, thus a polarization of the light according to a different angle plane than that of the incident light. However, if we go back to the Figure 6.6 we notice that the discussion does not mention the intensity of the absorption of light by the molecule. Indeed, we see in c) that the Fresnel diagram uses vectors of the same radius. But a molecule will not always send back an absorbed light with the same amplitude. That is to say that there is an intensity of absorption. When $\overrightarrow{W_r}$ and $\overrightarrow{W_l}$ are projected with different radii [96], the polarization is not linear, but elliptical as we can as can be seen in Figure 6.7 below.

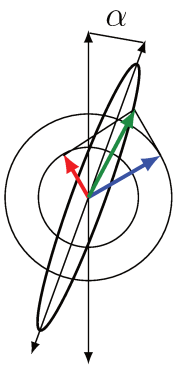
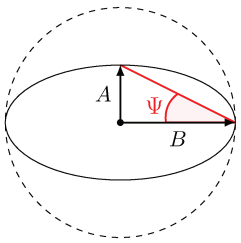


Figure 6.7: The construction of a Fresnel diagram allows to explain the elliptical polarization of a light resulting from an absorption intensity of $\overrightarrow{W_r}$ (blue) and $\overrightarrow{W_l}$ (red) by a molecule. The alpha angle is the rotation of the major axis of the ellipsis with respect to the polarisation axis of the incident light [96].

Two specific parameters are used to describe the circular dichroism phenomenon of a molecule. The first is the *dichroic molar absorption coefficient* $\Delta\varepsilon$, which is expressed in $mol^{-1}.cm^{-1}$. $\Delta\varepsilon$ is calculated as the difference between the absoption coefficients of left and right polarised light : respectively ε_l and ε_r. So its formula is :

$$\Delta\varepsilon = \varepsilon_l - \varepsilon_r \tag{6.2}$$

The last parameter is called ellipticity. In order to explain what this is, it is worth recalling that the light incident on the sample is circularly polarised, while the light exiting the sample is elliptically polarised. Then, if we consider a circle as a particular ellipse, we can characterise the flatness of the ellipse at the sample exit. To do this, the angle Ψ is determined from the calculation of its tangent, which involves dividing the smaller half-axis of the ellipse (A) by its larger half-axis (B) [96].
Since the ellipticity corresponds to a geometrical characterisation, thank to Figure 6.8 it can be seen that there is a maximum and a minimum value. The first is reached when $S = L$, that is when $\Psi = 45°$. In this case, the ellipse is in fact a circle, so the light is circularly polarized. The latter is reached when $S = 0$, i.e. when $\Psi = 0°$. This simply



(a)

$$\Psi = \arctan\left(\frac{A}{B}\right) \tag{6.3}$$

(b)

Figure 6.8: The ellipticity Ψ is (a) a geometric parameter determining the flatness of an ellipse, and is (b) calculated from the two half-axes of the ellipse.

corresponds to plane-polarised light. So we can say that linearly polarized light has an ellipticity of 0°.

However, the value of Ψ is usually very small, so it is expressed in milli-degrees (mdeg). This is why, one can consider that $\Psi = \tan(\Psi)$. Thus, another possibility to express the magnitude of the CD for this second parameter is to use the molar ellipticity $[\theta]$, expressed in $mdeg.cm^2/mol$:

$$[\theta] = \frac{\Psi}{C \times l} \tag{6.4}$$

Where C is the concentration [1] in $mol.L^{-1}$ and l is the optic path length in cm [96]. On the other hand, there is a variation of the molar ellipticity, applied to proteins and nucleic acids, it is the Mean residue molar ellipticity (MRME). Its calculation is based on the formula for molar ellipticity, but adds the number (n) of amino acid or nucleotide residues of the macromolecule of interest [104, 105].

$$MRME = \frac{\Psi}{C \times l \times n} \tag{6.5}$$

There is a relationship between the two parameters $\Delta\varepsilon$ and $[\Psi]$. Indeed, it has often been shown that for small values of Ψ, one can write the following equation [96, 106, 107] :

$$[\theta] = \frac{4500}{\pi} \times \log_e(10) \times \Delta\varepsilon \tag{6.6}$$

$$\approx 3298.2 \times \Delta\varepsilon \tag{6.7}$$

To finish, the CD spectrum is constructed by calculating the above parameters for each wavelength λ. For exemple, a spectra can be obtain by plotting $[\Psi]_\lambda$ in relation to λ.

[1] To remember $1L = 1000\ cm^3$. So the concentration in $mol.L^{-1}$ need to be divided by 1000 to convert it into $mol.cm^{-3}$ (equal to $mol.mL^{-1}$).

6.5 Origin of the G-quadruplex CD

Nucleic acids possess two main chromophores in the UV-Vis range: the ribose or deoxyribose sugar and the nitrogen bases. Purine and pyrimidine bases, have an electronic absorption at around 260 nm. But let's look more closely at the case of guanine, which is the characterizing base in G-quadruplexes. This base has two absorption bands in the UV-Vis region, involving two ($\pi \rightarrow \pi^*$) excited states at 248 nm and 279 nm [85], whose dipole moments are shown in Figure 6.9.

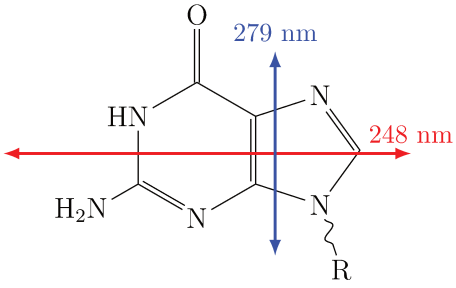


Figure 6.9: Representation of the two dipole transitions moments in guanine causing the relative electronic transitions. The most important is the one around 248 nm. *Adapted from Chaires and Graves [85].*

The other chromophore, the sugar, is characterized by an electrical transition around 180 nm. It also has a very interesting property in the context of circular dichroism, because it is an inherent chiral moiety that can assume D or L configuration (Figure 6.10) [108]. This property gives it the ability to provide a circular dichroism spectrum. However, the sugar chirality is usually not exploited in nucleic acid spectroscopic charcateristic since its absorption takes place in the far UV [96]. On the other hand, the secondary structure formed by nucleic acids procides a chiral arrangement, and hence induced a CD response. Specifically, the circular dichroism spectrum is derived from the electrical transition states of bases arranged in a helicoidal chiral structure. This effect is observed with double-helical DNA, as Z-DNA assumes a left-handed helix structure while both A-DNA and B-DNA assume a right-handed helix structure [109, 110, 111].

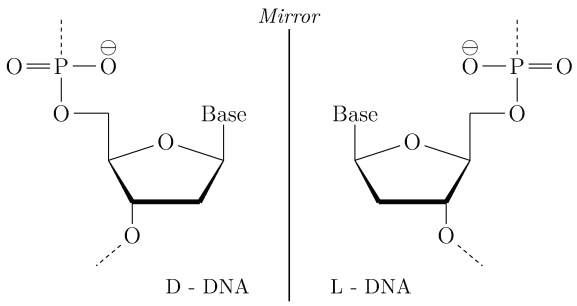


Figure 6.10: Charily of deoxyribose-phosphate sugar. *Adapted form [108].*

For similar reasons, also G-quadruplex structures adopt chiral right-handed or left-handed helical conformations (Figure 6.11) [85, 112]. However, this right-handed or left-handed helix folding does not seem to depend on the nucleotide sequence, but rather on environmental factors such as the presence of other molecules [113, 91, 114]. Interestingly, Winnerdy et al. [115] have shown experimentally that a G-quadruplex can adopt both types of helices forming what they name a " right- and left-handed hybrid G-quadruplex ". In conclusion, the electronic transitions of the guanines and their helical arrangement allows us to obtain a peculiar circular dichroism spectrum. As a matter of fact, CD signal is highly sensitive to the spoecific arrangements of supramolecular aggregates and hence it is a technique of choice for structural determination of nucleic acid arrangements.

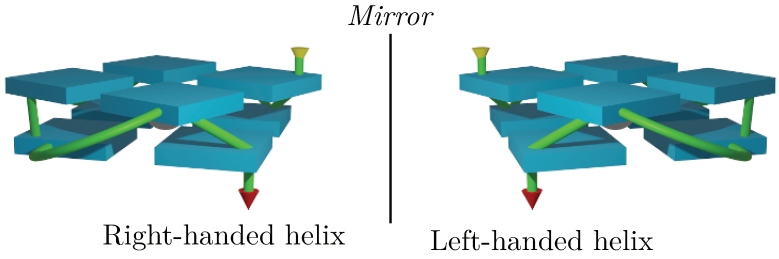


Figure 6.11: G-quadruplex is a chiral macromolecule.

6.6 G-quadruplex characterisation by CD

Circular dichroism spectroscopy is proving to be a very powerful technique for exploring G-quadruplex polymorphism. Each topology (arrangement of quartets and loops) has in fact its own spectrum. However, some generally applicable trends can be identified, as shown in Figure 6.12. For example, the spectrum of the parallel G-quadruplex is characterised by a negative peak at *ca.* 240 nm and a positive peak at *ca.* 260 nm. On the other hand, that of the hybrid G-quadruplex has an additional positive peak at *ca.* 290 nm. While the spectral signature of the antiparallel G-quadruplex differs from the previous two, showing a positive peak at *ca.* 240 nm, a negative peak at *ca.* 260 nm and a positive peak at *ca.* 290 nm [85, 78].

The spectral characteristics and their structural explanations have also been interpreted as in Table 6.1:

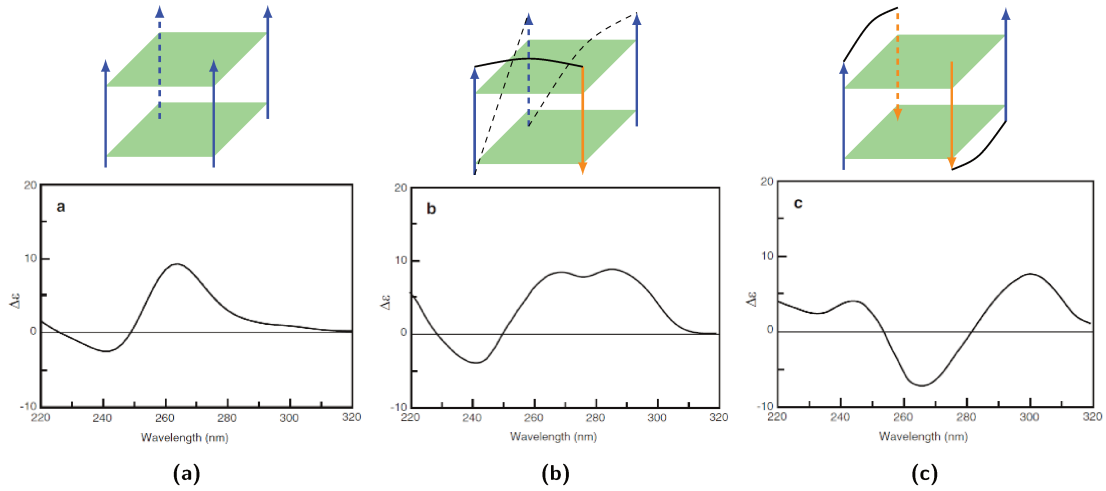


Figure 6.12: Examples of G-quadruplex CD spectra: **(a)** Tetrameric parallel G-quadruplex; **(b)** intramolecular hybrid G-quadruplex; **(c)** dimeric antiparallel G-quadruplex. *Modified from Chaires and Graves [85].*

Wavelength (nm)			Signification
240	*260*	*290*	
		+	Guanines stack with different glycosidic bond angles.
-	+		Guanines stack with the same glycosidic bond angles.
+	-		No guanines stacking with same glycosidic bond angles.

Table 6.1: Summary of G-quadruplexes circular dichroism spectra interpretation: + and - indicate positive and negative peaks, respectively. *Adapted from [85].*

CHAPTER 7

Structural parameters

Chapter contents

Daily practice shows that the development of various strategies for structural analysis of nucleic acids plays an important role in their classification. This started with the description of the helices of the different forms A-DNA, B-DNA and Z-DNA. Then the descriptions were refined to the point of describing the different conditions of the pairing between the bases [116, 117]. Today, software such as CURVES+ [118] or 3DNA [119] are able to provide information on the topology and base pairing of nucleic acids. However, even if the analysis of G-quadruplexes is possible, it is still limited to "general" topological

information. In fact, only few G-quadruplex-specific parameters are usually provided. However, several papers mention specific structural parameters for G-quadruplexes, we will focus on the article by Tsvetkov, Pozmogova, and Varizhuk [120], which gives several parameters specific to G-quadruplexes. Moreover, the authors provide a script for the VMD software [121] that allows to follow the evolution of those parameters along molecular dynamics trajectories.

Despite many efforts to define the structural parameters of G-quadruplexes, there is still a gap with respect to the description of the metal cation. Indeed, the totality of the parameters focus on the arrangement of the bases and the organisation of the sugar-phosphate chain. No method discusses the geometrical properties of the central cation with respect to the surrounding guanines. One exception is the article by Reshetnikov et al. [122], which uses the distance between the Centre of mass (COM) of the ion and the COM of the eight neighbouring O6 atoms.

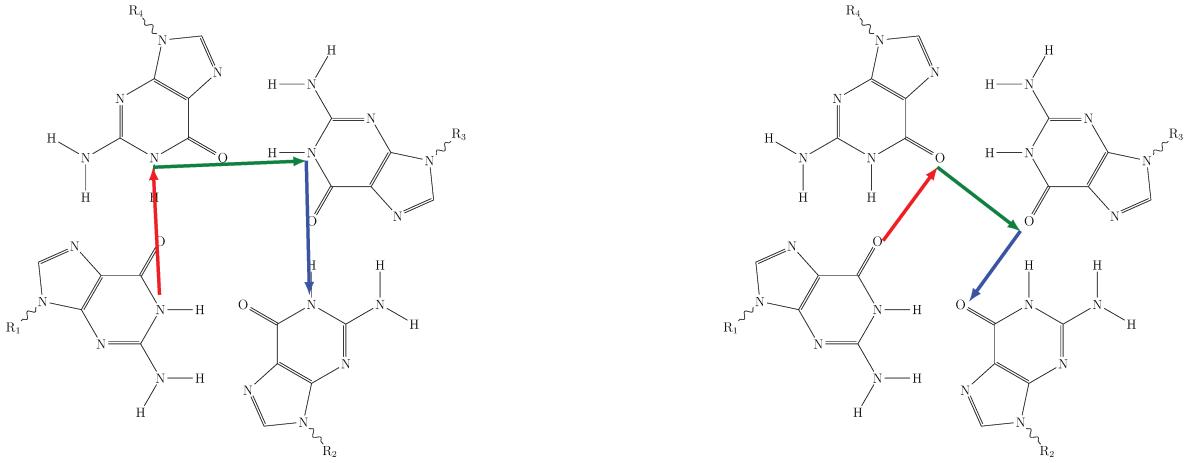
7.1 Angle parameters

7.1.1 Dihedral angle

A dihedral angle, also called a torsion angle, is constructed with three successive vectors that join a chain of four atoms A-B-C-D. More precisely, it is the angle formed between the plane formed by atoms A,B,C and that formed by atoms B,C,D [123]. The value of this type of angle is, in chemistry, a value between 0° and ±180°.

The dihedral angle is an indicator that highlights a out of plane bending of the G-tetrad. There are two specific variants. The strategy adopted by Mashimo et al. [124] is to construct a dihedral angle with the O6 atoms of each guanine involved in a quartet. The other strategy comes from Tsvetkov, Pozmogova, and Varizhuk [120] who propose to use the N1 atoms to construct the dihedral angle. Basically the two strategies are equivalent. The arrangement of guanines in a quartet gives a dihedral angle close to 0°. In fact, this value is never reached; it could only be reached if the four atoms were exactly coplanar. To better understand the construction of the dihedral angle within a G-tetrad, Figure 7.1 below represents graphically each of the two strategies. Later, we

will see other quartet bending parameters: "lengthwise" and "diagonal" quartet bending. Although they have their own characteristics, these three bending parameters give similar information.



(a) Dihedral angle involving four N1 atoms

(b) Dihedral angle involving four O6 atoms

Figure 7.1: Construction of the two dihedral angle variants in a guanine tetrad. In each representation, the dihedral angle is calculated between the red and blue vectors.

Mathematically the dihedral angle (Ξ) is calculated in the same way, whatever the type of atom, N1 or O6, is chosen. So the equation considers four atoms a, b, c and d as follows:

$$\Xi = \frac{180}{\pi} \arccos\left(\overrightarrow{n_{abc}} \cdot \overrightarrow{n_{bcd}}\right) \tag{7.1}$$

Where $\overrightarrow{n_{abc}}$ and $\overrightarrow{n_{bcd}}$ are the normals of the planes abc and bcd, respectively. Both are calculated by the same equation, used here with $\overrightarrow{n_{abc}}$ as example :

$$\vec{n_{abc}} = \left[\frac{\vec{r_{ab}}}{\|\vec{r_{ab}}\|} \times \frac{\vec{r_{bc}}}{\|\vec{r_{bc}}\|}\right] \tag{7.2}$$

$\vec{r_{bc}}$ corresponds to the vector joining two adjacent N1 or O6 atoms.

7.1.2 Guanine-Guanine angle

As explained above, a quartet (or tetrad) is a group of four almost coplanar guanines, arranged in a square arrangement. Therefore, the angle between the longest side of two adjacent guanines should be approximately 90° and between two opposite guanines approximately 180°. If the angles between the guanines change drastically, then also the geometry of the quartet changes and can become an arbitrary convex quadrilateral. Thus

the guanine-guanine angle allows to monitor a geometric deformation of the quartet.

Now, it is appropriate to detail the calculation of this parameter. We will refer to Figure 7.2 to get a visual appreciation of the explanation given below. First, the axis of a guanine ($\overrightarrow{r^i}_{C4C5-C8}$) is defined by drawing a vector between C8 and the midpoint of the distance between the two C4 and C5 carbons. This is repeated for each guanine in the same quartet, to assign them an axis, then the angle between these axes is simply measured.

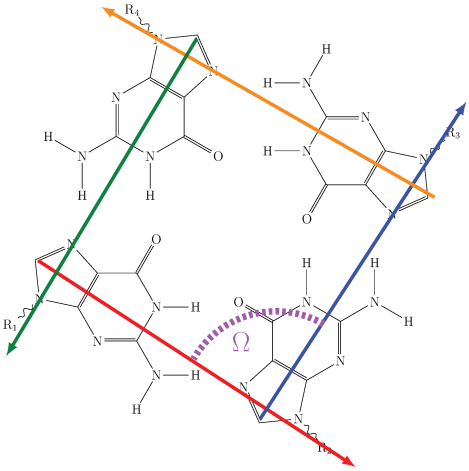


Figure 7.2: Ω is the angle between two guanine axis. In this example it's equal to 90° between red and blue vector, because they are adjacent. And it's equal to 180° between red and orange vector, because they are opposed.

Thus, the equation allowing to calculate the angle Ω may be written as:

$$\Omega_{i-j} = \frac{180}{\pi} \arccos\left(\frac{\overrightarrow{r^i}_{C4C5-C8}}{\left\| \overrightarrow{r^i}_{C4C5-C8} \right\|} \times \frac{\overrightarrow{r^j}_{C4C5-C8}}{\left\| \overrightarrow{r^j}_{C4C5-C8} \right\|} \right) \tag{7.3}$$

7.1.3 Normal of guanine - G-quadruplex Axis & Normal of two guanines angles

These two parameters allow us to define the orientation of a guanine with respect to the axis of the G-quadruplex, or to another guanine belonging to the same quartet. Together, they define a local planarity of the guanines in the tetrad, pointing out to some out of the plane deformations. This method requires a vector to be drawn between the centres of mass of two quartets called α and β, see Figure 7.3. In this way, we obtain an axis ($\overrightarrow{Z^{\alpha\beta}}$) which represents the vertical elongation of the G-quadruplex. But if the G-quadruplex has more than two quartets, such as three quartets, then the $\overrightarrow{Z^{\alpha\beta}}$ axis of

the firts quartet (α) is the vector joining the COMs of the α and β quartets, that of the second quartet $\overrightarrow{Z^{\beta\gamma}}$ is the vector joining the COMs of the β and γ quartets, so we consider that the axis of the γ quartet is the same as that of the β quartet which precedes it. Next, we need to establish the plane corresponding to the guanine. To do this, we draw two vectors: one connects the atoms N9 and N2 ($\overrightarrow{r^i}_{N9N2}$), while the other connects the atoms N9 and O6 ($\overrightarrow{r^i}_{N9O6}$). The two vectors thus created are tangent to the guanine, and form the plane corresponding to the guanine. Subsequently, the normal vector to the guanine plane $\overrightarrow{n_i}$ is drawn. This is a vector perpendicular to the tangent plane and passing through a point; it is calculated by the normalised vector product of the tangent vectors [125]. Tsvetkov, Pozmogova, and Varizhuk [120] provide the following equation:

$$\overrightarrow{n_i} = \left[\frac{\overrightarrow{r^i}_{N9N2}}{\left\| \overrightarrow{r^i}_{N9N2} \right\|} \times \frac{\overrightarrow{r^i}_{N9O6}}{\left\| \overrightarrow{r^i}_{N9O6} \right\|} \right] \tag{7.4}$$

Now the two essential elements are present, namely the vectors of the G-quadruplex axis and the normal vector on the guanine plane. Thus it is possible to calculate two parameters: the angle formed between the normal of the guanine and the G-quadruplex axis (φ), but also the angle formed between two normal vectors of two guanines ($\Delta\varphi$). The closer these two angles are to 0° the closer the guanine under study is to the plane of the quartet or the other guanine. They can be calculated using the following equations (Figure 7.3):

$$\varphi = \frac{180}{\pi} \arccos \left(\overrightarrow{n_i} \cdot \frac{\overrightarrow{Z^{\alpha\beta}}}{\left\| \overrightarrow{Z^{\alpha\beta}} \right\|} \right) \tag{7.5}$$

$$\Delta\varphi = \frac{180}{\pi} \arccos \left(\overrightarrow{n_i} \cdot \overrightarrow{n_j} \right) \tag{7.6}$$

This can also be seen graphically in Figure 7.3.

7.1.4 Twist angle

The helical twist angle, usually indicated as $\Theta^{\alpha\beta}$, represents the relative rotation angle of the G-quadruplex. More precisely, it determines the rotation of a quartet relative to the adjacent one. When extracted from of a molecular dynamics simulation, this parameter conveys information about the general flexibility of the guanine core. We have seen that G-quadruplexes are nucleic acid arrangements that can adopt a large

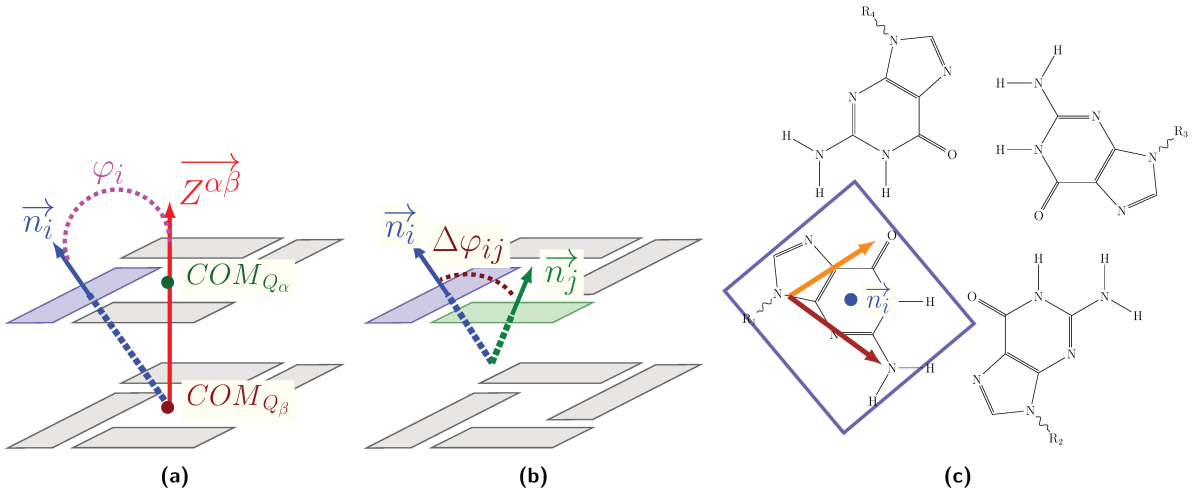


Figure 7.3: Plot of the construction of the angles φ (a) and $Delta\varphi$ (b), completed by the construction of the plane of a guanine and its normal (c).

variety of topologies. However, depending on its topology, a G-quadruplex has its own rigidity. That is, the greater the variation in the helical twist angle curve, the more flexible is the G-quadruplex. On the contrary, the smaller the variation, the stiffer the G-quadruplex.

The principle of measuring the twist angle is simple: measure the angle between two quartets. Lech, Heddi, and Phan [92] establish the value reference, i.e. 0, value of this angle according to the stacking of the O6 atoms. Indeed, the authors have arbitrarily chosen that the value 0° corresponds to the smallest separation distance of the O6 atoms of the stacked guanines. This is the arrangement illustrated their original contribution and reproduced here in Figure 7.4. This method is interesting and quite innovative. Nevertheless, it is not based on structural evidence. In fact, it requires a pre-established knowledge of the stacking of the quartets. Secondly, it makes difficult to compare the twist angle between two G-quadruplexes characterized by a different topology.

Other methods of measuring the helical twist angle have been reported. One of them also uses the O6 atoms of guanines. For each tetrad, these atoms globally form a planar square. Thus, it is sufficient to measure the angle formed between the squares of each quartet [78]. Reshetnikov et al. [126] proposes another method based on the use of the C1' atoms of the guanosine sugar. To do this, we draw a vector between the C1' atoms of two adjacent guanines in the first quartet. Then we do the same in the next quartet. In this way, we obtain two vectors, one for each quartet. The measure of the angle between these two vectors determines the twist angle. As a matter of fact, the method proposed

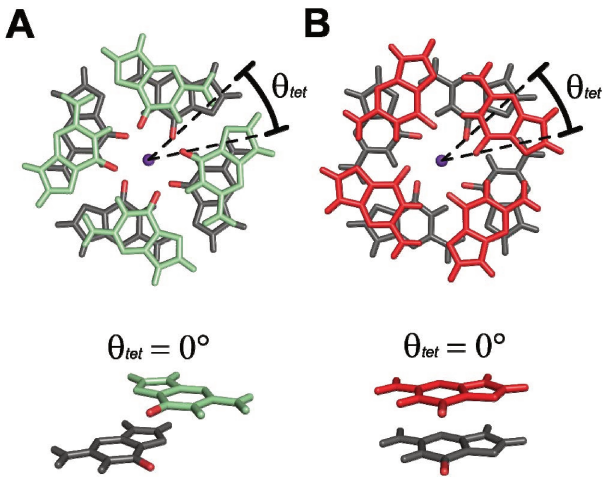
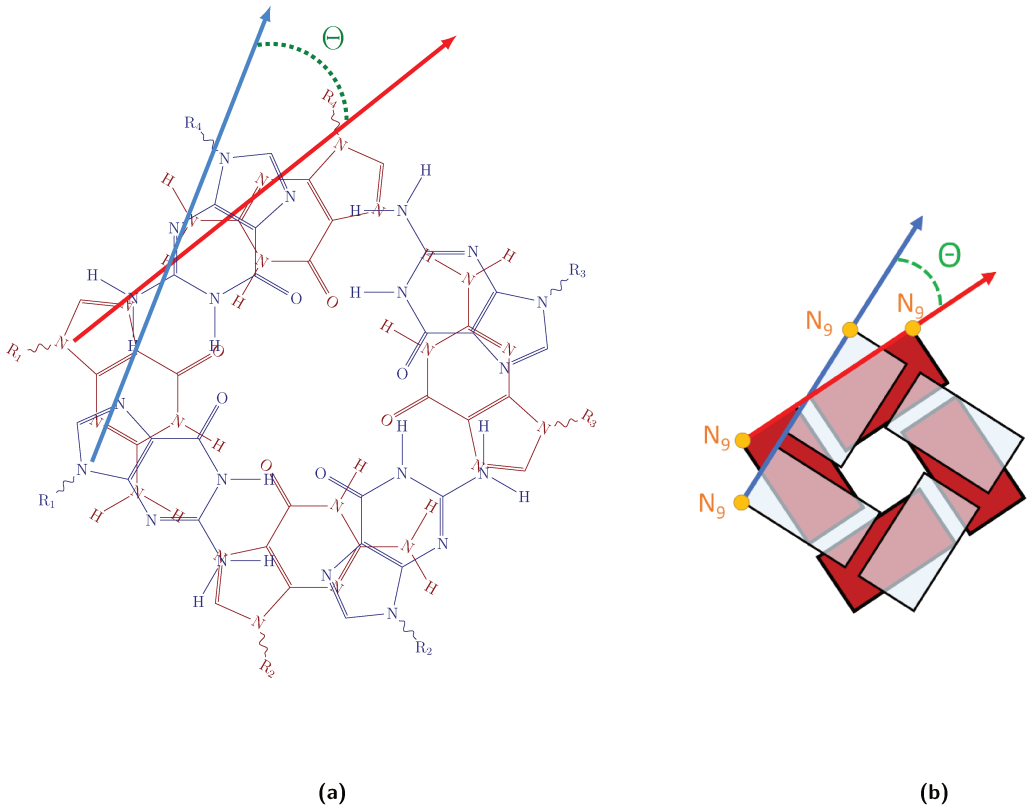


Figure 7.4: Illustration of the method established by Lech, Heddi, and Phan [92] to define the 0° of the twist angle for quartets with opposite polarity (a) or having the same polarity (b). *Figure reproduced from Lech, Heddi, and Phan [92].*

by Tsvetkov, Pozmogova, and Varizhuk [120] is equivalent to the previous one, but uses N1 atoms, as shown in Figure 7.5.



(a) (b)

Figure 7.5: Detailed (a) and simplified (b) representation of the method used by Tsvetkov, Pozmogova, and Varizhuk [120] to measure the helical twist angle (Θ) of G-quadruplexes. The angle is measured between the vectors connecting the N1 atoms of adjacent H-bonded guanines.

Contrary to the first one, this protocols introduce less arbitrary choices and are grounded on structural evidences, which allows them to compare the twist angle of G-quadruplexes independently of their quartet stacking, while they can still be easily

implemented in a software. So, the angle between the vectors of the α quartet and the β quartet is calculated using the following equation, where $\Theta_{ij}^{\alpha\beta}$ is the angle, $\overrightarrow{r_{ij}^{\alpha}}$ is the vector that join guanines i and j of the α quartet by the N1 atoms. $\overrightarrow{r_{ij}^{\beta}}$ is the same angle for the β quartet.

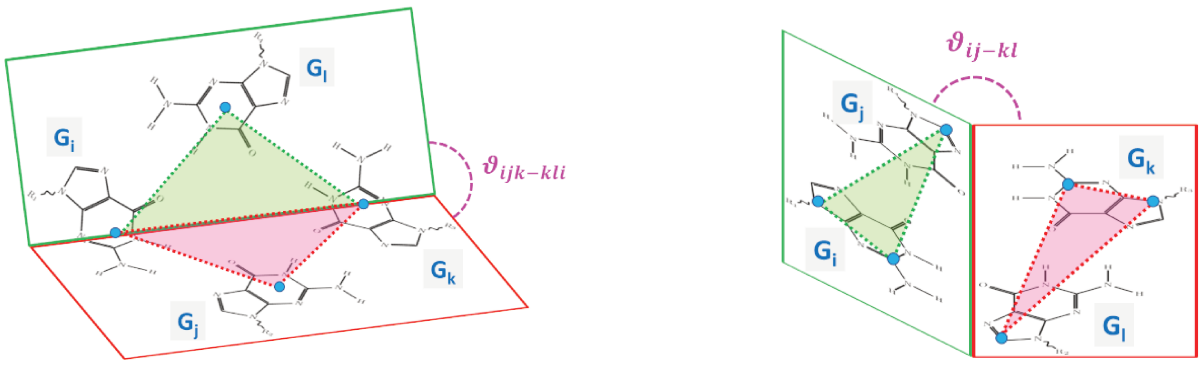
$$\Theta_{ij}^{\alpha\beta} = \frac{180}{\pi} \left(\frac{\overrightarrow{r_{ij}^{\alpha}}}{\left\| \overrightarrow{r_{ij}^{\alpha}} \right\|} \times \frac{\overrightarrow{r_{ij}^{\beta}}}{\left\| \overrightarrow{r_{ij}^{\beta}} \right\|} \right) \tag{7.7}$$

Indeed, this formula is not complete, because it does not refer to all guanines of the tetrad. So the formula of the twist angle is the arithmetic average of the four angles formed by all the guanines, namely:

$$\Theta^{\alpha\beta} = \frac{\Theta_{ij}^{\alpha\beta} + \Theta_{jk}^{\alpha\beta} + \Theta_{kl}^{\alpha\beta} + \Theta_{li}^{\alpha\beta}}{4} \tag{7.8}$$

7.1.5 "Diagonal" and "Lengthwise" bending angles

These two parameters both express the bending of a quartet. They give a more global view of the information provided by the Normal of Guanine - G-quadruplex Axis angle. They are however defined in the same way as the previous ones: the quartet is cut into two planes and the angle formed between them is measured. Then we draw the normal of each plane and measure the angle form by these two vectors.



(a) "Diagonal" bending

(b) "Lengthwise" bending

Figure 7.6: Representation of the construction of the guanine planes and the measurement of their angle, according to the two bending parameters. *Adapted from Tsvetkov, Pozmogova, and Varizhuk [120].*

The "Diagonal" bending parameter ($\vartheta_{ijk-kli}$) uses the center of mass of three guanines to form a plane. Consequently the two planes correspond to those formed by the triangles

connecting the centers of mass of guanines *i,j,k* and *k,l,i*. The normal of each plate is calculated by an identical formula to the one expressed here for the *i,j,k* plane :

$$\overrightarrow{n_{ijk}} = \left[\frac{\overrightarrow{n_{ij}}}{\|\overrightarrow{n_{ij}}\|} \times \frac{\overrightarrow{n_{jk}}}{\|\overrightarrow{n_{jk}}\|} \right] \tag{7.9}$$

Thus, the "Diagonal" bending angle is:

$$\vartheta_{ijk-kli} = \frac{180}{\pi} \arccos\left(\overrightarrow{n_{ijk}} \cdot \overrightarrow{n_{kli}}\right) \tag{7.10}$$

The "Lengthwise" bending parameter (ϑ_{ij-kl}) uses two vectors to define each plane. The first vector connects the N9 and C2 atoms of guanine *i*. While the second vector connects the N9 atom of guanine *i* to the C8 atom of guanine *j*. The *ij* plane is thus created; the *kl* plane is drawn in a similar way. Then the normal of both planes are calculated in the same way, detailed here for the *ij* plane:

$$\overrightarrow{n_{ij}} = \left[\frac{\overrightarrow{n^{i}_{N9C2}}}{\left\|\overrightarrow{n^{i}_{N9C2}}\right\|} \times \frac{\overrightarrow{n^{ij}_{N9C8}}}{\left\|\overrightarrow{n^{ij}_{N9C8}}\right\|} \right] \tag{7.11}$$

Finally, the "Lengthwise" bending angle is calculated following the equation:

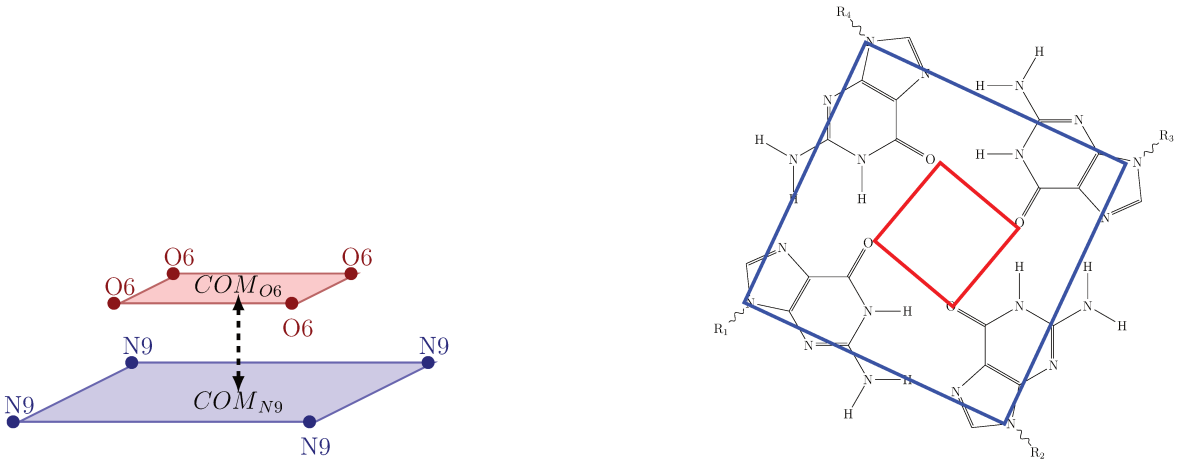
$$\vartheta_{ij-kl} = \frac{180}{\pi} \arccos\left(\overrightarrow{n_{ij}} \cdot \overrightarrow{n_{kl}}\right) \tag{7.12}$$

7.2 Distance and area parameters

7.2.1 Planarity of quartets

Planarity (ρ) corresponds to the fact that all the guanines of a quartet are coplanar. The distortion from planarity can be easily calculated by the distance between the center of mass of the tetragon formed by the four O6 atoms and the center of mass of the tetragon formed by the four N9 atoms [127], see Figure 7.7, using this equation:

$$\rho = COM(N9^{tetragon}) - COM(O6^{tetragon}) \tag{7.13}$$



(a) Separation of the O6 and N9 tetragons

(b) Construction of the O6 and N9 tetragons

Figure 7.7: Graphical representation of the calculation of the planarity of a quartet. *Adapted from Reshetnikov, Golovin, and Kopylov [127].*

7.2.2 Separation of G-tetrads

The separation distance d_{tet} is the distance between parallel quartets [92] in a G-quadruplex. It is obvious that tideally such a definition implies that the guanines in the quartet should be coplanar, see Figure 7.8. This parameter is equivalent to the separation distance of the stacked base pairs in double stranded DNA.

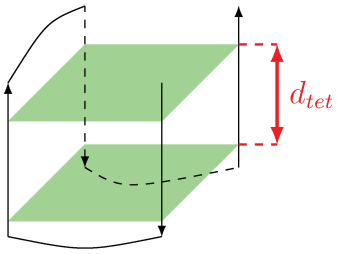


Figure 7.8: Schematic representation of the separation distance between G-tetrads.

The separation distance is calculated by the difference of the center of mass of each quartet stacked on the previous one along the axial elongation of the G-quadruplex:

$$d_{tet} = COM(quartet^{\alpha}) - COM(quartet^{\beta}) \tag{7.14}$$

7.2.3 Guanine-Quartet COMs and Guanine-G4 COMs distances

These are two very useful parameters because they describe the extrusion of each guanine from the quartet or from the G-quadruplex. They complement the planarity parameters in the description of G-quadruplex structural conservation and rigidity.

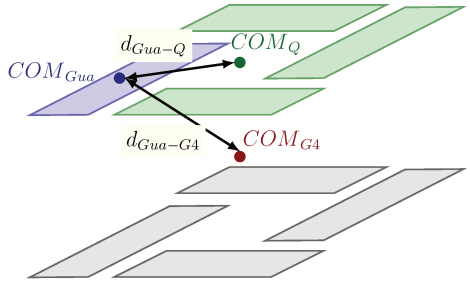


Figure 7.9: d_{Gua-Q} and d_{Gua-G4} are established according to the difference of two mass centers.

As highlighted in Figure 7.9, each parameter is a centre of mass distance. The first is the distance between the centre of mass of the guanine under study and the centre of mass of its quartet (Q). The second is the distance between the centre of mass of the guanine and that of the whole G-quadruplex. Thus we obtain the following equation:

$$d_{Gua-Q} = COM(guanine) - COM(quartet) \tag{7.15}$$

$$d_{Gua-G4} = COM(guanine) - COM(G-quadruplex) \tag{7.16}$$

7.2.4 Area of the O6 tetragon

Reshetnikov et al. [122] use this parameter to track the space between the four guanines in a tetrad. To do this, the four O6 atoms are linked to form a tetragon. Monitoring the evolution of its area makes it possible to follow any dynamical deformation of the guanine tetragon.

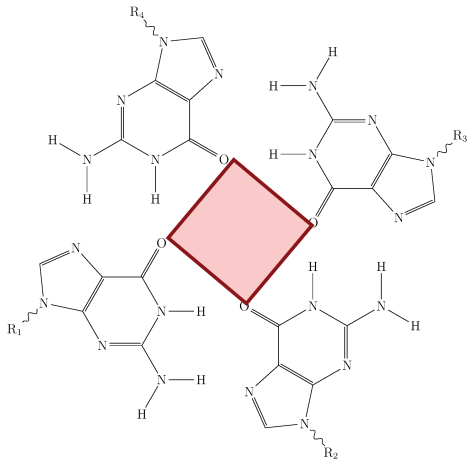


Figure 7.10: Plot the area of the tetragon formed by the four O6 atoms.

7.3 Compactness of G-quadruplexes

The radius of gyration is defined as the root-mean-square average of the distance of all scattering elements from the center of mass of the molecule [128]. It highlights the compactness of a molecule along a molecular dynamics simulation, because the smaller the value, the closer the atoms of the molecule remain relative to the global centre of mass. In fact, its monitoring in different conditions allows to see if an element favours the compaction, or decompaction, of a G-quadruplex. However, the radius of gyration while providing a description of the maintaining of the folding pattern of a given macrostructure, applies preferentially to globular entities. Figure 7.11 underlines the concept presented here, by comparing the radius of the sphere surrounding the molecule with the radius of gyration, which is smaller than the sphere surrounding the molecule.

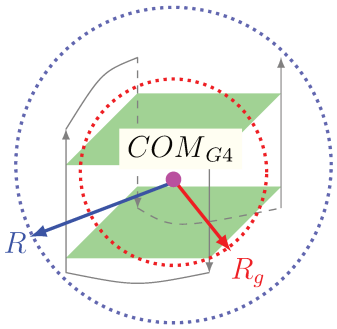


Figure 7.11: The difference between the radius of the sphere (R) enclosing the G-quadruplex and its radius of gyration (R_g).

3

Computational methodology

Simulation methods as defined by computational chemistry are able to provide information that may be experimentally difficult to obtain. This may be due to practical reasons: the difficulty to obtain a crystal, or it may be due to other methodological limitations such as systems too large to be analyzed by NMR. In the work presented here, simulations are used to answer a structural question centered around the impact of damages introduced at specific locations in G-quadruplexes, to understand the structural mechanisms of the interaction of its non-canonical forms with proteins, but also to expose a way to predict the folding type of a G-quadruplex. These objectives were achieved using different computational methods: classical and QM/MM molecular dynamics simulations and DFT calculations.

Molecular dynamics is a method allowing to simulate the temporal evolution of the structure or the reactivity of a molecular system. To do this, it uses models of molecules that are sufficiently complete to be useful, but sufficiently simple to be computable. Two main approaches limit the field of application of theoretical chemistry:

- Simulations using classical molecular mechanics. The equations of motion are numerically integrated for the nuclei of the atoms of the studied system. It relies on the use of *ad hoc* parameters, which depend on the system studied: geometry of the molecule and types of atoms. More precisely, molecular mechanics uses a set of potential functions that describe the energy of the system: these are called the force fields. This type of approach allows the calculation of trajectories of several thousand atoms over long times, of the order of microseconds (μs). Thus this type of simulation is used for the structural analysis of molecules. For example, its use allows to study the interaction between a protein and a ligand, or to observe the structural change of a protein or a nucleic acid.

- Simulations based on the principles of quantum physics: *ab initio*. It focuses on the electronic properties of the molecular system, allowing to simulate the formation and breaking of chemical bonds. Contrary to classic molecular mechanics, this type of simulation does not use any parameters and seeks to solve the Schrödinger equation for the molecular system. The absence of parameterization makes the method applicable to any type of molecular system. However, the use of this approach makes the calculations more expensive and the method is limited to a hundred atoms and to shorter simulation times. For example, this type of simulation is

used for the exploration of chemical reactivity that may take place in an enzyme.

Two other methods are intermediate between the classical method and the (purely) *ab initio* method. These are methods that integrate a certain degree of parameterization: fixed parameters are implemented in the equations to simplify the quantum representation of the molecular system. Consequently, the use of a parameterization lightens the calculations and gives access to longer times. The DFT method can use parametrization and is close to the *ab initio* method. On the contrary, the semi-empirical method diverges a little from the *ab initio* and is closer to the classical approach, while still giving access to chemical reactivity. Because it uses a parametrization depending on the geometry and the type of atom [129].

Table 7.1 below summarizes the properties of the four main methods presented in these paragraphs. But the different aspects of these methods are detailed in the following chapters.

Methods / **Features**	***ab initio***	**DFT**	**semi-empiric**	**empirical (molecular mechanics)**
Period of emergence of the method	1950's (resurgence with first computers)	1990's	1970's	1960's
First popular software using the method	GAUSSIAN 70 (1970)	ADF (1995)	MOPAC (1990)	MM2 (1977)
Theoretical formulation	Quantum (without parametrization)	Quantum (with parametrization)	Quantum (with chemical element by chemical element parameterization)	Classical (parametrization by type of atoms)
Modelled object	Small molecules	Large molecules	Large molecules	Macromolecules
Performance criteria	Representation in adequacy with quantum theory	Electronic prediction	Thermodynamic prediction	Conformational prediction

Table 7.1: Presentation of the characteristics of the four main methods used in theoretical chemistry. *Reproduced from Bensaude-Vincent and Eastes [129], personnal translation.*

CHAPTER 8

Classical molecular dynamics

I can calculate the movement of heavy bodies, but not the madness of crowds.

Sir Isaac Newton

Chapter contents

8.1 Validity of the classical treatment

To simulate a molecular system, it would be more rigorous to solve the Schrödinger equation corresponding to the system under study. But, it is impossible to perform this type of calculation on large systems, like proteins. The solution is to use Newton's law equation to simulate the motion of each nuclei on top of an empirical parametric potential. The evolution of the position of the nuclei over time gives a trajectory. Despite its approximation, this approach is justified by two important properties. First, the Born-Oppenheimer approximation, proposed in 1927 [130], states to the fact that the mass of the nuclei of atoms is much greater than that of the electrons. This makes it possible to consider the motion of the electrons and the motion of the nuclei separately. Based on this principle, molecular dynamics takes into account only the nuclei of atoms. Then, in most chemical systems, the De Broglie wavelength, recalled below, is smaller than the distance that separates an atom from its nearest neighbor. Therefore, nuclear quantum effects are essentially negligible [131, 132]. So, provided that the potential energy is well represented by the empirical potential, classical molecular dynamic is valid for structural evolution, but gives no information over bond breaking and chemical reactivity or excited states.

$$\lambda = \sqrt{\frac{h^2}{Mk_BT}} \tag{8.1}$$

Where M is the atomic mass and T correspond to the temperature, k_B is the Boltzmann constant and h symbolizes the Planck constant.

8.2 The phase space

The equations of molecular dynamics are described in phase space [133]. That is to say, an abstract space with $2f$ dimensions; in which f represents the number of degrees of freedom. More precisely, this space results from the set of physical quantities required to calculate the evolution of the studied system. In the framework of classical dynamics presented here, the physical quantities are the position and the corresponding

momentum; the phase space is therefore the space of states of the system. It is important to know the values of these quantities at a precise time space t. In this sense, any point in the phase space is specific to a state of the system under study [134].

Stöcker et al. [134] explains that a trajectory which is carried out in the space of phase re-situates the periodic movements which give the temporal evolution of the studied system. This means that the evolution of the system in the course of time is dependent on the position of the system in the phase space at a given instant t. But also that different trajectories cannot cross each other in the phase space. Otherwise it would not be possible to know with precision what is the trajectory followed by the system to arrive at this point.

We recall here that each atom is localized in space by 3 Cartesian coordinates (x,y,z). So, to describe a molecule of N atoms $3N$ coordinates, corresponding to $3N$ degrees of freedom, are necessary. Among these, 3 degrees of freedom correspond to rotational motion and 3 to translational motion of the whole molecule. Such motions are not involved in the vibration of the molecule. Consequently, the vibration modes of a molecule, involving changes in the relative distance and the angle between atoms, are $3N - 6$ (or $3N - 5$ for linear molecules) [131].

8.3 Equations of classical molecular dynamics

8.3.1 Newton's equations

Hamilton's equations are used to describe the motions of each atoms (i) of a molecule along a trajectory.

$$\begin{cases} r_i = \dfrac{dH}{dp_i} \\ \\ q_i = -\dfrac{dH}{dr_i} \end{cases} \tag{8.2}$$

The equations involve the momentum (p_i) and the position of the atom (r_i). But they can be reduced to the Newtonian equation of motion:

$$f_i = m_i \times a_i \tag{8.3}$$

In which the acceleration (a_i) is the derivative of the velocity (v_i), which is itself the derivative of the position r_i. So the acceleration is the second derivative of the position. Hence the following statement, where E is the energy of the system with respect to the position of the atom i :

$$\begin{cases} f_i = m_i \dfrac{d^2 r_i}{dt^2} \\ f_i = -\dfrac{dE}{dr_i} \end{cases} \tag{8.4}$$

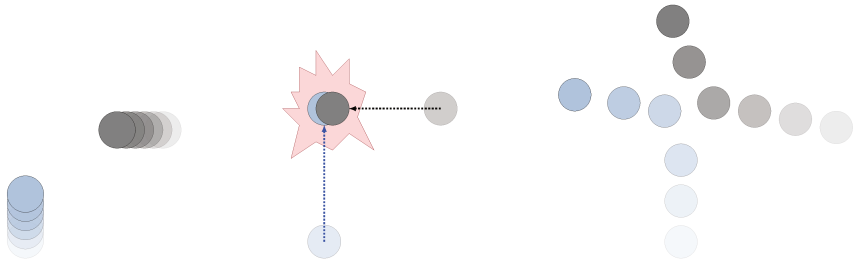
So a molecular dynamics trajectory is performed by the integration of Newton's equations, for which the force (f_i) emanates from the potential defined by the force field. [133].

To finish, it is important to note that in real systems the equations cannot be solved analytically so we need to integrate them numerically. This leads to the definition of a time step.

8.3.2 The integration step

The numerical integration of the equation of motion implies the definition of a discrete time advancement according to a time step (δt). The time step is a crucial parameter for the stability of the dynamics. More precisely, a correct time step must guarantee the conservation of the energy of the system [133, 131]. If the chosen value is too small then the overall computational cost will increase so much that the simulated time will be too short to obtain a correct trajectory, despite a more faithful approximation to the differential equations. But if the chosen value is too large, then there is a considerable increase in the energy of the system. In such a situation, the system is unstable and it is not possible to obtain a physically meaningful trajectory [133, 131]. Figure 8.1 explains these principles in a more visual way. In concrete terms, the choice of time step is made according to the movements with the highest frequency in the system. In fact, it is important that the time step is smaller than the period of these motions: $\delta t \ll v_{movements\ max}$. In molecular dynamics the motions with the highest frequencies are the vibrations of covalent bonds involving hydrogens. The choice of a time step of 1 fs

is therefore recommended [131, 132].



(a) δt too low **(b)** δt too big **(c)** δt suitable

Figure 8.1: Effect of poor and good integration step selections. *Modified from Fuchs [135].*

The heaviness of molecular dynamics computations often tests the computational speed of computers. The need to respond to the decrease in computation time has led to the development of several approaches whose goal is to increase the time step. We can mention the SHAKE [136] and SEETLE [137] algorithms. SHAKE applies constraints by adding a force on the covalent bonds of hydrogens, for example on a bond of type C—H. This action allows the use of a time step twice as high: it goes from 1 fs to 2 fs. SETTLE instead fixes the geometry of the solvent, water, molecules. Nevertheless, the use of constraints is not the only method to increase the time step. In fact these algorithms can be coupled to the Hydrogen Mass Repartitioning (HMR) and thus increase the time step up to 4 fs. This approach consists in increasing the mass of all the hydrogens, except those of water molecules, from 1.008 to 3.024 a.u. However, the total mass of the system is conserved because the mass added to the hydrogens is removed from the heavy atoms directly bound to them as show in 8.2 Indeed, the increase of the hydrogen mass will induce a slowing down of the higher frequency movements.

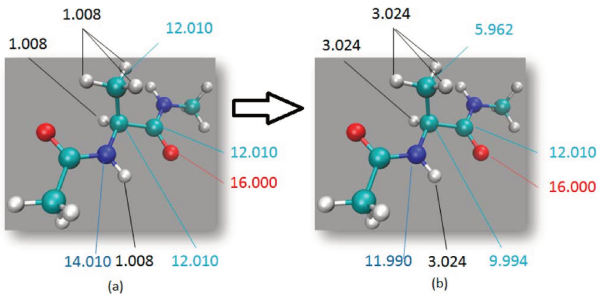


Figure 8.2: Example of the application of an HMR on a peptide. (a) Canonical dialanine peptide atomic masses (b) and after Hydrogen Mass Repartitioning *Reproduced from the article of Hopkins et al. [138]*

8.3.3 Cut-off

The application of the above equations does not require great computational complexity if the system is very small. But most of the time the size of the studied systems

requires to restrict the number of interactions remaining in the system. Doing so allows to reduce the computation time and the memory [131, 132]. The method needed here is called a cut-off. It consists in ignoring the interactions located at an arbitrary distance around a particle i. The application is quite simple. First we build the list of neighbors of the particle, using the $R_{neighbors}$ distance. This list is updated regularly, and thanks to it all the atoms that are located in the cut-off radius are established. Obviously, the cut-off radius R_{cutoff} is smaller than the one used to establish the list of neighbors (see Figure 8.3) [133, 131].

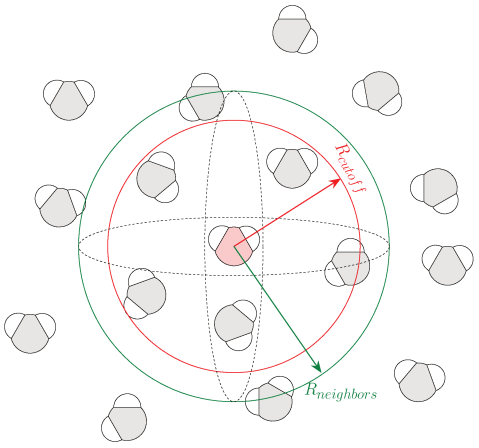


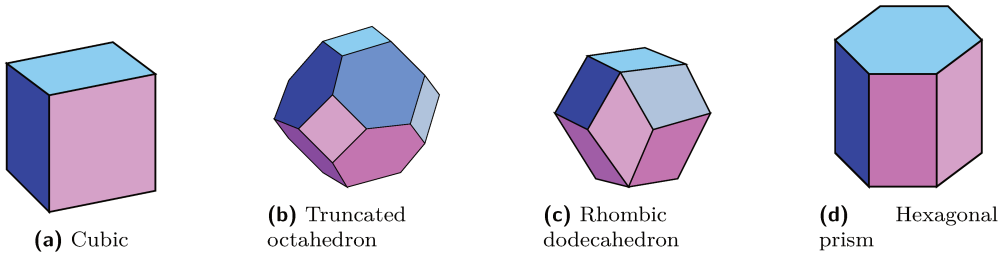
Figure 8.3: The particle i is an oxygen of a water molecule marked in red. The list of its neighbors is established with the $R_{neighbors}$ radius. Then the cutt-off R_{cutoff} is used to calculate the interaction between the particles inside this sphere.

8.4 Solvent, solvation and periodic boundary conditions

8.4.1 Using water as solvent

Water constitutes about 70% of the cells [139]; it is the solvent of biological systems. This is why the simulated molecules are solvated in water. In other words, water molecules are added all around the molecule to be simulated. The set is called a box and it has the shape of a geometrical polyhedron (Figure 8.4); the most common forms are the cube and the truncated octahedron. During the calculations, water is described by a force field parameterized especially for reproducing the properties of this liquid solvent. Some of the most used potentials for water are the TIP3P model (Transferable Intermolecular Potential 3 Points) or the SPC model (Single Point Charge). Bonomi and Camilloni [140] explains that these two models are very similar in that the water

molecule is described by three harmonic potentials, namely: the O——H bond, the H——H bond and the valence angle $\widehat{HOH}$ established between the three atoms.



(a) Cubic **(b)** Truncated octahedron **(c)** Rhombic dodecahedron **(d)** Hexagonal prism

Figure 8.4: The simulation boxes are polyhedra.

8.4.2 Periodic boundary conditions

The use of a simulation box implies the definition of a finite, and usually relatively small, system. However, the use of this type of object is not rigorous as regards the reproduction of thermodynamic quantities and produces edge effects. The solution is to virtually replicate the box in all directions, which produces a pseudo infinite system [133, 131]. However, there is a drastically increase of the interactions in the new system. Consequently, the use of a cut-off becomes necessary to limit the calculation overload. This means that a particle i in the minimal box is able to interact with another particle j in the minimal box, but also with a particle j' in a virtual box, if they are within the cut-off limit [133]. The use of cut-off also avoids unphysical self-interaction of the studied solute with its own image in a nearby box.

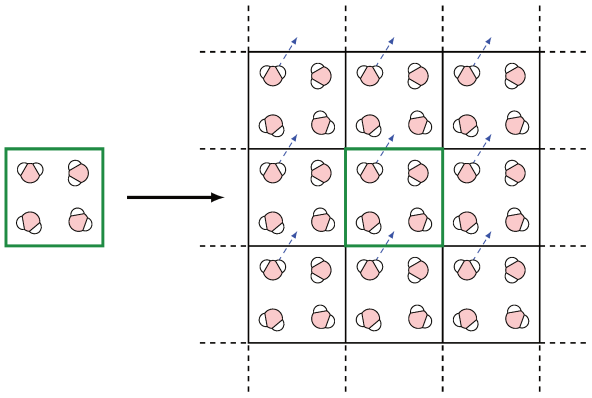
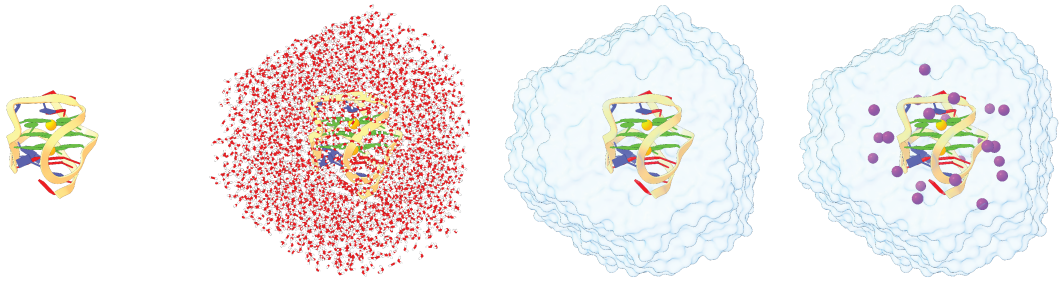


Figure 8.5: Periodic boundary condition of a simulation box.

8.4.3 Solvation and electrical neutrality

Molecular systems are electrically neutral, however this is not always the case with the solvated molecule. In this sense, it is necessary to ensure the electrical neutrality of the calculated system because of periodic boundary conditions. Electrical neutrality is ensured by the addition of ions: some water molecules are replaced by ions, until electrical neutrality is reached. The process of solvation and addition of ions is schematized by the Figure 8.6. Secondly, it goes without saying that the addition of ions is not exclusively limited to charge neutralization. Indeed it is possible to add cations and anions to obtain a certain salt concentration in the simulation box, namely the physiological concentration of 0.15 M.



(a) Molecule preparation **(b)** Solvatation **(c)** **(d)** Addition of ions

Figure 8.6: During the solvation process, water molecules are added around the molecule, while respecting the imposed geometrical shape; then the ions are added.

8.5 Statistical sets

The statistical sets represent thermodynamic constraints imposed by the external environment on the system [131, 132]. The thermodynamic parameters found in the canonical sets are: the number of atoms (N), the pressure (P), the temperature (T), the energy of the system (E), the volume of the box (V), and the chemical potential (μ). Canonical sets are intended to keep some of these parameters constant [131, 140]. Now we can expose the four most frequently encountered sets:

- **NVE** : Here the constant parameters are the number of atoms (N), the volume of the box (V) and the total energy of the system (E). Solving Newton's equation of motion in PBC gives NVE because the equations conserve total energy [131].

- **NVT** : In this set the constraint of constancy applies to the number of atoms (N), the volume (V) and the temperature (T) [131]. To obtain the NVT, it is necessary to apply a thermostat. To obtain the NVT, it is necessary to apply a thermostat. This means that the velocities must be sclaed to maintain the kinetic energy, and thus hence T constant.

- **NPT** : [131, 132] explains that this set is much more representative of the experimental conditions. In this case, the constraint of constancy is applied to the number of atoms (N), the pressure (P) and the the temperature (T). To obtain the NPT, it is necessary to use a barostat. This means that the volume is changed to keep the pressure constant.

CHAPTER 9

Parameterization of molecular force fields

Chapter contents

9.1 What is a molecular force field ?

Force field is a term that has become widely used, also in science fiction and fantasy movies, books and video games. In the latter context, its meaning is defined as a barrier of energy (physical or magical) serving as a passive protection system, such as a shield, intended to repel various forms of danger, or serving as a trap to enclose enemies [141].

In physics, the notion of force field is a portion of space-time in which a constant force is exerted and acts outside of contact on an object located in this field. The magnetic field or the gravitational fields are two examples.

In computational chemistry, the definition of force field is slightly different from that used in physics. In the context that interests us here: a force field is a function that uses the positions of the atoms to describe the variations of the potential energy surface of the molecule [140, 131].

9.2 Overview of empirical potential energy calculations

In molecular dynamics, a molecule is considered as a set of atoms linked together by chemical bonds. Each atom is described as a ball with its own radius, mass and partial charge (or a complete charge for monoatomic ions). Generally, we distinguish two types of interatomic interaction energies: the one between bound atoms and the one between unbound atoms, including also atoms of different molecules (intermolecular interactions). Thus the empirical potential energy ($E(r^N)$) for N particles system, depending on the Cartesian coordinates r_i, is the sum of these two terms [133, 131, 140].

$$E(r^N) = E_{bonded\ atoms} + E_{unbonded\ atoms} \tag{9.1}$$

This description is very minimalist. In fact, each of the two energies is itself the sum of several terms. The energy of bound atoms takes into account the elongation energy of

covalent bonds, as well as bond angle and dihedral angle strain energies. The interaction energy of unbound atoms includes the Van der Waals and the electrostatic components (the latter is also called Coulomb energy) [131, 133]. All these terms are explained below. Before continuing, it is important to understand that the terms of a potential energy function do not always have a straightforward physical or chemical meaning, instead they are meant to reproduce as much as possible the form of the potential energy surface. Some terms have been simply calibrated [133, 140] and their equilibrium values are obtained by comparison with data obtained by experimental measurements or by high level quantum chemical calculations.

9.3 Interaction between bound atoms

9.3.1 Elongation energy of covalent bonds

The elongation term is a parameter that establishes the energy potential necessary to stretch a covalent bond between two atoms. The bond oscillates around an equilibrium value for which the potential energy of the two atoms is minimal (called E_0) [142], see Figure 9.1. The oscillation always takes place in an acceptable zone, i.e. close to the equilibrium value, so that the values of distance (r) and equilibrium distance (r_0) are sufficiently close. The parameter E_{bond} is an analytical function which expresses the variation of the value of the bond distance around its equilibrium value, using a force constant of the bond (k_r). The most used equation is a harmonic function, which comes directly from the second order truncation of the Taylor expansion of the potential. Harmonic approximation is simple and requires a relatively small computation time [133].

$$E_{\text{bond}} = k_r \left(r - r_0\right)^2 \tag{9.2}$$

However, the harmonic function does not perfectly describe the stretching-compression of a chemical bond. In fact, the variation of energy as a function of distance, especially for large deviations from the equilibrium, is more accurately reproduced by the Morse dissociate potential function. We can visually compare the harmonic and the Morse functions in Figure 9.1. The Morse function should be used when theoretical research requires a more accurate level of calculation [133, 131]. But only harmonic terms are

used in the methodology of this work.

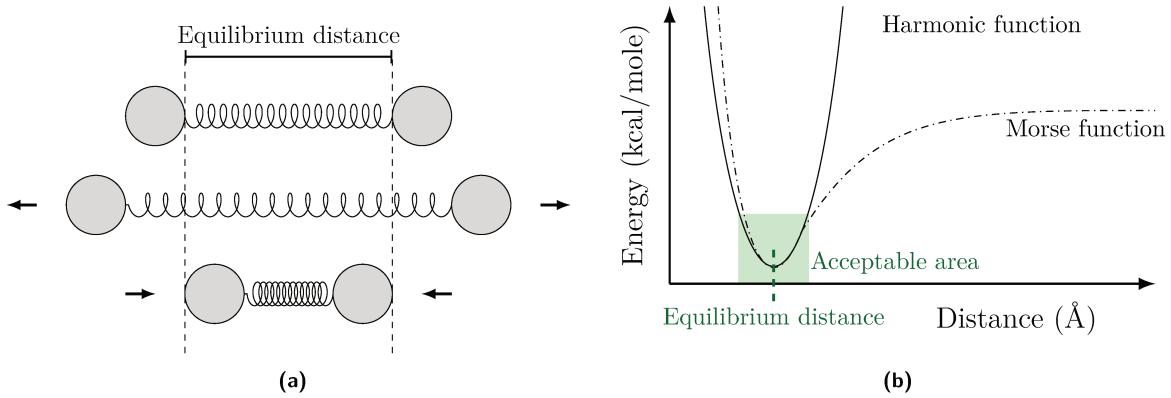


Figure 9.1: **(a)** A covalent bond can be considered as a sprong and the bonded atoms oscillate around an equilibrium distance, corresponding to the E_0 energy value, by attraction-repulsion interactions. **(b)** Comparison between the harmonic and the Morse functions. The harmonic (quadratic) function describes well the bond stretching for small deformations around the equilibrium value.

9.3.2 Angular strain energy

The (E_{angle}) energy describes the bending of the bond angle (θ) around an equilibrium value (θ_0), as shown in Figure 9.2. The relative equation describing this phenomenon is typically a harmonic function that uses a stiffness constant (k_θ). However, also in this case the bond angle bending energy can be improved by using cubic and higher order corrections, as seen for the stretching energy of covalent bonds [131, 132].

$$E_{\text{angle}} = k_\theta \left(\theta - \theta_0\right)^2 \tag{9.3}$$

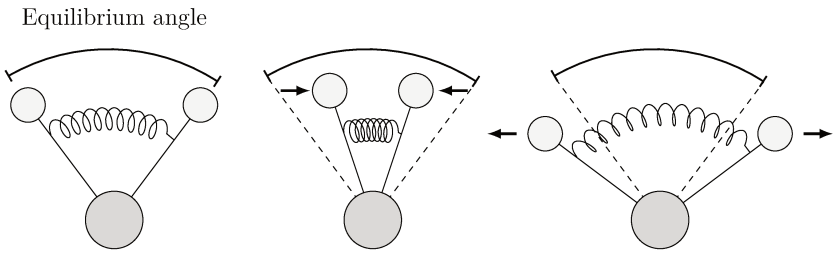


Figure 9.2: The bond angle oscillates around an equilibrium value.

9.3.3 Dihedral angles strain energy

The dihedral angle, or torsion terms, have been presented in the discussion of the structural parameters of G-quadruplexes. This potential isessential for the description of the flexibility of polymer chains. Indeed, it allows the mutual rotation of chemical moieties [140].

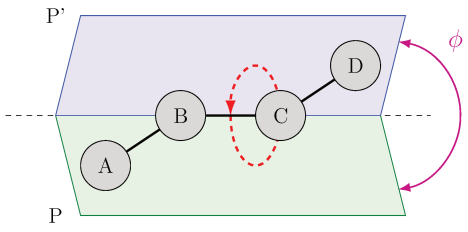


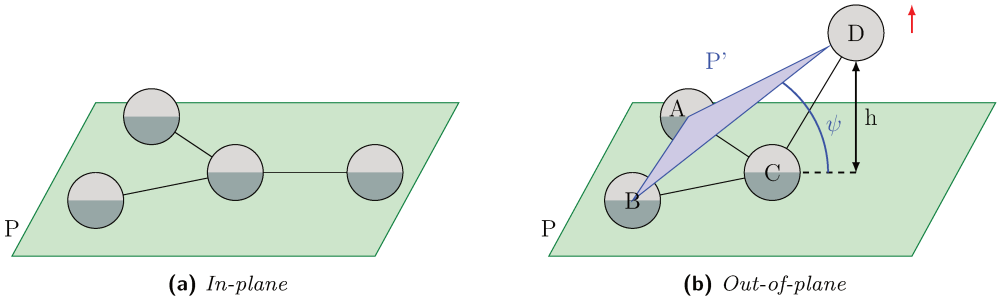
Figure 9.3: Schematic description of distortion of a dihedral angle, ϕ, between planes P and P', formed by the atoms ABC and BCD, respectively. The red dashed arrow symbolizes the rotation of atom D around the axis formed by atoms B and C. *Adapted from Chipot [143].*

The equation describing the potential energy for the rotation around a dihedral angle (E_{dihedral}) is shown below. It is a periodic function containing three parameters, known as i) the "height" of the torsion barrier ($V_n/2$), ii) the periodicity of the rotation – also called the multiplicity – which gives the number of minima that the function encounters during a complete rotation of 360° (n) and iii) the phase of the function (γ).

$$E_{\text{dihedral}} = \sum_n \frac{V_n}{2} \left[1 + \cos\left(n\phi - \gamma\right)\right] \tag{9.4}$$

9.3.4 Improper dihedral angle strain energy

Like the dihedral angle, the improper dihedral (or torsion) angle (ψ) is also established by four atoms, but arranged in a Y shape. In fact, it is essentially used to describe out of plane movement of atoms in coplanar systems, see Figure 9.4. Its usefulness becomes obvious when one wishes to force the coplanarity of the atoms constituting an aromatic ring, for example. Consequently, this potential is used to maintain an improper angle value as close as possible to 0 [133, 140]. Concretely, we consider four atoms ABCD chemically linked, where C is the central atom and D is the mobile atom. If the four atoms lie on the same plane P, then they are said to be coplanar (or in the *In-plane* configuration). But if the mobile atom D is outside the plane P, then the configuration of the four atoms is said to be *Out-of-plane*. The improper angle ψ is formed between the two planes: P and P', determined by atoms ABC and ABD, respectively.
MacKerell et al. [145] gives the following equation to calculate the potential of the improper dihedral angle, where k_ψ is the angular force constant at equilibrium.



(a) *In-plane* **(b)** *Out-of-plane*

Figure 9.4: Schematic description of the *In-plane* (a) and *Out of plane* (b) configurations and of the improper dihedral angle, ψ, between planes P and P', formed by the atoms ABC and ABD, respectively. *Adapted from Malliavin [144]*

$$E_{\text{Improper}} = k_{\psi} \left(\psi - \psi_0\right)^2 \tag{9.5}$$

9.3.5 Cross terms

Cross terms are potential energy functions that couple different energy terms. For example, the *stretch-bend* is a cross term that couples the elongation energy of a covalent bond with the angular strain energy. It is useful because the distance value of a covalent bond is often reduced during the opening of the valence angle. The equation describing the *stretch-bend* cross term is shown below, in which $k_{r\theta}$ is the equilibrium elongation and angular force constant. The use of cross terms, shown in Figure 9.5, allows to obtain a better description of the system dynamics. However their use heavily increases the computation time and furthermore requires often a consequent work of reparameterization of the force field [132, 133].

$$E_{\text{Bond-Angle}} = k_{r\theta} \left(r - r_0\right) \left(\theta - \theta_0\right) \tag{9.6}$$

9.4 Non-bonded interactions

Electrostatic and Van der Waals interactions concern both covalently bonded and non bonded atoms. These interactions can be described as shown in Figure 9.6:

- Intramolecular interactions, if they involve atoms of a molecule.
- Intermolecular interactions, if they involve atoms of different molecules.

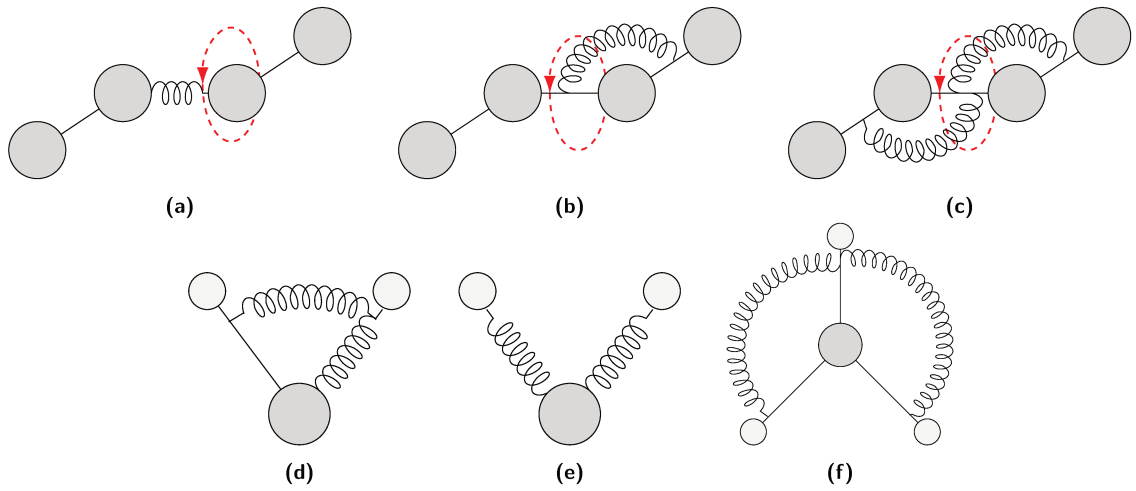


Figure 9.5: Examples of cross terms: stretch-torsion (a), bend-torsion (b), bend-bend-torsion (c), stretch-bend (d), stretch-stretch (e), bend-bend (f). *Modified from Chipot [143].*

The interactions between atoms connected by covalent bonds are described by the bond and angle strain energy terms described above. In this sense, covalently bonded atoms are *type 1-2* connected, atoms forming a bond angle are *type 1-3* connected and atoms forming a dihedral angle are *type 1-4* connected. Atoms with more than *type 1-4* connections can interact only through non-bonding interactions.

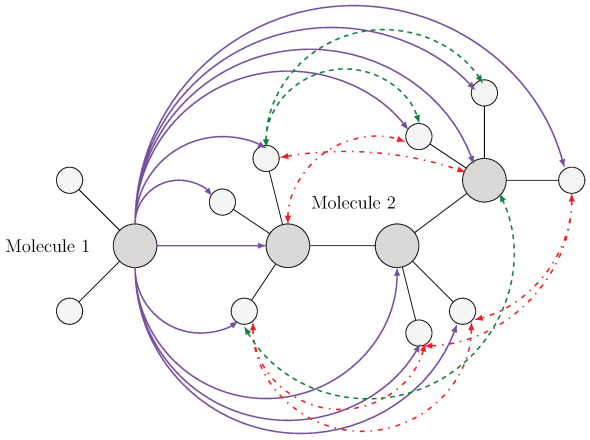


Figure 9.6: Possible interactions between non bonded atoms. The solid purple lines represent the intermolecular interactions; the green dashed lines represent the intramolecular interactions, within Molecule 2, involving atoms with more *type 1-4* connections; the red dash-dotted lines represent the intramolecular interactions, within the Molecule 2, involving atoms with *type 1-4* connections. *Modified from Chipot [143].*

The electrostatic potential energy function ($E_{\text{electrostatic}}$) involves atoms with partial point charges, or ions possessing a formal charge. The function is described by the classical Coulomb law. This involves the partial charges of each atom (q_i and q_j), the distance between them (r_{ij}), the dielectric permittivity of vacuum (ϵ_0) and the effective dielectric constant of the medium (ϵ_1) [133, 131].

$$E_{\text{electrostatic}} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \tag{9.7}$$

Van der Waals interactions take into account the attraction-repulsion occurring between non-bonded atoms. The most common description of the Van der Waals potential energy function is the Lennard-Jones equation, also called *6-12 potential*, as follows:

$$E_{\text{vdW}} = \varepsilon_{ij} \left[\left(\frac{R^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R^*_{ij}}{r_{ij}} \right)^{6} \right] \tag{9.8}$$

Where, R^*_{ij} is the equilibrium distance between the two interacting atoms, and ε_{ij} corresponds to the minimum energy of the function (the depth of the potential well) [133, 131, 132]:

9.5 AMBER force field

The molecular dynamics simulations performed for the research presented in this thesis used the potential energy function ($E(r^N)$) of the AMBER (*Assisted Model Building and Energy Refinement*) suite of biomolecular simulation programs, whose detailed explanations have been provided by Case et al. [146] and Salomon-Ferrer, Case, and Walker [147]. Bonomi and Camilloni [140] has provided a mathematical description of this force field, i.e. of the whole potential energy function $E(r^N)$, in which each term has been discussed previously. A more thorough explanation and discussion of the force field used for proteins and nucleic acids can be found in articles written by Kollman et al. [148], Ponder and Case [149], Cornell et al. [150], Cheatham and Case [151], Maier et al. [152], and Tian et al. [153].

$$\begin{aligned} E(r^N) = & \sum_{\text{bonds}} k_r \left(r - r_0\right)^2 + \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} \sum_{n} \frac{V_n}{2} \left[1 + \cos\left(n\phi - \gamma\right)\right] \\ & + \frac{1}{k^{1-4}_{\text{vdW}}} \sum_{\substack{i<j \\ \{i,j\}\in 1-4}} \varepsilon_{ij} \left[\left(\frac{R^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R^*_{ij}}{r_{ij}} \right)^{6} \right] + \frac{1}{k^{1-4}_{\text{Coulomb}}} \sum_{\substack{i<j \\ \{i,j\}\in 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \\ & + \sum_{\substack{i<j \\ \{i,j\}>1-4}} \varepsilon_{ij} \left[\left(\frac{R^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R^*_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{\substack{i<j \\ \{i,j\}>1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \end{aligned} \tag{9.9}$$

In this equation, there are two terms describing the interaction between non-bonded atoms: one for those separated by three chemical bonds, the so-called 1-4 contribution, and the other for atoms separated by more than three chemical bonds. The 1-4 contributions of the Van der Waal and electrostatic interactions are weighted by $1/k^{1-4}_{\text{vdW}}$ and by $1/k^{1-4}_{\text{Coulomb}}$, respectively.

9.6 *Ab initio* parameterization of AMBER force field for a general molecule.

The underlying physical laws necessary for the mathematical representation of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact solution of these laws leads to equations much too complicated to be soluble.

Paul Dirac – 1929

The AMBER force field contains a large set of parameters for a large number of atoms and is able to describe the structure and dynamical evolution of several biological molecules. However, when considering a new molecule or unit, like a DNA lesions such as a strand breaks or modified nucleotide or small ligands, it is often necessary to introduce new parameters for describing the properties of atoms and the bonds between atoms non present in the force field. This is typically done by performing quantum chemical calculations on the new molecules.

9.6.1 Geometry optimization

The potential energy function $E(r^N)$, also called potential energy (hyper)surface, allows to link the energy of a molecule with its geometry, i.e. with the position of its N nuclei. And whatever the method used to describe this function, the objective of the geometry optimization is always to find minima on the potential energy surface function [154]. In fact, the minimum of the function corresponds to a stable molecular configuration, thus to an equilibrium arrangement of the nuclei. Cramer [133] points out that

geometric optimization is in fact a mathematical problem, and then asks the following question to explain his statement: « How does one find a minimum in an arbitrary function of many variables ? ». In fact, the algorithm acts on a geometrical parameter, also called degrees of freedom (for example, a link or an angle), while keeping the others constant. When changing a geometrical parameter causes an increase in energy, then this indicates that the algorithm is pushing in the wrong direction and needs to be changed. On the contrary, if the modification of the geometrical parameter induces a decrease in energy, then the algorithm is pushing in the right direction. The algorithm will continue in this way until the decrease in energy is no longer possible, and it will explore the next geometrical parameter, until all degrees of freedom are completely explored. This represents an optimization procedure. The objective of the algorithm is to perform a succession of optimization steps until a point is reached where the energy cannot decrease any more [133]. From a mathematical point of view, a potential energy minimum corresponds to a point on the potential energy surface for which the forces are zero and the first derivative of the energy, i.e. the gradient, is also zero. The force constants are second derivatives which can then be used for the harmonic terms in the force field. They are calculated at each step are arranged in a Hessian matrix, which allows to estimate the curvature of the potential energy surface function [154]. For example, the Gaussian [155] software may use an optimization tool based on the Berny algorithm [156, 157, 158].

9.6.2 Assignment of atomic charges

This strategy is specific for Amber force field. Charmm and other force fields have different strategies.

Restrained electrostatic potential (RESP) method allow the calculation of partial atomic charges based on the local electrostatic potential [133], obtained by a charge grid in Gaussian. The RESP method, used in the present thesis work, assigns net charges to the atoms of a molecule, whose sum provides the total net charge [159]. In the present work, it is used through the Antechamber tool of the Amber suite [146], for the parameterization of force fields of organic molecules [160].

9.6.3 Force constants

To calculate bond and bond angle force constants, the easiest way is to use the Parmchk2 tool implemented in the AmberTools suite [146]. Its purpose is to find the force field parameters: geometrical parameters (angles and distances) and the force constant. Parmchk2 is powerful, but it is not always possible to assign parameters. In this case there are several solutions. It is possible to search manually in the database file used by Parmchk2, for example in the GAFF2 file. Otherwise, it is necessary to do a bibliographic search to find the missing parameters. The last options are to make quantum calculations of vibration frequencies or to use IR spectroscopy on samples containing the new molecule under consideration. The force constant (k) obtained by IR spectroscopy can be deduced from the vibration frequency using Hooke's law [161], where U is the potential energy, ν_{vibr} is the vibration frequency, and m_1, m_2 are the masses of atoms 1 and 2:

$$U(r) = \frac{1}{2}k\,(r - r_{eq})^2 \tag{9.10}$$

$$\nu_{vibr} = \frac{1}{2\pi} \times \sqrt{\frac{k}{\frac{m_1 \times m_2}{m_1 + m_2}}} \tag{9.11}$$

On the other hand, software like Gaussian [155] provides force constants by diagonalizing the hessian matrix and the corresponding vibration frequencies, assign also the corresponding vibration normal modes. [133].

CHAPTER 10

Biased molecular dynamics

Chapter contents

Simulations calculated on the basis of equilibrium molecular dynamics provide information on interactions, or structural transformations, that take place on a microsecond time scale (μs). However, there are events which take place on a higher order time scale. In such a situation, equilibrium molecular dynamics is not properly describing the physical process we want to study. The solution to such a situation is to renounce to the explicit time description of the phenomenon and replace it by a description of thermodynamic properties such as free energy barriers and profiles. To do this a bias is applied to the molecular dynamic simulation to allow the system to overcome the free energy barriers leading to the desired event, knowing the biased potential applied will allow then to reconstruct the free energy profile. The application of this strategy is done by using accelerated sampling methods, which is also known under the name of biased molecular dynamics.

Accelerated sampling methods use one or more collective variables, also named reaction coordinate, to reduce the dimensionality of the free energy surface. Then on obtains free energy profile depending on a reduc ed number of collective variables, which provide the description of the desired physical phenomenon, this is also called the Potential

of Mean Force (PMF). Biased molecular dynamic simulation may be performed using classical potential or on quantum approaches and hybrid. In this chapter I will explain the fundaments of metadynamics, eABF and the meta-eABF methods, which has been used in this work at classical level, and umbrella sampling used at the QM/MM level for the study of reactive events.

10.1 Metadynamics

Every molecular system has an energy landscape that represents the different possible states of this system. Each possible state corresponds to a minimum in the potential energy function of the system. Metadynamics , introduced in 2002 by Laio and Parrinello [162], is used to accelerate the exploration of the energy landscape of a molecular system, To do so, this method accelerates the time scale of rare events by adding a bias potential around one or two collective variables. Concretely, a Gaussian repulsive potential is added to the energy of the system. In doing so, there is a repulsion of the system out of the initial energetic state, towards another one, by passing through a transition state [163]. To explain it differently, the method prevents the system from returning to states already explored by molecular dynamics. Interestingly minus the sum of the repulsive potential converges towards the final free energy potential.

The metadynamics can be represented in the following way: a ball is at the bottom of a large container and to make it go out, we empty sand bags one by one until the height of the added sand makes the ball go out of the container.

The accuracy of the metadynamics is dependent on two things:

- The definition of the height and width of the Gaussians, which determines the accuracy of the method.
- The choice of the collective variables, because a wrong choice can prevent the observation of the phenomenon or lead to a wrong estimation of the free energy[163, 164].

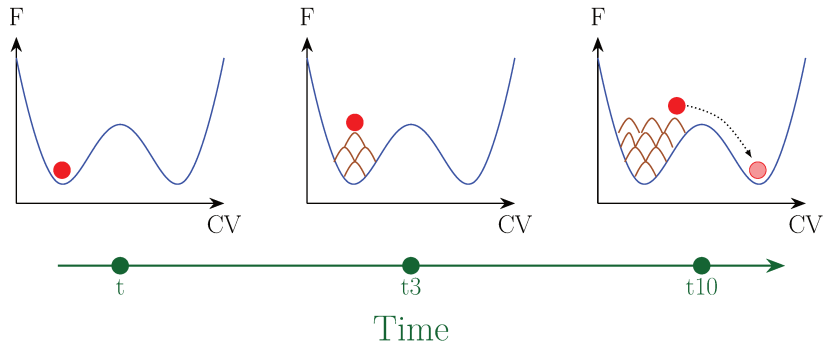


Figure 10.1: In a metadynamics calculation, the gaussians are added gradually over time. Little by little the system is repulsed and goes to another minima by passing through a transition state.

10.2 Difference in free energy between two states

Studies often seek to know the free energy between two states of a molecular system. In biology, this can be the energy required to make a structural change from a Y conformation of the protein to a Z conformation. For example, a protein channel can have two conformations: open or closed. The simplest method to achieve this is to perform a molecular dynamics to generate a set that contains a number of representative conformational configurations in the Y conformation (N_Y) and in the Z conformation (N_Z). Then difference of the free energy is calculated using the equation bellow [165]. This method has the advantage of being very simple. However, we can notice two important points limiting its implementation. First, it is necessary that the Y and Z conformations of the molecular system are not separated by a too large energy barrier. In fact, it is preferable to apply it to molecular systems whose $Y \rightarrow Z$ transition has a small energy barrier. Then, it is necessary that the frequency of presence of each event is sufficient in the set created. i.e. when the value of ΔF_{ZY} is low.

$$\Delta F_{ZY} = -k_B T ln \left[\frac{N_Z}{N_Y}\right] \tag{10.1}$$

10.3 Thermodynamic integration

When the system has a high energy barrier between the Y and Z conformations and/or when their frequency is low in the created set, then a thermodynamic integration can be performed. But, before giving the formula of thermodynamic integration, it is necessary to introduce the basis of the method. The Hamiltonian $H(p,r)$ of the system is also defined as a function of a coupling parameter (λ) between the two conformations of the molecular system: Y and Z. Consequently the Hamiltonian is written as $H(p,r,\lambda)$, where the parameter λ is chosen to correspond to the Hamiltonian of each conformation. This leads to :

- λ_Y corresponds to the Hamiltonian of conformation Y : $H(p,r,\lambda_Y) = H_Y(p,r)$
- λ_Z corresponds to the Hamiltonian of conformation Z : $H(p,r,\lambda_Z) = H_Z(p,r)$

The consequence of the expression of the Hamiltonian with respect to the parameter is that the derivative of of the free energy is also expressed as a function of this parameter. As a result, the free energy formula is written in the following equation [165]:

$$\frac{dF_\lambda}{d\lambda} = \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_i \tag{10.2}$$

In this equation, the system is forced to be confined according to a fixed value of lambda. As a result, it is necessary to calculate the average gradient of the Hamiltonian. This being said, we can explain the principle of thermodynamic integration. It is an empirical way to calculate the energy between state Y and state Z by integrating Equation 10.2 throughout the segment $[\lambda_Z, \lambda_Y]$ [165].

$$F(\lambda_Z) - F(\lambda_Y) = \int_{\lambda_Z}^{\lambda_Y} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_i d\lambda \tag{10.3}$$

The implementation of this method requires the use of a large number of calculation windows, in order to cover the whole energy profile of the molecular system. But this process is very costly in terms of computation time.

10.4 eABF and meta-eABF

To solve the computational time problem, other methods such as eABF and meta-eABF have been proposed. Their objective is to improve the formulation of the thermodynamic integration.

eABF method The Adaptive Biaising Force (ABF) method [166] directly biases the chosen collective variable. It is the origin of the extended Adaptive Biaising Force (eABF) method [167], yet the two methods differ. First, eABF relies on extending the molecular system q by a dummy collective variable λ, which fluctuates around the chosen collective variable ξ. This results in an extended molecular system (q, λ). λ has no physical character, is of mass m_λ, and changes progressively under the same temperature condition as the molecular system under study. Next, it is imperative to approximate the dummy collective variable λ to the chosen collective variable ξ. This step is done simply by coupling the two variables into a harmonic potential, namely : $\frac{k}{2}(\xi(q) - \lambda)^2$. After these explanations we can see that the application of the eABF method on a collective variable ξ, corresponds to the application of an ABF method on an extended collective variable ξ^{ext}. Here, it should be understood that $\xi^{ext}(q, \lambda) \equiv \lambda$ [167]. Finally, the eABF is a potential of mean force (PMF) method which has the advantage of being accurate and whose convergence can be calculated. However, it requires long computation times, as well as the establishment of computation windows.

meta-eABF method Fu et al. [168] propose a new framework to combine the advantages of metadynamics and eABF, in the strategy called meta-eABF. This method requires only a limited number of windows, in some cases just one, and provides accurate and converged results in usually shorter time then eABF. Concretely, the meta-eABF applies the metadynamics to sample the less important regions and applies the eABF on the more important regions.

10.5 Umbrella Sampling

Umbrella sampling is a method proposed by Torrie and Valleau [169] (see Figure 10.2). It varies a harmonic bias potential which is introduced on the coordinate of the path chosen to make the molecular system pass from state Y to state Z. Similar to other methods such as eABF, Umbrella sampling requires the use of several computational windows. But, here, it is very important to have a good overlap of the windows, i.e. between the distributions of the values of the chosen collective variable. Indeed, the sampling is done on the center of each window to provide a distribution set. The distribution sets of each window are biased by the introduced harmonic potential [170]. However, the energy profile of the $Y \rightarrow Z$ transition is found by applying the Weighted Histogram Analysis Method (WHAM) [171]. Concretely, WHAM combines all the windows of calculations and then removes the bias, which is known [172].

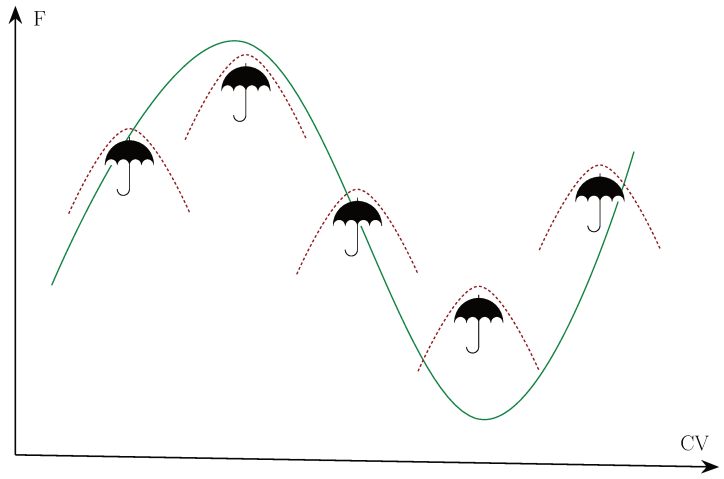


Figure 10.2: Schematic representation of the Umbrella sampling method. Each umbrella represents a calculation window.

CHAPTER 11

Methods for quantum chemical calculations

Chapter contents

Classical molecular dynamics is very adapted to the study of the conformational evolution of large molecular systems. However, this approach is useless when we are interested in enzymatic reactions, because it requires the study of chemical bond formation or breakage. The same is true for studying the spectroscopic properties of a molecule, because it is necessary to describe the electronic transition states. So when this type of study is considered, it is necessary to orient the calculations towards a quantum description of the molecular system, which necessarily involves solving the Schrödinger equation.

11.1 The Schrödinger equation

The classical approach describes atoms by considering only their nuclei while the electrons are treated implicitly via the force field. On the contrary, the quantum approach describes atoms by explicitly considering their nuclei and electrons. The quantum state of a system is a superposition of the configurations that this system can adopt. It is therefore impossible to describe precisely all the aspects of this system. The solution is to describe the system according to its wave function [173], which depends on the coordinates of nuclei and electrons (r). When we look for a stationary state independent of time $\Phi(\vec{r})$, the Schrödinger equation is written in the following terms [173]:

$$\hat{H}\Phi(\vec{r}) = E\Phi(\vec{r}) \tag{11.1}$$

In the time independent formalism, the Hamiltonian of a system composed of N electrons and M nuclei, contains the following terms: the electron kinetic energy ($\hat{T_e}$), the nuclear kinetic energy ($\hat{T_n}$), the interaction term between electrons and nuclei ($\hat{V_{ee}}$), the electrostatic repulsion between the electrons ($\hat{V_{ne}}$), the electrostatic repulsion between the nuclei ($\hat{V_{nn}}$). So the Hamiltonian equation may be written as:

$$\hat{H} = \hat{T_e} + \hat{T_n} + \hat{V_{ee}} + \hat{V_{ne}} + \hat{V_{nn}} \tag{11.2}$$

The terms of the above equation can be decomposed according to: the mass of the nuclei (M), the atomic number of the nuclei (Z), the electrons (i and j), the distance between the electrons r_{ij} , the electron-nucleus distance (r_{iA}) and the distance between two nuclei (r_{AB}), which gives [133]:

$$\hat{H} = -\sum_{i=1}^{N}\frac{\nabla_i^2}{2} - \sum_{A=1}^{M}\frac{\nabla_A^2}{2M_A} - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>1}^{M}\frac{Z_A Z_B}{r_{AB}} \tag{11.3}$$

Solving the time-dependent Schrödinger equation gives the time-dependent wave function $\Psi(\vec{r}, t)$ of a system. In this case the equation is written [173]:

$$\hat{H}\Psi(\vec{r}, t) = i\hbar\frac{\partial}{\partial t}\Psi(\vec{r}, t) \tag{11.4}$$

It is possible to relate the time-independent steady state ($\Phi(\vec{r})$ with the time-dependent wave function ($\Psi(\vec{r},t)$), by the following formula [173]:

$$\hat{H}\Psi(\vec{r},t) = \Phi(\vec{r},t)^{\frac{-iEt}{\hbar}} \tag{11.5}$$

However, the Schrödinger equation cannot be solved analytically because of the correlated motions of particles [133]. That's why several forms of approximations have been proposed.

11.2 The BornOppenheimer Approximation

The motion of the nucleus of an atom is much slower than that of its electrons, so that the electronic relaxation can be considered as instantaneous compared to the motion of the atoms. Thus, it is possible to decouple the two motions. This leads to consider that the nucleus has a fixed position and that the electrons are in motion. The Born-Oppenheimer approximation is therefore to calculate the wave function based on the decoupling of the motion of the nucleus and that of the electrons. By reasoning in this way, the Schrödinger equation is written for a polyelectronic system. It uses several terms: the kinetic energies of the electrons, the attraction between nuclei and electrons, the repulsion between electrons and a constant for a given set of fixed nuclear coordinates (V_n). But in practice V_n is often ignored. Note that by solving the Scrhödinger equation for different nuclear arrangement one explicitly obtains a potential energy surface, whose importance has been described earlier. In conclusion, the Born-Oppenheimer approximation gives the following equation [133]. But, this approximation gives a multivariate N-electron function that remains impossible to solve for polyelectronic atoms

$$\left(-\sum_i \frac{\hbar^2}{2}\nabla_i^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}}\nabla_i^2\right) \Psi\left(r_i, R_A\right) = E\Psi\left(r_i, R_A\right) \tag{11.6}$$

11.3 The Hartree-Fock approximation

Because electrons are fermions it's important to consider the spin-orbital asymmetry when exchanging two particles. So the polyelectron wave function (Ψ) is expressed as a Slater determinant [133, 131]:

$$\Psi\left(\xi_1, \xi_2, \cdots, \xi_i, \xi_j, \cdots, \xi_N\right) = \frac{1}{\sqrt{N!}} \begin{bmatrix} \chi_1(\xi_1) & \chi_2(\xi_1) & \cdots & \chi_N(\xi_1) \\ \chi_1(\xi_2) & \chi_2(\xi_2) & \cdots & \chi_N(\xi_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\xi_N) & \chi_2(\xi_N) & \cdots & \chi_N(\xi_N) \end{bmatrix} \tag{11.7}$$

11.3.1 The Hartree-Fock approximation

The Hartree-Fock approximation proposes to solve the Schrödinger equation by describing separately the interaction of an electron with the nuclei and the mean local electron repulsion field. To do this, this approximation is based on a single determinant, each term of which represents an orbital. In fact, a "pseudo" monoelectronic operator replaces the search for the eigenvalues and eigenfunctions of the Hamiltonian operator. It appears that the Schrödinger polyelectronic equation can be rewritten as a set of equations based on monoelectronic Hamiltonians (h_i), the energy of the i orbital (ϵ_i) and the the set of spin orbitals i (spinorbit) (ξ_i) [133]:

$$h_i \chi_i(\xi_i) = \epsilon_i \chi_i(\xi_i) \tag{11.8}$$

Classically the Hartree-Fock equation is written according to the Fock operator ($\hat{F}$) and the monoelectronic space function ($\varphi_i(\vec{r})$) [174]:

$$\hat{F} \varphi_i(\vec{r}) = \epsilon_i \varphi_i(\vec{r}) \tag{11.9}$$

The Fock operator is the monoelectronic Hamiltonian and it is expressed as follows [174, 133]:

$$\hat{F}_i = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j\geq 1}^{N}\left(2J_j(r_i) - K_j(r_i)\right) \tag{11.10}$$

This writing considers two things. First, the electron Coulomb repulsion, which is expressed through a Coulomb operator (J_j). Secondly, the Pauli principle which is expressed under an exchange operator (K_j) [133, 132]. But this writing remains heavy. Also, it can be rewritten by considering the kernel operator. This is a monoelectronic operator which describes the evolution of an electron in the field of the nuclei. It is calculated by the sum of the kinetic energy and the coulombic attraction energy by the nuclei [131, 132]:

$$\hat{h}_i^c = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} \tag{11.11}$$

The Fock operator is thus written as follows:

$$\hat{F}_i = \hat{h}_i^c + \sum_{i=1}^{N}\sum_{j\geq 1}^{N}\left(2J_j(r_i) - K_j(r_i)\right) \tag{11.12}$$

Finally, it must be pointed out that the Hartree-Fock approximation neglects the electronic correlation. This often leads to a very bad reproduction of the experimental data.

11.3.2 Concept of *atomic orbital basis*

An atomic orbital basis is a linear combination of functions centered on the nuclei of atoms. The functions are associated to the atomic orbitals and can be expressed in two ways. Either by the use of Slater function (Slater Type Orbital), or by the use of Gaussian function (Gaussian Type Orbital). [133, 132, 131].

Solving Schrödinger's equation means to reproduce exactly the molecular orbitals. To do this, it is necessary to use an infinite number of functions. That is to say that it would be necessary to use an infinite basic set. It is obvious that such a thing is technically impossible. This is why it is necessary to introduce approximations. Obviously, the larger the base set, i.e. with a large number of functions, the more precise the calculations will be, but the longer they will take. To answer the demand of this problem, several basis sets have been developed and then implemented in the code of quantum chemistry

software [133], like Gaussian [155]. As an example of base, we can quote 6-31G which uses a combination of six Gaussian for the internal electronic layers, and a combination of three Gaussian and a diffuse function for the valence electronic layers. While the base 6-31G**, more precise, adds 6 primitives of order 2 for the atoms other than hydrogen [133].

11.4 Density functional theory (DFT)

In 1998, the Nobel Prize in Chemistry was awarded to Walter Kohn *for his development of the density-functional theory* and to John A. Pople *for his development of computational methods in quantum chemistry* [175].

DFT is a method which consists in using the number of electrons per unit volume at the point in space with coordinate r, also named electronic density $n(r)$, of a N-electrons system to obtain its energy $E[n(r)]$. DFT method is based on two theorems established by Hohenberg and Kohn:

- In the ground state, the electron density $n_0(r)$ sets the external potential $V[n_0]$. Therefore, the electron density conditions the Hamiltonian of the system. Consequently, all observables of the system resulting from the electron density can be known: such as the wave function Ψ, the type of atom, or the electronic energy of the system E [133, 131].
- The minimization of the total system energy (variational principle) established for the wave function is transposable to the electron density. Thus if the spatial distribution of the electron density of the system is in the ground state, then the energy functional is minimal [133]:

$$\min_{n(r)\to n_0(r)} E[n(r)] = E[n_0(r)] = E_0 \tag{11.13}$$

In the DFT case, the Born-Oppenheimer approximation remains valid. So the Hamiltonian describing each system is written according to the same formula stated previously (Equation 11.3). For an N-electron system, the DFT problem is solved considering a fictious wavefunction representing non interacting particles but giving the exact electronic

density electronic density (Ψ), which determines the state of the system, and spin and space coordinates (τ_i) [131]:

$$n(r) = N \int |\Psi(\tau_1, \tau_2, \cdots, \tau_N))|^2 \, d\tau_r, d\tau_2, \cdots, d\tau_N \tag{11.14}$$

Then, it is important to specify the main characteristics of the electron density:

$$n(r \to \infty) = 0 \tag{11.15}$$

$$\int n(r) dr = N \tag{11.16}$$

The electronic energy of the system E is a density functional and is calculated in the following way, in the framework of DFT:

$$E = E[n_r] \tag{11.17}$$

In the ground state, the energy is a unique functional of n_0.

$$E = E[n_0] \tag{11.18}$$

The energy is expressed according to the kinetic energy ($\hat{T}[n(r)]$) and the electron/electron interaction energy ($\hat{V}_{ee}[n(r)]$), it is the term of repulsion between the electrons, and the electron/nucleus interaction energy ($\hat{V}_{ne}[n(r)]$) [131, 133]:

$$E[n_r] = \hat{T}[n(r)] + \hat{V}_{ee}[n(r)] + \hat{V}_{ne}[n(r)] \tag{11.19}$$

The equation is also written as [133]:

$$E[n_r] = \langle \Psi | \hat{T} + \hat{E}_{ee} | \Psi \rangle + \int dr \; v_{ext}(r) n(r) \tag{11.20}$$

For a system with N-electrons, kinetic energy ($\hat{T}[n(r)]$) and the electron/electron interaction energy ($\hat{V}_{ee}[n(r)]$ or $\hat{E}_{ee}$) are written as [133]:

$$\hat{T} = -\sum_{i=1}^{N} \nabla^2 \tag{11.21}$$

$$\hat{E}_{ee} = -\sum_{i=1}^{N}\sum_{j<1}^{N}\frac{1}{r_{ij}} \tag{11.22}$$

The addition of the two terms T and H gives rise to the universal Hohenberg-Kohn functional ($F_{HK}[n(r)]$). It can be written simply as follows [133]:

$$F[n(r)] = \langle\Psi[n(r)]|\hat{T} + \hat{E}_{ee}|\Psi\rangle \tag{11.23}$$

Which means that the energy is written:

$$E[n_r] = F[n(r)] + \int dr\ v_{ext}(r)n(r) \tag{11.24}$$

However, there is a complication: $F[n(r)]$ is unknown. The result is that $E[n(r)]$ cannot be calculated exactly. Fortunately, it is possible to solve this difficulty thanks to the second Hohenberg-Kohn theorem, which was mentioned above. Here, we simply recall that the ground state energy (E_0) is obtained for the electronic density when the system is at the ground state of ($n_0(r)$).

11.4.1 The Kohn-Sham equations

Kohn and Sham [176] calculate the energy functional of the system using an expression close to the Hartree-Fock, and making use of the fictious wavefunction following the equations (Equation 11.10). The only difference is that the exchange potential, in Hartree-Fock, is replaced by an exchange-correlation potential (V_{xc}) in the Kohn-Sham method [133].

$$\hat{F}_i^{KS} = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \int\frac{n(r')}{|r-r'|}dr' + V_{xc} \tag{11.25}$$

The Kohn-Sham operator ($\hat{F}_i^{KS}$) replaces the Fock operator ($\hat{F}$) in the Hartreee-Fock equationEquation 11.9, which gives:

$$\hat{F}_i^{KS}\varphi_i(\vec{r}) = \epsilon_i\varphi_i(\vec{r}) \tag{11.26}$$

So, in the Kohn-Sham theory the electronic energy of the fundamental state of the system is written [133]:

$$E[n(r)] = -\sum_{i}^{N}\frac{1}{2}\nabla^2 - \sum_{i}^{N}\sum_{k}^{M}\frac{Z_A}{r_{iA}} + \sum_{i}^{N}\frac{1}{2}\int\frac{n(r')}{|r-r'|}dr' + E_{xc}[n(r)] \tag{11.27}$$

Now it is necessary to explain that the exchange-correlation potential (V_{xc}) is the functional derivative of the exchange-correlation energy (E_{xc}), so it is written [133]:

$$V_{xc}[n(r)] = \frac{\delta E_{xc}[n(r)]}{\delta[n(r)]} \tag{11.28}$$

From a formal point of view, the Kohn-Sham equation is exact. However, $E_{xc}[n(r)]$ contains a kinetic contribution. But this poses a consequent problem: it cannot be known. Consequently this term must be approximated, this is the role of the functionals which have been developed for this purpose [133]. There are different families of functionals, some of them are detailed in the next subsection.

11.4.2 The three different families of functions

Local density approximation (LDA). It is the simplest functional. Because it uses the perfect gas model. This requires the assumption of a uniform electron density. The disadvantage of its use is that such an assumption misrepresents molecular and most atomic systems. In fact, it does not adequately reproduce the weak interactions: van der Walls and hydrogen bonds. Therefore, the dissociation energies are overestimated, while the bond lengths are underestimated [133].

Generalized Gradient Approximation (GGA). This type of functionals will introduce information on the $\nabla n(r)$ charge density gradient. This results in a better estimation of the binding energies, compared to the LDA [133].

Hybrid functionals. This type of exchange-correlation functional allows to correct the self interaction error generated in DFT. The introduction of percentage of the Hartree-Fock exact exchange eliminates the repulsion of an electron by itself. We can cite the example of B3LYP which is one of the most used hybrid functional[133].

11.5 Quantum mechanics coupled to molecular mechanics (QM/MM)

In the previous sections, the advantages and limitations of the classical and quantum approaches have been seen. The first one allows to simulate the structural behavior of large systems, the second one is able to simulate the formation and breaking of chemical bonds of smaller systems. But what to do when we want to study the reaction of an enzyme, for example? The solution is to use a hybrid method, named Quantum Mechanics/Molecular Mechanics (QM/MM), which aims at coupling the quantum with the classical approach. To do this, the method separates the system into two regions (Figure 11.1). One corresponding to a small number of atoms on which a quantum approach is performed (QM) to simulate the chemical reaction, while The other one includes all the other atoms on top of which a classical approach is used (MM), because they do not participate to the chemical reaction [177]. So, the advantage of QM/MM is the separation of the MM and QM parts of the system. In other words, the classical approach is performed only on the MM region, keeping the accuracy of the quantum approach on the QM region, which is the site of interest of the chemical reaction.

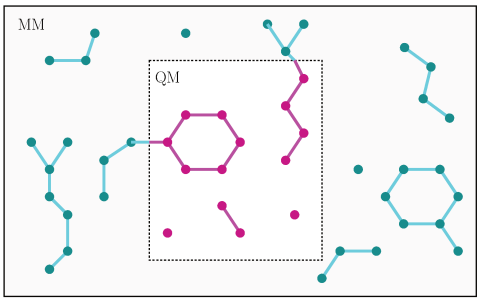


Figure 11.1: Representative diagram of a QM/MM coupling scheme.

In QM/MM the system is characterized by an effective Hamiltonian, which consists of three terms: QM Hamiltonian, MM Hamiltonian and coupling QM/MM Hamiltonian. Depending on the case, the QM/MM Hamiltonian can be added or subtracted to the other two. So the possible equations to obtain the effective Hamiltonian are [133]:

$$\hat{H}_{eff} = \hat{H}_{MM} + \hat{H}_{QM} + \hat{H}_{QM/MM} \tag{11.29}$$

$$\hat{H}_{eff} = \hat{H}_{MM} + \hat{H}_{QM} - \hat{H}_{QM/MM} \tag{11.30}$$

However, QM/MM presents a problem. Since the two parts of the system are treated at different level of theory, it becomes difficult to link the two parts together. This remark is more understandable if we consider a molecule where some atoms are located in the MM part, while the others are located in the QM part. For example, in the case of a nucleotide belonging to a DNA, it is possible to put the base in the QM part and the backbone in the MM part. By doing so, we realize a cut of the molecule at the level of the C — N bond between the sugar and the base. So two atoms involved in the same chemical bond are in the QM and tin the MM region, respectively. Thus, the problem comes from the description of the system at the border between the QM and the MM part [133].

To answer this problem, the simplest solution is to add a monovalent atom, usually hydrogen, to the center of the bond located at the boundary [133] to avoid unpaired electrons or unphysical charge distributions. This is a procedure proposed by M. Karplus. So, if the A — B bond is at the boundary, the application of this solution results in having an A — H in the MM part and an H — B bond in the QM part. Generally, it is better to add the atom in a bond where the A and B atoms are of the same kind, for example a C — C bond.

The use of QM/MM methods was first proposed in 1976 by Warshel and Levitt [177] and was the subject of a Nobel Prize awarded in 2013 to Martin Karplus, Michael Levitt and Arieh Warshel *for the development of multiscale models for complex chemical systems* [178].

4

Modeling and simulating G-quadruplexes

CHAPTER 12

in vitro model validation of G-quadruplexes structure

The study of nucleic acids often requires to know their three-dimensional structure. Several experimental techniques allow to access it, among the other NMR, X-ray crystallography, or crioelecrtonmicroscopy. However, not all structures are known, either because no study has been conducted, or because of experimental constraints. Computational approaches are able to predict the three-dimensional structures. Here, we can quote AlphaFold [179] which uses Deep Learning to predict the structure of proteins from their sequence. Unfortunately there is no software as powerful as this one for G-quadruplexes, and there are comparatively fewer structure solved. For example, a search in the NDB database [180], specifying only *Quadruple Helix* in the *Nucleic Acid Structural Conformation* section, returns 365 DNA and 38 RNA G-quadruplexes for a total of 403 structures. So if the structure of a G-quadruplex is not known, it becomes necessary to use a strategy to determine it. This is what is presented in the following article. More precisely, the research focuses on an *in silico* approach explaining the construction and validation of a model of G-quadruplex structure formed by the RG-1 RNA sequence of the SARS-CoV-2 genome. The model is first constructed using molecular homology and validated using a combination of classical molecular dynamics and QM/MM modeling. The validation strategy is based on the correspondence between the circular dichroism spectrum obtained experimentally by Zhao et al. [58] and the one that has been calculated on top of the structures obtained by our molecular dynamic simulations thanks to QM/MM TD-DFT calculations. In this approach the whole G-quadruplex core has been considered in the QM partition, other ways to calculate a theoretical circular dichroism spectrum of a macromolecular system exist based on Fragment Diabatization-based Excitonic (FrDEx) or Frenkel excitonic model [181].

https://doi.org/10.1021/acs.jpclett.1c0307

CHAPTER 13

Structural stability of G-quadruplexes

Chapter contents

Understanding the mechanisms affecting the stability of G-quadruplexes is, along with their identification, a subject of great interest. Several studies exist and show different approaches to study the structural stability of G-quadruplexes. They have highlighted the importance of the central ion, which is somehow the keystone of G-quadruplexes since its loss leads to unfolding, and that parallel topology is more easily unfolded than the antiparallel conformation [182]. Other studies have focused on the involvement of adenines in the structure of G-quadruplexes. First, the substitution of an adenine in the quartets does not prevent the formation of a G-quadruplex, but there is a position effect. Indeed, replacing a guanine of the central quartet by an adenine causes the formation of a G-quadruplex structure that is not very stable and bimolecular [183]. It has alsop been shown that the position and the number of adenines present in the loops of G-quadruplexes influence the formation and the stability of the folding [184, 185], and can drive the structural conformation of G-quadruplex [186]. Other studies have focused on the influence of loop size on the structure of G-quadurplexes. They reveal that large-size loop may influences the syn-anti orientation of guanines present in quartets [187]. But also that the convention that the maximum length of 7 bases for loops is overrated. Indeed, Guédin et al. [188] have shown that G-quadruplexes can be formed even in the presence of a loop exceeding 12 nucleotides; and that it is still possible to form a

relatively stable G-quadruplex with a 30 nucleotide loop. Further on this topic, loops do not always appear to be destructured as a single flexible strand. This means that the loops can also adopt a secondary structure: hairpin, and participate in the formation of a stable G-quadruplex [189]. Finally, the stability of G-quadruplexes should also be considered in the framework of DNA lesions, and in particular guanine oxidation and strand breaks. These are the main focus of this chapter.

13.1 Introduction of oxidized guanine into G-quadruplex

Guanines represent the keystone of G-quadruplexes, However, guanine is also the most sensitive nucleotide to reactive oxygen species, its most common oxidation product being 8-oxo-7,8-dihydroguanine (8oxoG) [33, 34, 35]. The introduction of this oxidized form of guanines in G-quadruplexes has been studied. It has been shown that the structural impact is dependent on the position of 8oxoG, but that telomeric G-quadruplexes remain globally stable [190]. The following paper presents a multidisciplinary approach using molecular dynamics calculations in combination with circular dichroism experiments and cellular immunofluorescence assays to assess the resistance of G-qudruplexes to oxidative lesions. The calculations show that the conformation of G-quadruplexes remains globally stable when a single lesion is introduced. Also that the introduction of a double lesion, i.e. of two 8oxoG, gives some unfolded structures even if the majority of the structures remain globally stable. These theoretical results are in perfect agreement with the circular dichroism curves which highlight the resistance of G-quadruplexes to an increasing concentration of hydrogen peroxide. Furthermore, immunofluorescence assays reveal a simultaneous increase between the amount of 8oxoG and G-quadruplex when cells are exposed to hydrogen peroxide.

The introduction of more than one 8oxoG in a G-quadruplex may seem unlike to occur in biological conditions since 8oxoG are even more sensitive to oxidation than guanine. So there should not be several 8oxoG in a G-quadruplex. From a certain point of view, an 8oxoG "protect" the structure from oxidation. However, it is quite possible that

oxidation occurs on the DNA strand before it folds into G-quadruplex. In this case, it is possible to have several 8oxoG in a G-quadruplex, especially if they are distant in the sequence.

Personal contribution My involvement in the paper was the part on molecular dynamics simulations, as well as their analysis. I also carried out all circular dichroism measurements.

https://doi.org/10.1002/chem.202100993

13.2 G-quadruplexes resistances to strand break damage

Strand breaks consist into the cleavage of the phosphodiester bond $—O—PO_2—O—$ of DNA backbone. This lesion can be produced by ionizing radiation [27], oxidative stress [191, 192], or enzymes [193]. In fact, there are several types of strand breaks and they are treated by different repair pathways [194]. On this subject, the bacterium *Deinococcus radiodurans* shows a remarkable resistance to DNA damage induced by ionizing radiation. First of all, it possesses a protein named DdrC which is able to colocalize at the level of multiple damage sites located on a single DNA strand [195]. This is not the only resistance strategy employed by the bacterium, as it possesses the ability to repair non-canonical forms of strand breaks ending in $5' - OH/3' - PO_4^-$ ends [196]. Further into the main topic of this thesis, the resistance of *D. radiodurans* also involves the presence of G-quadruplex in its genome [52]. Concerning the impact of ionizing radiation on G-quadruplexes we can point to the work of Kumari et al. [197], in which it is shown that damage is localized in the G-quadrplexes. When a strand break damage takes place in a loop, it causes it to open. But their work highlights the resistance of G-quadruplexes core to ionizing radiation and tends towards the idea that they provide radiation protection. The effect of ionizing radiation and stand breaks on the stability of G-quadruplexes is poorly documented, when one looks for a description at the atomic level. In this context, the following paper proposes a structural study on the effect of introducing one, or more, strand breaks. It is known that this type of lesions can produce two types of canonical $5' - PO_4^-/3' - OH$, and non-canonical $5' - OH/3' - PO_4^-$ ends [28]. Thus using molecular dynamics simulations, the study proposed here shows that telomeric G-quadruplex DNAs are resistant to canonical or non-canonical strand break, even when present at the quartet level

https://doi.org/10.3390/molecules27103256

5

G-quadruplex interactions with proteins

CHAPTER 14

G-quadruplexes can promote the dimerization of proteins.

The article presented here is part of our mobilization at the beginning of the 2020 global pandemic. The work focuses on the mechanism of infection of the SARS-CoV-2 virus. More precisely, the SARS Unique Domain (SUD) is a dimeric protein composed of two symmetrical monomers. It interacts with the RNA of the host cell, also in the form of G-quadruplex [198, 199]. The role of the SUD protein is to inhibit the apoptosis signaling pathway and hence promote cell survival, which leads to increased viral replication and evasion from the immune system [198]. Thus, SUD is an important topic of interest, since understanding the mechanisms of its interaction with G-quadruplexes may provide leads for the design of therapeutic molecules. The research presented below gives a structural explanation at the molecular level of these mechanisms, and in particular the role of G-quadrupolexes in promoting the active dimeric form. The analysis of molecular dynamics simulations reveals two possible modes of interaction. One is an interaction between G-quadruplex and a single monomer of the protein, the other is an interaction between G-quadruplex at the monomer/monomer interface. Furthermore, free energy calculations show that the dimeric interaction favors the active form of the protein.

Personal contribution I created the models of the two starting complexes between G-quadruplex and the SUD protein and performed the molecular dynamics simulations, carrying out a series of structural analysis on the trajectories.

https://dx.doi.org/10.1021/acs.jpclett.0c01097

CHAPTER 15

Specific recognition of *c-Myc* by DARPin 2E4

Several strategies have been developed in order to identify and localize G-quadruplexes *in cellulo*, for analytical or therapeutic purposes. Generally, these involve small molecules that interact with G-quadruplexes [200, 201, 202]. Also, antibodies have been designed for the same purpose [203], however, the design of antibodies is often complex. Another class of protein, named DARPins, has been identified as being able to interact preferentially with G-quadruplexes [204]. DARPins are a class of synthetic protein inspired by natural ankyrin proteins. More precisely, they are small repeating proteins consisting of a succession of ankyrin motifs: two helices and a loop. Each motif is called a module. DARPins have an N-terminal module and a C-terminal module, called caps. Between the caps, there are one or more internal modules that are designed according to the purpose [205]. DARPins have found multiple applications as cell marker, protein cap to facilitate crystalography or as therapeutic agent [206, 205, 207]. In 2014 Scholz, Hansen, and Plückthun [204] performed a study highlighting the recognition specificity of DARPin 2E4 towards the G-quadruplex of the *c-Myc* promoter. The specificity is extremely well defined experimentally, although the exact mechanism of recognition is still not known. Accordingly, the following paper presents an approach to provide an explanation of the mechanisms behind the specific recognition. Our strategy is based on homology modeling, docking and molecular dynamics simulations. By comparing the behavior of DARPin 2E4, specific to *c-Myc*, and that of DARPin 2G10, non-specific, towards three G-quadruplexes: *c-Myc*, *h-Telo* and *Bcl-2*. Analysis shows that the specificity of the 2E4/*c-Myc* interaction arises from the recognition of a structural motif induced by a peculiar folding of the *c-Myc* G-quadruplex.

doi.org/10.1002/chem.202201824

6

Conclusions

It is not enough to strike the ear and occupy the eyes, it is necessary to act on the soul.

Discours sur le style - Georges-Louis Leclerc, comte de Buffon - 1753 – Personal translation

G-quadruplex structure and dynamics represent a rich and varied research topic with potential applications ranging from medicine to biotechnology. These unique nucleic acid structures are the subject of much research and multiple national and international projects have emerged. The aim is not only to understand the biological mechanisms involving G-quadruplexes, but also to use them in innovative technologies. In this perspective, the works presented in this manuscript, realized during three years of thesis, aims at deepening the knowledge of G-quadruplex structural complexity and the related properties. In particular, we focus on different aspects of G-quadruplex, ranging from understanding their structural stability to their interactions with proteins.

The results obtained revealed that the use of multi-approach strategy allows to propose sounding models of G-quadruplex. In details, a sequence homology search allows to find the experimentally resolved structure of a G-quadruplex with similarity with the target. The latter is then used as a basis to reconstruct the G-quadruplex structure of interest. This model is then subjected to molecular dynamics simulations to find its conformational landscape. Finally, QM/MM calculations can be performed to obtain the corresponding theoretical circular dichroism spectrum, which can be easily compared with the experimental one. The structural model of G-quadruplex is valid if there is a good matching between the calculated and the experimental spectrum.

8-oxo-guanine or strand breaks have been introduced in the structure of telomeric G-quadruplex, *h-Telo*. These are two types of damage caused by reactive oxygen species and ionizing radiation, respectively. 8-oxo-guanine and strand breaks were introduced according to the same pattern: singularly or in clusters and at different positions. The trajectories resulting from molecular dynamics highlight the high structural stability of G-quadruplexes, even in poresence of lesions. Indeed, the introduction of strand breaks in the loops or in the quartets does not perturb the structural parameters or the global conformlation of the G-quadruplex. The introduction of one 8-oxo-guanine lesion slightly perturbs the structural parameters, but the overall structure remains stable. Only the

introduction of two 8-oxo-guanines can significantly distort the G-quadruplex structure and, in few cases, disrupt it.

The importance of G-quadruplex RNA in the dimerization of the SUD protein of SARS-CoV-2 was also investigated. The analysis of the molecular dynamics evolution shows two possible binding modes. The first involves a single monomer of the protein and G-quadruplex. The other binding mode involves two dimers of the protein for which the G-quadruplex is positioned on both monomers. A free energy calculation on the SUD/G-quadruplex complex shows that the dimeric interaction is the most stable and favors the assembly of the SUD protein in its active conformation. The understanding of this mechanism is very important and provides clues for the development of a possible treatment to fight effectively against the viral infection.

The specific recognition of the G-quadruplex of the c-Myc promoter by the DARPin 2E4 protein was finally. Previous experimental studies highlighted this specificity. However, no structural or sequence explanation has been proposed. As a consequence, the application of molecular docking followed by molecular dynamics simulations on several possible positions pointed out that the specificity of 2E4 towards *c-Myc* is due to the recognition of a structural motif adopted by this G-quadruplex.

Finally, the work presented in this PhD manuscript is part of a research effort enriching the knowledge on G-quadruplexes. It highlights the high resistance of these particular structures, as well as the importance and specificity of their interactions with proteins. Yet the perspectives of this work can be pushed further. Such as establishing the role of alkali metal cations in the stabilization of damaged G-quadruplexes, by calculating their interaction energy. Or, sought to perform similar work on the effect of 8-oxo-guanine and strand breaks can be implemented on G-quadruplex RNA. Thus, it will be possible to verify if there is a difference in stability between DNA and RNA. But it is also possible to consider the rational design of DARPins specific to certain G-quadruplex. This will allow to have an efficient material for the rapid identification or the cellular localization of G-quadruplexes. There is no lack of research avenues concerning this type of non-canonical structure of nucleic acids.

Afterwards, I would like to end this work with this few words from the music *Maintenant je sais* by Jean Gabin :

> [...]
> All my youth, I wanted to say « I know ».
> But, the more I searched, the less I knew.
>
> [...]
> I'm still at my window, I look out, and I wonder ?
>
> Now I know, I know that we never know !
>
> [...]
> That's all I know ! But this I know ... !

Personal translation.

7

Appendices

APPENDIX

Price – Premio Ricerc@STEBICEF

This Call for Papers is aimed at awarding no. 5 prizes of 500.00 euro each, in accordance with the provisions of the current Ricerc@STEBICEF Prize Regulations. Recipients The Call for Proposals is addressed to young PhD students, post-doctoral fellows, grant holders and contract holders who have distinguished themselves for their scientific merits by publishing, in the year 2021 (01.01.2021 - 31.12.2021), as first name or corresponding STEBICEF author, in journals classified in the first or second quartile, according to the Scopus and WOS indices, uploaded and validated on IRIS.

Awards received following the submission of this article:
Miclot T, Hognon C, Bignon E, Terenzi A, Marazzi M, Barone G & Monari A (2021) *Structure and Dynamics of RNA Guanine Quadruplexes in SARS-CoV-2 Genome. Original Strategies against Emerging Viruses.* J Phys Chem Lett 12: 1027710283

Università degli Studi di Palermo

Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche

Direttore

DIPARTIMENTO DI SCIENZE E TECNOLOGIE BIOLOGICHE CHIMICHE E FARMACEUTICHE (STEBICEF)

Nomina Vincitori Premio Ricerc@ STEBICEF 2022

***

IL DIRETTORE

VISTO lo Statuto dell'Università degli Studi di Palermo;

VISTO il D.R. rep. n. 3956/2021, prot.n. 97036 del 06.10.2021, con il quale il Prof. Vincenzo Arizza è stato nominato Direttore del Dipartimento STeBiCeF, per gli anni accademici 2021/2022, 2022/2023, 2023/2024;

VISTO il Regolamento del Premio Ricerc@STEBICEF, emanato con D.R., rep. n.2602 del 18.06.2021, prot. 63742;

VISTO il Bando Premio Ricerc@STEBICEF 2022, Rep. 103 del 13.04.2022, prot. 2957, per l'assegnazione di n.5 premi da 500,00 euro ciascuno, emanato secondo le prescrizioni del succitato vigente Regolamento;

VISTO in particolare l'Art. 3 del Bando, ai sensi del quale, *"un premio verrà assegnato per una pubblicazione interdisciplinare che coinvolga autori del Dipartimento STEBICEF afferenti a due Aree CUN differenti"*;

VISTO il verbale, assunto al prot.n 4797 del 01.06.2022, relativo alla seduta della Commissione AQ-Ricerca, che ha valutato le istanze di partecipazione pervenute e stilato una graduatoria di merito, ai sensi dell'Art. 3 del suddetto Bando;

DECRETA

di approvare la seguente graduatoria di merito:

1. Dott. Salvatore Emanuele Drago;
2. Dott.ssa Daniela Carbone;
3. Dott.ssa Teresa Faddetta;
4. Dott. Tom Miclot;
5. Dott. Gabriele La Monica;
6. Dott. Morena Anthony.

Risultano pertanto vincitori il Dott. Salvatore Emanuele Drago, la Dott.ssa Daniela Carbone, la Dott.ssa Teresa Faddetta, il Dott. Tom Miclot.

Il Direttore

Prof. Vincenzo Arizza

Viale delle Scienze, Ed. 16 - 90128 Palermo - +39 091 238 97111 - +39 091 238 67515
dipartimento.stebicef@unipa.it - dipartimento.stebicef@cert.unipa.it (pec) - https://www.unipa.it/dipartimenti/stebicef

APPENDIX

Interview

We know what we are, but know not what we may be.

Ophelia, in Hamlet, Act 4, Scene 5. William Shakespeare.

At the beginning of my second year of thesis, Dominique Daloz, director of the C2MP doctoral school, gave me the honor to propose me to realize a profile of my academic career and my PhD thesis. Questions was led by Caroline Sobolewski, PhD Communication Officer at the University of Lorraine.

The interview was published in the University of Lorraine's news journal, named *Factuel*. The following page presents a reproduction of this interview (in french), with the web links.

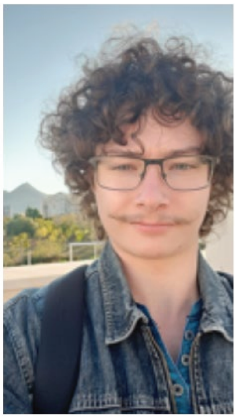
With these few lines, I would like to thank Dominique Daloz and Caroline Sobolewski and convey to them all my consideration.

NOS LABORATOIRES |

[Rencontre] Tom Miclot, doctorant au Laboratoire de physique et chimie théoriques

Publié le 3/11/2020

Tom Miclot, doctorant en deuxième année de thèse, nous présente son parcours et sa thèse réalisée en co-tutelle France-Italie.



Quel est ton parcours ?
Mon parcours d'étude est assez atypique. Après mon Bac technologie en Physique et Chimie de Laboratoire, j'ai choisi de m'orienter vers le BTS Biotechnologies au lycée Stanislas à Villers-Lès-Nancy. C'est lors de ces études que j'ai eu l'opportunité de partir faire un stage de deux mois dans un laboratoire, à l'Université de Limerick (Irlande). Puis, j'ai intégré la Faculté des Sciences de Nancy en 2015 où j'ai obtenu un DEUG et une Licence en Science de la Vie. Ensuite, j'ai continué mon cursus en m'inscrivant au Master Biotechnologies - option Ingénierie Moléculaire. Là, j'ai eu un premier contact avec la recherche théorique grâce à un stage de Master 1 au sein du Laboratoire de Physique et Chimie Théoriques. L'année d'après, j'ai préparé un sujet de mémoire sur les protéines répétées Armadillo, à l'Université de Zurich (Suisse). S'ajoute à cela mon entrée au Master d'Épistémologie, Logique et Histoire des Sciences (MADELHIS) à l'Université des Sciences Humaines et Sociales à Nancy. Je l'ai suivi à distance et parallèlement à la seconde année de mon premier master, ainsi que pendant ma première année de thèse en 2019. Aujourd'hui, je plonge dans ma deuxième année d'une thèse que j'effectue en cotutelle entre le Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche (Università degli Studi di Palermo, Italie), et le Laboratoire de Physique et Chimie Théoriques (FST Nancy, France).

Je suis également étudiant entrepreneur : je suis porteur d'un projet d'entreprise innovante qui est accompagné par le Pôle entrepreneuriat étudiant de Lorraine (PEEL). Nous formons une équipe de cinq personnes et nous travaillons sur une nouvelle technologie visant à éliminer plus facilement les perturbateurs endocriniens présents dans l'eau potable. Bien sûr, ce projet n'est encore qu'à un stade embryonnaire. Vu la portée du sujet, nous espérons le voir se concrétiser d'ici quelque temps.

Sur quelle thématique travailles-tu ?
Mon sujet de thèse porte sur l'étude des lésions des acides nucléiques (ADN et ARN) et de leurs interactions avec des protéines. Afin de mener à bien mes recherches, j'utilise des techniques informatiques qui opèrent comme un « microscope virtuel » permettant d'observer les molécules à l'échelle atomique. Concrètement, j'exploite des structures résolues d'acides nucléiques et de protéines qui proviennent de banques de données telles que la Protein Data Bank ou la Nucleic Acid Database. Ensuite, j'utilise des logiciels pour modifier la structure d'un G-quadruplex (une forme particulière d'ADN à 4 brins) en lui ajoutant des lésions. Puis, j'emploie d'autres outils informatiques capables de donner le comportement des molécules dans de l'eau. En fait, il s'agit de simulations numériques qui modélisent le mouvement des atomes au cours du temps. C'est ce que l'on nomme la dynamique moléculaire. Puis, lorsque cela est nécessaire, je réalise des calculs d'énergie libre sur les modèles issus des dynamiques. Réaliser ce type de calculs permet de connaître les grandeurs d'association entre deux molécules. Par exemple, cela permet de savoir si une protéine s'associe préférentiellement avec un ADN sain, ou avec un ADN lésé.

Pourquoi as-tu décidé de faire une thèse en co-tutelle France-Italie ?
Mes voyages internationaux en Irlande et en Suisse m'ont beaucoup appris, surtout sur la valeur de la Science et le savoir-faire bâti hors de France. Alors, c'est en cosmopolite que j'ai souhaité poursuivre une thèse à l'étranger. D'abord, j'ai pensé à l'Amérique ou à l'Australie. Puis, l'occasion d'une thèse en co-tutelle m'a été présentée par Antonio Monari que je connaissais déjà depuis mon stage de Master 1 sous sa supervision. Je fus donc séduit, car j'y ai vu l'opportunité de prendre part à un projet de recherche s'inscrivant dans une portée européenne et pluridisciplinaire, avec la promesse d'un double diplôme décerné par chaque université.
Outre le haut intérêt scientifique, la thèse en cotutelle est avant tout un échange : elle concrétise l'aspiration justifiée de chercheurs à fonder un projet commun et marque la volonté d'établir et de valoriser une science libre dans une société mondialisée. En fait, la co-tutelle symbolise la détermination de plusieurs acteurs visant à promouvoir l'ouverture de la science, tant à l'échelle locale que dans la dimension internationale.

Que fais-tu en Italie ?
Le LPCT à Nancy est spécialisé en recherche théorique, mais le laboratoire italien dispose d'un équipement expérimental important et me permet d'étudier et de caractériser les structures G-quadruplex. En fait, modèles théoriques et expériences se complètent.

http://factuel.univ-lorraine.fr/node/15443

References

[1] Jocelyn E. Krebs, Elliott S. Goldstein, and Stephen T. Kilpatrick. *Lewin's genes XII*. Burlington, MA: Jones & Bartlett Learning, 2018. ISBN: 9781284104493.

[2] Friedrich Miescher. "Ueber die chemische Zusammensetzung der Eiterzellen". In: *Medicinisch-chemische Untersuchungen: 4*. August Hirschwald, 1871, pp. 441–462. URL: https://books.google.it/books?id=YJRTAAAAcAAJ.

[3] Paul O. P. Ts'o and J. Eisinger. *Basic principles in nucleic acid chemistry*. New York: Academic Press, 1974. ISBN: 9780127019017.

[4] J. D. Watson and F. H. C. Crick. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". In: *Nature* 171.4356 (1953), pp. 737–738. DOI: 10.1038/171737a0.

[5] M. Spencer. "The stereochemistry of deoxyribonucleic acid. II. Hydrogen-bonded pairs of bases". In: *Acta Crystallographica* 12.1 (1959), pp. 66–71. ISSN: 0365-110X. DOI: 10.1107/S0365110X59000160.

[6] K. Hoogsteen. "The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine". In: *Acta Crystallographica* 16.9 (1963), pp. 907–916. ISSN: 0365110X. DOI: 10.1107/S0365110X63002437.

[7] John P. Bartley, Tom Brown, and Andrew N. Lane. "Solution Conformation of an Intramolecular DNA Triplex Containing a Nonnucleotide Linker: Comparison with the DNA Duplex". In: *Biochemistry* 36.47 (1997), pp. 14502–14511. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi970710q.

[8] C. W. Hilbers et al. "The Hairpin Elements of Nucleic Acid Structure: DNA and RNA Folding". In: ed. by Fritz Eckstein and David M. J. Lilley. Vol. 8. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 56–104. ISBN: 9783642786686. DOI: 10.1007/978-3-642-78666-2_4.

[9] A. T. Phan. “Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix”. In: *Nucleic Acids Research* 30.21 (2002), pp. 4618–4625. ISSN: 13624962. DOI: 10.1093/nar/gkf597.

[10] Dipankar Sen and Walter Gilbert. “A sodium-potassium switch in the formation of four-stranded G4-DNA”. In: *Nature* 344.6265 (1990), pp. 410–414. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/344410a0.

[11] R. Garrett and Charles M. Grisham. *Biochemistry*. 4th ed. Belmont, CA: Brooks/Cole, Cengage Learning, 2010. ISBN: 9780495109358.

[12] P. Yakovchuk. “Base-stacking and base-pairing contributions into thermal stability of the DNA double helix”. In: *Nucleic Acids Research* 34.2 (2006), pp. 564–574. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkj454.

[13] James D. Watson, ed. *Molecular biology of the gene*. Seventh edition. Boston: Pearson, 2014. ISBN: 9780321762436.

[14] Takashi Ohyama, ed. *DNA conformation and transcription*. Molecular biology intelligence unit. Georgetown, Tex. : New York, NY: Landes Bioscience ; Springer Science Business Media, 2005. ISBN: 9780387255798.

[15] Romaric Forêt. *Dico de bio*. 3e éd. Bruxelles [Paris]: De Boeck, 2012. ISBN: 9782804171452.

[16] Anthony J. F. Griffiths et al. *Introduction to genetic analysis*. Eleventh edition. New York, NY: W.H. Freeman & Company, a Macmillan Education imprint, 2015. ISBN: 9781464109485.

[17] B Ewald, D Sampath, and W Plunkett. “Nucleoside analogs: molecular mechanisms signaling cell death”. In: *Oncogene* 27.50 (2008), pp. 6522–6537. ISSN: 0950-9232, 1476-5594. DOI: 10.1038/onc.2008.316.

[18] Deepa Sampath, V Ashutosh Rao, and William Plunkett. “Mechanisms of apoptosis induction by nucleoside analogs”. In: *Oncogene* 22.56 (2003), pp. 9063–9074. ISSN: 0950-9232, 1476-5594. DOI: 10.1038/sj.onc.1207229.

[19] Jay P. Parrish et al. “DNA Alkylation Properties of Yatakemycin”. In: *Journal of the American Chemical Society* 125.36 (2003), pp. 10971–10976. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja035984h.

[20] N. Shrivastav, D. Li, and J. M. Essigmann. "Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation". In: *Carcinogenesis* 31.1 (2010), pp. 59–70. ISSN: 0143-3334, 1460-2180. DOI: 10.1093/carcin/bgp262.

[21] Tod Duncan et al. "Reversal of DNA alkylation damage by two human dioxygenases". In: *Proceedings of the National Academy of Sciences* 99.26 (2002), pp. 16660–16665. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.262589799.

[22] Eric Krueger et al. "Modeling and Analysis of Intercalant Effects on Circular DNA Conformation". In: *ACS Nano* 10.9 (2016), pp. 8910–8917. ISSN: 1936-0851, 1936-086X. DOI: 10.1021/acsnano.6b04876.

[23] James E Cleaver. "UV damage DNA repair and skin carcinogenesis". In: *Frontiers in Bioscience* 7.4 (2002), pp. d1024–1043. ISSN: 10939946, 10934715. DOI: 10.2741/A829.

[24] Thierry Douki and Jean Cadet. "Individual Determination of the Yield of the Main UV-Induced Dimeric Pyrimidine Photoproducts in DNA Suggests a High Mutagenicity of CC Photolesions". In: *Biochemistry* 40.8 (2001), pp. 2495–2501. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi0022543.

[25] Jean-Luc Ravanat, Thierry Douki, and Jean Cadet. "Direct and indirect effects of UV radiation on DNA and its components". In: *Journal of Photochemistry and Photobiology B: Biology* 63.1-3 (2001), pp. 88–102. ISSN: 10111344. DOI: 10.1016/S1011-1344(01)00206-8.

[26] H Nikjoo et al. "Radiation track, DNA damage and responsea review". In: *Reports on Progress in Physics* 79.11 (2016), p. 116601. ISSN: 0034-4885, 1361-6633. DOI: 10.1088/0034-4885/79/11/116601.

[27] W D Henner et al. "gamma Ray induced deoxyribonucleic acid strand breaks. 3' Glycolate termini." In: *Journal of Biological Chemistry* 258.2 (1983), pp. 711–713. ISSN: 00219258. DOI: 10.1016/S0021-9258(18)33104-1.

[28] Clemens Von Sonntag et al. "Radiation-Induced Strand Breaks in DNA: Chemical and Enzymatic Analysis of End Groups and Mechanistic Aspects". In: *Advances in Radiation Biology*. Vol. 9. Elsevier, 1981, pp. 109–142. ISBN: 9780120354092. DOI: 10.1016/B978-0-12-035409-2.50009-6.

[29] Tom Miclot et al. “Never Cared for What They Do: High Structural Stability of Guanine-Quadruplexes in the Presence of Strand-Break Damage”. In: *Molecules* 27.10 (2022), p. 3256. ISSN: 1420-3049. DOI: 10.3390/molecules27103256.

[30] Junji Morita et al. “Sequence specific damage of DNA induced by reducing sugars”. In: *Nucleic Acids Research* 13.2 (1985), pp. 449–458. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/13.2.449.

[31] T A Kunkel. “Mutational specificity of depurination.” In: *Proceedings of the National Academy of Sciences* 81.5 (Mar. 1984), pp. 1494–1498. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.81.5.1494.

[32] Bruce K. Duncan and Jeffrey H. Miller. “Mutagenic deamination of cytosine residues in DNA”. In: *Nature* 287.5782 (Oct. 1980), pp. 560–561. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/287560a0.

[33] Steen Steenken and Slobodan V. Jovanovic. “How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution”. In: *Journal of the American Chemical Society* 119.3 (1997), pp. 617–618. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja962255b.

[34] N. R. Jena and P. C. Mishra. “Mechanisms of Formation of 8-Oxoguanine Due To Reactions of One and Two OH Radicals and the H_2O_2 Molecule with Guanine: A Quantum Computational Study”. In: *The Journal of Physical Chemistry B* 109.29 (2005), pp. 14205–14218. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp050646j.

[35] Geneviève Pratviel and Bernard Meunier. “Guanine Oxidation: One- and Two-Electron Reactions”. In: *Chemistry - A European Journal* 12.23 (2006), pp. 6018–6030. ISSN: 0947-6539, 1521-3765. DOI: 10.1002/chem.200600539.

[36] William L. Neeley and John M. Essigmann. “Mechanisms of Formation, Genotoxicity, and Mutation of Guanine Oxidation Products”. In: *Chemical Research in Toxicology* 19.4 (2006), pp. 491–505. ISSN: 0893-228X, 1520-5010. DOI: 10.1021/tx0600043.

[37] R. Wood. “Which DNA polymerases are used for DNA-repair in eukaryotes?” In: *Carcinogenesis* 18.4 (1997), pp. 605–610. ISSN: 14602180. DOI: 10.1093/carcin/18.4.605.

[38] Sara K Martin and Richard D Wood. "DNA polymerase ζ in DNA replication and repair". In: *Nucleic Acids Research* 47.16 (2019), pp. 8348–8361. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz705.

[39] Wen-Jin Wu, Wei Yang, and Ming-Daw Tsai. "How DNA polymerases catalyse replication and repair with contrasting fidelity". In: *Nature Reviews Chemistry* 1.9 (2017), p. 0068. ISSN: 2397-3358. DOI: 10.1038/s41570-017-0068.

[40] T. Izumi and I. Mellon. "Chapter 17 - Base Excision Repair and Nucleotide Excision Repair". In: *Genome Stability*. Ed. by Igor Kovalchuk and Olga Kovalchuk. Boston: Academic Press, 2016, pp. 275–302. ISBN: 9780128033098.

[41] Tae-Hee Lee and Tae-Hong Kang. "DNA Oxidation and Excision Repair Pathways". In: *International Journal of Molecular Sciences* 20.23 (2019), p. 6092. ISSN: 1422-0067. DOI: 10.3390/ijms20236092.

[42] Ralph Scully et al. "DNA double-strand break repair-pathway choice in somatic mammalian cells". In: *Nature Reviews Molecular Cell Biology* 20.11 (2019). ISSN: 1471-0080. DOI: 10.1038/s41580-019-0152-0.

[43] Michael R. Lieber. "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway". In: *Annual Review of Biochemistry* 79 (2010), pp. 181–211. ISSN: 1545-4509. DOI: 10.1146/annurev.biochem.052308.093131.

[44] Nees Jan van Eck and Ludo Waltman. "Software survey: VOSviewer, a computer program for bibliometric mapping". In: *Scientometrics* 84.2 (2010), pp. 523–538. ISSN: 0138-9130, 1588-2861. DOI: 10.1007/s11192-009-0146-3.

[45] Jessica J. King et al. "DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells". In: *Journal of the American Chemical Society* 142.49 (2020), pp. 20600–20604. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.0c11708.

[46] Daniela Rhodes and Hans J. Lipps. "G-quadruplexes and their regulatory roles in biology". In: *Nucleic Acids Research* 43.18 (2015), pp. 8627–8637. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkv862.

[47] Paulina Prorok et al. "Involvement of G-quadruplex regions in mammalian replication origin activity". In: *Nature Communications* 10 (2019), p. 3274. DOI: 10.1038/s41467-019-11104-0.

[48] Katrin Paeschke et al. “Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo”. In: *Nature Structural & Molecular Biology* 12.10 (2005), pp. 847–854. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/nsmb982.

[49] Shankar Balasubramanian, Laurence H. Hurley, and Stephen Neidle. “Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?” In: *Nature Reviews Drug Discovery* 10.4 (2011), pp. 261–275. ISSN: 1474-1776, 1474-1784. DOI: 10.1038/nrd3428.

[50] Sefan Asamitsu et al. “Perspectives for Applying G-Quadruplex Structures in Neurobiology and Neuropharmacology”. In: *Int. J. Mol. Sci.* 20.12 (2019), p. 2884. DOI: 10.3390/ijms20122884.

[51] Norifumi Shioda et al. “Targeting G-quadruplex DNA as cognitive function therapy for ATR-X syndrome”. In: *Nature Medicine* 24.6 (2018), pp. 802–813. ISSN: 1078-8956, 1546-170X. DOI: 10.1038/s41591-018-0018-6. (Visited on 07/15/2022).

[52] Nandhini Saranathan and Perumal Vivekanandan. “G-Quadruplexes: More than just a kink in microbial genomes”. In: *Trends in microbiology* 27.2 (2019), pp. 148–163.

[53] Mathieu Métifiot et al. “G-quadruplexes in viruses: function and potential therapeutic applications”. In: *Nucleic acids research* 42.20 (2014), pp. 12352–12366.

[54] Emanuela Ruggiero and Sara N Richter. “G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy”. In: *Nucleic acids research* 46.7 (2018), pp. 3270–3283.

[55] Jinzhi Tan et al. “The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes”. In: *PLoS pathogens* 5.5 (2009).

[56] Jinzhi Tan et al. “The SARS-unique domain(SUD) of SARS coronavirus is an oligo (G)-binding protein”. In: *Biochemical and biophysical research communications* 364.4 (2007), pp. 877–882.

[57] Yuri Kusov et al. “A G-quadruplex-binding macrodomain within the SARS-unique domain is essential for the activity of the SARS-coronavirus replication–transcription complex”. In: *Virology* 484 (2015), pp. 313–322.

[58] Chuanqi Zhao et al. “Targeting RNA GQuadruplex in SARSCoV2: A Promising Therapeutic Target for COVID19?” In: *Angewandte Chemie* 133.1 (Jan. 2021), pp. 436–442. ISSN: 0044-8249, 1521-3757. DOI: 10.1002/ange.202011419. (Visited on 07/15/2022).

[59] Shruti Mishra et al. “Guanine Quadruplex DNA Regulates Gamma Radiation Response of Genome Functions in the Radioresistant Bacterium *Deinococcus radiodurans*”. In: *Journal of Bacteriology* 201.17 (Sept. 2019). Ed. by Victor J. DiRita. ISSN: 0021-9193, 1098-5530. DOI: 10.1128/JB.00154-19. (Visited on 07/15/2022).

[60] Vikas Yadav et al. “G Quadruplex in Plants: A Ubiquitous Regulatory Element and Its Biological Relevance”. In: *Frontiers in Plant Science* 8 (2017), p. 1163. ISSN: 1664-462X. DOI: 10.3389/fpls.2017.01163.

[61] Xiaofei Yang et al. “RNA G-quadruplex structures exist and function in vivo in plants”. In: *Genome Biology* 21.1 (2020), p. 226. ISSN: 1474-760X. DOI: 10.1186/s13059-020-02142-9.

[62] Emilia Puig Lombardi and Arturo Londoño-Vallejo. “A guide to computational methods for G-quadruplex prediction”. In: *Nucleic Acids Research* 48.1 (2020), pp. 1–15. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz1097.

[63] Chun Kit Kwok and Catherine J. Merrick. “G-Quadruplexes: Prediction, Characterization, and Biological Application”. In: *Trends in Biotechnology* 35.10 (2017), pp. 997–1013. ISSN: 01677799. DOI: 10.1016/j.tibtech.2017.06.012. (Visited on 07/15/2022).

[64] Scott L. Forman et al. “Toward Artificial Ion Channels: A Lipophilic G-Quadruplex”. In: *Journal of the American Chemical Society* 122.17 (2000), pp. 4060–4067. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja9925148.

[65] Lucas T. Gray et al. “G-quadruplexes Sequester Free Heme in Living Cells”. In: *Cell Chemical Biology* 26.12 (2019), 1681–1691.e5. ISSN: 24519456. DOI: 10.1016/j.chembiol.2019.10.003.

[66] Nisreen Shumayrikh and Dipankar Sen. “HemeG-Quadruplex DNAzymes: Conditions for Maximizing Their Peroxidase Activity”. In: *G-Quadruplex Nucleic Acids*. Ed. by Danzhou Yang and Clement Lin. Vol. 2035. New York, NY: Springer New

York, 2019, pp. 357–368. ISBN: 9781493996650. DOI: 10.1007/978-1-4939-9666-7_22.

[67] Prabhpreet Singh et al. "G-quartet type self-assembly of guanine functionalized single-walled carbon nanotubes". In: *Nanoscale* 4.6 (2012), p. 1972. ISSN: 2040-3364, 2040-3372. DOI: 10.1039/c2nr11849a.

[68] Loic Stefan and David Monchaud. "Applications of guanine quartets in nanotechnology and chemical biology". In: *Nature Reviews Chemistry* 3.11 (2019), pp. 650–668. ISSN: 2397-3358. DOI: 10.1038/s41570-019-0132-0.

[69] Jean-Louis Mergny and Dipankar Sen. "DNA Quadruple Helices in Nanotechnology". In: *Chemical Reviews* 119.10 (2019), pp. 6290–6325. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.8b00629.

[70] Donald Voet, Judith G Voet, and Guy G Rousseau. *Biochimie*. Ed. by De Boeck. 2010. ISBN: 9782804147952.

[71] N. B. Leontis. "The non-Watson-Crick base pairs and their associated isosteric-ity matrices". In: *Nucleic Acids Research* 30.16 (2002), pp. 3497–3531. ISSN: 13624962. DOI: 10.1093/nar/gkf481.

[72] Jerry Donohue and Kenneth N. Trueblood. "Base pairing in DNA". In: *Journal of Molecular Biology* 2.6 (1960), pp. 363–371. DOI: 10.1016/S0022-2836(60)80047-2.

[73] Evgenia N. Nikolova et al. "Transient Hoogsteen base pairs in canonical duplex DNA". In: *Nature* 470.7335 (2011), pp. 498–502. DOI: 10.1038/nature09775.

[74] Heidi S. Alvey et al. "Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics". In: *Nature Communications* 5.1 (2014), p. 4786. DOI: 10.1038/ncomms5786.

[75] Maxim D. Frank-Kamenetskii and Sergei M. Mirkin. "Triplex DNA Structures". In: *Annual Review of Biochemistry* 64.1 (1995), pp. 65–95. DOI: 10.1146/annurev.bi.64.070195.000433.

[76] Hala Abou Assi et al. "i-Motif DNA: structural features and significance to cell biology". In: *Nucleic Acids Research* 46.16 (2018), pp. 8038–8056. DOI: 10.1093/nar/gky735.

[77] Martin Gellert, Marie N. Lipsett, and David R. Davies. "HELIX FORMATION BY GUANYLIC ACID". In: *Proceedings of the National Academy of Sciences* 48.12 (1962), pp. 2013–2018. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.48.12.2013.

[78] Stephen Neidle and Shankar Balasubramanian, eds. *Quadruplex nucleic acids*. RSC biomolecular sciences. Cambridge: RSC Pub, 2006. ISBN: 9780854043743.

[79] Jihye Moon et al. "Effects of deficient of the Hoogsteen base-pairs on the G-quadruplex stabilization and binding mode of a cationic porphyrin". In: *Biochemistry and Biophysics Reports* 2 (2015), pp. 29–35. DOI: 10.1016/j.bbrep.2015.03.012.

[80] F. Zaccaria, G. Paragi, and C. Fonseca Guerra. "The role of alkali metal cations in the stabilization of guanine quadruplexes: why K^+ is the best". In: *Physical Chemistry Chemical Physics* 18.31 (2016), pp. 20895–20904. DOI: 10.1039/C6CP01030J.

[81] Besik I. Kankia and Luis A. Marky. "Folding of the Thrombin Aptamer into a G-Quadruplex with Sr^{2+} : Stability, Heat, and Hydration". In: *Journal of the American Chemical Society* 123.44 (2001), pp. 10799–10804. DOI: 10.1021/ja010008o.

[82] Eric Largy, Jean-Louis Mergny, and Valérie Gabelica. "Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability". In: *The Alkali Metal Ions: Their Role for Life*. Ed. by Astrid Sigel, Helmut Sigel, and Roland K. O. Sigel. Vol. 16. Cham: Springer International Publishing, 2016, pp. 203–258. DOI: 10.1007/978-3-319-21756-7_7.

[83] Pascale Hazel et al. "Loop-Length-Dependent Folding of G-Quadruplexes". In: *Journal of the American Chemical Society* 126.50 (2004), pp. 16405–16415. DOI: 10.1021/ja045154j.

[84] M. Clarke Miller et al. "Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3 Sequence of $d(AG_3(TTAG_3)_3)$". In: *Journal of the American Chemical Society* 132.48 (2010), pp. 17105–17107. DOI: 10.1021/ja105259m.

[85] Jonathan B. Chaires and David Graves, eds. *Quadruplex nucleic acids*. Topics in current chemistry 330. Springer Verlag, 2013. ISBN: 9783642347429.

[86] Daisuke Miyoshi, Hisae Karimata, and Naoki Sugimoto. “Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions”. In: *Journal of the American Chemical Society* 128.24 (2006), pp. 7957–7963. DOI: 10.1021/ja061267m.

[87] Mateus Webba da Silva et al. “Design of a G-Quadruplex Topology through Glycosidic Bond Angles”. In: *Angewandte Chemie International Edition* 48.48 (2009), pp. 9167–9170. DOI: 10.1002/anie.200902454.

[88] Sarah Burge et al. “Quadruplex DNA: sequence, topology and structure”. In: *Nucleic Acids Research* 34.19 (2006), pp. 5402–5415. DOI: 10.1093/nar/gkl655.

[89] Markus Meier et al. “Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge”. In: *Nucleic Acids Research* 46.10 (2018), pp. 5319–5331. DOI: 10.1093/nar/gky307.

[90] Vineeth Thachappilly Mukundan and Anh Tuân Phan. “Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences”. In: *Journal of the American Chemical Society* 135.13 (2013), pp. 5017–5028. DOI: 10.1021/ja310251r.

[91] Poulomi Das et al. “Bulges in left-handed G-quadruplexes”. In: *Nucleic Acids Research* 49.3 (2021), pp. 1724–1736. DOI: 10.1093/nar/gkaa1259.

[92] Christopher Jacques Lech, Brahim Heddi, and Anh Tuân Phan. “Guanine base stacking in G-quadruplex nucleic acids”. In: *Nucleic Acids Research* 41.3 (2013), pp. 2034–2046. DOI: 10.1093/nar/gks1110.

[93] Colin A. Ronan. *Histoire mondiale des sciences*. 2. éd. Points Sciences 129. Paris: Éd. du Seuil, 2000. ISBN: 978-2-02-036237-5.

[94] Richard Feynman, Matthew Sands, and Robert B. Leighton. *The Feynman Lectures on Physics Vol. II Ch. 1: Electromagnetism*. Caltech. URL: https://www.feynmanlectures.caltech.edu/II_01.html (visited on 02/16/2021).

[95] Viorel Sergiesco. *CHAMP, physique*. URL: https://www.universalis.fr/encyclopedie/champ-physique/ (visited on 02/16/2021).

[96] Jean-Claude Maurizot. “Particules élémentaires et interactions fondamentales”. In: *Techniques de l’ingénieur*, R6470 V2 (2009).

[97] James Clerk Maxwell. “VIII. A dynamical theory of the electromagnetic field”. In: *Royal Society* 155 (1865). DOI: https://doi.org/10.1098/rstl.1865.0008.

[98] Michel Ney. “Bases de lélectromagnétisme”. In: *Techniques de l’ingénieur*, E1020 V1 (2004).

[99] *CODATA Internationally recommended 2018 values of the Fundamental Physical Constants.*

[100] Richard Feynman, Matthew Sands, and Robert B. Leighton. *The Feynman Lectures on Physics Vol. I Ch. 1: Polarization.* Caltech. URL: https://www.feynmanlectures.caltech.edu/I_33.html (visited on 02/19/2021).

[101] *Dictionnaire de l’Académie française. Dichroïsme.* Académie française. URL: https://www.dictionnaire-academie.fr/article/A9D2405 (visited on 02/24/2021).

[102] M Hamon et al. *Méthodes spectrales et analyse organique.* 2nd ed. Vol. 3. Chimie Analytique. Paris: Masson, 1998. ISBN: 9782225835070.

[103] Gregory Roos and Cathryn Roos. “Isomers and Stereochemistry”. In: *Organic Chemistry Concepts.* Elsevier, 2015, pp. 43–54. ISBN: 9780128016992. DOI: 10.1016/B978-0-12-801699-2.00003-1. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780128016992000031.

[104] *Circular Dichroism Unit Conversion. 4.Ultraviolet/visible spectroscopy.* Biozentrum Biophysics Facility, 2012. URL: https://www.biozentrum.unibas.ch/fileadmin/redaktion/05_Facilities/01_Technology_Platforms/BF/Protocols/CD_Unit_Conversion.pdf.

[105] *Circular Dichroism Spectroscopy Overview.* JASCO Inc. URL: https://jascoinc.com/learning-center/theory/spectroscopy/circular-dichroism-spectroscopy/ (visited on 03/25/2021).

[106] Alison Rodger. “Circular Dichroism Spectroscopy: Units”. In: *Encyclopedia of Biophysics.* Ed. by Gordon C. K. Roberts. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 316–317. ISBN: 9783642167119. DOI: 10.1007/978-3-642-16712-6_647. URL: http://link.springer.com/10.1007/978-3-642-16712-6_647.

[107] *MODEL J-715 Spectrophotometer. Hardware Manual.* P/N:0302-0265A. JASCO Corporation, 1995.

[108] Roberto Corradini et al. “Chirality as a tool in nucleic acid recognition: Principles and relevance in biotechnology and in medicinal chemistry”. en. In: *Chirality* 19.4 (2007), pp. 269–294. DOI: 10.1002/chir.20372.

[109] Walter A. Baase and W.Curtis Johnson. “Circular dichroism and DNA secondary structure”. In: *Nucleic Acids Research* 6.2 (1979), pp. 797–814. DOI: 10.1093/nar/6.2.797.

[110] Cindy A. Sprecher, Walter A. Baase, and W. Curtis Johnson. “Conformation and circular dichroism of DNA”. In: *Biopolymers* 18.4 (1979), pp. 1009–1019. DOI: 10.1002/bip.1979.360180418.

[111] Tomoo Miyahara, Hiroshi Nakatsuji, and Hiroshi Sugiyama. “Helical Structure and Circular Dichroism Spectra of DNA: A Theoretical Study”. en. In: *The Journal of Physical Chemistry A* 117.1 (2013), pp. 42–55. DOI: 10.1021/jp3085556.

[112] Wan Jun Chung et al. “Structure of a left-handed DNA G-quadruplex”. In: *Proceedings of the National Academy of Sciences* 112.9 (2015), pp. 2729–2733. DOI: 10.1073/pnas.1418718112.

[113] Boshi Fu et al. “Right-handed and left-handed G-quadruplexes have the same DNA sequence: distinct conformations induced by an organic small molecule and potassium”. In: *Chemical Communications* 52.65 (2016), pp. 10052–10055. DOI: 10.1039/C6CC04866H.

[114] Antonella Virgilio et al. “Unprecedented right- and left-handed quadruplex structures formed by heterochiral oligodeoxyribonucleotides”. In: *Biochimie* 93.7 (2011), pp. 1193–1196. DOI: 10.1016/j.biochi.2011.04.007.

[115] Fernaldo Richtia Winnerdy et al. “NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes”. In: *Nucleic Acids Research* 47.15 (2019), pp. 8272–8281. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz349.

[116] Wilma K Olson et al. “A standard reference frame for the description of nucleic acid base-pair geometry”. In: *Journal of Molecular Biology* 313.1 (2001), pp. 229–237. DOI: 10.1006/jmbi.2001.4987.

[117] Richard E. Dickerson. “Definitions and Nomenclature of Nucleic Acid Structure Parameters”. In: *Journal of Biomolecular Structure and Dynamics* 6.4 (1989), pp. 627–634. DOI: 10.1080/07391102.1989.10507726.

[118] R. Lavery et al. “Conformational analysis of nucleic acids revisited: Curves+”. In: *Nucleic Acids Research* 37.17 (2009), pp. 5917–5929. DOI: 10.1093/nar/gkp608.

[119] X.-J. Lu. "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures". In: *Nucleic Acids Research* 31.17 (2003), pp. 5108–5121. DOI: 10.1093/nar/gkg680.

[120] Vladimir Tsvetkov, Galina Pozmogova, and Anna Varizhuk. "The systematic approach to describing conformational rearrangements in G-quadruplexes". In: *Journal of Biomolecular Structure and Dynamics* 34.4 (2016), pp. 705–715. DOI: 10.1080/07391102.2015.1055303.

[121] William Humphrey, Andrew Dalke, and Klaus Schulten. "VMD – Visual Molecular Dynamics". In: *Journal of Molecular Graphics* 14 (1996), pp. 33–38.

[122] Roman V. Reshetnikov et al. "Cation binding to 15-TBA quadruplex DNA is a multiple-pathway cation-dependent process". In: *Nucleic Acids Research* 39.22 (2011), pp. 9789–9802. DOI: 10.1093/nar/gkr639.

[123] *Compendium of Chemical Terminology Gold Book*. Version 2.3.3. International Union of Pure and Applied Chemistry, 2014. URL: https://goldbook.iupac.org/files/pdf/goldbook.pdf.

[124] Tomoko Mashimo et al. "Folding Pathways of Human Telomeric Type-1 and Type-2 G-Quadruplex Structures". In: *Journal of the American Chemical Society* 132.42 (2010), pp. 14910–14918. DOI: 10.1021/ja105806u.

[125] Horst Stöcker and Horst Stöcker. *Toutes les mathématiques et les bases de l'informatique*. OCLC: 1033430082. 2005. ISBN: 9782100702633.

[126] Roman Reshetnikov et al. "Structural Dynamics of Thrombin-Binding DNA Aptamer d(GGTTGGTGTGGTTGG) Quadruplex DNA Studied by Large-Scale Explicit Solvent Simulations". In: *Journal of Chemical Theory and Computation* 6.10 (2010), pp. 3003–3014. DOI: 10.1021/ct100253m.

[127] R. V. Reshetnikov, A. V. Golovin, and A. M. Kopylov. "Comparison of models of thrombin-binding 15-mer DNA aptamer by molecular dynamics simulation". In: *Biochemistry (Moscow)* 75.8 (2010), pp. 1017–1024. DOI: 10.1134/S0006297910080109.

[128] Gale Rhodes. *Crystallography made crystal clear: a guide for users of macromolecular models*. 3rd ed. Complementary science series. Amsterdam ; Boston: Elsevier/Academic Press, 2006. ISBN: 9780125870733.

[129] Bernadette Bensaude-Vincent and Richard-Emmanuel Eastes. *Philosophie de la chimie*. Collection Philosophie des sciences. Louvain-la-Neuve Paris: De Boeck supérieur, 2020. ISBN: 9782807305663.

[130] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 389.20 (1927), pp. 457–484. DOI: 10.1002/andp.19273892002.

[131] Andrew R. Leach. *Molecular modelling: principles and applications*. 2nd ed. Harlow, England ; New York: Prentice Hall, 2001. ISBN: 9780582382107.

[132] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. 2nd ed. Computational science series 1. San Diego: Academic Press, 2002. ISBN: 9780122673511.

[133] Christopher J. Cramer. *Essentials of computational chemistry: theories and models*. 2nd ed. Chichester, West Sussex, England ; Hoboken, NJ: Wiley, 2004. ISBN: 9780470091821.

[134] Horst Stöcker et al. *Toute la physique*. 2nd ed. Paris: Dunod, 2007. ISBN: 9782100511815.

[135] P. Fuchs. *Modélisation Moléculaire. Cours3: Méthodes dexploration de la surface dénergie potentielle*. Université Paris Diderot, 2007. URL: https://www.dsimb.inserm.fr/~fuchs/M1BI/Cours3_MD.pdf.

[136] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of Computational Physics* 23.3 (1977), pp. 327–341. ISSN: 00219991. DOI: 10.1016/0021-9991(77)90098-5.

[137] Shuichi Miyamoto and Peter A. Kollman. "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models". In: *Journal of Computational Chemistry* 13.8 (1992), pp. 952–962. ISSN: 1096-987X. DOI: 10.1002/jcc.540130805.

[138] Chad W. Hopkins et al. "Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning". In: *Journal of Chemical Theory and Computation* 11.4 (2015), pp. 1864–1874. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/ct5010406.

[139] Geoffrey M. Cooper. *The cell: a molecular approach*. 2. ed. Washington, DC: ASM Press, 2000. ISBN: 9780878931064.

[140] Massimiliano Bonomi and Carlo Camilloni, eds. *Biomolecular simulations: methods and protocols*. Methods in molecular biology 2022. New York, NY: Humana Press, 2019. ISBN: 9781493996070.

[141] Jeff Prucher, ed. *Brave new words: the Oxford dictionary of science fiction*. Oxford ; New York: Oxford University Press, 2007. ISBN: 9780195305678.

[142] Paul Arnaud. *Chimie générale: les cours de Paul Arnaud*. 8th ed. Paris: Dunod, 2016. ISBN: 9782100743674.

[143] Christophe Chipot. *Les methodes numériques de la dynamique moléculaire*. École de modélisation des macromolécules biologiques, 2002. URL: http://ecole.modelisation.free.fr/cours_chipot.pdf.

[144] Thérèse Malliavin. "Simulations de dynamique moléculaire en biochimie". In: *Techniques de l'ingénieur*, AF6043 V1 (2003).

[145] A. D. MacKerell et al. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins ". In: *The Journal of Physical Chemistry B* 102.18 (1998), pp. 3586–3616. DOI: 10.1021/jp973084f.

[146] David A. Case et al. "The Amber biomolecular simulation programs". In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1668–1688. DOI: 10.1002/jcc.20290.

[147] Romelia Salomon-Ferrer, David A. Case, and Ross C. Walker. "An overview of the Amber biomolecular simulation package: Amber biomolecular simulation package". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.2 (2013), pp. 198–210. ISSN: 17590876. DOI: 10.1002/wcms.1121.

[148] Peter Kollman et al. "The development/application of a minimalist organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data". In: *Computer Simulation of Biomolecular Systems*. Ed. by Wilfred F. van Gunsteren, Paul K. Weiner, and Anthony J. Wilkinson. Dordrecht: Springer Netherlands, 1997, pp. 83–96. DOI: 10.1007/978-94-017-1120-3_2.

[149] Jay W. Ponder and David A. Case. "Force Fields for Protein Simulations". In: *Advances in Protein Chemistry*. Vol. 66. Elsevier, 2003, pp. 27–85. ISBN: 9780120342662. DOI: 10.1016/S0065-3233(03)66002-X.

[150] Wendy D. Cornell et al. “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules”. In: *Journal of the American Chemical Society* 117.19 (1995), pp. 5179–5197. DOI: 10.1021/ja00124a002.

[151] Thomas E. Cheatham and David A. Case. “Twenty-five years of nucleic acid simulations”. In: *Biopolymers* (2013), n/a–n/a. DOI: 10.1002/bip.22331.

[152] James A. Maier et al. “ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB”. In: *Journal of Chemical Theory and Computation* 11.8 (2015), pp. 3696–3713. DOI: 10.1021/acs.jctc.5b00255.

[153] Chuan Tian et al. “ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution”. In: *Journal of Chemical Theory and Computation* 16.1 (2020), pp. 528–552. DOI: 10.1021/acs.jctc.9b00591.

[154] James B. Foresman and AEleen Frisch. *Exploring chemistry with electronic structure methods*. Third edition. Wallingford, CT: Gaussian, Inc, 2015. ISBN: 9781935522034.

[155] M. J. Frisch et al. *Gaussian16 Revision C.01*. Gaussian Inc. Wallingford CT. 2016.

[156] H. Bernhard Schlegel. “Optimization of equilibrium geometries and transition structures”. In: *Journal of Computational Chemistry* 3.2 (1982), pp. 214–218. DOI: 10.1002/jcc.540030212.

[157] H. Bernhard Schlegel. “Estimating the hessian for gradient-type geometry optimizations”. In: *Theoretica Chimica Acta* 66.5 (1984), pp. 333–340. ISSN: 0040-5744, 1432-2234. DOI: 10.1007/BF00554788.

[158] Chunyang Peng et al. “Using redundant internal coordinates to optimize equilibrium geometries and transition states”. In: *Journal of Computational Chemistry* 17.1 (1996), pp. 49–56. DOI: 10.1002/(SICI)1096-987X(19960115)17:1<49::AID-JCC5>3.0.CO;2-0.

[159] Christopher I. Bayly et al. “A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model”. In: *The Journal of Physical Chemistry* 97.40 (1993), pp. 10269–10280. ISSN: 0022-3654, 1541-5740. DOI: 10.1021/j100142a004.

[160] Piotr Cieplak et al. “Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins”. In: *Journal of Computational Chemistry* 16.11 (1995), pp. 1357–1377. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.540161106.

[161] William Kemp. *Organic spectroscopy*. 3rd ed. Palgrave, 1991. ISBN: 9780333519547.

[162] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.202427399.

[163] Giovanni Bussi and Alessandro Laio. “Using metadynamics to explore complex free-energy landscapes”. In: *Nature Reviews Physics* 2.4 (2020), pp. 200–212. ISSN: 2522-5820. DOI: 10.1038/s42254-020-0153-0.

[164] Frank Noé and Cecilia Clementi. “Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods”. In: *Current Opinion in Structural Biology* 43 (2017), pp. 141–147. ISSN: 0959440X. DOI: 10.1016/j.sbi.2017.02.006. (Visited on 08/30/2022).

[165] Conrad Shyu and F. Marty Ytreberg. “Reducing the bias and uncertainty of free energy estimates by using regression to fit thermodynamic integration data”. In: *Journal of Computational Chemistry* (2009), NA–NA. ISSN: 01928651, 1096987X. DOI: 10.1002/jcc.21231.

[166] Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. “Adaptive biasing force method for scalar and vector free energy calculations”. In: *The Journal of Chemical Physics* 128.14 (2008), p. 144120. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.2829861.

[167] Adrien Lesage et al. “Smoothed Biasing Forces Yield Unbiased Free Energies with the Extended-System Adaptive Biasing Force Method”. In: *The Journal of Physical Chemistry B* 121.15 (2017), pp. 3676–3685. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/acs.jpcb.6b10055.

[168] Haohao Fu et al. “Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys”. In: *The Journal of Physical Chemistry Letters* 9.16 (2018), pp. 4738–4745. ISSN: 1948-7185, 1948-7185. DOI: 10.1021/acs.jpclett.8b01994.

[169] G.M. Torrie and J.P. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of Computational Physics* 23.2 (1977), pp. 187–199. ISSN: 00219991. DOI: 10.1016/0021-9991(77)90121-8.

[170] Johannes Kästner. “Umbrella sampling: Umbrella sampling”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.6 (2011), pp. 932–942. ISSN: 17590876. DOI: 10.1002/wcms.66.

[171] Shankar Kumar et al. “THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.540130812.

[172] Benoît Roux. “The calculation of the potential of mean force using computer simulations”. In: *Computer Physics Communications* 91.1-3 (1995), pp. 275–282. ISSN: 00104655. DOI: 10.1016/0010-4655(95)00053-I. (Visited on 09/01/2022).

[173] Jean-Louis Basdevant. *15 leçons de mécanique quantique*. Louvain-la-Neuve (Belgique) Paris: De Boeck supérieur, 2019. ISBN: 9782807321786.

[174] Claude Leforestier. *Introduction à la chimie quantique*. Paris: Dunod, 2014. ISBN: 9782100081998.

[175] *The Nobel Prize in Chemistry 1998*. URL: https://www.nobelprize.org/prizes/chemistry/1998/8811-the-nobel-prize-in-chemistry-1998/ (visited on 09/03/2022).

[176] W. Kohn and L. J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Physical Review* 140.4A (1965), A1133–A1138. ISSN: 0031-899X. DOI: 10.1103/PhysRev.140.A1133.

[177] A. Warshel and M. Levitt. “Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme”. In: *Journal of Molecular Biology* 103.2 (), pp. 227–249. ISSN: 00222836. DOI: 10.1016/0022-2836(76)90311-9.

[178] *The Nobel Prize in Chemistry 2013*. URL: https://www.nobelprize.org/prizes/chemistry/2013/summary/.

[179] John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (2021), pp. 583–589. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-03819-2.

[180] Buvaneswari Coimbatore Narayanan et al. “The Nucleic Acid Database: new features and capabilities”. In: *Nucleic Acids Research* 42.D1 (2014), pp. D114–D122. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkt980.

[181] Hugo Gattuso, Xavier Assfeld, and Antonio Monari. “Modeling DNA electronic circular dichroism by QM/MM methods and Frenkel Hamiltonian”. In: *Theoretical Chemistry Accounts* 134.3 (2015), p. 36. ISSN: 1432-881X, 1432-2234. DOI: 10.1007/s00214-015-1640-8.

[182] Ana Elisa Bergues-Pupo et al. “Role of the central cations in the mechanical unfolding of DNA and RNA G-quadruplexes”. In: *Nucleic Acids Research* 43.15 (2015), pp. 7638–7647. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkv690.

[183] Martin Tomako, Michaela Vorlíková, and Janos Sagi. “Substitution of adenine for guanine in the quadruplex-forming human telomere DNA sequence G3(T2AG3)3”. In: *Biochimie* 91.2 (2009), pp. 171–179. ISSN: 03009084. DOI: 10.1016/j.biochi.2008.07.012.

[184] Brenna A. Tucker et al. “Stability of the Na^+ Form of the Human Telomeric G-Quadruplex: Role of Adenines in Stabilizing G-Quadruplex Structure”. In: *ACS Omega* 3.1 (2018), pp. 844–855. ISSN: 2470-1343, 2470-1343. DOI: 10.1021/acsomega.7b01649.

[185] Phillip A. Rachwal, Tom Brown, and Keith R. Fox. “Sequence effects of single base loops in intramolecular quadruplex DNA”. In: *FEBS Letters* 581.8 (2007), pp. 1657–1660. ISSN: 00145793. DOI: 10.1016/j.febslet.2007.03.040.

[186] Martina Lenari ivkovi, Jan Rozman, and Janez Plavec. “AdenineDriven Structural Switch from a Two to ThreeQuartet DNA GQuadruplex”. In: *Angewandte Chemie International Edition* 57.47 (2018), pp. 15395–15399. ISSN: 1433-7851, 1521-3773. DOI: 10.1002/anie.201809328. (Visited on 07/17/2022).

[187] Beatrice Karg and Klaus Weisz. “Loop Length Affects *Syn* *Anti* Conformational Rearrangements in Parallel GQuadruplexes”. In: *Chemistry A European Journal* 24.40 (2018), pp. 10246–10252. ISSN: 0947-6539, 1521-3765. DOI: 10.1002/chem.201801851.

[188] Aurore Guédin et al. “How long is too long? Effects of loop size on G-quadruplex stability”. In: *Nucleic Acids Research* 38.21 (2010), pp. 7858–7868. ISSN: 1362-4962, 0305-1048. DOI: 10.1093/nar/gkq639.

[189] Subramaniyam Ravichandran et al. “The effect of hairpin loop on the structure and gene expression activity of the long-loop G-quadruplex”. In: *Nucleic Acids Research* 49.18 (2021), pp. 10689–10706. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab739.

[190] Stas Bielskut, Janez Plavec, and Peter Podbevek. “Impact of Oxidative Lesions on the Human Telomeric G-Quadruplex”. In: *Journal of the American Chemical Society* 141.6 (2019), pp. 2594–2603. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.8b12748. (Visited on 07/19/2022).

[191] Natacha Driessens et al. “Hydrogen peroxide induces DNA single- and double-strand breaks in thyroid cells and is therefore a potential mutagen for this organ”. In: *Endocrine-Related Cancer* 16.3 (2009), pp. 845–856. ISSN: 1351-0088, 1479-6821. DOI: 10.1677/ERC-09-0020.

[192] Cynthia J. Burrows and James G. Muller. “Oxidative Nucleobase Modifications Leading to Strand Scission”. In: *Chemical Reviews* 98.3 (1998), pp. 1109–1152. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/cr960421s.

[193] Lucia Lauková et al. “Deoxyribonucleases and Their Applications in Biomedicine”. In: *Biomolecules* 10.7 (2020), p. 1036. ISSN: 2218-273X. DOI: 10.3390/biom10071036.

[194] Keith W. Caldecott. “Single-strand break repair and genetic disease”. In: *Nature Reviews Genetics* 9.8 (2008), pp. 619–631. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg2380.

[195] Anne-Sophie Banneville et al. “Structural and functional characterization of DdrC, a novel DNA damage-induced nucleoid associated protein involved in DNA compaction”. In: *Nucleic Acids Research* 50.13 (2022), pp. 7680–7696. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkac563.

[196] Brad J. Schmier and Stewart Shuman. “Deinococcus radiodurans HD-Pnk, a Nucleic Acid End-Healing Enzyme, Abets Resistance to Killing by Ionizing Radiation and Mitomycin C”. In: *Journal of Bacteriology* 200.17 (2018). Ed. by William W. Metcalf. ISSN: 0021-9193, 1098-5530. DOI: 10.1128/JB.00151-18.

[197] Nitu Kumari et al. “G-quadruplex Structures Contribute to Differential Radiosensitivity of the Human Genome”. In: *iScience* 21 (2019), pp. 288–307. ISSN: 25890042. DOI: 10.1016/j.isci.2019.10.033. (Visited on 07/24/2022).

[198] Jinzhi Tan et al. “The SARS-Unique Domain (SUD) of SARS Coronavirus Contains Two Macrodomains That Bind G-Quadruplexes”. In: *PLoS Pathogens* 5.5 (2009). Ed. by Félix A. Rey, e1000428. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1000428.

[199] Margaret A. Johnson et al. “SARS Coronavirus Unique Domain: Three-Domain Molecular Architecture in Solution and RNA Binding”. In: *Journal of Molecular Biology* 400.4 (2010), pp. 724–742. ISSN: 00222836. DOI: 10.1016/j.jmb.2010.05.027.

[200] V. Dhamodharan and P. I. Pradeepkumar. “Specific Recognition of Promoter G-Quadruplex DNAs by Small Molecule Ligands and Light-up Probes”. In: *ACS Chemical Biology* (2019), acschembio.9b00475. ISSN: 1554-8929, 1554-8937. DOI: 10.1021/acschembio.9b00475.

[201] Debasish Dutta et al. “Cell penetrating thiazole peptides inhibit c-MYC expression via site-specific targeting of c-MYC G-quadruplex”. In: *Nucleic Acids Research* 46.11 (2018), pp. 5355–5365. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gky385.

[202] Joana Figueiredo et al. “Ligands as Stabilizers of G-Quadruplexes in Non-Coding RNAs”. In: *Molecules* 26.20 (2021), p. 6164. ISSN: 1420-3049. DOI: 10.3390/molecules26206164.

[203] Saniya M Javadekar et al. “Characterization of G-quadruplex antibody reveals differential specificity for G4 DNA forms”. In: *DNA Research* 27.5 (2020), dsaa024. ISSN: 1340-2838, 1756-1663. DOI: 10.1093/dnares/dsaa024. (Visited on 07/26/2022).

[204] Oliver Scholz, Simon Hansen, and Andreas Plückthun. “G-quadruplexes are specifically recognized and distinguished by selected designed ankyrin repeat proteins”. In: *Nucleic Acids Research* 42.14 (2014), pp. 9182–9194. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gku571. (Visited on 07/26/2022).

[205] Andreas Plückthun. “Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy”. In: *Annual Review of Pharmacology and Toxicology* 55.1 (2015), pp. 489–511. ISSN: 0362-1642, 1545-4304. DOI: 10.1146/annurev-pharmtox-010611-134654. (Visited on 07/26/2022).

[206] Peer RE Mittl, Patrick Ernst, and Andreas Plückthun. "Chaperone-assisted structure elucidation with DARPins". In: *Current Opinion in Structural Biology* 60 (2020), pp. 93–100. ISSN: 0959440X. DOI: 10.1016/j.sbi.2019.12.009.

[207] Robert C Münch et al. "DARPins: An Efficient Targeting Domain for Lentiviral Vectors". In: *Molecular Therapy* 19.4 (2011), pp. 686–693. ISSN: 15250016. DOI: 10.1038/mt.2010.298.

Abstract in french

La structure et la dynamique des G-quadruplex représentent un sujet de recherche riche et varié dont les applications potentielles vont de la médecine à la biotechnologie. Ces structures uniques d'acides nucléiques font l'objet de nombreuses recherches et de multiples projets nationaux et internationaux ont vu le jour. L'objectif est non seulement de comprendre les mécanismes biologiques impliquant les G-quadruplexes, mais aussi de les utiliser dans des technologies innovantes. Dans cette perspective, les travaux présentés dans ce manuscrit, réalisés au cours de trois années de thèse, visent à approfondir la connaissance de la complexité structurelle des G-quadruplexes et des propriétés associées. En particulier, nous nous concentrons sur différents aspects des G-quadruplex, allant de la compréhension de leur stabilité structurelle à leurs interactions avec les protéines.

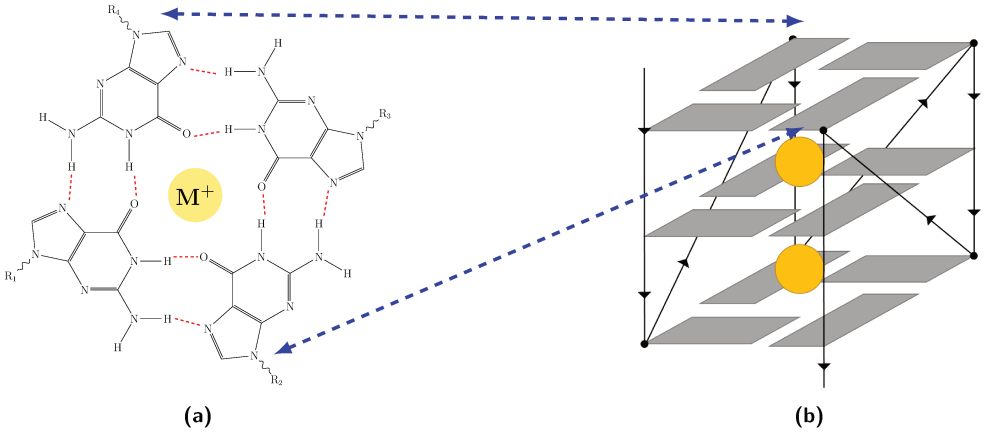


Figure 1: **(a)** Quatret de guanine **(b)** superposé à deux autres quatets dans un arrangement G-quadruplex stabilisé par la présence de cations de métaux alcalins (M^+ représentés par des sphères jaunes).

Les résultats obtenus ont révélé que l'utilisation d'une stratégie multi-approche permet de proposer des modèles solides de G-quadruplex. En détail, une recherche d'homologie de séquence permet de trouver la structure résolue expérimentalement d'un G-quadruplex ayant une similarité avec la cible. Cette dernière est ensuite utilisée comme base pour reconstruire la structure du G-quadruplex d'intérêt. Ce modèle est ensuite soumis à des simulations de dynamique moléculaire pour trouver son paysage conformationnel. Enfin, des calculs QM/MM peuvent être effectués pour obtenir le spectre de dichroïsme circulaire théorique correspondant, qui peut être facilement comparé au spectre expérimental. Le modèle structurel du G-quadruplex est valide s'il y a une bonne correspondance entre le spectre calculé et le spectre expérimental.

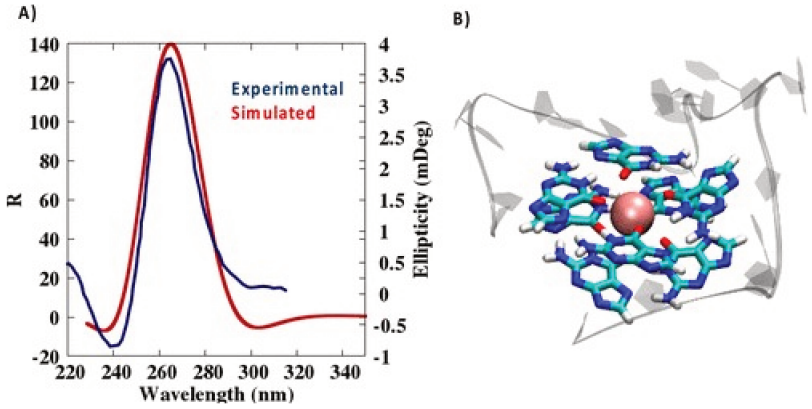


Figure 2: (A) Spectre ECD expérimental, d'après Zhao et al. [58] et simulé de la séquence de l'ARN RG-1. Notez que le spectre simulé a été décalé de manière homogène de 40 nm. Le spectre simulé est obtenu par QM/MM au niveau TD-DFT avec la fonctionnelle M06-2X et la base 6-31G(d), en calculant 60 états excités. (B) La partition QM choisie est mise en évidence dans la représentation de la boule et du bâton, tandis que le squelette G4 et les bases pendantes sont représentés dans le dessin animé.

La 8-oxo-guanine ou les ruptures de brin ont été introduites dans la structure du G-quadruplex télomérique, *h-Telo*. Il s'agit de deux types de dommages causés respectivement par les espèces réactives de l'oxygène et les rayonnements ionisants. La 8-oxo-guanine et les ruptures de brin ont été introduites selon le même schéma : de manière singulière ou en grappes et à des positions différentes. Les trajectoires issues de la dynamique moléculaire mettent en évidence la grande stabilité structurale des G-quadruplexes, même en présence de lésions. En effet, l'introduction de ruptures de brins dans les boucles ou dans les quatuors ne perturbe pas les paramètres structuraux ni la conformation globale du G-quadruplexe. L'introduction d'une lésion de 8-oxo-guanine perturbe légèrement les paramètres structuraux, mais la structure globale reste stable. Seule l'introduction de deux 8-oxo-guanines peut déformer significativement la structure du G-quadruplex et, dans quelques cas, la perturber.

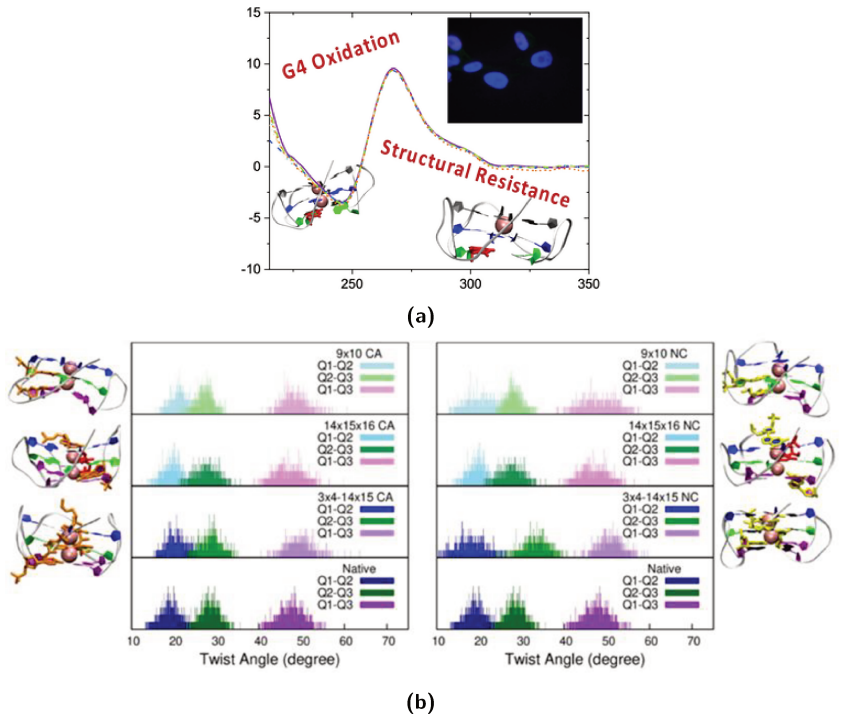


Figure 3: (a) La stabilité structurelle des quadruplexes de guanine en présence de lésions oxydatives est révélée par des simulations de dynamique moléculaire, la spectroscopie CD et l'immunofluorescence. (b) Distribution de l'angle de torsion pour des structures représentatives de l'ADN endommagé (panneaux supérieurs) par rapport à celles de la structure native (panneaux inférieurs).

L'importance de l'ARN G-quadruplex dans la dimérisation de la protéine SUD du SRAS-CoV-2 a également été étudiée. L'analyse de l'évolution de la dynamique moléculaire montre deux modes de liaison possibles. Le premier implique un seul monomère de la protéine et le G-quadruplex. L'autre mode de liaison implique deux dimères de la protéine pour lesquels le G-quadruplex est positionné sur les deux monomères. Un calcul d'énergie libre sur le complexe SUD/G-quadruplex montre que l'interaction dimérique est la plus stable et favorise l'assemblage de la protéine SUD dans sa conformation active. La compréhension de ce mécanisme est très importante et fournit des indices pour le développement d'un éventuel traitement permettant de lutter efficacement contre l'infection virale.

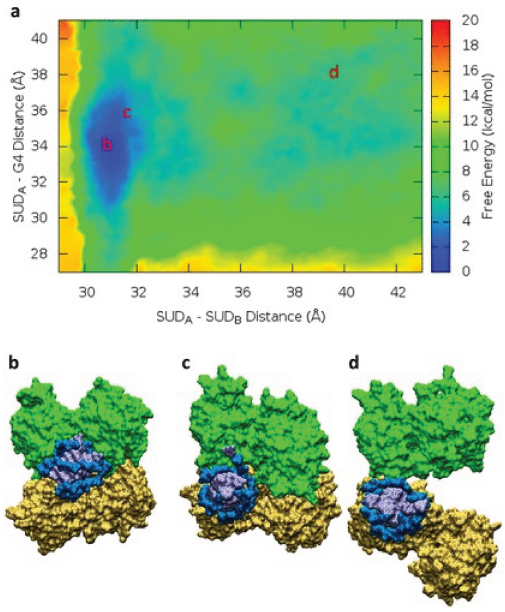


Figure 4: (a) Profil d'énergie libre 2D décrivant l'interaction avec l'ARN et la dimérisation de SUD. Des instantanés représentatifs sont également fournis, décrivant le minimum principal (b), le minimum secondaire (c) et une conformation SUD ouverte (d). La position des instantanés sélectionnés sur la carte PMF est également indiquée en rouge.

Enfin, la reconnaissance spécifique du G-quadruplex du promoteur c-Myc par la protéine DARPin 2E4. Des études expérimentales antérieures ont mis en évidence cette spécificité. Cependant, aucune explication structurelle ou séquentielle n'a été proposée. En conséquence, l'application de docking moléculaire suivi de simulations de dynamique moléculaire sur plusieurs positions possibles a mis en évidence que la spécificité de 2E4 envers le *c-Myc* est due à la reconnaissance d'un motif structural adopté par ce G-quadruplex.

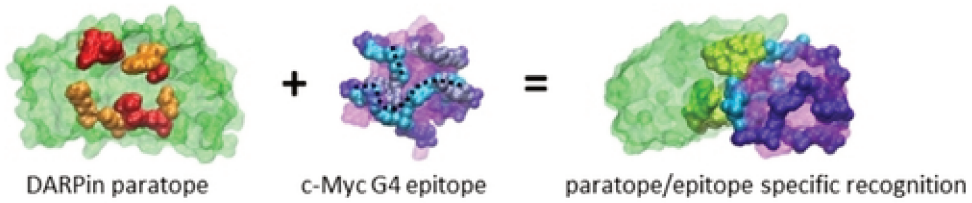


Figure 5: Les DARPins sont des protéines capables de reconnaître spécifiquement les G-quadruplexes. Leur petite taille combinée à leur facilité de conception en fait un bon concurrent des anticorps pour l'identification et la localisation des G-quadruplexes. En utilisant des calculs de dynamique moléculaire, nous montrons que la sélectivité des DARPins envers les G-quadruplexes est obtenue par la reconnaissance d'un motif structurel adopté par le pliage du filament d'ADN.

Enfin, le travail présenté dans ce manuscrit de thèse s'inscrit dans un effort de recherche enrichissant les connaissances sur les G-quadruplexes. Il met en évidence la grande résistance de ces structures particulières, ainsi que l'importance et la spécificité de leurs interactions avec les protéines. Pourtant, les perspectives de ce travail peuvent être poussées plus loin. Par exemple, établir le rôle des cations de métaux alcalins dans la stabilisation des G-quadruplexes endommagés, en calculant leur énergie d'interaction. Ou encore, chercher à réaliser un travail similaire sur l'effet de la 8-oxo-guanine et des ruptures de brins peut être mis en uvre sur l'ARN G-quadruplexe. Ainsi, il sera possible de vérifier s'il existe une différence de stabilité entre l'ADN et l'ARN. Mais il est également possible d'envisager la conception rationnelle de DARPins spécifiques à certains G-quadruplex. Cela permettra de disposer d'un matériel efficace pour l'identification rapide ou la localisation cellulaire des G-quadruplexes. Les pistes de recherche ne manquent pas concernant ce type de structure non canonique des acides nucléiques.

Modeling the influence of DNA lesion on the regulation of gene expression

Tom Miclot

Nucleic acids are organic macromolecules that result from the polymerization of nucleotides. These molecules are generally considered as the support of the genetic information. Two families of nucleic acids are currently known: DNA and RNA. From a structural point of view, the most popular form is the double helix of DNA. However, other forms exist and among them are the G-quadruplex. This is a folding of the DNA, or RNA, in an area rich in guanines. These form quadruplex of guanines, which are stacked on top of each other and are stabilized by a central cation. G-quadruplex structures are increasingly studied. This is not surprising since their biological role involves the regulation of genetic mechanisms. They are notably involved in the regulation of the cell cycle, but they also play a role in cancer, certain neurological or viral diseases. The aim of this PhD thesis is to study G-quadruplex using theoretical chemistry tools. The three years of work raise very important points for the research on G-quadruplex. First, the modeling of a theoretical G-quadruplex structure can be achieved by sequence homology and validated by calculations of a theoretical circular dichroism spectrum. Consequently, it is possible to use these tools to propose and use a G-quadruplex structure if it is not yet experimentally solved. Then, the work done shows that G-quadruplex form a very stable folding since they are globally conserved even when 8-oxo-guanine or strand breaks lesions are introduced at the quartets. Then, the paper focuses on the interaction between G-quadruplex and proteins. It highlights the important role of G-quadruplex RNA in the infection of the viral pathogen SARS-CoV-2. This RNA promotes the dimerization of the SUD protein of the virus, which in turn is responsible for the disruption of the immune system. Finally, this thesis provides a structural explanation for the specific interaction between the DARPin 2E4 protein and the G-quadruplex of the c-Myc promoter.

Keywords G-quadruplex, Interaction DNA-protein, Damaged DNA, Molecular dynamic