



# DNA photosensitization : modeling the interaction between light and complex molecular systems

Hugo Gattuso

## ► To cite this version:

Hugo Gattuso. DNA photosensitization : modeling the interaction between light and complex molecular systems. Theoretical and/or physical chemistry. Université de Lorraine, 2017. English. NNT : 2017LORR0101 . tel-01735331

**HAL Id: tel-01735331**

**<https://theses.hal.science/tel-01735331>**

Submitted on 15 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Thèse présentée en vue de l'obtention du grade de

Docteur de l'Université de Lorraine

Spécialité Chimie

---

**Photosensibilisation de l'ADN: modélisation des  
interactions entre la lumière et les systèmes moléculaires complexes**

**DNA photosensitization: modeling the interaction between  
light and complex molecular systems**

---

**Hugo Gattuso**

Soutenue publiquement le **6 juillet 2017**,  
Faculté des Sciences et Technologies de Vandœuvre-lès-Nancy, France

Membres du jury:

**Rapporteurs:**

Mme. Célia Fonseca Guerra	Professeure	Vrije Universiteit, Amsterdam
Mr. Mathieu Linares	Professeur	Royal Institute of Technology, Stockholm

**Examineurs:**

Mr. Richard Lavery	Directeur de Recherche	CNRS/Université de Lyon I
Mme. Elise Dumont	Maître de conférence	Ecole Normale Supérieure de Lyon
Mr. Antonio Monari	Maître de conférence	Université de Lorraine

**Directeur de thèse:**

Mr. Xavier Assfeld	Professeur	Université de Lorraine
--------------------	------------	------------------------

# Contents

<b>I</b>	<b>Acknowledgments/Remerciements</b>	<b>ii</b>
<b>II</b>	<b>Outreach activities</b>	<b>iv</b>
0.1	Introduction . . . . .	v
0.2	Pretexte . . . . .	v
0.3	Esperlurette . . . . .	vi
0.4	MT180 . . . . .	viii
<b>III</b>	<b>List of publications</b>	<b>xii</b>
<b>IV</b>	<b>Introduction générale</b>	<b>1</b>
<b>V</b>	<b>General introduction</b>	<b>6</b>
<b>VI</b>	<b>Main concepts and systems</b>	<b>11</b>
<b>1</b>	<b>Photochemical and photophysical processes</b>	<b>12</b>
1.1	Photochemistry . . . . .	12
1.2	Spectroscopies . . . . .	15
1.2.1	UV-visible absorption . . . . .	15
1.2.2	Circular dichroism . . . . .	18
1.2.3	Fluorescence . . . . .	18
1.3	Photosensitization . . . . .	19
1.4	Photodynamic therapy . . . . .	20
<b>2</b>	<b>DNA</b>	<b>23</b>
2.1	Introduction . . . . .	23
2.2	Fundamental building blocks of DNA . . . . .	24
2.3	Double helix . . . . .	25
2.4	DNA structural parameters . . . . .	27



2.5	DNA photosensitivity . . . . .	27
2.6	DNA photolesions . . . . .	28
2.7	Repair processes . . . . .	29
<b>VII</b>	<b>Methods</b>	<b>31</b>
<b>3</b>	<b>Molecular Mechanics</b>	<b>32</b>
<b>4</b>	<b>Molecular Dynamics</b>	<b>34</b>
4.1	Integration of equations of motion . . . . .	34
4.2	Periodic Boundary Conditions (PBC) . . . . .	35
4.3	Ewald summation and Particule Mesh Ewald (PME) . . . . .	37
4.4	Biased Molecular Dynamics . . . . .	38
4.5	Alchemical simulations . . . . .	39
<b>5</b>	<b>Quantum chemistry</b>	<b>41</b>
5.1	The six postulates of quantum mechanics . . . . .	41
5.1.1	Postulate 1: the wave function $\Psi$ . . . . .	41
5.1.2	Postulate 2: Hermitian operators . . . . .	42
5.1.3	Postulate 3: eigenstates and eigenvalues . . . . .	42
5.1.4	Postulate 4: expectation of the wave function . . . . .	42
5.1.5	Postulate 5: Time dependent Schrödinger equation . . . . .	43
5.1.6	Postulate 6: wave function collapse . . . . .	43
5.2	The Schrödinger equation in quantum chemistry . . . . .	43
5.3	Born-Oppenheimer approximation . . . . .	44
5.3.1	Variational Principle . . . . .	44
5.4	Ground state . . . . .	44
5.4.1	Hartree-Fock approximation . . . . .	44
5.4.2	Density Functional theory (DFT) . . . . .	48
5.4.2.1	Principle . . . . .	48
5.4.2.2	The electronic density . . . . .	48
5.4.2.3	The Hohenberg-Kohn theorems . . . . .	49
5.4.2.4	The Kohn-Sham equations . . . . .	50
5.4.2.5	LDA . . . . .	51
5.4.2.6	GGA . . . . .	52
5.4.2.7	meta-GGA . . . . .	52
5.4.2.8	Hybrid GGA and hybrid meta-GGA . . . . .	52
5.4.2.9	Double hybrids . . . . .	53
5.5	Excited states . . . . .	54

5.5.1	Time Dependent-Hartree Fock (TDHF)	54
5.5.2	Time Dependent-Density Functional Theory (TDDFT)	55
5.5.2.1	The Runge-Gross theorem	55
5.5.2.2	Time-Dependent Kohn-Sham equation	55
5.5.3	Frenkel excitons theory	56
<b>6</b>	<b>QM/MM</b>	<b>58</b>
6.1	QM/MM energy	58
6.1.1	QM/MM embeddings	59
6.1.1.1	Mechanical Embedding (ME)	60
6.1.1.2	Electrostatic Embedding (EE)	60
6.1.1.3	Polarisable Embedding (PE)	60
6.2	QM/MM boundary	60
6.2.1	Link Atom (LA)	60
<b>VIII</b>	<b>Results and discussions</b>	<b>62</b>
<b>7</b>	<b>Introduction</b>	<b>63</b>
<b>8</b>	<b>Electronic Circular Dichroism (ECD) modeling</b>	<b>64</b>
8.1	ECD of B-DNA	64
8.2	ECD of G-quadruplexes	71
<b>9</b>	<b>Binding free energy of Benzophenone</b>	<b>80</b>
<b>10</b>	<b>Photosensitization of DNA</b>	<b>91</b>
10.1	Benzophenone: Hydrogen abstraction	91
10.2	Nile Blue, Nile Red: Electron transfer	101
10.3	BMEMC: Radical species	111
10.4	Pyo: Triplet-Triplet Energy Transfer (TTET)	123
<b>11</b>	<b>Damaged DNA: structure and recognition</b>	<b>134</b>
11.1	Cluster abasic sites	134
11.1.1	Recognition by APE1	147
11.2	64-PP and CPD	150

<b>IX</b>	<b>Conclusion générale</b>	<b>162</b>
<b>X</b>	<b>General conclusion</b>	<b>166</b>
<b>XI</b>	<b>Appendices</b>	<b>170</b>
A	DNA structural parameters	171
B	ECD of B-DNA	174
C	ECD of G-quadruplexes	175
D	Binding free energy of Benzophenone	179
E	Benzophenone: Hydrogen abstraction	183
F	Nile Blue, Nile Red: Electron transfer	186
G	BMEMC: Radical reactive species	189
H	Pyo: Triplet-Triplet Energy Transfer (TTET)	191
I	Cluster abasic sites	195
J	6-4PP and CPD	199

# Part I

## Acknowledgments/Remerciements

First of all, I want to warmly thank Antonio and Xavier who are without doubt the pillars of the work accomplished during these three years. I feel privileged to have been supervised by both of you; being introduced to your scientific networks, the conferences, the beers together . . .

Also, I want to thank everyone from the TMS team (now LPCT) for their kindness and our discussions about my scientific problems: Gerald, Mounir, Maria-Chiara, Fabien, Christian, and Severine (without who this lab would fall apart). And the volatiles (PhDs, postdocs and interns): Tony, Maura, Antoine, Ilke, Marina, Maryline, Wanlei, Audrey, Elena, Tony, Elodie, Cecilia, Elodie, Saber, Houda and certainly more. Special mention to the team of Zozos for the entertainment they provided me: Francois, Chris, Andreea, Sebastien and Carole. Since I am leaving soon, they will be happy to know that I will stop disturbing their long "scientific" discussions in the coffee room of the 6th floor.

Also, I would like to express my gratitude to the amazing people who spiced up my PhD: Benjamin for a 2 years long deskmate relationship, amazing times together and interesting discussions at the *Quincaillerie* (we were like two dogs sitting at the top of a hill, looking at the horizon); Thibaud aka *the belgian* for our great friendship and all the things that came with it, parties, trips and more parties; Marco for our great scientific collaboration and his unexpected help to sometimes put things in perspective; Angelo, the best guide of Sicily, first an amazing scientific match between us that soon became a strong friendship; to Fernanda for our adventures together, road trips and discoveries, especially for having initiated me to rollerblading; the Phi-Science team of course that was after all always here when I needed to relax from the first day of my PhD to the last, Delphine, Clara, Manon, Ismael, Belo, Greg, Nata . . .; I also feel privileged having built such an amazing friendship with my dear PhD fellows, I definitely wish to see them again one day, Lara, Hugo, Christopher, Max, Vincent, Gauillaume, Bisch, Stephane and Yannis; and to my roommate JS, the greatest apartment brewer in this world.

Also a quick thank to Knarr who carried me fiercely across mountains and valleys without dropping a single screw on the way (so far).

Et pour finir je tiens à remercier mes plus proches amis, Marie-Chantal et Etienne qui ont toujours été là pour moi et avec qui j'ai partagé tant de choses. A Clémence, j'espère que nos chemins se recroiseront un jour. A ma sœur, aka *sakos*, qui a su m'apporter une aide précieuse grâce à ses talents d'artiste et enfin à mes parents sans qui rien de tout cela n'aurait été possible. Je tiens à profondément les remercier de m'avoir toujours soutenu dans mes décisions et surtout de m'avoir transmis une légèreté d'esprit me qui m'a permis d'affronter toutes les épreuves de la thèse. Quoi qu'il en soit, je tiens à m'excuser pour les coups occasionnés par toutes ces années d'études, comme dirait mon père, "*Ca n'a pas été gratuit!*".

## Part II

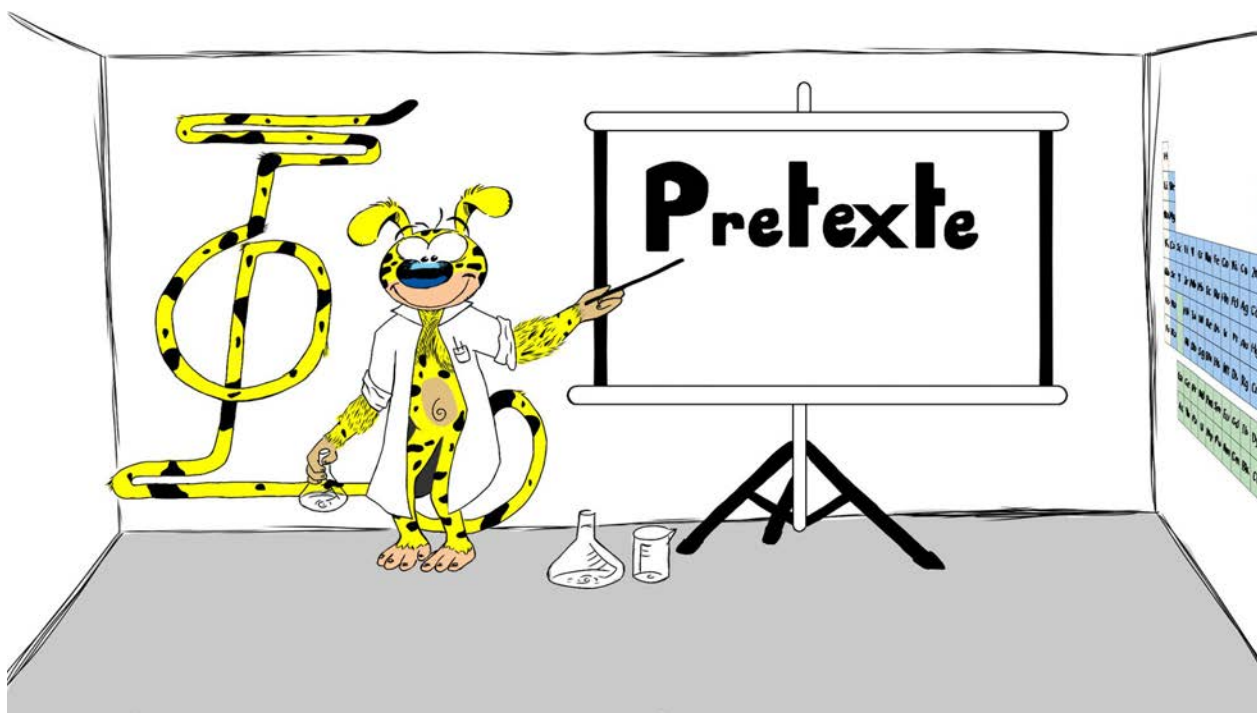
### Outreach activities

## 0.1 Introduction

The aim of this manuscript is to reflect the work that has been realized during these three years, and since I have been highly involved in several initiatives of scientific diffusion and popularization, I will quickly present them in the following pages. Since two of these initiatives have a popularization aim they can be considered by the reader as an "appetizer" of my research project before starting the main core of this work.

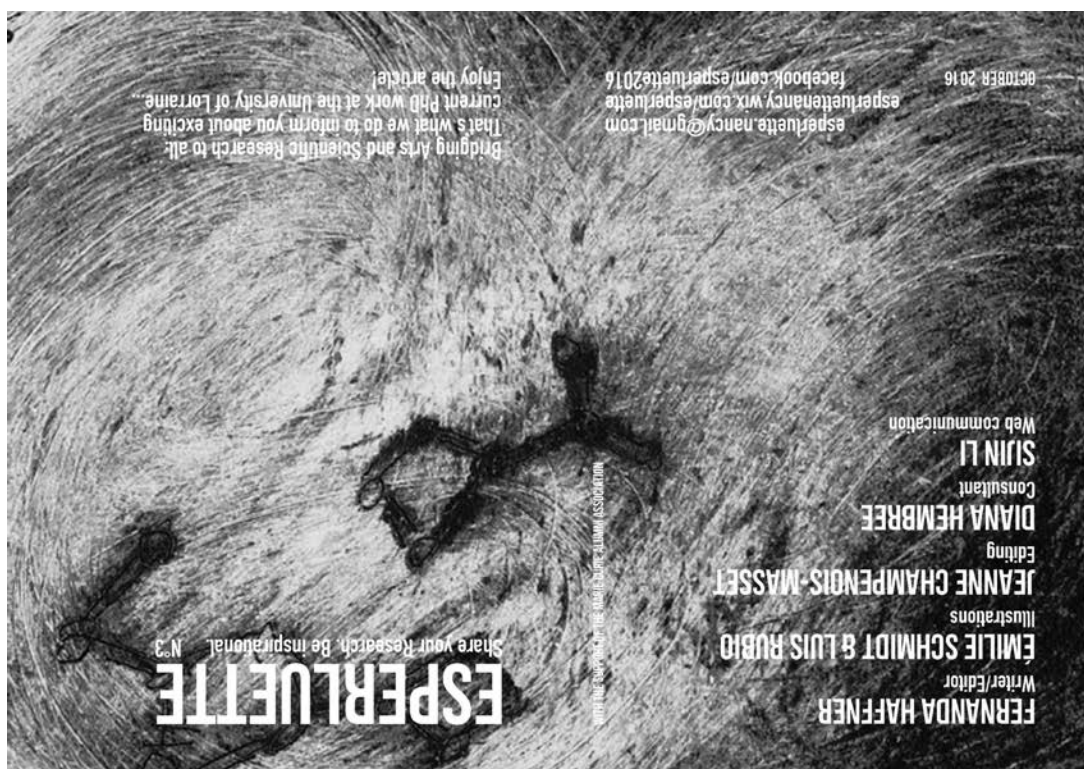
## 0.2 Pretexte

In a desire to enhance the interactions between bachelor/master students and researchers, with the help and ideas of two other motivated PhD students, we created the "Pretexte" series of scientific popularization conferences. A simple concept: every month in a bar 4 to 5 PhD students have 10 minutes each to explain their PhD projects to a broad audience. Simple, but highly efficient in sharing scientific concepts to the students, each edition has been enjoyed by the 60 to 80 participants who could discover the science surrounding them in these laboratories they sometimes barely heard of before. Moreover, these events were often the first experience in scientific popularization for the PhD students who volunteered, thus it has been a great training for them.



## 0.3 Esperluette

Popularization events are great to share science but they are ephemeral. Luckily enough, other initiatives ought to popularize scientific projects of PhD students in a long term fashion. Indeed it is the aim of "Esperluette", which brings together arts and sciences in a lampoon like document. On one side, the project of a PhD student is presented and popularized, on the other side the research is pictured through an artistic point of view. The edition focused on my PhD project is presented bellow:



### THE FUNKY INTERACTIONS OF CHEMICALS AND DNA

**STUDENT DOMAIN** Hugo Gattuso  
Theoretical Chemistry

#### VOCABULARY AID

*DNA* - Deoxyribonucleic acid

*Synthesized* - Obtained via a chemical process

*Nucleobases* - The fundamental units of the genetic code alphabet.

The entire human genome is composed by 3 273 000 000 bases.

Have you, too, come to the conclusion that our genes are a fragile masterpiece? A molecule called *DNA* constitutes this remarkable jewelry, which carries all the information about our development, functioning and reproduction. Wow! Surely you keep questioning yourself about what could happen to the DNA, and thus to humankind, once it gets in contact

with certain chemicals. If your answer is that eventually these chemicals would transform us into Cthulhu monsters, you've got it partly right -- thanks to Hollywood and science fiction. But seriously, if these chemicals interact with the molecule that stores our genetic code, how harmful can this be?

Well, that depends on the chemical. Most of the *synthesized* unnatural chemicals, such as pesticides, for example, are metabolized and assimilated by our bodies, causing no noticeable harm for our health at first or at all. However, after reaching the internal compartment of cells, the nucleus, some of these chemicals can freely interact with the DNA and form long-living aggregates. Depending on the chemical structure of the molecule in question and the famous double-helical arrangement of the DNA double-strand, specific interactions between these molecules may occur, resulting in changes in the DNA structure. Messing with DNA can be pretty terrible news, but this is where scientists come into the picture. They enter the understated game of chemical interactions through molecular modeling -- that is -- computational simulations.

"My field is known as theoretical chemistry. One of its great advantages is that it allows me to cover chemistry, physics, biology,

mathematics and informatics. Everything is done in an enormous calculation cluster. In comparison to scientists working on experiments, I can start a long simulation and go for a beer. Moreover, I only need a wifi spot to work in order to connect to the calculation machines. Imagine how cool it would be to do science in a bar! It is a perfect area of research for the ones who do not want to get their hands dirty with chemicals, but still want to study them".

In my PhD, I simulate how certain molecules could interact with parts of the DNA -- 40 *nucleobases* only. The whole DNA is not taken into consideration in the calculations, mostly due to the computer capacity. This approach is plausible in terms of scientific robustness, since in reality small molecules interact only with a maximum of a dozen nucleobases. Besides, in such systems, you usually need two days of calculations for 0.0001 second of real time.

Via computational simulations, scientists can get precious information at a molecular level, which can often not be elucidated via chemical experiments. I am Hugo Gattuso, a PhD student working in the SRSMC laboratory of the University of Lorraine. My project is funded by the university's research department and the French government.



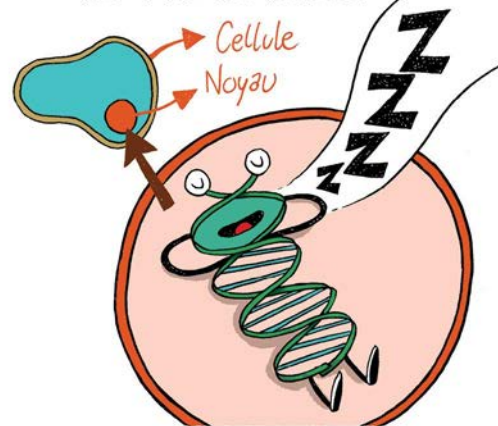


## 0.4 MT180

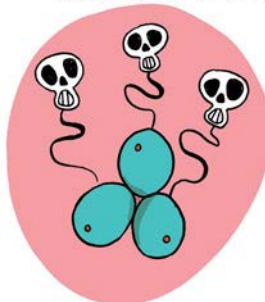
And finally, I have been involved in the well known contest of science popularization, "Ma thèse en 180 secondes" (MT180). Having reached the final of the *Université de Lorraine*, a comic of my research has been created for this particular occasion:



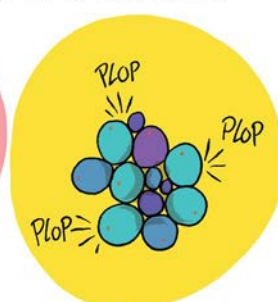
Si on respecte les temps d'exposition, notre ADN est relativement bien protégé au cœur de nos cellules.



Par contre, si les rayons parviennent à entrer dans le noyau, ils peuvent atteindre l'ADN et l'irradier.



ADN DÉTRUIT  
= MORT DES CELLULES

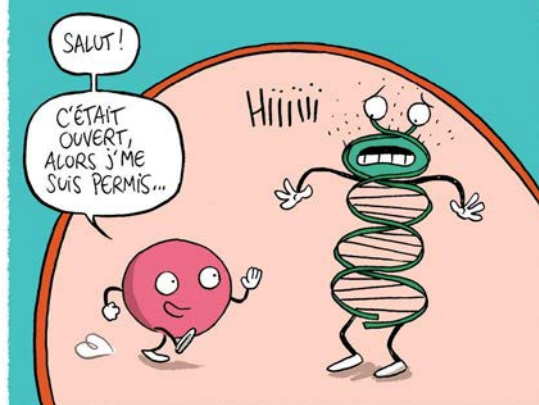


ADN MUTANT  
= DÉVELOPPEMENT ANARCHIQUE

Le code génétique bugge, les cellules prolifèrent de façon incontrôlable: c'est le cancer.



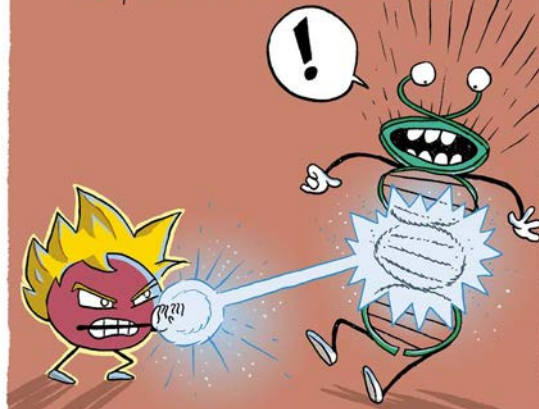
Dans notre quotidien, certaines molécules étrangères à notre corps peuvent s'introduire dans le noyau des cellules.



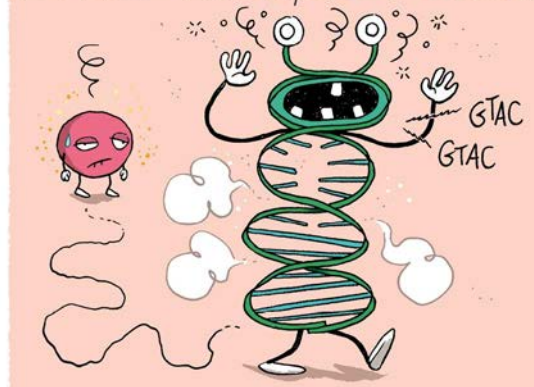
Et si, par malheur, elles arrivent à capter l'énergie lumineuse au travers de la paroi; les conséquences sont terribles.



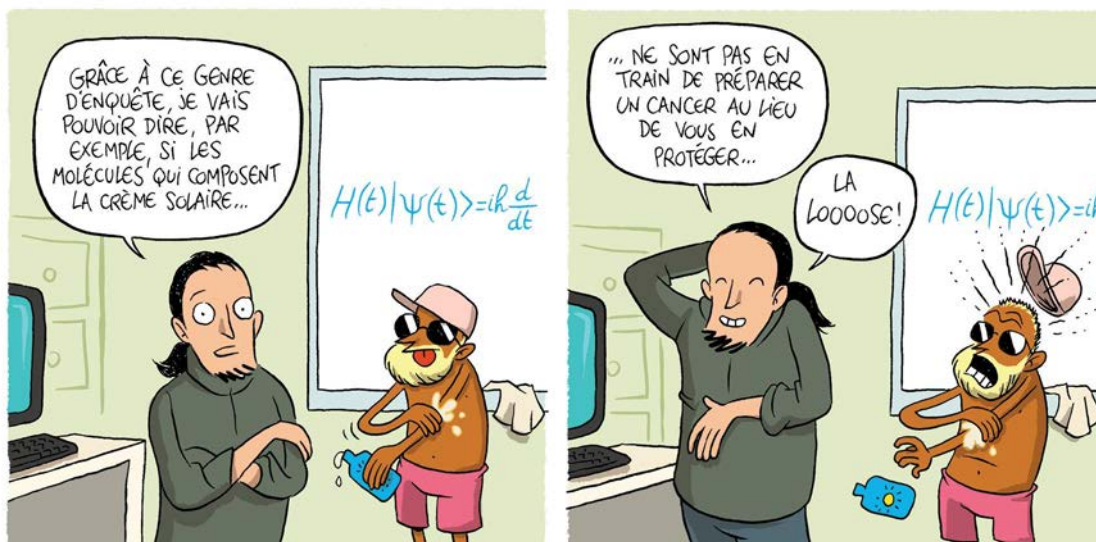
Toute cette énergie accumulée est déchargée sur l'ADN alors qu'il se croquait à l'abri du soleil.



Plus intensément que s'il avait été directement exposé aux rayons du soleil, il va se détruire ou muter, ce qui aura bien évidemment un impact sur la cellule.







The contest itself gathered 11 candidates who had to present their PhD projects in 180 seconds in front of a vast audience. During this exciting evening I succeeded to win the Public Award which recognizes the quality of the presentation. You can rewatch it online at the following clickable [link](https://videos.univ-lorraine.fr/index.php?act=view&id=4618) (url: <https://videos.univ-lorraine.fr/index.php?act=view&id=4618>).



## Part III

### List of publications

**List of publications included in this manuscript:**

- A F. Dehez, **H. Gattuso**, E. Bignon, C. Morell, E. Dumont, A. Monari *Conformational polymorphism or structural invariance in DNA photoinduced lesions: implications for repair rates*, **Nucleic Acid Research**, 2017, 45(7), 3654-3662.
- B **H. Gattuso**, E. Durand, E. Bignon, C. Morell, A. G. Georgakilas, E. Dumont, C. Chipot, F. Dehez, A. Monari, *Repair Rate of Clustered Abasic DNA Lesions by Human Endonuclease: Molecular Bases of Sequence Specificity*, **The Journal of Physical Chemistry Letters**, 2016, 7(19), 3760-3765.
- C E. Bignon, **H. Gattuso**, C. Morell, F. Dehez, A. G. Georgakilas, A. Monari, E. Dumont, *Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion*, **Nucleic Acids Research**, 2016, 44(18), 8588-8599.
- D **H. Gattuso**, A. Spinello, A. Terenzi, X. Assfeld, G. Barone, A. Monari, *Circular Dichroism of DNA G-Quadruplexes: Combining Modeling and Spectroscopy To Unravel Complex Structures*, **The Journal of Physical Chemistry B**, 2016, 120(12), 3113-3121.
- E **H. Gattuso**, E. Dumont, C. Chipot, A. Monari, F. Dehez, *Thermodynamics of DNA: Sensitizer recognition. Characterizing binding motifs with all-atom simulations*, **Physical Chemistry Chemical Physics**, 2016, 18(48), 33180-33186.
- F **H. Gattuso**, V. Besancenot, S. Grandemange, M. Marazzi, A. Monari, *From non-covalent binding to irreversible DNA lesions: nile blue and nile red as photosensitizing agents*, **Scientific Reports**, 2016, 6(28480).
- G **H. Gattuso**, E. Dumont, M. Marazzi, A. Monari, *Two-photon-absorption DNA sensitization via solvated electron production: unraveling photochemical pathways by molecular modeling and simulation*, **Physical Chemistry Chemical Physics**, 2016, 18(27), 18598-18606.
- H M. Marazzi, M. Wibowo, **H. Gattuso**, E. Dumont, D. Roca-Sanjuán, A. Monari, *Hydrogen abstraction by photoexcited benzophenone: consequences for DNA photosensitization*, **Physical Chemistry Chemical Physics**, 2016, 18(11), 7829-7836.
- I E. Bignon, **H. Gattuso**, C. Morell, E. Dumont, A. Monari, *DNA Photosensitization by an "Insider": Photophysics and Triplet Energy Transfer of 5-Methyl-2-pyrimidone Deoxyribonucleoside*, **Chemistry: A European Journal**, 2015, 21(32), 11509-11516.
- J **H. Gattuso**, A. Xavier, A. Monari, *Modeling DNA electronic circular dichroism by QM/MM methods and Frenkel Hamiltonian*, **Theoretical Chemistry Accounts**, 2015, 134(3), 1-8.

**List of publications not included in this manuscript:**

- K **H. Gattuso**, A. Monari, M. Marazzi *Photophysics of Chlorin e6: From One- and Two-Photon Absorption to Fluorescence and Phosphorescence*, **RCS Advances**, 2017, 7(18), 10992-10999.
- L T. Etienne, **H. Gattuso**, C. Michaux, A. Monari, X. Assfeld, E. A. Perpète, *Fluorene-imidazole dyes excited states from first-principles calculations - Topological insights*, **Theoretical Chemistry Accounts**, 2016, 135(4), 1-11.
- M M. Marazzi, **H. Gattuso**, A. Monari, *Nile blue and Nile red optical properties predicted by TD-DFT and CASPT2 methods: static and dynamic solvent effects*, **Theoretical Chemistry Accounts**, 2016, 135(3), 1-11.
- N **H. Gattuso**, T. Duchanois, V. Besancenot, C. Barbieux, X. Assfeld, P. Becuwe, P. C. Gros, S. Grandemange, A. Monari, *Interaction of Iron II Complexes with B-DNA. Insights from Molecular Modeling, Spectroscopy, and Cellular Biology*, **Frontiers in Chemistry**, 2015, 3(67).



## Part IV

### Introduction générale

La **photosensibilisation de l'ADN** est un concept scientifique sur lequel repose l'intégralité du travail qui a été réalisé pendant ces trois années de thèse. Il peut être tout simplement décrit comme étant l'augmentation de l'impact d'une source d'énergie lumineuse véhiculée par un composé appelé photosensibilisateur sur la structure moléculaire de l'ADN (principalement des molécules organiques telles que celles de la famille des porphyrines ou des chlorines) [1].

On retrouve ce phénomène dans le traitement de cancers et de maladies de la peau sous la forme d'une technique appelée photothérapie dynamique [2, 3].

L'évolution des photosensibilisateurs a recouvert une large variété de composés, d'éléments chimiques et de structures moléculaires [4, 5]. On retrouve principalement dans la première génération les médicaments commerciaux HpD[6] et Photofrin[7] qui sont de la famille des porphyrines. Ensuite, la deuxième génération, bien plus diverse, a permis une augmentation de l'efficacité et de la sélectivité des traitements avec comme molécule principale le Foscan[8]. Et enfin, la troisième génération, basée sur des porphyrines contenant un atom métallique en leurs centres améliorant leurs capacité à capter l'énergie lumineuse améliorant ainsi leur efficacité [9].

Malgré ces recherches, un manque en sélectivité, biocompatibilité et adaptabilité lors des traitements est à déplorer et des efforts supplémentaires sont nécessaires au développement et à la découverte de nouveaux photosensibilisateurs. Sachant que *chaque cancer est différent* [10], il apparaît évident que la recherche de nouveaux composés ayant une plus grande efficacité va se heurter à de nombreux obstacles.

Cette problématique vieille de plus d'un siècle s'est de nos jours étendu à tous les domaines de recherche qui pourraient de près ou de loin apporter de nouvelles idées mais surtout une meilleure compréhension des mécanismes sous-jacents. Le domaine de la modélisation moléculaire, qui permet de décrire et d'étudier des phénomènes moléculaires à l'échelle atomique voire électronique. C'est un outil qui apporte une compréhension des résultats expérimentaux sur les phénomènes liés à la photosensibilisation de l'ADN [11, 12].

De plus, puisque les phénomènes biochimiques sont affectés par leurs environnement aux niveaux local et global, ils nécessitent d'être décrits à plusieurs échelles, de l'électron jusqu'à l'ensemble macromoléculaire. Dans l'ADN, et plus spécifiquement pour l'étude de sa photosensibilisation, il est important d'avoir d'un côté une compréhension détaillée de la structure électronique des bases et des agents sensibilisants impliqués dans le processus et d'un autre côté une description complète à plus large échelle de l'environnement. La stratégie multi-échelles rassemble différentes méthodes calculatoires en fonction du niveau de détails qui est requis, c'est à dire la propriété qui est étudiée.

Premièrement, l'utilisation de la dynamique moléculaire a permis d'étudier des brins d'ADN (relativement courts en comparaison avec l'intégralité d'un brin dans

un chromosome) dans ses multiples arrangements [13] (ADN de forme A, B et Z, G-quadruplexes, ...) et dans des environnements qui peuvent être considérés comme étant relativement proche des conditions biologiques, incluant les molécules d'eau explicites servant de solvants, une concentration physiologique en ions, allant jusqu'à l'interaction avec des protéines et des enzymes. L'augmentation des capacités de calculs a permis l'étude de systèmes d'acides nucléiques toujours plus grands, d'abord seulement quelques paires de bases [14] qui n'étaient pas encore pertinentes à cette époque pour soutenir la recherche sur la photosensibilisation de l'ADN jusqu'à aujourd'hui où les simulations des systèmes composés de dizaines de paires de bases peuvent être modélisés sur des périodes de temps comparables à celles de phénomènes biologiques [15]. De plus l'amélioration constante des champs de forces, à la fois pour la description de propriétés structurelles et dynamiques de l'ADN [16] et plus important encore la simulation de composés moléculaires organiques [17] (tels que les photosensibilisateurs) a permis de renforcer la qualité et la légitimité des résultats de modélisation moléculaire pour l'étude des systèmes biologiques.

Deuxièmement, la mécanique quantique et plus spécifiquement le domaine de la chimie quantique permet une description de systèmes moléculaires au niveau électronique. Les études mécaniques quantiques apportent une compréhension directe des mécanismes de photosensibilisation qui sont basés sur l'activation par la lumière, c'est à dire la population d'états excités [12, 18, 19]. Tout a commencé par l'étude des propriétés électroniques d'acides nucléiques isolés [20] (dans le vide ou dans des solvants implicites) ou en interaction [21] (paires de bases,  $\pi$ -stacking). Par exemple des croisements inter-systèmes [22] ou des recombinaisons de charges [23]. De plus, l'interaction entre des photosensibilisateurs et l'ADN et leurs possibles mécanismes de photoréactions sont activement étudiés [24]. La principale limitation de ces méthodes est leurs besoins en capacités de calcul, ce qui est directement relié à la prise en compte de l'environnement de tels systèmes pour pouvoir envisager une comparaison directe avec des données expérimentales. Aussi la composition de ces systèmes, comme la présence d'atomes métalliques par exemple est un autre inconvénient à l'utilisation des techniques de mécanique quantique tant la photophysique de tels systèmes est bien souvent complexe.

Finalement, il apparaît évident de considérer la combinaison des deux méthodes présentées ci-dessus si on souhaite s'attaquer à l'étude de processus électroniques entre des agents sensibilisants et des ensembles d'acides nucléiques tout en prenant en compte l'environnement. Cette stratégie appelée QM/MM, pour *Quantum Mechanics/Molecular Mechanics*, a été reconnue par un prix Nobel de chimie en 2013 remis à Martin Karplus, Michael Lewitt and Arieh Warshel pour leur aide au développement, à l'implémentation et à la validation des méthodes QM/MM tout au long

de leurs carrières [25]. De nos jours la QM/MM a atteint une maturité suffisante pour pouvoir être utilisée lors de l'étude de phénomènes chimiques, physiques et biologiques [26–28]. De ce fait, il est apparu évident d'appliquer cette méthode afin de réaliser les recherches inscrites dans ce projet de thèse.

Avec le concept de "photosensibilisation de l'ADN" comme coeur d'étude, nous nous sommes intéressés aux sujets exposés ci-dessous (ne seront présentés dans le corps de ce manuscrit que ceux qui lui apportent la meilleure cohérence).

- La première étape a été d'étudier et de comprendre les propriétés photo-physiques des agents sensibilisants. Plus spécifiquement nous avons réalisé une investigation des composés nile bleu et nile rouge (**M**) ainsi qu'un photosensibilisateur de la rhodopsine qui est une protéine bovine impliquée dans la vision, la chlorine E6 (**K**).
- Ensuite il fut impératif de déterminer la capacité de ces molécules à s'aggréger de manière stable avec l'ADN, pour une durée suffisamment importante pour permettre l'activation de processus électroniques liés à la photosensibilisation. En effet, la plupart des projets qui ont été menés ont commencés par l'étude des interactions molécules/ADN dans différentes conformations. Par exemple on peut citer la liaison non-covalente de complexes de fer en intercalation et dans le petit sillon de l'ADN (**N**) et d'autres composés photoactifs tels que nile bleu, nile rouge (**F**) ou un carbazole modifié, BMEMC (**G**).
- Directement lié au point précédent, il fut important de déterminer les stabilités relatives de ces molécules dans leurs cavités dans l'ADN. Ainsi, nous avons développés un protocole calculatoire basé sur des transformations alchimiques afin d'obtenir les énergies libres d'association de ces complexes. Cette méthodologie a été testée sur l'interaction de la benzophénone avec un dodécamer d'ADN (**E**) soit en double insertion, soit glissée dans le petit sillon. Les résultats ont montrés la capacité de ce protocole à fournir une évaluation directe des stabilités relatives entre différents modes d'interactions.
- En parallèle à ce travail, nous avons créé un outil de post-traitement de calculs quantiques afin de modéliser une autre propriété clé de l'ADN et de ses interactions avec des ligands, le dichroïsme circulaire. En effet, dans le cadre de la théorie des excitons de Frenkel, nous avons pu reproduire le signal de dichroïsme circulaire d'ensembles d'acide nucléiques tels que l'ADN dans sa forme B (**D**) mais aussi des formes plus complexes comme les G-quadruplexes (**J**). Ces données spectroscopiques modélisées peuvent être directement reliées à la structure des systèmes moléculaires et dans le future nous pourrions

fournir des indications cruciales quant à la comparaison de différents modes d'interactions ou même entre différentes structures primaires et secondaires.

- Le coeur de ce travail a porté sur l'étude des possibles mécanismes de photosensibilisation, c'est à dire les réorganisations électroniques qui pourraient se dérouler après excitation par une source lumineuse. En effet nous avons cherchés à déterminer les possibles transferts électroniques entre deux photosensibilisateurs connus le nile bleu et le nile rouge ainsi que BMEMC et l'ADN (**F**, **G**), le profil énergétique de l'abstraction d'un hydrogène d'une base ou d'un sucre de l'ADN par la molécule benzophénone dans son état triplet (**H**) mais aussi un transfert d'énergie entre une nucleobase modifiée photoactivable et une thymine proche (**I**). Ces études ont montrés que le mode d'interaction, la séquence du brin d'ADN considéré et l'environnement jouent un rôle majeur dans le processus de photosensibilisation par ces composés.
- L'étape suivante a été de nous engager dans des études approfondies de l'influence de lésions de l'ADN (dimère de cyclopyrimidine (CPD, **A**), pyrimidine(6-4)pyrimidone (6-4PP, **A**), 5-méthyl-2-pyrimidone (Pyo, **I**) et les nucléotides apuriques/apyridiniques (sites AP, **C**)) sur la structure locale et globale de la double hélice ainsi que ses comportements dynamiques et ses propriétés photophysiques. Ces résultats ont pu être corrélés avec des données expérimentales telles que la courbure du brin ou bien des activités enzymatiques de réparation.
- Et enfin, le phénomène de réparation de l'ADN par des protéines qui implique la reconnaissance du double brin a été étudié. En effet, nous nous sommes concentrés sur la reconnaissance de brins d'ADN contenant des clusters de sites AP par une endonucléase humaine (APE1, **B**). Les résultats, en accord avec les données expérimentales ont pu montrer une fois de plus la capacité de la modélisation moléculaire à assister la compréhension de processus chimiques et biologiques fondamentaux.

Tous ces sujets ont convertis un large ensemble de domaines liés premièrement à la création de dommages dans les acides nucléiques, deuxièmement leurs influences sur l'ADN et la reconnaissance ADN/protéine menant à leur anihilation. Je pense que nous avons su apporter de nouvelles compréhensions des phénomènes photophysiques liés à la photosensibilisation ainsi que de nouvelles méthodes et outils qui permettront des avancées dans l'étude de l'ADN et de ses nombreuses propriétés.

# Part V

## General introduction

**DNA Photosensitization** is the scientific concept on which relies all the work that has been achieved during these three years. It can be simply described as the enhancement of the sensitivity of DNA by an external element, called a photosensitizer (mostly organic molecules such as porphyrins and chlorins) [1]. The process of photosensitization itself can also be employed to target other important macromolecules present in living organisms such as proteins and membranes [29].

This phenomenon is nowadays paramount in cancer and skin treatments known as photodynamic therapy (1.4) [2, 3].

The evolution of photodrugs has exploited a tremendous variety of molecular elements, structures, and compositions [4, 5]. Starting from the first generation composed of the two commercial compounds HpD[6] and Photofrin[7], mainly porphyrin-like molecules, followed by the second generation, more diverse, with a higher efficiency and selectivity, in which is found one of the main photosensitizers in nowadays treatments, Foscan[8]. Finally the third generation, focused on metalated porphyrins which had increased light harvesting thus higher activity[9].

However, a lack in selectivity, biocompatibility and adaptability to every medical cases underline the need for a supplementary effort in the discovery and development of photosensitizers. Since *every cancer is different* [10], it appears evident that the exploration for new compounds of better efficiency will face many obstacles.

This more than centennial challenge has nowadays spread over all the fields of research that might closely or remotely bring new points of view, new discoveries, new information and more importantly new understandings of the underlying mechanisms. In the field of molecular modeling, the capacity to obtain atomistic and electronic descriptions of molecular phenomena has been for a long time a useful tool to bring understanding to experimental evidences of DNA photosensitization[11, 12].

Moreover since many biochemical phenomena are affected by both local and global properties, they require to be described in a multiscale fashion, from electrons to atomic macromolecular arrangements. In DNA, and more specifically in the field of DNA photosensitization, it is necessary to have on one hand a thorough understanding of the electronic structure of nucleobases and sensitizing agents involved in the process and on the other hand a complete large scale description of solvation environments and the remaining of the double strand. In fact the multiscale strategy spans different computational methods depending on the scale that is considered, thus on the properties that are looked up.

On one side, the use of Molecular Mechanics (MM) and Molecular Dynamics (MD) has provided the possibility to study first relatively short (compared to the entire genome) DNA double-strand in its many arrangements[13] (A-, B-, Z-DNA, G-quadruplexes, hairpins, ...) in environments that can be considered as relatively

close to biological ones, including water molecules as solvent, in the presence of physiological concentrations of ions and even in interaction with proteins or enzymes. The increase of computing capacities has offered the opportunity to study larger and larger nucleic acid systems, starting from only several base pairs [14] which was not relevant at that time to support the research about DNA photosensitization, until nowadays, state-of-the-art investigations of dozens of base pairs that can be achieved for lengths of time directly comparable to those of biological phenomena[15]. Furthermore the improvement of force fields, both for the description of the DNA structure and its dynamic behavior [16] and more importantly for the reproduction of molecular organic compounds (photosensitizers) properties[17] have offered the possibility to tackle with increasing accuracy many biological cases.

On the other side, Quantum Mechanics (QM) and more specifically the field of Quantum Chemistry is the scientific field that provides a description of molecular systems at the electronic level. QM studies directly support the understanding of photosensitization mechanisms which rely on an activation by light, that is the population of electronic excited states[12, 18, 19]. These studies started with the investigation of nucleic acid electronic properties, isolated (in vacuum or in implicit solvation) [20] or in interaction [21] (base pairs,  $\pi$ -stacking) which even expressed a more complex behavior, for example intersystem crossings[22] and charge recombinations [23]. Furthermore the interaction and direct light triggered mechanisms in nucleic acids and sensitizers is being extensively studied [24]. The limitation of these methods is mainly due to their computational cost, which is directly related to the necessity to take into account the environment of such systems in order to obtain data that can be ultimately compared to experimental results. Moreover, the compositions of the systems, the number of nucleic acids that has to be considered, the presence of metal atoms in the photosensitizer are other methodological drawbacks.

In fact, the combination of both QM and MM methods have to be considered if one wants to tackle the investigation of electronic processes between sensitizing agents and nucleic acid ensembles while taking into account complex environments. This method named QM/MM, for Quantum Mechanics/Molecular Mechanics, acquired its reputation for excellence in 2013 with a Nobel prize of chemistry that awarded the work of Martin Karplus, Michael Lewitt and Arieh Warshel on the development, implementation and validation of QM/MM throughout their careers [25]. Indeed, nowadays, QM/MM can be considered as having reached a sufficient maturity to be employed in tackling chemical, physical and biological problematics[26–28], thus it appeared evident to focus on this strategy to carry out the projects of these past three years.

Starting with the idea of "DNA photosensitization" as a core concept, we tackled



many related topics presented bellow (in the manuscript I will present only the projects that provides the best coherence ).

- Indeed, the first step has been to study and understand the photophysical properties of these sensitizing agents, more specifically one can cite the organic compounds nileblue and nilered (**M**) and another sensitizing agent acting on the membrane rhodopsin protein, chlorin E6 (**K**).
- Then it was necessary to rationalize the capacity of a compound to generate stable aggregates with biomacromolecules that will live long enough to allow the electronic process to occur. In fact, most of the projects started by the investigation of DNA/drug binding configurations, we can cite for example the interaction of iron complexes through different binding modes (**N**) and other organic chemicals such as nileblue, nilered (**F**) or a modified carbazole, BMEMC (**G**).
- Directly related to this, since it is necessary to study the overall stabilities of these molecules in there binding pockets in DNA, we developed a protocol based on alchemical transformations to compute free energies of association. It has been applied on the interaction of benzophenone in a DNA dodecamer (**E**). The results showed the ability of this new methodology to provide a direct comparison of binding mode stabilities that shall in the future permit advances in the development of more specific drugs.
- In conjunction to this work, we also created a tool to model another key property of DNA and its behavior in contact with ligands, its circular dichroism. Indeed, in the framework of the Frenkel exciton theory we managed to reproduce the circular dichroism signal of nucleic acid ensembles such as B-DNA (**D**) and G-quadruplexes (**J**). Such optical data can directly be related to structures of molecular systems and in the future we might be able to provide strong evidences of the validity of a binding conformation compared to another or even primary and secondary structures of more complex systems.
- Subsequently to the binding of a molecule in DNA, the possible photosensitizing pathways, hence the photochemistry and photophysics have been addressed. Indeed, we investigated the possible electron transfer mechanisms of three known photosensitizers of DNA, nileblue, nilered and BMEMC (**F**, **G**), the energetic of the abstraction of a hydrogen from a sugar or a base of the DNA structure by the benzophenone molecule activated in its triplet state (**H**) and also the triplet energy transfer induced to a nucleobase by an

endogenous phototriggerable modified nucleobase (**I**). The results we have obtained demonstrated that the binding configuration, the DNA sequence and the environment play a major role on the photosensitization process.

- More recently, we also engaged thorough investigations of the influence of DNA lesions (cyclopurimidine dimers (CPD, **A**), pyrimidine(6-4)pyrimidone (64-PP, **A**), 5-methyl-2-pyrimidone (Pyo, **I**) and apurinic/apyridinic nucleotides (AP-sites, **C**)) on the double helix global and local structure as well as its dynamic behavior or photophysical properties.
- And as a final step, we started studying the phenomenon of lesion repair by proteins that involves the description of the recognition of damaged DNA strands. So far we mainly focused on the recognition of double strands containing several types of AP cluster lesions by a human endonuclease enzyme (APE1, **B**). The results, corroborating experimental data, and again confirmed the efficiency of molecular modeling to assist the understanding of fundamental biological processes.

All these subjects spanned the fields of nucleic acid damages, from their creation to their annihilation. We achieved to provide understandings, new methodologies and tools that will allow to go forward in the study of DNA and its properties.

## Part VI

### Main concepts and systems

# Chapter 1

## Photochemical and photophysical processes

### 1.1 Photochemistry

The general concept of photochemistry [30] relies on the fact that when a molecular system is hit by a light source, which is in fact an electromagnetic wave (or photons based on the wave-particle duality), reorganizations of its electronic structure can be triggered, thus modifying its potential energy surface and inducing chemical reactions, energy absorptions/emissions, ionizations, ... It is a field of chemistry employing the capacity of molecules to absorb light and reach electronic excited states of higher energy. Photochemistry is ruled by two main laws:

**Grotthuss-Draper law** a photochemical reaction occurs when and only when light is absorbed. Even if this statement seems obvious, it actually signifies that a light source is mandatory and that its energy should be close enough to the energy gap between the molecule ground and excited states so that absorption can occur.

**Stark-Einstein law** every photon that a molecular ensemble absorbs induces an electronic excitation.

Once a molecule has absorbed light, it does not necessarily evolve toward a chemical reaction, instead radiative phenomenons can occur. Indeed, the excess of energy can be released either as heat (internal conversion) or by the emissions of photons (fluorescence and phosphorescence).

The possible electronic processes following electronic excitation are presented in Figure 1.1 in a diagram called Jablonski diagram [31] that describes the energy levels of the ground state  $S_0$ , the singlet excited states  $S_1$  and  $S_2$  and a triplet state  $T_1$ . It is a schematic representation of the electronic and vibrational excited states, their energy levels and the possible transitions between them.

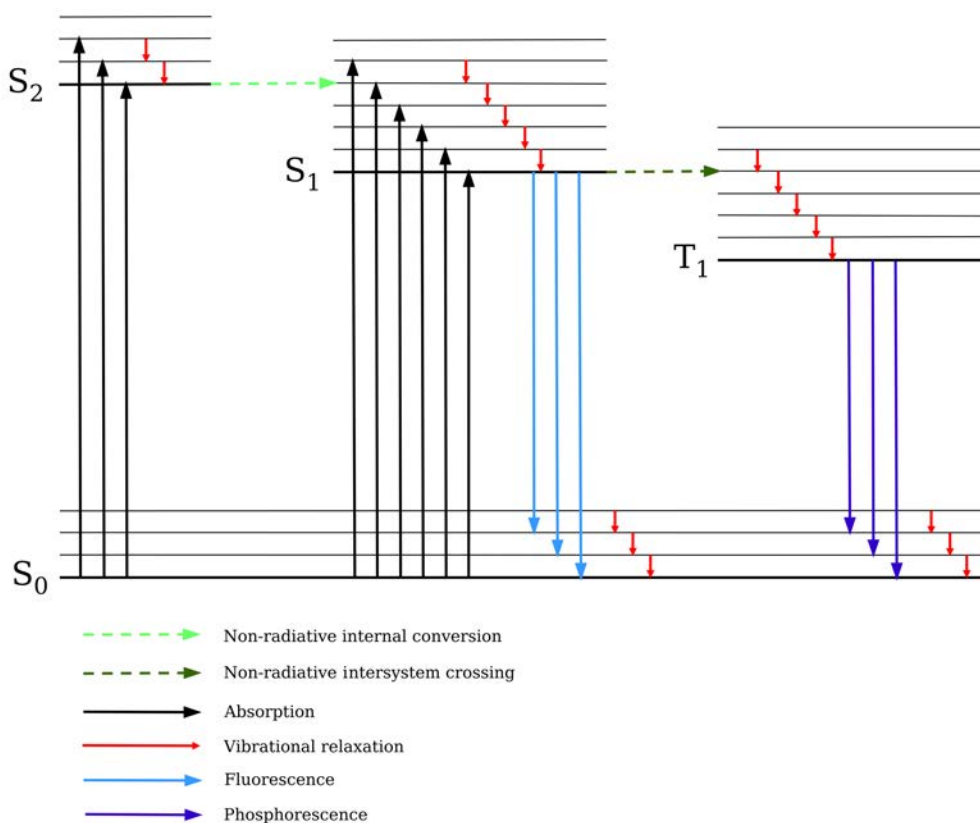


Figure 1.1: Jablonski diagram

First of all, the vertical arrows between states (black, light blue and dark blue) are called Franck-Condon transitions [32]. They are considered to be of a time scale infinitely larger than the motion of nuclei meaning that the atomic coordinates do not evolve during the process. The arrows correspond to:

**black arrow:** the electronic transition from a lower to a higher energy level of same multiplicity is called absorption. The time scale of this transition is approximately  $10^{-15}$  second.

**light blue arrow:** the emission of a photon during the transition between two states of same multiplicity is called fluorescence. The lifetime of the electronic excited states before emission is from  $10^{-9}$  to  $10^{-7}$  second, since the system has to relax first to the excited level of lowest energy before the electronic transition to the ground state. This is due to the fact that internal conversion and vibrational relaxation occurs.

**dark blue arrow:** the emission of a photon due to an electronic transition between states of different multiplicities (triplet and singlet) is called phosphorescence. The life time of the triplet state is around  $10^{-3}$  second since the triplet  $\rightarrow$  singlet transitions (and vice versa) are formally forbidden and can only occur due to spin-orbit and vibronic couplings between the states.

**red arrow:** this process is called vibrational energy relaxation. A system out of equilibrium on a vibrational excited state can relax to lower energy levels in the same electronic state through vibration modes. It is a non radiative energy dissipation process that releases the excess of energy as kinetic energy. The relaxation to the lowest vibrational state will occur between  $10^{-14}$  and  $10^{-11}$  second.

**light green dashed arrow:** when vibrational states of two different electronic states are close in energy, an internal conversion can occur. It is a non-radiative process that transfers energy to vibration modes generating heat and allowing the decay to lower electronic states without emission.

**dark green dashed arrow:** this forbidden (made possible thanks to the overlap of the vibrational energy levels of the two states and their spin-orbit coupling) non-radiative electronic transition is called intersystem crossing. It is the fundamental phenomenon that permit the population of triplet states of molecular systems.

Every photochemical phenomenon can be quantified by its quantum yield[33] which can be defined as the number of molecules that have reacted divided by the number of absorbed photons (or in the case of fluorescence, the number of emitted photons divided by the number of absorbed ones).

Photochemistry and photophysical processes are found to be the fundamental core of many biological mechanisms such as:

- Photosynthesis in plants [34];
- vision through the photoswitching of the rhodopsin chromophore [35];
- production of melanin and vitamin D [36];
- DNA repair mechanisms (flavoproteins) [37].

Nowadays, many industrial and medical processes related to our daily life rely on photochemistry. For example:

- The analysis of the composition of samples (field of spectroscopy, section 1.2);
- polymerization[38];
- photodynamic therapy for cancer treatment[39];
- synthesis of chemical compounds [40].

Furthermore once a molecule has been photoactivated, reaching an electronic excited state, it can transfer its excess of energy to an adjacent molecular system. This phenomenon is called photosensitization and will be discussed in details in section 1.3.

## 1.2 Spectroscopies

One of the most common way to study photochemical and photophysical properties of light driven processes is to analyze their energetic spectrum of absorption or emission upon irradiation. The methods employed in such aim are called spectroscopies and they can be sorted depending on the energy range that is considered (Table 1.1). In fact, depending on the field of research, this energy might either be depicted as a wavelength, or a frequency. This relation is given by the Planck-Einstein equation:

$$E = h\nu = \frac{hc}{\lambda} \quad (1.1)$$

where E is the energy (J), h the Planck constant ( $6.63 \times 10^{-34}$  J.s), c the celerity of light ( $3 \times 10^8$  m/s) and  $\lambda$  the wavelength (m).

Wavelength	Spectrum region	Method
$> 10^7$	Radio-waves	Nuclear Magnetic Resonance (NMR)
$10^5 - 10^7$ nm	Microwaves	Rotationnal spectroscopy
$800 - 10^5$ nm	Infrared (IR)	near-IR and IR spectroscopies
$100 - 800$ nm	UV-visible light	UV-visible spectroscopy
$0.01 - 100$ nm	X rays	X rays spectroscopy
$< 0.01$ nm	$\gamma$ rays	$\gamma$ spectroscopy

Table 1.1: Principal spectroscopic methods and their energy ranges.

We will focus here on UV-visible absorption, fluorescence and circular dichroism since they are the main spectroscopic tools employed to study most phenomenons related to photosensitization. Indeed, they offer the possibility to obtain information about molecular excitations, ligand binding, structural modifications, energy transfers ... [41, 42]

### 1.2.1 UV-visible absorption

As described in Figure 1.1 (black arrows), when a sample containing light sensitive molecules is hit by electromagnetic waves, especially in the ultra-violet and visible wavelength range, absorption of energy will occur resulting on electronic excitations (namely electronic transitions from lower occupied energy levels to higher's). The

absorption efficiency of the irradiating beam by the sample will depend on the vicinity between its energy and the energetic excitation gap of the excited state energy levels.

The absorption intensity ( $I$ ) of a light beam of wavelength  $\lambda$  traveling through a given sample of length ( $l$ ) will depend on the absorption coefficient ( $\epsilon$ ) and molar concentration ( $C$ ) of the molecules it contains, as well as the initial intensity ( $I_0$ ):

$$I = I_0 e^{-\epsilon' C} = I_0 10^{-\epsilon C} \quad (1.2)$$

with

$$\epsilon' = \epsilon \ln(10) \approx 2.303 \epsilon \quad (1.3)$$

The value of the absorption coefficient  $\epsilon$ , also known as molar attenuation coefficient mainly depends on the molecules in the sample and the wavelength of the excitation beam.

From this, the Beer-Lambert law states that the absorbance ( $A$ ) depends linearly on the molar concentration of the sample, for a given length ( $l$ ), such as:

$$A = -\log \frac{I}{I_0} = \epsilon' l C \quad (1.4)$$

Experimentally, the principle of UV-visible spectroscopy requires to have a cuvette located inside a spectrophotometer and containing the studied sample, then a continuous light of known intensity  $I_0$  is shone through it and the resultant light of intensity  $I$  is recorded. Using equation 1.4, the ratio between  $I$  and  $I_0$  provides the absorbance of the solution.

For example, it is common to use UV-visible absorbance titration to observe the interaction of a molecule with DNA [43]. For a given concentration of studied molecule, an increasing amount of DNA is added in a solution and the resulting spectra peaks shifts will indicate to formation of aggregates. In the case of intercalation of metal complexes in DNA, hypochromism and bathochromism shifts are commonly observed due to the formation of  $\pi$ -stacking interactions between the ligand and the nucleobases [44].

Nevertheless, the Beer-Lambert law does not always prevail. Indeed, in the case of high molar concentrations, the absorption will be saturated and even higher concentrations will not induce any modification of the spectrum absorption bands. This is due to a 100% absorption of the incident beam by the solution. Moreover, a non-homogeneous solution will also not be ruled by the Beer-Lambert law



since the concentration can not be properly evaluated. And finally, the absorbance of molecules that react and are denatured under UV-visible light will also not be properly described since the studied compound has reacted during the irradiation.

To go even further, some molecules have the ability to absorb not one, but two photons simultaneously (and even more in certain cases), this is called Two-Photon Absorption (TPA) [45]. In that case the  $S_0 \rightarrow S_1$  transition is done absorbing two photons of lower energy which sum equals the energy gap (Figure 1.2).

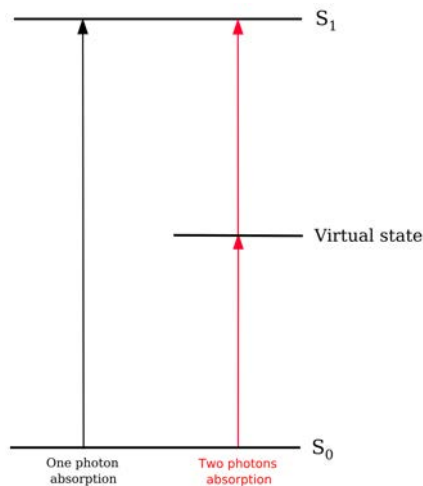


Figure 1.2: Jablonski diagram of the  $S_0 \rightarrow S_1$  transition through one and two photons absorption.

In the field of photosensitization and photodynamic therapy TPA is a phenomenon that proves to be highly promising since it allows the activation of light driven processes (such as the production of  $^1O_2$  [46], discussed in 1.4) in the therapeutic window which spans between 650 and 1350 nm, that is the near-infrared region. In this interval of energy the tissues does not absorb thus the probability to reach and photosensitize deeper lesions is much higher.

The particularity of TPA is that compared to one photon absorption, it is a non-linear optical process. The Beer-Lambert rule does not apply here and the absorption intensity becomes:

$$I = \frac{I_0}{1 + \beta C \sigma I_0} \quad (1.5)$$

where  $\beta$  is the TPA extinction coefficient (similar to  $\epsilon$  in one photon absorption),  $C$  the molar concentration and  $\sigma$  the cross-section. This last variable is a quantity in GM unit (Goeppert-Mayer) that allows to quantify the TPA efficiency.

### 1.2.2 Circular dichroism

The principle of Circular Dichroism (CD) relies on the fact that optically active chiral chemicals absorb left and right circularly polarized light with different intensities. The resulting signal that is measured is the difference between the absorbance of left and right polarized light:

$$\Delta A = A_G - A_D \quad (1.6)$$

The rotatory strength correspond to the coupling between the electric and magnetic dipoles:

$$R_{ab} = Im(\langle \Psi_a | \mu | \Psi_b \rangle \cdot \langle \Psi_b | m | \Psi_a \rangle)$$

CD is a technique routinely employed to study secondary structures and structural modifications in bio-macromolecular systems such as proteins and nucleic acids arrangements [47–49]. Indeed, even though DNA is composed of non-chiral residues (nucleotides), its global double helical form gives rise to a specific CD signal. For example, it is possible to distinguish between different sequences of DNA [48], or even G-quadruplexes polymorphisms [50]. In the case of proteins, it is the specific arrangement of the protein secondary structure that gives rise to the CD signal. In that case, 3 main structures can be defined, alpha helices, beta sheets and coils having all distinguishable specific CD spectra [49].

### 1.2.3 Fluorescence

Fluorescence as described in Figure 1.1 is the emission of photons by a molecular system in an excited state decaying to a lower energy state. Moreover, due to internal conversion phenomena, the Kasha’s rule states that the transition will often be from the first excited state  $S_1$  toward the ground state  $S_0$ .

As UV-visible absorption, it can be employed to study the interaction of ligands with biomolecules such as DNA and more specifically [51], it is able to monitor electronic phenomena such as electron transfers [52]. Indeed, the signal of fluorescent molecules can be quenched (deactivated) by the interaction with the environment, by example through the electron transfer from a guanine nucleobase [53].

### 1.3 Photosensitization

In the field of biology and biochemistry, the reaction of a photoactivated compound with an adjacent molecular system, such as DNA, membranes or proteins, is called photosensitization [54], (it allows a system that is not sensitive to light to still indirectly be activated and modified through an energy vector molecule). Such molecule is called a photosensitizer, or sometimes simply sensitizer[55]. Since its discovery, photosensitization received high scientific interests and developments due to the fact that it is the main mechanism involved in many cancers and tissues diseases productions but also in their treatments[2] (see photodynamic therapy described in 1.4).

Usually, and especially in phototherapy, the main underlying photophysical phenomenon of photosensitization is the population of the drug's triplet state. Indeed, two main pathways exist in biological environments:

**Type I** the driving force of this photosensitization is the direct electron transfer toward the sensitizer. Most of the time it is described as the activation of solvent or oxygen molecules by electron transfer ( $O_2 \rightarrow O_2^-$  and  $sensi \rightarrow sensi^+$ ), which will lead to the production of oxidizing reactive species. In fact this sensitization can be even more direct since if the environment is hypoxic (absence of oxygen molecules), the electron can be transferred to or from the substrate itself, directly producing endogenous reactive radical species. Since an oxidization occurs in type I sensitization, the active molecule reacts and does not necessarily relaxes to its fundamental state, in contrast of type II photosensitization where the photosensitizer can be regenerated [56].

**Type II** the sensitizer is activated first to a singlet excited state, then populates a triplet state through intersystem crossing. Finally its energy is transferred to a nearby oxygen molecule which activates its singlet reactive form ( $^3O_2 \rightarrow ^1O_2$ ). The energetic levels of this reaction are presented in Figure 1.3. Most of the current photosensitization therapies are achieved through type II oxygen activation employing porphyrins molecules [56–59].

In biological environments, the species that are eager to be photosensitized by triplet oxygen are mainly unsaturated molecular entities:

- Nucleic acid bases of DNA and RNA [60];
- lipids of cellular membranes[29];
- amino acids in proteins (tryptophan, tyrosine, histidine, methionine and cysteine) [61];

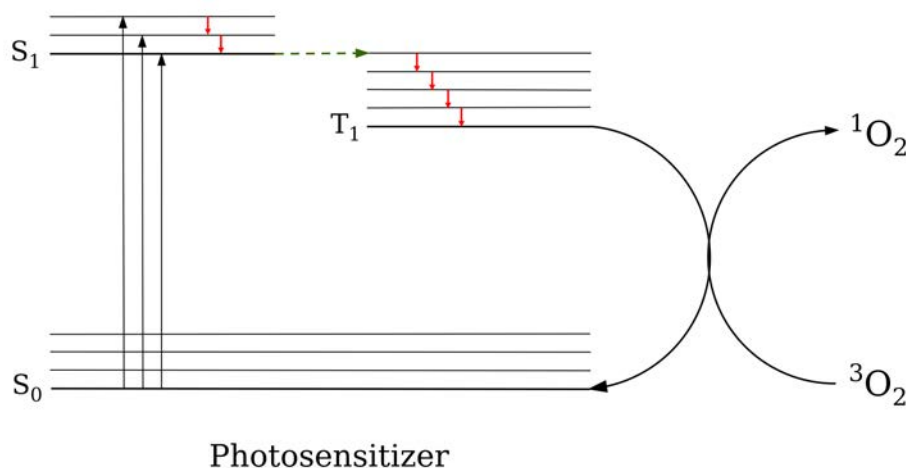


Figure 1.3: Jablonski diagram of the type II photosensitization mechanism.

- other cofactors (retinal in the rhodopsin proteins) [62].

The research for new photosensitizing agents is a highly active field since the ideal case in which the molecule would be completely specific, with a triplet quantum yield of 1 and no side effects to the human body does not exist yet. The most promising compounds that are currently being developed are composed of metal nanoparticles and sensitizer that have more complex photoactivation pathways (such as TPA), they shall provide in the future great improvements in the field of PhotoDynamic Therapies (PDT) [63–65].

## 1.4 Photodynamic therapy

Photodynamic therapy (PDT), also called photochemotherapy is a medical treatment to treat localized living tissue damages (mainly cancers, tumors and skin diseases) through the use of photosensitizing agents that are activated by non-ionizing light sources. The first depictions of the use of light to heal human afflictions, *heliotherapy*, dates back to 3000 years ago when Greeks and Egyptians described the exposure of the human body to sunlight as a treatment. In our current modern society, this therapy evolved to become what is called phototherapy. In that case the direct action of UV light has a rather poor efficiency and only the skin can be treated in such way. Indeed, UV-vis light does not penetrate deep into living tissues and biomolecules are relatively stable to light irradiation.

PDT, through the use of a photosensitizer is much more efficient than *phototherapy*. It has been discovered by serendipity in 1898 by a student of the university of Munich who observed that the molecule acridine could function as a toxic photosensitizer since its action depended on the light source and its time of exposure. Later on, in 1903, the first PDT treatments were performed on humans, treating lupus,

skin afflictions and cancers using eosin as a sensitizing agent.

Along the development and evolution of the photosensitizing molecules, porphyrins [66], that are highly conjugated cyclic compounds, were found to be the best candidates for photosensitization. Indeed, their high production ratio of reactive  $^1\text{O}_2$  [67], their relatively high biocompatibility [68] (with a central chelated  $\text{Mg}^{2+}$  it is known as chlorophyll and is fundamental for plant respiration, with a central  $\text{Fe}^{2+}$  it is called heme and carries oxygen in our blood vessels, see Figure 1.4) and their light absorption in the red window [69] (where the tissues are mainly transparent) enhanced the interest toward these compounds in the development of new photosensitizers. Nowadays the most common PDT sensitizers are haematoporphyrin and its derivatives (Figure 1.4).

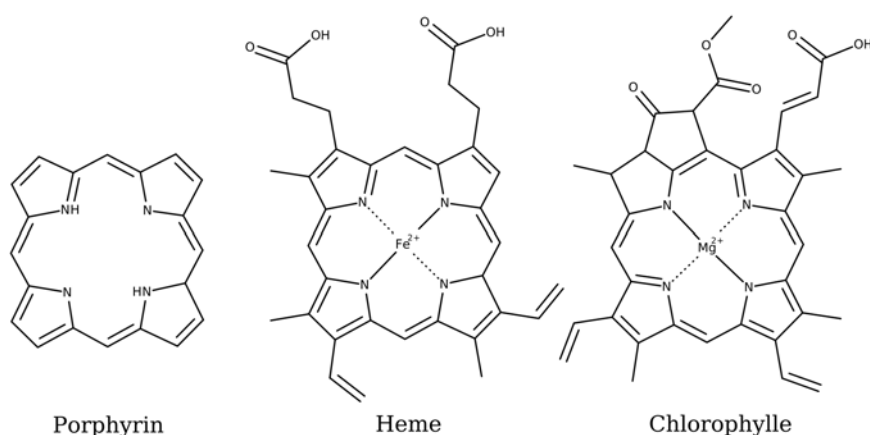


Figure 1.4: Main structure of the porphyrin photosensitizer and two derivatives, the heme and chlorophyll molecules

The main features of an efficient photosensitizer are:

- Absorption in the therapeutic window and a high quantum yield;
- biocompatibility;
- first triplet state energy higher than the one of singlet reactive oxygen;
- selectivity to tissues and biological entities;
- narrow absorption band.

The research for new photosensitizing agents is wide and spans every features presented above. Usually type I and type II photosensitization reactions coexist for most drugs on the market with a ratio depending on their photophysical properties. Still, the most common pathway is type II and in presence of  $\text{O}_2$ .

In the last ten years, new highly promising properties have been developed for PDT agents. Indeed, the last generation of photosensitizers is now combined with

either saccharides to enhance the specificity of their carriage or with peptide group to improve their cell uptake. Moreover, a new generation of sensitizers that would have a TPA activation is currently under study by many research groups [70, 71]. It would increase the depth of treatments in living tissues and its resolution in space [72]. Finally, improvements have been obtained on the life time of the triplet state for Ruthenium complexes, thus increasing the activation of  $^1O_2$  which greatly enhanced its efficiency to induce cancerous cells death [73, 74].

# Chapter 2

## DNA

### 2.1 Introduction

Deoxyribonucleic acid (DNA) is without doubt the most famous and important biological macromolecule. It stores the genetic code necessary for the creation and reproduction of every living organism.

Most of DNA is located inside the cell nucleus (nuclear DNA) but it is also found in the mitochondria (mitochondrial DNA). As represented in Figure 2.1, in the cell, DNA is most of the time coiled around histone proteins to form nucleosomes. This process of compaction is necessary since the 3.2 billion DNA bases would have a total length of approximately 11 meters. To go even beyond, nucleosomes adopt a relative macromolecular arrangement to form chromatin. Furthermore, these chromatin ensembles are again compacted to form the famous chromosomes.

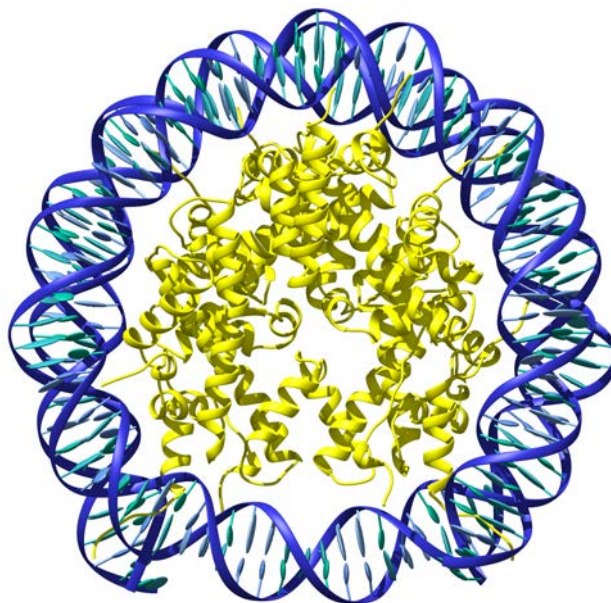


Figure 2.1: DNA double strand wrapped around a histone protein. Protein Data Bank (PDB): 5AV5 [75].

## 2.2 Fundamental building blocks of DNA

DNA is a biopolymer made of monomers called *nucleotides* which are composed of three organic chemical entities:

- Pyrimidines (thymine, cytosine) and purine (adenine, guanine) nitrogen bases;
- pentose sugars (deoxyribose in DNA);
- phosphate ( $\text{PO}_4^-$ ) groups.

As presented in 2.2, the pentose sugar works as an anchors to fix the base to the backbone of the polymer. Then the nucleotides are conventionnaly arranged in a 3' - 5' sequence where the phosphate group of each monomer (attached to the C5' of the sugar) is linked to the C3' of the following sugar moiety.

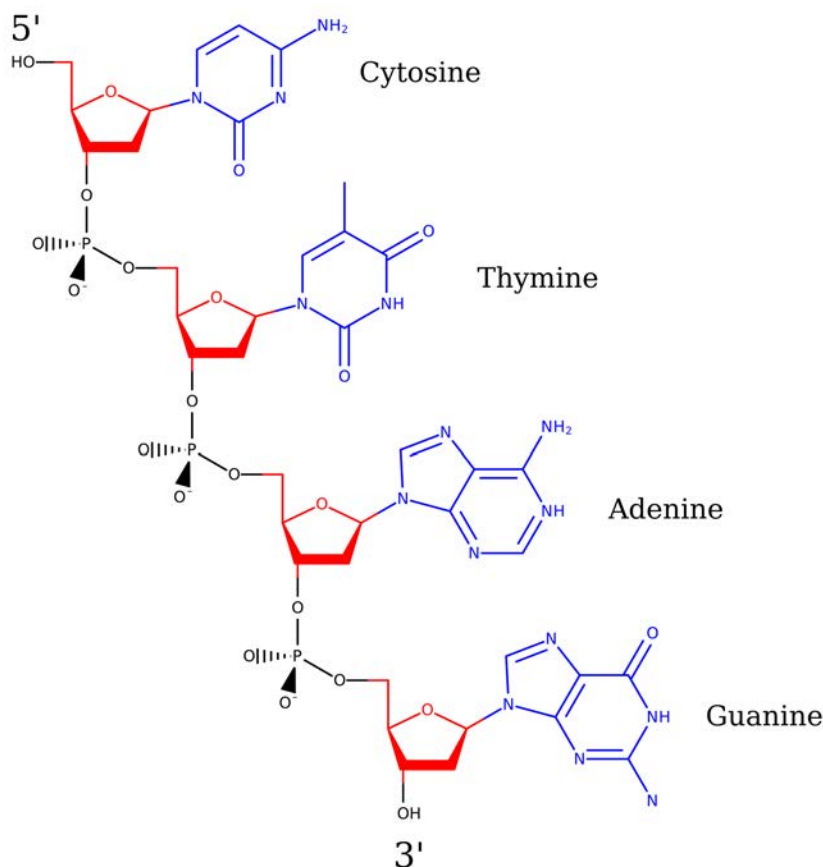


Figure 2.2: The four nucleotides of DNA in their 3' - 5' orientation, in red the deoxy-pentose sugars, in blue the bases and in black the phosphate bridges.

The purinic and pyrimidinic bases are, in DNA, always associated pairwise, cytosine with guanine and thymine with adenine (see 2.3). These couples, called *nucleobases*, can be considered as the four letters of the genomic alphabet (A...T, T...A, G...C, C...G).



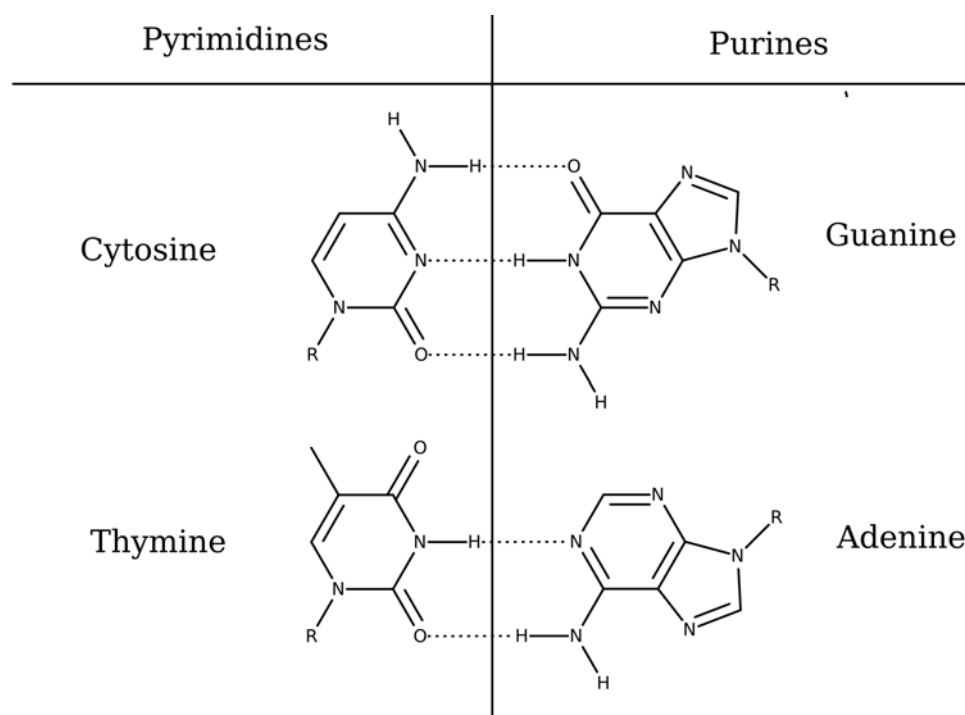


Figure 2.3: Watson-Crick hydrogen bonding pattern between adenine and thymine, and cytosine and guanine.

## 2.3 Double helix

The first successful resolution of the DNA structure has been realized by Francis Crick and James Watson in 1953[76]. Using X-ray diffraction of crystals of DNA, they discovered the particular right handed double helical arrangement of the two nucleic acid strands. One should not forget to mention Rosalind Franklin who acquired the data and hence should be glorified for this discovery as much as her colleagues.

As depicted in Figure 2.4, double helices of DNA are known to exist in three main arrangements, the A-, B- and Z-form. The following table summarizes their main characteristics:

Table 2.1: Main parameters of the three helical polymorphisms of the DNA double strand, A-, B- and Z-DNA.

Geometric property	A-DNA	B-DNA	Z-DNA
Helicity	right handed	right handed	left handed
Base pair per turn	11	10.5	12
Rise between base pairs (Å)	2.6	3.4	3.7
Diameter (Å)	23	20	18

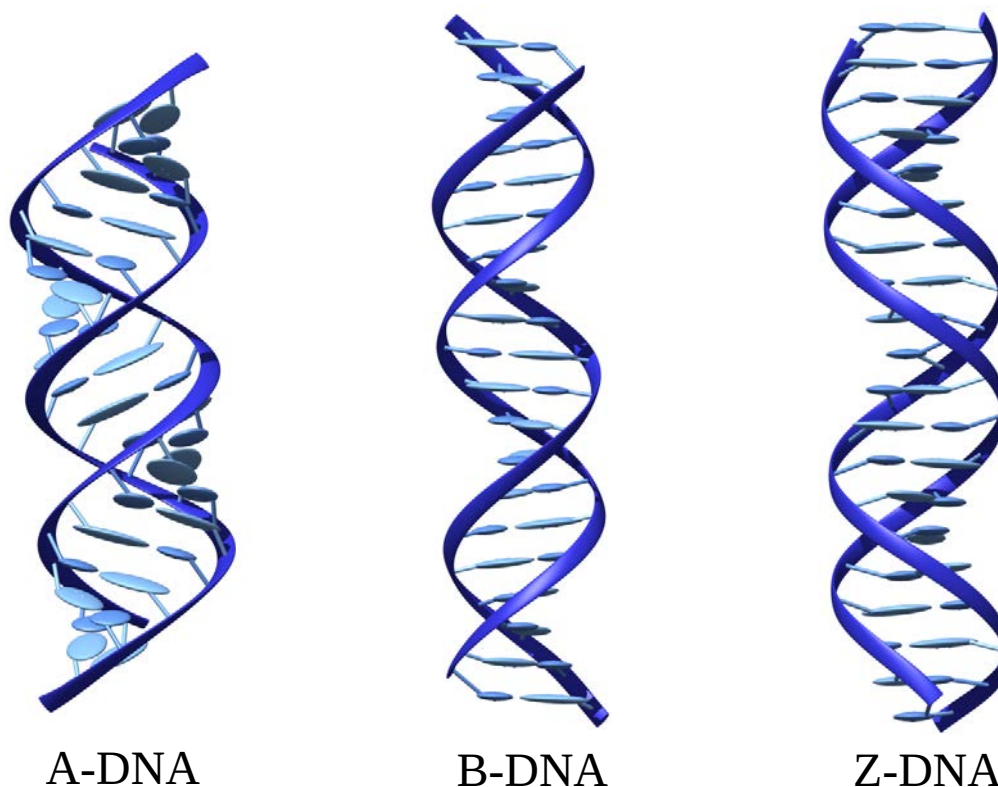


Figure 2.4: The three main polymorphisms of double helical DNA, A-DNA, B-DNA and Z-DNA.

The overall DNA double helical structure is stabilized by 3 main chemical interactions:

**Hydrogen bonds** DNA is composed of two non covalently linked strands which cohesion is mainly maintained due to hydrogen bonding patterns between the two bases. The number of hydrogen bonds in base pairs depends on the base pair, there are two for adenine-thymine and three for cytosine-guanine as depicted in Figure 2.3 [77].

**$\pi$ -stacking** Thanks to the aromaticity of the four DNA bases and their internucleobases arrangements, their stacking creates a stabilizing interaction made of hydrophobic effects and Van der Waals forces called  $\pi$ -stacking [77].

**Charge-charge interactions** Since both backbone are highly negatively charged due to the phosphate groups, the minor groove region would be highly destabilizing, but in reality, it can contain positively charged chemicals or polar interacting molecules ( $\text{Mg}^{2+}$  and other biological ions, water solvent molecules, lysine amino-acids that are used as anchors for the numerous proteins and enzymes interacting with DNA) thus this -/+/- alternating charges interaction becomes stabilizing [78].

## 2.4 DNA structural parameters

Even though DNA in its double helical form is highly flexible, an ensemble of well defined parameters can be employed to efficiently and completely describe its structure and behavior. Indeed, these parameters can either characterize nucleotides (intra base pairs parameters), stacked base pairs (inter base pair parameters) and grooves (width and depth of the major and minor grooves). All schemes of parameters are given in Appendix A.

These parameters can be efficiently obtained by accessible tools such as Curves+ [79] and 3DNA [80]. The information that can be extracted from such data are of high interest in the study of compounds interacting with DNA, global and local deformations of the strand, interaction with proteins and even more specific phenomena such as base flipping for example.

## 2.5 DNA photosensitivity

Since it is present in every cells of every living animal, DNA is facing the influence of sunlight [81]. Indeed, the electronic aromaticity of the bases causes their highly efficient light absorption in the UV, with a maximum at approximately 260 nm [82]. Their absorption spans the UVC region as well as all the UVB wavelengths, and also have a small absorption probability in the UVA due to excitonic couplings. In order to prevent the photoreaction and degradation of the nucleobases, thus inducing mutations or cell death, the chemical structure of the absorbing nitrogen bases evolved toward a highly photostable ensemble. Photostability in DNA is assured through several mechanisms:

**Fluorescence** The nucleobases excited states decay at the subfemtosecond time scale [83]. The direct fluorescent deexcitation has, both in isolated nucleobases or in DNA, a low quantum yield, respectively  $10^{-4}$  and  $3 \cdot 10^{-4}$ . This means that the excess of energy is mainly released in a non radiative way through IC and ISC.

**Non-radiative deexcitation** It has been demonstrated that the non-radiative relaxation of nucleic acids excited states differs between monomers and mono- or double-strands. In the case of monomers, the  $S_0 \rightarrow S_1$  transition of  $\pi - \pi^*$  nature will either decay directly through a conical intersection between the two potential energy surfaces (60%) or evolve toward a  $^1n - \pi^*$  state followed by a ISC to the triplet state (picosecond timescale) [84–86]. In the case of base stacks and double strands, charge transfer states and charge recombination is considered to play the highest role in the decay toward  $S_0$ ;

**Charge recombination** When the excited nucleobase is embedded in a single or double strand, the  $\pi$ -stacking interactions and hydrogen bonds of the Watson-Crick arrangement play a modulating role in the excited-states lifetimes. Indeed, for single strands an electron can be transferred to the adjacent base and induce the formation of radical ions. Then charge recombination prevents any further reactions and allows both bases to relax to their canonical states [23]. In double strands, it is the hydrogen pairing that plays the highest role. After excitation a charge separation may occur between the two nucleobases of the base pair followed by a proton transfer. Then, through conical intersections between the charge transfer state and the ground state, the system can decay non-radiatively to its ground state.

## 2.6 DNA photolesions

Even if the photostability of DNA is high and prevents the formation of lesions and damages, since it is composed of approximately 3 billion nucleobases, the occurrence of this phenomenon cannot be ruled out. For example, the increasing number of skin cancers is due to the direct exposure of the skin to UV light, and especially the photosensitization of DNA and the photochemical reactions of the nucleobases inducing photolesions. The main DNA lesions due to irradiation are the cyclobutane pyrimidine dimers (CPD), the 6-4 photoproduct (6-4PP) and the Dewar photoproduct (DewPP) [87, 88], represented in 2.5.

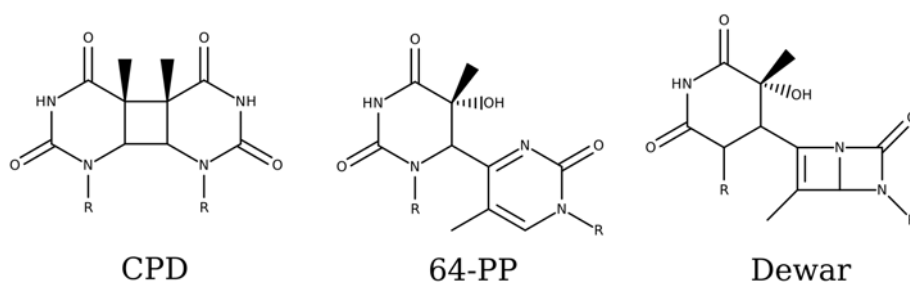


Figure 2.5: The three most common UV induced DNA lesions, CPD, 64-PP and Dewar product.

Type II photosensitization of DNA, via activation of singlet oxygen or reactive radical species can produce other carcinogenic and mutagenic lesions called oxidative lesions. In that case a tremendous amount of lesions have been found so far [89]. The most common ones being 8-oxo-guanine, 5-hydroxyhydantoin, FapyGua (2,6-diamino-4-hydroxy-5-formamidopyrimidine) and even DNA-protein crosslinks (Figure 2.6) [90].

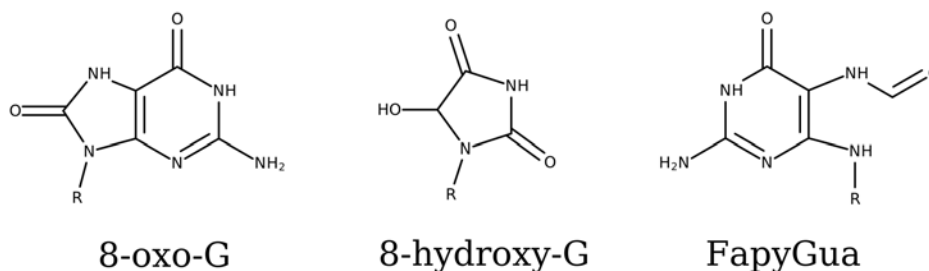


Figure 2.6: The three most common oxidative nucleobases lesions.

## 2.7 Repair processes

The main feature of the efficient repairing processes of DNA is based on the fact that its nucleobases are composed of two complementary strands that carry the same information, and if one of the two nucleotides is missing or damaged, the repairing mechanism only needs to remove the non canonical nucleotide and replace it with the complementary one. On one side, the repair mechanism can be direct (without having to break the backbone to remove the lesion and replace it by a standard nucleotide), this is called direct reversal mechanism. Here an enzyme called photolyase recognizes a photolesion (for example CycloPyrimidine Dimers, CPD) and reverses it to the two canonical pyrimidines. This mechanism specific of bacteria, fungi and some animals does not exist anymore in humans. Instead, the repairing process of a lesion is done by two excision mechanisms, Nucleotide Excision Repair (NER) and Base Excision Repair (BER)[91].

**NER** It repairs bulky lesions, such as the photolesions CPD and 6-4PP that have an important influence on the DNA structure. The complete mechanism is quite complex, involving 9 different proteins or enzymes and has two main pathways, either a global genomic (GG-NER) that repairs in both transcribed and untranscribed DNA strands and a transcription coupled (TC-NER) that only repairs lesions that prevents the transcription of DNA.

**BER** This mechanism has evolved to tackle a large number of non bulky base lesions. The complete repairing pathway possesses different reaction routes depending on the lesion, thus involving different enzymes. It is mainly executed in 5 steps (see Figure 2.7):

- The mismatched base or the lesion is recognized and removed by a DNA glycosylase. It creates an abasic site;
- an AP endonuclease incises the phosphodiester group;
- a lyase enzyme removes the dangling phosphodiester;

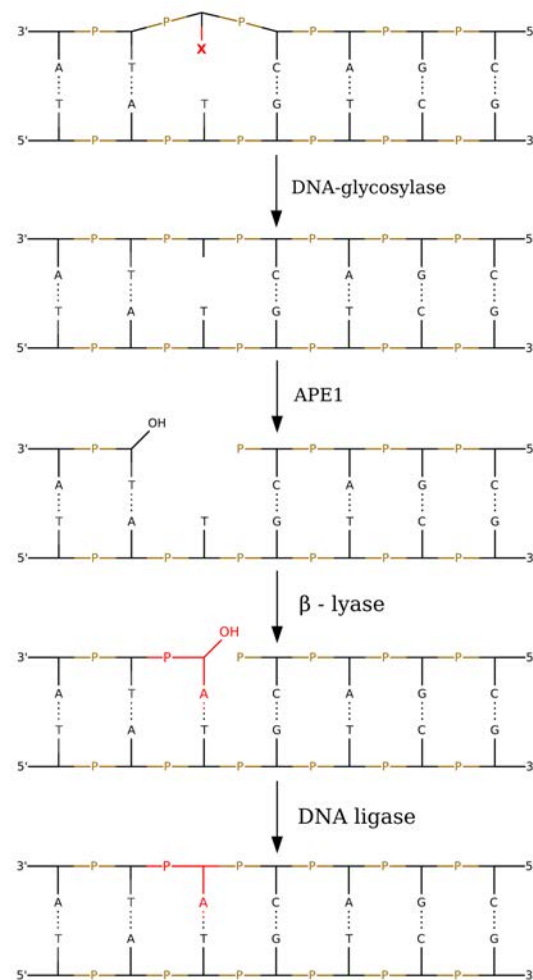


Figure 2.7: BER mechanism, the mismatched base or lesion is indicated by **X**.

- a DNA polymerase ( $\text{Pol}\beta$ ) inserts the corresponding missing nucleotide ;
- the 3' part of the new nucleotide is linked to the previous 5' phosphate group of the strand.

A defect in these repair mechanisms induces high sensibility to UV irradiations and leads most notably to immunological and neurological disorders.

# **Part VII**

## **Methods**

# Chapter 3

## Molecular Mechanics

In the framework of the Newtonian's classical mechanics, Molecular Mechanics (MM) is a method to obtain the energy of molecular systems given an ensemble of atomic positions, masses and charges. The interactions between atoms are here represented by a set of harmonic force constants (springs) and the potential energy surface by what is called a classical Force Field (FF). These interactions can be of bonded (bonds, angles, dihedrals and improper dihedrals) or non-bonded (Coulomb and van der Waals) natures and the potential energy surface  $V_{MM}$  of a molecular system composed of  $N$  atoms with corresponding coordinates  $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$  can be represented by the addition of each term such as:

$$V_{MM}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = V_{bonds} + V_{angles} + V_{dihedrals} + V_{impropers} \dots \\ \dots + V_{vanderWaals} + V_{electrostatics} \quad (3.1)$$

where the first three terms can be written as:

$$V_{bonds} = \sum_{bonds} k_r (r - r_0)^2 \quad (3.2)$$

$$V_{angles} = \sum_{angles} k_\theta (\theta - \theta_0)^2 \quad (3.3)$$

$$V_{dihedrals} = \sum_{dihedrals} \frac{V_n}{2} (1 + \cos(n\phi - \gamma)) \quad (3.4)$$

where  $k_r$  and  $k_\theta$  are the force constants describing the harmonic potential,  $r_0$  and  $\theta_0$  the central value of this potential,  $V_n$  the dihedral force constant of angle  $\phi$  and phase  $\gamma$ .



An improper term is also applied for some very specific cases when the planarity of a given molecular function should be kept. For example due to electron delocalization in rings or peptide bonds.

Then the non-bonded terms accounting for long-range interactions are described as on one hand a Coulombic potential between atomic charges and on the other hand a 12-6 Lennard-Jones potential for van der Waals. They are written as:

$$V_{vanderWaals} = \sum_i \sum_{j>i} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (3.5)$$

$$V_{electrostatics} = \sum_i \sum_{j>i} \frac{q_i q_j}{\epsilon r_{ij}} \quad (3.6)$$

where  $A_{ij}$  and  $B_{ij}$  are parameters of the 12-6 Lennard-Jones potential. This potential is employed to reproduce the Morse like potential of the van der Waals interactions between two atoms. The attractive part of this function is varying in  $1/r^6$  when the repulsive part is varying in  $1/r^{12}$ . This simplified way to compute such a complex interaction is mainly employed because of its low computational cost, indeed the repulsive part has the same mathematical expression as the attractive part, hence they can be obtained simultaneously. Finally, the electrostatic potential corresponds to the Coulomb interaction between the charges of the two atoms, with  $q_i$  and  $q_j$  the respective charges of particles  $i$  and  $j$ ,  $r_{ij}$  the distance between them and  $\epsilon$  the dielectric constant of the medium.

# Chapter 4

## Molecular Dynamics

As presented previously, MM methods have a clear computational interest since they allow to study large molecular systems such as proteins or nucleic acids arrangements. But its statical description of the energy neglects temperatures and time evolutions which is of a main importance to study biomolecular phenomenons. That is why, in order to compute motions and thermodynamic properties, Molecular Dynamics (MD) has been developed. It relies on the integration of the classical Newton's second law of motion and forces/velocities will be generated at a step-wise manner using algorithms described in [92]. MD is nowadays the main tool to study large scale molecular systems over biologically and mechanically relevant time intervals. Moreover, thanks to a constant and fast improvement of computing capacities, of the algorithm efficiencies, and recently the use of graphic processor units (GPU), MD went from its first biological test case of a protein in vacuum [93] toward nowadays standard studies reaching  $\mu s$  of computations for systems containing tens of thousands atoms. It even reached the famous all-atom HIV capsid (64 millions of atoms) simulation lasting some hundreds of nanosecond, performed on a Blue Waters supercomputer [94].

MD is a multidisciplinary field of research since it gives structures and properties information in chemistry (aggregate formation, surface interaction ...), biology (proteins, DNA structures, peptide folding, ligand binding ...) and material sciences (polymers, structures after impact ...) [13, 95].

### 4.1 Integration of equations of motion

Molecular dynamics is based on classical Newton's laws of motion, especially the second one:

$$\frac{\partial^2 x_i}{\partial t^2} = \frac{F_{x_i}}{m_i} \quad (4.1)$$

$$m_i \frac{\partial^2 x_i}{\partial t^2} = -\frac{\partial V_{x_i}}{\partial x_i} \quad (4.2)$$

where  $x_i$  is the position of the atom  $i$  of mass  $m_i$  submitted to a force  $F_{x_i}$  for a particle moving in one dimension. Moreover, this force can be determined by the calculation of the first differential of the potential energy equation given by the force field formula 3.1.

In most molecular dynamics softwares, the molecular dynamic simulations are conducted by numerically solving the differential equations of motions with finite difference methods. The concept of these methods is to split the integration into small intervals of a fixed time length  $\partial t$ . Then, the total force applied on each atom for a given configuration at a time  $t$  is computed as the vector sum of its interactions with every other atom. Now that the forces are generated, the accelerations can be easily obtained. And finally, with positions, velocities and accelerations at time  $t$  the system can evolve toward the next step at time  $t + \partial t$  and so on.

In order to achieve a better stability, the Amber software employs a common algorithm for simulations called *leapfrog* and is described as follows:

$$\mathbf{r}(t + \partial t) = \mathbf{r} + \partial t \mathbf{v} \left( t + \frac{1}{2} \partial t \right) \quad (4.3)$$

$$\mathbf{v} \left( t + \frac{1}{2} \partial t \right) = \mathbf{v} \left( t - \frac{1}{2} \partial t \right) + \partial t \mathbf{a}(t) \quad (4.4)$$

In this algorithm the velocities at a time  $t + \frac{1}{2} \partial t$  are calculated from the ones at time  $t - \frac{1}{2} \partial t$  and the accelerations at time  $t$ . And from equation 4.3, the positions at time  $t + \partial t$  are generated from the previous velocities and the positions at time  $t$ . And finally to synchronize the positions and velocities (respectively calculated at times  $t + \partial t$  and  $t + \frac{1}{2} \partial t$ ), the velocity at time  $t$  is determined as follows:

$$\mathbf{v}(t) = \frac{1}{2} \left[ \mathbf{v} \left( t + \frac{1}{2} \partial t \right) - \mathbf{v} \left( t - \frac{1}{2} \partial t \right) \right] \quad (4.5)$$

## 4.2 Periodic Boundary Conditions (PBC)

The computational cost of MD simulations is dependent on the numbers of particles. If one wants to consider a protein in solution, the number of particles that should be taken into account in the calculation compared to the size of the protein would be basically infinite, thus leading to impossible simulations. In order to bypass this

issue, the system can be placed in a finite sphere of explicit solvent molecules, but then important border effects will appear if the solute reaches the edges of the sphere during the MD run leading to non-meaningful results. So to avoid the border effect problem and still reduce the computational cost of the simulations using reasonable solvation box sizes, one can rely on the use of the so called Periodic Boundary Conditions (PBC), Figure 4.1. The aim of this tool is to surround a box of explicit particles, with replicated images of itself in all directions to give a periodic array. The principle of this method, described in 2D on the Figure 4.1, is to replace a particle, leaving the box in one side, by an identical particle in the opposite side, thus, the number of particles remains the same.

The simplest periodic system is the cubic box, surrounded by 6 similar replica, but it is possible to reduce the cost of the simulation by using a more complex shape. In the studies related to this manuscript, the structure of the periodic box is a truncated octahedron. It is surrounded by 8 neighbors and have half of the volume of a cubic box which reduces much more the costs of simulations. The only inconvenience of this method is that the system has to be strictly smaller than the length of the box, otherwise an overlap between the solute molecules in the periodic cells will occur resulting in non-physical results.

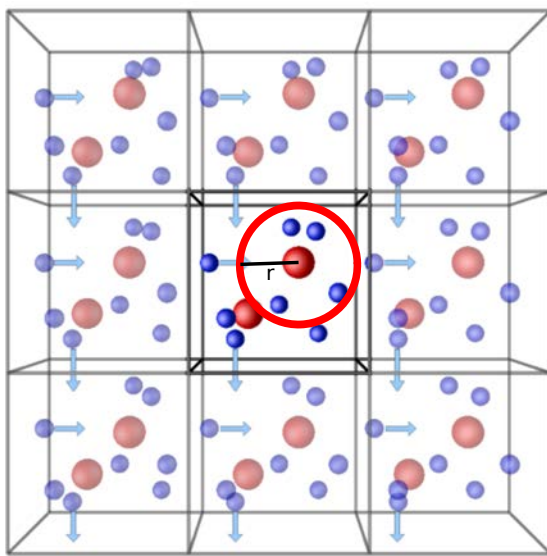


Figure 4.1: Representation of the periodic boundary conditions for a box containing an ensemble of particles. Readapted from <http://isaacs.sourceforge.net/phys/pbc.html>.

The calculation of long range interactions (Coulomb and van der Waals) will be done between every pair of atoms if no other simplification is added to the simulation. One can easily understand that in this case, increasing the number of particles in a system will quickly increase the computational cost. A first approximation would be to add what is commonly called a *cut-off* of a radius  $r_C$  as represented as a red circle

in 4.1. Here the interactions will be turned off for atom to atom distances greater than this value chosen value. Since the Lennard-Jones potential is decreasing as  $1/r^6$  for van der Waals interactions, this approximation does not add a significant error to the calculation. But in the case of Coulombic interactions which decreases at the  $1/r$  rate, the final error would be too important to be neglected, thus another methodology must be employed that is called Particle Mesh Ewald (PME).

### 4.3 Ewald summation and Particule Mesh Ewald (PME)

Based on the Ewald summation that allows the computation of charge and dipole interactions in periodic systems, the so-called Particle Mesh Ewald (PME) algorithm has been developed to improve the efficiency of MD simulations and remove the bias created when a simple cut-off is used for Coulombic interactions.

The Coulombic potential  $V_C$  for  $N$  particles in a periodic box of size  $L$  is given by:

$$V_C = \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{4\pi\epsilon_0 |r_{ij} + \mathbf{n}|} \quad (4.6)$$

where  $q_i$  and  $q_j$  are the charges of the particules  $i$  and  $j$  that are separated by distance  $r_{ij}$  and  $\mathbf{n}$  is the translation vector of the periodic cell:

$$\mathbf{n} = (n_x L, n_y L, n_z L) \quad (4.7)$$

where  $x$ ,  $y$  and  $z$  are the Cartesian coordinates unit vectors.

The calculation of 4.6 can become a bottleneck for MD simulations since  $V_C$  converges extremely slowly. To overcome this issue, the Ewald summation is employed. It is a mathematical trick to rewrite the potential as a sum of terms converging faster. For the simplest case of a  $1/r$  interaction the Ewald summation gives:

$$\frac{1}{r} = \frac{f(r)}{r} - \frac{1 - f(r)}{r} \quad (4.8)$$

Here  $f(r) = \text{erfc}(\alpha r)$  where  $\alpha$  is called Ewald splitting parameter. This function must be chosen so that  $1/r$  will decay rapidly for small  $r$  and slowly for long  $r$ .

As presented in 4.2, each point charge of the lattice will be assigned in the direct space a first Gaussian function of same magnitude and opposite sign and in the reciprocal space a second one of same magnitude that will cancel the first one.

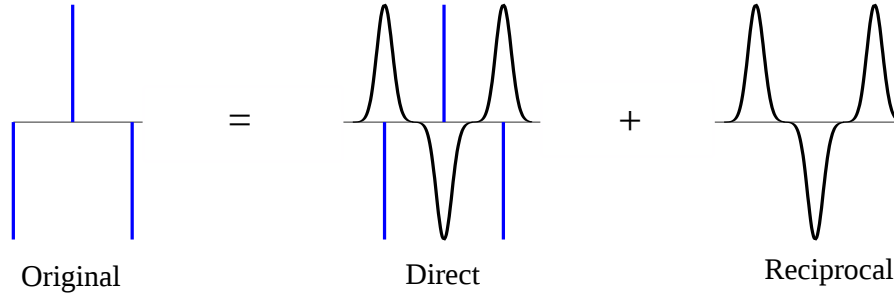


Figure 4.2: Representation of the Ewald summation procedure for three charges represented along an arbitrary axis.

Usually these Gaussian distribution functions are written as:

$$\rho_i(r) = \frac{q_i \alpha^3}{\pi^{(3/2)}} \exp(-\alpha^2 r^2) \quad (4.9)$$

Then the Coulombic energy potential can be separated in 3 distinct terms,  $V_{sr}$  the short range direct space contribution,  $V_{lr}$  the reciprocal long range Fourier space contribution and  $V_{self}$  which removes the interaction of each Gaussian with itself when these functions are summed. The formula 4.6 will be split as:

$$V_C = V_{sr} + V_{lr} + V_{self} \quad (4.10)$$

where,

$$V_{sr} = \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j \operatorname{erfc}(\alpha |r_{ij} + \mathbf{n}|)}{4\pi\epsilon_0 |r_{ij} + \mathbf{n}|} \quad (4.11)$$

$$V_{lr} = \frac{1}{2L^3} \sum_{\mathbf{k} \neq 0} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{k^2 \epsilon_0} \exp\left(-\frac{k^2}{4\alpha^2} \cos(\mathbf{k}) \cdot \mathbf{r}_{ij}\right) \quad (4.12)$$

$$V_{self} = -\frac{\alpha}{\sqrt{\pi}} \sum_i^N \frac{q_i^2}{4\pi\epsilon_0} \quad (4.13)$$

Here  $\mathbf{k}$  is the reciprocal vector as  $\mathbf{k} = 2\pi\mathbf{n}/\mathbf{L}$ .

## 4.4 Biased Molecular Dynamics

MD simulations in their standard formalism presented before only allow to sample conformational energy wells that are not separated by high energy barriers in the

global potential energy surface. This usually forbids to reach many states that would be of interest in the study of biological mechanisms (for exemple base flipping in DNA or protein reorganisation). The main strategy to bypass such problematic aspect is to add biases that forces to artificially sample a transition between two states, thus helping to overcome high energy barriers. The main methodologies that have been developed in that aim are:

**Potential of Mean Force (PMF)** The path between two states is divided in several windows (distance, angle, dihedrals for example), and each window is sampled by means of an umbrella sampling simulation. The system is forced to stay close to a chosen conformation. Finally, the ensemble of windows are recoupled to obtain the free energy path of the considered phenomenon (using utilities such as WHAM [96]).

**Replica exchange** Also called Parallel Tempering, is based on the exchange of configurations between several copies of the same system but at different temperatures. From this, configurations that requires higher temperatures to be reached can be obtained.

**Metadynamics** The method is used to explore the conformational space of a system by filling every free energy well so that the system will evolve by itself toward other states until it has travelled on the full free energy landscape [97].

**Alchemical transformations** Another methodology to obtain thermodynamical data and free energies of transitions between two configurational states is to performe a stepwise series of simulations through alchemical pathways. It is detailed in the following section.

## 4.5 Alchemical simulations

Since MD simulations are able to generate numerous ensembles of configurations, it is possible to obtain important thermodynamic information such as free energies.

The free energy difference  $\delta A$  between two states a and b can be obtained as an average over the ensemble of configurations obtained through long samplings of the two characteristic Hamiltonians of both states:

$$\delta A_{a \rightarrow b} = -k_B T \ln \left\langle \exp \left( -\frac{H_b(\mathbf{r}, \mathbf{p}) - H_a(\mathbf{r}, \mathbf{p})}{k_B T} \right) \right\rangle \quad (4.14)$$

where  $k_B$  is the Boltzmann constant,  $T$  the temperature and  $(\mathbf{r}, \mathbf{p})$  the coordinates and momentum. This approach is really convenient for systems in a thermodynamically stable conformation, but to obtain the free energy pathway between

states a and b, the coupling parameter  $\lambda$  is introduced. It will allow to sample over non-physical intermediate states and will connect from a value 0 for state a, until a value of 1 for state b. The requirement of such computation is to obtain a good convergence of the final results that are highly dependent on the quantity of sampled conformations (time length of the simulation) and the wideness of the windows between each intermediate states. It gives for an ensemble of k different configurations:

$$\delta A_{a \rightarrow b} = -k_B T \int_0^1 \ln \left\langle \exp \left( -\frac{H_b(\mathbf{r}, \mathbf{p}, \lambda_k) - H_a(\mathbf{r}, \mathbf{p}, \lambda_k)}{k_B T} \right) \right\rangle_\lambda d\lambda \quad (4.15)$$

Such approach is called Thermodynamical Integration (TI). Thanks to such methodologies, the free energy between two states with a completely unknown pathway can be obtained going through alchemical steps.

Alchemical methods, even though they are highly demanding in human and machine time, can provide results directly comparable to experimental data, thus, once validated, greatly helps in the understanding of many biological phenomena that cannot be reached and understood without simulation. For example binding of ligands in proteins or DNA, protein-DNA recognition . . . .



# Chapter 5

## Quantum chemistry

In the previous chapter we demonstrated the advantages of Molecular Mechanics and Molecular Dynamics to tackle large molecular systems, to study their structure and dynamic evolutions over time as well as their thermodynamical properties. But in these approximated methods, atoms are only described by masses and charges interacting through a series of springs, in fact the electrons are not considered explicitly. Such thing entirely removes the possibility to study electron mediated processes such as chemical reactions or the computation of excited states. That's why, Quantum Mechanics (QM) are used to take into account electrons explicitly. The following chapter aims to describe the methods from the scientific field of quantum chemistry employed in this manuscript, in term of electronic description of the fundamental state of molecules (Hartree Fock and Density Functional Theory) and the computation of their excited states giving access to spectroscopic properties (Time Dependent-Density Functional Theory).

Even if the determination of electronic properties of molecular systems seems a much more accurate solution than MM and MD for example, these calculations are not straightforward and require a huge amount of computation resources which becomes quickly a bottleneck when the size of the studied systems increases.

Quantum Mechanics is ruled by six postulates:

### 5.1 The six postulates of quantum mechanics

#### 5.1.1 Postulate 1: the wave function $\Psi$

Every system considered in the quantum mechanics framework will be described by what is called a wave function symbolized by the famous  $\Psi$ . This function is dependent to the coordinates  $\mathbf{r}$  of the particles and is dependent on the time  $t$ , written as  $\Psi(\mathbf{r}, t)$ . Moreover, from this function one can obtain the probability to find a particle in the elementary volume  $d\tau$  at the position  $\mathbf{r}$  and at time  $t$  using

$$\Psi^*(\mathbf{r}, \mathbf{t})\Psi(\mathbf{r}, \mathbf{t})d\tau$$

And logically, the probability to find one particle somewhere in a infinite volume will be equal to 1, this is called the normalization condition:

$$\int_{-\infty}^{+\infty} \Psi^*(\mathbf{r}, \mathbf{t})\Psi(\mathbf{r}, \mathbf{t})d\tau = 1 \quad (5.1)$$

### 5.1.2 Postulate 2: Hermitian operators

One can assign a QM Hermitian operator for every physical observable in classical mechanics. For exemple:

Table 5.1: Physical observables and their corresponding quantum operators.

Observable	Observable Symbol	Quantum Operator	Operation
position	$\mathbf{r}$	$\hat{\mathbf{r}}$	times $\mathbf{r}$
momentum	$\mathbf{p}$	$\hat{\mathbf{p}}$	$-i\hbar \left( \hat{i} \frac{\partial}{\partial x} + \hat{j} \frac{\partial}{\partial y} + \hat{k} \frac{\partial}{\partial z} \right)$
kinetic energy	T	$\hat{T}$	$-\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right)$
potential energy	$V(\mathbf{r})$	$\hat{V}(\mathbf{r})$	times $V(\mathbf{r})$
total energy	E	$\hat{H}$	$-\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(\mathbf{r})$

### 5.1.3 Postulate 3: eigenstates and eigenvalues

If one considers an observable  $A$ , the eigenvalues of the corresponding operator  $\hat{A}$  as defined in the third postulate will be the only possible results of the measurement of  $A$ . This satisfies the eigenvalue equation:

$$\hat{A}\Psi = A\Psi \quad (5.2)$$

### 5.1.4 Postulate 4: expectation of the wave function

The expectation value of the observable, for a given operator  $\hat{A}$  is:

$$\langle A \rangle = \int_{-\infty}^{\infty} \Psi^* \hat{A} \Psi d\tau \quad (5.3)$$

### 5.1.5 Postulate 5: Time dependent Schrödinger equation

The time dependence of a quantum system can be obtain by solving the time-dependent Schrödinger equation:

$$\hat{H}(\mathbf{r}, t)\Psi(\mathbf{r}, t) = i\hbar \frac{\partial}{\partial t}\Psi(\mathbf{r}, t) \quad (5.4)$$

For a stationary time-independent state, the Schrödinger equation simply reads:

$$\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \quad (5.5)$$

### 5.1.6 Postulate 6: wave function collapse

After observation, the wave function of a system can be reduced to a single eigenstate, which was initially a superposition of eigenstates.

## 5.2 The Schrödinger equation in quantum chemistry

In quantum chemistry, the particles characterized by wave functions are the nuclei of molecules and their electrons. In this case, if we consider a system composed of  $N$  electrons and  $M$  nuclei, the Hamiltonian of 5.5 is:

$$\hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{Ne} + \hat{V}_{ee} + \hat{V}_{NN} \quad (5.6)$$

in which  $\hat{T}_e$  is the kinetic energy term for the electrons,  $\hat{T}_N$  is the kinetic energy term for the nuclei,  $\hat{V}_{Ne}$  is the interaction term between electrons and nuclei,  $\hat{V}_{ee}$  is the repulsion term between electrons and  $\hat{V}_{NN}$  is the repulsion term between nuclei. These terms are decomposed as:

$$\hat{H} = -\sum_{i=1}^N \frac{\nabla_i^2}{2} - \sum_{A=1}^M \frac{\nabla_A^2}{2M_A} - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{r_{AB}} \quad (5.7)$$

with  $M_A$  the mass of nuclei  $A$  and  $Z_A$  its atomic number,  $i$  and  $j$  electrons with a distance between them of  $r_{ij}$ ,  $r_{iA}$  the electron nucleus distance and  $r_{AB}$  the distance between two nuclei.

## 5.3 Born-Oppenheimer approximation

Considering that the motion of the nuclei is by orders of magnitude slower than electrons, the Born-Oppenheimer approximation allows to separate their kinetics, thus equation 5.5 can be written for a polyelectronic system as:

$$\left( -\sum_{i=1}^N \frac{\nabla_i^2}{2} - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \right) \Psi(\mathbf{r}_i, \mathbf{R}_A) = E \Psi(\mathbf{r}_i, \mathbf{R}_A) \quad (5.8)$$

Here the Hamiltonian is only composed of the electrons kinetic energies, the attraction term between nuclei and electrons and the repulsion term between electrons.

### 5.3.1 Variational Principle

The energy obtained solving the time-independent Schrödinger equation for any wave function  $\Phi$  will always be greater than or equal to the *exact* ground state energy  $E_0$ :

$$\langle H \rangle \equiv \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0 \quad (5.9)$$

## 5.4 Ground state

### 5.4.1 Hartree-Fock approximation

The Hartree-Fock approximation which is a single determinant theory aims to find the best set of spin-orbitals  $\chi_i$  such that the wave function  $\Psi_0$  for a system of  $N$  electrons:

$$|\Psi_0\rangle = |\chi_1 \chi_2 \cdots \chi_i \chi_j \cdots \chi_N\rangle \quad (5.10)$$

is the best approximation to the ground state. The spin-orbital of an electron is a way to take into account its spin such as the product between a molecular orbital ( $\phi_i$ ) and the spin function ( $\alpha$  or  $\beta$  respectively for spin up or spin down) is defined as:

$$\chi_i(\xi_i) = \chi_i(\mathbf{r}_i, \sigma_i) = \phi_i(\mathbf{r}_i)\alpha \quad \text{or} \quad \chi_j(\xi_j) = \chi_j(\mathbf{r}_i, \sigma_j) = \phi_i(\mathbf{r}_i)\beta \quad (5.11)$$

with  $\mathbf{r}_i$  the electron coordinates,  $\sigma$  the spin and  $\xi$  a variable containing both information.

The wave function of the system can be described by the product of monoelectronic wave functions that are antisymmetric. The product of spin-orbitals is called the Hartree product:

$$\Psi_{HP}(\xi_1, \xi_2, \dots, \xi_i, \xi_j, \dots, \xi_N) = \chi_1(\xi_1)\chi_2(\xi_2) \cdots \chi_i(\xi_i)\chi_j(\xi_j) \cdots \chi_N(\xi_N) \quad (5.12)$$

And the antisymmetry will be satisfied using a so-called Slater determinant written:

$$\Psi(\xi_1, \xi_2, \dots, \xi_i, \xi_j, \dots, \xi_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\xi_1) & \chi_2(\xi_1) & \cdots & \chi_N(\xi_1) \\ \chi_1(\xi_2) & \chi_2(\xi_2) & \cdots & \chi_N(\xi_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\xi_N) & \chi_2(\xi_N) & \cdots & \chi_N(\xi_N) \end{vmatrix} \quad (5.13)$$

The polyelectronic Schrödinger equation in 5.8 becomes a set of monoelectronic equations based on monoelectronic Hamiltonians  $h_i$ :

$$h_i \chi_i(\xi_i) = \epsilon_i \chi_i(\xi_i) \quad (5.14)$$

where  $\epsilon_i$  is the energy of the spin-orbital  $i$ . Moreover the monoelectronic Hamiltonian  $h_i$  is:

$$F_i = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j \geq i}^N (2J_j(\mathbf{r}_i) - K_j(\mathbf{r}_i)) \quad (5.15)$$

which is composed of the kinetic energy term, the electron-nuclei interaction term and a bielectronic term which expresses the mean field applied on each electron by its neighbors. This last expression considers the electronic Coulombian repulsion through a Coulomb operator  $J_i$  and an exchange operator  $K_i$  representing the Pauli principle. These two operators read:

$$J_i(\xi_i)\chi_i(\xi_i) = \int d\tau \chi_j^*(\xi_j) \frac{1}{r_{ij}} \chi_j(\xi_j) \chi_i(\xi_i) \quad (5.16)$$

$$K_i(\xi_i)\chi_i(\xi_i) = \int d\tau \chi_j^*(\xi_j) \chi_i(\xi_j) \frac{1}{r_{ij}} \chi_i(\xi_i) \quad (5.17)$$

Considering that the electronic kinetics and electron-nucleus interaction term can be put together in the core Hamiltonian  $h_i^c$ , 5.15 is considered as the Fock operator and reads in the set of spin-orbits  $\chi_i$ :

$$F = h_i^c + \sum_{i=1}^N \sum_{j \geq i}^N (2J_j(\mathbf{r}_i) - K_j(\mathbf{r}_i)) \quad (5.18)$$

To make this equation more practical to solve, the concept of basis set or namely the *Linear combination of Atomic Orbitals* (LCAO) has been introduced by Roothaan, where the Molecular Orbitals (MO) are represented by a linear combination of atomic orbitals (AO) in the case of a closed shell system:

$$\phi_i = \sum_{\mu} c_{\mu i} \chi_{\mu} \quad (5.19)$$

In the framework of the AO, 5.18 becomes for  $\mu$  and  $\nu$  AOs:

$$F_{\mu\nu} = \int \phi_{\mu}^* F(1) \phi_{\nu} d\mathbf{r}_1 \quad (5.20)$$

$$= \int \phi_{\mu}^* h^c(1) \phi_{\nu} d\mathbf{r}_1 + \sum_j^{N/2} \int \phi_{\mu}^* (2J_j(1) - K_j(1)) \phi_{\nu} d\mathbf{r}_1 \quad (5.21)$$

$$= H_{\mu\nu}^c + \sum_j^{N/2} 2 \langle \mu\nu | jj \rangle - \langle \mu j | j\nu \rangle \quad (5.22)$$

Then the linear combination can be inserted:

$$F_{\mu\nu} = H_{\mu\nu}^c + \sum_j^{N/2} \sum_{\lambda\sigma} c_{\lambda a} c_{\sigma a} (2 \langle \mu\nu | \lambda\sigma \rangle - \langle \mu\sigma | \lambda\nu \rangle) \quad (5.23)$$

$$= H_{\mu\nu}^c + \sum_{\lambda\sigma} P_{\lambda\sigma}^T (\langle \mu\nu | \lambda\sigma \rangle - \frac{1}{2} \langle \mu\sigma | \lambda\nu \rangle) \quad (5.24)$$

$$= H_{\mu\nu}^c + G_{\mu\nu} \quad (5.25)$$

with bielectronic integrals:

$$\langle \mu\nu | \lambda\sigma \rangle = \int \int \phi_\mu^*(1) \phi_\nu(1) \frac{1}{r_{12}} \phi_\lambda^*(2) \phi_\sigma(2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (5.26)$$

$$\langle \mu\sigma | \lambda\nu \rangle = \int \int \phi_\mu^*(1) \phi_\sigma(1) \frac{1}{r_{12}} \phi_\lambda^*(2) \phi_\nu(2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (5.27)$$

Considering that the Fock operator depends on the total density matrix  $P^T$  defined as:

$$P_{\mu\nu}^T = 2 \sum_i^N c_{\lambda i} c_{\sigma i}^* \quad (5.28)$$

The matrix form of the Roothaan Hartree-Fock is given by:

$$\mathbf{FC} = \mathbf{SCE}. \quad (5.29)$$

The basis functions can be orthogonalized by introducing the orthogonalization matrix  $\mathbf{X}$  such as:

$$\mathbf{X}^\dagger \mathbf{S} \mathbf{X} = \mathbf{I} \quad (5.30)$$

where  $\mathbf{I}$  is the identity matrix. And

$$\mathbf{F}' = \mathbf{X}^\dagger \mathbf{F} \mathbf{X} \quad (5.31)$$

To obtain the energy  $\mathbf{E}$  of the system, one must rely on an auto-coherent iterative algorithm called Self-Consistent Field (SCF):

- 1 Providing information about the molecular system: charge, coordinates, atoms, basis set.
- 2 Calculation of molecular integrals ( $\mathbf{S}, \mathbf{G}, \mathbf{H} \dots$ )
- 3 Diagonalizing overlap matrix and construction of  $\mathbf{X}$
- 4 First guess of  $\mathbf{P}$  using a trial wave function
- 5 Computation of the Fock matrix  $\mathbf{F}$
- 6 Expression of  $\mathbf{F}$  in an orthogonal basis:  $\mathbf{F}' = \mathbf{X}^\dagger \mathbf{F} \mathbf{X}$

- 7 Diagonalizing  $\mathbf{F}'$  to obtain  $\mathbf{C}'$  and the molecular orbital energies
- 8 Express  $\mathbf{C}'$  such as  $\mathbf{C} = \mathbf{X}\mathbf{C}'$  and computation of the new density matrix
- 9 If convergence is not reach, go back to point 5.

The first disadvantage of the HF method is that it neglects the electron correlation, leading to large deviations compared to experimental results. To overcome this, post-HF methods have been developed such as the Moller-Plesset perturbation theory that treats electron correlation as a perturbation of the Fock operator or to go even further, some methods use linear combinations of Slater determinants to obtain a multi-configurational wave function (the description of such methods is beyond the scope of this manuscript). Moreover, the quality of the results will depend on the chosen basis set with a closer result to the exact energy for larger ones, but increasing sometimes considerably the computation resources needed.

The main and most famous alternative to HF and post-HF methods is Density Functional Theory which treats both exchange and correlation energies. The following Chapter is dedicated to this method.

## 5.4.2 Density Functional theory (DFT)

### 5.4.2.1 Principle

Instead of using the wave function of a system to obtain it's energy, an alternative way is to employ it's electronic density, this method is called Density Functional Theory (DFT). It has mainly been developed by John A. Pople and Walter Kohn who received a Nobel prize in 1998 for their development of the density-functional theory and computational methods in quantum chemistry. The Hamiltonian describing each system is exactly the same as formula 5.7 since the Born-Oppenheimer approximation still holds.

### 5.4.2.2 The electronic density

Instead of the wave function, DFT focuses on the electronic density written as  $n(\mathbf{r})$  which is given for a system of  $N$  electrons and a normalized  $\Psi$ :

$$n(\mathbf{r}) = N \int |\Psi(\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_N \cdots \mathbf{r}_N)|^2 d\mathbf{r}_1 d\mathbf{r}_2 \quad (5.32)$$

and the ground state density  $n_0(\mathbf{r})$  can be calculated for a corresponding wave function  $\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ . In fact the energy is a unique functional of  $n_0$ . The main features of the electronic density are:



$$n(\mathbf{r} \rightarrow \infty) = 0 \quad \text{and} \quad \int n(\mathbf{r}) d\mathbf{r} = N \quad (5.33)$$

The ground state energy in DFT can be obtained as:

$$E_0 = E[n_0] = \langle \Psi[n_0] | \hat{T} + \hat{V} | \Psi[n_0] \rangle \quad (5.34)$$

where  $\langle \Psi[n_0] | \hat{V} | \Psi[n_0] \rangle$  can be written:

$$V[n_0] = \int d\mathbf{r} V[n_0(\mathbf{r})] d\tau \quad (5.35)$$

#### 5.4.2.3 The Hohenberg-Kohn theorems

The first Hohenberg-Kohn theorem demonstrates that the Hamiltonian operator of a many-body problem can be uniquely determined by the electron density.

It states that any system composed of interacting particles under the influence of an external potential  $v_{ext}(\mathbf{r})$ , this potential is uniquely determined by the ground state electronic density  $n_0(\mathbf{r})$  which in turns defines the ground state wave function  $\Psi_0$  and ultimately yields the ground state energy.

This first theorem allows to write the ground state energy as a function of the ground state density:

$$E_0 = E[n_0(\mathbf{r})] = \langle \Psi_0 | \hat{T} + \hat{E}_{ee} | \Psi_0 \rangle + \int d\mathbf{r} v_{ext}(\mathbf{r}) n_0(\mathbf{r}) \quad (5.36)$$

here  $\hat{T}$  is the kinetic energy of the electrons,  $E_{ee}$  the electron-electron repulsion such as for  $N$  interacting electrons:

$$\hat{T} = - \sum_{i=1}^N \nabla^2 \quad (5.37)$$

and

$$\hat{E}_{ee} = - \sum_{i=1}^N \sum_{j < i}^N \frac{1}{\mathbf{r}_{ij}} \quad (5.38)$$

Both can be summed to obtain the universal Hohenberg-Kohn functional  $F[n(\mathbf{r})]$  that for a wave function  $\Psi$  depending on the density  $n(\mathbf{r})$  is:

$$F[n(\mathbf{r})] = \langle \Psi[n(\mathbf{r})] | \hat{T} + \hat{E}_{ee} | \Psi[n(\mathbf{r})] \rangle \quad (5.39)$$

Here  $F[n(\mathbf{r})]$  being unknown,  $E[n(\mathbf{r})]$  cannot be calculated. Through the use of the second Hohenberg-Kohn theorem, this problem can be tackled. Indeed, it states that the exact ground state density of a many particles system minimizes the energy functional  $E[n(\mathbf{r})]$  in a variational way. Meaning that the exact ground-state energy  $E_0$  is obtained for the exact ground state density  $n_0(\mathbf{r})$ :

$$\min_{n(\mathbf{r}) \rightarrow n_0(\mathbf{r})} E[n(\mathbf{r})] = E[n_0(\mathbf{r})] = E_0 \quad (5.40)$$

#### 5.4.2.4 The Kohn-Sham equations

To be able to find the ground state density  $n_0(\mathbf{r})$  for a system of non-interacting electrons as described in the Hohenberg-Kohn theorems, Walter Kohn and Lu Sham [98] suggested to split the Hohenberg-Kohn functional (5.39), such as the functional energy of the system  $E[n(\mathbf{r})]$  can be written as:

$$E[n(\mathbf{r})] = T_S[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] + \int d\mathbf{r} v_{ext}(\mathbf{r})n_0(\mathbf{r}) \quad (5.41)$$

which is equivalent to a fictitious system of non-interacting electrons that has the same density and energy as the real system. In this equation,  $T_S[n(\mathbf{r})]$  corresponds to the kinetic energy of non-interacting electrons,

$$T_S[n(\mathbf{r})] = -\frac{1}{2} \sum_{i=1}^N \nabla^2 \quad (5.42)$$

then  $E_H[n(\mathbf{r})]$  is the Coulomb or Hartree energy functional,

$$E_H[n(\mathbf{r})] = \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (5.43)$$

and finally  $E_{xc}[n(\mathbf{r})]$  is called the exchange-correlation energy obtained using the exchange correlation potential  $v_{xc}[n(\mathbf{r})]$  such as:

$$v_{xc}[n(\mathbf{r})] = \frac{\partial E_{xc}[n(\mathbf{r})]}{\partial n(\mathbf{r})} \quad (5.44)$$

5.41 can be rewritten as:

$$E[n(\mathbf{r})] = T_S[n(\mathbf{r})] + \int d\mathbf{r} v_{eff}(\mathbf{r})n_0(\mathbf{r}) \quad (5.45)$$

where the effective Kohn-Sham potential appears and is composed of:

$$v_{eff}[n(\mathbf{r})] = v_{ext}[n(\mathbf{r})] + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{xc}[n(\mathbf{r})] \quad (5.46)$$

And finally, for a set of Kohn-Sham orbitals, of energy  $\epsilon_i$ :

$$E = \sum_i^N \epsilon_i - E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] - \int \frac{\partial E_{xc}[n(\mathbf{r})]}{\partial n(\mathbf{r})} n(\mathbf{r}) d\mathbf{r} \quad (5.47)$$

Even though the Kohn-Sham equation is formally exact, the exchange-correlation  $E_{xc}[n(\mathbf{r})]$  is only known for a few systems [99] and will have to be approximated for most cases. So far 5 main families of functionals have been developed going from the accuracy of the Hartree-Fock theory until the exact chemical result. This stepwise evolution is composed of the so-called LDA, GGA, meta-GGA, hybrid-GGA/meta-GGA and double hybrids. Each family will be quickly presented in the next subsections.

#### 5.4.2.5 LDA

The first known approximation of the exchange-correlation energy is the so-called Local Density Approximation (LDA). Here,  $E_{xc}[n(\mathbf{r})]$  will be considered as the one of a single electron in an homogeneous gas of electrons of density  $n(\mathbf{r})$ . The LDA energy is:

$$E_{xc}^{LDA}[n(\mathbf{r})] = \int \epsilon_{xc}(n(\mathbf{r})) n(\mathbf{r}) d\mathbf{r} \quad (5.48)$$

where the exchange-correlation energy  $\epsilon_{xc}(n(\mathbf{r}))$  of the considered electron can be split in an exchange term and a correlation term:

$$\epsilon_{xc}(n(\mathbf{r})) = \epsilon_x(n(\mathbf{r})) + \epsilon_c(n(\mathbf{r})) \quad (5.49)$$

where  $\epsilon_x(n(\mathbf{r}))$  is the exchange energy of an electron in a uniform electron gas of density  $n(\mathbf{r})$ :

$$\epsilon_x(n(\mathbf{r})) = -\frac{3}{4} \left( \frac{3n(\mathbf{r})}{\pi} \right)^{\frac{1}{3}} \quad (5.50)$$

and  $\epsilon_c(n(\mathbf{r}))$  is not explicitly known but can be obtained through highly accurate quantum Monte-Carlo simulations of homogeneous electrons gases [100].

The main drawbacks of the use of LDA exchange-correlation energy is that it

works poorly in reproducing van der Waals interactions and hydrogen bonds which are crucial in many chemical applications.

To go even beyond in this approximation it can be applied to the spin-density  $n_\alpha(\mathbf{r})$  and  $n_\beta(\mathbf{r})$  respectively for spin up and spin down, it is called Local Spin-Density Approximation (LSDA):

$$E_{xc}^{LSDA}[n(\mathbf{r})] = \int \epsilon_{xc}(n_\alpha(\mathbf{r}), n_\beta(\mathbf{r}))n(\mathbf{r})d\mathbf{r} \quad (5.51)$$

#### 5.4.2.6 GGA

The next famous approach to obtain the exchange correlation energy is not to consider only the density at a given point  $\mathbf{r}$ , but to take also into account the information of the charge density gradient  $\nabla n(\mathbf{r})$ . It is called the Generalized Gradient Approximation (GGA).

$$E_{xc}^{GGA}[n(\mathbf{r})] = \int \epsilon_{xc}(n_\alpha(\mathbf{r}), n_\beta(\mathbf{r}), \nabla n_\alpha(\mathbf{r}), \nabla n_\beta(\mathbf{r}))n(\mathbf{r})d\mathbf{r} \quad (5.52)$$

One can cite in this family of GGA functional BLYP, BP86, PBE, BPW91, HCTH which have been extensively used.

#### 5.4.2.7 meta-GGA

The third improvement in the development of GGA functionals was to include the second derivative of the density, thus including a dependence on the kinetic energy density.

#### 5.4.2.8 Hybrid GGA and hybrid meta-GGA

This main characteristic of this class of functionals first introduced by Becke in 1993 [101] is to include a percentage of exact exchange calculated using the Hartree-Fock method. This exact exchange expression is for two orbitals  $i$  and  $j$ :

$$E_x^{HF} = \frac{1}{2} \sum_{i,j} \int \int \Psi_i^*(\mathbf{r})\Psi_j^*(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \Psi_i(\mathbf{r})\Psi_j(\mathbf{r}') d\mathbf{r}_1 d\mathbf{r}_2 \quad (5.53)$$

with  $|\mathbf{r} - \mathbf{r}'|$  the distance between the two particles. Then by fitting a set of exchange and correlation explicit density functionals over experimental data one obtain some of the most famous and most accurate exchange-correlation functional, by example B3LYP (which is employed in approximately 80% of DFT studies), PBE0 [102] or B3PW91 [103]. The exchange correlation term in these functionals

develops as:

$$E_{xc}^{B3LYP} = E_x^{LDA} + a(E_x^{HF} - E_x^{LDA}) + b(E_x^{GGA} - E_x^{LDA}) + \dots \\ \dots E_c^{LDA} + c(E_c^{GGA} - E_c^{LDA}) \quad (5.54)$$

$$E_{xc}^{B3PW91} = E_{xc}^{LSDA} + a(E_x^{HF} - E_x^{LSDA}) + b\delta E_x^{BP88} + c\delta E_c^{PW91} \quad (5.55)$$

where  $a = 0.20$ ,  $b = 0.72$  and  $c = 0.81$  in both 5.54 and 5.55.

And,

$$E_{xc}^{PBE0} = \frac{1}{4}E_x^{HF} + \frac{3}{4}E_x^{PBE} - E_c^{PBE} \quad (5.56)$$

where  $E_x^{PBE}$  and  $E_c^{PBE}$  are respectively the exchange and correlation terms of the PBE functional which is from the GGA family.

Then the family of hybrid meta-GGA is mainly composed of the M06 suite developed by the Truhlar group [104, 105] (M06-L, M06, M06-2X, M06-HF). They are empirically fitted but still constrained to a uniform electron gas. Their major improvement is made over one of the most pathological problem of DFT until the hybrid GGA functionals, that is taking into account dispersion forces.

#### 5.4.2.9 Double hybrids

Finally the last class of functionals is called double hybrids or Range-Separated Hybrids (RSH). It includes a separation between long and short inter-electron distance:

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \frac{1 - \text{erf}(\mu|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} + \frac{\text{erf}(\mu|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} \quad (5.57)$$

here  $\mu$ . Here the exchange correlation contains a perturbation term improving the diffuse orbitals. Indeed, this perturbation term also allows to compute more accurately the weak dispersion forces.

$$E_{xc} = (1 - a)E_x^{GGA} + aE_x^{HF} + (1 - b)E_c^{GGA} + bE_c^{MP2} \quad (5.58)$$

where  $\mu$ ,  $a$  and  $b$  are fitted parameters that varies depending on the functional.

The main examples here are CAM-B3LYP, wB97XD or B2-LYP which give the most accurate results but increases considerably the computation time which is also

the main disadvantage of these functionals compared to all the previously presented families.

## 5.5 Excited states

Photochemical, photobiological and photophysical processes can only be simulated through the calculation of the excited states of a given many-particles system. It allows to obtain information about light-driven reactions, absorption and fluorescence spectra (excitation and emission energies),

I will present in the following section the formalism of two methods employed to obtain excited states, namely Time-Dependent Hartree-Fock (TDHF) and Time Dependent-Density Functional Theory (TD-DFT).

### 5.5.1 Time Dependent-Hartree Fock (TDHF)

In order to study time dependent mechanisms, an approximation of the time dependent Schrödinger equation has been formalized by Dirac in 1930 [106]. Using the linear-response TDHF equations, one can obtain electronic excited states, thus electronic spectra and properties.

From 5.4, where the operator  $\hat{H}$  is also time dependent:

$$\hat{H}(\mathbf{r}, t) = \hat{H}(\mathbf{r}) + \hat{V}(\mathbf{r}, t) \quad (5.59)$$

where  $\hat{V}(\mathbf{r}, t)$  is a single-particle operator dependent on the time.

Then the Hartree-Fock equation can be written in a time dependent fashion:

$$F(\mathbf{r}, t)\Psi(\mathbf{r}, t) = i\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) \quad (5.60)$$

The excitations can be obtained in a matrix form based on a non-Hermitian eigenvalues equation. Using a time-dependent perturbation formalism of the first order, one can obtain the linear response TDHF.

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{vmatrix} \begin{vmatrix} \mathbf{X} \\ \mathbf{Y} \end{vmatrix} = \omega \begin{vmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{vmatrix} \begin{vmatrix} \mathbf{X} \\ \mathbf{Y} \end{vmatrix} \quad (5.61)$$

where the different terms are:

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + \langle ia||bj \rangle \quad (5.62)$$

$$B_{ia,jb} = \langle ia || bj \rangle \quad (5.63)$$

with  $i, j$  and  $a, b$  respectively the occupied and virtual orbitals.

## 5.5.2 Time Dependent-Density Functional Theory (TDDFT)

### 5.5.2.1 The Runge-Gross theorem

Analogously to the first Hohenberg-Kohn theorem stating that there is a direct correlation between the time independent external potential and a unique electron density  $n(\mathbf{r})$ , the Runge-Gross theorem, written by Erich Runge and Eberhard Gross in 1984 [107], demonstrates the same relation in a time dependent manner (the unique relationship between the time-dependent potentials  $v_{ext}(n(\mathbf{r}, t))$  and the time-dependent electronic densities  $n(\mathbf{r}, t)$ )

$$n(\mathbf{r}, t) \longleftrightarrow v_{ext}(n(\mathbf{r}, t)) \quad (5.64)$$

And the time-dependent Schrödinger equation has for solution the wave function  $\Psi(\mathbf{r}, t)$ :

$$H(\mathbf{r}, t)\Psi(\mathbf{r}, t) = i\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) \quad (5.65)$$

that can be rewritten dependent on the electronic densities:

$$H(n(\mathbf{r}, t))\Psi(n(\mathbf{r}, t)) = i\frac{\partial}{\partial t}\Psi(n(\mathbf{r}, t)) \quad (5.66)$$

### 5.5.2.2 Time-Dependent Kohn-Sham equation

Similarly to DFT, 5.45 and 5.46 can also be rewritten as dependent of the time:

$$E[n(\mathbf{r}, t)] = T_S[n(\mathbf{r}, t)] + \int d\mathbf{r} v_{eff}(\mathbf{r}, t)n_0(\mathbf{r}, t) \quad (5.67)$$

where the effective Kohn-Sham potential is composed of:

$$v_{eff}[n(\mathbf{r}, t)] = v_{ext}[n(\mathbf{r}, t)] + \int d\mathbf{r}' \frac{n(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} + v_{xc}([n(\mathbf{r}, t)]) \quad (5.68)$$

Then the time dependent Schrödinger equation is solved:

$$(T_s + v_{eff})\Psi_i(\mathbf{r}, t) = i\frac{\partial}{\partial t}\Psi_i(\mathbf{r}, t) \quad (5.69)$$

And the linear response of the TD-DFT formalism is similar to the one of TDHF (5.61) with the **A** and **B** term now including the exchange correlation term  $V_{xc}$ :

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + (ia|jb) + (ia|V_{xc}|jb) \quad (5.70)$$

$$B_{ia,jb} = (ia|jb) + (ia|V_{xc}|jb) \quad (5.71)$$

### 5.5.3 Frenkel excitons theory

Investigating the electronic properties of large biomacromolecular systems such as DNA and proteins, and more specifically their spectroscopic properties, requiring to compute the excited states, is still a difficult task since their study using QM methods requires large amounts of resources. For example, the circular dichroism signal of DNA arises from its double-helical structure [48] which requires to consider at least 2 base pairs (even more for more accurate results) that is at least 50 atoms. It appears evident that new methodologies and protocols has to be developed in order to tackle such problematic [108, 109]. In that aim, the Frenkel excitons theory can be employed. It relies on the coupling of localized excited states of interacting chromophores in order to obtain approximate excitons. If one considers the simplest model of two interacting chromophores  $\phi_i$   $\phi_j$ , the global excited states will be built by the coupling of the monomers excited states:

$$|\Psi_{\pm}\rangle = c_{\pm}^a |\phi_i^{\dagger}\phi_j^0\rangle \pm c_{\pm}^b |\phi_i^{\dagger}\phi_j^{\dagger}\rangle \quad (5.72)$$

where  $|\phi_i^{\dagger}\phi_j^0\rangle$  is an excitation localized on the monomer  $i$ . The effective Hamiltonian of the system can be defined by the excitation energies  $\epsilon$  on its diagonal and coupling terms in the off-diagonal elements.

In this formalism, the matrix elements can simply be written using the following semi-empirical equation:

$$H_{ii} = \epsilon_i; H_{ij} = \langle\phi_i^{\dagger}|\hat{H}|\phi_j^{\dagger}\rangle \quad (5.73)$$

where  $\mu_i$  and  $\mu_j$  are the transition dipole moments of monomers  $i$  and  $j$ , and  $R_{ij}$  the vector distance between the two centers of charges. After diagonalization of  $H$



the oscillator strength of the exciton  $k$  can be obtained as:

$$f_k = \sum_{i,j} (\mu_i \cdot \mu_j) U_{ik} U_{jk} \quad (5.74)$$

and the rotatory strength is:

$$r_k = \sum_{i,j} \epsilon_k [R_{ij}(\mu_i \cdot \mu_j)] U_{ik} U_{jk} \quad (5.75)$$

It is necessary to note here that in the case of the modeling of electronic circular dichroism, the three components of  $r_k$  in the three directions of space need to be averaged in order to recover the actual signal.

# Chapter 6

## QM/MM

### 6.1 QM/MM energy

Biochemical phenomena such as rhodopsin residue switches that is necessary for vision, mechanisms of the NER and BER repair processes for example happens inside large macromolecular environments but the electronic reorganization itself only occurs for tenth or hundreds of atoms. Their study using QM methods, including explicitly the electrons involved, is straightforward if the system that is considered is extracted of it's environment and studied isolated. But these models lack the electrostatic influence of either the protein/nucleic environment or the surrounding solvent molecules and ions, and also the mechanical influence of the atoms on the studied phenomena. In order to assess both aspects, the electronic mechanism and the large scale environment, the so called and famous QM/MM methodology has been created [110]. Represented in 6.1 for the case of the chelerythrine DNA sensitizer intercalated in DNA, in the QM/MM framework, the system is split in two distinct regions, one QM composed of a relatively small number of atoms where the main electronic phenomena occurs, and a MM including the remaining external environment. Thanks to the development and capacity of QMMM, Warshel, Levitt and Karplus have been awarded a Nobel prize for the creation and development of QM/MM [25].

The entire system is then not only characterized by a QM Hamiltonian  $H_{QM}$  and a MM one  $H_{MM}$  for both regions, but a third term must be added to treat the frontier and interactions between them, the QM/MM Hamiltonian  $H_{QM/MM}$ , thus the effective Hamiltonian of the system will be written:

$$H_{eff} = H_{QM} + H_{MM} + H_{QM/MM} \quad (6.1)$$

This is called the additive scheme since  $H_{QM/MM}$  is *added* to the effective one. But it can also be subtractive for some methodologies where the QM and MM parts

have a different definition:

$$H_{eff} = H_{QM} + H_{MM} - H_{QM/MM} \quad (6.2)$$

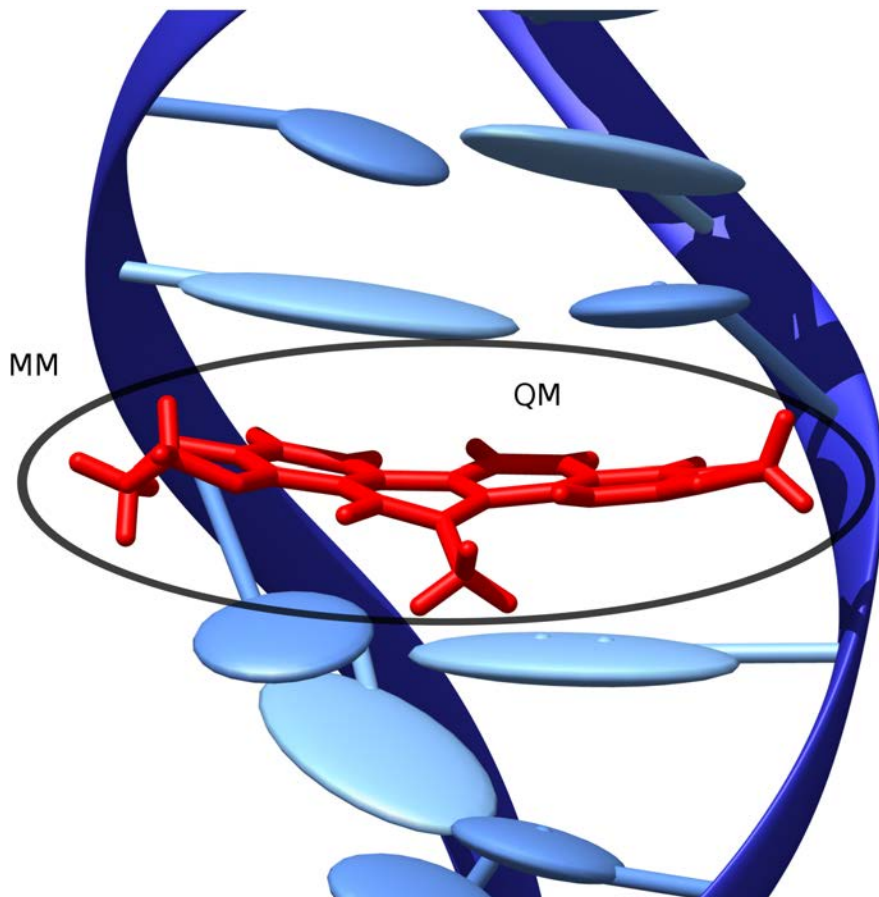


Figure 6.1: Representation of the QM (red) and MM (blue) regions in the case of a DNA sensitizer in intercalation mode.

We shall focus here on the additive scheme 6.1 since it is the one employed in QM/MM interface. In that situation, the MM energy is calculated for the outer system, then the QM energy calculation is done on the inner region and finally the frontier interaction energy is added by calculating  $E_{QM/MM}$ . The advantage of that procedure is that the classical mechanics calculation is only performed for the MM region, keeping the accuracy of the QM calculation for the important electronic description of the active site.

### 6.1.1 QM/MM embeddings

As presented before, the surrounding plays an important role in the studied electronic mechanisms. QM/MM allows to decouple the mechanical and electrostatic influences of the environment in three main strategies described hereinafter.

#### 6.1.1.1 Mechanical Embedding (ME)

The mechanical embedding strategy only considers the mechanical influence of the MM atoms on the QM region for the calculation of the QM/MM interactions. The charges of the MM atoms are neglected and do not influence the wave function of the QM area.

#### 6.1.1.2 Electrostatic Embedding (EE)

The next step is to consider the MM atom charges and take them into account on the QM wavefunction, hence QM energy. This strategy is predominant since its effect on the QM energy accuracy is much higher than the ME, and compared with the polarizable embedding, its implementation is much more straightforward and the simulation time is much less affected.

#### 6.1.1.3 Polarisable Embedding (PE)

And finally the polarizable embedding is a mean to add a counter-effect of the QM charges on the MM atoms and so on iteratively until complete polarization of both parts. To simplify this costly strategy, one can simply consider the MM region as a polarizable continuum that will play the role of the surrounding electrostatic interaction, it is called the Electronic Response of the Surrounding (ERS) [111].

## 6.2 QM/MM boundary

For the case of a small organic endogenous molecule interacting with DNA non-covalently, only electrostatic and van der Waals effects are considered, in that case the implementation of QM/MM methods is rather straightforward. But when two atoms involved in the same chemical bond are one in the QM region and the other in the MM region, strategies must be found to describe correctly the system at the frontier. In the studies related to this manuscript the strategy that has been employed is the so-called link atom described below.

### 6.2.1 Link Atom (LA)

The Link Atom strategy is the most commonly employed method to treat the frontier of QM and MM regions having one or several covalent bonds since it is the easiest to implement and many computing codes have implemented it. The idea is to add a monovalent atom (usually a hydrogen atom), at the center of the frontier bond, that will be on one side treated in the QM part as bonded to the first QM atom and then considered in the MM part as linked to the first MM atom. To avoid

overpolarization of the QM part in the case of EE and PE, the LA should be added as much as possible between two atoms of the same nature (C-C, N-N for example), so that they both have relatively similar charges. An example of QM/MM frontier treatments for nucleic acids is presented in Figure 6.2 bellow.

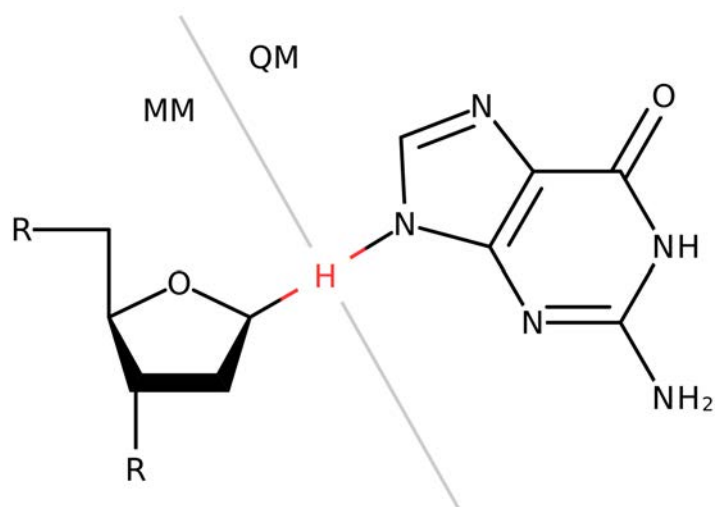


Figure 6.2: Position of the link atom (a hydrogen in this case) for a QM/MM frontier between a nucleobase and its sugar.

## Part VIII

### Results and discussions

# Chapter 7

## Introduction

**Structure globale** Cette partie a été organisée de manière à donner une vue d'ensemble des projets qui ont été menés sur la photosensibilisation de l'ADN. Cela commence par la description des outils développés pour comprendre la structure de l'ADN et plus spécifiquement ses interactions avec des agents sensibilisants; i) la modélisation du dichroïsme circulaire électronique de biomacromolécules; ii) l'énergie libre d'interaction de ces molécules avec l'ADN. Ensuite, les voies de photosensibilisation de molécules activables par la lumière (benzophenone, nile blue, nile red, BMEMC et Pyo) sont décrits ainsi que leurs interactions avec l'ADN. Et pour finir une analyse complète des effets structuraux et dynamiques de ces lésions sur l'ADN et leurs impacts sur la reconnaissance par des protéines, c'est-à-dire leurs activités de réparation, est présenté.

**Global structure** The following chapters have been organized so that it gives a complete overview of the projects carried out in relation with DNA photosensitization. It starts with the description of the tools developed to understand the DNA structure and more specifically its interaction with drugs; i) the modeling of electronic circular dichroism of biomacromolecular systems; ii) the binding free energy of sensitizers in DNA. It is directly followed by the investigation of the photosensitizing pathways of several light activable molecules (benzophenone, nile blue, nile red, BMEMC and Pyo) and their interaction configurations with DNA. Moreover, their damaging processes are presented. A thorough analysis of the structural and dynamical effects of the lesions embedded in DNA and their impact on protein recognition, thus repair activities, is also described.

# Chapter 8

## Electronic Circular Dichroism (ECD) modeling

### 8.1 ECD of B-DNA

#### Introduction

Due to the presence of weak stabilizing interactions in the DNA double helix; namely  $\pi$ -stacking and hydrogen bonding (2.4); its structure and dynamic is still the matter of debates in many fields of research such as biology, chemistry and biophysics [112–115]. The necessity to have a clear understanding of this complex behavior[116], its relation with the DNA base pairs sequences and its many structure polymorphisms[117, 118] arises most notably from its high impact on numerous biological processes related to DNA repair, transcription, replication, . . . Moreover, the influence of the biological environment and in particular nucleosomal compaction has a crucial impact on the structure and dynamic whose study by some experimental techniques such as crystallography or spectroscopies may be complex and time consuming.

The most common experimental technique to study the macromolecular arrangement of DNA is the so called Electronic Circular Dichroism (ECD) [48, 119–121]. Indeed, at the local level (nucleotide), DNA is not chiral thus does not have a dichroic signal, but due to its global helical arrangement, the base pairs stacks gives rise to specific ECD signals [121–125], that are highly sensitive to either the molecular structure (sequence), polymorphism (double-strand, G-quadruplex, . . .) and environment (ions, solvent, proteins, . . .). In fact, all the information about the structure is embedded in this signal and the link between both is not straightforward and still needs more understandings and precise descriptions.

Such aim can be tackled employing molecular modeling. Indeed, on one side the structure and dynamic properties of DNA can be studied using MD simulations and



on the other side its optical properties, that are excited states, can be obtained by QM calculations. Combining both, in a QM/MM fashion, is an appealing methodology that can, compared to experimental results, brings a new point of view on the elucidation of ECD signal/structures relation. Such methodology has gained interest in the scientific community and will require more investigations to prove its efficiency and impact[124, 126].

Nevertheless, the sensitivity of ECD is such that the modeling of the environment has to be considered with great care, since it affects the nucleobases spectroscopic properties as already reported [123, 127–130]. The system becomes even more complex if we consider that the nucleobases of DNA are an ensemble of interacting chromophores. In that case an effective Hamiltonian is compulsory since the size of the system is too large to be entirely treated by QM methods. Here the Frenkel exciton theory[108, 109, 131] and its related Hamiltonian can be of a great interest since it allows to decouple an ensemble of interacting chromophores in units of reasonable sizes in which the excited states will be obtained in the QM/MM framework. We sampled the conformational space of the two DNA heteropolymers using MD simulations and the excited states were obtained by TD-DFT. Their coupling through the Frenkel Hamiltonian resulted in a final ECD signal that can be directly compared to extended QM calculations on similar models.

## Computational details

Two Watson-Crick paired double-helical B-DNA strands (poly(dAT) and poly(dCG) of 15 pairs each) were created in silico using the nucleic utility code implemented in Amber[132]. A water solvation box of 10 Å buffer was used to embed the strands (truncated octahedron). Water molecules were modeled using TIP3P model while Na<sup>+</sup> cations have been added to ensure the neutrality of the systems. The molecular dynamics simulations were performed using the AMBER99 force field with the bsc0 correction, developed for DNA [133]. Periodic boundary conditions and the particle mesh Ewald summation have been used throughout. Global simulations were divided into five steps. A first step consisted in the relaxation of the water box conformation over 8000 steps (4000 using the steepest descent algorithm and 4000 with the conjugated gradient), followed by the relaxation of DNA and solvent. The system was then heated up from 0 to 300 K in the NVT ensemble and equilibrated in the NPT ensemble to reach a pressure of 1 atm during 20 ps. Finally, 10 ns production runs were carried out again in the NPT ensemble. From the previous trajectories, 40 snapshots were extracted in order to compute the vertical electronic excitations of each DNA base using our QM/MM method 10.0 [111, 134]. The QM/MM frontier was placed at the junction between the sugar and the nucleobase, and the dangling

bond was treated with the link-atom scheme[111]. The quantum parts of the computations were conducted at the time-dependent density functional theory (TDDFT) level using the M06-2X[105] density functional and the 6-311+G(d) basis set[135]. In the case of poly(dAT), the energies of the four first excited states of each base were determined as well as their electronic transition dipole moments, while the six first excited states were calculated for the poly(dCG). Excitation energies and transition dipole moments were used to build the Frenkel Hamiltonian and simulate the ECD of the DNA, the formalism has been described in 5.5.3. To assess the validity of the model, ECD spectra obtained with the Frenkel model have been compared to the ones obtained from the ab initio QM/MM calculations of a total of four DNA nucleobases constituting two nearby base pairs. Representative structures obtained from the molecular dynamic trajectories are reported in Figure 8.1 with the bases placed in the QM partition evidenced in licorice representation. Excited states have been calculated on top of ten snapshots extracted from the corresponding molecular dynamic trajectories. TDDFT level was used with the 6-311+G(d) basis set and the M06-2X functional. Thirty-two excited states were calculated for poly(dAT) and 48 states for poly(dCG), this correspond to highest excitation energies of about 6.5 and 6.6 eV, respectively.

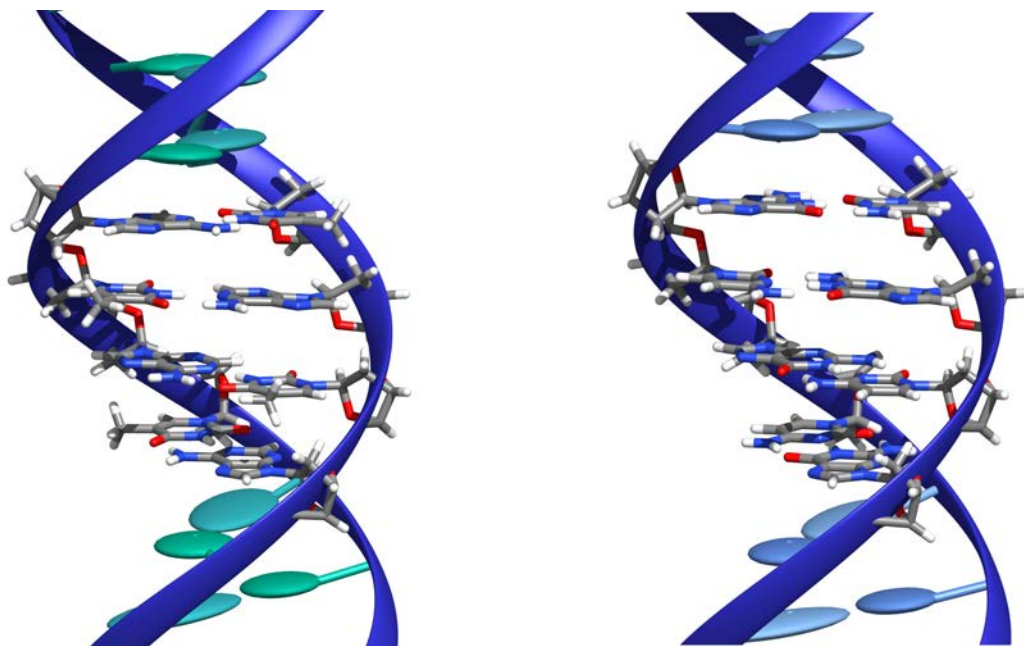


Figure 8.1: Representative snapshots of the DNA structures of the two double strands poly[d(GC)] on the right and poly(dAT) on the left. The central base pairs included in the Frenkel Hamiltonian are evidenced in licorice representation.

It is important to underline that for both the ab initio protocol and the Frenkel semiempirical Hamiltonian, the excited states of each individual monomer have been calculated using a QM/MM approach. More particularly, and thanks to the orig-

inal electrostatic response of the environment (ERS) technique [111] mechanical, electrostatic and polarizable embeddings are accounted for.

All the QM/MM computations were performed using a local modified version of Gaussian 09, Revision B0.1 [136], coupled with Tinker [137]. Spectra band shapes have been obtained convoluting each vertical transition of each snapshot with a Gaussian function of full width at half-length of 0.2 eV. Frenkel Hamiltonian ECD spectra have been obtained using a private code to post-process Gaussian 09 outputs and publicly available under GPL license. DNA global deformation parameters have been calculated using the Curves+ code [79].

## Results and discussion

In order to assess the performance of the simple Frenkel Hamiltonian approach in Figure 8.2, we report the comparison of the ECD spectra obtained considering a subsystem composed of the two central base pairs, treated with the Frenkel effective Hamiltonian or calculating at ab initio QM/MM level (QM approach) the excited states for the four bases supersystem. For both double strands, one may notice that the QM approach leads to systematically red-shifted maxima compared to the effective Hamiltonian. Indeed, it appears that the maxima are systematically shifted by about 10-15 nm in a consistent way among different bands and different strands.

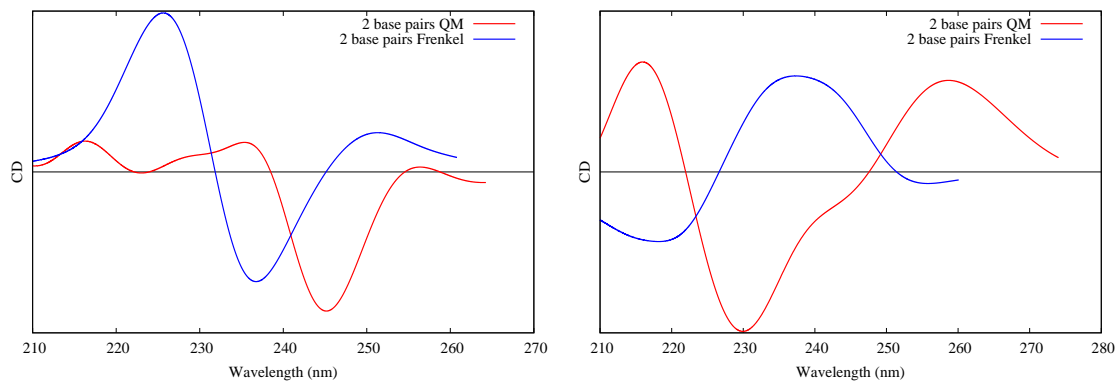


Figure 8.2: ECD spectra for a poly(dAT) (left panel) and poly(dCG) (right panel) obtained with the Frenkel Hamiltonian or a full QM/MM approach on the two central base pairs. Wavelengths in nm, intensities in arbitrary units

On the other hand, the band shapes for both systems appear to be quite correctly reproduced. In the case of poly(dAT) strand, one can see the low-intensity positive maxima at longer wavelengths followed by a strongly negative band (at about 245 nm). Note, however, that the low-energy positive maximum is much more intense when considering the Frenkel approach. Poly(dGC) presents a more complex structure with a large positive band at lower energy (260-240 nm) followed by a negative band at about 230 nm; note, however, that the relative intensity of

the negative band when using the Frenkel Hamiltonian appears reduced compared to the full QM/MM one. It is, however, evident, and despite some discrepancies, that even the crude Frenkel approximation used here is able to correctly recover the main features of DNA ECD provided the energy shift is taken into account. Note also that all the transitions are mainly of valence  $\pi$ - $\pi^*$  nature, charge-transfer transitions appearing only at higher transition energies; this fact also justifies the use of the Frenkel Hamiltonian that does not take into account charge-transfer. In Figure 8.3, we report the analysis of the convergence with the number of monomers, i.e., the DNA nucleobases included in the Frenkel model Hamiltonian for the two representative DNA fragments. To facilitate comparison, the intensities have been renormalized per number of base pairs. From the spectra reported there, we may evidence a relatively fast convergence of the Frenkel ECD spectra. Indeed, if the simplest model (i.e., the two base pairs) only gives a very approximated picture of the spectrum, the differences between the four and eight base pairs model are much smaller. More in detail in the case of the poly(dAT) spectrum, we may evidence the presence of a very low-intensity positive band at around 250-260 nm followed by a negative band at about 230 nm. The differences in the excitation wavelengths given by the four or eight base pair models are extremely negligible.

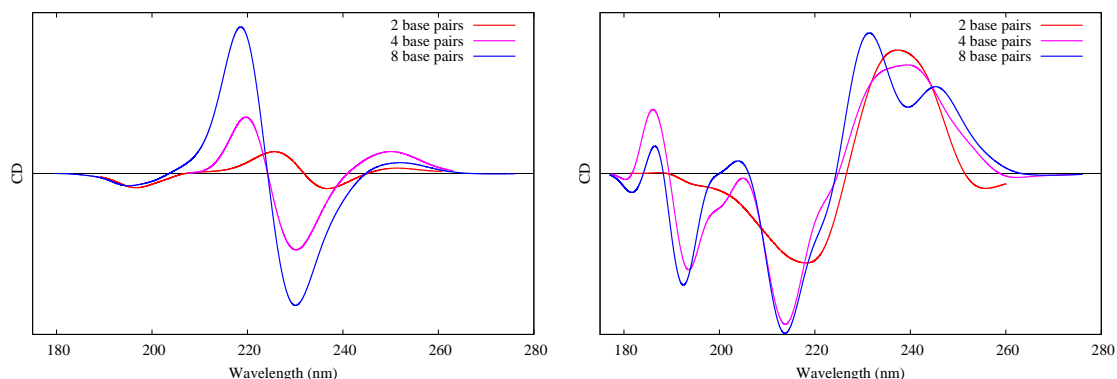


Figure 8.3: Simulated ECD spectra using the Frenkel Hamiltonian for poly(dAT) (left panel) and poly(dCG) (right panel). An increasing number of bases were included in the model Hamiltonian to check the convergence. Wavelengths in nm, intensities in arbitrary units

The spectrum of the poly(dCG) double strand appears again much more complicated, in particular in this case the low-energy band is much larger, extending from 260 to 230 nm, and presents an almost double maxima structure. The first band is followed by a negative band peaking at about 218 nm. Once again, the differences between the eight and four bases model are quite negligible, even if it has to be noted that the eight bases model seems much more suited to capture the complex structure of the first band. Nonetheless, we may quite safely conclude that the four and eight bases models appear as converged and are able to catch the main features

of the DNA ECD spectra.

Indeed, even if a direct experimental comparison appears difficult since ECD spectra are extremely sensible for instance to salt concentration, we may note that our simulated spectra reproduce quite nicely experimental data as the ones reported by Kypr et al. [48], in particular concerning the presence of the low-intensity positive band followed by the important negative one for poly(dAT) strands as well as the large and complex positive band shown by the poly(dCG) strand. It has, however, to be cited that the intensity of the lower-energy positive band evidenced in the experimental poly(dAT) spectrum is strongly reduced in our model.

In all the previous calculations, we have reported results only for bases embedded in the central regions of the double strand, i.e., far from the borders. However, since we are using a finite strand, we may wonder whether the influence of border effects can induce important modification in the calculated ECD spectra. In Figure 8.4, we report the ECD spectra obtained including all the 15 base pairs constituting our two model strands in the Frenkel Hamiltonian. We can immediately see two very different situations, indeed the poly(dAT) strand presents a spectrum that is absolutely coherent with the features evidenced by the eight base pairs model. On the other hand, the poly(dCG) strand gives now a spectrum that looks almost specular compared to the one obtained with the eight base pairs that satisfactorily reproduced the experimental measurements.

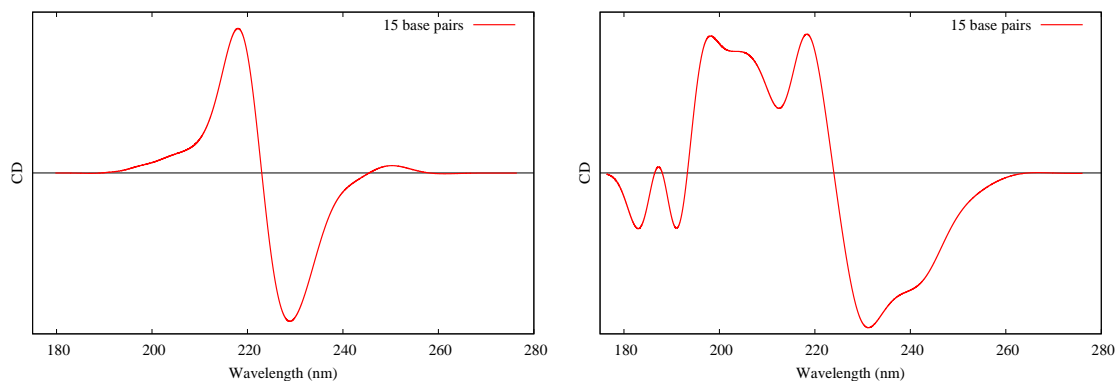


Figure 8.4: Simulated ECD spectra calculated including all the 15 base pairs in the Frenkel Hamiltonian. Poly(dAT) in the left panel and poly(dCG) in the right panel, respectively. Wavelengths in nm, intensities in arbitrary units.

Indeed, the reason for such discrepancy should not be found in a bad convergence of the model, and on the contrary, the 15 base pairs spectrum is indeed plagued by border effects due to the terminal bases. This fact is rationalized in Figure 8.5 in which we see that the general features of the 15 base pairs spectrum are strongly dominated by the edge bases. Indeed, the external bases at the 3' and 5' extremities have a total different spectroscopic signature compared to the central ones and

are characterized by a negative band at higher wavelengths and a positive one at lower ones. Since incidentally the border bases ECD spectrum happens to have an absolute intensity that is almost double compared to the one calculated from the internal bases, the 15 bases spectrum will not be representative of a general and longer poly(dCG) B-DNA fragment.

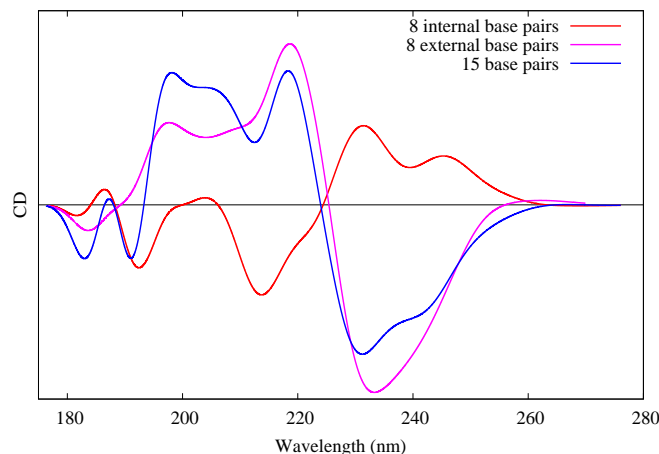


Figure 8.5: Calculated ECD spectra for the poly(dCG) strand considering the 15 base pairs, only the eight central base pairs and the eight base pairs at the extremities. In the last case, we took four base pairs at the 3' end and the four others at the 5' border. Wavelengths in nm, intensities in arbitrary units.

Although no real ejection from the stacked structure of the terminal bases is observed all along the dynamic trajectory, we may observe a larger deviation from the ideal helical structure in the case of the poly(dCG) strand than in the case of the poly(dAT).

This aspect can be also be quantified using the global intra- and inter-base DNA deformation parameters[79]. Indeed, the terminal base of poly(dCG) strands experiences an average deviation from an ideal B-strand of about  $20^\circ$  for the "buckle" and "propel" angle and of about  $14^\circ$  for the "opening" and in contrast, central bases are not exceeding  $10^\circ$  deviations. The same behavior is found for the inter-base parameters with about  $20^\circ$  deviations for "tilt" and "roll" and  $15^\circ$  for the twist angles in the case of the poly(dCG) border bases. On the other hand, all the poly(dAT) parameters experience very limited deviations from the ideal B-DNA structure (smaller than  $10^\circ$ ). Anyway, one would need longer molecular dynamics, with timescale approaching the  $\mu\text{s}$ , as well as longer DNA strands to extract all the structural information related to the border and sequence effect.

Anyway from that last analysis, it is evident that extreme care should be taken to avoid border effects using sufficiently long enough DNA fragments. Luckily enough due to the fast convergence of the Frenkel Hamiltonian ECD with the number of bases pairs, this aspect should not impose too heavy constraints in the calculation

of circular dichroism spectra.

## Conclusion

We have reported the simulation of ECD spectra for two different representative DNA double strands. Our proposed protocol consisted in the combination of molecular dynamic to obtain a reasonable sampling of the configuration space explored by the DNA fragments with QM/MM to obtain the properties of excited states of single chromophores. By comparison with larger QM treatment of multichromophoric ensemble of DNA nucleobases, we have validated the extremely simple Frenkel Hamiltonian approach to obtain a good simulation of ECD spectra. We have also evidenced the necessity to take extreme care to avoid unphysical border effects that could give unreliable spectra. However, we have shown that the convergence with the number of base pairs included in the effective Frenkel Hamiltonian is relatively fast, and hence, it allows an easy and efficient calculation of those properties.

The validation of our protocol gives us a general strategy to study and interpret ECD spectra of complex DNA aggregates, for instance, it will allow us to model the differences induced in DNA spectrum upon interactions with sensitizers, as well as to efficiently study induced circular dichroism. Another possible extension of the protocol will be the study of non-canonical DNA structures such as G-quadruplex, presented in the following chapter.

## 8.2 ECD of G-quadruplexes

### Introduction

G-quadruplexes (G4), first discovered using X-ray diffraction in 1962[138], are considered to be the second most important DNA morphism after B-DNA [139–143]. These *quadruple* helices are composed of guanine tetramers stacks stabilized by the so-called Hoogsteen hydrogen bond network[144]. Between each pairs of plane, the eighth bases form a highly negatively charged region which is always stabilized by a cationic residue, either  $\text{Na}^+$ ,  $\text{K}^+$  or  $\text{NH}_4^+$  [145]. These entities mainly found in telomers or gene transcription areas[146] play a crucial role in cell viability and aging. Indeed, at each replication the DNA strand will be shortened in the telomer region. Then, when the length of the telomeric part has been reduced to a certain critical point, the cell undergo apoptosis [147]. However since telomerase enzymes are involved in cell aging, a defect in their activity would cause cell "immortality" thus leading to cancers[148]. Moreover, due to their peculiar arrangement, new drugs en-



tirely specific to G4 can be developed, thus reducing the cytotoxicity of currently employed chemotherapeutic chemicals[149–154], enhancing the interest toward G4 targeting. Indeed the stabilization of G4 results in the inhibition of telomerase and hence prevents the cell immortality. From their implication in biological processes and their numerous morphisms, a better understanding of G4 structures, dynamics and interactions with ligands is required. To simplify the picture, three representative systems have been considered from each main families of G4, parallel (PDB: 1KF1), hybrid (PDB: 2HY9) and antiparallel (PDB: 143D) (Figure 8.6).

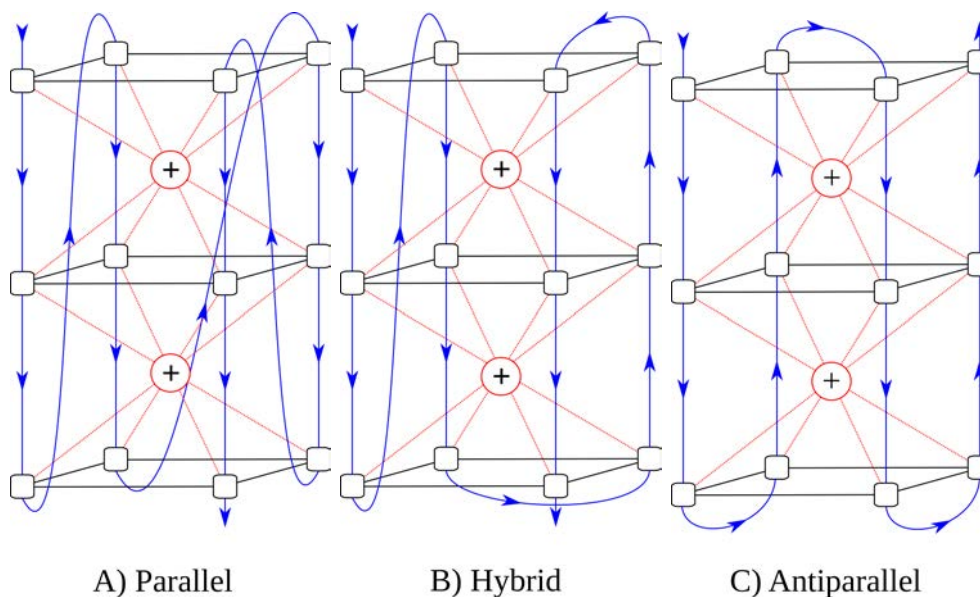


Figure 8.6: Schematic representation of the DNA base and backbone arrangement in typical G4 conformations: (A) parallel, (B) hybrid, and (C) antiparallel.

What characterizes the three main forms of G4 is the orientation of the bases composing its core. The simplest conformation is parallel in which all the guanines are "anti" and are stabilized by a potassium ion. Then hybrid and antiparallel exhibit a more complex pattern since the orientation of the bases is either "syn" or "anti" and they are respectively stabilized by potassium or sodium cations(Figure 8.7). Finally, even though these cores can be considered as extremely rigid, the environments have a major impact on both their stability and polymorphisms. Indeed, a modification of the ion concentration, the temprerature or the presence of a crowding agent will induce transitions from one form to another[155].

Using our previously described protocol (Figure 8.1) we modeled the ECD spectrum of the three G4 polymorphisms. We aim to, first, validate once again the ability of the combination between QM/MM and Frenkel excitons to reproduce ECD signals of biomacromolecules, second provide new understandings of the relation structure/signal of G4. In the future this tool could help in a full comprehension of the interaction of ligands with G4, hence helping in the development of specific



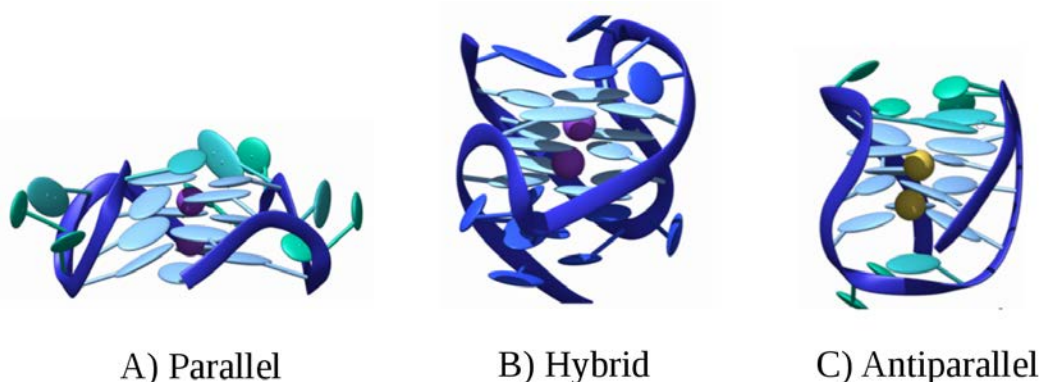


Figure 8.7: Representative structures of the G-quadruplexes in cartoon rendering extracted from the molecular dynamic trajectories. Backbone is represented as ribbon, with basis as ellipsoidal objects. Stabilizing cations are in van der Waals representation, in purple for  $K^+$  and yellow for  $Na^+$  cations. The bases forming the G4 core are in light blue, flanking bases in dark blue and light green: (A) parallel, (B) hybrid, and (C) antiparallel.

therapeutic agents.

## Computational details

All through the present contribution, the simulation of the ECD has been performed at the QM/MM level following a preliminary exploration of the G4 conformational space by MD techniques. In order to achieve a good convergence of the sampling, a rather large number of snapshots has been taken into account. For that reason a relatively small QM partition is necessary in order to avoid computational burdens that are too heavy. To achieve such a compromise, we used the so-called Frenkel excitonic model to take into account the interactions between the chromophores, i.e., the DNA nucleobases. For the reader's convenience we briefly recall the basic foundation of Frenkel excitonic coupling and in particular its application to supramolecular ECD determination[156].

G-quadruplex conformations, parallel (PDB ID: 1KF1), hybrid (2HY9), and antiparallel (143D), were taken from the Protein Data Bank. For each conformation, molecular dynamics (MD) simulations were performed with the AMBER software[132], using the AMBER99 force field with the bsc0 correction[133]. First, the two stabilizing  $K^+$  (in the case of hybrid type) and  $Na^+$  (in the case of antiparallel type), absent in the PDB data file, were added manually between each G-quadruplex plane. Then, oligonucleotides were solvated in a truncated octahedron box with edges of approximately 76 Å. A TIP3P water model was used, and  $K^+$  or  $Na^+$  (in the case of the antiparallel) was added to neutralize the systems. Particle mesh Ewald (PME) and periodic boundary conditions were used throughout. A total of 8000 steps of minimization have been performed to remove close contacts in the G4

and with solvent molecules. The first half of the steps have been performed using steepest descent procedure and the second half with conjugate gradient algorithm. The molecular dynamic simulation has been started with a 200 ps thermalization to reach 300 K in the NVT ensemble. Then, the system is equilibrated in the NPT ensemble over 200 ps in order to achieve a pressure of 1 atm. Finally, 100 ns production runs were performed in the NPT ensemble for each G4 structure. From the trajectories, 100 snapshots were extracted every nanosecond to compute the vertical electronic excitations using a QM/MM method. Different QM partitions can be envisaged as illustrated in Figure 8.8 corresponding to isolated nucleobases (A in Figure 8.8), entire row, with guanine bases corresponding to a G4 stem (B) or the whole planes containing the 4 guanines (C). As seen in appendix (Figure C.3), it definitively appears that the best reproduction of circular dichroism band shape is obtained when using the row partition.

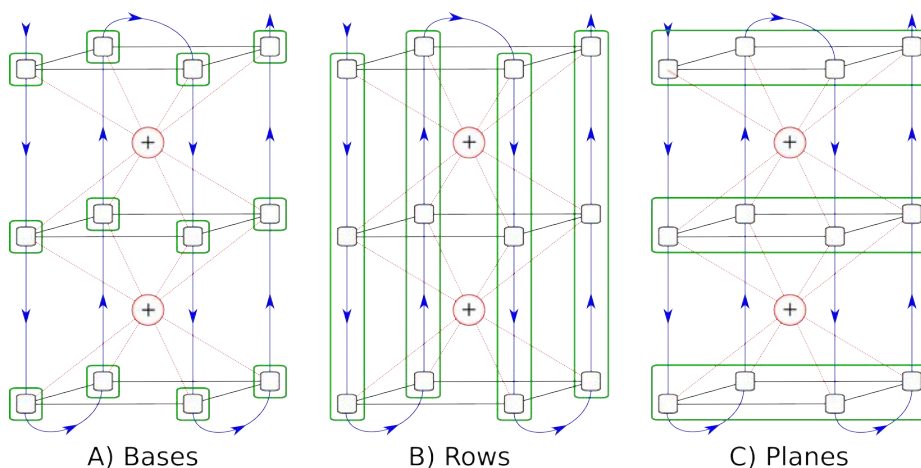


Figure 8.8: Different possible QM partitions for the calculation of G4 ECD in the framework of the Frenkel Hamiltonian. The bases comprising each green box constitute the basic QM monomers on top of which the Frenkel Hamiltonian is built.

The QM/MM frontier was placed at the junction between the sugar and the guanine base, and the dangling bond was treated with the link atom scheme. Time-dependent density functional theory (TDDFT) with the M06-2X functional<sup>[105]</sup> and 6-31G(d,p) basis set is used to describe the different excited states. Meta-GGA M06-2X has been chosen for its capacity in reproducing  $\pi$ -stacked systems and for its good performance with charge-transfer states that may be artificially overstabilized by hybrid functionals for coupled chromophores. The excitation energies of the first 12 excited states of each QM unit were determined together with their electronic transition dipole moments. Test cases performed calculating 12 or 20 excitations per monomers showed no significant deviations. Excitation energies and transition dipole moments were used to build the Frenkel Hamiltonian and simulate the electronic circular dichroism spectra of the different quadruplex structures,

see formalism at 5.5.3. QM/MM calculations were performed using a modified version[111, 134] of Gaussian 09[136], coupled with Tinker[137]. Spectral band shapes have been obtained convoluting each snapshot with a Gaussian function of full width at half-length of 0.3 eV. Note that excitation energies, dipole moments, Frenkel energies, and rotatory strengths have been individually calculated for each snapshot, and the resulting ECD vertical transitions have been convoluted to obtain the final spectrum.

The details on the experimental acquisition of the G4 circular dichroism spectra is given in appendix C.

## Results and discussion

### G4 structural parameters

The G-rich strand of human telomere DNA 5'-AGGGTTAGGGTTAGGGTTAGGG-3' adopts three different conformations, with typical CD signals, depending on the environment[157]. Significant structural parameters for the three conformations (parallel, hybrid, and antiparrallel) have been extracted from the MD trajectories. A global stability all along the trajectory can be evidenced for each system; in particular, as expected for G4, the guanine core region is quite rigid while a much greater mobility is observed for the flanking bases that may in certain cases form a more labile  $\pi$ -stacking interaction with the G4 core. This aspect can be easily inferred by the root-mean-square deviations (RMSDs) reported Figure 8.9; in particular, we may see that the core region has RMSD close to, or even lower than, 1.00 Å; on the other hand the global RMSD may reach a value of 3.00 Å. Antiparallel conformation (Figure 8.9C) is a partial exception; indeed, in this case the flanking bases are much less flexible, resulting in a very low total RMSD.

In Figure 8.12 we report the distribution of the distance between the G4 core planes in the three different conformations. As expected, also due to the rigidity of the G4 core, the distributions have an almost ideal Gaussian shape and are relatively sharp and peaked. The average distance value is around 3.5 Å for parallel and hybrid conformers, while the antiparallel has a slightly larger average distance of about 3.55 Å. Note that the slight difference between the two distances is largely due to small differences in the local environment experienced by the two border planes and most probably by the effect of the flanking bases. The asymmetry between the two edges is larger for the parallel conformations that experience the larger variation of the distance distributions. This aspect also correlates well with the RMSD. Indeed, the antiparallel conformations, that have less flexible flanking bases, present less variability.

Another important feature related to the G4 structure is the occurrence of Hoog-

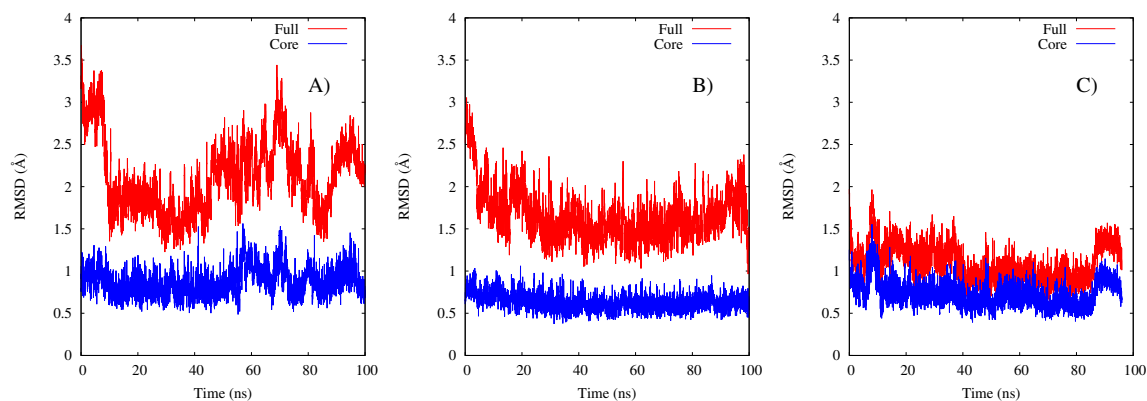


Figure 8.9: Time series of the root-mean-square deviation (RMSD) from the average structure for the three G4 conformations: (A) parallel, (B) hybrid, and (C) antiparallel. Full RMSD is reported in red while the RMSD relative to the core guanines only is in blue.

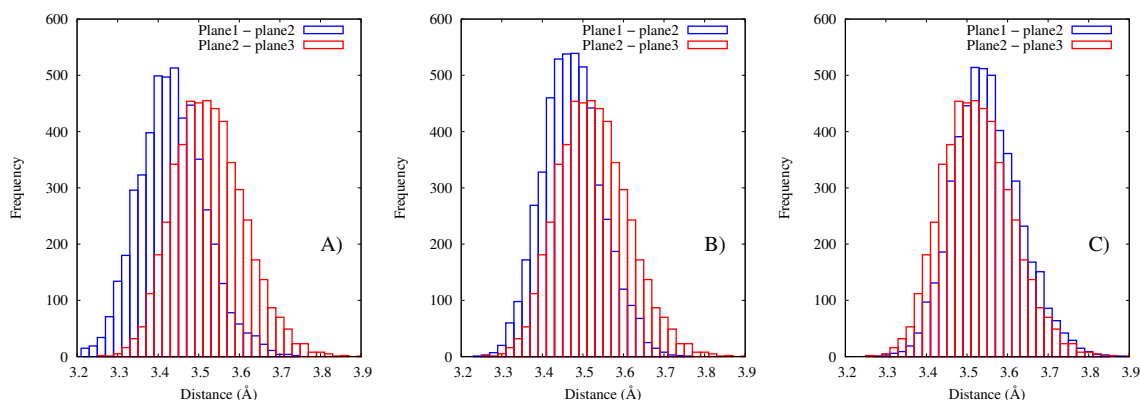


Figure 8.10: Distribution of the distances between consecutive G4 planes as extracted from the MD trajectory. Note that planes are numbered sequentially from the top, see Supporting Information for details: (A) parallel, (B) hybrid, and (C) antiparallel.

steen hydrogen bonds, determining the formation and the stability of the core planes. The average distribution of the O-H distances per plane along the MD trajectory is reported in Figure 8.11. Although once again the stability of the different conformations is confirmed, a rather different behavior may now be evidenced. In particular, a larger deviation between the edge and the central planes is now clearly visible. Indeed, the distribution of the hydrogen bond distances in the central plane is constantly broader for the three conformations. Furthermore, the distribution for the central plane also peaks at a 0.2 Å larger distance than the edge units. This larger variability, that seems correlated to a sort of larger breathing of the central plane (Figure C.5), is due to a larger mobility of the backbone allowing larger vibrational motions, as well as to the influence of the single strand terminal bases. However, all the distances remain very representative of persistent hydrogen bonds and oc-

cur in between 2.0 and 2.3 Å. Moreover, their behavior is extremely similar for the three conformers, with the only small difference occurring for the central plane of the antiparallel structure that presents a maximum shift of 0.1 Å compared to the other two conformers. From the structural analysis we may safely conclude that MD is able to correctly reproduce stable and coherent structures and behavior of the different G4 conformers.

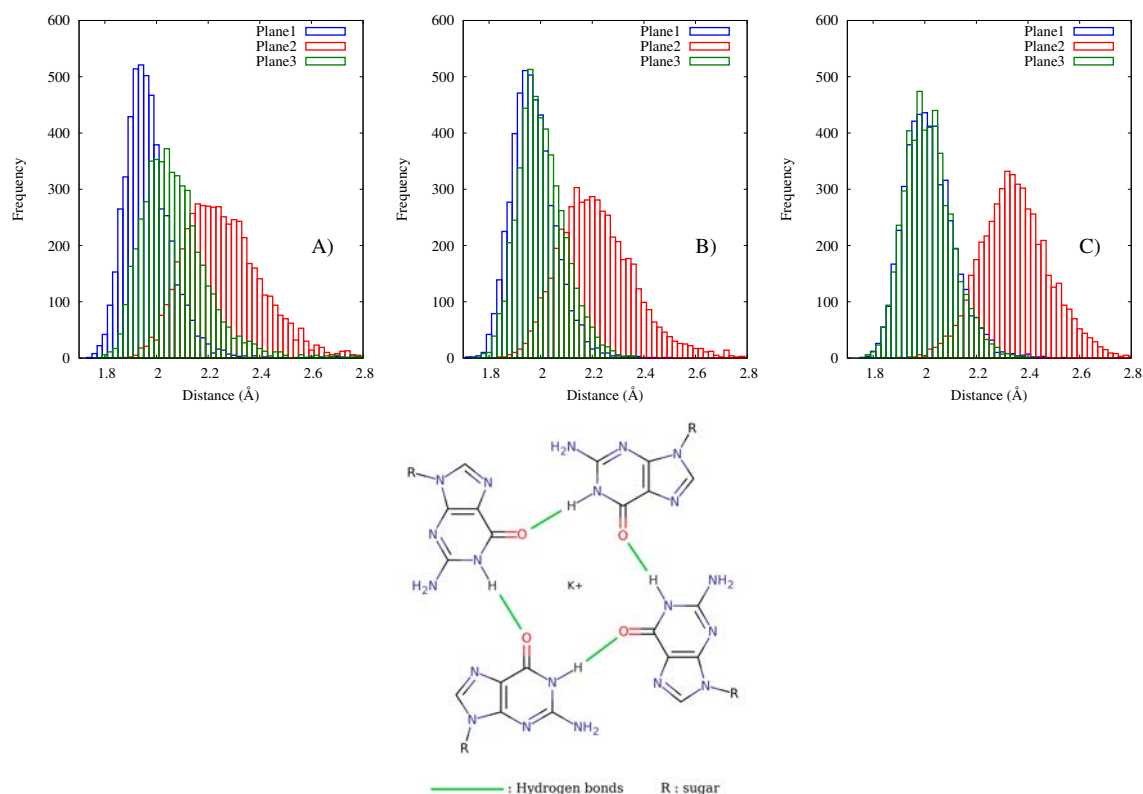


Figure 8.11: Distribution of the distances between the atoms involved in the Hoogsteen hydrogen bonds extracted from the MD trajectory and averaged for each G4 plane: (A) parallel, (B) hybrid, and (C) antiparallel.

### G4 ECD spectra

The comparison between the calculated and measured ECD spectra for the human telomere region in different configuration and in the region above 250 nm is reported in Figure 8.12. First of all, we may notice that the three different conformers present a specific spectroscopic signature that allows the straightforward identification of the different structures. In particular, the parallel conformation (Figure 8.12A) presents a very well-resolved and rather narrow positive band peaking at about 260 nm. On the other hand, the other conformers show more complex spectral features. This is somehow expected for the hybrid structure (Figure 8.12B) since in that case the variability is larger compared to the rather uniform parallel one. The ECD

band is now quite broad covering a large spectral domain going from 265 to 295 nm. Finally, the antiparallel configuration (Figure 8.12C) shows a pretty distinctive signature with a narrower positive band at around 300 nm followed by a poorly resolved shoulder at around 270 nm and a negative less intense band at 260 nm.

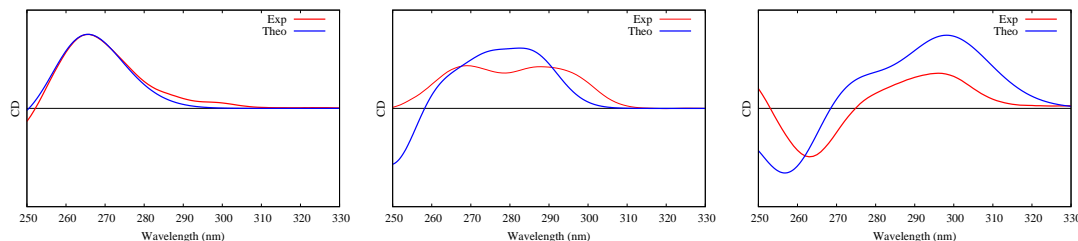


Figure 8.12: Computed (red) and experimental (blue) spectra for the three G4reproduce the maxima wavelengths: (Left) parallel, (Center) hybrid, (Right) antiparallel.

Although the theoretical excitation energies need to be shifted by values between 0.30 and 0.75 eV to match the correct experimental maxima, a procedure commonly employed in the calculation of ECD spectra of similar systems[126, 156], the global band shape is quite well-reproduced for the three different cases. Since we are here dealing with different arrangements of the same molecular building blocks, it appears most probable that the necessity to use different energy shifts should be ascribed to the simplifications made in the Frenkel Hamiltonian rather than to TDDFT deficiencies in catching the electronic nature of the individual excited states. Remarkably, the shape of the parallel systems almost perfectly matches the experimental one while slightly larger differences are observed for hybrid and antiparallel configurations. In particular, the double maximum of the hybrid systems is less well-resolved in the simulated spectrum, while the intensity of the negative peak of the antiparallel G4 is overestimated. However, the reproduction of the experimental spectra allows us to unambiguously assign the configurations on the basis of the sampling performed by MD. This is particularly important since it validates the possibility to use a combination of QM/MM and MD to retrieve structural informations on noncanonical DNA structures and to compare the obtained results with straightforward experimental measurements. On the other hand the combination of theoretical and computational results confirms that the inclusion of crowding agents induces a transition of the human telomere from a hybrid to a parallel structure. The larger deviations from the experiment observed for hybrid and antiparallel configurations are presumably due to the presence of  $K^+$  cations that, compared to  $Na^+$ , induce larger deviations in the G4 structure. It is also to be noted that the main characteristics of the ECD bands are due to the guanine plane organization and not to the effect of the flanking bases; indeed, their inclusion on the Frenkel procedure produces

only a totally negligible change in the spectroscopical features. By inspecting the results reported in Figure C.1, it also appears evident that the higher energy (lower wavelengths) part of the spectrum is less well-reproduced. This can be related to the inherent difficulty of TDDFT to reproduce higher energy excited states but also to the fact that the final ECD spectra are constituted by the delicate equilibrium between positive and negative transitions; the truncation of the excitation manifold based on the number of computed states can have a rather strong border effect on the limiting part of the spectrum. Finally, the comparison of the calculated absorption maxima of guanine alone (240 nm) compared to that of the G4 (274 nm) allows us to give an estimation of the average strength of the excitonic coupling in G4 structures that amounts to about 0.65 eV.

## Conclusion

The combination of spectroscopy and molecular modeling is now becoming an invaluable tool in chemical and molecular biology in order to unravel all the fine mechanisms of biological processes, including drug actions. Thanks to a combined experimental and theoretical approach we have proven that it is possible to link the macroscopic ECD spectra to the all atom structure and dynamics of important and complex aggregates such as G4 from human telomeres. In particular we have been able to identify and assign the corresponding spectroscopic signatures of all the different conformations presented by the polynucleotides. It is noteworthy that our relatively easy and fast protocol allows us to disentangle the rich density of information embedded in the ECD spectra, and hence participate in straightforward and much more powerful structural resolution processes. In particular it will be important to extend the protocol to the study of the interaction between G4, in different conformations, and interacting drugs. The analysis of the native and induced circular dichroism mapped with the structures obtained from MD trajectories will allow us to discriminate between different binding modes and configurations. This in turn will be helpful in the rational design of novel G4-based anticancer drugs. In the future we plan to expand the study determining ECD of DNA interacting with different compounds, both organic and inorganic, and also comparing noncanonical structures, like G4 and B-DNA.



# Chapter 9

## Binding free energy of Benzophenone

### Introduction

The interaction of the photosensitizer benzophenone (BNZ) with DNA, described in the work of Monari et al. [122], showed extremely varied photochemical and photophysical mechanisms upon irradiation with properties mainly depending on its binding conformation, either MinGC or DBI (Figure 9.1). Furthermore BNZ may be regarded as a paradigmatic sensitizer and a good model system to study DNA non covalent interactions. But a crucial point is missing to fully understand and rationalize the possible mechanisms of DNA photosensitization by BNZ, its stability in the binding pocket, hence the ratio between the two binding modes. So far, experimental techniques have not been able to obtain such information, but with the help of molecular modeling and computational chemistry, one can obtain for exemple thermodynamic properties of such systems. In our case, we are interested in the reproduction of the binding free energy of BNZ in DNA.

Already, several DNA/ligand system have been studied employing statistical approaches based on alchemical transformations, either Free-Energy Perturbation (FEP)[158–160] or Thermodynamical Integration (TI)[160]. The results obtained on several minor groove binders (distamycin[161] or netropsin[162]) showed interesting relative binding free energy differences, but the short time scale simulation they employed was not statistically relevant in term of sampling thus resulting in thermodynamic quantities of low quality. To increase the precision of such methods, it is necessary to greatly increase the sampling time through long MD simulations. The basic concept of alchemical transformations is going from an equilibrium state to another which in fact can be realized using classic MD simulations, mainly by introducing an ensemble of restraints that, if well defined, will force the system to evolve efficiently and sample the conformational space between these two states thus



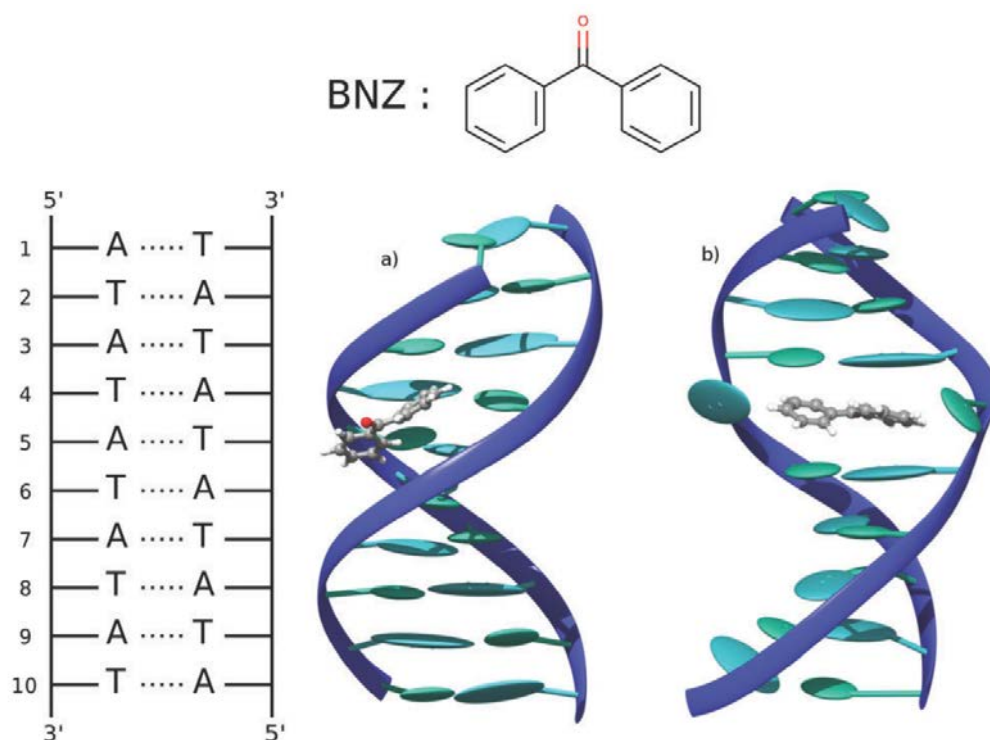


Figure 9.1: Cartoon representations of (a) BNZ in the minor groove binding mode (b) BNZ in the double insertion mode.

giving the opportunity to compute the free energy of the transformation. A set of restraints is often added on translational, orientational and conformational degrees of liberty to reduce the entropy between the binder and the receptacle[163–165]. Technically, the host is both decoupled and then recoupled from its binding site (for the host-guest interaction energy) or from water (for the solvation energy) in its geometrically restrained position. The loss of entropy created in this case will be recovered in subsequent free energy simulations. The accuracy of such methodology has already been proven by Gumbart et al [165] in which a protein-ligand binding free energy has been evaluated through a classic alchemical route.

Here we evaluated the binding energies of both binding conformation of BNZ in B-DNA employing the alchemical transformation methodology. Then the result is compared and rationalized on the available experimental and computational data already available.

## Computational details

### DNA/BNZ systems

Minor groove (MinGB) and double insertion (DI) binding modes have been evidenced and characterized by Dumont et Monari[122]. In the present study, each initial conformation of DNA-BNZ complexes was taken from unbiased molecular

dynamics simulations. In each case, the most representative conformation of a B-DNA double strand decamer composed of alternating thymine-adenine base couples, namely poly(dA-dT)-poly(dA-dT), in complex with BNZ was considered. As depicted in Figure 9.1 BNZ was inserted at the level of base pair 5 for both minor groove binding and double insertion. DNA-BNZ complexes and both DNA and BNZ unbound states were embedded in a cubic box of explicit water molecules. For all cases, the edge of the initial box was 60 Å for a total of about 6600 water molecules[166]. Potassium cations were added as counter ions to neutralize the global charge of the DNA-containing systems.

### MD simulations

DNA double strand was described using the Amber 99 force field[167] including the bsc0 correction[133]. BNZ was modeled using the generalized amber force field (GAFF)[17] consistently with the strategy used elsewhere[122]. Water molecules were described using the TIP3P model[168].

All molecular dynamics simulations were performed using the NAMD program[169], in the NPT ensemble. The pressure (1 atm) and the temperature (300 K) were kept constant using, respectively, the Langevin dynamics and the Langevin piston[170] algorithms. Long-range electrostatic interactions were taken into account using the Particle-Mesh-Ewald method[171]. Van der Waals and electrostatic interactions were truncated for distances longer than 9 Å. An integration time step of 2 fs was used for all simulations. Both systems were first minimized for 100 steps using the conjugate gradient algorithm in order to remove close contacts between atoms. DNA heavy atoms were first restrained by means of an harmonic potential. Restraints were slowly released within 10 steps of 100 ps and a trajectory of 40 ns was produced. The structural reorganization of the 10 base-pair double-stranded DNA upon BNZ binding was analyzed using Curves+[79].

### Alchemical transformations

The standard binding free energy  $\Delta G^0$  of BNZ to DNA was evaluated following the alchemical cycle described in Figure 9.2. For coupling and decoupling BNZ from the hydrated DNA environment or from the bulk, the position of BNZ with respect to DNA was harmonically restrained by means of a set of spherical coordinates ( $r, \theta, \phi$ ), whereas its relative orientation was restrained using three Euler angles ( $\Theta, \Phi, \Psi$ ) (Figure 9.3) using the collective variable software[172] implemented in NAMD[169]. Furthermore, the double-stranded DNA was harmonically restrained to its bound state conformation thanks to a root mean-square deviation (RMSD) collective variable. The force constant of each harmonic potential has been chosen as

low as possible to maintain the system in a conformation identical to that observed in equilibrium simulations (Table D.1). Assignment of the groups of atoms used in the definition of the geometric restraints is arbitrary (Figure 9.3 and Figure D.1), namely P1, P2 and P3 for DNA, and L1, L2 and L3 for BNZ, within the limit that these groups of atoms are not aligned, and their relative position in space fluctuates only marginally. Under these premises, one can show that the binding free energy is an invariant of the geometric restraints.  $\Delta G^0$  can be computed as

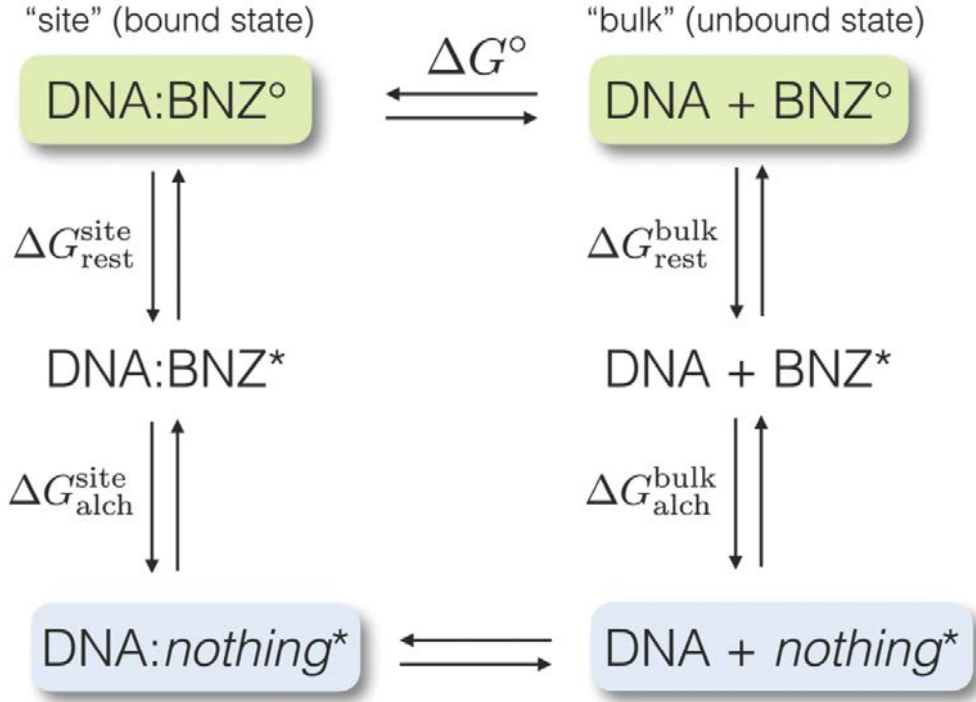


Figure 9.2: Alchemical cycle to determine the binding free energy of the DNA/BNZ complex. \* refers to a restrained state.

$$\Delta G^0 = \Delta G_{rest.}^{site} + \Delta G_{alch.}^{site} - \Delta G_{rest.}^{bulk} - \Delta G_{alch.}^{bulk} \quad (9.1)$$

The purpose of introducing geometrical restraints in the alchemical free-energy calculations is twofold. First, appropriately chosen harmonic potentials acting on translational, orientational and conformational degrees of freedom are expected to reduce appreciably the change in configurational entropy associated with the binding process, making sampling of the latter amenable to finite-length, equilibrium simulations[163, 165]. Second, by tethering it to the binding site, the substrate is prevented from drifting astray as it is decoupled from its environment, which would otherwise violate thermodynamic micro-reversibility[173–175].

The free energy differences associated with the coupling/uncoupling of BNZ from the hydrated double-strand DNA decamer and bulk water,  $\Delta G_{alch.}^{site}$  and  $\Delta G_{alch.}^{bulk}$ .

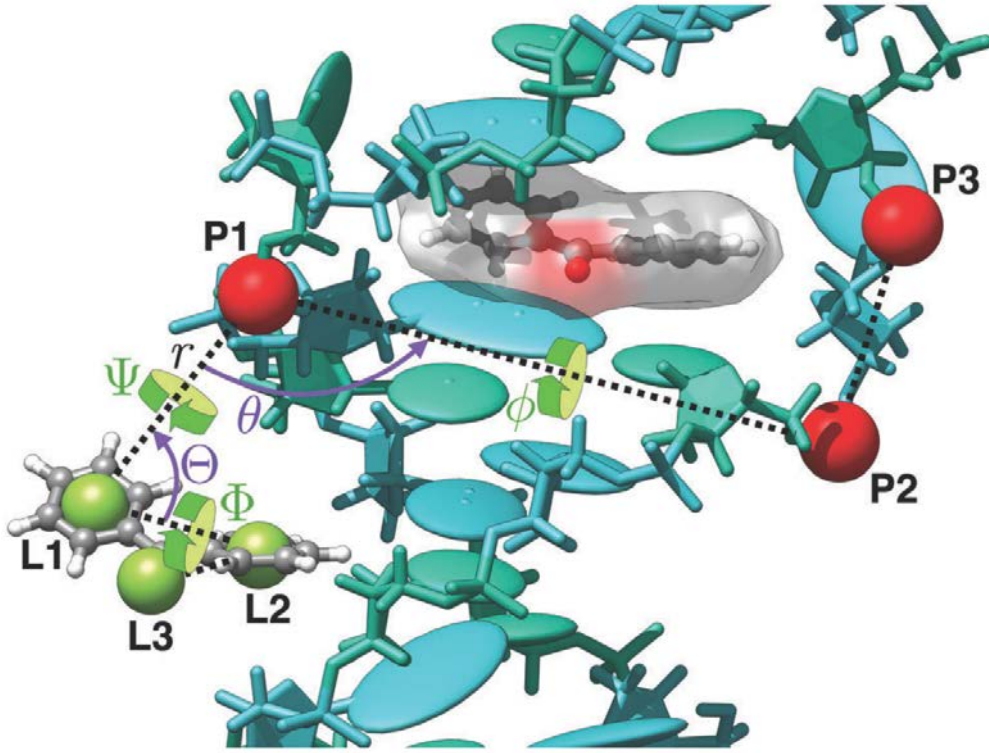


Figure 9.3: Representation of the set of parameters defining the position of BNZ in the double insertion mode. For both DNA and BNZ, a triplepoints, P1, P2, P3 (red spheres) and L1, L2, L3 (green spheres), respectively, is defined arbitrarily. The position of BNZ is defined by three spherical coordinates ( $r(\text{P1-L1})$ ,  $\theta(\text{L1-P1-P2})$ ,  $\phi(\text{L1-P1-P2-P3})$ ), whereas its relative orientation is described by three Euler angles ( $\Theta(\text{P1-L1-L2})$ ,  $\Phi(\text{P1-L1-L2-L3})$ ,  $\Psi(\text{P2-P1-L1-L2})$ ). For clarity, inserted BNZ (gray surface) has been split out from the complex.

respectively, were computed within the Free Energy Perturbation (FEP) framework[158]. The reversible work,  $\Delta G_{rest.}^{site}$  and  $\Delta G_{rest.}^{bulk}$ , associated with BNZ positional/orientational restraints and with DNA conformational restraint, were evaluated using thermodynamic integration (TI) simulations or analytically[164]. The latter contributions are obtained by summing up the contributions of each individual restraint:

$$\Delta G_{rest.}^{site} = \Delta G_{RMSD}^{site} + \Delta G_r^{site} + \Delta G_{\Theta}^{site} + \Delta G_{\phi}^{site} + \Delta G_{\Phi}^{site} + \Delta G_{\Psi}^{site} + \Delta G_{\Theta}^{site} \quad (9.2)$$

and

$$\Delta G_{rest.}^{bulk} = \Delta G_{RMSD}^{bulk} + \Delta G_{r,\theta,\varphi}^{bulk} + \Delta G_{\Theta,\Phi,\Psi}^{bulk} \quad (9.3)$$

$\Delta G_{r,\theta,\varphi}^{bulk}$  and  $\Delta G_{\Theta,\Phi,\Psi}^{bulk}$  are obtained through numerical integration of analytical expressions (see D). In practice, the free-energy differences between the different

states were evaluated through a series of transformations between non-physical intermediate states (windows) connecting the initial to the final state by means of a coupling parameter  $\lambda$ , varying from 0 to 1. All FEP and TI simulations were performed in both decoupling and coupling directions except for the DNA unbound state system in which the work associated with the RMSD restraint,  $\Delta G_{RMSD}^{bulk}$ , was averaged over a set of 14 independent uncoupling simulations. For FEP simulations, the maximum-likelihood estimator of the free-energy difference was evaluated using data collected in both directions by means of the Bennett Acceptance Ratio (BAR) using the ParseFEP plugin[176] of VMD[177]. Both  $\Delta G_{alch.}^{site}$  and  $\Delta G_{alch.}^{bulk}$  FEP simulations were performed under the same conditions but using different stratification strategies tailored to ensure proper convergence of the free-energy estimates. Each  $\lambda$  window consists of 200 ps of equilibration followed by 800 ps of production. For  $\Delta G_{alch.}^{site}$ , 200 windows were used in the case of the double insertion mode for both coupling and uncoupling simulations (400 ns), whereas 80 windows were employed for minor groove binding (160 ns).  $\Delta G_{alch.}^{bulk}$  was obtained by stratifying the coupling and uncoupling of BNZ to the bulk in 63 windows (126 ns). Contributions to  $\Delta G_{rest.}^{site}$  (eqn (2)) were determined through TI simulations stratified over 24 windows in both coupling and uncoupling directions (48 ns). For  $\Delta G_{rest.}^{bulk}$  (eqn (3)),  $\Delta G_{RMSD}^{bulk}$  was obtained through a series of 14 TI simulations performed only in the uncoupling direction (334 ns), whereas  $\Delta G_{r,\theta,\varphi}^{bulk}$  and  $\Delta G_{\Theta,\Phi,\Psi}^{bulk}$  were calculated through numerical integration of analytical expressions[163, 178]. In the case of FEP simulations, a soft-core potential was employed in which the interatomic distance,  $r$ , used in the Lennard-Jones potential, was shifted  $r_2 \rightarrow r_2 + 5 * (1 - \lambda)$ . Furthermore, in order to avoid the so-called "end-point catastrophe", electrostatic interactions of vanishing or appearing particles were linearly coupled to the simulation for  $\lambda$ , varying from 0 to 0.5. At  $\lambda$  values greater than 0.5, electrostatic interactions were fully decoupled from the simulation.

## Results

### Structural analysis for DNA:BNZ binding modes

We provide here a characterization of DNA/BNZ binding modes based on specific parameters, namely the opening of the base pairs, Figure 9.5, and the  $\zeta$  dihedral of the DNA backbone, Figure 9.4, both compared for base pairs 4, 5 and 6 showing the deviation from their canonical position. First of all, the time evolution of the opening parameter shows good stability in the position of the two ejected bases with an angle of around  $120^\circ$ . The higher intensity in the distribution of this value for base pair 5 compared to 4 and 6 is due to the lack of stabilization of these two nucleobases. In fact while in a canonical environment hydrogen bondings and

$\pi$ -stacking interactions stabilize strongly the structure of DNA, in that case the bases are dangling in the water bulk and thus are able to move freely. Moreover, this double ejection can also be clearly described by the movement induced in the  $\zeta$  dihedral angle of the backbone structure. To follow the ejection of bases, the dihedral flips over and goes from between  $-60^\circ$  and  $-80^\circ$  to around  $90^\circ$ .

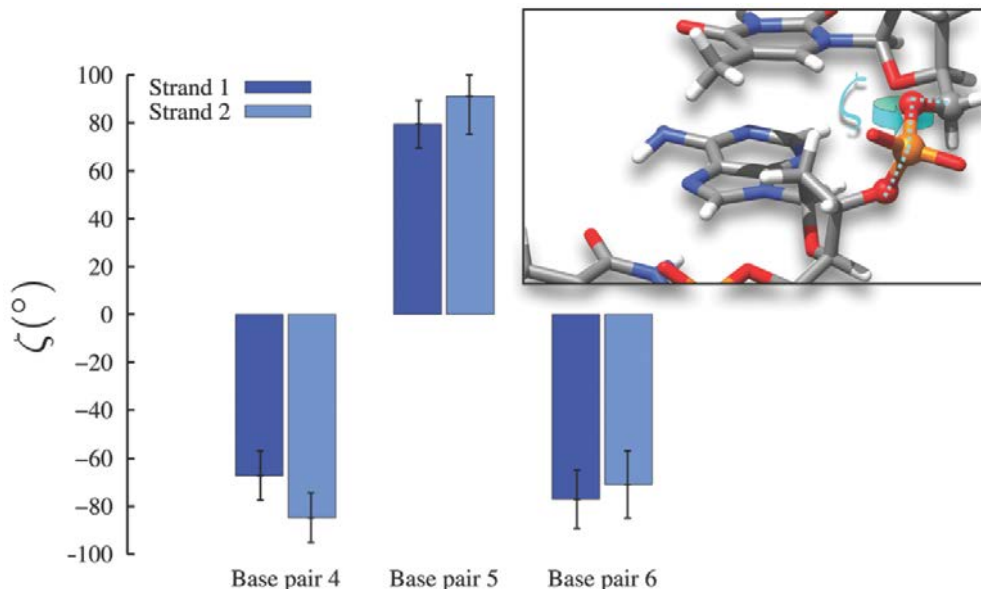


Figure 9.4: Variation over time of the opening structural parameter in the double insertion mode.

In the case of minor groove binding both local and global DNA structures are much less affected by interaction with BNZ. Only a slight deformation of the backbone width and length can be evidenced during the simulation. Moreover along the 40 ns of molecular dynamic simulation, BNZ slides in the minor groove to reach another equivalent and stable binding position. This sliding phenomenon is not specific to BNZ and has been described previously[179]. The study reported the existence of multiple binding stations, thermodynamically equivalent and separated by free energy barriers of  $6 \text{ kcal.mol}^{-1}$  at most, suggesting that sliding is a likely process.

### Binding free energies

Individual contributions (Figure 9.1, 9.2, 9.3) to the free energy of binding of BNZ to DNA are reported in Table 9.1 for both minor groove and double insertion binding motifs. Systematic errors have been estimated from the hysteresis between backward and forward alchemical transformations with the exception of the work arising from restraining, in the bulk, the conformation adopted by DNA in the double insertion complex ( $\Delta G_{RMSD}^{bulk}$ ). Starting from DNA canonical conformation, coupling the latter restraint did not allow the recovery of the target conformation where two bases are



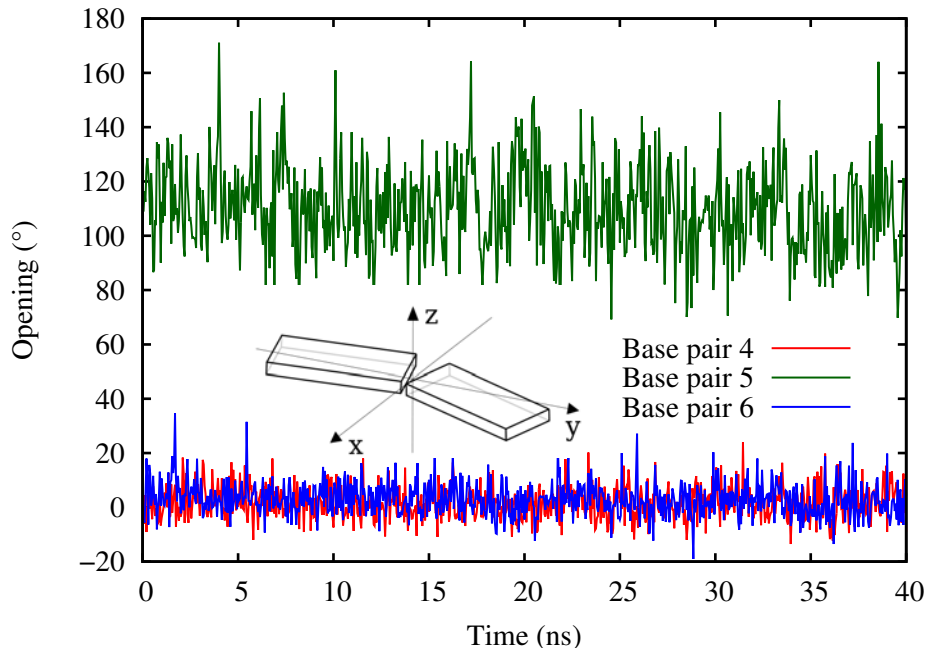


Figure 9.5: Average  $\zeta$  dihedral parameter over the entire molecular dynamic simulation for both strands in the DI binding mode.

expelled out from the double stranded structure. This issue stems essentially from the high level of degeneracy associated with the RMSD collective variable.  $\Delta G_{RMSD}^{bulk}$  was therefore averaged over a series of 14 independent alchemical transformations, where the RMSD restraint was decoupled.

Free-energy contributions of BNZ positional and orientational restraints are small in the bound state,  $\sim 1$  kcal.mol $^{-1}$  or less, in line with the very soft harmonic potentials employed in the simulations (Table D.1 and Table D.2). The hysteresis between the corresponding coupling and uncoupling simulations never exceeds at most a few tenth of kcal.mol $^{-1}$  (Figure D.2). The bulk contributions,  $\Delta G_{r,\theta,\phi}^{bulk}$  and  $\Delta G_{\Theta,\Phi,\Psi}^{bulk}$  computed analytically, are about one order of magnitude larger than their bound state counterparts. Restraining the conformational space of the double stranded DNA in the bound state contributes by  $\sim 10$  kcal.mol $^{-1}$  ( $\Delta G_{RMSD}^{site}$ ) for both minor groove and double insertion binding motifs. Because the conformation of DNA in the bulk is identical to that observed in the complex where BNZ is bound to the minor groove  $\Delta G_{RMSD}^{bulk}$  is close to  $\Delta G_{RMSD}^{site}$ . Conversely, maintaining the distorted conformation adopted by DNA when BNZ is doubly inserted requires twice more energy in the unbound state than in the complex. In the case of the double insertion mode, both  $\Delta G_{RMSD}^{bulk}$  and  $\Delta G_{RMSD}^{site}$  are affected by a relatively large systematic error compared to the other contributions, 1.7 and 0.9 kcal.mol $^{-1}$ , respectively (Figure D.3).

Free-energy associated with coupling BNZ to the hydrated DNA double strand,  $\Delta G_{alch.}^{site}$ , is twice larger for minor groove binding than for double insertion (Figure

Table 9.1: Binding free energy contributions for DI and MinGB binding motifs.

Contributions	DBI		MinGB	
	Free energies (kcal.mol <sup>-1</sup> )	Time (ns)	Free energies (kcal.mol <sup>-1</sup> )	Time (ns)
$\Delta G_{RMSD}^{site}$	12.8 ± 0.9	48	8.6 ± 0.3	48
$\Delta G_r^{site}$	0.1 ± 0.0	-	0.1 ± 0.0	-
$\Delta G_\Theta^{site}$	0.8 ± 0.0	-	0.9 ± 0.0	-
$\Delta G_\Phi^{site}$	1.1 ± 0.0	-	0.9 ± 0.0	-
$\Delta G_\Psi^{site}$	1.0 ± 0.1	-	0.7 ± 0.1	-
$\Delta G_\theta^{site}$	0.2 ± 0.0	-	0.6 ± 0.0	-
$\Delta G_\phi^{site}$	0.6 ± 0.0	-	0.4 ± 0.0	-
$\Delta G_{alch.}^{site}$	-15.5 ± 0.1	400	-30.0 ± 0.1	160
$\Delta G_{alch.}^{bulk}$	-35.8 ± 0.0	126	-35.8 ± 0.0	126
$\Delta G_{RMSD}^{bulk}$	23.4 ± 0.0	14 * 24	10.3 ± 0.3	48
$\Delta G_{r,\theta,\phi}^{bulk}$	4.7	-	4.6	-
$\Delta G_{\Theta,\Phi,\Psi}^{bulk}$	7.2	-	6.6	-
$\Delta G^0$	1.6 ± 1.9	910	-3.6 ± 0.4	382

D.4). Backward and forward estimates of  $\Delta G_{alch.}^{site}$  and  $\Delta G_{alch.}^{bulk}$  present a marginal hysteresis, resulting in very small errors (see Table 9.1 and Figure D.5). Convergence of all FEP simulations was checked thanks to the ParseFep plugin[176] of VMD (Figure D.6).

## Discussion

The precise determination of the binding free energy between DNA and a paradigmatic, yet not totally resolved photosensitizer, is of extreme importance to allow the prediction or rationalization of the subsequent chemical and photochemical reactions and hence to assess the possible further DNA damages as it will be clearly illustrated in the following chapters. Indeed it has been shown that the photochemical pathways strongly depend on the coupling between DNA and the photodrugs, and hence on the structure and the binding mode[123, 180]. In particular, in the case of benzophenone, we will show that hydrogen abstraction from the backbone, a minor photosensitization mechanism, is possible only in minor groove-binding[180], while triplet-triplet energy transfer is in principle possible from both modes[123]. Furthermore, as shown experimentally by Miranda’s group on artificial model systems allowing to clearly dictate the relative position between benzophenone and thymine nucleobases[181], the structure of the BNZ:thymine aggregates may open the door to the formation of excited state excimer complexes that may ultimately



alter the final product distribution also in the case of energy transfer driven phenomena. Hence, the method we provide here allow a preliminary and quantitative discrimination that can orientate the following studies on the photo-reactivity of the drug, as such its importance cannot be underestimated. As a matter of fact it is useful to remind that benzophenone is not only a very useful model system for the study of DNA photosensitization but is also a widespread pollutant, notably present in cigarette smoke. Ketoprofen, a very common antinflammatory drug, also share the core structure, and the photoactive core, with benzophenone hence it can have similar binding modes and subsequently induce similar DNA (photo)sensitization. Our results have allowed for the first time to resolve the benzophenone:DNA binding modes. Specifically we have shown that benzophenone tends to favor minor groove binding that shows a binding free-energy of about  $-3.6 \text{ kcal.mol}^{-1}$  (Table 9.1). It is important to note that the overall statistical error, related to the conformational space sampling, is particularly low and accounts for  $0.4 \text{ kcal.mol}^{-1}$ . There is no experimental binding constant available for BNZ:DNA; however our data are consistent with DNA binding affinities measured for larger organic minor groove binders - e.g.  $-7.7 \text{ kcal.mol}^{-1}$  for Hoescht 3328 and  $-8.6 \text{ kcal.mol}^{-1}$  for Netropsin[182]. Since the possibility of energy transfer from the minor groove binding mode has been clearly evidenced[123], especially via the photophysical population of the benzophenone  $T_2$  excited state[180, 183] our results are consistent with the well known photosensitization characteristics of benzophenone. On the other hand the double insertion mode appears much less favorable; indeed not only the binding energy has been found to be slightly positive ( $1.6 \text{ kcal.mol}^{-1}$ ), but also the formation of double-insertion is most probably characterized by a much larger kinetic barrier. However, due to the relatively high statistical errors ( $1.9 \text{ kcal.mol}^{-1}$ ) and difficult sampling of the complex DNA reorganization, double insertion cannot be totally ruled out. We may however certainly affirm that the former is, at most, participating only with a very minor, and even marginal role in the chemistry of benzophenone interacting with DNA.

## Conclusion

We have reported the thermochemistry, and in particular binding free energy, for the interaction between DNA and the paradigmatic sensitizer benzophenone. The free energy has been obtained using atomistic molecular dynamics and in particular using recently developed methods allowing proper sampling of the conformational space. In particular we have been able to clearly resolve between the two possible interaction structures hypothesized previously, i.e. minor groove binding and double insertion. Indeed, the lack of experimentally resolved structures leaves open the

possibility of the coexistence of the two modes that happen to have very different photochemical and photophysical behaviors. Our simulation, on the other hand, has clearly pointed toward a strong preference for minor groove binding that should, hence, be considered as the predominant species. From a more methodological point of view we have been able to present a rigorous method to provide thermochemical values of complex DNA/drug aggregates. The great flexibility of B-DNA results in very difficult sampling, which in some cases may lead to relatively large errors. This is also due to the fact that many degrees of freedom, in particular, on the backbone dihedrals, are coupled and entangled in a non-trivial way, while global variables such as RMSD may lead to degeneracy.

# Chapter 10

## Photosensitization of DNA

### 10.1 Benzophenone: Hydrogen abstraction

#### Introduction

The next key aspect that has to be rationalized is the relative stability of bound photosensitizers in DNA in order to differentiate the most probable photosensitization routes that might be followed upon excitation of the photodrug. In this approach we studied a known photosensitizing agent of DNA, that exhibited a series of mechanisms, Benzophenone (BP)[184]. This compound is known to interact with DNA through two competitive binding modes[122], either minor groove binding or double insertion, which is a particular binding configuration in DNA in which the molecule replaces a couple of nucleobases, both being ejected outside the canonical Watson-Crick base pairing. Another important information, necessary to understand the action of BP on DNA is to determine the energetics (free energies) of the competition between these two binding modes, this will be presented and discussed in 9. Experimentally, BP, absorbing light in the UVA region, is able to induce several lesions in DNA through triplet-triplet energy transfer, electron transfer, strand breaks and hydrogen abstraction[184]. Moreover, the hydrogen abstraction sensitization can be followed by an electron-proton transfer from a nearby thymine[185]. It has been demonstrated in our laboratory that both binding configurations can lead to the population of its triplet state[123]. Moreover, the complexity of BP photo-physics[186] has required a tremendous amount of studies, both computational and experimental [187–191].

As stated before, the photosensitization of DNA by BP starts with the population of its triplet state. Here we focussed on the possibility of a hydrogen abstraction from this electronic state, Figure 10.1, either from the closest sugar of the DNA backbone (S) in the case of minor groove binding, or from a methyl of the nearby thymine (Thy) when BP is double inserted. Following the hydrogen abstraction,

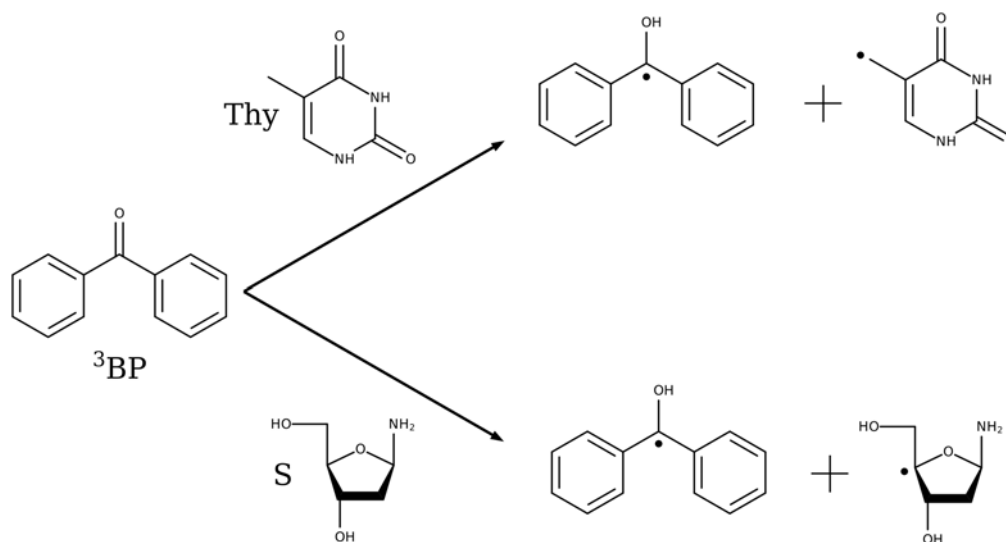


Figure 10.1: Sketch of the molecular structure of BP and the reaction mechanisms for H-abstraction by  $^3\text{BP}$  from thymine (Thy) and sugar (S).

reactions might occur in the DNA structure that would produce lesions such as strand breaks or base oxydations. Using TD-DFT methods, we rationalized the possible mechanisms of hydrogen abstraction from Thy and S by BP in its first triplet state, more specifically we obtained the free energy of the reactive species, the product and the transition state, allowing to understand the energetic barrier that has to be overcome for such mechanism.

## Computational details

The two model systems (BP/S and BP/Thy) were studied in the gas phase and water solution. The solvent was taken into account applying the integral equation formalism variant (IEF) within the Polarizable Continuum Model (PCM)[192]. Reactants, transition states, and product geometries were optimized in the gas phase using density functional theory (DFT), by the hybrid meta-exchange-correlation M06-2X functional.[105] In order to take into account the basis set effect, the 6-31+G(d) and the more accurate 6-311++G(d,p) basis sets were employed. Showing that the M06-2X/6-31+G(d) level of theory is in remarkable agreement with the higher M06-2X/6-311++G(d,p) level, the former was used to calculate free energy profiles, within the rigid rotor and harmonic oscillator approximations. Ab initio multiconfigurational methods were also employed, calculating the vertical excitation energy of the M06-2X/6-311++G(d,p) optimized structures using second-order perturbation theory applied to the complete active space self-consistent field theory (CAS-PT2),[193],[194] i.e. a CAS-PT2//M06-2X approach. The use of a multiconfigurational active space method allowed us to appropriately assess the effects of

both static and dynamic correlations as well as to tackle spin contamination issues for the open-shell products. As a matter of fact this procedure also offers a sound validation of the DFT approach. For both BP/S and BP/Thy systems, the CAS was selected so as to include 14 electrons in 13 molecular orbitals. In particular, the  $n$  orbital centered on the oxygen atom of BP, the  $\pi$ - and  $\pi^*$ -orbitals mainly localized in the carbonyl group of BP, the  $\pi$ -HOMO and the  $\pi^*$ -LUMO of Thy (or the  $\sigma$ -HOMO and the  $\sigma^*$ -LUMO of S), and additional four couples of  $\pi$ - and  $\pi^*$ -orbitals were included (see details in the Figure E.1). Moreover, in the case of BP/Thy, the RAS-PT2 method[195],[196] was employed to account for all the valence  $\pi$ - and  $\pi^*$ -orbitals, including the five most relevant orbitals in the RAS2 space, 10  $\pi$ -orbitals in the RAS1 space and 8  $\pi^*$ -orbitals in the RAS3 space. A maximum of two holes and two electrons were allowed in the RAS1 and RAS3 spaces, respectively, to generate the configuration state functions, i.e. RAS-PT2(26,2,2;10,5,8). The performance of this level of theory in the characterization of the triplet energy-transfer process and its comparison with the previously employed RAS-PT2(20,2,2;7,5,11) level[123] are given in the Figure E.2. Both RAS-PT2 approaches give rise to the same results, within the error of the method. The RAS-PT2 method was not employed for the BP/S system, since S (unlike Thy) does not own a  $\pi$ -conjugated system. In all CAS-PT2 and RAS-PT2 calculations, the ANO-S-VDZP basis set was employed, as it was previously shown to correctly describe the BP/Thy system[123]. Additionally, the basis set superposition error (BSSE) was calculated at DFT, CAS-PT2, and RAS-PT2 levels (see details in the E). All DFT calculations were performed using the Gaussian 09 program,[136] while all CAS-PT2 and RAS-PT2 calculations were performed using the Molcas 8 code[197].

All molecular dynamics simulations were carried out using the GPU CUDA version of amber15[132]. Starting structures for both double insertion and minor groove binding cases were obtained from the stable structures published by Dumont and Monari.46 DNA-BP complexes were placed in a  $67 \text{ \AA}^3$  side length truncated octahedron, containing 5575 TIP3P[168] water molecules. Periodic boundary conditions were applied to simulate the bulk phase and a cutoff was applied to turn off the interactions between atoms that are farther than  $9.0 \text{ \AA}$  apart. In order to neutralize the system, 18  $\text{K}^+$  cations were placed in the simulation box. The DNA strand was modeled using the amber99 force field parameters66 including the bsc0 correction67 developed to specifically reproduce backbone dihedrals of nucleic systems. For BP, the GAFF parameters[17] were used for bonded interactions and the atomic charges were calculated using the RESP[198] fitting protocol using Hartree Fock quantum calculations using the 6-31G(d) basis set commonly used for organic compounds. After an energy minimization of 8000 steps, the first 4000 steps using the steepest descent algorithm, followed by 4000 steps of the conjugate gradient algorithm, the

molecular dynamics started with a 200 ps heating to bring the system to 300 K in the NVT ensemble, followed by a 200 ps equilibration MD in the NPT ensemble at 300 K and 1 atm to obtain the correct density. Finally 100 ns of MD simulation production were performed in order to validate the stability of the binding mode and extract the relevant O-H distances. Along all MD simulations the shake algorithm was applied to bonds involving hydrogens. Moreover Langevin dynamics were used to maintain the temperature and during the production run the pressure was kept fixed using the Monte Carlo thermostat. The MD simulation protocol was also validated analyzing the behavior of the DNA double strand in the absence of BP. Although minimal salt concentration and a short strand were used all along the trajectories we evidenced only negligible deviations from B-DNA ideal structural parameters compatible with thermal fluctuation deviations. Finally QM/MM simulations were performed using a local modified[134] version of the Gaussian09 code including mechanical, electrostatic and polarizable embedding[111]. For the QM/MM calculations, and coherently with the MD we used the amber99 force field to model the environment, while DFT with the M06-2X functional and the 6-31+G(d) basis set was chosen for the QM part. The QM region was composed of BP together with sugar or thymine, while covalent dangling bonds were treated with the link-atom formalism. More details of the QM/MM partition are given in the Figure E.3.

## Results and discussion

The energy profiles ( $\Delta E$ ) for the H-abst from Thy obtained at different levels of theory are reported in Table 10.1. The energetic baseline is taken as the energy of the  $\pi$ -stacked reactants, namely <sup>3</sup>BP and the thymine moiety. DFT calculations provide an energy barrier of about 0.50 eV (11.5 kcal.mol<sup>-1</sup>), while the stabilization energy for the product, i.e. the interacting biradical system, is quite high and amounts to ca. 0.80 eV (18.4 kcal.mol<sup>-1</sup>). Remarkably enough, the results are quite insensitive to the basis set and going from an augmented double- $\zeta$  to a double-augmented-triple- $\zeta$  basis set has only a very small effect (less than 0.05 eV) on both activation energies and the driving force. The same relative small effect is evidenced in the case of the inclusion of the water environment by means of a continuum model PCM. Indeed the energy profile is only slightly altered, with the activation energy reduced by 0.05 eV.

CAS-PT2 results, as well as RAS-PT2 results, definitively confirm the above pictured scenario. The activation energy is now found to be 0.1 eV higher compared to the DFT estimate. Furthermore, almost the same results are obtained for CAS-SCF and RAS-SCF reference wavefunctions. We would also like to highlight that

Table 10.1: Relative energies and free energies (eV) for the abstraction of hydrogen from thymine by BP. Results are shown for the isolated molecules, reactants, transition states, and products computed at different levels of theory. The values in parentheses are computed in the solvent (water) phase.

Method	<sup>3</sup> BP/ <sup>1</sup> Thy	R	TS	P	<sup>2</sup> BP/ <sup>2</sup> Thy
		En.	$\Delta E$		
M06-2X	0.40	0.00	0.50	-0.77	-0.29
6-31+G(d)	(0.38)	(0.00)	(0.44)	(-0.78)	(-0.38)
M06-2X	0.40	0.00	0.50	-0.80	-0.23
6-311++G(d,p)	(0.39)	(0.00)	(0.44)	(-0.80)	(-0.38)
CAS-PT2 (14,13)	0.56	0.00	0.59	-0.71	-0.22
RAS-PT2 (26,2,2;10,5,8)	0.37	0.00	0.60	-0.84	-0.48
		Free En.	$\Delta E$		
M06-2X		0.00	0.43	-0.72	
6-31+G(d)		(0.00)	(0.39)	(-0.70)	

the effect of basis set superposition errors (BSSE) that could be rather important in the case of a  $\pi$ -stacked system has been assessed through a counterpoise correction. As can be evidenced from the tables reported in E these effects are extremely small for DFT calculations, while they are larger in the case of the perturbative treatment; however it only affects the stabilization energy of the reactants/products aggregates compared to the isolated fragments, while the effect on the activation barrier is absolutely negligible. Finally, the calculation, at the DFT level, of the free energy profile ( $\Delta G$ ) using the rigid rotor approximation, particularly the inclusion of entropic effects slightly lowers the activation energy barrier that is now found at 0.43 eV (9.9 kcal.mol<sup>-1</sup>) at 298 K and the DFT level. It has however to be reminded that the rigid rotor model is a useful but quite crude approximation as such free energy values are only semi-quantitative. More sound values could be obtained by a statistical sampling of the conformational space performed by ab-initio MD. However, in the present case it is remarkable that rigid-rotor free-energies basically confirm the purely energetic picture.

The same considerations can be drawn in the case of C4'-hydrogen abstraction on the 2-deoxyribose moiety (Table 10.2). Here, only the abstraction of the most reactive hydrogen has been considered coherently with the fact that the C4'-hydrogen is the most exposed to the interaction with BP when the latter is bound to DNA, as shown in the subsequent section. Once again the energy profile does not crucially depend on the choice of the basis set. An even lower activation energy barrier has been found (0.40 eV (9.2 kcal.mol<sup>-1</sup>) in the gas phase and 0.35 eV (8.1 kcal.mol<sup>-1</sup>) in water) coupled to a driving force (exothermicity) accounting for almost 1.00 eV (23.0 kcal.mol<sup>-1</sup>). In this case, the difference in the CAS-PT2 barrier is slightly larger

than for abstraction of the methyl carbon of thymine, and most notably CAS-PT2 confirms the ease of H-abst, with an even lower barrier than the one estimated using DFT. This being said, the observed differences lie within the error bars of the different methods. Note that in the case of S abstraction we did not resolve to RAS-PT2 calculations since in this case the sugar moiety does not involve any  $\pi$ -orbital and consequently, the active space size and the choice are much less problematic. On the other hand, the inclusion of entropic factors and the calculation of Gibbs free energy  $\Delta G$  profiles, estimated via the rigid rotor approximation, contribute to lower the  $\Delta G$  by a slightly larger amount than in the case of Thy abstraction. However, since the barriers are quite low we can still safely say that the two approaches give absolutely coherent pictures. Indeed, the DFT free energy activation barrier for S abstraction is found to be less than 0.3 eV while the driving force, i.e. the product stabilization remains close to 1.0 eV.

Table 10.2: Relative energies and free energies (eV) for the abstraction of hydrogen from sugar by BP. The results are shown for the isolated molecules, reactants, transition states, and products computed at different levels of theory. The values in parentheses are computed in the solvent (water) phase.

Method	<sup>3</sup> BP/ <sup>1</sup> Thy	R	TS	P	<sup>2</sup> BP/ <sup>2</sup> Thy
		En.	$\Delta E$		
M06-2X	0.30	0.00	0.43	-0.94	-0.29
6-31+G(d)	(0.25)	(0.00)	(0.32)	(-0.98)	(-0.34)
M06-2X	0.31	0.00	0.46	-0.95	-0.29
6-311++G(d,p)	(0.25)	(0.00)	(0.35)	(-1.00)	(-0.34)
CAS-PT2 (14,13)	0.12	0.00	0.13	-1.32	-0.83
		Free En.	$\Delta E$		
M06-2X		0.00	0.28	-0.95	
6-31+G(d)		(0.00)	(0.22)	(-0.93)	

In Figure 10.2 we simultaneously compare the free energy profiles for Thy and S H-abst in the gas phase and solution. Both activation energies appear small, with the reaction involving sugar favored by a barrier reduced by more than one-third compared to the thymine case. As already underlined, solvent effects generally tend to lower, although slightly, the free energy barrier. Interestingly, such barriers are absolutely consistent with thermally allowed and fast reaction at room temperature. This aspect is also consistent with the catalytic activity of BP and more generally arylketones.[199–201] On the other hand the negligible difference observed in the gas phase and solution is also significant in the framework of DNA photosensitization. Indeed DNA being a complex supramolecular arrangement, the gas phase can be considered more representative of the hydrophobic core where Thy abstraction should occur; while the backbone, and notably sugar, is much more exposed to



water.

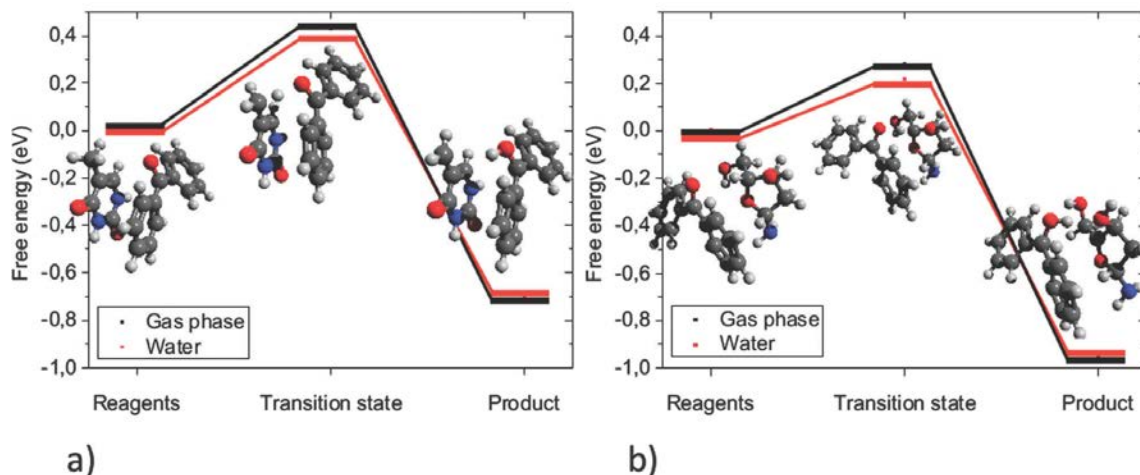


Figure 10.2: Free energy ( $\Delta G$ ) of the H-abstraction reaction from the thymine (a) and the sugar (b) by  $^3\text{BP}$ . The optimized structures for the reactants, products, and transition states are also given as insets.

as seen in Figure 10.2, the reactions proceed without significant geometrical distortion. Indeed, the reaction coordinate is easily identified as the distance between the carbonyl oxygen and the transferred hydrogen atoms for both reactions. Notably the variation of the former distance is coupled with a planarization of the carbon methyl atoms losing the hydrogen. These carbon atoms formally go from a  $\text{sp}^3$  to a  $\text{sp}^2$  hybridization.

From the results obtained for the model systems it appears that the H-abst process is a kinetically and thermodynamically favored reaction, that hence should play a major role in DNA photosensitization processes, and should be at least competitive with triplet photosensitization. However, in order to investigate reactivity in complex organized systems, even before considering the reaction free energy profiles, one should assess the existence and the probability of potentially reactive conformations. As stated previously two non-covalent interaction modes between BP and DNA have been characterized: minor groove binding (MGB) and double-insertion (DI). Two representative sketches extracted from the MD trajectories are given in Figure 10.3. Clearly, the two modes are absolutely non-equivalent in terms of further possible photochemical channels. Indeed, MGB places the carbonyl group of BP in the vicinity of the sugar hydrogens that can potentially be extracted. Conversely the Thy methyl groups are embedded in the DNA hydrophobic core and hence they are kept far away from the BP reactant. The situation is reversed in the case of DI where BP, embedded in the DNA core, may access Thy hydrogens but is precluded to persistently interact with S. Hence the interaction mode will precisely dictate the possible photochemical channels with H-abst being possible from Thy only in the

case of DI, and from S only for MGB.

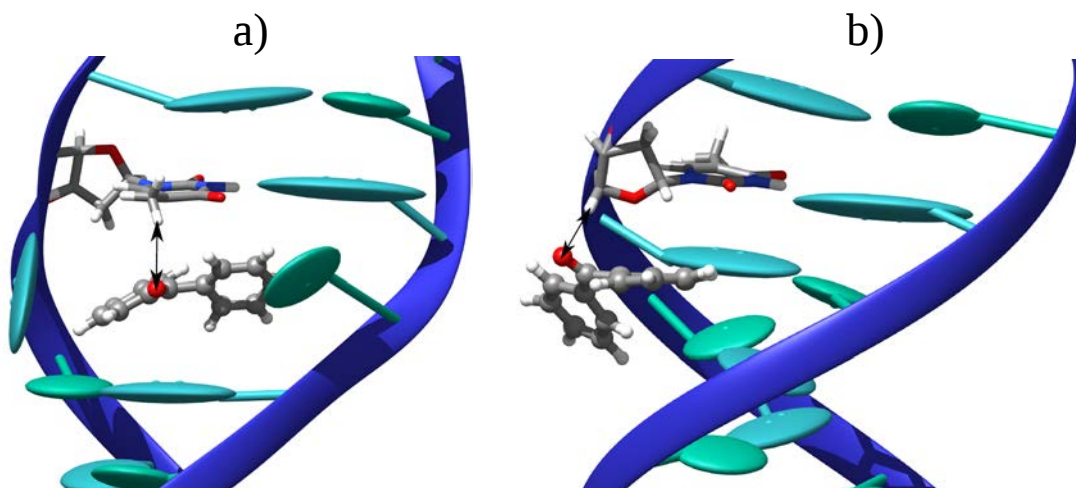


Figure 10.3: Two MD representative snapshots of potentially reactive conformations (a) double-insertion leading to interaction with methyl of thymine and (b) minor-groove binding interacting with backbone sugar.

In Figure 10.3 we have reported two conformations showing short oxygen-hydrogen approaching distances and hence possibly evolving toward reactivity. However, a more quantitative assessment of the probability of occurrence of such conformations can be found in 10.4. Here we report the distribution of the distances between the carbonyl oxygen of BP and each one of the three methyl hydrogens and the sugar hydrogen, in the cases of DI and MGB, respectively.

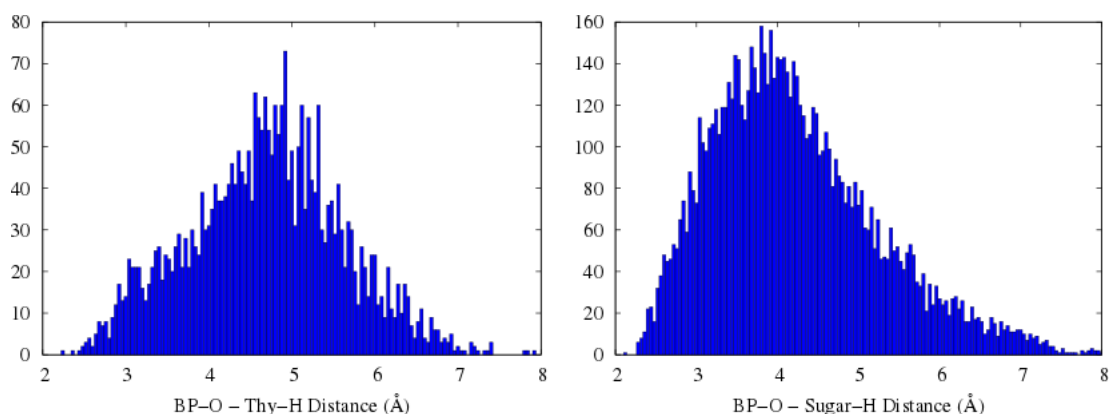


Figure 10.4: Distribution of crucial  $\text{H}\cdots\text{O}$  distances extracted from MD simulation. The left panel (double inserted BP) reports the distance between the carbonyl oxygen of BP and each of the three thymine methyl hydrogens. The right panel (minor groove binding) shows the distance between the carbonyl oxygen of BP and the C4'-hydrogen of the deoxyribose moiety.

The two distributions appear as quite dissimilar and while DI has an almost ideal gaussian shape, in the case of MGB it is much closer to a Lorentzian curve. However, both distributions present maxima at O-H distances of 4.0 Å, i.e. at

distances far too large to allow reactivity. Nonetheless, both distributions are broad and small but no negligible tails exist with distances going down to 2.0 Å, i.e. in a region that we may roughly estimate as giving rise to potentially reactive geometries. Once again the two conformations are not equivalent with MGB, leading to H-abst from S, showing a slightly larger probability of encounter than DI, leading to extraction from Thy. It has to be recalled that the former distributions are obtained from ground state MD, while H abstraction will take place in the BP triplet manifold. This can be significant especially for Thy abstraction since such  $\pi$ -stacked systems can form triplet exciplexes that will lower the distance between the two units. Globally, the low encounter probability justifies the non-prevalence of H-abst in DNA photosensitization by BP despite the very low activation barriers. Moreover, H-abst will be in competition with the very efficient and fast triplet sensitization[123]. This is particularly true for the DI situation for which all the pathways leading to triplet-sensitization happen to be almost barrierless, hence even considering the shortening of the distance between the interacting fragments due to excimer coupling H-abst appears quite unlikely in that case. On the other hand, the presence of some conformations with short O-H distances coupled with the kinetic efficiency of H-abst could suggest that the BP-DNA photochemistry may still present hydrogen abstraction reactions, in particular from the sugar moieties.

From a kinetic point of view the precise determination of the free energy profiles for the macromolecular complex and flexible systems like DNA would require the use of specifically tailored and accelerated QM/MM molecular dynamics. However the use of static QM/MM methods has allowed to enlighten some aspects related to the DNA environmental effects on the H-abst energy profiles. Also, resorting to a static QM/MM approach allows a direct comparison with previous results concerning the barrier for TTET. Indeed, in the case of MGB and H-abst from S, by using QM/MM DFT optimization of products and reactants, we noticed only a very modest decrease of the reaction energetic driving force, i.e. the energy of the products compared to the reactants, that accounts for 0.80 eV (18.4 kcal.mol<sup>-1</sup>) while the barrier is not strongly enhanced compared to the gas phase or water solution. In the case of DI, and consequently H-abst from Thy, because of the rather confined environment in which the reaction takes place, we observed the emergence of a steric clash for the products that dominate over the stabilization of the diradical complex observed for the isolated and solvated systems. Even if the destabilization obtained from our QM/MM results has to be considered just as a preliminary result, we can certainly say that H-abst in DI will imply a larger geometrical relaxation of the DNA environment, that will further diminish the probability of the reaction to occur. H-abst being more favorable from the sugar moiety than from Thy, our results would suggest that some BP-induced photochemical pathways could be related to

the production of strand breaks.

All in all, our results point toward the fact that H-abst should be considered as an important mechanism in the gas phase or water. Nevertheless, coherent with the experimental observation its occurrence in DNA is strongly limited because of the reduced probability of reactant encounters, competition with triplet sensitization, and probably geometrical constraints imposed by the DNA structure. Indeed, although DNA is flexible and can accommodate rather large structural changes this would require a non-negligible free-energy contribution.

## Conclusion

We have designed a multiscale model of an ubiquitous and biologically important chemical reaction, namely hydrogen abstraction, performed by a common photosensitizer (benzophenone) on DNA constituents, thymine and backbone sugar. Quantum chemical calculations performed on model isolated systems, as well as in implicitly described water solvent, unambiguously showed that the H-abst reaction is thermodynamically favored, while no significant kinetic energy barrier reduces its probability. Those results have been confirmed and validated at both DFT and multiconfigurational CAS-(RAS-)PT2 levels of theory. However, the presence of the DNA macroenvironment, as evidenced by the analysis of MD trajectories, significantly alters the reaction occurrence probability. Indeed, the DNA organized environment acts, in a certain sense, like a filter largely diminishing the encounter probability of the reactants even when BP is forming stable non-covalent interacting aggregates with DNA, as evidenced by the low occurrence of short O-H distances all along the MD. This effect is also probably enhanced since H-abst is in competition with triplet-triplet energy transfer that has been proved to easily proceed via efficient photophysical channels. Furthermore, a preliminary QM/MM investigation of the reaction happening in the complex environment seems to point toward the presence of steric constraints imposed by DNA for the DI mode. All in all, both QM and MD simulation results seem to slightly favor H-abst from the sugar in MGB interaction mode. This fact is also quite remarkable since triplet photosensitization is much more favored when BP is in DI than in MGB. Our results have clearly rationalized a complex photochemical pathway related to DNA photosensitization, in particular explaining the smaller role played by this mechanism in the global sensitization.

## 10.2 Nile Blue, Nile Red: Electron transfer

### Introduction

As stated in 1.3, photodynamic therapy is predominantly performed through type II photosensitization, that is the production of reactive singlet oxygen using porphyrin like molecules. Even though the development of such therapy has been thorough over the years and is now efficiently preformed in a daily basis, its limiting aspects are problematic for many specific cases. Indeed, for solid tumors in which the diffusion of molecular oxygen is rather low, making type II photosensitization inapplicable, photosensitizers with other photophysical pathways must be developed. On one side, a strong absorption in the near-infrared and infrared[202] (the therapeutic transparent window) is necessary to trigger the photosensitizer in such environment, on the other side, direct photosensitization through electron transfer (type I photosensitization[3]) that does not necessarily requires the presence of oxygen and will be effective even in hypoxia conditions[203] has to be considered.

We determined the optical properties of two fluorescent dyes, nile blue (NB) and nile red (NR) Figure 10.5, which exhibit interesting photophysical properties[180]. Indeed, both are already employed in the field of cellular biology[204],[205]. Using TD-DFT we evidenced the necessity to include dynamic and vibrational effects in order to reproduce the experimental UV spectra of these compounds.

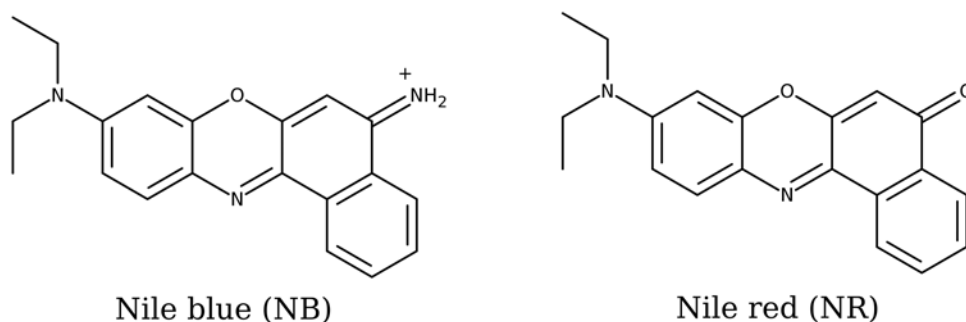


Figure 10.5: Nile blue and nile red chemical structure.

A photosensitizing effect of NB on DNA has been evidenced by Hirakawa et al.[206]. They hypothesized a type I like mechanism where NB interacting with DNA would be able to trigger an electron transfer from a nearby guanine, followed by an electronic reorganization that brings NB back to its fundamental state. Then the radical cationic guanidinium has the ability to evolve toward a damaged nucleobase.

As a preliminary step we want to understand the possible binding conformations

of NB and NR to a canonical B-DNA double strand. NB being a cationic species and NR neutral, the first would be eager to interact in a Minor Groove Binding (MinGB) fashion (in which the molecule slips inside the negatively charged moiety of the minor groove of DNA). Then both being rather small and flat conjugated species, they might both be DNA intercalators, forming a complex stabilized through  $\pi$ -stacking interactions. For each stable aggregate, we evaluated the possible charge transfer from a close guanine to the sensitizer. Indeed we built *in silico* a heteropolymer of cytosine-guanine sequence bound to either NB and NR both in MinGB and intercalation. It is in fact known that guanine has the highest probability to be involved in electron transfers, thus promoting type I photosensitization[207],[208]. From this, we ought to evaluate the photosensitizing properties of NB and NR in their interacting configurations. Also, since the environment proved to have an important effect on the optical properties of both dyes, we took into account the molecular environment, either hydrophobic in intercalation or surrounded by charges in the minor groove. To do so we employed hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) techniques, with a QM description of the photosensitizer and the closest guanine.

We validate our hypothesis on the capacity of guanine to be of a major role in this photosensitization process by comparing the experimental optical properties of the two heteropolymers guanine-cytosine and adenine-thymine in interaction with either NB or NR.

## Computational details

All simulations were performed on a TIP3P[168] water solvated poly(dG-dC) hexadecamer, including counterions. The stability of each DNA-photosensitizer complex was validated after a pre-equilibration step and a final 100 ns MD simulation production. The amber99 force field parameters[167] with bsc0 correction[133] were applied to the DNA strand, while GAFF parameters[17] coupled to RESP fitted atomic charges were considered for NB and NR. The MD simulations were carried out using the GPU CUDA version of amber15[132], combined with Curves+ as a tool to analyze DNA structural distortions caused by the sensitizers[79].

Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations were performed to correctly model electronic excited states. The QM region comprises the sensitizer (NB or NR) and the nearest guanine nucleobase. The time dependent-density functional theory was used in the framework of a TD-DFT/Tinker protocol for excited state optimization, based on electrostatic embedding and hydrogen link atom schemes. More in detail, the hybrid meta exchange-correlation M06-2X functional[209] was selected, with a 6-31+G(d) basis set, as it can properly describe



charge transfer mechanisms. Moreover, it is in agreement with ab-initio multiconfigurational CASPT2 calculations on NB and NR[180]. All QM/MM calculations were performed with a locally modified[111],[134] version of the Gaussian09 code [136], coupled to Tinker [137].

## Results and discussion

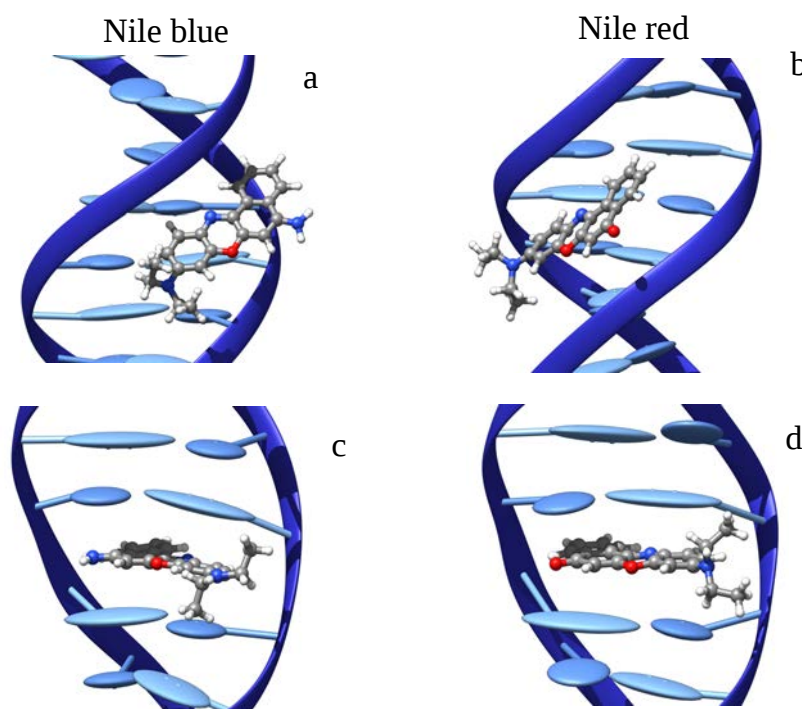


Figure 10.6: DNA binding interaction modes: selected MD snapshots of NB in minor groove binding (a) and intercalated (c). NR in minor groove binding (b) and intercalated (d).

The first step in assessing the potentially different photochemistry of DNA sensitizers is to unravel its or their binding modes with the biological system. From the MD simulation it turned out that both NR and NB present two stable interaction modes with the DNA double strand (Figure 10.6). Indeed, in both cases one can observe the presence of minor groove binding and intercalation. The positive charge held by NB can certainly justify minor groove binding due to electrostatic interactions with the charged phosphate groups in the backbone. In the case of NR minor groove binding is mostly driven by a combination of dispersion and labile hydrogen bonds interactions. Intercalation, on the contrary, is largely driven by dispersion interactions and the  $\pi$ -stacking between sensitizers and the nucleobases in the intercalation pocket. Even if costly calculations of binding free energies would be necessary to ultimately assess the relative stability of the modes (see 9), one can presume that NB would probably favor minor groove binding instead of intercala-

tion. In this latter case NB's positive charge will be surrounded by the hydrophobic environment of the DNA core; for the opposite reasons we could probably state that NR will be more prone towards intercalation.

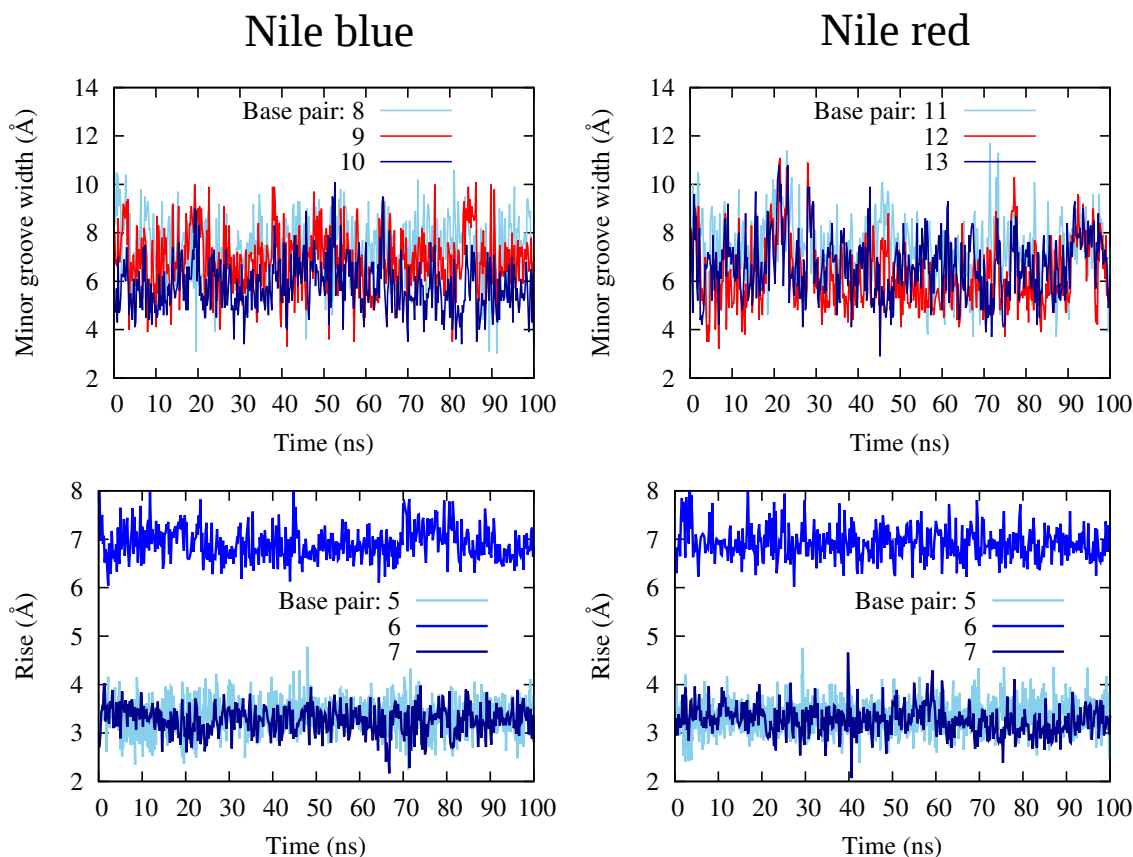


Figure 10.7: DNA structural analysis: time series for the minor groove width for NB and NR in minor groove binding and rise parameter for NB and NR in intercalation. The chosen base pairs are the ones interested by the interaction with the sensitizers together with one non-interacting pair as reference.

The structural deformations induced on DNA by the interactions with the sensitizers are quite standard for those kind of DNA non-covalent binding modes and almost indistinguishable between the two sensitizers.

Almost no peculiar deformation in the minor groove parameters can be evidenced for the minor groove binding (Figure 10.7 top), while in the case of intercalation the distance between the bases (rise parameter in Figure 10.7) is doubled for the bases in between which NB or NR slips. This feature is quite common in the case of intercalation and allows the creation of the intercalation pocket able to accommodate the sensitizer.



**Absorption spectra:**

The calculated absorption spectra for NB and NR in interaction with DNA (minor groove binding and intercalation mode) are reported in Figure 10.8. The main optical characteristic of the two dyes are quite well reproduced, in particular the low energy intense peak corresponding to the  $S_0 \rightarrow S_1$  transition that will dominate the subsequent photophysics.

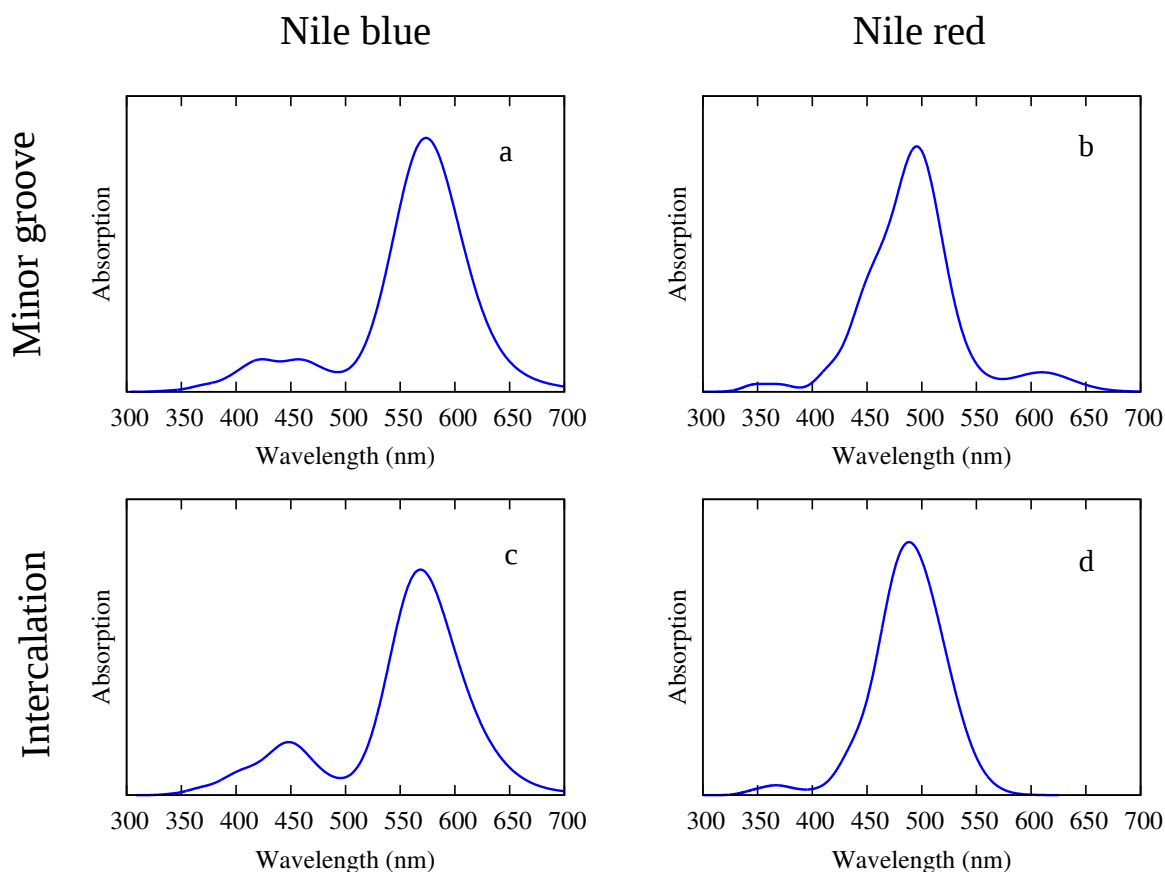


Figure 10.8: Simulated absorption spectra of NB and NR interacting with B-DNA. NB in minor groove binding (a) and intercalated (c). NR in minor groove binding (b) and intercalated (d).

It is also noteworthy to see that the interaction with DNA has only a very slight effect on the position of the absorption peaks. Indeed, both interaction modes give, for the same sensitizers, almost the same excitation energy that also happens to be quite close to the one obtained in aqueous environment[180], indicating a quantitative agreement between implicit and explicit water solvent models.

Finally, and to prove the interaction with DNA, we report the absorption spectra for NB and NR obtained with different concentrations of DNA (Figure 10.9). The change of the intensity of the NB (NR) characteristic band with the increasing concentration of DNA is clearly indicative of the development of non-covalent

interactions and hence of the fact that both dyes are able to give rise to stable non-covalent aggregates with DNA. On the other hand, the negligible change in the position of the maximum also confirms our theoretical results. The consistent red-shift compared to our theoretical results can be adequately recovered by introducing ab-initio electronic dynamic correlation, indeed CASPT2 shifts the absorption maximum of NR and NB of 0.3 and 0.2 eV respectively[180]. Interestingly, it is also evident that NB absorption maximum falls in the phototherapeutic window (620-850 nm), while both NB and NR long-wavelength tails significantly overlap with the phototherapeutic window.

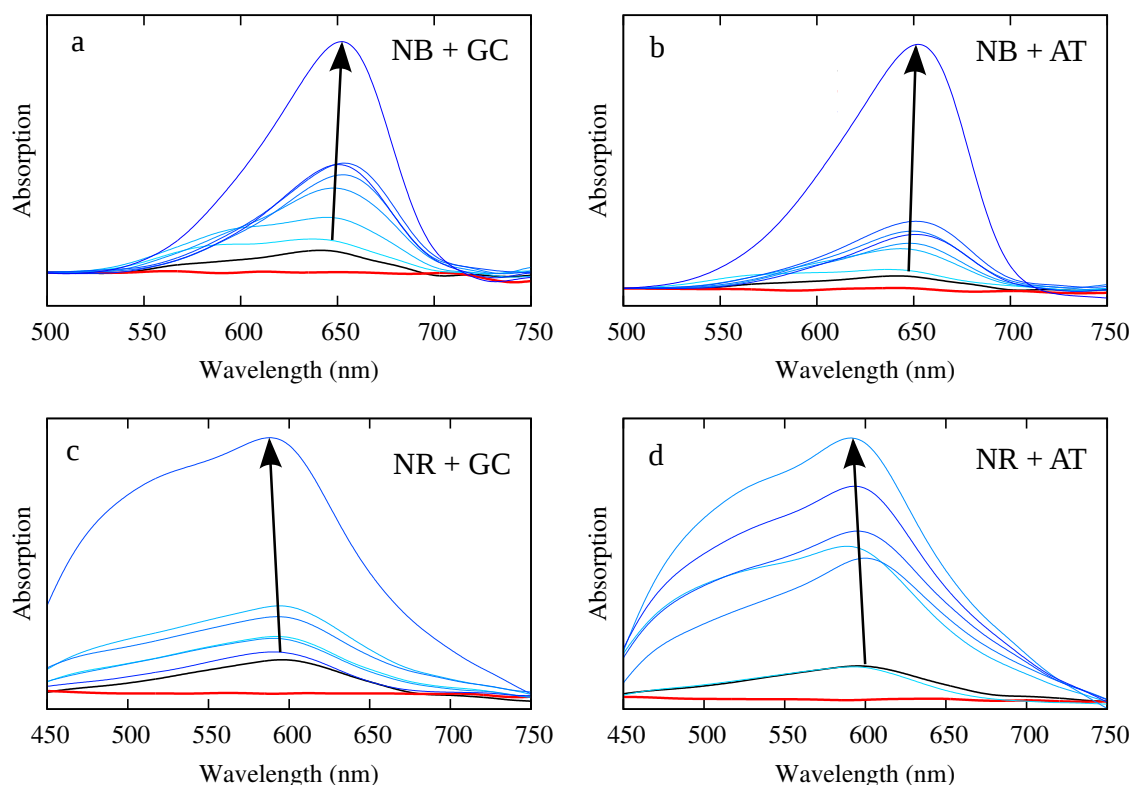


Figure 10.9: Experimental absorption spectra of NB (a,b) and NR (c,d) interacting with poly(dA-dT) and poly(dG-dC) DNA strands. Black and red curves represent the absorption of solvated sensitizer (NB and NR) and DNA, respectively; blue curves represent different DNA concentrations: from the lowest, represented by light blue curves, to the highest shown in dark blue (see Methods). The sensitizer concentration is kept constant.

### Charge transfer mechanisms:

In order to assess the photoinduced charge transfer mechanism by NB and NR we have performed simulations at the QM/MM level, including in the QM partition the sensitizer and one nearby guanine. Subsequently the geometry of the lowest excited states has been optimized in order to obtain the corresponding Jablonski diagram and hence to infer the photochemical pathways (Figure 10.10). In particular, and

already at the Franck-Condon region, in addition to the local transitions we evidence the presence of a long-range charge transfer state. Its analysis, in terms of natural transition orbitals (NTO) and  $\Phi_S$  index [210] (see Figure F.3 and Figure F.4 in appendix for details), confirms that the state correspond to an electron transfer from the guanine to the sensitizer.

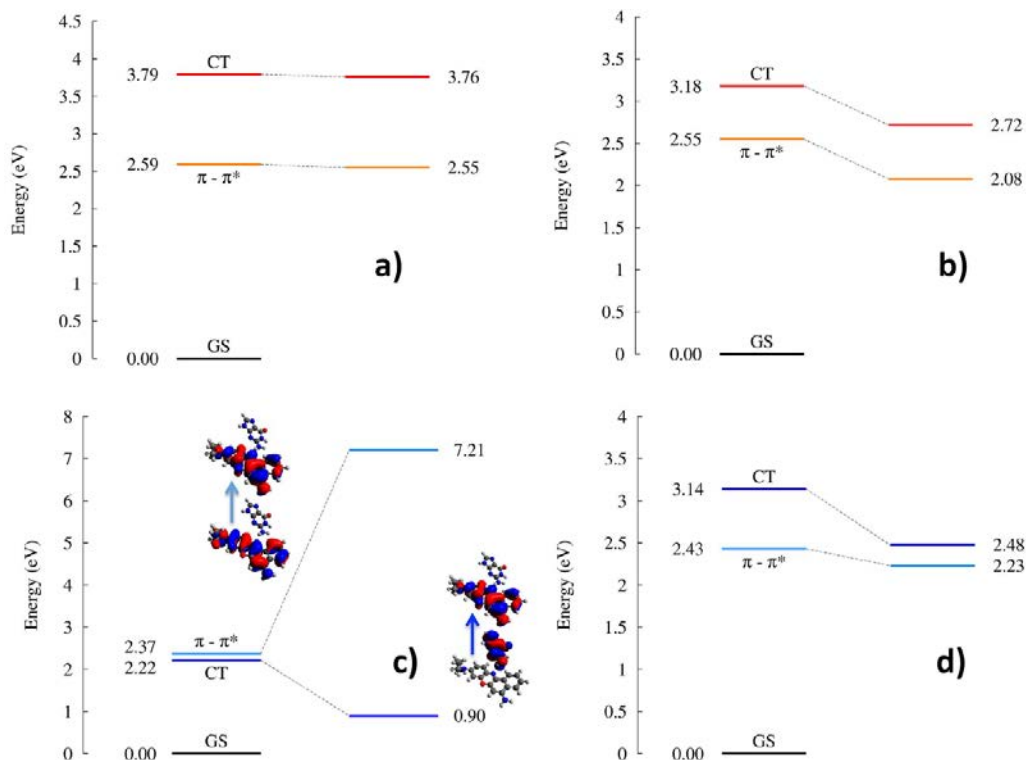


Figure 10.10: Jablonski (energy) diagram, showing the evolution of locally excited ( $\pi - \pi^*$ ) and charge transfer states (CT), from the Franck-Condon region to the respective excited state minima. (a) NR in minor groove binding, (b) NR in intercalation, (c) NB in minor groove, (d) NB in intercalation. Energies for the optimized structures are reported with reference to the energy of the ground state at the lowest excited state minimum. NTOs for NB in minor groove are given in panel (c) (Figures F.3, F.4).

In the case of NR the charge transfer state is quite high in energy: at Franck-Condon the charge transfer state is 0.65 eV (intercalation) and 0.86 eV (minor groove binding) higher than the lowest-energy local excited state. Even upon optimization the charge transfer state always lays higher in energy than the local  $\pi - \pi^*$  state and their relative energy only changes negligibly for both interaction modes. As a consequence we may conclude that the sensitization through electron transfer is precluded for NR, since the population of the charge transfer state is most unlikely to happen for energetic reasons.

On the contrary NB shows a much interesting pattern. Indeed, in the case of intercalation the charge transfer state at Frank-Condon is higher in energy (0.71 eV) and upon optimization it does not cross the lowest-energy local excited state, how-

ever the energy difference is strongly reduced and the states become quasi-degenerate laying only 0.25 eV far apart. Even more appealing, the lowest-energy singlet excited state for minor groove binding is the charge-transfer, lying 0.15 eV lower than the local state at Frank-Condon. This energetic order is preserved, and even amplified, upon optimization of both local and charge-transfer states. The local  $\pi$ - $\pi^*$  excited state is now destabilized while the charge transfer one is strongly stabilized ultimately leading to an energy difference of about 6.0 eV. This situation could seem quite unlikely since one could expect that the  $\pi$ -stacking between NB and guanine, characteristic of intercalation, should favor the coupling of the two chromophores and hence the charge transfer. However, it has to be recalled that in the case of intercalation charge separation should take place in a strongly hydrophobic environment, hence being less favorable than the one happening in the hydrophilic and charged minor groove region.

Our results clearly point toward the fact that charge transfer from guanine to NB is feasible, but is environmentally controlled: it can take place only in the case of minor groove binding. Since the local  $\pi$ - $\pi^*$  state is always much optically brighter than the charge transfer, one may expect the photophysics to proceed in the following way: light absorption will induce the transition to the local  $\pi$ - $\pi^*$  state ( $S_2$  for NB in minor groove binding), subsequently the system will go through an internal conversion that is expected to be fast due to the low energy difference and the short distance between the chromophores, leading to the population of the charge transfer state ( $S_1$ ). This will hence promote the charge separation with the formal production of a guanidinium cation that will subsequently evolve to induce further DNA lesions.

### Fluorescence spectra:

The presence of electron transfer in the case of NB is also confirmed by performing experimental fluorescence spectra in presence of poly(dA-dT) or poly(dG-dC) DNA duplexes. Indeed, since electron transfer should be possible only from guanine, one can expect a strong fluorescence quenching in the case of guanine rich strands, while the poly(dA-dT) strand should have a differential effect. This can be seen in Figure 10.11a): it is evident that the fluorescence of NB in presence of poly(dG-dC) is almost entirely quenched, hence confirming the presence of an accessible charge transfer state that strongly limits the life-time of the potentially emissive local  $\pi$ - $\pi^*$  state.

However, as also reported in Figure F.5, the fluorescence behavior of NB interacting with poly(dA-dT) is much more complex. Indeed after a quenching happening at small concentrations one witnesses a sharp enhancement of the emission intensity and hence of the quantum yield, that becomes even higher than the one of solvated NB alone. This fact is usually interpreted with the presence of two interaction modes

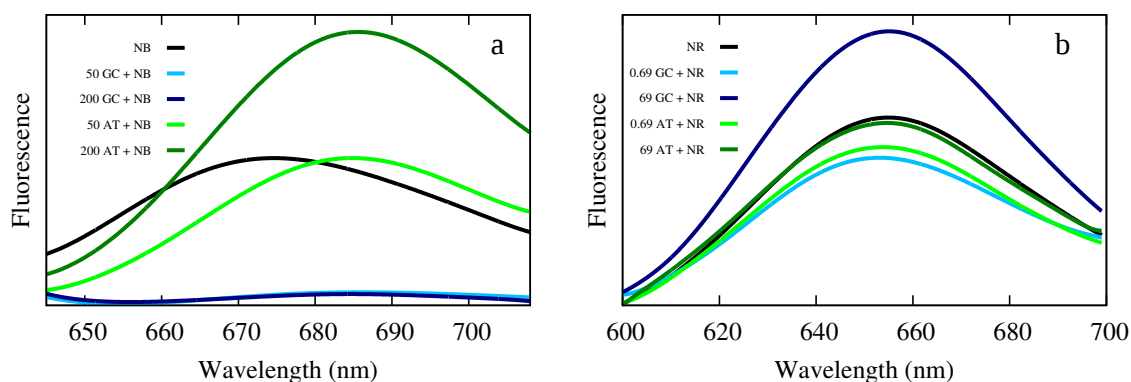


Figure 10.11: NB (a) and NR (b) experimental fluorescence spectra. Black curve refers to the fluorescence of the solvated NB ( $10\ \mu\text{M}$ ) or NR ( $690\ \mu\text{M}$ ). Green curves and blue curves refers to the fluorescence of the sensitizer in presence of poly(dG-dC) or poly(dA-dT) double strand, respectively. DNA concentrations are given in ng/L and the concentration of the sensitizer is kept constant.

[211]. The quenching at low concentration is usually seen as the mark of electrostatic (groove binding) interaction modes that at higher concentration are in competition with intercalation. This aspect is also confirmed by Figure 10.12 showing the MD simulation of a solution of DNA double strand in which an excess of NB is present. The spontaneous formation of stable aggregates and clusters of NB in the minor groove after 50 ns is clearly evidenced by a representative snapshot of the 100 ns dynamics.

Those  $\pi$ -stacked structures may be regarded as the ones responsible for the initial quenching of the fluorescence, in accordance with the observed loss of emission intensity in the case of stacked organic fluorophores. Intercalation that was not spontaneously observed probably due to a higher kinetic barrier could, most probably, become competitive at higher concentration and be responsible of the subsequent fluorescence enhancement. On the other hand, as reported in Figure 10.11b, NR fluorescence is not quenched by the poly(dG-dC) strand, and both DNA sequences show the characteristic fluorescence enhancement at high concentrations. Globally it is evident that the fluorescence measurement strongly confirms our prevision of the presence of two competitive binding modes, as well as the photoinduction of electron transfer through guanine by NB correlated to the almost total fluorescence quenching.

## Discussion/Conclusion

With a combined use of molecular modeling and experimental spectroscopy we have unraveled the different interactions modes, and photochemical pathways of two organic dyes (NB and NR) interacting with B-DNA. In particular we have evidenced

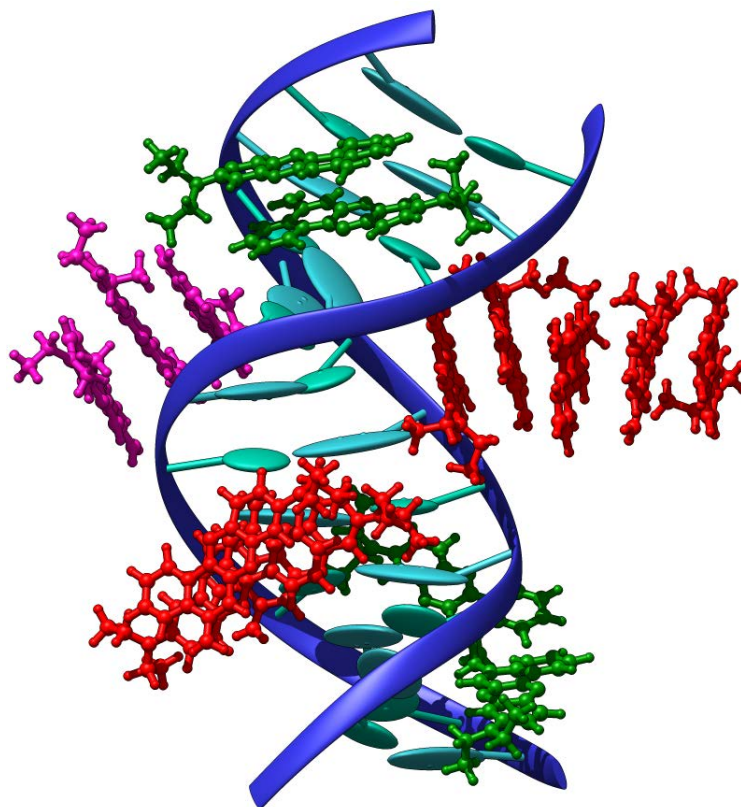


Figure 10.12: Representative snapshot of a DNA double strand in presence of an excess concentration of NB. Note the spontaneous formation of minor-groove binding as well as of  $\pi$ -stacked aggregates. The different interaction conformations are represented with different color of NB.

the presence for both drugs of two competitive interaction modes, namely intercalation and groove binding. The subsequent QM/MM exploration of the different excited states has shown that while NR is unable to induce any sensitization of the DNA through electron transfer, NB may indeed give rise to a net transfer of an electron from DNA to the dye. However, the sensitization properties of NB are strongly dependent on its molecular environment and in particular on its binding mode. Indeed, while in intercalation, probably because of the embedding in a hydrophobic environment, charge separation is unfavorable. On the contrary, a highly feasible and energetically favored pathway for electron transfer is evidenced in the case of groove binding. This suggests that NB, also due to its absorption in the red part of the spectrum, could be used as a sensitizer agent in hypoxia phototherapy to induce DNA lesions via an electron transfer mechanism (Type I).

Absorption and fluorescence spectra have confirmed the presence of two competitive interaction modes. Furthermore, the almost complete quenching of fluorescence by guanine rich sequences, compared to the more complex behavior in the case of poly(dA-dT), is strongly indicative of a charge transfer sensitization. In the following we plan to completely characterize the time-evolution of the different involved

excited state by using non-adiabatic molecular dynamics, and hence have access to characteristic times connected with the sensitization process. Unraveling at electronic and molecular level the mechanism hidden behind sensitization, and also its dependence on the environment, will ultimately greatly help the rational design of novel (photo)-therapeutic agents interacting with DNA.

## 10.3 BMEMC: Radical species

### Introduction

Finding new photosensitizers that would work in the therapeutic window would greatly help the treatment of either deep tumors or solid ones. In that aim, developing compounds that has both the ability to be stable in interaction with DNA and and possess a high Two Photon Absorption (TPA) activation that would allow the use of IR light in photodynamic therapy. Furthermore, compared to the usual type II photosensitization using porphyrins and mainly based on the production of reactive singlet oxygen  $^1\text{O}_2$ , it would be a great step forward to develop TPA photosensitizers that would work in hypoxia environment since their mode of action would not require molecular oxygen activation. In fact a recently synthesized strong TPA absorber[3] with a A- $\pi$ -D- $\pi$ -A structure where the donor unit is a carbazole chemical function, showed the ability to photosensitize DNA. Indeed, the dicationic 3,6-bis[2-(1-methylpyridinium)-]9-methylcarbazole (BMEMC) (Figure 10.13) has evidenced strand breaks formations upon infrared activations even in hypoxia conditions.

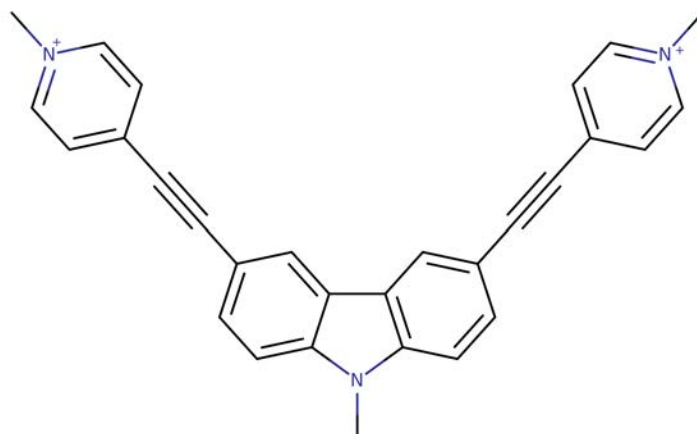


Figure 10.13: Molecular structure of the dicationic BMEMC DNA sensitizer.

Experimentally, neither the interaction with DNA and the photophysical electronic mechanisms leading to the sensitization have been described. Moreover, since it has been demonstrated that the production of radical species can induce the formation of strand breaks in DNA[212–214], one should consider the production of



solvated electrons and radical ions in the mechanism. On one side, we investigated the possible binding modes of BMEMC in DNA through MD simulations giving insights on the most probable sensitizing pathways. Then we realized QM investigation of the photophysical properties of BMEMC, mainly One-Photon Absorption (OPA) and TPA efficiencies, depending on its environment, first in a water solution or bound to DNA.

## Computational details

All along the study we resorted to two different protocols to access the optical and photophysical properties of the BMEMC dye, namely static and dynamic approaches [123, 129, 210].

In the case of the static approach, and for BMEMC in solution only, we optimize at density functional theory (DFT) level the geometry of the ground state; environmental effects are modeled using the polarizable continuum model (PCM)[192]. From the equilibrium geometry the absorption spectrum is obtained within the time-dependent DFT (TD-DFT) and the response formalism as vertical transition. Geometry optimization was performed using the Gaussian 09 D.01 code [136], B3LYP and the Pople double- $\zeta$  6-31G(d,p) basis set. OPA was obtained at the TD-DFT/CAM-B3LYP/6-31+G(d) level using Gaussian 09. Note that the results have been checked against vertical transitions obtained using CAM-B3LYP and 6-31++G(d,p) basis sets; with the differences being less than 0.1 eV the smaller basis set has been kept for the following study. The TPA cross-section was obtained using the DFT quadratic response formalism [215] as implemented in the Dalton 2016 code [216]. In order to mimic the experimental band shape vertical transitions have been convoluted using a Gaussian function of a full-width at half-length (FWHL) of 0.2 eV. Finally, the Jablonski diagram was obtained by calculating the energy of BMEMC in its ground state ( $S_0$ ) and first excited state ( $S_1$ ), and of the ionized species having lost an electron, at  $S_0$  and  $S_1$  equilibrium geometry respectively. In order to take into account the solvated electron effects the energy of the cation has been scaled by the value of the vertical electron affinity of the solvated electron. The  $S_1$  equilibrium geometry has been obtained by TD-DFT optimization using CAM-B3LYP and 6-31G(d,p) basis sets.

In the case of the dynamic approach the conformational space of the chromophore is preliminarily explored using MD, subsequently independent snapshots are selected and for each one of them the excited state energy and oscillator strength (cross-section) are computed as vertical transitions, with the environment being treated at the QM/MM level. The final spectrum is obtained by convoluting the vertical transitions of each snapshot. 100 snapshots are used for OPA and 20 for TPA and



a convolution with the Gaussian function of a FWHM of 0.1 eV was applied. The dynamic approach, differently from the static one, is also able to retrieve the optical properties of non-homogeneous systems such as BMEMC interacting with DNA. The nature of the excited state has been evidenced using the natural transition orbital (NTO) formalisms [217–219]. NTOs have been obtained using a locally produced code (NancyEX) [218, 219].

All MD simulations, in a water box and in interaction with a 14 base-pair long polydA-polydT B-DNA, were carried out using the GPU cuda version of amber15[132]. DNA-BMEMC complexes were placed in a  $67 \text{ \AA}^3$  side length truncated octahedron, containing TIP3P67 water molecules. Periodic boundary conditions were applied to simulate the bulk phase and a cutoff was applied to turn off interactions between atoms more than  $9.0 \text{ \AA}$  apart. In order to neutralize the system, potassium  $\text{K}^+$  cations were placed in the simulation box. A 13 base-pair oligonucleotide was modeled with the amber99 force field parameters[167] including the bsc0 correction[133] developed to specifically reproduce backbone dihedrals of nucleic systems. The BMEMC potential was represented via the GAFF parameters[17], used for bonded interactions and the atomic charges were calculated using the RESP71 fitting protocol and Hartree-Fock quantum calculations with the 6-31G(d) basis set commonly used for organic compounds. Preliminary MM energy minimization of 8000 steps was carried out, the first 4000 steps using the steepest descent algorithm followed by 4000 steps using the conjugate gradient algorithm. MD started with 200 ps heating to bring the system to 300 K in the NVT ensemble, followed by a 400 ps equilibration MD in the NPT ensemble at 300 K and 1 atm to obtain the correct density. Finally a 100 ns production MD simulation was performed in order to validate the stability of the binding modes and to extract the relevant radial distribution functions. Along all MD simulations the shake algorithm was applied to bonds involving hydrogens. Moreover Langevin dynamics were used to maintain the temperature and during the production run the pressure was kept constant using a Monte Carlo thermostat. The MD simulation protocol was also validated analyzing the behavior of the double-stranded DNA in the absence of BMEMC. Although minimal salt concentration and a short oligonucleotide were used all along the trajectories we evidenced only negligible deviations from B-DNA ideal structural parameters compatible with thermal fluctuation deviations. To model the OPA spectrum, QM/MM simulations were performed using a local modified[134] version of the Gaussian 09 code[136] including mechanical, electrostatic and polarizable embedding[111]. It has been shown that polarizable effects are marginal for this system. For the QM/MM calculations, and coherently with the MD, we used the Amber ff99 force field to model the environment, while DFT using the CAM-B3LYP functional and the 6-31G(d) basis set was chosen for the QM part

(BMEMC cation). TPA spectra have been obtained using the same conditions as for OPA with the Dalton2016 code[216] and in particular the polarizable embedding module to treat the environmental effects at the QM/MM level[220].

## Results and discussion

### Water solution

The experimental absorption spectrum of BMEMC is characterized by an intense and relatively large band at around 420 nm, furthermore a shoulder can be observed at around 350-380 nm[3]. The intense band is normally associated with the  $S_0 \rightarrow S_1$  transition that is optically bright. The TD-DFT calculated absorption spectrum is shown in Figure 10.14. In particular in Figure 10.14 a) we report the spectrum calculated statically as vertical transition from the Franck-Condon geometry, while the solvent effects are taken into account using the PCM method. One can see that although general agreement with experimental results still holds, the absorption maximum appears to be slightly blue-shifted. This aspect can be associated with the absence of vibrational and dynamic effects inherent in the static approach, as already evidenced for other large  $\pi$ -conjugated aromatic systems and in particular carbazole moieties[129]. Indeed, if we calculate the absorption spectrum convoluting vertical transitions from geometries extracted from a MD simulation of BMEMC in a water box the previously observed blue-shift is corrected and now the calculated results remarkably match the experimentally measured spectrum.

The reasons for the influence of dynamical and vibrational effects can be related to the presence of large amplitude, low frequency, vibrations corresponding to the out of plane bending of the carbazole unit, as confirmed by the calculation of the ground state harmonic vibrational normal modes. As is apparent from the analysis of the NTOs in Figure 10.15c the first excited state is of  $\pi$ - $\pi^*$  nature and delocalized over the carbazole and the bridge unit. However, while the occupied NTO (hole density) shows a bonding interaction in the fused carbazole three rings, the virtual NTO exhibits an antibonding pattern. As a consequence out-of-plane vibrations will destabilize more the ground state compared to the excited state, hence producing a global red-shift. It is noticeable that although  $S_1$  is globally local a very slight charge-transfer can be observed with a small density accumulation toward the methylpyridinium edges. On the other hand, the charge transfer characteristic is much more evident in the case of the  $S_2$  state with a larger electron displacement from the carbazole to the acceptor edges. The larger charge transfer nature of the  $S_2$  state can also explain the lower oscillator strength compared to the brighter  $S_1$ . At the same time the marginal participation of the carbazole moiety in the excited state electronic density distribution also explains why the  $S_0 \rightarrow S_2$  transition is

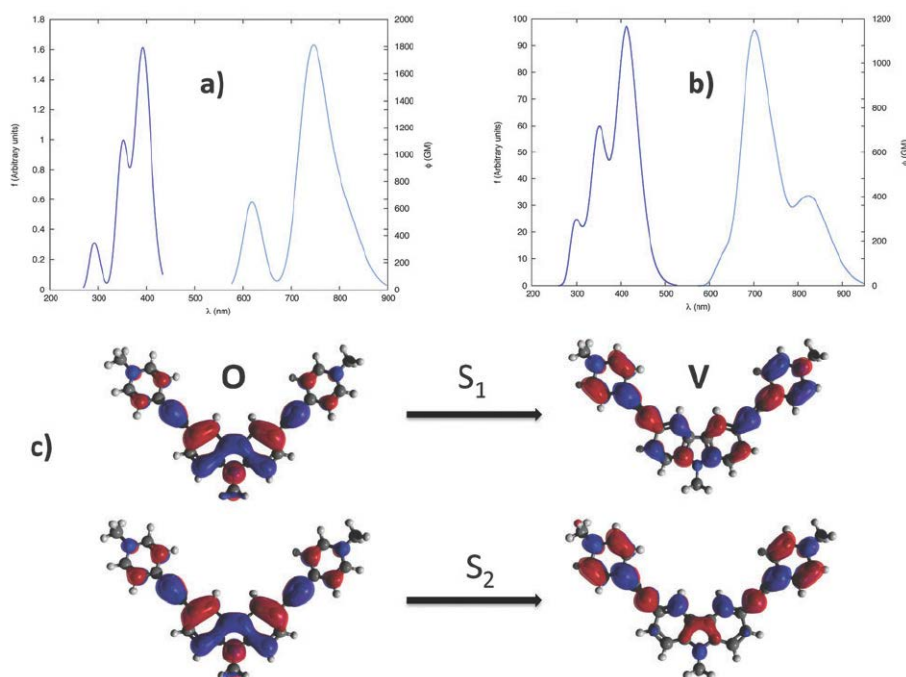


Figure 10.14: Optical properties and excited state characterization of BMEMC in aqueous solution. (a) OPA (dark blue) and TPA (light blue) calculated from the static approach with the environment modeled by PCM (b) OPA and TPA calculated at the QM/MM level as convolution of the MD snapshots. (c) The occupied (O) and virtual (V) NTOs describing the first two excited states.

much less affected by vibrational effects than the  $S_0 \rightarrow S_1$  one.

The TPA absorption cross section as a function of the wavelength is also reported in Figure 10.14 a) and b) for the static and dynamic approaches, respectively. Once again coherently with experimental results[3], we may observe a quite intense maximum at around 700 nm, with the TPA cross section exceeding 1000 GM. While those results confirm that BMEMC is an efficient TPA absorber in the near infrared region some aspects deserve to be properly commented on. First of all, and as compared to OPA, now the  $S_2$  state is much brighter than  $S_1$  almost by a factor three (Figure 10.14b). This fact can be rationalized by the larger participation of the bridge and of the pyrimidinium edges in the  $S_2$  state since the latter, due to extended conjugation, are characterized by a larger hyperpolarizability. As a matter of fact, since  $S_2$  is dominating the TPA spectra the influence of dynamic effects is now different compared to OPA. Indeed, since in the static approach the  $S_0 \rightarrow S_1$  transition is blue shifted while  $S_0 \rightarrow S_2$  is almost unaltered, in the static spectrum one can observe a large band with only a shoulder around 800 nm. Also, the effect of convolution and the partial overlap between  $S_1$  and  $S_2$  bands induces an artificial very slight red-shift of the TPA maximum. On the contrary in the dynamic approach, the  $S_1$  artificial blue shift having been corrected, the two peaks are now much more separated and as a result one can now distinctively observe the peak due

to the first excited state at around 810 nm, while the TPA maximum due to  $S_2$  lies at 700 nm. Remarkably enough, in the case of the dynamic approach we obtain a TPA cross section of around 370 and 500 GM at 800 and 760 nm, respectively, to be compared with the reported experimental values of 352/492 GM[3].

In order to elucidate the proposed photophysical channel, and in particular the production of the BMEMC radical cation and solvated electrons[3] we report in Figure 10.15 the Jablonski diagram, calculated both at the Franck-Condon and  $S_1$  equilibrium geometries. Note that the calculated energy of the radical cation has been scaled by the well known energy of the solvated electron in the bulk, whose electron affinity is estimated at 3.3-3.6 eV[214]. At Franck-Condon the ionized systems lies around 0.5 eV lower than the  $S_1$  state, and this energetic order is preserved also at the  $S_1$  equilibrium geometry, i.e. under the most favorable conditions to the non-ionized state, although the gap is diminished (around 0.2 eV). As a consequence, we can clearly estimate that after the population of  $S_1$  through direct absorption in the case of SPA, or via internal conversion from  $S_2$  in the case of TPA, the system will rapidly and efficiently undergo ionization, most probably barrierless, to produce a solvated electron and a radical cation. Hence, two species potentially able to sensitize DNA and produce irreversible lesions will be obtained upon irradiation. The electronic structure of the radical cation can also be inferred from the spin density reported in Figure 10.15 b). We may observe that the unpaired electron appears to be widely delocalized over the carbazole and the bridge region, although a large density is centered on the carbazole nitrogen atom. This could imply the presence of a reactive, yet sufficiently long-lived species to reach the DNA backbone and induce strand-breaks.

### Interaction with DNA

In order to assess the photophysics and photochemistry of BMEMC in the presence of DNA it is of fundamental importance to unravel the stable interaction modes. Indeed, from an experimental point of view little is known about the effective mode of interactions. Intercalation is invoked on the basis of mostly indirect and speculative observations for BMEMC[3] or similar carbazole sensitizers[221–223]. The analysis of our MD trajectories has allowed us to identify four main largely stable interaction modes as reported in Figure 10.16. In particular we may observe the presence of a rather classical intercalation mode with the carbazole moiety perfectly slipping between two base pairs and the positive charged pyridinium edges pointing toward the bulk and the negative phosphates (Figure 10.16 a). The former mode is similar to the one observed for similar systems, namely  $\beta$ -carboline harmaline[129]. A second intercalative pattern has also been observed of which a representative conformation

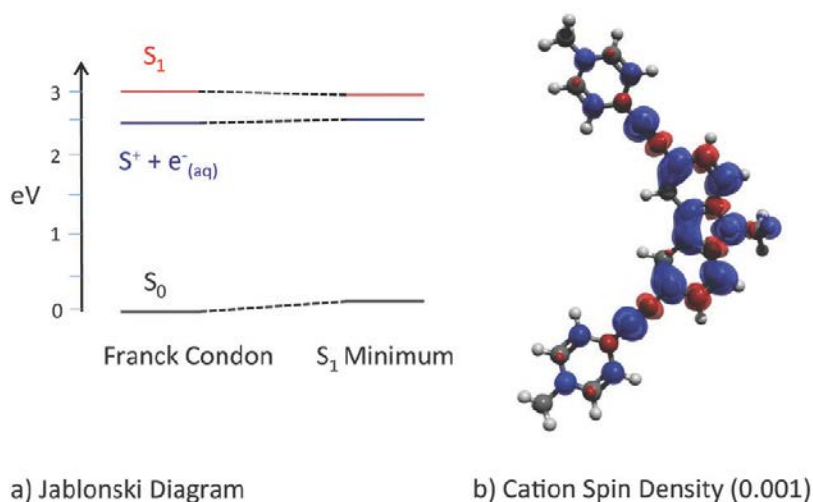


Figure 10.15: Photophysical properties of BMEMC in water solution. (a) Jablonski diagram representing the energy of the ground ( $S_0$ ), first excited state ( $S_1$ ), and the radical cation and the solvated electron ( $S^+ + e^-_{(aq)}$ ). (b) Spin density isosurface for the BMEMC radical cation.

is reported in Figure 10.16 b). In this case the participation of the pyridinium edges is more pronounced with the two bridges pointing toward the minor and the major groove respectively. Although both conformations appear to be stable during the whole MD trajectory it has to be pointed out that the formation of the latter is most probably associated with a much larger barrier since it necessitate to slide one of the charged edges inside the hydrophobic DNA core. Furthermore, the binding free energy will most probably also be less favorable since it is associated with larger DNA deformations. Finally two minor groove binding modes have been observed as seen in Figure 10.16 c) and d). In one case the carbazole moiety points toward the bulk while the pyridinium edges are surrounded by the negatively charged phosphates. It has to be noted that the angle between the two bridges almost perfectly matches the width of the B-DNA groove allowing for largely favorable dispersive interactions. In the second minor groove binding mode the carbazole and only one of the charged edges lie in the minor groove, while the second bridge points toward the bulk. Although once again no significant steric hindrance is observed and favorable dispersive interactions are present, from an electrostatic point of view in this case only one of the two charged bridges directly interacts with the negatively rich phosphate backbone, hence the driving force will most probably be inferior compared to the previous minor groove binding mode that maximizes favorable electrostatic interactions. However, the situation appears to be much more complicated than the one previously hypothesized in particular because of the presence of a plurality of competitive binding modes that cannot be a priori ruled out.

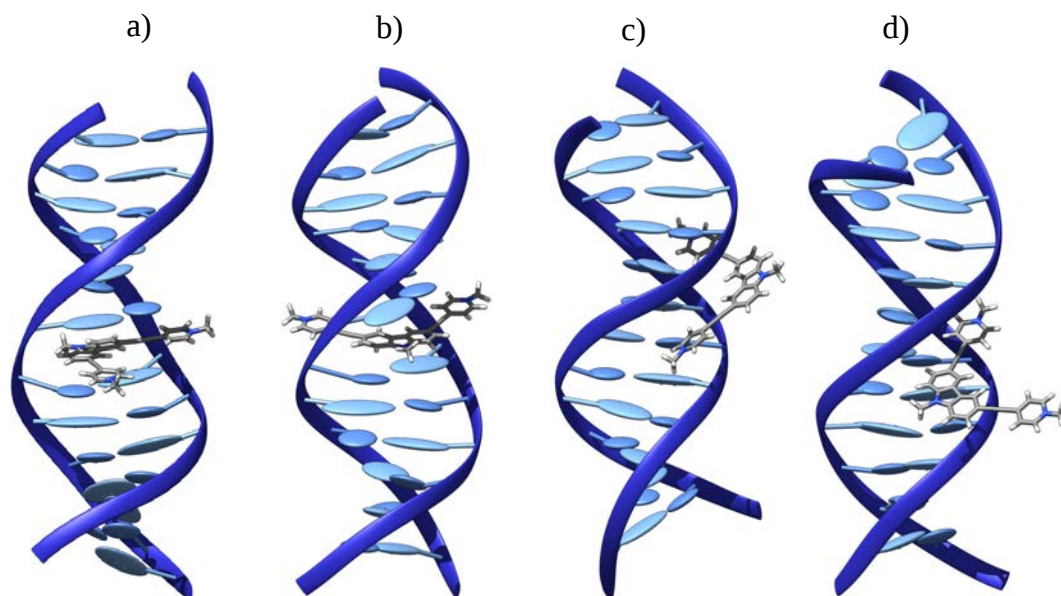


Figure 10.16: Representative structures from the MD simulations for the stable DNA-BMEMC binding modes. (a) Intercalation. (b) Alternative intercalation. (c) Minor groove binding. (d) Alternative minor groove binding.

Both the OPA and TPA spectra calculated for the different interaction modes are reported in Figure 10.18. Once again consistently with the situation reported for water solution the agreement with experimental results[3] is convincing. Most notably the interaction with DNA induces no significant modifications of the optical properties, the difference between the four interaction modes are also quite marginal and mostly evidenced in the TPA response. In particular once again  $S_1$  is the brightest state in OPA while  $S_2$  has the maximum intensity in TPA. The absorption maximum peaks at around 400 nm in the case of OPA while the maximum in the TPA cross section appears at around 700 nm, with the auxiliary peak due to  $S_1$  absorption appearing at around 800 nm. Interestingly enough, the shape of the  $S_1$  band in TPA seems to be much affected by the environment, in particular the band is broad and much less well defined in intercalation than in minor groove binding. This fact can be ascribed both to the modified electrostatic environment experienced in the hydrophobic pocket in the center of the DNA core and to the more constrained environment (mechanical embedding). As far as the photophysical pattern is concerned the global picture (see Figure 10.17) is the same as for the aqueous solution.

In particular one can see that the energy of the radical cation plus the solvated electron is generally lower than that of the  $S_1$  state pointing toward a spontaneous, fast and efficient photoionization independently of the specific interaction mode.

After the photoionization event two reactive species are produced in the vicinity of DNA, namely the solvated electron and the radical cation. The observed strand



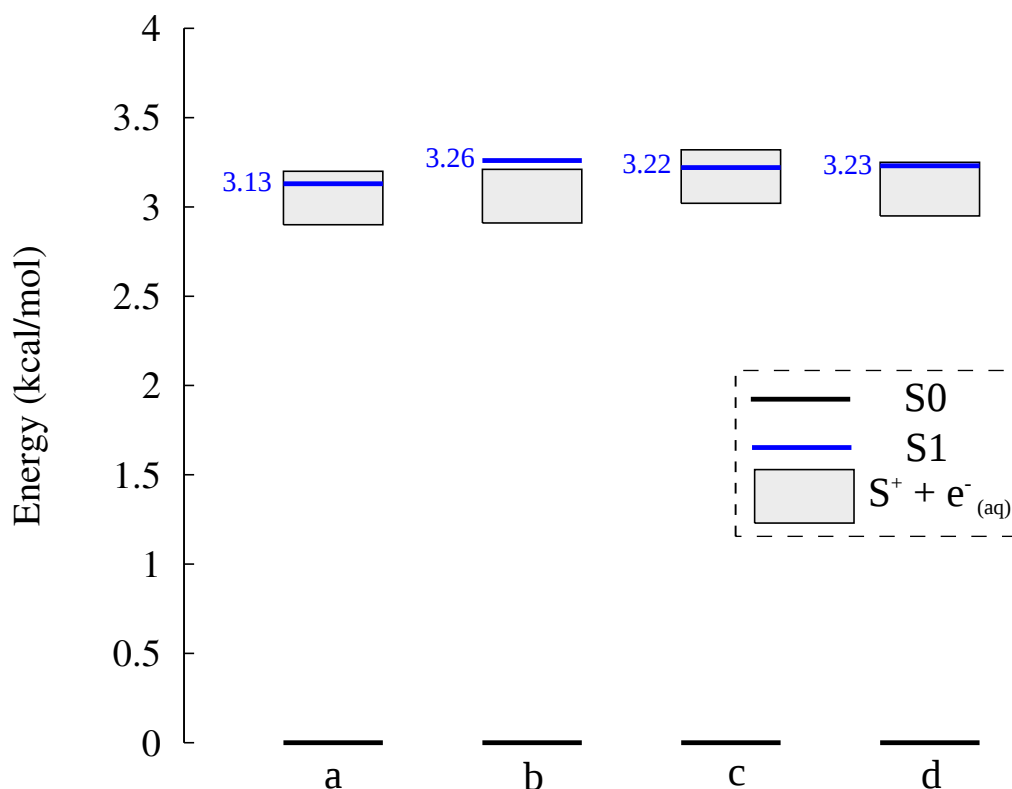


Figure 10.17: Jablonski Diagram for the photoionization processes extracted from 20 MD snapshots for the four stable interaction modes. A) Intercalation b) Alternative intercalation c) Minor groove binding d) Alternative minor groove binding. The square represents the energy of the solvated electrons in the limiting 3.3 and 3.6 eV.

breaks may arise from the direct action of low-energy electrons. Indeed, and contrarily to what was supposed in the past solvated electron may not only attack DNA nucleobases, but also induce reactions with the backbone resulting in the breaking of the strand[224, 225]. In addition one may consider the possible reactivity of the BMEMC radical cation. One possible and most probable subsequent pattern could be the sugar hydrogen abstraction that, especially in the case of the abstraction of the 5' H atom may lead to strand separation at the phosphate level. The high reactivity of radical cations is well established, however a key factor influencing the occurrence of the strand break is also the encounter probability between the reactants, namely the sensitizer and the DNA backbone hydrogens[180]. The encounter probability can be estimated for the different interaction modes by calculating the radial distribution function ( $g(r)$ ) between the BMEMC carbazole nitrogen atom and the hydrogens of the sugars as reported in Figure 10.19. We may observe that for each of the modes, with the partial exclusion of the alternative intercalation, we may find sugar H atoms close to the radical cation. In particular peaks in the  $g(r)$  curve are observed to be centered at around 3.0 Å with significant tails extending up to 2.0 Å. Those distances are certainly indicative of a high encounter probability

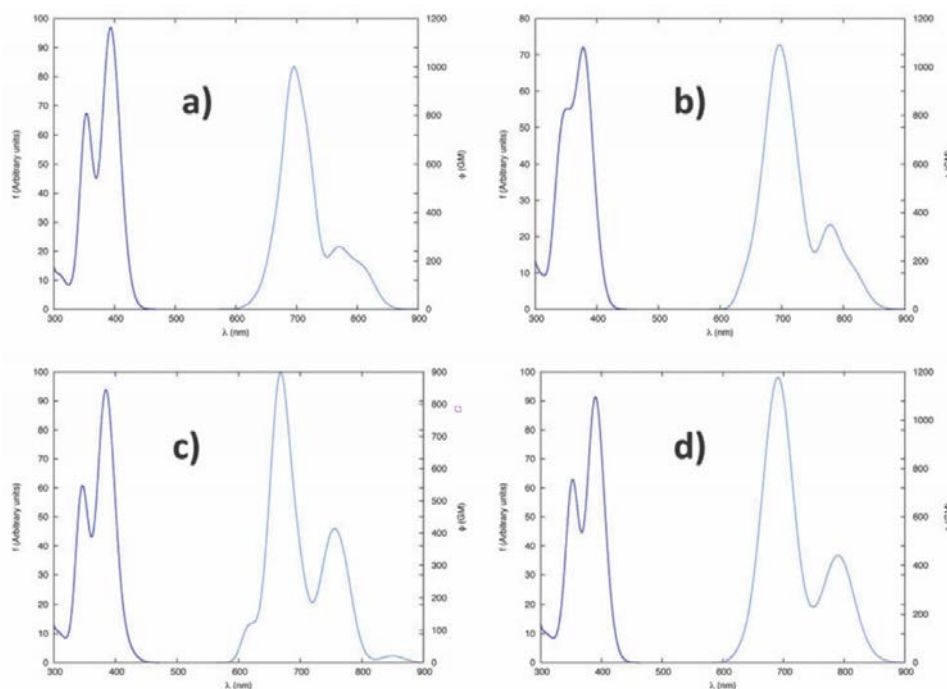


Figure 10.18: TD-DFT calculated spectra for BMEMC interacting with DNA. (a) Intercalation. (b) Alternative intercalation. (c) Minor groove binding. (d) Alternative minor groove binding. See Figure 10.16 for the corresponding structures.

that could suggest a relatively high probability of the abstraction reaction to occur. On the other hand we may see some differences in the  $g(r)$  as a function of the interaction mode. If we exclude the anyway less probable alternative intercalation mode we may see that minor groove binding tends to place the 5' and 4' hydrogen atoms closer to the radical cation, in particular, in the case of the alternative minor groove binding (see Figure 10.20 for a pictorial description of the naming convention). Those H atoms are the ones whose abstraction is most likely to result in a strand break. On the other hand in the case of intercalation the 1' atom happens to be the closest one to the reactive center and at distances consistent with its abstraction. This channel will end up favoring the emergence of abasic sites, even if the latter pathway could require the presence of molecular oxygen. Globally, it appears evident that the distance distribution between the BMEMC radical cation and the DNA backbone is coherent with a possible reactivity leading to the emergence of strand breaks, and this in particular is the case of minor groove binding. This fact coupled to the simultaneous presence of a solvated electron, as well as to the high TPA cross section and its favorable photoionization, allows to consider the former as an efficient and potentially extremely useful sensitizer absorbing in the near infrared and able to perform DNA sensitization under hypoxia conditions.



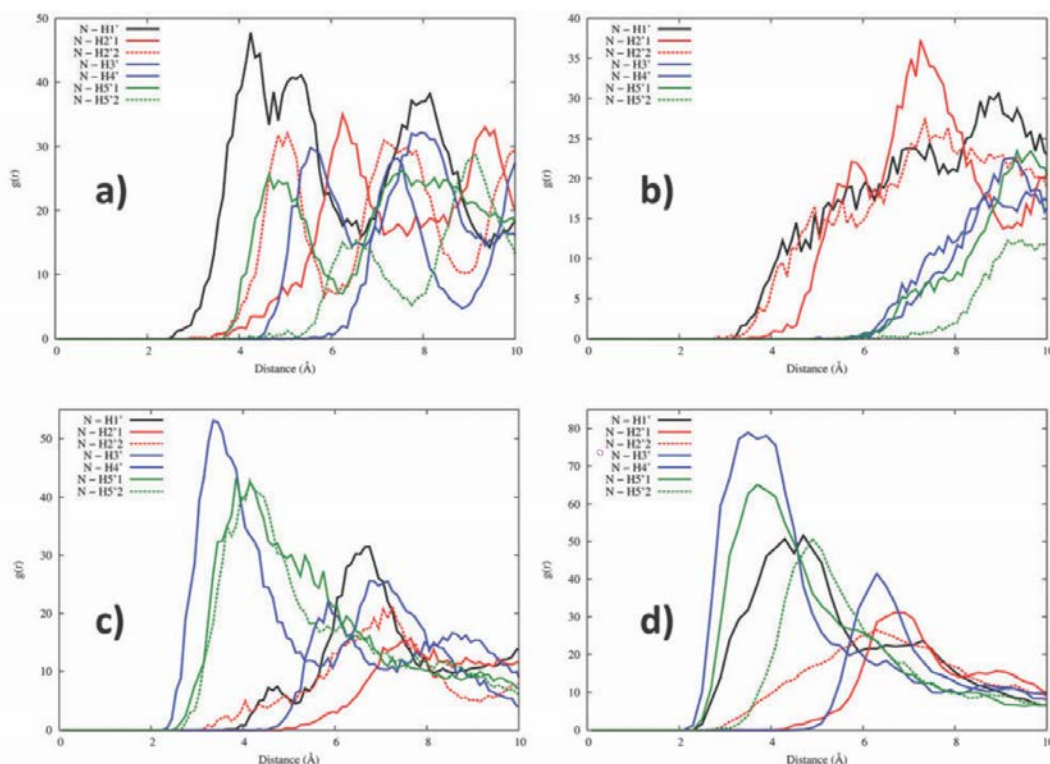


Figure 10.19: Radial distribution functions as extracted from the MD simulation between the BMEMC carbazole's nitrogen and the sugar hydrogen atoms. (a) Intercalation. (b) Alternative intercalation. (c) Minor groove binding. (d) Alternative minor groove binding.

## Conclusion

We report an extensive study on the linear and non-linear optical properties as well as on the further photophysical behavior of a novel dye, BMEMC, interacting with DNA and proposed for photodynamic therapy of deep solid hypoxic tumors. Using molecular dynamics simulation we have evidenced four competitive interaction modes between BMEMC and B-DNA. In particular, and in addition to what supposed experimentally we have evidenced the stable interaction through minor-groove binding. One- and two-photon absorption spectra have been obtained confirming that BMEMC acts as a very efficient sensitizer in the near infra-red region, i.e. at wavelengths that are able to deeply penetrate human tissues. It has also been evidenced that in OPA  $S_1$  is the brightest state, while the maximal cross section in TPA is for the  $S_0 \rightarrow S_2$  transition. Vibrational and dynamic effects have been taken into account thanks to the coupling of QM/MM methods to MD simulations to explore the conformational space. Remarkably enough a different influence of vibrational effects for TPA and OPA has been observed. This was also pointed out for different systems, using a quantum vibronic approach by Rizzo and coworkers[226]. Moreover, the photophysical pathways responsible for the sensitization has been clearly evidenced. We have proved that upon light absorption BMEMC is

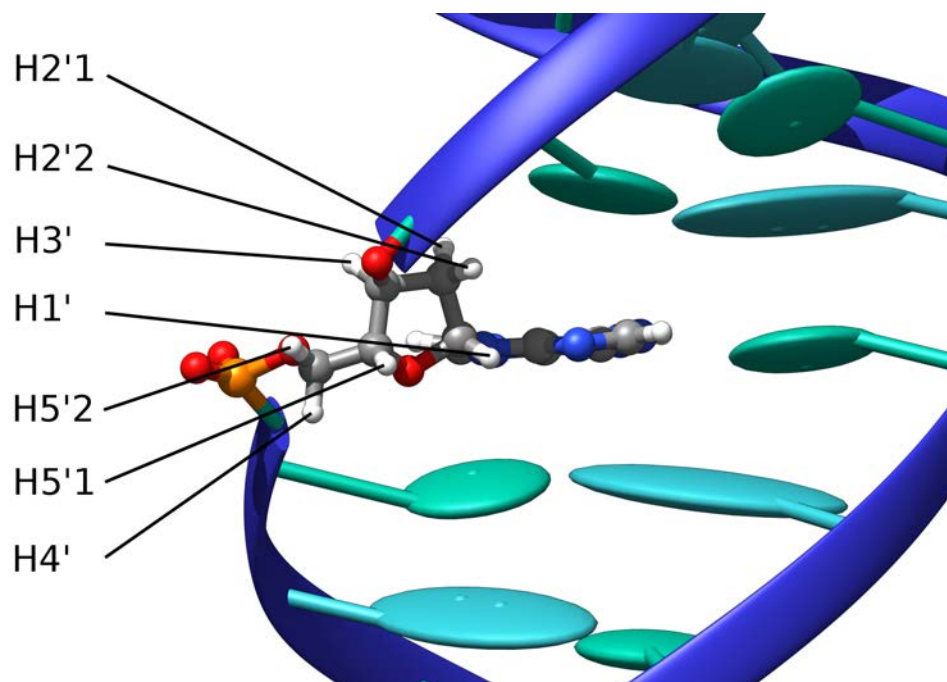


Figure 10.20: Cartoon representation illustrating the sugar hydrogen atoms naming convention. The sugar and phosphate backbones are represented in ball and stick while the other DNA constituents are in cartoon representation.

able to undergo spontaneous ionization to yield a radical cation and a solvated electron. This in turn results in potentially reactive species close to the DNA backbone, which can give rise to strand-breaking reactions. We have also demonstrated that the distance between BMEC and backbone hydrogens is compatible with a high encounter probability, and hence with reactivity. Interestingly, while intercalation seems to favor abstraction of hydrogen H1', which should lead to abasic sites, minor groove binding points toward the possible abstraction of hydrogen atoms (5' and 4') known to lead to strand breaks. In the following we plan to perform a full QM and QM/MM study of the reactivity of the BMEC radical cation in order to assess the kinetic barriers involved. The use of QM/MM accelerated molecular dynamics can also be envisaged to obtain fully converged free energy profiles. Those studies provide a clear molecular, atomistic and electronic description of the action mechanisms of a novel class of phototherapeutic prodrugs, and hence can be of invaluable help in the future rational design of phototherapeutic, near infra-red absorbing drugs, specifically targeting hypoxic tissues, characteristic of invasive solid tumors.

## 10.4 Pyo: Triplet-Triplet Energy Transfer (TTET)

### Introduction

Photosensitization of DNA has been demonstrated to be not only caused by external sources, but can also be induced by endogenous compounds, that can be considered as Trojan horses [227]. For example, 64-PP and more specifically its highly delocalized subunit 5-methyl-2-pyrimidone (Pyo) (Figure 10.21) is able to absorb light in the UVA region [228]. Compared to canonical nucleobases, Pyo is able to modify the absorption bands of DNA and extend it to higher wavelength. Moreover, Pyo containing double strands can overcome efficient triplet energy transfers to a closely thymine as demonstrated by Miranda et al. [227]. This electronic process is known to be able to trigger photosensitization mechanisms that might subsequently induce other lesions in DNA. Both type II and triplet transfer have been postulated for  $^3\text{Pyo}^*$  as well as the production of reactive oxidative species [227, 229, 230]. Indeed,  $^3\text{Pyo}^*$  is able to produce reactive singlet oxygen ( $^1\text{O}_2$ ) and hydroxyl radicals, evaluated by electron-paramagnetic resonance (EPR) analysis[227].

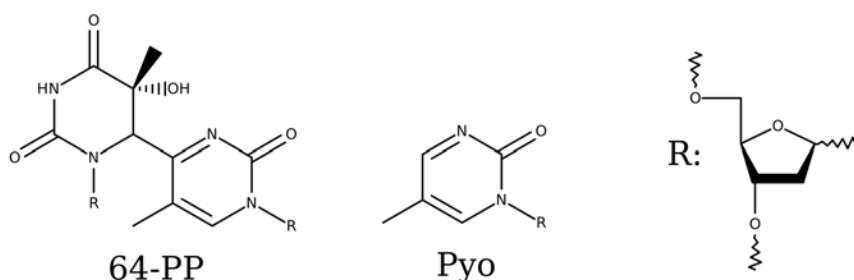


Figure 10.21: Molecular structure of the (6-4)PP and dPyo subunits.

The concept of introducing non canonical nucleobases in the DNA sequence that would act as endogenous photosensitizers seems to be a promising angle to develop new generations of sensitizing agents [231]. For example, the 6-thioguanine and 5-fluorocytosine have been patented in such aim and are used as anti-fungi [232]. Considering the vicinity between the Pyo unit to a thymine before or after in the sequence will help understand the triplet-transfer mechanism  $^3\text{Pyo} \rightarrow ^3\text{T}$  that is efficient due to Dexter mechanisms.

Moreover, understanding the structural and dynamic influence of Pyo on the highly flexible DNA moiety, its stability regarding the  $\pi$ -stacking and hydrogen bonding interactions is crucial if one wants to later on study the possible photosensitization pathways. Using state of the art molecular modeling techniques we investigated the two main aspects of the impact of Pyo in DNA. First we employed thorough Molecular Dynamic (MD) studies on the deformations of the dou-

ble strand containing Pyo compared to B-DNA, then thanks to hybrid quantum-mechanics/molecular-mechanics (QM/MM) techniques[122, 123] we described the electronic properties of Pyo, and more specifically its triplet form  $^3\text{Pyo}$  as well as its intrinsic absorption and fluorescence spectra. And finally we modeled the energetic levels of the triplet-triplet energy transfer to the thymine.

## Computational details

The Amber14 molecular-dynamics software package was used for all the calculations [132]. The force-field parameters were taken from the GAFF[17] and ff99bsc0[133] sets. The atom-point charges for the Pyo nucleobase were generated by using the restrained electrostatic potential [198] charge model of the Antechamber program. The Pyo (P) nucleobase was incorporated into the standard B-DNA conformation of the 10-bp double-stranded sequence d(CGCATAPACGC):d(GCGTAATGCG) at the underlined position. The chosen sequence is representative of a (6-4)PP lesion-containing oligomer; in our case, the cyclobutane dimer was replaced by Pyo and one thymine unit. The system was subsequently neutralized by adding 18  $\text{K}^+$  ions. The modified DNA molecule with its counterions was placed in a box ( $60.9 \times 60.4 \times 63.4 \text{ \AA}^3$ ) containing 5662 TIP3P water molecules[168]. The Particle-Mesh-Ewald method with a cutoff of 9  $\text{\AA}$  was used to calculate the electrostatic interactions. Throughout the simulation, periodic-boundary conditions were employed to eliminate undesirable edge effects. The molecular-dynamics simulations were performed by using the MPI version of the Particle-Mesh Ewald molecular-dynamics method (PMEMD).

The system geometry was initially optimized to remove bad contacts. Then, heating from the initial temperature of 0 to 300 K was performed in a thermalization MD run. The temperature was kept constant during the MD simulations by using the Andersen thermostat. A MD run of 100 ns was performed after equilibration to ensure equilibrium sampling.

The UV/Vis spectra of Pyo in DNA were computed at the hybrid quantum mechanics/molecular mechanics (QM/MM) level by using a local interface[111, 134], between the Tinker [137] and Gaussian09 [136] programs. The quantum partition consisted of the chromophoric unit Pyo, a linking atom was placed to cap the covalent bond with the sugar unit. The total spectrum was obtained as a convolution of 200 snapshots extracted from the MD trajectory to account for the dynamic and vibrational effects[127]. The first 30 excited states were obtained at the TD-DFT level of theory by using the 6-31+G(d) basis set[29] and the hybrid M06-2X30 functional. Polarizable embedding was modeled by taking into account the electronic response of the surrounding (ERS)[111]. The geometry of the first singlet and triplet excited states, respectively, was optimized to model the fluorescence and phosphorescence

properties. Finally, the shapes of the spectra were obtained in both cases by convoluting each vertical transition with a Gaussian function of full-width at half-length of 0.2 eV.

The triplet-triplet energy transfer between the artificial nucleobase Pyo and the close-by  $\pi$ -stacked thymine (T5) moiety was also investigated at the QM/MM level. Obviously, in that case, the QM partition was extended to comprise both Pyo and vicinal thymine T5. The transfer coordinate was chosen as the connecting path between the two optimized structures where the triplet state was located on Pyo  $Q(^3\text{Pyo}-^1\text{Thy})$  or on the thymine unit  $Q(^1\text{Pyo}-^3\text{Thy})$ . A simplified coordinate[123]  $Q(\xi)$  was used, with  $\xi$  going from 0 to 1, according to the following formula:

$$Q(\xi) = \xi Q(^3\text{Pyo} - ^1\text{Thy}) + (1 - \xi) Q(^1\text{Pyo} - ^3\text{Thy}) \quad (10.1)$$

To compare the triplet transfer in the case of the isolated nucleotide in aqueous solution, we built a model system composed of  $\pi$ -stacked 5M2P and thymine units, for which we applied the same approximate reaction coordinate following the necessary geometry optimization. In the case of aqueous solution, the environment was modeled by using the polarizable continuum model (PCM).

The nature of the triplet excited states along the simplified coordinate was analyzed by using the NTO formalism[210, 219].

Spin-orbit coupling was calculated at the TD-DFT level on the Pyo ground-state equilibrium geometry by using the Dalton 2015 code[216], and the mean field approximation.

## Results and discussion

To unravel all the aspects of the potential dPyo DNA triplet photosensitization, we started by analyzing the structural and dynamical properties of this species by means of molecular-dynamics (MD) trajectories. These simulations showed, with no ambiguity, the stability of the Watson-Crick pairing along the double-stranded (ds)-DNA, with only slight global deviations from the ideal helical structure.

### Structural features of the Pyo-containing decamer

The global helical structure is conserved along the entire trajectory of 100 ns. The Pyo artificial nucleobase, or its paired adenine base, could have been expected to be more labile because the hydrogen-bonding network between P6 and the paired A15 (Figure 10.22) is partially destroyed. The mobility of these bases was quantified by the time evolution of the distance d1, which is between the closest nitrogen atoms of the P6-A15 base pair (see H.1), which clearly shows the typical behavior of a

dual regime characterized by the oscillation between two conformations. In the first conformation, the value of  $d1$  is rather small, around 3.5 Å, and is also stabilized by a hydrogen bond between the Pyo oxygen atom and the amine nitrogen atom of the A15 base. In the second conformation, the value of  $d1$  is slightly larger with an average of about 9.0 Å. Therefore, partial opening and ejection of the Pyo nucleobase is assumed to take place.

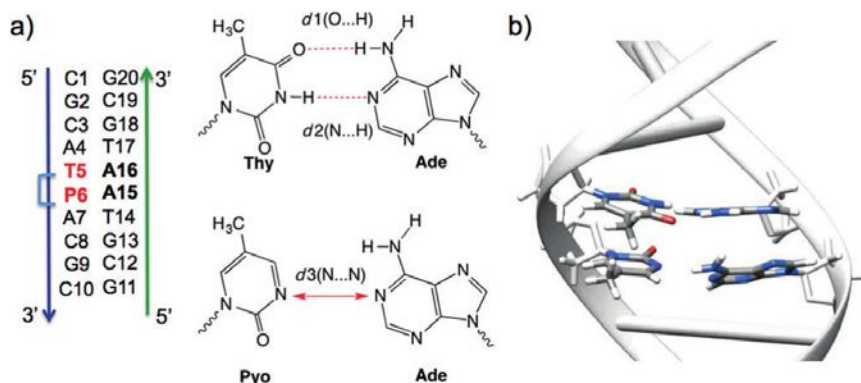


Figure 10.22: a) Sequence of the 10-mer in which dPyo is embedded as the sixth nucleobase. The Watson-Crick hydrogen-bond pairing between thymine (Thy) and adenine (Ade) no longer comes into play upon substitution by dPyo. Energy transfer toward the thymine T5 unit is investigated: b) representative structure extracted from the MD simulation (the T5-A16 and P6-A15 base pairs are shown as a licorice representation). The water molecules have been omitted for visualization purposes.

The partial disruption of the Watson-Crick network remains local and does not affect the T5-A16 pairing, with an average distance  $d2$  between the hydrogen and nitrogen atoms in T5 and A16, respectively, of 2.09 Å. The distance  $d3$  between the oxygen and hydrogen atoms in T5 and A16, respectively, is also typical of a hydrogen bond centered at around 2.08 Å.

The global bending of the double strand (Figure 10.23; as defined in [79]) remains limited by never exceeding 45° and averaging at 15°. After the occurrence of occasional larger deviations, the double helix rapidly evolves toward a much smaller bending. Beyond this overall stability of the Pyo-containing sequence, a closer inspection of the intra- and interbase parameters [79] reveals more subtle influences of the T → Pyo mutation (Figure 10.24).

Indeed, the partial disruption of the hydrogen-bonding network induced by Pyo does play a role. In particular, if the "stagger" parameters of the Pyo-containing base pairs have a much smaller deviation than the terminal base pairs, the "shear" and "stretch" parameters of the sixth base pair are globally larger than the other base pairs, even if in most of the cases they are still comparable to the deviations of the edge basis. This fact can be easily attributed to the less efficient locking of Pyo, which results in its higher mobility. This picture is even much more evident



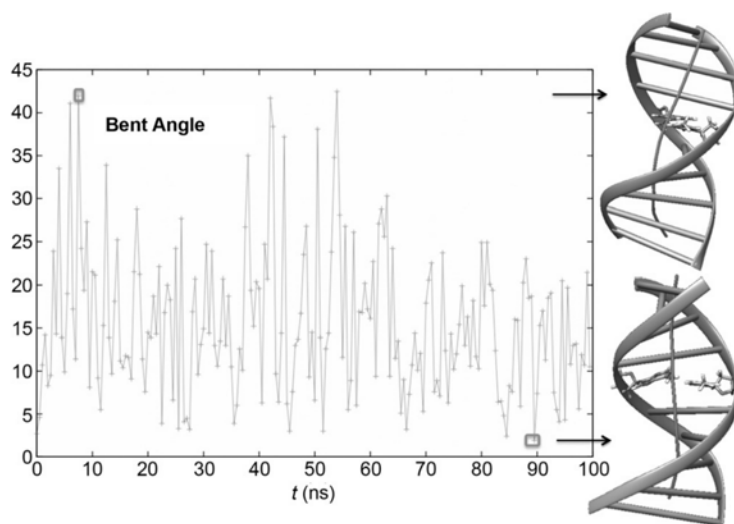


Figure 10.23: Time evolution of the global bending of the DNA strand. Two representations of the DNA double strand and the helical axis are given for the limiting conformations. The dPyo-containing base pair is displayed as a licorice representation

for the rotational intrabase parameters, with the particularly stunning case of the "propeller" that rises up to  $100^\circ$ .

Although the general picture of the interbase parameters is similar to the intrabase parameters, we may notice that the deviations of the "rise" and "twist" parameters are less pronounced for the base pair including Pyo. This effect could be related to the necessity of maintaining an optimal  $\pi$ -stacking interaction, which becomes even more important because of the hydrogen-bonding network weakening. On the contrary, the "roll" and "shift" parameters experience larger deviations, again probably due to the higher mobility of the Pyo nucleobase, due to the partial disruption of the hydrogen-bonding network. However, remarkably enough, this average value of deviation masks the dynamical oscillation between the two conformations. Indeed, the time evolution of the deviation of the "twist" parameter from the ideal B-DNA for the 6-7 base-pair couple perfectly matches the time evolution of d1 and hence can be regarded as an indicator of the transition between the two conformations (as reported in the Supporting Information). Nevertheless, this refined analysis also strongly confirms the overall stability of the Pyo-containing DNA strand, hence strongly supporting its efficient behavior as an artificial nucleobase.

### Spectroscopic properties

The intrinsic fluorescent properties of 5-methyl-2-pyrimidone deoxyribonucleoside have been characterized by using high-level *ab initio* calculations, while neglecting the influence of the DNA environment. The experimental absorption and emission spectra of Pyo-containing DNA have been reported [227, 233]. The calculated TD-

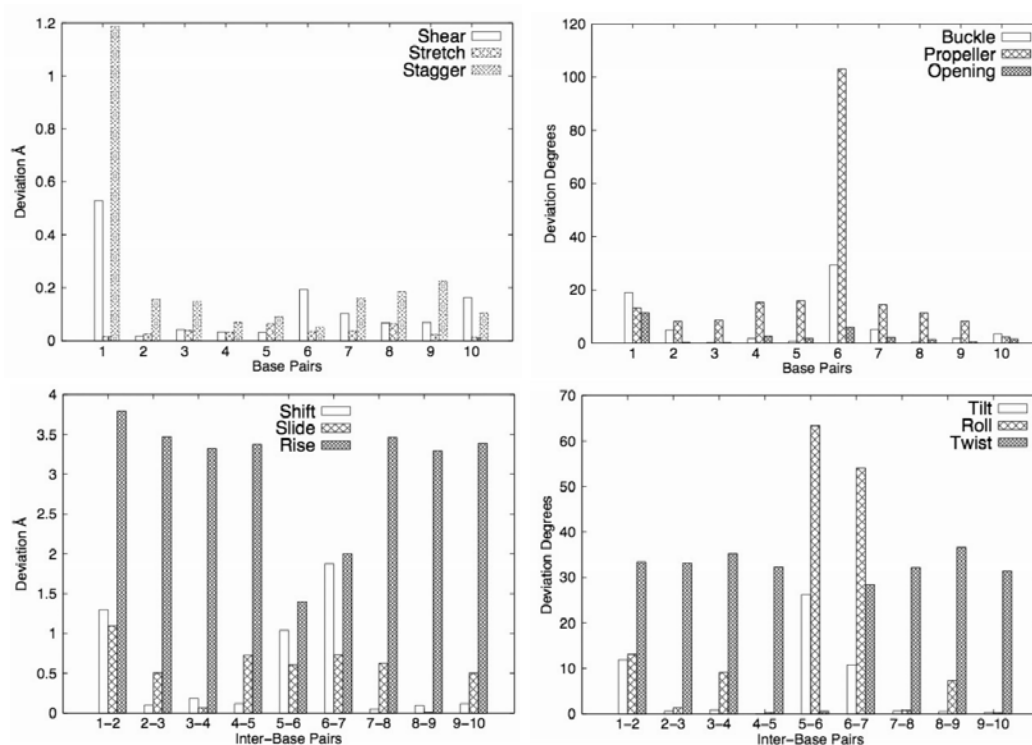


Figure 10.24: Average deviations from the ideal B-DNA structure for the intra (top) and interstrand (bottom) parameters.

DFT absorption spectrum of Pyo embedded in a B-DNA environment is reported in Figure 10.25 (left panel). It appears evident that a rather intense band at about 310 nm dominates the UVA spectrum of the artificial nucleobase. The fluorescence spectrum shows a tail and shoulder at around 460 nm, which is due to dynamic and vibrational effects optimization; furthermore, some of the snapshots are consistently more red-shifted because of the DNA-environmental constraints.

On one hand, this outcome confirms the experimental results reasonably well [227, 233, 234] for the mononucleoside and the DNA embedding. On the other hand, they corroborate the capacity of Pyo, and consequently of the (6-4)PP photolesion, to extend the DNA active spectral window, hence potentially increasing phototoxicity. Note that the absorption band appears rather asymmetrical, with a marked shoulder at about 290 nm due to vibrational effects. Unsurprisingly, the first excited state is described as a  $\pi$ - $\pi^*$  transition that extends over the entire conjugated ring. This outcome is also an indirect confirmation that our sampling of the ground-state conformations and our excited-state characterization protocol are sufficiently solid and robust. The calculated fluorescence spectrum is also well reproduced (Figure 10.25, right panel), and we found a maximum absorption at about 350-360 nm. The slight blueshift ( $\Delta\lambda = 20 - 30$  nm) relative to the experimental values has to be ascribed to the lack of solvent relaxation and can be seen as systematic, as we have analyzed in depth elsewhere for similar systems. Note that the fluorescence



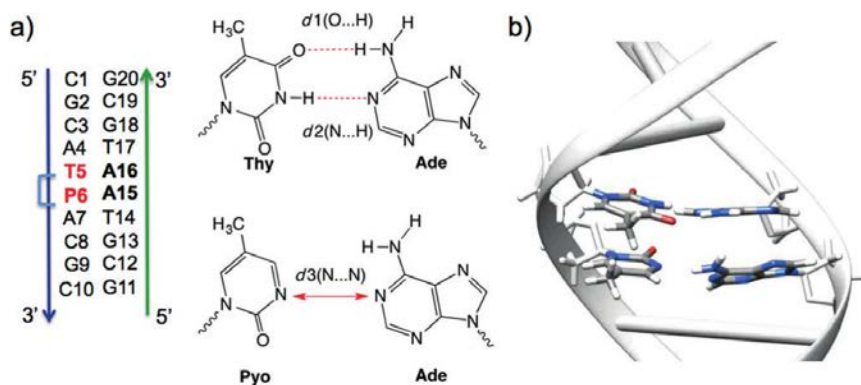


Figure 10.25: a) Sequence of the 10-mer in which dPyo is embedded as the sixth nucleobase. The Watson-Crick hydrogen-bond pairing between thymine (Thy) and adenine (Ade) no longer comes into play upon substitution by dPyo. Energy transfer toward the thymine T5 unit is investigated: b) representative structure extracted from the MD simulation (the T5-A16 and P6-A15 base pairs are shown as a licorice representation). The water molecules have been omitted for visualization purposes.

also presents a rather asymmetric curve, with now an even more pronounced shoulder, which becomes almost a second maximum absorption, at about 420 nm. This asymmetry is also paralleled by a greatly extended fluorescence tail that reaches the visible range, although with a very low intensity. Finally, the phosphorescence maximum has been estimated to be approximately 530 nm, which once again compares satisfactorily with the experimental results.<sup>16</sup> The possibility of a relatively easy and fast population of the triplet manifold is also confirmed by the high value of the spin-orbit coupling. Indeed, the  $S_1 - T_1$  coupling in the Franck-Condon region has been estimated to be  $28 \text{ cm}^{-1}$ . Such a value coupled with the small singlet-triplet energy gap is sufficient to induce the population of  $T_1$  in the ultrafast regime. To sum up, the spectroscopic features of Pyo, consistent with the results of Miranda and co-workers for the mononucleoside, are all concordant in establishing the picture of a possible Pyo-driven DNA triplet photosensitization.

### Triplet energy transfer

To assess for the capacity of Pyo to act as an insider sensitizer, modeling of its photophysical properties and, in particular, transfer toward thymine is compulsory. Hence, we resorted to a simplified coordinate (see the 10.4 for details) to estimate the possible pathways that lead to the energy transfer. First, it has to be noted that in the case of Pyo the different transfer pathways are much more sensitive to the chosen initial conditions than for other external sensitizers such as benzophenone [122, 123]. We delineate two representative cases from which some hints can be inferred (Figure 10.26). In most cases, the transfer is associated with a barrier, which is still of an order of magnitude (0.3-0.4 eV) that can be overcome in a relative

short time, as recently postulated for similar energetic barriers.<sup>18</sup> Nevertheless, the barrier is strongly dependent on the initial conditions, and in some cases it can reach values close to 1.0 eV.

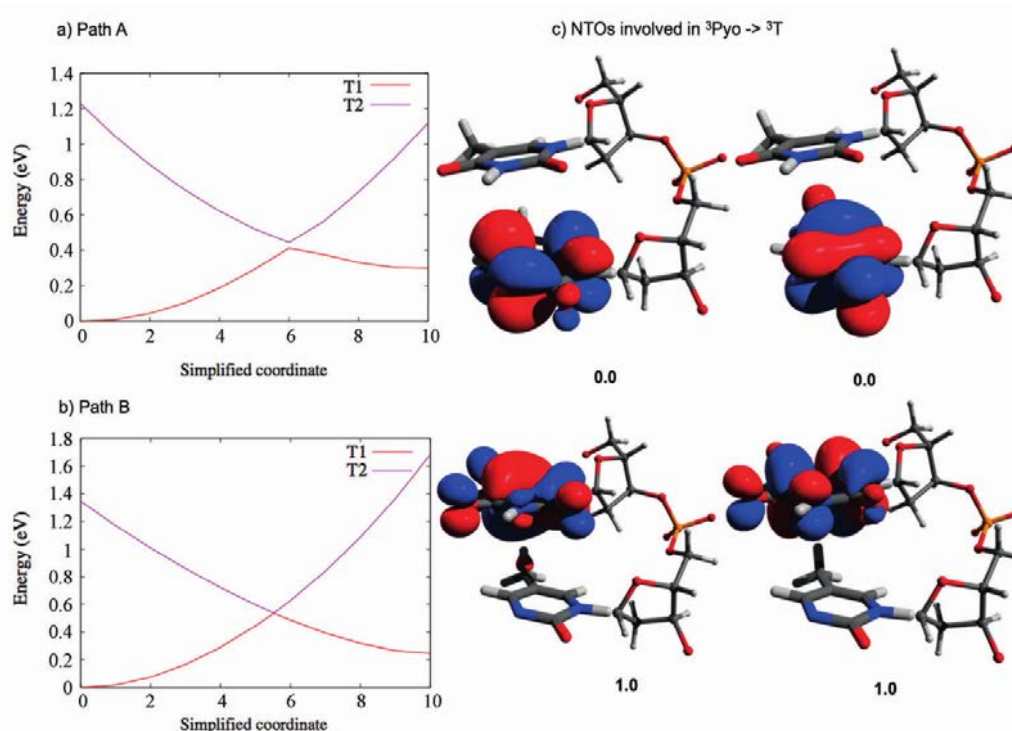


Figure 10.26: a, b) Energy profile along the  $Q(x)$  approximate coordinate for two limiting paths (different initial conditions). c) Natural-transition orbitals (NTOs) for  $x = 0.0$  (up) and  $x = 1.0$  (bottom) show the transfer of the triplet state between the two interacting monomers.

Second, the relative stability of  $^3\text{Pyo}$  and  $^3\text{T}$  deserves a comment. From the simple point of view of potential energy,  $^3\text{Pyo}$  appears to be about 0.2-0.3 eV more stable than  $^3\text{T}$ . However, these results do not take into account other factors, such as entropic influences, which could change the stability order. Furthermore, the presence of a nonzero barrier and a small energetic driving force also justifies the fact that the life-time of  $^3\text{Pyo}$  is long enough to allow the experimentally observed production of  $^1\text{O}_2$ .<sup>5a</sup> The energetic profile in B-DNA should also be compared to the one mimicking the nucleoside in aqueous solution, for which experimental results exist.<sup>5a</sup> To this end, we built a model system comprised of  $\pi$ -stacked 5M2P and thymine and we studied the triplet transfer in water and the gas phase, respectively (Figure 10.27). In both cases, the thymine-centered triplet state is energetically favored by about 1.0 eV, whereas a barrier of the same order of magnitude than the barrier observed for B-DNA still holds. This fact justifies the experimentally observed sensitization in the model systems.

Beyond the energetic point of view, the elucidation of the nature of the triplet state also deserves consideration. Notably enough, we experienced an almost perfect

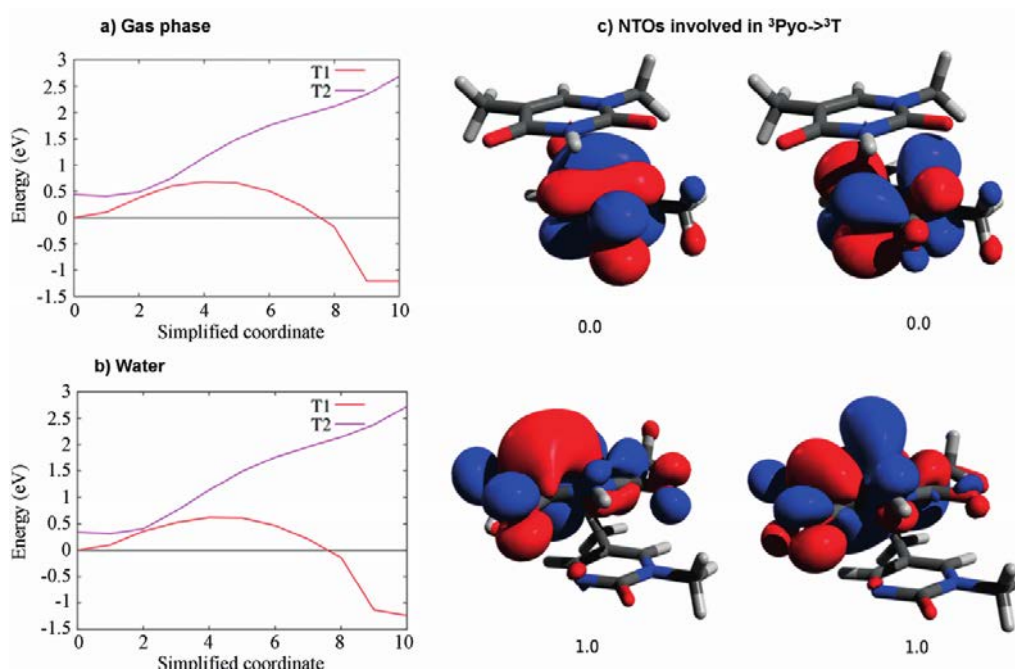


Figure 10.27: a,b ) Energy profile along the  $Q(x)$  approximate coordinate for p-stacked Pyo and thymine moieties in the gas phase and aqueous solution. c) NTOs for  $x = 0.0$  (up) and  $x = 1.0$  (bottom) shows the transfer of the triplet state between the two interacting monomers.

crossing in some cases between the  $T_1$  and  $T_2$  potential-energy curves in a situation somehow reminiscent of a conical intersection (Figure 10.26 b). This occurrence can point to a conformation that should be able to funnel the energy transfer. Indeed, because of the proximity of the two monomers, we are undoubtedly confronted with a Dexter energy-transfer process, which is known to be driven by the overlap between the wavefunctions of the interacting monomers [122, 123, 229, 235]; namely, the higher the overlap, the better the transfer efficiency. In the case of the  $^3\text{Pyo} \rightarrow ^3\text{Thy}$  energy transfer, we face a huge and extended mixing of the electronic densities of the two monomers, as clearly evidenced by the NTOs [210, 219]. The electronic-density mixing is particularly evident when the  $T_1$  and  $T_2$  potential-energy curves actually cross (Figure 10.28).

This electronic-density mixing leads to an impressively high overlap of the wavefunctions (estimated 0.76 electrons). The latter estimation has been straightforwardly obtained on the basis of the NTO square overlap. As a comparison, the efficient energy transfer of benzophenone was driven by a wavefunction overlap of only about 0.1 eV. When dealing with Pyo "internal" photosensitization, we are therefore confronted with a case in which the Dexter energy transfer is expected to be extremely favored and highly efficient; consequently, the sensitization can be thought to be strongly "kinetically" favored. On the other hand, the prevalence of the Dexter energy transfer implies a strong distance dependence. Indeed, the

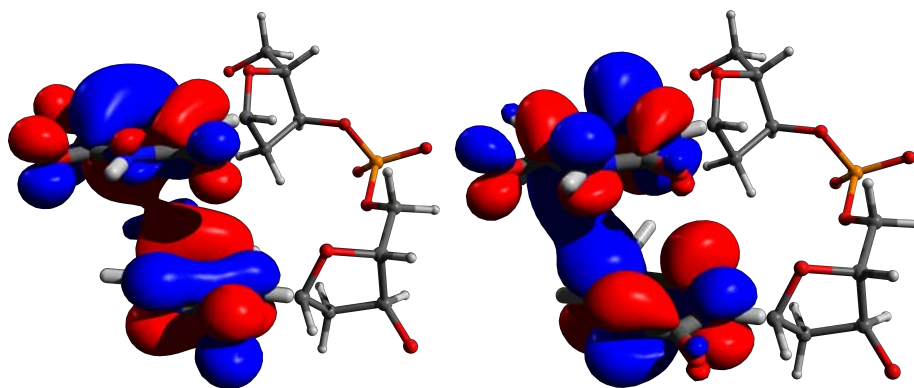


Figure 10.28: NTOs at the T1/T2 crossing point (Figure 10.26;  $x = 0.5$ ) show a total delocalization of the wavefunction between the two monomers, hence confirming the high efficiency of the energy transfer.

Dexter triplet transfer decays exponentially with the monomers distance, hence its efficiency will become negligible for nucleobases that are not proximal.

## Conclusion

We have unraveled the structural and photophysical properties of an artificial nucleobase (Pyo) embedded in a B-DNA representative pentadecamer. The Pyo nucleobase can be thought as an exogenous artificial unit or as the light-active fragment of the 64-PP photolesion. As far as the first aspect is concerned, we have clearly evidenced by means of molecular-dynamics simulations that the inclusion of the Pyo unit does not induce significant global structural modification of the B-DNA helix. However, partial disruption of the Watson-Crick pairing results in a higher mobility of the artificial nucleobase. As much as the photophysics is concerned, we have shown that Pyo is characterized by efficient absorption in the UVA region. Henceforth, it may act as an internal photosensitizer by extending the absorption region of native DNA. Upon absorption, Pyo excited states may evolve toward the population of the triplet state and subsequent photosensitization. The triplet-energy transfer from Pyo to the nearby thymine units is possible, is energetically favored for nucleotides in water, and can not be excluded in DNA, even if less favorable since it requires to overcome an barrier, hence the DNA environment compared to water may be seen as a sort of self-protection. Furthermore, the Dexter energy-transfer mechanism should be particularly efficient because the overlap of the wavefunctions of the two monomers is extremely high. Our study has confirmed the possible effectiveness of DNA sensitization by the Pyo nucleobase. Although care should be taken because the conformations and configurations, and in particular the  $\pi$ -stacking interactions, can significantly differ, this behavior means that the (6-4)PP photolesion actually acts as a "Trojan horse" inside the DNA sequence and can provoke

further lesions. This fact may in turn give rise to a situation in which repair is known to be extremely critical. The relatively high value of the spin-orbit coupling, thus favoring Pyo triplet population, and the presence of a barrier to the transfer also justifies the action of Pyo as a type-II sensitizer that produces singlet oxygen, in which case guanine would be the preferential nucleobase target. Furthermore,  $^3\text{Pyo}$  could also extract hydrogen atoms from 2-deoxyribose units because previous studies have shown that excited pyrimidone derivatives have access to this photochemistry[236, 237]. Furthermore, because the Pyo nucleobase does not modify the structure of DNA considerably, one may hypothesize its use as an exogenous nucleobase analogue for subsequent phototherapy purposes. However, this theoretical hypothesis should be checked carefully because triplet photosensitization processes are known to be mutagenic and carcinogenic themselves and metabolic pathways to incorporate Pyo in DNA are not known.

# Chapter 11

## Damaged DNA: structure and recognition

### 11.1 Cluster abasic sites

#### Introduction

The excision of a DNA base, either as a consequence of the exposure to ionizing radiations, and the consequent activation of radical species such as  $\text{OH}\cdot$  or following biomolecular pathways such as repairing (NER, BER 2.7 )[\[238\]](#), produces one of the most common DNA lesion, the so-called apurinic/apyrimidinic site, namely AP site. This lesion exist in two different cyclic chiral forms, R and S, that can switch from one to the other through the ring oppening. Even though the cyclic form is predominant (99% [\[239\]](#)), its open form has been described as the possible cause of interstrand cross-links [\[240\]](#) lesions which are one of the most deleterious lesions.

The production of radical oxidative species in a localized area can produce not only one isolated lesion be a series of damages accumulated in the DNA structure, they are called clustered lesions. Their low repair rate compared to single lesions make them much more dangerous [\[241\]](#). Among the family of clustered lesions, double strand breaks (DSBs), oxidised guanines and AP sites have been reported [\[242\]](#). The comprehension of their effect on the DNA dynamic properties, flexibility, repair rate and replication is still open and requires more investigations[\[243–245\]](#).

The case of clustered AP sites[\[246, 247\]](#) interestingly showed different repair rates by *Escherichia coli* endonuclease IV (Nfo) and the corresponding human endonuclease APE1 depending on the relative position of the lesions in the sequence of nucleotides. As an example, NMR and MD simulations[\[248\]](#) have evidences extrahelical positions for both 5' and 3' AP sites, but for the first they were located in the minor groove while the second ones where in the major.

We used here classical MD simulations that already proved to be efficient in



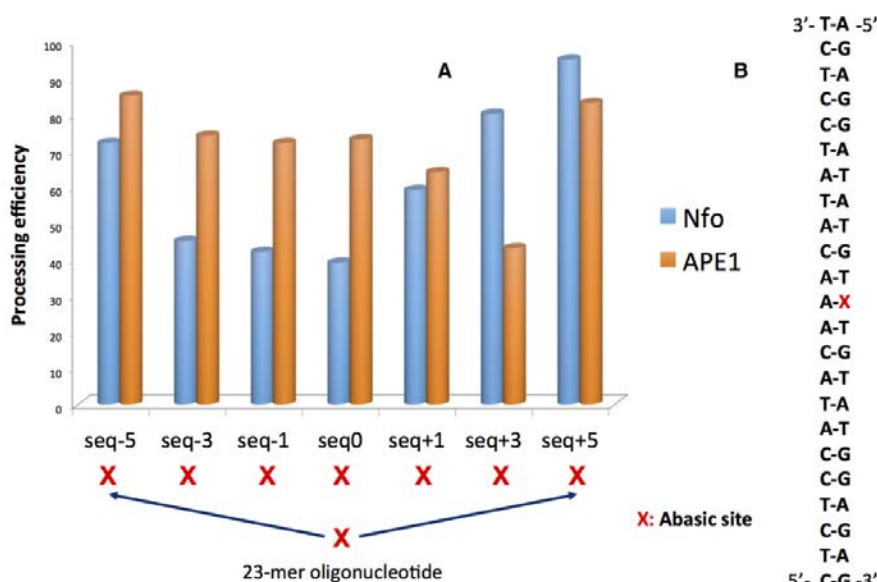


Figure 11.1: (A) Repair rate of the different clustered AP oligomer using Nfo (*Escherichia coli*) and APE1 (human) enzymes, respectively. (B) The original DNA double strand featuring the position of the AP marked as X. The clustered AP containing lesions have been produced inserting a secondary AP site in the complementary strand and up to five residues apart from the original lesion, see also Scheme 1. Data are adapted from references [246, 247].

studying damaged DNA double strands (abasic sites[249–251], bulky adducts [242, 252], cross-links [253] or photolesions[123, 254]). We especially focused on 7 clustered AP sites as reported by Georgakilas et al.[246, 247] (Figure 11.1) in which they obtained the repair rates of each situation. As showed in Figure 11.2 we created in silico a 35 base pairs long DNA double stand in which we incorporated clusters of AP sites locates at  $\pm 0, 1, 3, 5$  (for exemple the 0 system is composed of facing AP sites). The results of our MD simulations, directly compared to the experimental activities will help in the understanding of the relation between structural deformations induced by these lesions and enzymatic repair efficiencies.

## Computational details

**Oligonucleotide structure:** The choice for the specific oligonucleotide sequence presented in Figure 11.2 was based on the analytical existing results for AP DNA cluster processing for both Nfo and APE1 enzymes as described in the original publications by Georgakilas et al. [246, 247]. One of the reason behind the choice of this specific sequence is that it did form no hairpin loops and was thermally stable in solution even after the induction of the AP site. Current theoretical calculations agree with this initial finding and additionally using BLAST (blast.ncbi.nlm.nih.gov) sequence alignment analysis it was found that the specific sequence shared great

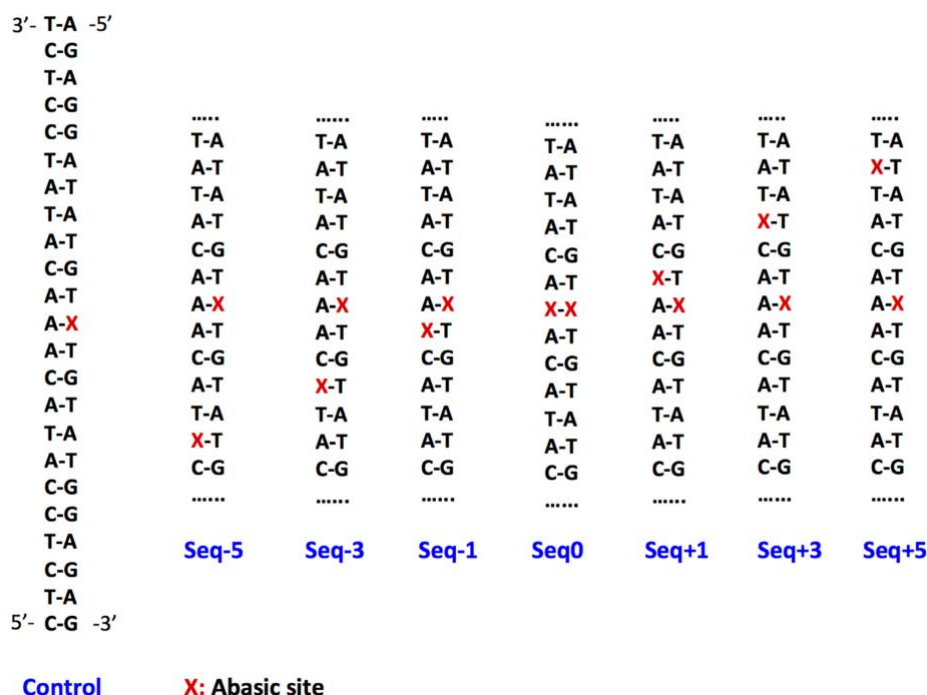


Figure 11.2: Representation of the sequence of the studied oligonucleotides as taken from [246, 247].

homology with human chromosomal regions.

**Computational protocol:** The Amber12 software[132] package and the Amber ff99 force field[17, 167] including bsc0 corrections[133] were used for all classical MD simulations. Seven oligonucleotides were built on top of canonical B-DNA sequences produced with the Nucleic Acids Builder module of Amber[132]. Abasic sites were inserted at specific positions to generate the seven sequences corresponding to the synthesized oligonucleotides of reference[246], with the xleap module. The parameters for the AP site were generated with antechamber and parmcheck subprograms, and atom point charges were computed using the RESP protocol[17]. Sodium cations ( $\text{Na}^+$ ) were added in order to neutralize the systems, which were immersed in a truncated octahedral TIP3P[255] water box, counting between 18 272 and 22 405 water molecules depending on the system. Each AP-containing oligonucleotide structure was first minimized in a 10 000 steps simulation, including 5000 steps of steepest descent. Then, a thermalization step was performed to heat each system from 0 to 300 K in 20 ps. The temperature was kept constant during the following steps using Langevin thermostat with a collision frequency  $\gamma_{\text{ln}}$  of  $1 \text{ ps}^{-1}$ . A first 100 ps equilibration run was performed in NPT conditions, followed by a second one in NVT conditions. Finally, a 400 ns production was executed with constant pressure.

In order to check how the cluster lesions affect the global B-DNA structure,



we also computed the overall bending of the whole clustered AP oligomers and compared it to the control, i.e. the strand containing only one AP site. We used the bending measurement method as defined within the Curves+ program[79]. To assess residency times for hydrogen bonds, we fix a cut-off on the  $H \cdots O$  of  $H \cdots N$  distances of 2.4 Å. We adopt the criterion that extrahelical positions correspond to a distance  $C1' \cdots C1'$  greater than 14 Å, as proposed by Lavery et al.[250]. This criterion was used to infer the percentages of extrahelicity for abasic sites and other pyrimidines. Furthermore, we also calculated the minor groove occupancy by the extrahelical AP; details are given in appendix (Table I.1 and Figure I.1).

Note that due to the equilibrium between the closed and open form AP sites can exist as  $\alpha$ - or  $\beta$ -anomers. However, as pointed out both theoretically[256] and via NMR determinations[257] the two isomers give rise to duplexes having the same structural characteristics. This aspect will be computationally too demanding to analyze systematically, yet preliminary MDs on seq0 and seq+1 show a conservation of the structural and dynamical properties.

## Results

MD simulations were performed over 400 ns, for seven sequences and a reference one as depicted in Figure 11.1 and Figure 11.2. We described hereafter the main structural outcomes obtained for these oligonucleotides. Particular attention will also be given to the global bending and extra-helical position of the AP sites as reported in Table 11.1.

Table 11.1: Average structural descriptors for the damaged oligonucleotides over 400 ns and the latest 40 ns along MD simulations. The global parameter such as bending is not strongly affected by the presence of the cluster lesion. On the other hand the extrahelicity of the AP site, which is usually considered a signature of this class of lesions, is strongly altered and shows a remarkable clustered AP position dependency that can partially correlate with the processing rate.

Sequence	Bend (400 ns)	Bend (40 ns)	AP extrahelicity
seq-5	$34.1 \pm 18.7^\circ$	$25.2 \pm 13.4^\circ$	AP7: 12.4% ; AP35: 12.5%
seq-3	$38.3 \pm 19.3^\circ$	$24.9 \pm 19.3^\circ$	AP9: 36.6% ; AP35: 27.7%
seq-1	$31.1 \pm 16.5^\circ$	$23.6 \pm 14.1^\circ$	AP11: 2.5% ; AP35: 3.9%
seq-0	$34.3 \pm 17.8^\circ$	$35.5 \pm 18.2^\circ$	AP12: 81.2% ; AP35: 81.2%
seq+1	$27.0 \pm 14.4^\circ$	$27.4 \pm 13.5^\circ$	AP13: 2.1% ; AP35: 4.2%
seq+3	$35.3 \pm 19.4^\circ$	$33.7 \pm 18.1^\circ$	AP15: 13.2% ; AP35: 3.9%
seq+5	$31.5 \pm 16.5^\circ$	$27.3 \pm 13.6^\circ$	AP17: 30.3% ; AP35: 28.0%
Reference	$34.1 \pm 17.2^\circ$	$25.3 \pm 11.8^\circ$	AP35: 35.9%

We first describe the situation where the two AP sites face each other (seq0). Because of the constraints imposed by the DNA backbone, the damaged sites are too distant to associate together by HB as the B-helix evolves in time. Instead,

the oligonucleotide deploys an alternative strategy of stabilization by excluding the two AP sites (i.e. forcing their extrahelical residency), this in turn enables to stack T11 and T13, and A34 and A36. The former conformation is easily accessed and the structural reorganization occurs in less than 2 ns. The two Watson-Crick base pairs T11:A36 and T13:A34, hence maintain their pairing all along the simulation. As a result of the peculiar stabilization strategy the two AP sites present the same high extrahelical residency of 80%. Yet the two AP sites are still directed toward the core of the helix due to transient HB interactions with O4' and N3 of proximal adenines (see Figure 11.3C). The 23-bp oligonucleotide presents a locally narrower minor groove in correspondence of the AP sites due to the B helix compression subsequent to the ejection of the lesions, but no marked deviation of the backbone. In contrast, backbone deformation leading to bulge have been observed in the case of single AP site and extensively studied[258–261]. The absence of bulge is an important feature that one can relate to the lack of repair of seq0: the duplex structural reorganization, acts as such to make the damaged oligonucleotide extremely close to a non-damaged 22-bp B-DNA strand with, in its center, two dangling extrahelical AP sites. Consequently, not only recognition is made harder but a repair would imply the disruption of the newly formed and stable T11  $\cdots$  T13 and A34  $\cdots$  A36 stacking (Figure 11.3), i.e. an energetic penalty corresponding to  $\sim 10$  kcal.mol<sup>-1</sup>. However, even though the structural factors are certainly important the corresponding barrier is not too high and indeed seq0 is only barely less efficiently excised than seq-1 or seq-3 by Nfo. Concerning the evolution of the strand one may also note that the bending assumes an average value of  $35.5 \pm 18.2^\circ$  over the last 40 ns, which is only slightly more pronounced than the single-AP-containing control sequence ( $25.3 \pm 11.8^\circ$ ).

As the two AP sites are shifted by one base pair upstream (staggered 3'), i.e. for seq-1 a very distinct stabilization scheme is evidenced (Figure 11.4) with the two AP sites now occupying only scarcely extrahelical positions ( $\sim 3\%$ , see Table 11.1). Figure 11.4C reveals that AP11 first pairs with A36:N1, through the terminal hydroxyl hydrogen H1, until a most stable HB with the vicinal O5' atom occurs  $\sim 330$  ns. AP35 also involves a stable HB with A12:N1 up to 60 ns, and reforms interactions at 250 ns, also implying A34:N3. As a consequence all along the MD trajectory one can observe a swapping between the two HB patterns. The presence of the two AP sites initially isolates the orphan nucleobases A12 and A36 that are stabilized partly by HB formation. However, A36 reinstate favorable  $\pi$ -stacking interactions by intercalating between G10 and A12. This interstrand stacking of three purines triggers stabilization (compared to a situation where pyrimidines would be involved) and is accompanied by a noticeable local narrowing of the double helix (C1'  $\cdots$  C1' distances of  $11.2 \pm 1.3$  Å, versus  $18.2 \pm 0.5$  Å in absence of lesions). The geometrical

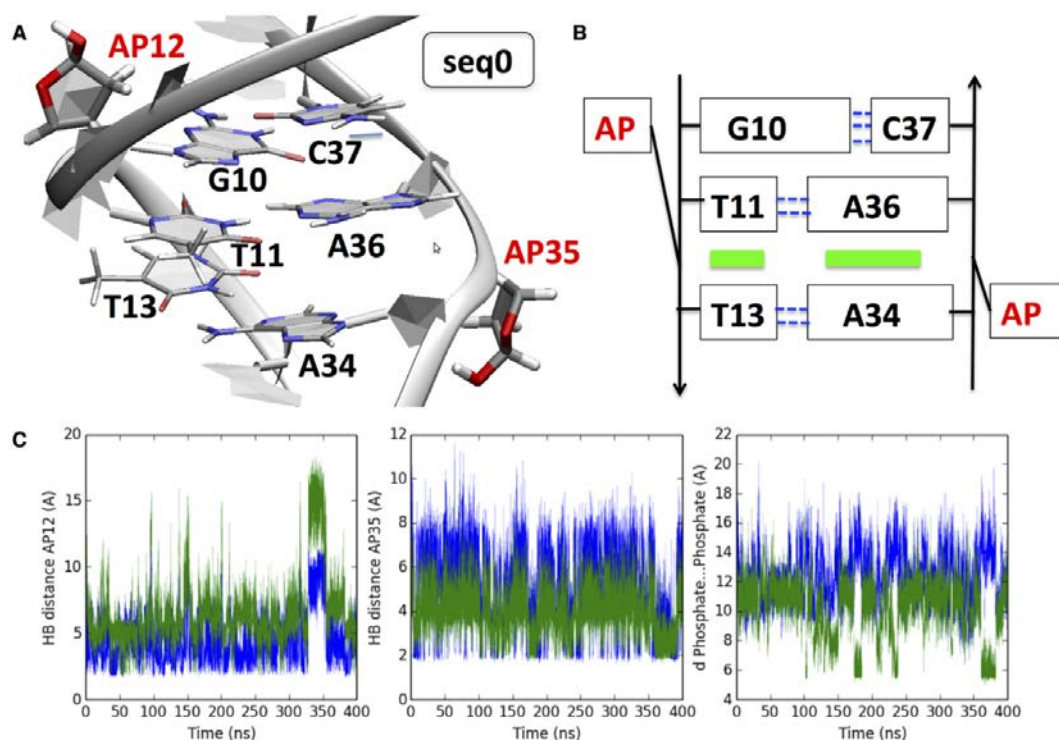


Figure 11.3: (A) Cartoon representation of the final structure obtained after 400 ns for seq0 and (B) scheme of the non-covalent interactions reshaping the oligonucleotide: green boxes indicate new  $\pi$ -stacking, and dashed blue lines indicate Watson-Crick hydrogen bonds. (C) Time evolution of HB and phosphate-phosphate distances. Left panel: blue line AP12H1...T11O4 distance, green line AP11H1...A36N1. Middle panel: AP35H1...A12N1 distance blue line, AP35H1...A34N3 distance green line. Right panel: blue line A12P...C37P distance, green line T15P...A36P.

effects induced by this peculiar arrangement are propagated to the nearby bases: C37 adopts an extrahelical position at 136 ns up to the end of the simulation, the former configuration is moreover locked by one strong hydrogen bond with one phosphate oxygen A36:O2P (Figure I.2). Furthermore, we see that seq-1 does not exhibit a significant deviation of the overall bending when compared to the control sequence.

In the case of seq+1, i.e. the reciprocal downstream duplex featuring the two abasic sites at 13th and 35th positions, we observe very different structural features, with notably no extrusion of pyrimidines (Figure 11.5). AP13 and AP35 sites rapidly seek to form hydrogen bonds with the orphan A12 and A34 bases, and most notably the two AP sites do not adopt extrahelical position. AP13 is hydrogen bonded to A12:N3, but this interaction is disrupted after around 280 ns as AP13 favors a pairing with A34:N1 and also a more transient interaction with a proximal O5' oxygen of the sugar moiety. A12, through its nitrogen N1, becomes pivotal in favoring a rearrangement ultimately ending up in the formation of a HB for AP35. Also G14 ends forming a hydrogen bond with the AP13 residue further stabilizing its position (see Figure 11.5). One identifies a stacking between A12, A34 and G14,

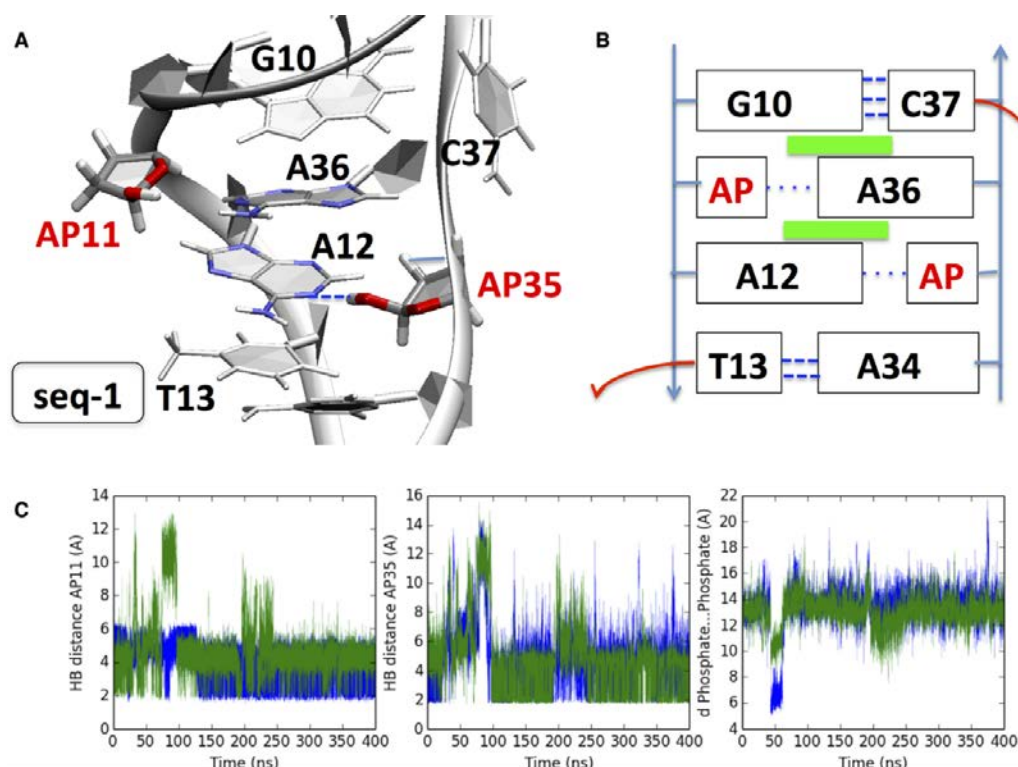


Figure 11.4: (A) Cartoon representation of the final structure obtained after 400 ns for seq-1. (B) Diagram showing the interaction pattern: Green rectangles indicate transverse  $\pi$ -stacking, and dashed blue lines new hydrogen bonds (thick ones for Watson-Crick pairing). (C) Left panel: blue line AP11H1...AP11O5 distance, green line AP11H1...A36N1. Middle panel: blue line AP35H1...A12N1 distance, green line AP35H1...A34N3 distance. Right panel: blue line A12P...C37P distance, green line T13P...A36P. The two AP sites are mostly directed within the helix, owing to HB with vicinal nucleobases: the cytosine C37 adopts an extrahelical position. A36 develops interstrand stacking with G10 and A12.

but with only a limited impact on the C1' ... C1' distances compared to seq-1, that correlates with the non-exclusion of the vicinal pyrimidines. The duplex exhibits a bend angle of  $27.0 \pm 14.4^\circ$ , almost constant along the trajectory.

As the AP sites are more widely separated, like in seq+3, the duplex still evolves to develop interstrand purine stacking, for A12 and A34 as shown in Figure 11.6 (bottom). This implies a moderate constriction of the B-helix reflected by C1' ... C1' distances up to  $\sim 16.3 \pm 1.1$  Å. We also note that the Watson-Crick pairing of the two base pairs in between the defects, T13...A34 and G14 ... C33 is maintained. The AP15 and AP35 damages adopt transiently extrahelical positions accounting for 13.2% versus 3.9%: hence they are only barely solvent exposed, and their -OH group transiently points toward an HB partner available in the surrounding (G14 and A12, respectively). On the other hand the orphan A32 base adopts an intra-helical position and develops attractive stacking interactions with G14. The DNA backbone is flexible enough to accommodate interstrand stacking, and the two



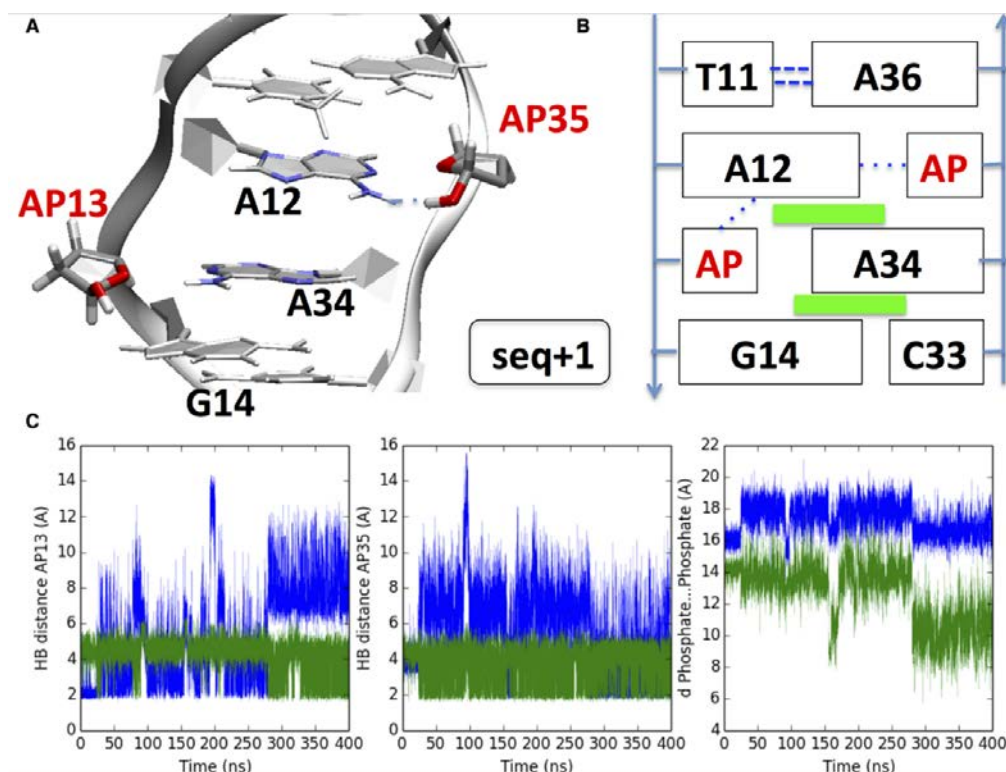


Figure 11.5: (A) Cartoon representation of the final structure obtained after 400 ns for seq+1 and (B) diagram illustrating the non-covalent interactions and position of the bases. Dispersion interactions are represented as green rectangles, while dotted blue lines represent hydrogen bonds. (C) Left panel: blue AP13H1...A12N3 distance, green line AP13H1...AP13O5. Middle panel: blue line A12N1...AP35H1 distance, green line AP35O5...AP35H1 distance. Right panel: blue line A12:P - A36:P distance, green line AP13 - AP35. The two AP sites are mostly directed within the helix, owing to HB with vicinal nucleobases; however, an interstrand stacking between A12...A34...G14 is observed.

central base pairs are left unaltered in their Watson-Crick pairing. Interestingly the structure is slightly more bent than the reference sequence by more than  $8^\circ$ .

The structural evolution of the 23-bp duplex seq-3 (Figure 11.6 top) featuring AP sites as 9th and 35th residues is also driven by the need to fill up the space left unoccupied by the excised thymines. Differently from seq+3 the pairing between the central non-damaged base pairs (T11...A36 and G10...C37) rapidly experience a disruption at  $\sim 34$  and  $64$  ns ending up to the exclusion of C37 at  $\sim 280$  ns and T13 at  $\sim 320$  ns. This conformation constitutes a very distinctive feature of seq-3 and presents specifically complex features, such as the stabilization of the two ejected pyrimidines between 76 and 244 ns via a single HB and extended intra- and inter-strand  $\pi$ -stacking developing from T11 to C33 base. Not surprisingly, due to the complex structural rearrangements the duplex exhibits polymorphism as evidenced by the bending that can reach values up to  $75^\circ$ .

Further increasing the separation between the two AP sites leads to a situa-

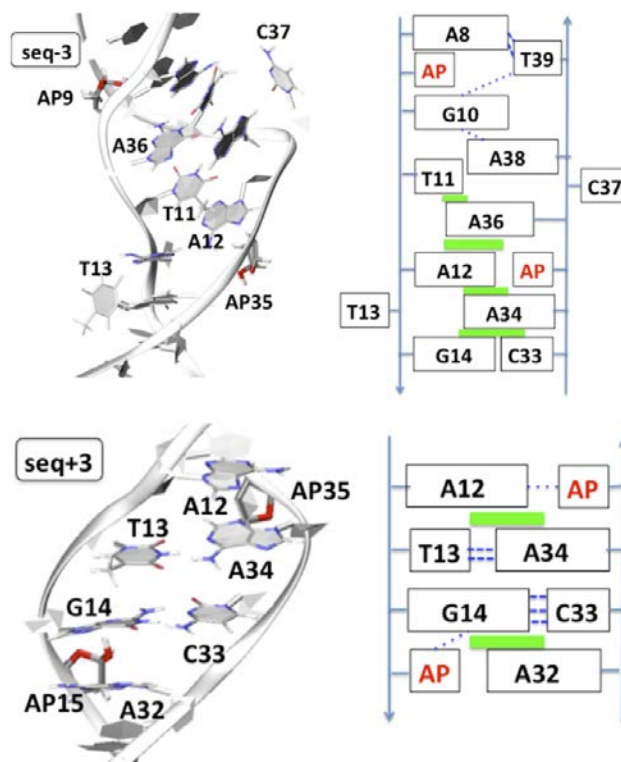


Figure 11.6: Cartoon representation of the final structure obtained after 400 ns for seq-3 (top) and seq+3 (bottom). A schematic representation is given at the right where green squares indicate  $\pi$ -stacking, and dashed blue lines new hydrogen bonds (thick ones for Watson-Crick pairing).

tion where they begin to behave independently and partially loose the correlation between the AP sites. The structure of seq-5 (Figure 11.7 left panel) follows the rather well established behavior of isolated AP sites although we observe an extrahelical occupation of  $\sim 12.5\%$ , i.e. a noticeably lower extrahelicity than for isolated AP sites. The AP sugar on the other hand explores a broad range of sugar pucker angles giving rise to a well established DNA bulge typical of AP strands[258]. AP35 exhibits many HB partners along the MD trajectory, yet mostly interacts with the oxygen of A34. A similar interaction pattern is found for AP7, which interacts with its O5' despite the steric constraint it induces to the backbone. The stacking remains identical to the one of the reference structure, with no deviation for the orphan A12 and A40 and no interstrand stacking. The DNA bending is only weakly perturbed ( $25.2 \pm 13.4^\circ$ ) and assumes values close to the ones experienced by single AP sites.

The symmetrical sequence seq+5 (Figure 11.7 right panel) shares common overall features, with no marked reorganization such as the ones observed for closer AP sites. The extrahelical occupancy accounts for roughly 30% along our simulations, this value being extremely close to the one of the control double-strand. As for the case of seq-5 it also shows a distinctive bulge all along the trajectory due to the backbone deformation in correspondence to the two AP sites. The Watson-Crick

pairing of the four separating base pair is maintained all along the 400 ns trajectory, and the bending of the helix remains limited ( $27.3 \pm 13.6^\circ$ ).

## Discussion/Conclusion

AP sites can occur either as a result of regular oxidative stress or ionizing radiation, or as an intermediate product resulting from the removal of an oxidized base lesion by a DNA glycosylase. Unrepaired AP sites are highly toxic and can carry mutagenic potential since they are non-coding and they can result in transversions due to the erroneous incorporation of DNA bases opposite to the AP site in replication or repair[262, 263]. Whereas the occurrence of one isolated AP site has been more widely studied and is now well characterized, quantifying bistranded clustered DNA damage induction and repair turns out to be more difficult. In addition, it is very important to address the relationship between clustered AP sites and DSBs. Accumulating experimental but also theoretical evidences, suggest that, the processing of a DSB can be impeded or inhibited by the presence of non-DSB neighboring lesions and the same stands for the processing of an oxidized base or AP site due to the presence of breaks or non-DSB lesions[264–266]. Understanding the processing of an AP site within a cluster lesion is considered of major importance. Most notably, different in vitro repair studies have shown that processing of complex lesions follows a sequential established hierarchy in the case of an AP site(s) and an 8-oxodG or SSB etc. Specifically, it is suggested that the AP sites and the resulting SSB(s) are processed and repaired before any DNA glycosylase binds and excises the neighboring base lesions[245, 262, 266–268], thus leading to the so-called "DSB avoidance" and repair retardation of the lesions belonging to a cluster[241]. However, "DSB avoidance" is not universal and is mostly likely to occur for cluster composed of oxidative damaged bases it is not always happening in case of bistranded AP site or complex, uracil-containing lesions[269, 270]. Hence, elucidating at a molecular level the structural deformations induced on DNA by complex lesions patterns, as well as the interplay with repair enzymes recognition, is of clear importance and relevance.

To examine the effects of such complex damages on DNA structure, we have analyzed a series of cluster abasic lesions-containing oligonucleotides differing by the number of separating base pairs (up to four) and the strand orientation. We stress out the need for hundreds of ns dynamics to capture the inherent complex reorganization of the duplex and the multiple conformations spanned by these complex systems. In contrast with experimental techniques such as NMR, our simulations allow probing transient interactions such as weak hydrogen bonds, which may ultimately dictate the global structure rearrangements, as well as the conformational flexibility of the systems. We evidence the following general trends: (i) the duplex

evolves, as much as possible, to restore an ideal B-helicity with a global bending close to  $30^\circ$ , this aspect is also coherent with the NMR results recently found by Zalesak et al. on DNA strand featuring clustered AP and 8-oxoguanine damages[271]. However (ii) interstrand dispersive interactions take place and play a major role in stabilizing the AP sites (iii) the local constriction of the helix, i.e. a narrower minor groove, correlates with the number of separating base pairs, (iv) AP positions can be locked by strong hydrogen bonding, and (v), when separated by 4 base pairs, the AP sites behave almost independently.

Clusters of closely spaced oxidative DNA lesions present challenges to the cellular repair machinery, especially in the case of bacterial organisms[246, 247, 272]. As expected the closest the AP sites the more difficult the repair-enzyme processing with seq0 being the less efficiently repaired one. However, a stunning low repair ratio has been evidenced for some sequences such as seq-3 and to a lesser extent seq-1 suggesting the emergence of peculiar interactions and structural deformations strongly affecting the recognition patterns between damaged DNA strands and proteins. On the contrary, human endonuclease experiences a globally higher processing rate and a less pronounced dependency on the secondary AP position. Our results provide a set of conformations constituting a preliminary step to reveal the interactions of clustered AP sites with repair enzymes. Furthermore, clustered AP sites embrace a combinatorial biochemistry, and their complex structural, dynamical and thermodynamical[273, 274] properties may be strongly sequence dependent.

Human APE1 and its counterpart the bacterial enzyme Nfo exhibit a striking difference in their repair capacity of clustered AP sites. With the exception of only one particular sequence (seq+3), APE1 has a higher processing rate than the Nfo protein (Figure 11.1). This suggests different evolutionary pathways[275], which are probably rooted in different interaction patterns between the oligonucleotides and the repair proteins. As revealed by the crystal structures [276], APE1 has an extended contact area between the enzyme and the DNA, furthermore strong electrostatic interactions between positively-charged residues and the negatively-charged DNA backbone induces rather large structural deformations of the helix. This fact was also elegantly revealed by Beloglazova et al. [277] in their study of thermodynamic, kinetic and structural basis for recognition of AP site by APE1. Hence, in the case of APE1 specific interaction and DNA deformation dominates, and hence the repair ratio cannot be correlated with the duplex deformation. On the other hand interaction with Nfo[278] is much more local and, as a consequence, the DNA structure is almost not constrained by the protein. Hence, in the case of Nfo, the structural modifications of the DNA duplex will directly influence the stability of the protein/nucleic acid aggregate and hence will largely correlate with the observed process efficiency. However, one should be aware of the fact that the process rate



depends on many complex factors, comprising the extrusion of the damaged base, its accommodation in the active pocket and the subsequent interaction developing with adjacent bases, therefore an absolute quantitative correlation cannot be drawn on the basis of structural factors, only.

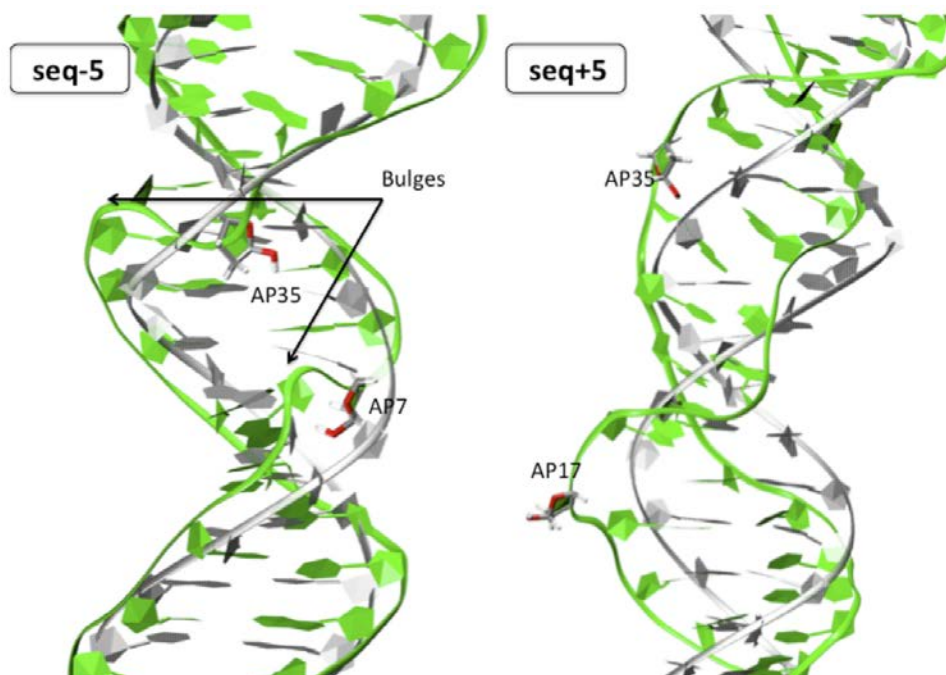


Figure 11.7: Cartoon representation of seq-5 (left) and seq+5 (right) after 400 ns of classical simulations. Both structures exhibit backbone deformation.

Consistently with previous observations and computational studies, our data clearly show that the distortion induced by a single AP site is not inducing a permanent gap or hole in the strongly coupled and organized double helix. Instead, and also taking advantage of the well-known flexibility of the DNA backbone, the macromolecule experiences a considerable local rearrangement to restore the maximal number of attractive non-covalent interactions. Two of the most paradigmatic modifications are the significant extrahelical position of the AP site and the induction of an important kink and bulge in the helix. Normally those two features are recognized as flags triggering the repair process[253].

Concerning the dynamic behavior of clustered AP sites, our results suggest a complex situation, associated with serious consequences on the efficiency of repair, especially in the case of Nfo. Furthermore, the induced structural deformations are strongly entangled with the specificity inherent to a single DNA sequence and hence cannot be captured by using only one simple descriptor. As an interesting example, and in agreement with previous NMR analysis[248], we found that the extrahelical behavior is strongly dependent on the specific clustered AP. Furthermore, while in seq0 and seq-3 the extruded AP sites occupy preferentially the minor groove, in the

case of  $\text{seq}\pm 1$  and  $\text{seq}\pm 5$  the major groove occupancy predominates,  $\text{seq}+3$  has a particular behavior with only one AP site preferentially occupying the minor groove. Consistently with the experimental observation[246, 247], we may safely conclude that the presence of four separating base pairs represents the limit under which two AP sites can no longer be considered as independent. Consequently, the repair rate for such well separated cluster sites is quite high and, even in the case of Nfo, close to 100% for  $\text{seq}+5$ , since no peculiar structural distortion hamper the recognition by the repair enzyme. Even if a general conclusion can be drawn, it is however important to note, that some sequence effects still exist, and  $\text{seq}-5$  has a slightly lower repair ratio, which we relate to a lower percentage of extrahelicity for both AP sites ( $\sim 12\%$ , versus  $\sim 30\%$  for  $\text{seq}+5$  and  $\sim 33\%$  for single AP sites).

$\text{Seq}0$  is the oligomer showing the less efficient repair by Nfo, yet the extrahelicity of the two AP sites is extremely high, close to 80%. The ejection of the AP sites is accompanied by a strong deformation of the DNA that consistently shrinks in order to reinstate  $\pi$ -stacking between the vicinal base pairs surrounding the two APs. This deformation is accompanied by a narrowing of the minor groove, in absence of any important kinking. Hence,  $\text{seq}0$  is characterized by a global structure significantly different from the one shown by a conventional AP site, thus precluding the recognition by the bacterial repair enzyme that indeed necessitates interaction with the minor groove.

Low extrahelicities and important local deformations characterize the duplexes  $\text{seq}\pm 1$ . However, once again the  $\text{seq}-1$  is somehow peculiar since non-damaged nucleobases too assume extrahelical position. This fact, also reported based on NMR structures[248, 257, 279], once again imposes structural modifications in the groove region, i.e. in the regions developing interactions with the Nfo enzyme, hence the repair ratio is even lower than for the corresponding  $\text{seq}+1$ .

However, the most striking position dependence is the one happening between  $\text{seq}+3$  and  $\text{seq}-3$ . Indeed,  $\text{seq}+3$  is almost perfectly repaired by Nfo and its efficiency is comparable with  $\text{seq}-5$ . Interestingly enough, all the structural parameters, particularly the extrahelicity and the kink, are extremely similar for the two cases. On the other hand the strong structural deformation, evidenced for  $\text{seq}-3$  with the ejection of non-damaged bases, and, the development of an extended interstrand stacking between different base pairs, result in a very low repair ratio, which could be the consequence of a weak interaction of the altered DNA with Nfo.

The strong asymmetry in Nfo repair efficiency observed experimentally is dictated by the various local environments experienced by the AP sites and the nearby nucleobases. In particular, an extremely complex pattern of non-covalent interactions between the double-helix constituents comes into play, going up to the development of transverse stacking. Even if it is not the sole parameter playing an

important role, extrahelicity propensity and conversely the presence of hydrogen bonds or transverse stacking locking the AP site, is certainly strongly correlated to the Nfo repair ratio of a given cluster lesion. While the local embedding and structural deformations around a cluster site are fundamental to understand its repair propensity, global deformations such as DNA bend angle do not have any strong influence. All these effects ultimately sum up to give a clear relation between the Nfo repair efficiency and the separation of the AP sites.

### 11.1.1 Recognition by APE1

As presented hereinbefore the low processing rate of APE1 for seq0 and especially seq-3 (Figure 11.2, Figure 11.1) is characterized by a strong local deformation with shrinking of the DNA duplex for seq0 and a complicated stabilization pattern involving the ejection of non-damaged purines for seq-3[280]. This inference is also justified by the observation that the bacterial enzyme contact with DNA is quite local (Figure 11.1) and, as a consequence, is not able to induce strong structural deformation and mechanical constraints on the DNA duplex.

A molecular tweezers, capped at its end respectively by one arginine and one methionine residues (Figure 1.4), stabilizes further the DNA-protein complex. In the case of the isolated enzyme (Figure 11.8a), this molecular tweezers is open and the distance between the terminal amino acids fluctuates dramatically around 12 Å, hence leaving an easy access for the DNA strand to reach the active site. In stark contrast, when DNA is present, the molecular tweezers closes around the oligomer, the distance between the terminal edges drops down to 7 Å and the amplitude of the oscillations strongly diminishes. The closure of the tweezers upon complexation with DNA is once again driven by the interaction of positive arginine with the negative backbone; upon equilibrium, the tweezers is kept closed also thanks to the interaction between the electron rich sulfur of methionine and the arginine forming the tweezers' tips.

The behavior of APE1 enzyme free and in interaction with the single AP-containing DNA strand is depicted in Figure 11.8. The DNA/APE1 contact region is quite extended (see Figure 11.1) and is constituted chiefly of basic arginine and lysine residues. The positively charged amino acids provide an extensive electrostatic stabilization of the negatively charged DNA backbone. As a consequence the root-mean-square deviation (RMSD) of the DNA strand with respect to the crystallographic structure remains low throughout the trajectory. The stabilization of the DNA structure is mainly due to the electrostatic interactions between the negatively charged DNA backbone and the positively charged amino acids constituting the recognition region. Furthermore, DNA structure and position are also locked by

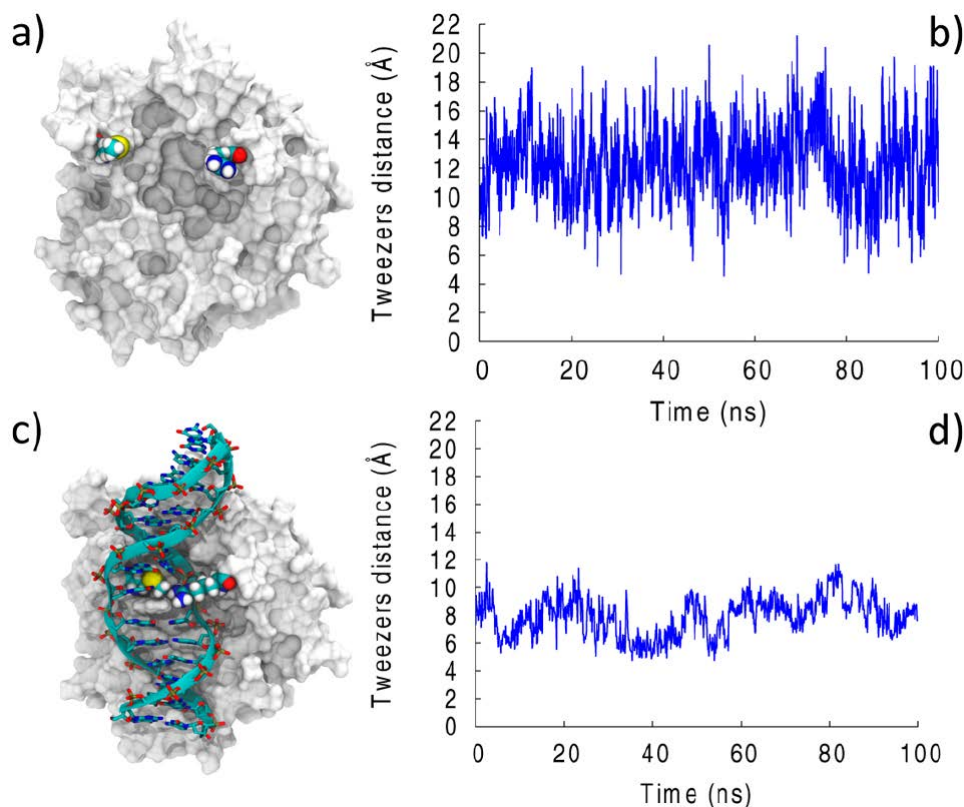


Figure 11.8: Representative snapshot and time evolution of the molecular tweezers width for the free APE1 (a,b) and for the complex between APE1 and the single AP containing DNA strand (c,d).

the closure of the molecular tweezers that hence minimize structural deformations.

When introducing a second lesion to form clustered AP lesions, we have observed in all the models simulated in this study the formation of persistent complexes (representative conformations are reported in Figure 1.5 and Figure 11.9) spanning all 100 ns. This is congruent with the relative high repair rate measured for human APE1. As already observed in all cases studied here, the APE1 enzyme induces strong geometrical constraints on the DNA oligomer, and indeed, we may see that its RMSD with respect to the reference crystalline form remains low and close to that found for the isolated AP-containing strand. In particular, the RMSD is much lower than that for the solvated strands, which are known to exhibit a strong structural deformation (Figure 11.10). As already evidenced in the case of isolated nucleotides[280], the deviation in the RMSD is not only due to the AP and it involves the complementary base and the nearby residues, up to at least five base pairs distance. The case of seq0, that is, the sequence exhibiting the larger structural deformation in solution, is particularly striking. This fact may also explains why seq0 is highly repaired by APE1 thanks to the imposed geometrical constrains, whereas its larger structural deformation probably makes its recognition by Nfo particularly difficult.

The only exception to this general trend is observed for seq+3 for which the

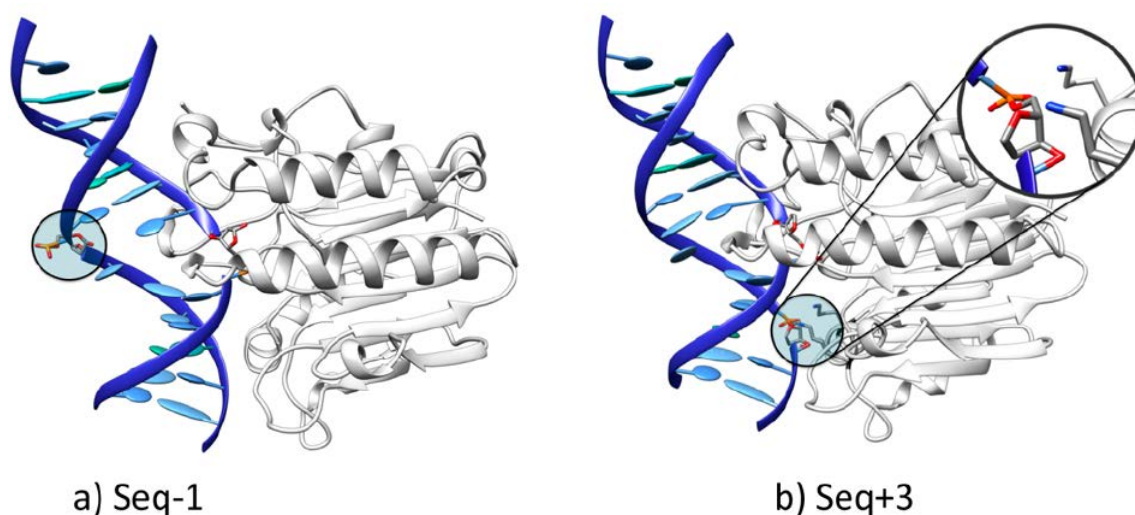


Figure 11.9: Representative snapshots for the seq-1 (a) and seq+3 (b) DNA/APE1 complex. DNA and protein are represented in carton. AP sites are evidenced in licorice representation. The interaction between the secondary AP site of seq+3 and the recognition region of APE1 is evidenced in the zoom of panel b.

RMSD of the DNA/APE1 complex almost overlaps with that of the DNA strand in solution, hence showing a much larger deviation from the ideal crystal structure. Interestingly enough, seq+3 is the only sequence for which the processing rate is lower than 50% and for which repair resistance can be correlated to a worse recognition by APE1, resulting in a less stabilized complex. The molecular bases for poor recognition can be related to the fact that seq+3 is the only sequence in which the second AP site is in direct interaction with the APE1 recognition region (Figure 11.9b), whereas in all the other cases, the secondary lesions are exposed toward the bulk. Indeed, the AP site having a strong propensity to adopt extrahelical position, it strongly perturbs the DNA/APE1 complex when interacting with seq+3. Another evidence of this perturbation can be seen in the behavior of the molecular tweezers (Figure 11.11), as mirrored by the representative conformations and time evolution of the edge distance for seq+3. One may clearly observe a partial opening of the molecular tweezers in the course of the MD trajectory, which may once again safely conclude in favor of a less stable arrangement.

The high efficiency of APE1 in processing clustered DNA lesions compared to bacterial endonucleases is clearly related to its capacity to induce strong mechanical constraints on the damaged DNA strands, and hence, to compensate for the large structural deviation engendered by these complex lesions. In particular, the presence of an extended positively charged region interacting with the DNA backbone and a tweezers closing around the DNA strand is crucial to ensure an adequate recognition and repair rate. Certain sequences (seq+3) exhibit a dramatically lower repair rate due to the perturbation induced by the secondary lesions on the recognition region,



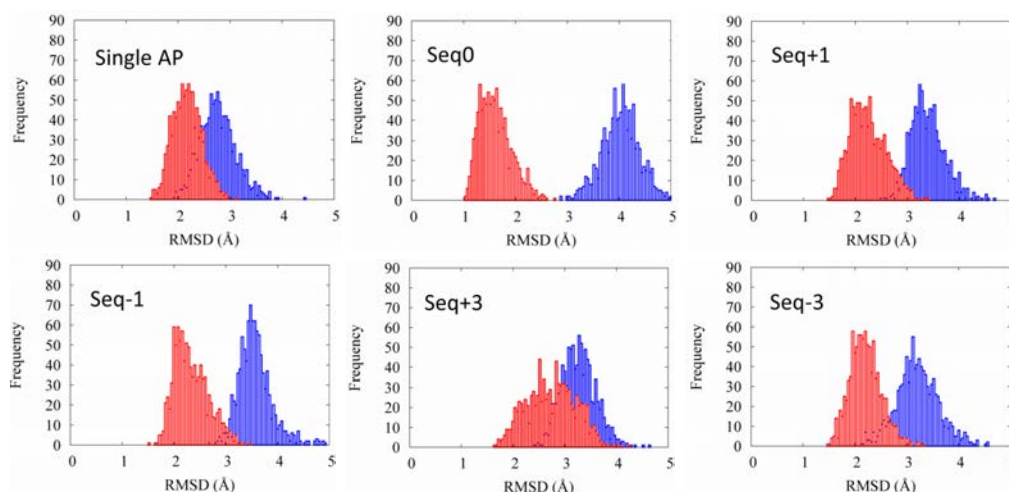


Figure 11.10: Distribution of the DNA root-mean-square deviation (RMSD) with respect to the crystal structure for the single containing and clustered AP lesions, inside the APE1 complex (red) and in solution (blue). The corresponding repair rates are given in Figure 11.1.

and in particular, on the stability of the closed molecular tweezers. The processing of AP clustered DNA is of high biological importance also because all DNA base lesions will be eventually converted to AP sites during reparation. The accumulation of complex AP DNA sites with increased processing efficiency by APE1 may in one hand reduce toxic AP lesions but on the other hand can lead to increased possibility of de novo DSB formation. For example, when two neighboring AP sites (<15-20 base-pair distance) are processed simultaneously by AP endonuclease, the occurrence of mutations and genomic instability may be increased through the generation of additional DSBs. Molecular modeling and simulations are complementary to the experimental studies to ultimately achieve the complete rationalization of the in vivo repair mechanisms of clustered DNA lesions, in particular focusing on the molecular basis of repair selectivity. Hence, the present investigation also paves the way not only to improve our understanding of DNA repair mechanisms but also to potentially design selective novel APE1 inhibitors, whose present efficiency is controversial[281, 282], ultimately being able to enhance radiotherapy efficiency.

## 11.2 64-PP and CPD

### Introduction

Even if the constitutive parts of DNA shows many protective pattern against UV irradiation 2.4, the appearance of photodamages such as cyclobutane pyrimidine dimers (CPD) and/or pyrimidine(6-4)pyrimidone 64-PP (Figure 11.12)[283], which are carcinogenic lesions, cannot be ruled out and must be studied to understand

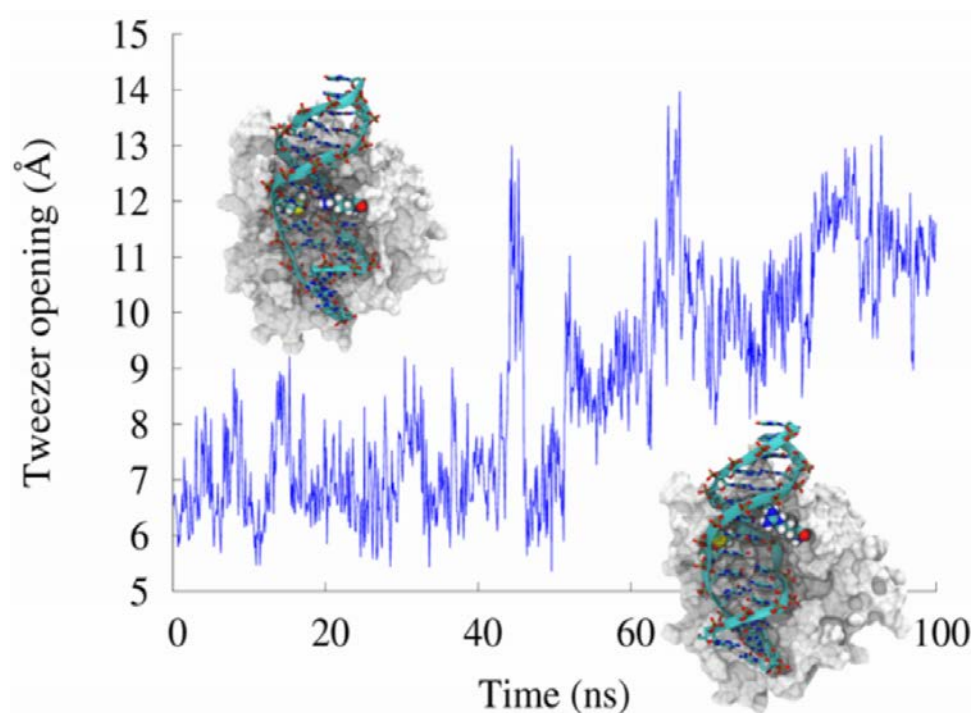


Figure 11.11: Representative conformations and time evolution of the molecular tweezers width for seq+3 showing the partial opening and the loss of stabilizing interactions.

their biological impacts. The possible action of 64-PP as a Trojan horse has also been analyzed in Chapter 10.4.

The ratio of CPD/64-PP in DNA upon UVB irradiation is approximately 80/20% [284, 285]. Indeed, 6-4PP formation requires the population of higher excited states. Moreover, both lesions are known to be repaired through the NER mechanism [286], but the rationalization of the molecular mechanisms and recognition of the damaged DNA strand [287], at the structural but also dynamical level, will help understand their repair rates [280, 288–290] and in particular the low efficiency of CPD [291] repair compared to 64-PP [284].

So far, structures of DNA double strands containing 64-PP has been obtained with X-ray or NMR techniques, especially complexed with enzymes [291], free in water [292–294] and even coiled around an histone [290]. The use of Förster Resonant Energy Transfer (FRET) also gave insights on the DNA bending, in particular providing the end-to-end distance of the strand [295]. These last results are in disagreement with the ones from X-ray and NMR cited previously. Indeed, FRET suggest a low bending compared with canonical strands while the others techniques described values up to  $44^\circ$ . Even if the interaction with histones can induce strong deformations of the global DNA structure, the result of free strands in water has to be rationalized using simulations and molecular modeling. Yet, MD has already been employed on solvated double strands and complexed with enzymes [284, 290,



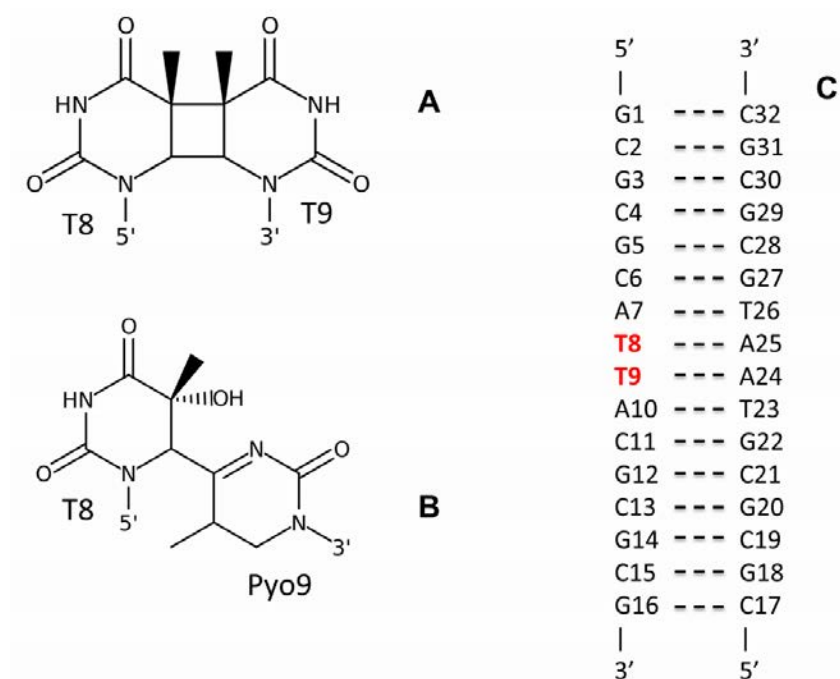


Figure 11.12: Molecular formula of CPD (A) and 64-PP (B). The 16 bp DNA double strand is also schematically reported with the bases' numbering (C). The two adjacent thymine at positions 8 and 9 (boldfaced in red) are the ones dimerizing to form the photolesions. In the case of 64-PP, the ninth nucleobase (T9) will be denoted Pyo9, being transformed into a pyrimidone unit.

[291, 296]. A study realized by Kollman et al. [296] resulted in the observation of a low bending for 64-PP, in agreement with FRET experiments, but the simulation time was not sufficient to observe the ensemble of conformational polymorphisms of the flexible DNA ensemble. In the case of CPD, experiments[297] and theory[298] already agreed on the rather low DNA bending. But still, values up to 30° have been reported<sup>34</sup> thus proving the necessity for more thorough investigations. Moreover, the repair rate of such photolesions is highly influenced by both the DNA deformation which impacts the protein recognition [287] and its flexibility [299].

Here we employed  $\mu$ s timescale MD simulations on two DNA double strands containing either CPD or 64-PP (Figure 11.12). Compared with the canonical B-DNA behavior, we provided insights on the repair rates of both photolesions based on the flexibility, polymorphisms or stiffness of each case.

## Computational details

For all the subsequent study we consider a 16-bp DNA strand (see Figure 11.12); the choice of a longer oligonucleotide instead of the 10-bp resolved by Lee et al., has been done to avoid instability at the oligomer terminal bases. Starting from the same oligomer the lesion was produced by dimerization of the thymine at position

8 and 9, to obtain one CPD or a 64-PP, respectively. All atom MD simulations were performed with Amber16 (45) and NAMD [169] codes for the three systems, B-DNA, CPD and 64-PP. DNA nucleobases were represented by the parmbsc0 force field [133, 167]. The performance of the parmbsc0 force field has also been tested against the recently developed bsc1 [300] corrections and shown to perform analogously. In particular over the 150 ns dynamics we observed similar transitions between conformers as the ones described with parmbsc0 force field. The parameters for the CPD and 64-PP lesion were assigned according to the standard RESP protocol [17]. We note that they are in very good agreement with the ones derived by Kollman et al. [296]. Water molecules were represented by TIP3P model and 30 potassium cations were added to the simulation box to assure neutrality.

After thermalization and equilibration production MD has been performed in the constant pressure and temperature (NPT) ensemble (300 K, 1 atm). In the case of 64-PP a 2  $\mu$ s trajectory has been performed in order to properly explore the conformational space evidencing the dynamical equilibrium between different conformers. In order to characterize the global DNA behavior the DNA bending angle was calculated along the three trajectories with Curves+ [79]. The compactness of the structures was assessed by the calculation of the solvent accessible surface area (SASA) [301] obtained post-processing the MD trajectories with the measure command of VMD [177] and a probe radius of 1.4 Å. The local deformation was further explored by analyzing the extent of  $\pi$ -stacking between the nucleobases adjacent to the lesion as obtained by the distance between the aromatic rings.

## Results

In a first step we assess the structural distortion of the 16-bp duplex containing a photolesion through two global parameters namely the DNA helical bending, and the SASA. Helical bending in particular also allows for a straightforward comparison with experimental results. In Figure 11.13, we report the distributions of the bending angle as calculated by Curves+ (50) in the case of the native B-DNA and in presence of CPD or 64-PP. CPD bending distribution almost perfectly overlaps the one of the control double-stranded sequence, with maximum values at around 20° and 18°, respectively. Instead, the situation is stunningly different for 64-PP (Figure 11.13C), where the bending is much more dispersed. Indeed, one can evidence two peaks: a first one centered at around 20°, overlapping with the control B-DNA sequence and the CPD containing duplex, and a second one appearing at larger values (40°). We also point the presence of a less populated yet clear shoulder at even higher bending angles.

SASA values, presented in Figure 11.13, globally show similar tendencies as for

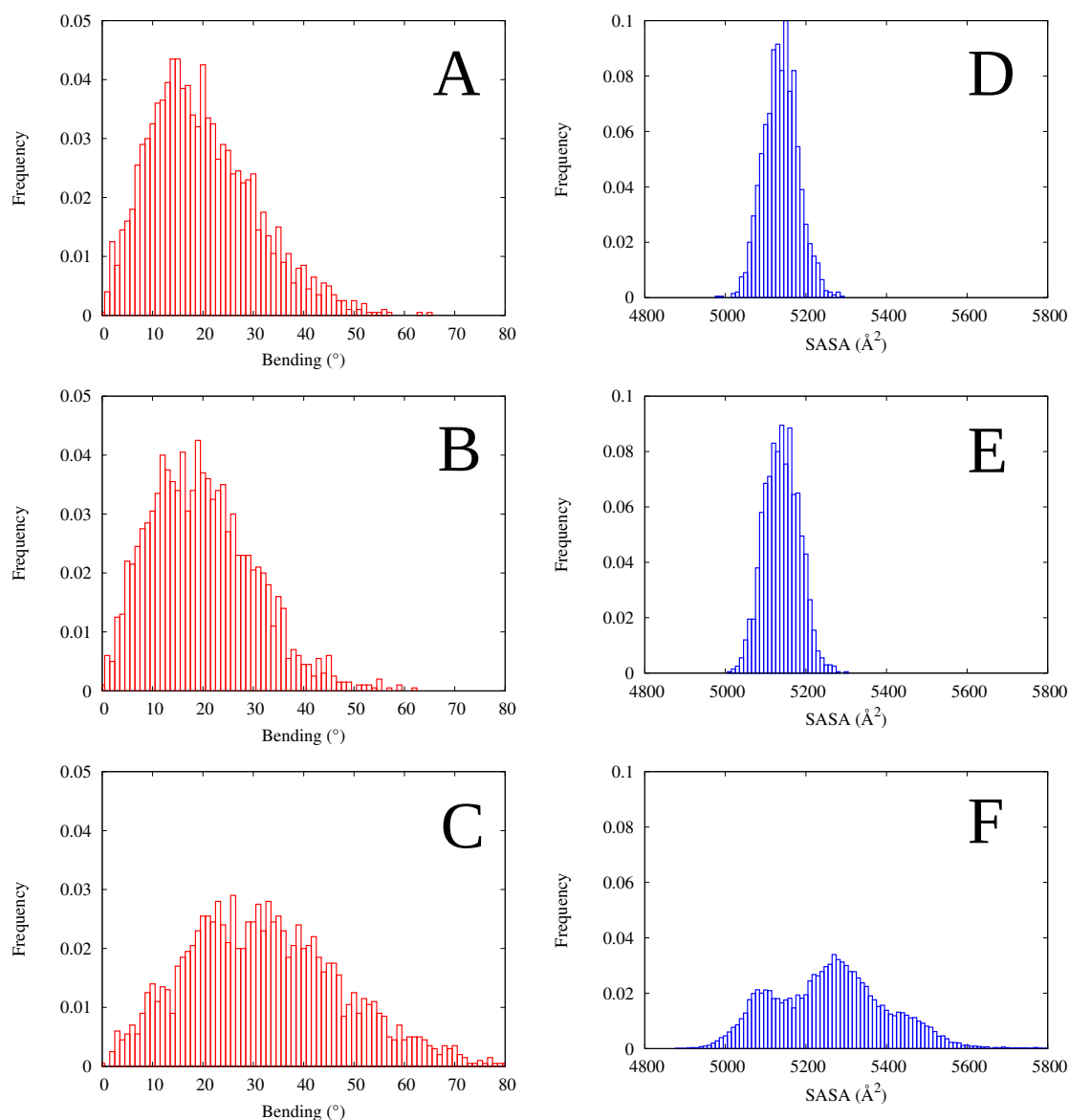


Figure 11.13: Distribution of the total DNA bending (in red) for B-DNA (A), CPD (B) and 64-PP (C); and of the SASA (in blue) for B-DNA (D), CPD (E) and 64-PP (F).

the bending and clearly points in favor of marked deformations only in the case of 64-PP. Once again CPD and B-DNA distributions almost perfectly overlaps both in terms of the maximum value and of the band shape. On the other hand, three maxima, two well resolved peaks and a shoulder, can be evidenced for 64-PP. The solvent accessibility values indicate that 64-PP spans an ensemble of conformations, some of them being quite compact (low SASA) and other presenting a large surface area (high SASA).

In Figure 11.14 these tendencies are exemplified by the comparison of the representative structures of B-DNA and CPD containing oligomers (that indeed appears indistinguishable), with two highly distorted 64-PP conformations characterized by a low and high bending, and by different compactness due to the formation of hollow

structures.

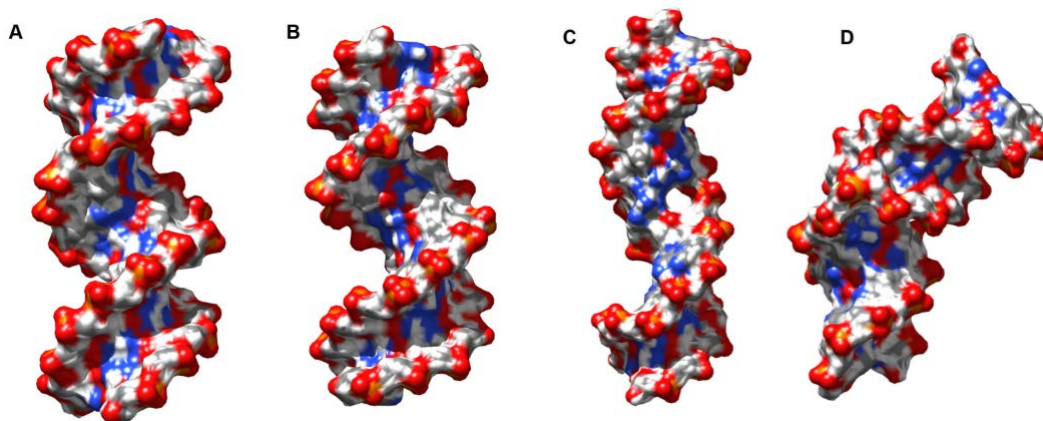


Figure 11.14: Representative snapshots rendered in surface mode for B-DNA (A), CPD (B) and 64-(PP) in a low bending high SASA conformation (C) and in a high bending low SASA one (D).

In order to assess the convergence of 64-PP simulation we also performed parallel tempered replica exchange [302] representing a total sampling time of about 5  $\mu$ s. The distribution of bending and SASA obtained for 64-PP are reported in Figure 11.15. Globally the enhanced sampling procedure confirms the previous results and in particular the polymorphism exhibited by 64-PP containing strand, in particular both high and low bending values are once again represented. The SASA distribution is also extremely broad and point toward the coexistence of different structures exhibiting different compactness, however the peak for low SASA, i.e. compact structure, is much less well resolved, indicating a lower population of the corresponding conformers. Indeed, the relative population of the different structures appears difficult to reproduce and would necessitate a prohibitively computational effort, nonetheless the global consideration over the dynamic equilibrium leading to polymorphism still holds.

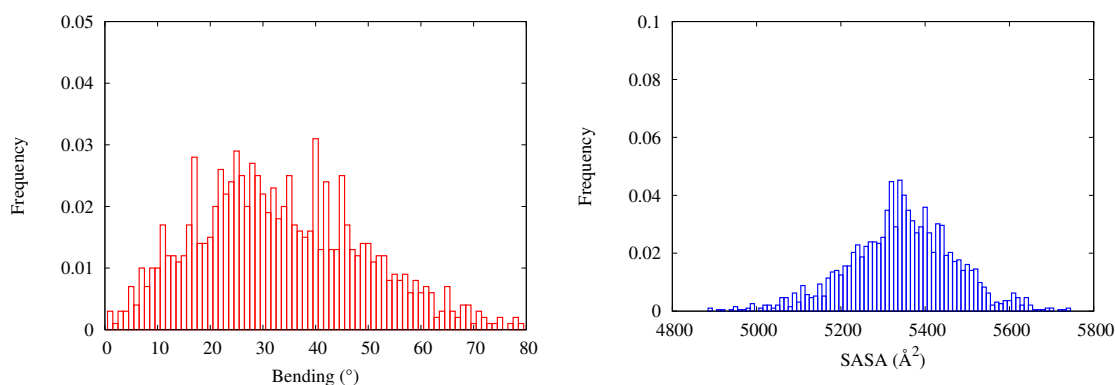


Figure 11.15: Distribution of the total DNA bending (red) and SASA (blue) obtained from replica exchange parallel tempering.

Hence, a first stunning difference between CPD and 64-PP can be unequivocally invoked, i.e. the structural invariance of the former compared to the important deformation, and the flexibility, induced by the latter. As indicated by the global descriptors 64-PP exhibits polymorphism and thus deserves a closer attention to clearly identify the complex behavior induced by this lesion. Figure 11.16 provides the time evolution of the distance between the nucleobases in the vicinity of the 64-PP lesion (Figure J.1). This parameter has been chosen since it allows to quantify the extent of the formation of  $\pi$ -stacking interactions. Along the 2  $\mu$ s of the trajectory 64-PP explores a rather large conformational space with a complex equilibrium between different conformations, characterized by different  $\pi$ -stacking interaction patterns between the nucleobases. Quite importantly, the equilibrium is extremely dynamic, the stable conformations are separated by some transition regions, and the conformers can have multiple occurrences during the time series. In partial contrast to the global parameters (bending and SASA) the use of a more local descriptor clearly indicates the presence of at least five distinct conformations. However, the use of unconstrained MD does not allow to estimate the absolute population rates between the conformers nor to a priori exclude the existence of other possible stable structures.

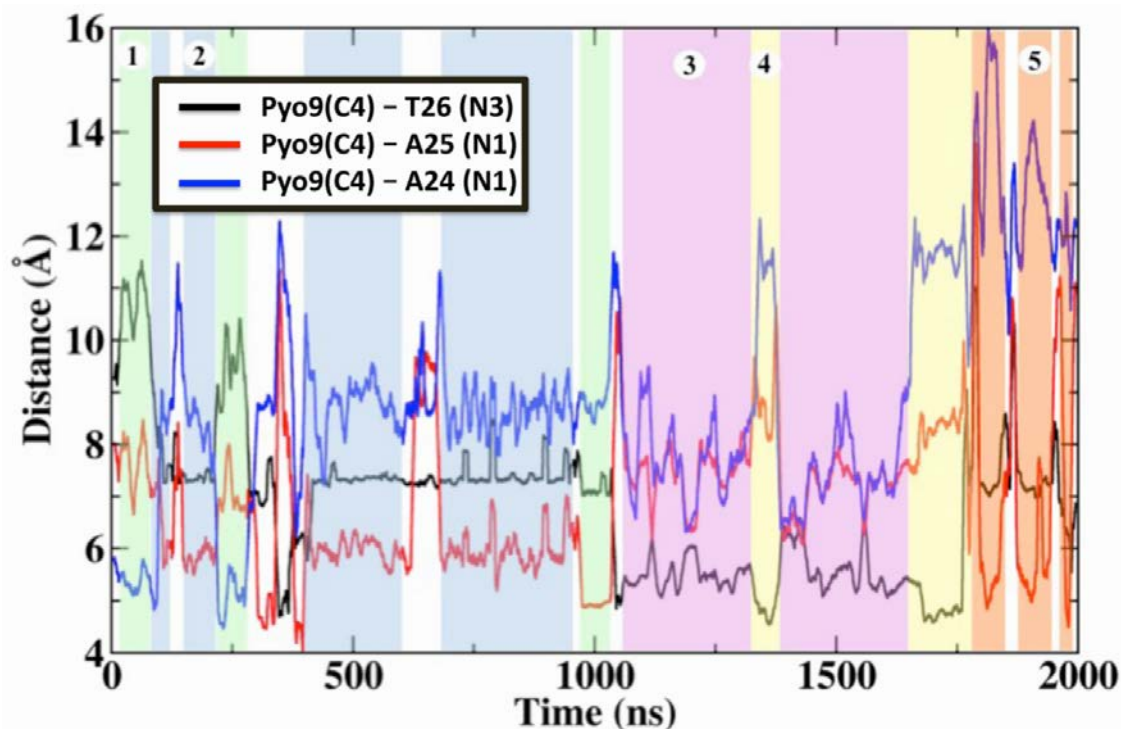


Figure 11.16: Time evolution of internucleobases distances measured between Pyo9 and A24, A25 and T26, characterizing the different  $\pi$ -stacking formations and hence the increased flexibility and polymorphism upon formation of 64-PP. The regions corresponding to different conformations are indicated by colored background.

Figure 11.17 provides representative snapshots of all the five stable equilibrium



conformations together with schematic diagrams depicting the interactions taking place and stabilizing the specific conformations. It is important to stress out the considerable structural differences as well as the high variability of the coupling between the different nucleobases. Hence, one can safely conclude that 64-PP does not present a well-defined unique structure but rather a complex equilibrium between different conformations whose structure and properties strongly differ.

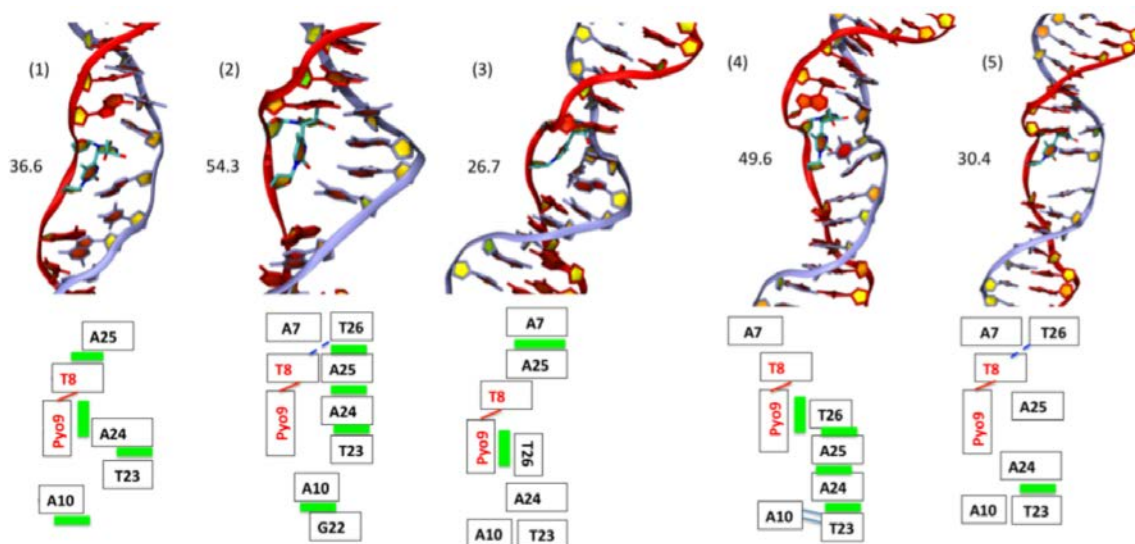


Figure 11.17: Cartoon representations of five representative conformation observed along the 2  $\mu$ s trajectory. The schematic depiction of the interaction patterns is given in the bottom figures for each conformer, where green boxes represent  $\pi$ -stacking and blue lines hydrogen bonds. The value of the bending angle for each conformer is indicated in parenthesis.

We seek to identify the rather complex reorganization pattern induced by 64-PP and the non-covalent interactions that come into play to stabilize the duplex after the induction of the covalent-linkage. Cartoon representations are given in Figure 11.17, together with schemes depicting the spatial arrangement of the nucleobases: hereafter the thymine at position 9 will be indicated as Pyo9 since it is the one forming the pyrimidone moiety (Figure 11.12). Indeed, the large conformational landscape explored at the  $\mu$ s range arises from the fact the duplex can rearrange to favor interstrand dispersive interactions (green boxes in Figure 11.17) and hydrogen bonding with offset nucleobases (dashed blue line), which in some cases can lead to the exclusion of vicinal nucleobases.

Conf-(1) maintains a spatial proximity between the damaged T8 and the originally complementary adenine A24, at a distance of  $\sim 4\text{\AA}$  which actually corresponds to the  $\pi$ -stacking. A24 can indeed no longer be Watson-Crick paired with Pyo9 and instead the helix readapts by a local shrinkage to afford a stacking between Pyo9 and A24, the latter behaving as an orphan nucleobase.

The duplex then evolves toward Conf-(2) which is stabilized by conserving an extended stacking on the 64-PP complementary strand involving thymines and adenines (from T23 up to T26) while interstrand dispersive interactions are no more present. Interestingly, this conformation corresponds to a maximal curvature, as verified by the bend angle ( $54.3^\circ$ ).

After 1  $\mu s$ , the duplex rearranges to adopt a more straight and compact conformation Conf-(3) with interstrand  $\pi$ -stacking between Pyo9 and T26, but also between A7 and A25. These interactions narrows the B-helix and contributes to decrease the bend ( $26.7^\circ$ ), whereas no new hydrogen bonds comes into play to stabilize the damaged duplex.

At around 1.8  $\mu s$  a conformation rather similar to Conf-(2), that we will call Conf-(4), is found to be characterized by a proximity between the damaged Pyo9 and T26. Also this conformation presents an extended stacking on the lesion's complementary strand going from T23 to T26 and a high bending ( $49.6^\circ$ ). However, differently from Conf-(2), the stacking region is shifted downward.

Oppositely, Conf-(5) is the first structure exhibiting a gap between A24 and A25, which lies at about 7 Å, while T8 gets close to the A7-T26 base pair still locked in the Watson-Crick pairing. Ultimately, the alcohol function of T8 interacts with the oxygen O2 of T26 to form a stable hydrogen-bond. Conf-(5) duplex presents a globally low bending just slightly higher than the average value for the control B-DNA sequence.

The tendency of A24 to seek opportunistic interactions to compensate for the lack of pairing with the complementary thymine is a constant all along our dynamics. The same behavior is also experienced by Pyo9 that in particular seeks to adopt conformations allowing dispersive interactions with A25. The coupled perturbation of the arrangements of the two strands propagates up to A10 and G22 which may develop non-conventional  $\pi$ -stackings interaction, while the Watson Crick pairing with T23 and C11 is maintained. Conf-(2) and Conf-(4) are both characterized by large bend angles of  $\sim 50^\circ$ , yet they differ markedly by the local structural rearrangement. Indeed, Conf-(4) is pushing A7 to an extrahelical position allowing to maintain C6...G23 and A10...T23 pairings. Note also that, Conf-(4) and Conf-(5) clearly exhibit a hole structure in correspondence of the lesions, leading to high values of SASA; the same behavior, even if less pronounced, is also be evidenced for Conf-(2). Correspondingly, the complex nucleobase rearrangement, necessary to accommodate the intrastrand cross-links created between Pyo9 and T8 leads to marked variations of the total bend angle. However, in all the cases, bulges and helical distortion in the region corresponding to the lesions are clearly identified. The existence of rather large structural deformations and a large reorganization of the helical parameters, in particular in the vicinity of the lesion, are also confirmed



by the time-series and the distribution of some relevant DNA structural parameters extracted with Curves+ [79], such as the minor and major groove width and depth (Figure J.2). Indeed the base-pairs surrounding the lesions (from 7 to 10) appears as the most influenced and exhibit a strong variation correlating well with the distribution of the distances previously analyzed. Most notably, one can observe a general shrinking of the major groove especially evident for Conf-(2) and Conf-(3) while Conf-(1), Conf-(4), and especially Conf-(5) are characterized by a slight enlargement of the major groove width. Minor groove, on the contrary, appears less influenced and correlation extracted from this parameters less straightforward, even though the value of the groove depth in the correspondence of the lesion peaks to zero indicating an important shrinking of the strand.

## Discussion and conclusion

The behavior of the two most common DNA photoproducts, CPD and 64-PP, has been fully characterized by long timescale, all-atom simulations. In particular, it has been clearly evidenced that CPD induces a minimal structural deformation. The CPD containing duplex is indeed absolutely comparable to the native B-DNA, both in terms of bending and SASA distribution. In addition, only one stable conformation is observed and all the local and global structural parameters are much similar to the ones of the undamaged strand. CPD is one of the only lesions in B-DNA implying the formation of two covalent bonds tethering the adjacent thymines: rather logically, this confers more rigidity and locks the structure, decreasing the rotational freedom.

On the other hand, the situation is totally reversed in the case of 64-PP. Indeed, 64-PP is characterized by a dramatic structural deformation centered around the lesion and propagating to a large part of the duplex. At least three maxima can be evidenced in the distributions of both SASA and bending angle, pointing toward the coexistence of compact and surface exposed structures, as well as bent and straight arrangements.

Our simulations unify the complex and somewhat conflicting panorama of experimentally based 64-PP, and CPD structural resolution. Indeed, while the general CPD low bending is recovered, as shown by the most probable angle, the distribution width clearly indicates that values close to  $30^\circ$  cannot be ruled out and can occur at room temperature. On the other hand, in the case of 64-PP our results show that both NMR [294] and FRET [295], i.e. high and low bending, are possible due to its polymorphisms and depending on the specific conformations. Indeed, I also report in appendix the distribution of the end-to-end distance of the DNA double strand we observe the presence of two peaks at distance of 55 or 60 Å that correlates well

with the values reported in [295] for strands having the same number of base pairs. The NMR predicted bending value of around  $40^\circ$  corresponds to one of the maximum of the bending distribution, on the other hand the FRET results are consistent with the second maximum appearing at lower angle and leading to bending comparable to the ones of native B-DNA.

Recently DNA polymorphism also in the case of non-damaged strands has been proved for instance by circular dichroism techniques [48] or by molecular simulation [303], it has also been correlated to the insurgence of different pathologies in particular when involving gene regulatory areas such as CpG islands [16] or telomeres [304, 305].

As it is clearly shown by our simulations, 64-PP is actually experiencing structural polymorphism that ends up in a dynamic equilibrium between different conformations whose local structural parameters differ significantly. Such polymorphism is definitively striking also when compared to the more predictable structural rearrangements produced by oxidatively-induced intrastrand cross-link lesions [122, 253]. The former lesions rearrange to restore B-helicity and as a consequence they lead to less extreme bending values and a globally more rigid structure.

The observed contrasting structural behavior may indeed be correlated to the repair and replication rate of the two lesions, and hence their toxicity. Indeed, CPD extremely low repair rate may be seen as a direct consequence of the absence of relevant structural modifications compared to B-DNA. This fact indeed, hampers the lesion's recognition since a common recognition pattern is the structural deformation induced by the damage [299]. In a simplified way, one may say that CPD actually masks itself and hides between the undamaged B-DNA, hence escaping recognition and repair. However, its rigidity also constitutes a blockage to the replication process that ultimately results in the limited mutagenic power especially as compared to 64-PP [306].

64-PP on the other hand adopts a totally different strategy, the large structural deformation experienced by this lesion is indeed consistent with the mobilization of the NER repair machinery and the lesions is processed much more efficiently than CPD [307]. However, due to the observed polymorphism no single stable structure can be evidenced, and the lesion is constantly shifting back and forth between different conformations, and as a consequence the recognition by enzymes can be made more difficult. This is particular true since some of the conformers present very different local deformations and interaction patterns, as well as different global parameters such as bending and surface accessibility. Furthermore, the flexibility and polymorphisms of 64-PP also correlates with the absence of unyielding obstruction to the DNA replication process [308] leading to the accumulation of mutations at the photolesion spot. Indeed, the easy replication of 64-PP damaged strand is

recognized as one of the reason of 64-PP high toxicity and mutagenicity [309], and the observed polymorphism and the different nucleobase interaction patterns may also explain the large spectrum of observed mutations.

It is however important to underline that we may anticipate a different behavior of 64-PP when complexed with histones in a nucleosomal environment, as surmised by Akihisa et al. [290]. In particular the constraints due to interacting proteins and the compact nucleosomal environment may restrict the accessibility of the explored conformational landscape, still the influence of flexibility on the replication obstruction will hold.

## Part IX

### Conclusion générale

Grâce à des collaborations efficaces, cette thèse a été fructueuse en terme de sujets qui ont été abordés et d'avancées qui ont été réalisées dans le domaine de la photosensibilisation de l'ADN et plus globalement les propriétés de l'ADN. Rapidement, les sujets qui ont été décrits dans le corps de ce manuscrit, commençant par les propriétés photochimiques et photophysiques des agents photosensibilisants en contact avec l'ADN ainsi que leurs énergies libres d'interactions, passant par le développement d'un outil pour la modélisation du dichroïsme circulaire, jusqu'à l'influence des lésions sur la structure et la flexibilité de l'ADN et enfin leurs reconnaissance par des enzymes. Une description plus détaillée des résultats principaux, ci-dessous, offre une vue d'ensemble de ces études.

- Pour commencer, une avancée importante a été réalisée au travers du développement d'un outil capable de modéliser le signal de dichroïsme circulaire électronique (ECD) de biomolécules. Il a été utilisé avec succès sur des doubles brins d'ADN dans sa forme B ainsi que sur des G-quadruplexes. Cet outil permet d'assister efficacement l'étude de la structure d'ensembles d'acides aminés à partir de la reproduction de leurs données spectroscopiques.

*Dans le future nous souhaiterions aller plus loin que la seule étude d'acides nucléiques. En effet, sachant que l'interaction de molécules avec l'ADN, mais aussi avec les protéines provoque la modification de leurs spectres ECD, il sera intéressant de comprendre et valider des configurations d'interactions simplement par la comparaison de nos résultats avec des données expérimentales. Ensuite, il sera possible de s'attaquer à des systèmes toujours plus complexes tels que les ensembles d'acides aminés, ce qui permettra d'obtenir de nouvelles perspectives dans l'étude des structures de protéines.*

- En corrélation avec le point précédent, a été développé un protocole qui permet d'obtenir et ainsi comparer directement l'énergie libre d'interaction de la benzophenone dans l'ADN soit dans le petit sillon, soit en double insertion. En utilisant des simulations de dynamique moléculaire et des transformations alchimiques, il a été montré que l'interaction petit sillon est la plus probable en terme d'énergie libre avec une valeur absolue  $2 \text{ kcal/mol}^{-1}$  inférieure à l'interaction petit sillon. De plus, les résultats ont pu fournir une meilleure compréhension des mécanismes très complexes de photosensibilisation de l'ADN par la benzophenone. Ce protocole est d'un grand intérêt pour la recherche de composés plus spécifiques et efficaces pour les traitements médicaux tels que la photothérapie dynamique.

*Les études d'interaction ADN/molécules pourront maintenant être assistées par une première détermination des modes d'interactions les plus stables ce qui va améliorer et accélérer les investigations de mécanismes de photosensibilisation.*

- Ensuite, les barrières énergétiques des voies d'abstraction d'un atome d'hydrogène par la benzophenone dans son état triplet, provenant d'un sucre ou d'une base de l'ADN (dans ce cas une thymine) ont été déterminées. Les résultats ont principalement montrés que l'abstraction d'un hydrogène d'un sucre est légèrement plus favorable dans le cas d'une interaction petit sillon.

*De nombreuses questions sont toujours ouvertes quant aux propriétés photophysiques de la benzophenone en interaction avec l'ADN et il est nécessaire de pouvoir y répondre. Plus spécifiquement il sera intéressant d'utiliser d'autres techniques afin d'obtenir des données qui pourront corroborer avec les résultats expérimentaux, par exemple avec des potentiels de force moyenne pour obtenir les profils d'énergie libre des agrégations pour aller plus loin dans l'étude des voies de photosensibilisation de la benzophenone.*

- Aussi les possibles modes d'interaction des composés chimiques BMEMC, nile bleu et nile rouge ont été décrits. Lors de cette étude il a été mis en évidence que seul nile bleu est capable de générer des transferts d'électron depuis une guanine à proximité et seulement s'il est en interaction dans le petit sillon. De plus BMEMC a montré un comportement sensibilisant de par sa capacité à induire des lésions après avoir été excité par une absorption de deux photons et sans la présence d'oxygène. En effet, il a été démontré qu'en interaction avec l'ADN ou en solution dans une boîte de molécules d'eau explicites, BMEMC peut produire des électrons solvatés qui peuvent par la suite être délétères pour l'ADN.

*Sachant que nous avons acquis une expertise dans l'étude des interactions entre composés photosensibilisants et leurs possibles mécanismes photophysiques, il est envisageable de travailler au développement d'autres molécules, plus efficaces et plus spécifiques.*

- A l'aide de simulations de dynamique moléculaire approfondies, l'influence de trois importantes lésions de l'ADN a été étudié. Premièrement 6-4PP augmente fortement la flexibilité du double brin ce qui lui permet d'explorer un ensemble conformationnel très important sur des temps de simulations allant jusque 2  $\mu$ s. Dans le cas de CPD, la lésion a un comportement proche de celui des bases canoniques et la déformation du brin lésé n'est pas distinguable en comparaison avec l'ADN-B. Finalement, les sites abasiques et plus spécifiquement les clusters de sites abasiques ont des effets très différents sur la structure et le comportement de l'ADN en fonction de leurs positions relatives dans la séquence, tel que cela a été démontré expérimentalement.

*Basé sur les résultats importants de ces études, l'intérêt pour les lésions de l'ADN est tel qu'il semble nécessaire de comprendre de façon plus complète*

*l'influence de ces bases non-canoniques. Cela pourra être achevé en développant par exemple un jeu de paramètres pour chacun de ces composés qui servira à simplifier et accélérer leur étude.*

- Enfin, l'interaction de double brins lésé contenant des clusters de sites abasiques avec l'endonucléase humaine de réparation, APE1, a montré une corrélation claire entre le taux de réparation de ces lésions et la reconnaissance du brin. En effet un cas spécifique dans lequel les deux sites abasiques sont séparés par trois bases dans la séquence a montré une déstabilisation importante des points de contacts avec l'ADN ainsi corroborant la faible activité enzymatique dans le processus de réparation de ce genre de dommages.

*A partir de là, deux chemins peuvent être suivis. D'un côté l'étude du mécanisme réactionnel du clivage de l'encrage de la base dans le brin au niveau du groupement phosphate, et de l'autre une étude de brins contenant d'autres types de lésions en interaction avec leurs protéines de réparation respectives.*

Ce manuscript ouvre la voie à des études supplémentaires liées à la photosensibilisation de l'ADN. A partir de maintenant, chacun peut simplement suivre/utiliser les outils et protocoles développés dans le but d'apporter sa propre contribution à ce vaste sujet de recherche.



## Part X

### General conclusion

Thanks to a strong collaboration network, this thesis has been fruitful in terms of topics that have been tackled and advances that have been provided to the field of DNA photosensitization and more globally to the research on DNA properties. Indeed a quick overview of the work that has been realized, going from photophysical and photochemical properties of DNA interacting with photosensitizers and their binding free energies, through the study of ways to model the electronic circular dichroism of DNA, toward the influence of lesions in the DNA structure and flexibility and finally reaching the recognition by enzymes, a rather complete investigation related to my PhD project has been provided. A more detailed description of the main results, provided below, offers a complete overview of these studies.

- First of all, a step forward has been realized through the development of a tool to model the electronic circular dichroism of biomacromolecular systems. It has been successfully utilized to reproduce the ECD spectra of B-DNA double-strands and G-quadruplexes. This tool offer a new way to assist the study of DNA structure through spectroscopic information.

*In the future we ought to go beyond the sole study of nucleic acid ensemble alone. Indeed, since the interaction of DNA with molecules, proteins and more gives rise to modifications of its ECD spectra, we shall rationalize and validate binding configurations simply through a comparison with experimental data. Then, I wish to tackle more complex systems such as amino acids ensembles and open the door to a vast field of research that is protein structures.*

- Secondly, this first study is strongly correlated to the methodological development of a protocol that allowed to directly compare binding free energies of benzophenone in DNA, interacting through either minor groove binding or double insertion. Using molecular dynamic simulations and alchemical transformations, it has been showed that the first is the most probable with an energy difference of 2 kcal/mol<sup>-1</sup>. Moreover it provides new data on the understanding of the complex photosensitizing mechanisms of benzophenone. Such protocol is of interest in the research for new specific target compounds in medical treatments such as photodynamic therapy.

*Every further study of DNA/drug complexes can now be assisted by a a first investigation of the main binding modes of the molecule that shall simplify and speed up the investigation of their photosensitization mechanisms.*

- Still focusing on benzophenone, the energetic barriers and mechanisms of hydrogen abstraction by benzophenone in its triplet state from either a sugar of the DNA moiety or a base (in that case thymine) have been provided. The results mainly showed that the abstraction for the sugar is slightly more favorable.

*Numerous open questions on the photophysical properties of benzophenone and its interaction with DNA has still to be addressed. More specifically it will be interesting to employ more advanced techniques to correlate experimental data on the competitive photosensitization routes of benzophenone such as electron transfer for example.*

- Then, we also addressed the possible interaction modes and photosensitizing pathways of nileblue and nilered in DNA. In this work we evidenced that only the first is able to generate electron transfers from a nearby guanine, and more interestingly, only in minor groove binding. Moreover we also provided interesting insights on the action of BMEMC, a new type of DNA sensitizer that can induce lesions through two photon absorption and without the presence of molecular oxygen. It has been demonstrated that in interaction with DNA, BMEMC always has the possibility to produce solvated electrons that could subsequently be deleterious for DNA. Also from the analysis of the structure of the different interaction modes we hypothesized the hydrogen that could most probably be abstracted by the drug, leading either to a base excision or a strand break.

*Since we obtained an expertise in the interaction of molecules with DNA and their possible photophysical and photochemical mechanisms, we ought to employ it for the investigation of new compounds with higher TPA efficiency and more specific interactions.*

- Employing extensive molecular dynamic simulations, the influence of three main nucleobase lesions of DNA has been studied. First 6-4PP enhances the flexibility of DNA and induces strong deformations and highly distorted polymorphisms of DNA all along a 2  $\mu$ s simulation. In the case of CPD, the lesion is more or less invariant to canonical nucleobases and the global deformation of the damaged strand is barely distinguishable compared to B-DNA. Finally, we observed that abasic sites and more specifically clustered abasic sites have rather different influences on the DNA structure and behavior depending on their relative positions in the sequence as demonstrated experimentally.

*Based on the important results described above, the interest toward DNA damages made it evident that we have to go toward a stronger understanding of the influence of these non-canonical nucleobases. It might be achieved through the development of a database of parameters for each compounds that could speed up and simplify similar studies.*

- Finally, subsequently to the study of cluster abasic sites in DNA, we also showed that their effects on the interaction with an endonuclease repair en-

zyme, APE1, is in clear correlation with the experimental repair rate of this protein. Indeed, for one specific case in which the two lesions are three nucleotides apart, the experimental results have been matched showing that the contact points between the two supramolecular entities are destabilized thus leading to a lower repair activity.

*From here two main paths could be followed. On one hand the investigation of the reaction mechanisms that induces the cleavage of the lesions, and on the other hand the interaction of such damaged DNA double strand with other repair enzymes.*

This manuscript paves the way to further studies related to DNA photosensitization. From now on, anyone could follow and understand the presented tools, methodologies and protocols in order to achieve their own research investigations on topics related related to this global context.

# Part XI

## Appendices

# Appendix A

## DNA structural parameters

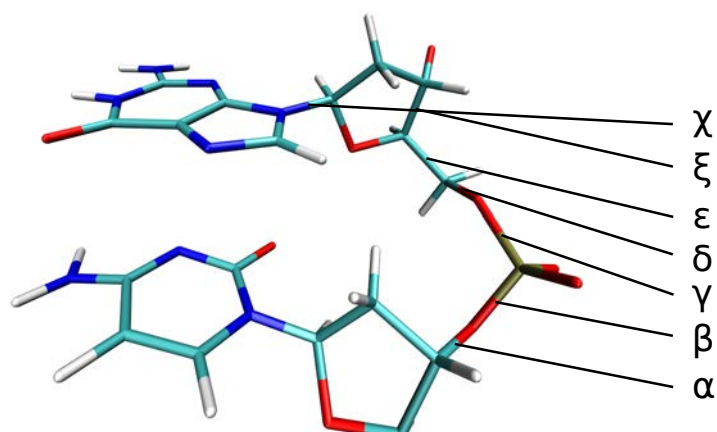


Figure A.1: DNA backbone dihedrals.

Table A.1: Inter base pairs parameters.

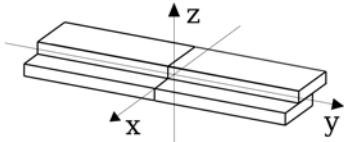
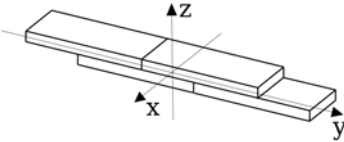
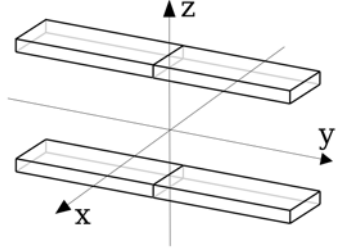
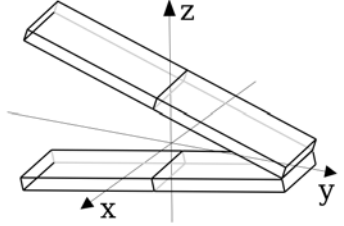
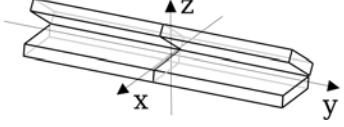
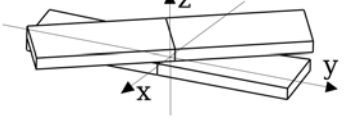
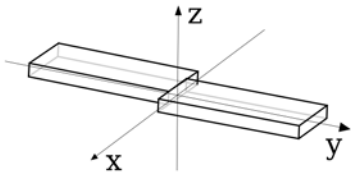
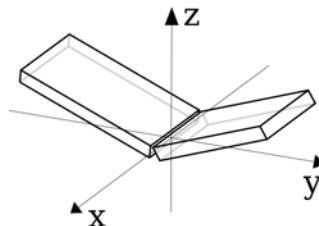
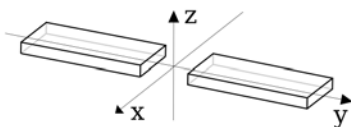
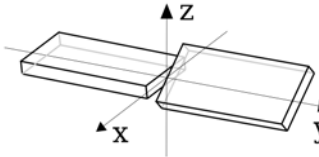
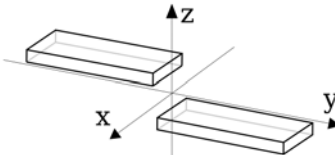
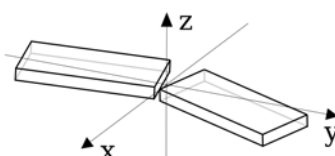
Parameter (B-DNA average canonical value)	Schematic representation
Shift	
Slide	
Rise	
Tilt	
Roll	
Twist	



Table A.2: Intra base pairs parameters.

Parameter (B-DNA average canonical value)	Schematic representation
Shear	
Buckle	
Stretch	
Propeller	
Stagger	
Opening	

# Appendix B

## ECD of B-DNA

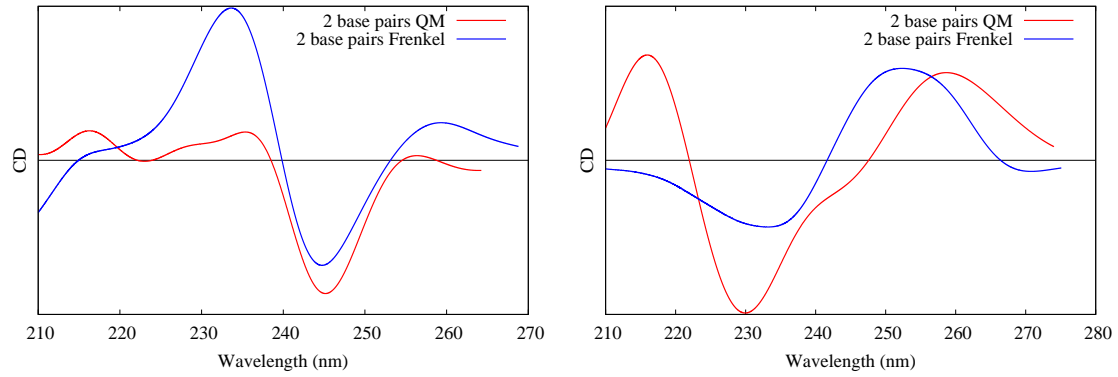


Figure B.1: ECD spectra for a poly(dAT) (right panel) and poly(dCG) (left panel) obtained with the Frenkel Hamiltonian or a full QM/MM approach on the 2 central base pairs. The Frenkel Hamiltonian spectra have been scaled by 8 nm for poly(dAT) and 15 nm for poly(dCG) to match the excitation wavelengths of the QM/MM approach. Wavelengths in nm, intensities in arbitrary units.

# Appendix C

## ECD of G-quadruplexes

Circular dichroism spectra were recorded on a Jasco J-715 spectropolarimeter, using 1 cm path-length quartz cells. The telomeric sequences:

5'-AGGGTTAGGGTTAGGGTTAGGG-3' (hTelo)

and

5'-AAAGGGTTAGGGTTAGGGTTAGGGAA-3 (hTelo-2HY9)

were purchased from IDT (Integrated DNA Technologies, Belgium) at HPLC purity grade. The lyophilized strand was first diluted in Milli-Q water to obtain 100  $\mu$ M stock solution. This was then diluted to the desired concentration, and the experiments were carried out in buffer (pH 7.4) containing 50 mM Tris-HCl and 100 mM salt (KCl or NaCl) in the absence or presence of PEG 200 as indicated. The oligonucleotides were folded into their G-quadruplex conformation by heating the solutions up to 90°C for 5 min and then by slowly cooling at room temperature. The exact concentration of the samples was checked by measuring the absorbance at 260 nm and using the appropriate extinction coefficient value provided by the manufacturer (228 500 for hTelo and 278 200  $\text{M}^{-1}.\text{cm}^{-1}$  for hTelo-2HY9). The antiparallel structure, characterized by a negative peak near 260 nm and a positive peak near 295 nm, was obtained annealing the hTelo sequence in 100 mM NaCl buffered solutions; the hybrid (mixed parallel/antiparallel) structure, characterized by a negative peak near 240 nm and a positive peak near 290 nm with a shoulder at 270 nm, was obtained using hTelo-2HY9 in 100 mM KCl buffered solutions. Finally, the all parallel structure, showing a negative peak near 240 nm and a positive peak near 265 nm, was obtained using hTelo in 100 mM KCl buffered solution with 40% (w/v) of the crowding agent PEG 200. Solvents and reagents were all commercial and used without further purification.

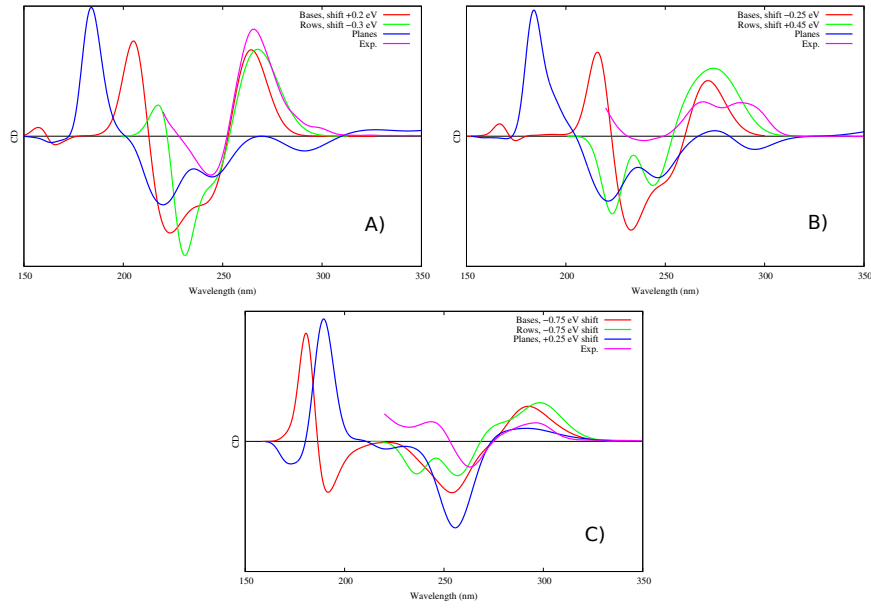
**Full ECD comparison with different QM partitions.**

Figure C.1: Computed and experimental circular dichroism spectra for each QM partition scheme as presented in the main text. A) Parallel, B) Hybrid and C) Antiparallel. Bases alone (red), rows of 3 bases (green), planes of 4 bases (blue) and experimental (purple). Each computed spectra has been shifted by a value indicated on the figure to fit with the experimental spectra.

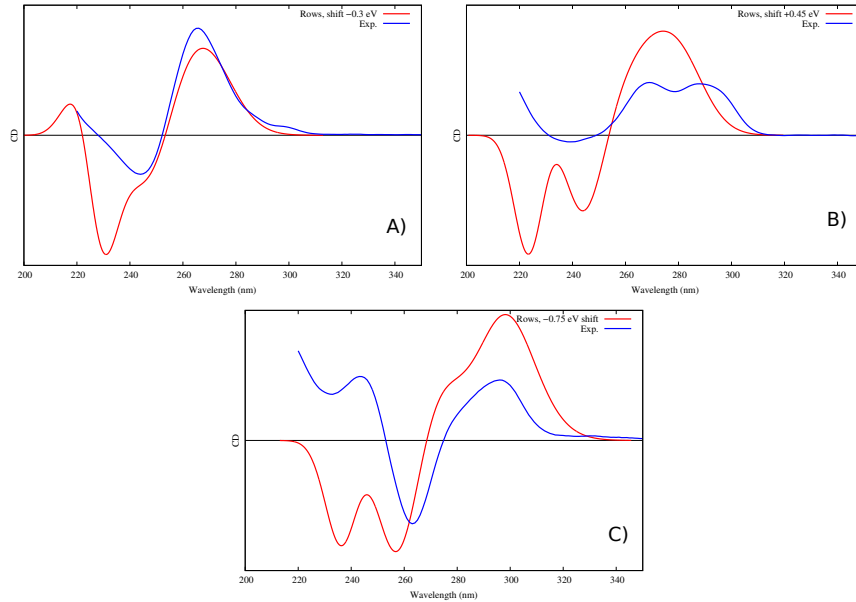
**Full circular dichroism spectrum.**

Figure C.2: Computed (red) and experimental (blue) spectra of A) Parallel, B) Hybrid and C) Antiparallel including higher energy transitions.

CD spectra of the hybrid G4 for different width of the convolution gaussians.

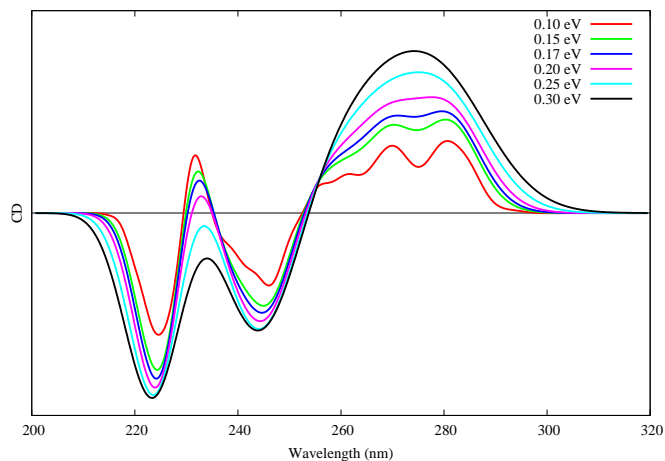


Figure C.3: Circular dichroism spectra of hybrid G4 with different width of the convolution gaussians. It is noticeable that the two characteristic positive bands of the hybrid type appear when the peak width decreases.

### Effect of the environment on the CD spectra.

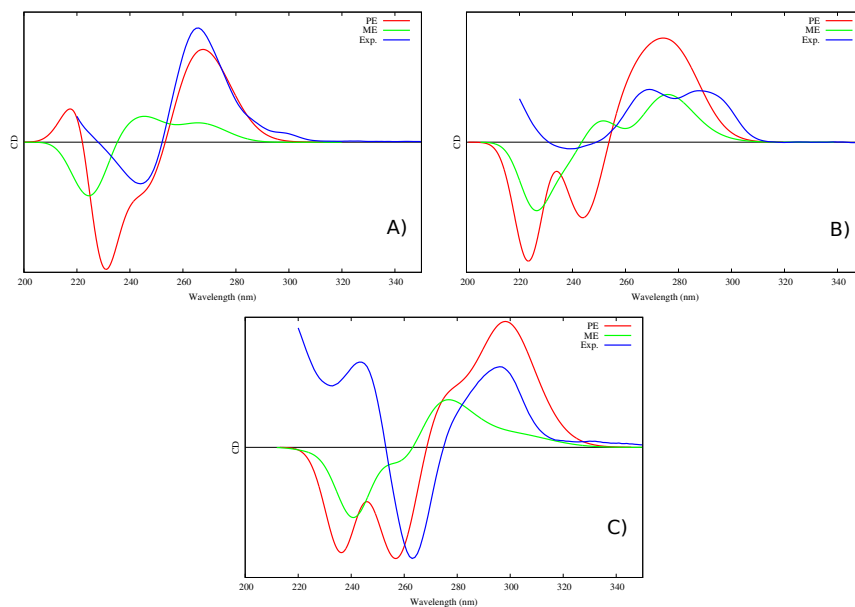


Figure C.4: Comparison of CD spectra taking into account the environment using a polarizable embedding (red) and considering the mechanical embedding only (green), compared with the experimental spectrum (blue). The shift applied on each computed spectrum is the same for both computed spectra. A) Parallel, B) Hybrid and C) Antiparallel.

## Planes to ions distances.

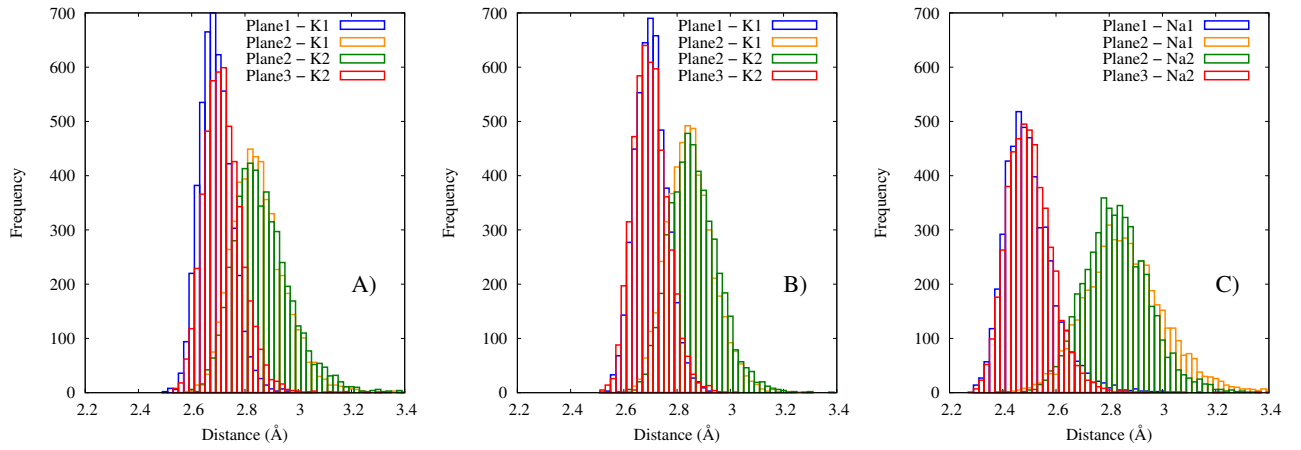


Figure C.5: Distribution of the distances between the guanine center of mass and the stabilizing ions calculated along the MD trajectory. A) Parallel, B) Hybrid, C) Antiparallel. In the inlay we report the G4 schematic structure with the definition of the distances and of the planes.

## Appendix D

### Binding free energy of Benzophenone

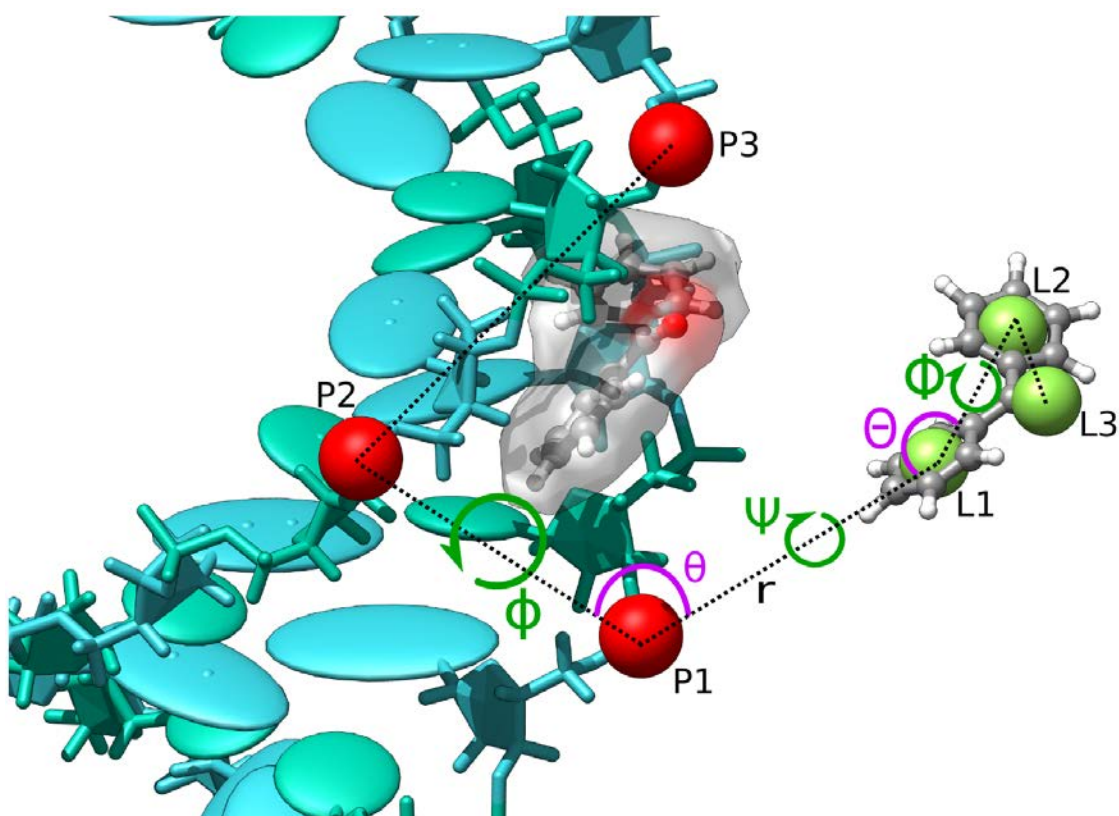


Figure D.1: Representation of the set of parameters defining the position of BNZ bound to the minor groove. For both DNA and BNZ, a triplet of points, P1,P2,P3 (red spheres) and L1,L2,L3 (green spheres) respectively, is defined arbitrarily. The position of BNZ is defined by three spherical coordinates ( $r$  (P1-L1),  $\theta$  (L1-P1-P2),  $\phi$  (L1-P1-P2-P3)) whereas its relative orientation is described by three Euler angles ( $\Theta$  (P1-L1-L2),  $\Phi$  (P1-L1-L2-L3),  $\Psi$  (P2-P1-L1-L2)). For clarity, inserted BNZ (gray surface) has been splitted out of the complex.

r	L1-P1 = 7.7 Å	$k_r = 1.0 \text{ kcal.mol.Å}^{-2}$
$\Theta$	P1-L1-L2 = 160°	$k_\Theta = 0.1 \text{ kcal.mol.Å}^{-2}$
$\Phi$	P1-L1-L2-L3 = -45°	$k_\Phi = 0.1 \text{ kcal.mol.Å}^{-2}$
$\Psi$	P2-P1-L1-L2 = -10°	$k_\Psi = 0.1 \text{ kcal.mol.Å}^{-2}$
$\theta$	L1-P1-P2 = 27°	$k_\theta = 0.1 \text{ kcal.mol.Å}^{-2}$
$\phi$	L1-P1-P2-P3 = 32°	$k_\phi = 0.1 \text{ kcal.mol.Å}^{-2}$
RMSD	$k_{RMSD} = 15 \text{ kcal.mol.Å}^{-2}$	

Table D.1: Harmonic restraints applied on DNA/BNZ in the DBI binding mode. Parameter equilibrium values and force constants are given.

r	L1-P1 = 5.7 Å	$k_r = 1.0 \text{ kcal.mol.Å}^{-2}$
$\Theta$	P1-L1-L2 = 77°	$k_\Theta = 0.1 \text{ kcal.mol.Å}^{-2}$
$\Phi$	P1-L1-L2-L3 = -100°	$k_\Phi = 0.1 \text{ kcal.mol.Å}^{-2}$
$\Psi$	P2-P1-L1-L2 = -174°	$k_\Psi = 0.1 \text{ kcal.mol.Å}^{-2}$
$\theta$	L1-P1-P2 = 74°	$k_\theta = 0.1 \text{ kcal.mol.Å}^{-2}$
$\phi$	L1-P1-P2-P3 = 9°	$k_\phi = 0.1 \text{ kcal.mol.Å}^{-2}$
RMSD	$k_{RMSD} = 7 \text{ kcal.mol.Å}^{-2}$	

Table D.2: Harmonic restraints applied on DNA/BNZ in the minor groove binding mode. Parameter equilibrium values and force constants are given.

Analytical expressions for estimating the contribution of positional and orientational restraints to  $\Delta G^{bulk}$  rest. The constant  $C^0$  insures conversion to the standard state concentration ( $= 1/1661\text{Å}^3$ ). Equilibrium values and force constants of the harmonic potentials are given in the tables hereinbefore.

$$\Delta G_{r,\theta,\phi}^{bulk} = -k_b T \ln(F_t C^0) \quad (\text{D.1})$$

$$F_t = \int_0^\infty dr r^2 \int_0^\pi d\theta \sin\theta \int_{-\pi}^\pi d\phi e^{-\beta u(r,\theta,\phi)} \quad (\text{D.2})$$

$$u(r, \theta, \phi) = \frac{1}{2}(k_r(r - r_0)^2 + k_\theta(\theta - \theta_0)^2 + k_\phi(\phi - \phi_0)^2) \quad (\text{D.3})$$

$$\Delta G_{\Theta,\Phi,\Psi}^{bulk} = -k_b T \ln(F_r C^0) \quad (\text{D.4})$$

$$F_r = \frac{1}{8\pi^2} \int_0^\pi d\Theta \sin\Theta \int_{-\pi}^\pi d\Psi \int_{-\pi}^\pi d\Phi e^{-\beta u(\Theta,\Phi,\Psi)} \quad (\text{D.5})$$

$$u(\Theta, \Phi, \Psi) = \frac{1}{2}(k_\Theta(\Theta - \Theta_0)^2 + k_\Psi(\Psi - \Psi_0)^2 + k_\Phi(\Phi - \Phi_0)^2) \quad (\text{D.6})$$



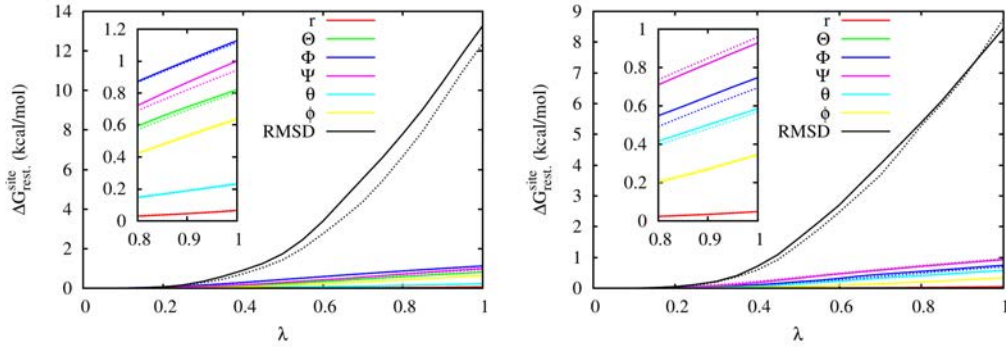


Figure D.2: Free energy contributions of each geometrical restraints for the DI (left) and MinGB (right) binding modes.

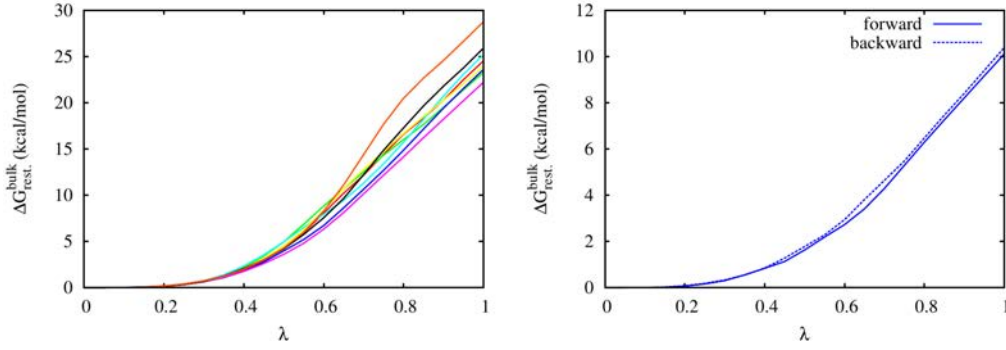


Figure D.3: Free energy contributions of the RMSD restraint in the unbound form of (left) DI mode (14 forward runs) and (right) MinGB mode (forward and backward runs).

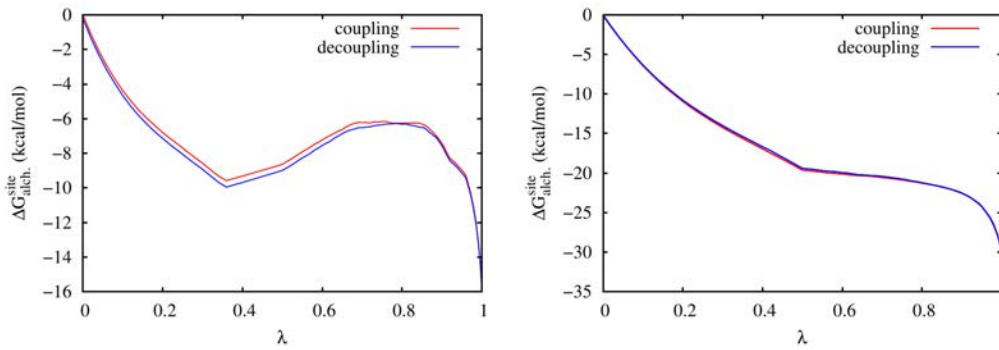


Figure D.4: Free energy perturbation changes for the coupling and decoupling of BNZ in the bound form of DNA for (left) DI mode and (right) MinGB binding mode.

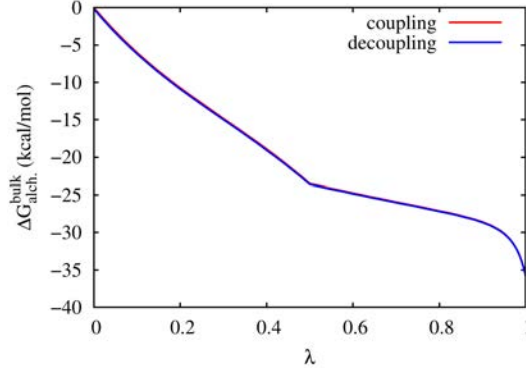


Figure D.5: Free energy perturbation changes for the coupling and decoupling of BNZ from the bulk.

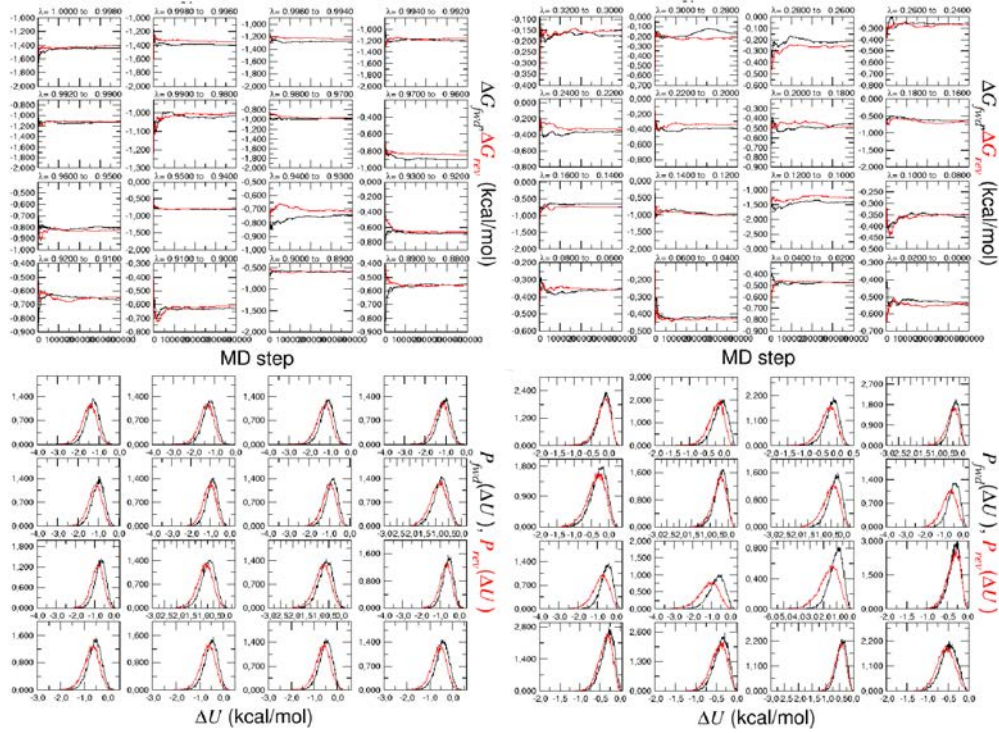


Figure D.6: Forward and backward free energy changes (top) and probability distributions (bottom) as given by ParseFEP for the first 16 windows and last 16 windows for decoupling BNZ from DNA in the MinGB binding mode.

# Appendix E

## Benzophenone: Hydrogen abstraction

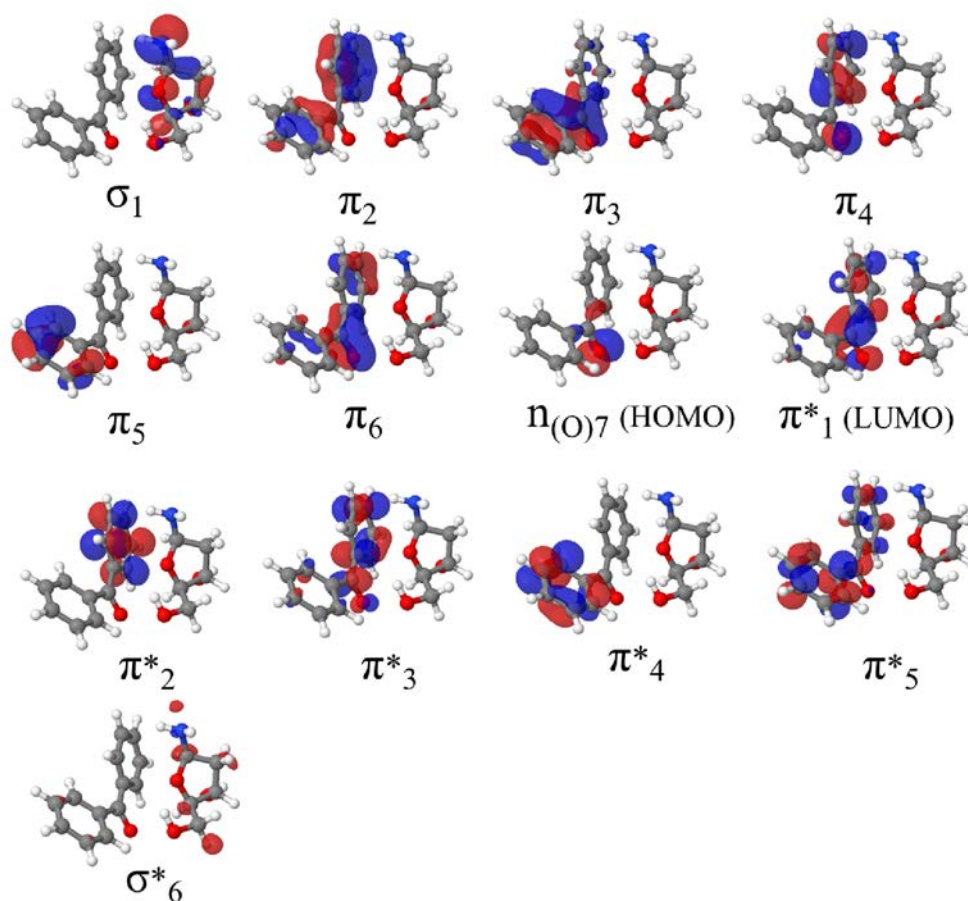


Figure E.1: CAS-SCF natural orbitals of the BP/S system.

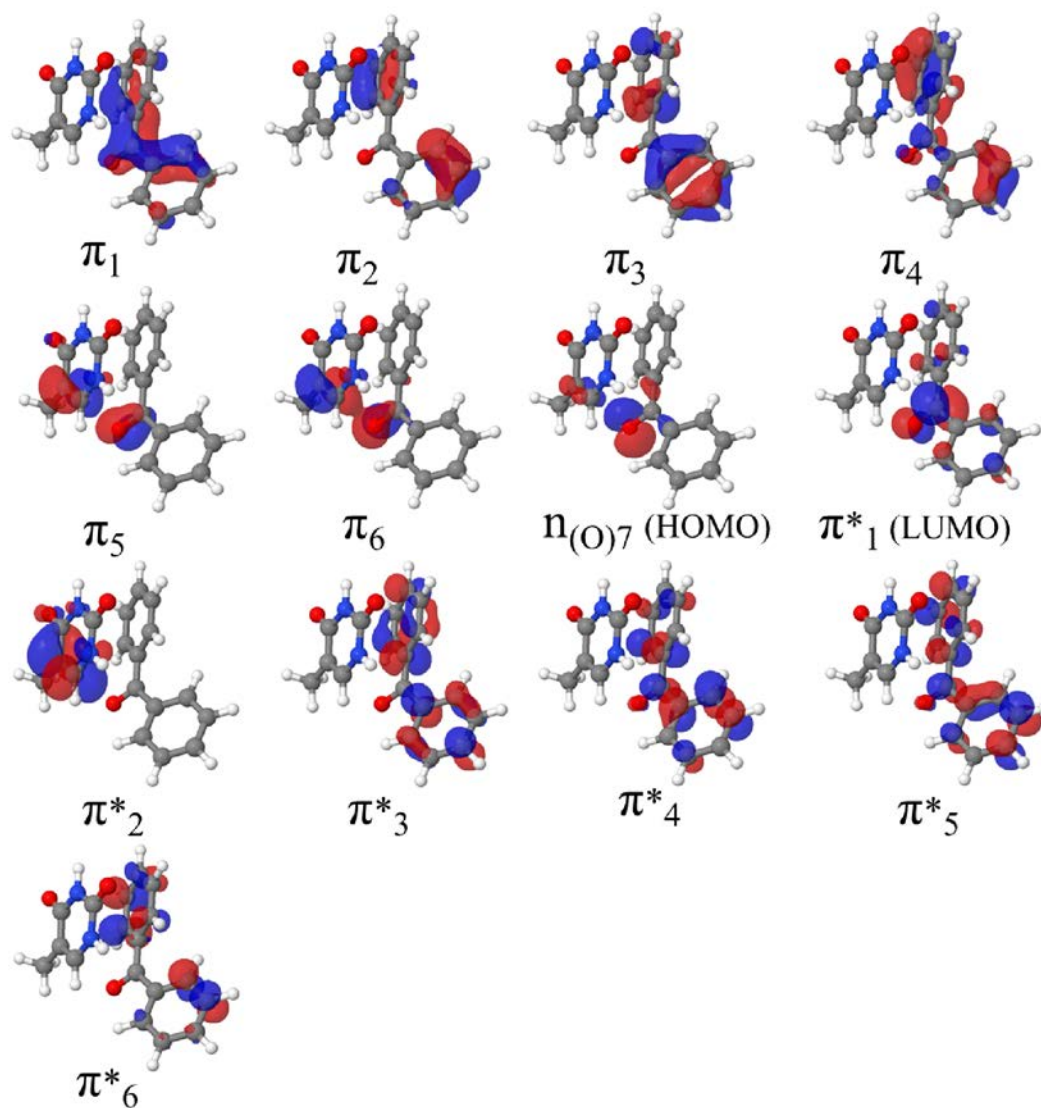


Figure E.2: CAS-SCF natural orbitals of the BP/Thy system.

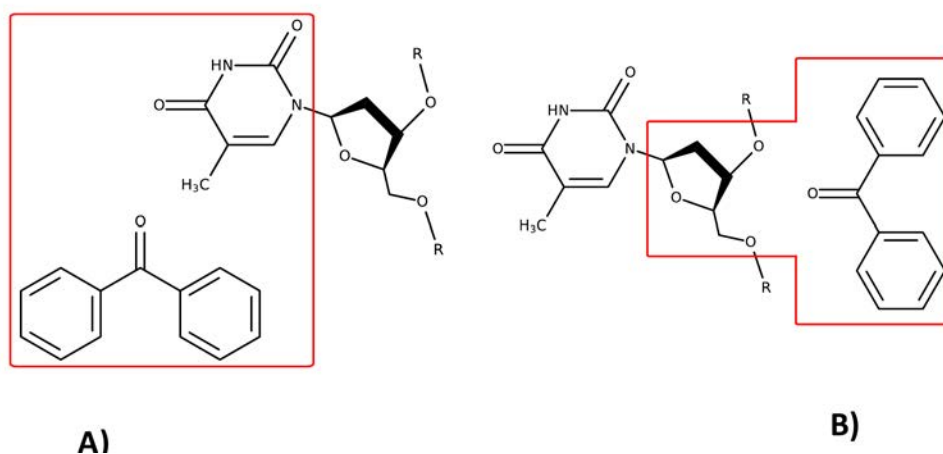


Figure E.3: Scheme of the QM/MM partition for the study of H-abst from Thy (A) and Sugar (B). The atoms inside the red squares have been treated at QM level. The covalent bonds intersecting the red squares represent the QM/MM frontier and have been saturated with link atoms.

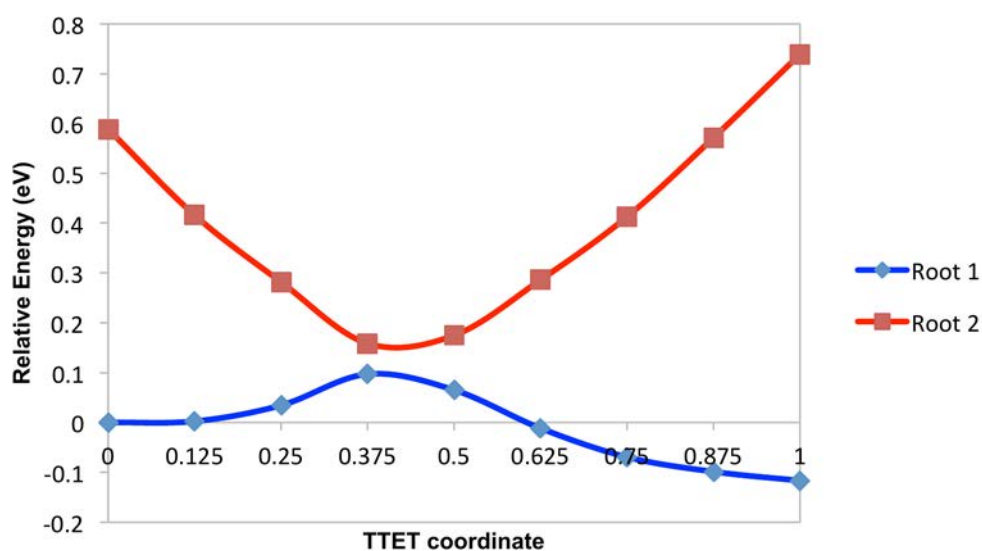


Figure E.4: Jablonski Diagram for the photoionization processes extracted from 20 MD snapshots for the four stable interaction modes. A) Intercalation b) Alternative intercalation c) Minor groove binding d) Alternative minor groove binding. The square represents the energy of the solvated electrons in the limiting 3.3 and 3.6 eV.



# Appendix F

## Nile Blue, Nile Red: Electron transfer

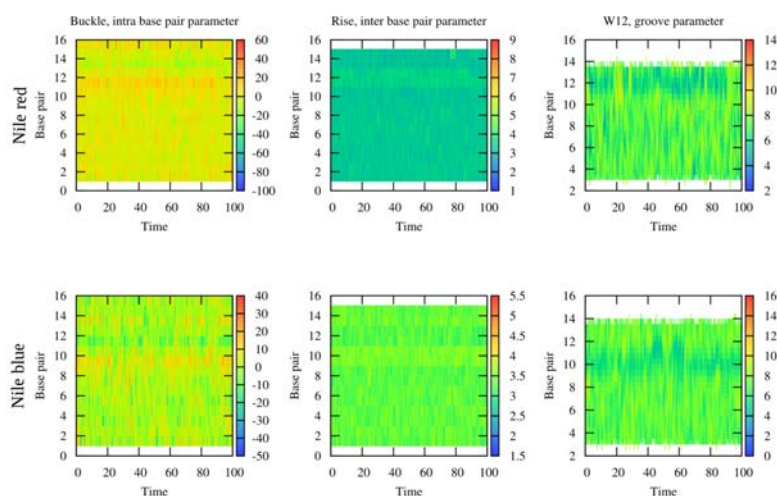


Figure F.1: DNA-NR (top) and DNA-NB (bottom) main geometrical parameters involved in the minor groove binding mode.

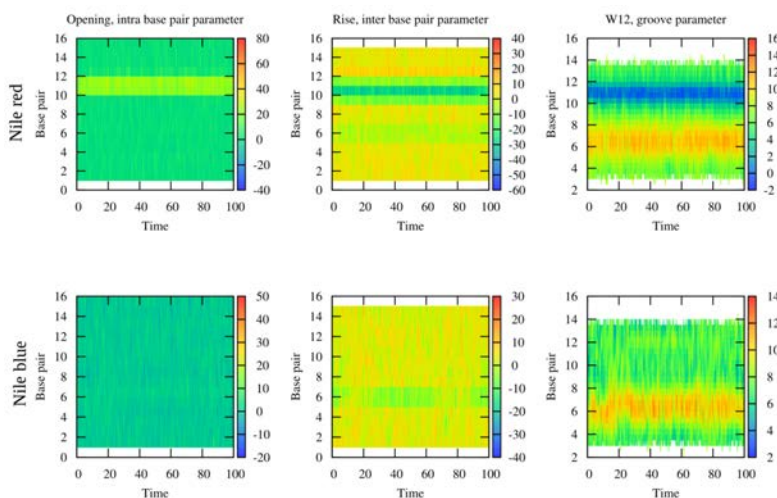


Figure F.2: DNA-NR (top) and DNA-NB (bottom) main geometrical parameters involved in the intercalation mode.

**Molecular orbitals:** The natural transition orbitals involved in the description of the electronic transitions to the locally excited and charge transfer excited state are shown in Figures F.3 (for NB) and F.4 (for NR). For the reader convenience we remind that NTO analysis is based in a singular value decomposition of the transition density matrix that allows to represent the density rearrangement upon excitation using only one orbital couple. The occupied orbital represent the regions from where electronic density have been depleted (i.e. a sort of hole density), while virtual orbital refers to the regions where electrons are accumulated. The nature of the excited states were also assessed using the  $\Phi_S$  index, that assumes values close to 1.0 for local states and close to 0.0 for charge transfer ones. In our case the local states had values close to 0.8 while charge-separated ones assumed values of around 0.4. NTOs and  $\Phi_S$  have been calculated using the locally developed Nancy\_Ex code freely available on the net under GPU license (<https://sourceforge.net/projects/nancyex/>)

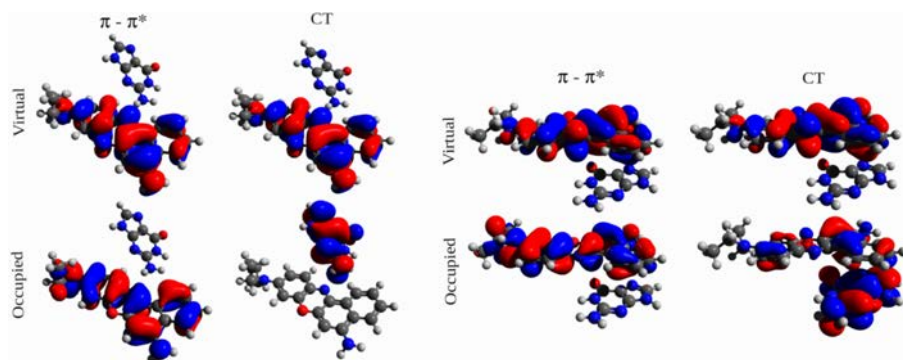


Figure F.3: Occupied and virtual natural transition orbitals involved in the electronic transition to the lowest-energy local excited state ( $\pi - \pi^*$ ) and to the charge transfer state (CT) of the DNA-NB complex in minor groove binding (left) and intercalation (right)

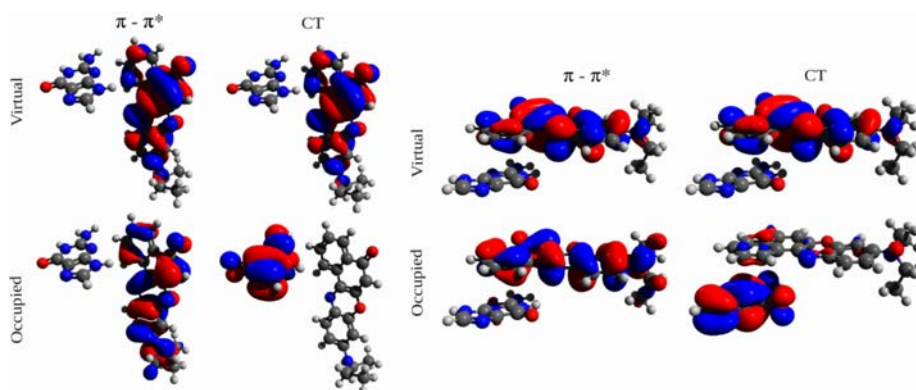


Figure F.4: Occupied and virtual natural transition orbitals involved in the electronic transition to the lowest-energy local excited state ( $\pi - \pi^*$ ) and to the charge transfer state (CT) of the DNA-NR complex in minor groove binding (left) and intercalation (right).

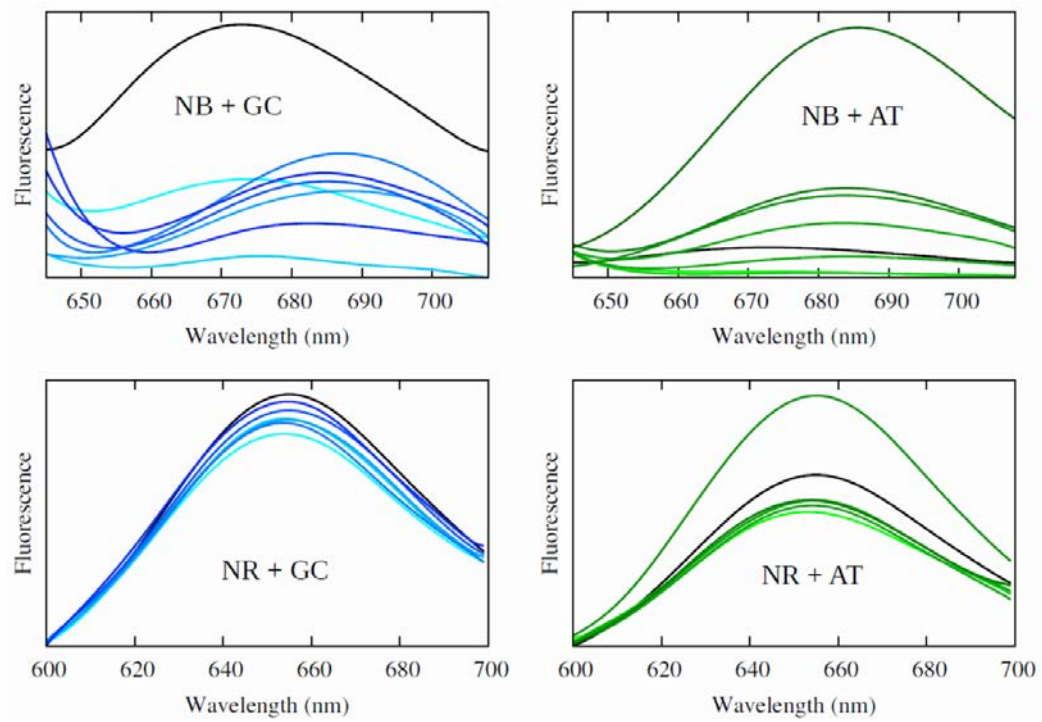


Figure F.5: Experimental fluorescence spectra of NB and NR in solution with poly(dG-dC) and poly(dA-dT)(GC and AT, respectively). All concentrations listed in the Methods section of the main text are shown, from the lowest (light blue for GC; light green for AT) to the highest one (dark blue for GC; dark green for AT).



# Appendix G

## BMEMC: Radical reactive species

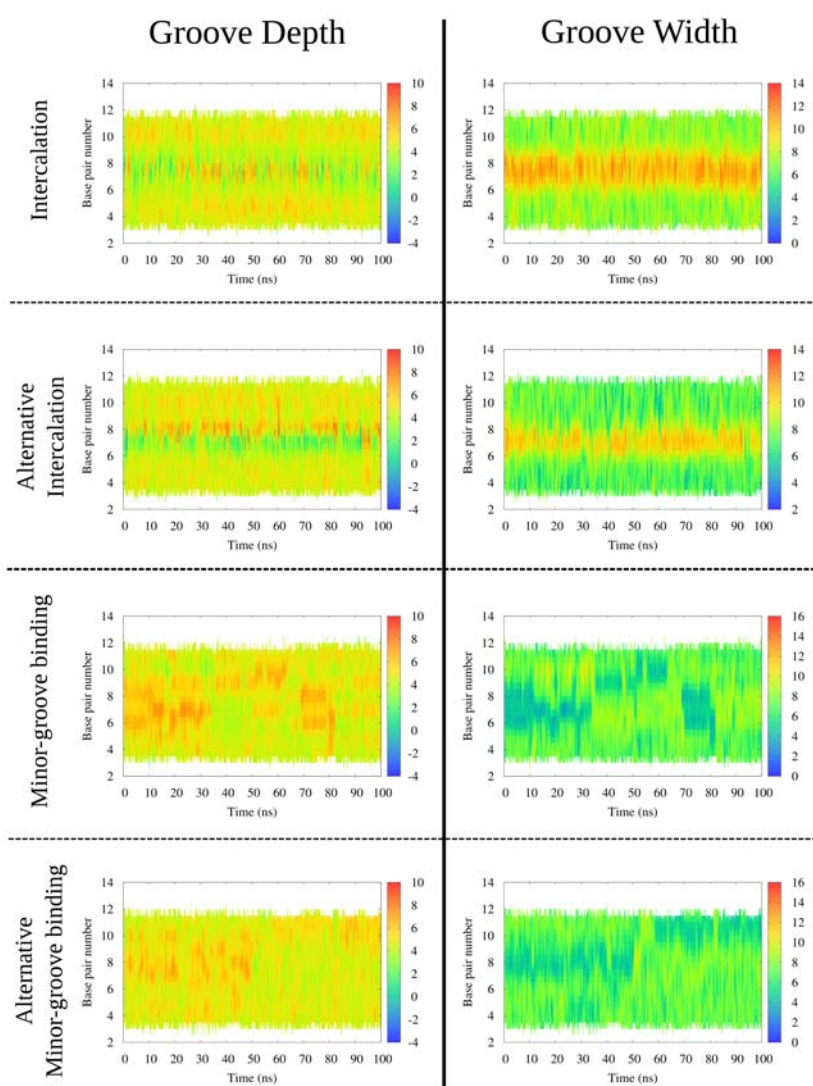


Figure G.1: DNA minor groove width and depth for the four stable interaction mode as obtained using Curves+ code.

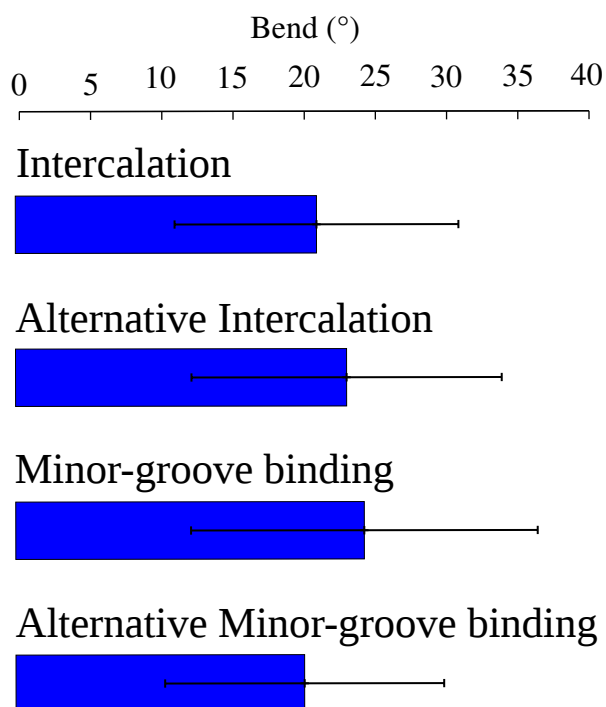


Figure G.2: B-DNA helix bending for the four interaction modes obtained with the Curves+ code.

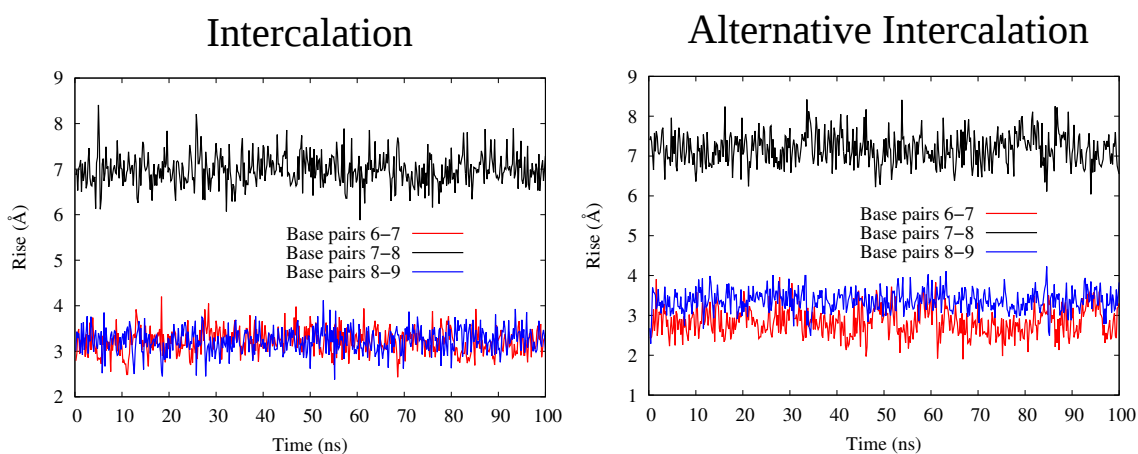


Figure G.3: Time series of the base-pair rise for the intercalation and alternative intercalation obtained with the Curves+ code BMEMC is between base pair 7 and 8.

# Appendix H

## Pyo: Triplet-Triplet Energy Transfer (TTET)

Time Evolution of the Key Intra-Base Distances P6-A15 and T5-A16

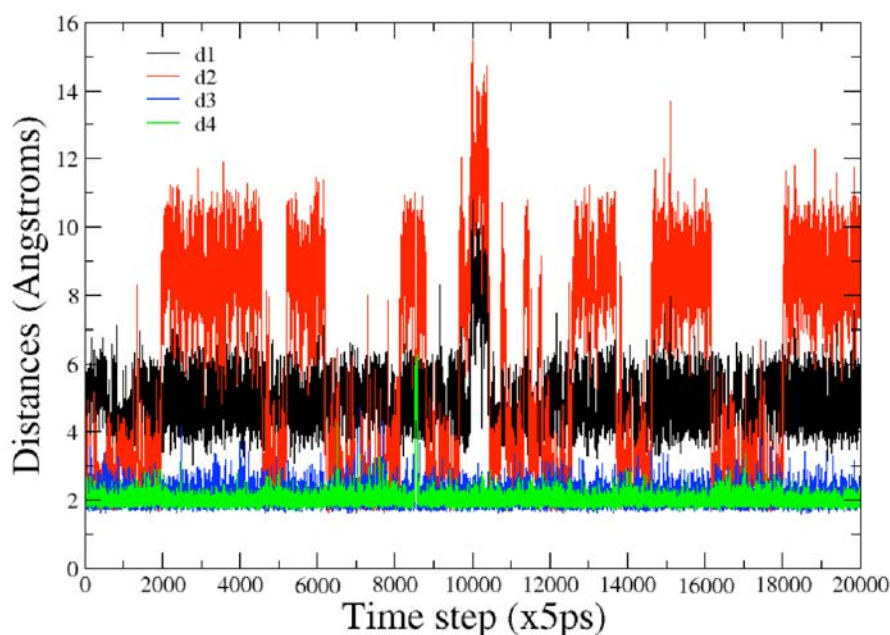


Figure H.1: Time evolution of the distances d1 between the closest nitrogen atoms of P6 and A15, and d2 corresponding to the distance between O on P6 and amino H on A16. For comparison, the time evolution of the distance between O on T5 and H of A16 (d3) and between H on T5 and N on A16 (d4) are reported too.

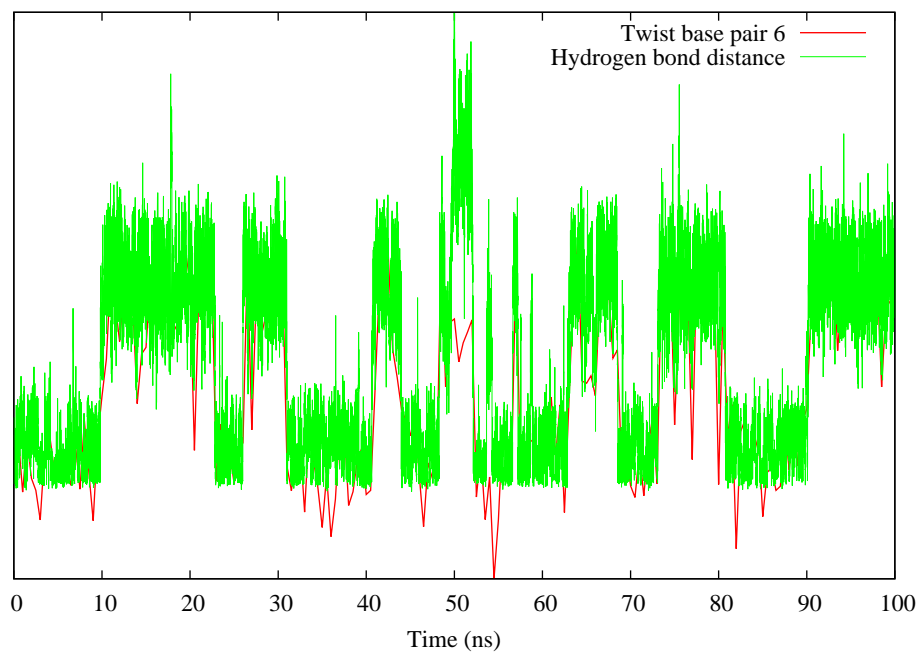


Figure H.2: Comparison of the time evolution of  $d_1$  and the deviation from ideal value of the twist interbase parameter for base pair couple 6-7.

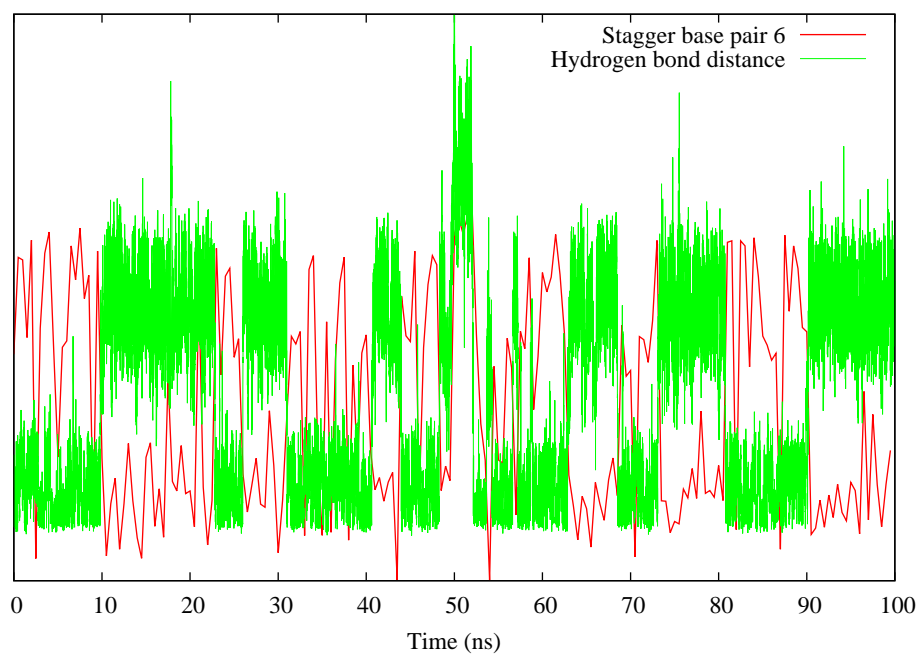


Figure H.3: Comparison of the time evolution of  $d_1$  and the deviation from ideal value of the stagger interbase parameter for base pair couple 6-7.

**Exchange-Correlation Functionals Performance** To assess for the performance of the exchange correlation functionals used to model spectroscopic and photochemical properties we calculated the QM/MM absorption spectrum with hybrid and long-range corrected functionals, as well as the meta-GGA M06-2X. All the functionals give satisfactory results and draw a coherent picture of the absorption spectrum.

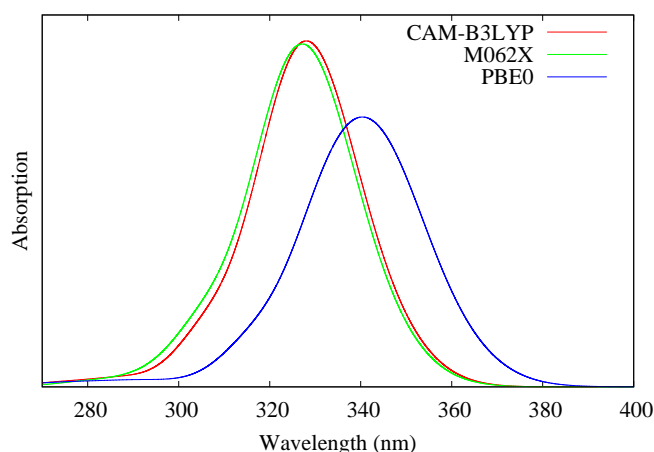


Figure H.4: Calculated absorption spectra for the Pyo nucleobase in B-DNA. TD-DFT with different functionals has been used. The protocols to obtain the spectra are the same as the one described in the main text.

**ECD spectra** Electronic circular dichroism spectrum has been calculated convoluting 30 vertical transitions obtained from 10 snapshots extracted from the MD trajectory. The spectra have been calculated at QM/MM TD-DFT level using M06-2X as exchange-correlation functional and the 6-31+G(d) basis set. The QM partition comprised the artificial Pyo and its paired base (P6-A15) and the closest base pairs T5-A16 and A7-T14. This different strategy compared to the absorption when only Pyo was included in the QM partition is necessary to take into account the supramolecular induced circular dichroism of helical structures such as B-DNA. Furthermore most of the ECD signal coming from DNA nucleobases absorption, their inclusion in the QM partition is necessary and also allows to infer for its eventual structural deformation.

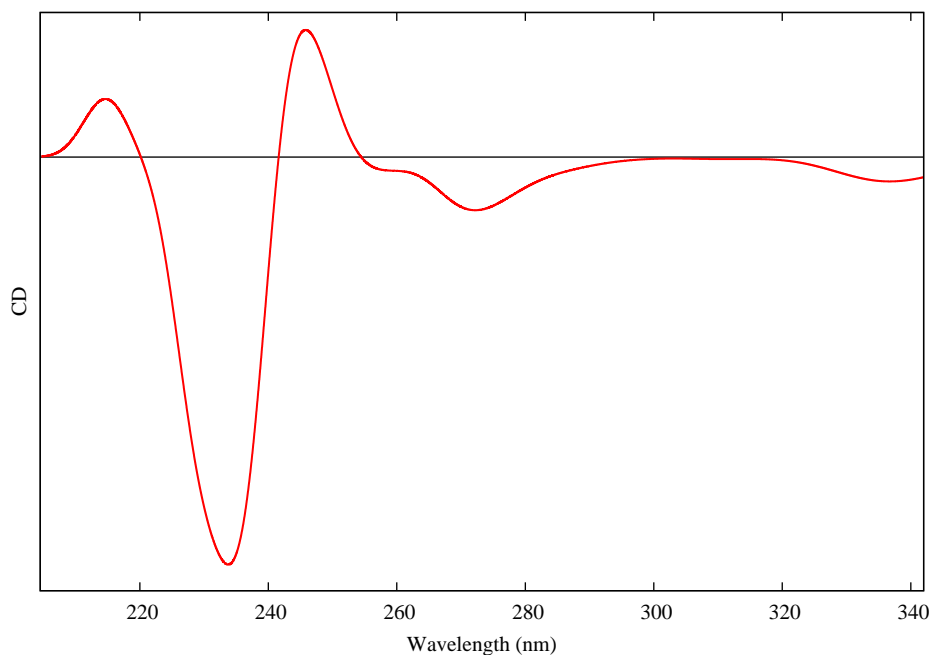


Figure H.5: Calculated ECD Spectrum for the Pyo containing DNA double strand.

# Appendix I

## Cluster abasic sites

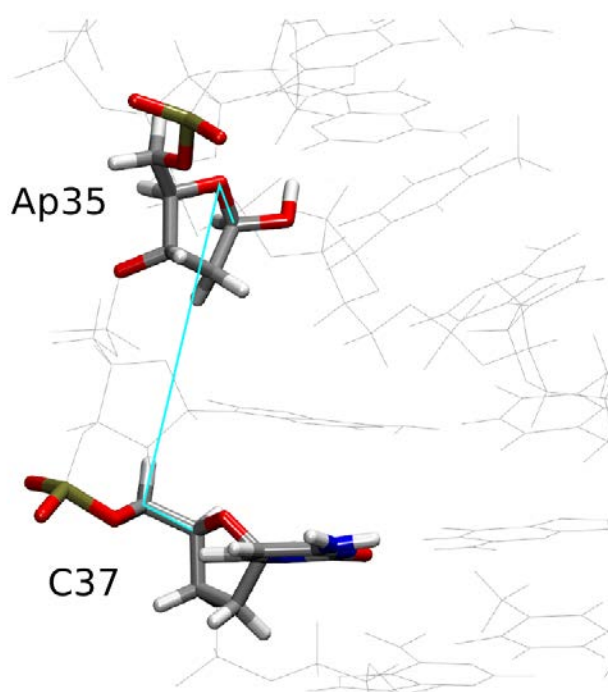


Figure I.1: Representation of the atoms used to calculate the dihedral angle and estimate minor groove occupancy.

		% minor groove occupancy
seq0	Ap12	92.9
	Ap35	99.9
seq+1	Ap12	17.4
	Ap35	14.4
seq-1	Ap12	18.1
	Ap35	7.9
seq+3	Ap12	76.4
	Ap35	16.0
seq-3	Ap12	92.5
	Ap35	73.4
seq+5	Ap12	28.0
	Ap35	21.5
seq-5	Ap12	39.1
	Ap35	41.1
seqC	Ap12	6.4

Table I.1: Minor groove occupancy of the extrahelical AP sites.

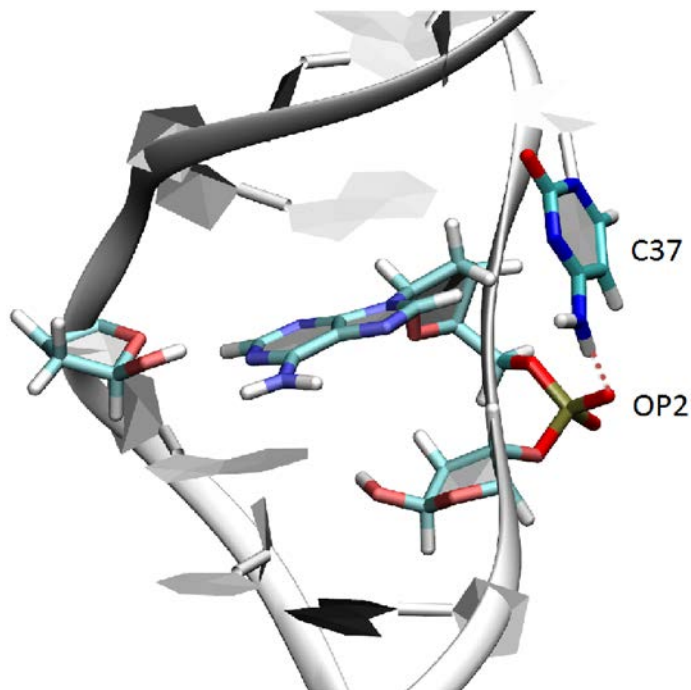


Figure I.2: Representative MD snapshot showing the occurrence of an HB between the ejected C37 pyrimidine and the backbone oxygen. The latter locks C37 in an extrahelical position.



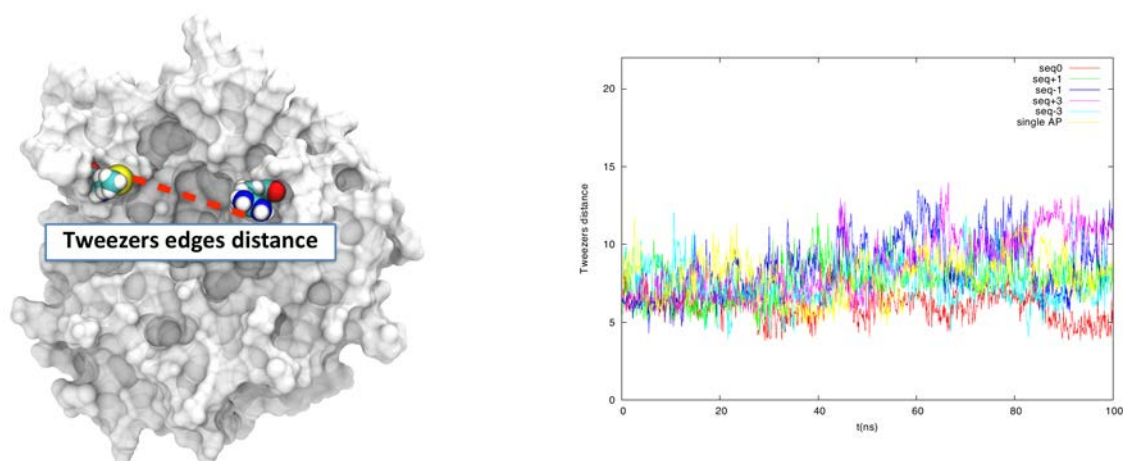


Figure I.3: Definition of the tweezers' opening, taken as the distance between the sulfur atoms of the Methionine and the  $\gamma$ -Carbon atom of the Arginine and time series of the tweezers' edges distance for all the for all the DNA/APE1 complexes simulated in this work.

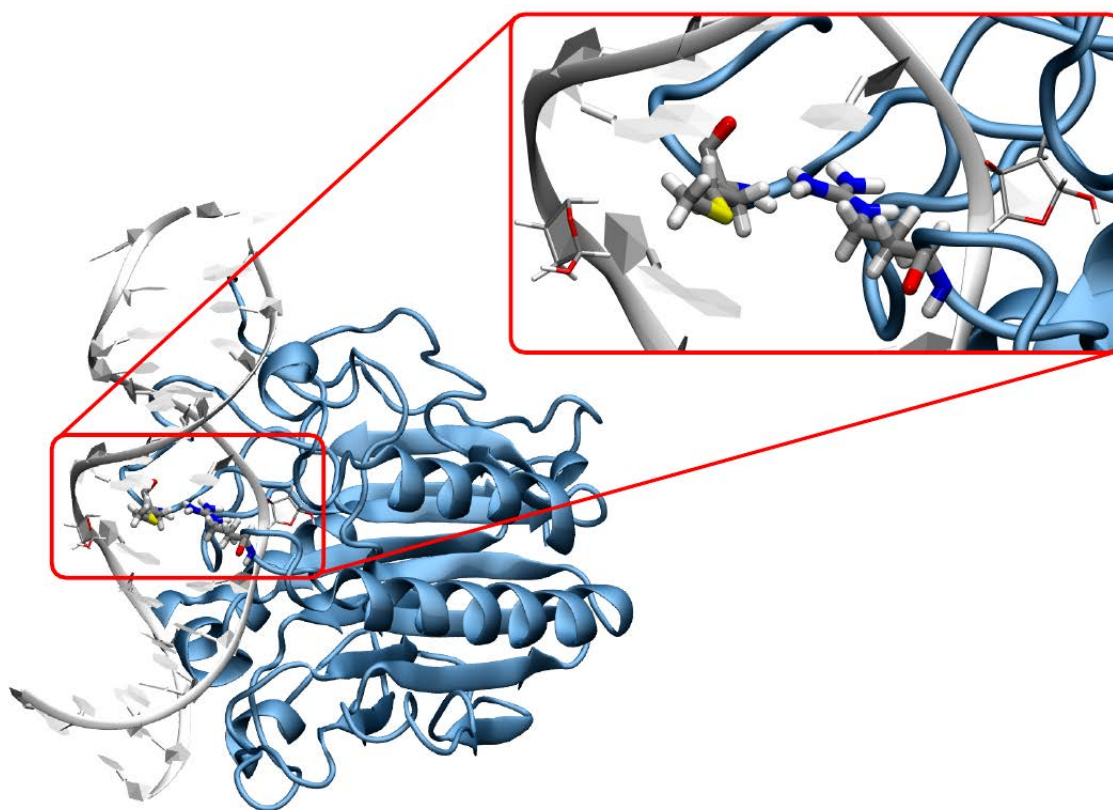


Figure I.4: A representative snapshot of the DNA APE1 complex showing the tweezers and its edges.

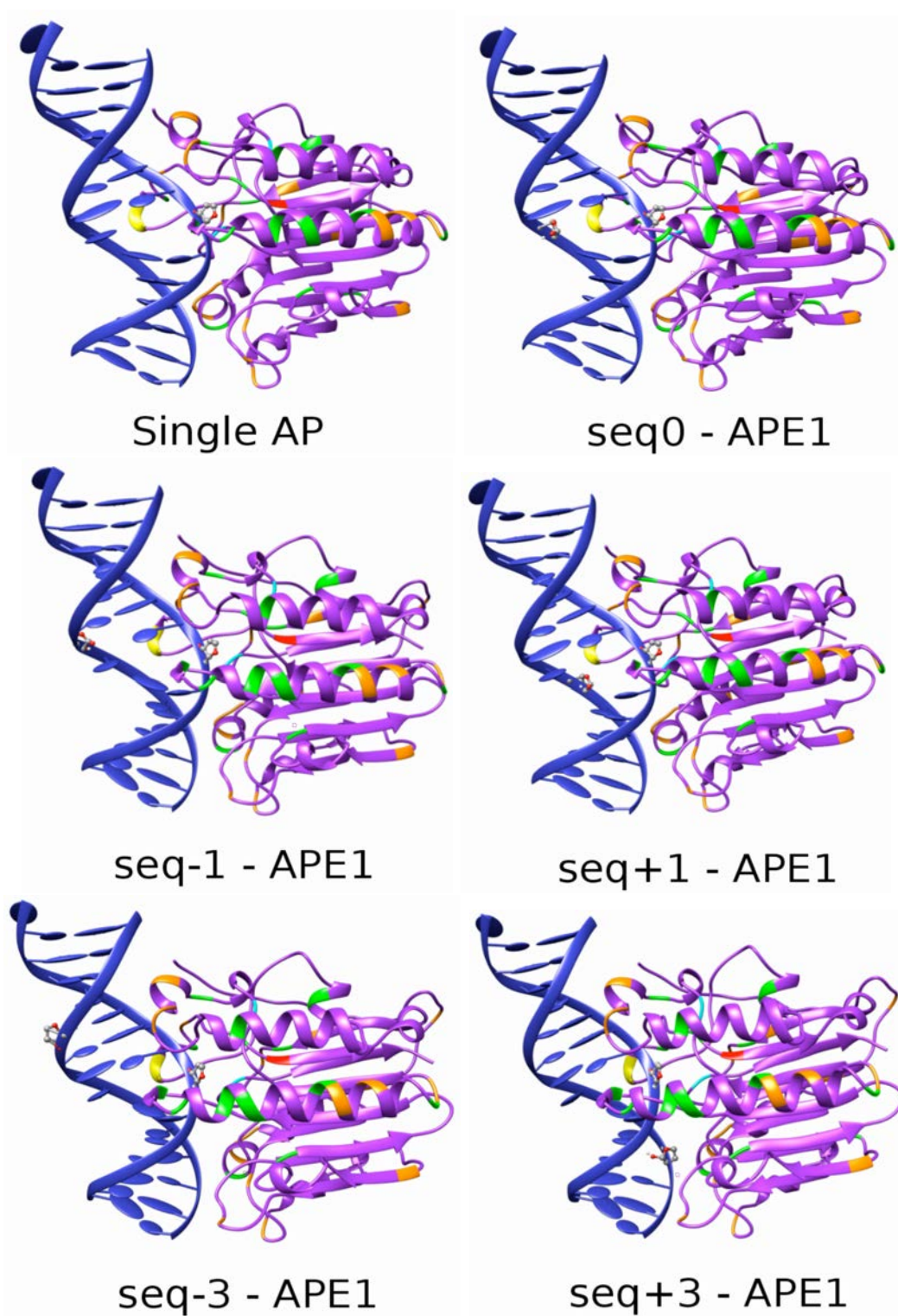


Figure I.5: Representative snapshots of DNA/APE1 complexes. DNA and protein are represented in carton. AP sites are evidenced in cpk representation.

# Appendix J

## 6-4PP and CPD

**MD methodology** All classical molecular dynamics simulations were performed using NAMD suite of programs, using ff99bsc0 and gaff force fields. Three 16 base pairs double strand oligomers have been built, the first one containing 64-PP lesions at the 8th and 9th position, the second one containing CPD at the same position, and finally an ideal non-damaged B-DNA for control. The DNA strands were built with the nab module. RESP charges and parameters were generated for 6-4PP and CPD lesions' sites according to the standard Antechamber protocol. Then, the system was placed in an octahedral TIP3P water box of 12 Å buffer, with 30 potassium ( $K^+$ ) cations to ensure the medium's neutrality. For CPD and B-DNA containing sequences, 100 ns classical molecular dynamics were performed after an equilibration procedure we described hereafter. On the contrary, for the 6-4PP sequence the production run reached 2  $\mu s$ . This choice was justified by the fact that CPD and B-DNA are not showing extended polymorphism and hence their inherent structural characteristics were already captured at the shorter time-scale. Conversely, the flexibility and the plurality of conformations exhibited by 6-4PP necessitated a much longer sampling. Although, our dynamic is not sufficient to estimate the relative position of the different conformations it definitively allowed to evidence the dynamic equilibrium and the interconversion between the conformers leading to polymorphism. Equilibration was performed with the following protocol for all the sequences: 10000 steps minimization, including 5000 steps of steepest descent, was first performed in order to adapt our system to the force field. Then, temperature was increased from 0K to 300 K in a 20 ps thermalization run. Langevin thermostat with a  $1.0 \text{ ps}^{-1}$  collision frequency  $\gamma_{ln}$  was used to keep it constant during the remaining of the simulation. Then, a first 100 ps equilibration run in NPT was performed, followed by a second one in NVT conditions. Finally, production was run in the NPT thermodynamic ensemble. Replica exchange simulations have been carried out with NAMD. Using the same protocol as the one employed for the unbiased MD simulation, an ensemble of 64 copies of the solvated DNA double-

strand featuring a 6-4PP lesion was evolved via replica exchange MD for 75 ns, representing a total sampling time of 4.8  $\mu$ s. Target temperatures of the 64 replicas ranged from 290 K to 418 K, a temperature step of 2 K was chosen to ensure an exchange success rate of about 30%. Exchange attempts between adjacent replicas were made every 200 fs.

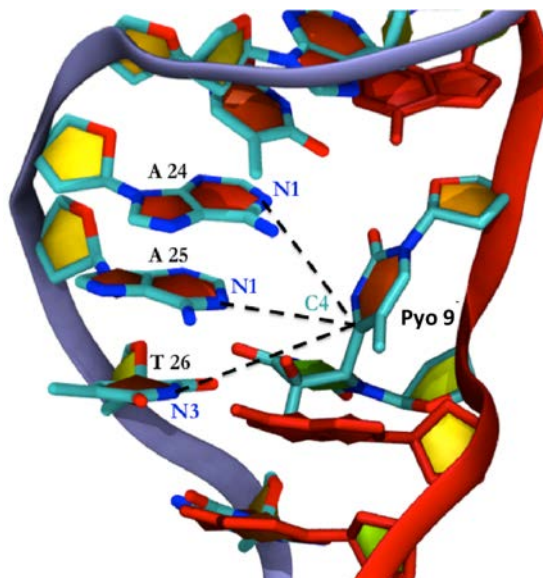


Figure J.1: Definition of the distances used to identify the extent of p-stacking between 6-4PP and the nearby bases. In particular the stacking with the pyrimidine (Pyo9) unit is considered for the three closest nucleobases belonging to the opposite strand (A24, A25, T26).

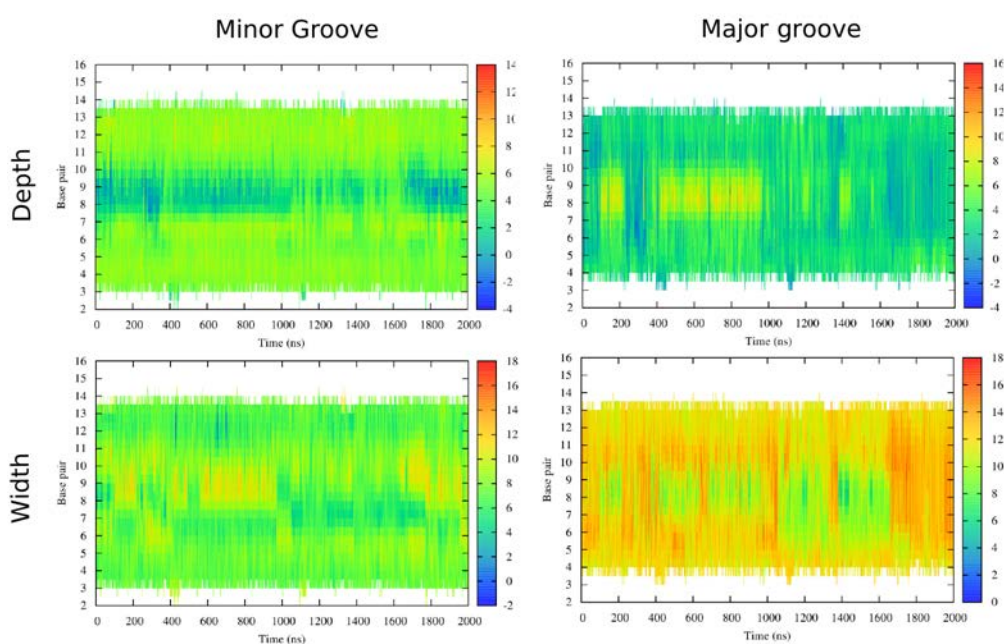


Figure J.2: Time series maps of the evolution of the minor (left) and major (right) groove width and depth.



# Bibliography

- [1] Y. Hiraku et al. *Photochemistry and Photobiology* (2007); (83)1: 205–212.
- [2] D. E. Dolmans, D. Fukumura, and R. K. Jain. *Nature Reviews Cancer* (2003); (3)5: 380–387.
- [3] Y. Zheng et al. *RSC Advances* (2015); (5)1: 770–774.
- [4] J. G. Levy. *Seminars in Oncology* (1994); (21)6: 4–10.
- [5] H. Abrahamse and M. R. Hamblin. *Biochemical Journal* (2016); (473)4: 347–364.
- [6] O. J. Balchum, D. R. Doiron, and G. C. Huth. *Progress in Clinical and Biological Research* (1984); (170) 727–745.
- [7] B. W. Henderson et al. *Cancer Research* (2000); (60)3: 525–529.
- [8] C. Hopper. *European Journal of Cancer* (2001); (37) S38–S39.
- [9] L. B. Josefsen and R. W. Boyle. *Metal-based Drugs* (2008); (2008)
- [10] P. J. Stephens et al. *Nature* (2012); (486)7403: 400–404.
- [11] J. J. Serrano-Pérez, L. Serrano-Andrés, and M. Merchán. *Chemical Physics* (2008); (347)1: 422–435.
- [12] L. Shen and H.-F. Ji. *Journal of Photochemistry and Photobiology B: Biology* (2008); (91)2: 96–98.
- [13] A. Pérez, F. J. Luque, and M. Orozco. *Accounts of Chemical Research* (2012); (45)2: 196–205.
- [14] G. L. Seibel et al. *Biophysics* (1985); (82) 6537–6540.
- [15] R. Galindo-Murillo et al. *Journal of Chemical Theory and Computation* (2016); (12)8: 4114–4127.
- [16] P. D. Dans et al. *Nucleic Acids Research* (2014); (42)18: 11304–11320.
- [17] J. Wang et al. *Journal of Computational Chemistry* (2004); (25)9: 1157–1174.
- [18] D.-Z. Chen et al. *Journal of Photochemistry and Photobiology A: Chemistry* (2005); (170)1: 37–43.

- [19] W. Lin et al. *Photochemical and Photobiological Sciences* (2006); (5)4: 422–425.
- [20] J. Leszczynski. *International Journal of Quantum Chemistry* (1992); (44)19: 43–55.
- [21] M. S. de Vries, E. Nir, and K. Kleinermauns. *Nature* (2000); (408)6815: 949–951.
- [22] S. Mai et al. *Chemical Physics* (2017); (482) 9–15.
- [23] N. Gorczak et al. *The Journal of Physical Chemistry B* (2015); (119)24: 7673–7680.
- [24] P. Marquetand et al. *Molecules* (2016); (22)1: 49.
- [25] W. Thiel and G. Hummer. *Nature* (2013); (504)7478: 96–97.
- [26] M. Chehel Amirani and T. Tang. *Physical Chemistry Chemical Physics* (2015); (17)11: 7564–7575.
- [27] U. Ryde. *Methods in Enzymology* (2016); (577) 119–158.
- [28] J. Aranda et al. *ACS Catalysis* (2016); (6)5: 3262–3276.
- [29] R. Santus and J. P. Reyftmann. *Biochimie* (1986); (68)6: 843–848.
- [30] R. C. Axtmann. *AIChE Journal* (1982); (28)5: 869–869.
- [31] A. Jablonski. *Nature* (1933); (131)3319: 839–840.
- [32] J. Franck and E. G. Dymond. *Transactions of the Faraday Society* (1926); (21) 536.
- [33] L. S. Forster and R. Livingston. *The Journal of Chemical Physics* (1952); (20)8: 1315–1320.
- [34] D. I. Arnon. *Proceedings of the National Academy of Sciences of the United States of America* (1971); (68)11: 2883–2892.
- [35] A. Wand et al. *Journal of the American Chemical Society* (2011); (133)51: 20922–20932.
- [36] M. F. Holick. *Federation proceedings* (1987); (46)5: 1876–1882.
- [37] K. S. Conrad, C. C. Manahan, and B. R. Crane. *Nature Chemical Biology* (2014); (10)10: 801–809.
- [38] A. Sawut et al. *Carbohydrate Polymers* (2014); (101) 231–239.
- [39] D. E. Dolmans, D. Fukumura, and R. K. Jain. *Nature Reviews Cancer* (2003); (3)5: 380–387.
- [40] N. Hoffmann. *Chemical Reviews* (2008); (108)3: 1052–1103.

- [41] M.-H. Zhang et al. *Journal of Photochemistry and Photobiology A: Chemistry* (1996); (96)1-3: 57–63.
- [42] Z. Wang, J. Zhao, and S. Guo. *CHIMIA International Journal for Chemistry* (2015); (69)9: 524–529.
- [43] M. Deiana et al. *PloS one* (2015); (10)6: e0129817.
- [44] N. Shahabadi, S. Mohammadi, and R. Alizadeh. *Bioinorganic Chemistry and Applications* (2011), 429241.
- [45] M. T. Beels et al. *Physical Review A* (2015); (91)4: 043818.
- [46] S. B. Chen et al. *Journal of the American Chemical Society* (2016); (138)33: 10382–10385.
- [47] N. Purdie. *Circular Dichroism and the Conformational Analysis of Biomolecules* (1996); 738: Plenum Press:New York.
- [48] J. Kypr et al. *Nucleic Acids Research* (2009); (37)6: 1713–1725.
- [49] N. J. Greenfield. *Nature Protocols* (2006); (1)6: 2876–2890.
- [50] S. Paramasivan, I. Rujan, and P. H. Bolton. *Methods* (2007); (43)4: 324–331.
- [51] A. G. Kozlov, R. Galletto, and T. M. Lohman. *Methods in Molecular Biology* (2012); (922) 55–83.
- [52] T. Takada, M. Fujitsuka, and T. Majima. *Proceedings of the National Academy of Sciences of the United States of America* (2007); (104)27: 11179–11183.
- [53] M. Makarska-Bialokoz. *Journal of Luminescence* (2014); (147) 27–33.
- [54] M. A. Quintero. *Ars Pharmaceutica* (2000); (41)1: 27–46.
- [55] R. R. Allison et al. *Photodiagnosis and Photodynamic Therapy* (2004); (1) 27–42.
- [56] M. S. Baptista et al. *Photochemistry and Photobiology* (2017), 10.1111/php.12716.
- [57] T. Atsumi et al. *Archives of Oral Biology* (2001); (46)5: 391–401.
- [58] P. Pospíšil. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* (2009); (1787)10: 1151–1160.
- [59] S. Bouillaguet et al. *Photomedicine and Laser Surgery* (2010); (28)4: 519–525.
- [60] L. E. Xodo, S. Cogoi, and V. Rapozzi. *Future Medicinal Chemistry* (2016); (8)2: 179–194.
- [61] C. K. Remucal and K. McNeill. *Environmental Science and Technology* (2011); (45)12: 5230–5237.

- [62] M. Kono, P. W. Goletz, and R. K. Crouch. *Biochemistry* (2008); (47)28: 7567–7571.
- [63] G. Jori. *Journal of Photochemistry and Photobiology. B, Biology* (1996); (36)2: 87–93.
- [64] H. Kataoka et al. *Annals of Translational Medicine* (2017); (5)8:
- [65] V. A. Lapina et al. *Journal of Applied Spectroscopy* (2016); (83)3: 344–349.
- [66] R. Bonnett and M. Berenbaum. *Ciba Foundation Symposium* (1989); (146) 40–53.
- [67] H. Kolarova et al. *General physiology and biophysics* (2008); (27)2: 101–105.
- [68] Z. Huang. *Technology in cancer research and treatment* (2005); (4)3: 283–293.
- [69] I. Yoon, J. Z. Li, and Y. K. Shim. *Clinical Endoscopy* (2013); (46)1: 7–23.
- [70] J. D. Bhawalkar et al. *Journal of Clinical Laser Medicine and Surgery* (1997); (15)5: 201–204.
- [71] J. Schmitt et al. *Chemistry - A European Journal* (2016); (22)8: 2775–2786.
- [72] C. Monnereau and C. Andraud. *SPIE Newsroom* (2012).
- [73] J. Fong et al. *Photochemical Photobiological Sciences* (2015); (14)11: 2014–2023.
- [74] G. Shi et al. *Coordination Chemistry Reviews* (2015); (282-283) 127–138.
- [75] M. Wakamori et al. (2015).
- [76] J. D. Watson and F. H. C. Crick. *Nature* (1953); (171)4356: 737–738.
- [77] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii. *Nucleic Acids Research* (2006); (34)2: 564–574.
- [78] Z.-J. Tan and S.-J. Chen. *Biophysical Journal* (2006); (90)4: 1175–1190.
- [79] R. Lavery et al. *Nucleic Acids Research* (2009); (37)17: 5917–5929.
- [80] X.-J. Lu and W. K. Olson. *Nucleic Acids Research* (2003); (31)17: 5108–5121.
- [81] A. Sancar. *Chemical Reviews* (2003); (103)6: 2203–2238.
- [82] B. O. Roos. In: 2007, John Wiley & Sons, Inc.
- [83] S. Blaser et al. *CHIMIA International Journal for Chemistry* (2014); (68)4: 260–263.
- [84] J.-M. L. Pecourt, J. Peon, and B. Kohler. *Journal of the American Chemical Society* (2000); (122)38: 9348–9349.
- [85] J. M. Pecourt, J. Peon, and B. Kohler. *Journal of the American Chemical Society* (2001); (123)42: 10370–10378.



- [86] L. Serrano-Andrés and M. Merchán. *Journal of Photochemistry and Photobiology C* (2009); (10) 21–32.
- [87] J. S. Taylor. *Accounts of Chemical Research* (1994); (27)3: 76–82.
- [88] S. Mouret et al. *Proceedings of the National Academy of Sciences of the United States of America* (2006); (103)37: 13765–13770.
- [89] M. S. COOKE et al. *The FASEB Journal* (2003); (17)10: 1195–1214.
- [90] M. Dizdaroglu. *FEBS Letters* (1993); (315)1: 1–6.
- [91] T. S. Dexheimer. In: 2013, Springer.
- [92] W. F. Van Gunsteren and H. J. C. Berendsen. *Molecular Simulation* (1988); (1)3: 173–185.
- [93] J. A. McCammon, B. R. Gelin, and M. Karplus. *Nature* (1977); (267)5612: 585–90.
- [94] J. Zhao et al. *Chemical Society Reviews* (2013); (42)12: 5323–5351.
- [95] M. O. Steinhauser and S. Hiermaier. *International Journal of Molecular Sciences* (2009); (10)12: 5135–5216.
- [96] S. Kumar et al. *Journal of Computational Chemistry* (1992); (13)8: 1011–1021.
- [97] A. Laio and M. Parrinello. *Proceedings of the National Academy of Sciences of the United States of America* (2002); (99)20: 12562–12566.
- [98] P. Hohenberg and W. Kohn. *Physical Review* (1964); (136)3B: 864–871.
- [99] E. Matito et al. (2016), arXiv:1608.02553.
- [100] D. M. Ceperley and B. J. Alder. *Physical Review Letters* (1980); (45)7: 566–569.
- [101] A. D. Becke. *Journal of Chemical Physics* (1993); (98)5648: 5648–5652.
- [102] C. Adamo, G. E. Scuseria, and V. Barone. *The Journal of Chemical Physics* (1999); (110)13: 6158–6170.
- [103] C. Lee, W. Yang, and R. G. Parr. *Physical Review. B, Condensed Matter* (1988); (37)2: 785–789.
- [104] E. G. Hohenstein, S. T. Chill, and C. D. Sherrill. *Journal of Chemical Theory and Computation* (2008); (4)12: 1996–2000.
- [105] Y. Zhao and D. G. Truhlar. *Theoretical Chemistry Accounts* (2008); (120)1-3: 215–241.
- [106] P. Dirac. *The Principles of Quantum Mechanics* (1930), Oxford University Press.

- [107] E. Runge and E. K. U. Gross. *Physical Review Letters* (1984); (52)12: 997–1000.
- [108] F. C. Spano et al. *Journal of the American Chemical Society* (2007); (129)22: 7044–7054.
- [109] V. I. Prokhorenko, D. B. Steensgaard, and A. R. Holzwarth. *Biophysical Journal* (2003); (85)5: 3173–3186.
- [110] A. Warshel and M. Levitt. *Journal of Molecular Biology* (1976); (103)2: 227–249.
- [111] A. Monari, J. L. Rivail, and X. Assfeld. *Accounts of Chemical Research* (2013); (46)2: 596–603.
- [112] Y. L. Lyubchenko. *Cell Biochemistry and Biophysics* (2004); (41)1: 75–98.
- [113] R. Galindo-Murillo and I. Cheatham Thomas E. *ChemMedChem* (2014); (9)6: 1252–1259.
- [114] B. Heddi and P. Anh Tuan. *Journal of the American Chemical Society* (2011); (133)25: 9824–9833.
- [115] R. Lavery et al. *Nucleic Acids Research* (2010); (38)1: 299–313.
- [116] A. Perez et al. *Nucleic Acids Research* (2008); (36)7: 2379–2394.
- [117] A. T. Phan and J.-L. Mergny. *Nucleic acids research* (2002); (30)21: 4618–4625.
- [118] F. Dehez et al. *Nucleic Acids Research* (2017); (45)7: 3654–3662.
- [119] D. M. Gray, A. R. Morgan, and R. L. Ratliff. *Nucleic Acids Research* (1978); (5)10: 3679–3695.
- [120] L. M. Nielsen, S. V. Hoffmann, and S. B. Nielsen. *Photochemical and Photobiological Sciences* (2013); (12)8: 1273–1285.
- [121] M. Vorlickova et al. *Chirality* (2012); (24)9: 691–698.
- [122] E. Dumont and A. Monari. *Journal of Physical Chemistry Letters* (2013); (4)23: 4119–4124.
- [123] E. Dumont et al. *Journal of Physical Chemistry Letters* (2015); (6)4: 576–580.
- [124] F. Di Meo et al. *Journal of Physical Chemistry Letters* (2015); (6)3: 355–359.
- [125] M. Kubista, B. Akerman, and B. Norden. *Journal of Physical Chemistry* (1988); (92)8: 2352–2356.
- [126] D. Loco et al. *Physical Chemistry Chemical Physics* (2016); (18)2: 866–877.

- [127] T. Very et al. *Physical Chemistry Chemical Physics* (2012); (14)36: 12496–12504.
- [128] T. Very et al. *Chemistry - a European Journal* (2014); (20)40: 12901–12909.
- [129] T. Etienne et al. *Journal of Physical Chemistry B* (2013); (117)17: 4973–4980.
- [130] M. Huix-Rotllant et al. *Photochemistry and Photobiology* (2015); (91)2: 323–330.
- [131] B. Nieto-Ortega et al. *Chemical Communications* (2012); (48)73: 9147–9149.
- [132] D. Case et al. *AMBER2015* (2015).
- [133] A. Pérez et al. *Biophysical Journal* (2007); (92)11: 3817–3829.
- [134] N. Ferré and X. Assfeld. *Journal of Chemical Physics* (2002); (117)9: 4119–4125.
- [135] R. Krishnan et al. *Journal of Chemical Physics* (1980); (72)1: 650–654.
- [136] M. J. Frisch et al. (2009). Gaussian Inc. Wallingford CT 2016.
- [137] J. W. Ponder and F. M. Richards. *Journal of Computational Chemistry* (1987); (8) 1016–1024.
- [138] M. Gellert, M. Lipsett, and D. Davies. *Proceedings of the National Academy of Sciences of the United States of America* (1962); (15)48: 2013–2018.
- [139] G. W. Collie and G. N. Parkinson. *Chemical Society Reviews* (2011); (40)12: 5867–5892.
- [140] S. Balasubramanian, L. H. Hurley, and S. Neidle. *Nature Reviews Drug Discovery* (2011); (10)4: 261–275.
- [141] H. J. Lipps and D. Rhodes. *Trends in Cell Biology* (2009); (19)8: 414–422.
- [142] J. L. Huppert. *Chemical Society Reviews* (2008); (37)7: 1375–1384.
- [143] T.-m. Ou et al. *ChemMedChem* (2008); (3)5: 690–713.
- [144] J. R. Williamson, M. K. Raghuraman, and T. R. Cech. *Cell* (1989); (59)5: 871–880.
- [145] A. E. Bergues-Pupo et al. *Nucleic Acids Research* (2015); (43)15: 7638.
- [146] S. Neidle and G. N. Parkinson. *Current Opinion in Structural Biology* (2003); (13)3: 275–283.
- [147] X. Zhang et al. *Genes and Development* (1999); (13) 2388–2399.
- [148] S. Neidle and G. Parkinson. *Nature Reviews Drug Discovery* (2002); (1)5: 383–393.

- [149] A. Terenzi et al. *RSC Advances* (2014); (4)63: 33245.
- [150] A. Biancardi et al. *Chemistry* (2014); (20)24: 7439–7447.
- [151] M. Yin et al. *Chemical Communications* (2012); (48)52: 6556–6558.
- [152] Q. Yuan et al. *Angewandte Chemie - International Edition* (2013); (52)52: 13965–13969.
- [153] S. Alcaro et al. *Journal of Medicinal Chemistry* (2013); (56)3: 843–855.
- [154] A. Spinello, G. Barone, and J. Grunenberg. *Physical Chemistry Chemical Physics* (2016); (18)4: 2871–2877.
- [155] C. C. Hardin et al. *Biochemistry* (1992); (31)3: 833–841.
- [156] H. Gattuso, X. Assfeld, and A. Monari. *Theoretical Chemistry Accounts* (2015); (134)3:
- [157] Y. Xue et al. *Journal of the American Chemical Society* (2007); (129)36: 11185–11191.
- [158] C. Chipot and A. Pohorille. *Free energy calculations* (2007), Springer.
- [159] L. Landau and E. Lifshitz. *Statistical Physics* (1938); (120) Oxford Univ. Press, New York.
- [160] J. Kirkwood. *Journal of Chemical Physics* (1935); (3) 300–313.
- [161] S. Singh, D. Wemmer, and P. Kollman. *Proceedings of the National Academy of Sciences of the United States of America* (1994); (3) 7673–7677.
- [162] J. Dolenc et al. *Nucleic Acids Research* (2005); (32) 725–733.
- [163] H. Woo and B. Roux. *Proceedings of the National Academy of Sciences of the United States of America* (2005); (102)19: 6825–6830.
- [164] J. C. Gumbart, B. Roux, and C. Chipot. *Journal of Chemical Theory and Computation* (2012); (9)8: 3789–3798.
- [165] J. C. Gumbart, B. Roux, and C. Chipot. *Journal of Chemical Theory and Computation* (2013); (9)1: 794–802.
- [166] B. M. Zeglis, V. C. Pierre, and J. K. Barton. *Chemical Communications* (2007); 44: 4565–4579.
- [167] W. D. Cornell et al. *Journal of the American Chemical Society* (1995); (117)19: 5179–5197.
- [168] W. L. Jorgensen et al. *The Journal of Chemical Physics* (1983); (79)2: 926.
- [169] J. Phillips et al. *Journal of Computational Chemistry* (2005); (26)16: 1781–1802.

- [170] S. Feller et al. *Journal of Chemical Physics* (1995); (103)11: 4613–4621.
- [171] T. Darden, D. York, and L. Pedersen. *Journal of Chemical Physics* (1993); (98)12: 10089–10092.
- [172] G. Fiorin, M. L. Klein, and J. Henin. *Molecular Physics* (2013); (111)22-23: 3345–3362.
- [173] S. B. Dixit and C. Chipot. *Journal of Physical Chemistry A* (2001); (105)42: 9795–9799.
- [174] J. Hermans and S. Shankar. *Israel Journal of Chemistry* (1986); (27)2: 225–227.
- [175] M. K. Gilson et al. *Biophysical Journal* (1997); (72)3: 1047–1069.
- [176] P. Liu et al. *Journal of Chemical Theory and Computation* (2012); (8)8: 2606–2616.
- [177] W. Humphrey, A. Dalke, and K. Schulten. *Journal of Molecular Graphics* (1996); (14)1: 33–38.
- [178] Y. Yu, P. Privalov, and R. Hodges. *Biophysical Journal* (2001); (81)3: 1632–1642.
- [179] A. V. Vargiu et al. *Biophysical Journal* (2008); (94)2: 550–561.
- [180] M. Marazzi, H. Gattuso, and A. Monari. *Theoretical Chemistry Accounts* (2016); (135)3: 1–11.
- [181] P. Miro et al. *Chemistry - A European Journal* (2015); (21)47: 17051–17056.
- [182] S. Shaikh, S. Ahmed, and B. Jayaram. *Archives of Biochemistry and Biophysics* (2004); (429)1: 81–99.
- [183] L. Favero, G. Granucci, and M. Persico. *Physical Chemistry Chemical Physics* (2016); (18)15: 10499–10506.
- [184] M. C. Cuquerella et al. *Accounts of Chemical Research* (2012); (45)9: 1558–1570.
- [185] T. Delatour et al. *Journal of Photochemistry and Photobiology B: Biology* (1998); (44)3: 191–198.
- [186] D.-C. Sergentu et al. *Physical Chemistry Chemical Physics* (2014); (16)46: 25393–25403.
- [187] M. A. El-Sayed and R. Leyerle. *The Journal of Chemical Physics* (1975); (62)4: 1579–1580.
- [188] S. Yabumoto, S. Sato, and H. Hamaguchi. *Chemical Physics Letters* (2005); (416)1-3: 100–103.

- [189] M. Batley and D. R. Kearns. *Chemical Physics Letters* (1968); (2)6: 423–425.
- [190] Y. Matsushita, Y. Kajii, and K. Obi. *Journal of Physical Chemistry* (1992); (96)11: 4455–4458.
- [191] F. Weigend and R. Ahlrichs. *Physical Chemistry Chemical Physics* (2005); (7)18: 3297–3305.
- [192] J. Tomasi, B. Mennucci, and R. Cammi. *Chemical Reviews* (2005); (105)8: 2999–3093.
- [193] K. Andersson et al. *Journal of Physical Chemistry* (1990); (94)14: 5483–5488.
- [194] K. Andersson, P. Malmqvist, and B. O. Roos. *Journal of Chemical Physics* (1992); (96)2: 1218.
- [195] P. Malmqvist et al. *The Journal of Chemical Physics* (2008); (128)20: 204109.
- [196] V. Sauri et al. *Journal of Chemical Theory and Computation* (2011); (7)1: 153–168.
- [197] F. Aquilante et al. *Journal of Computational Chemistry* (2016); (37)5: 506–541.
- [198] W. D. Cornell et al. *Journal of the American Chemical Society* (1993); (115)7: 9620–9631.
- [199] M. Fagnoni et al. *Chemical Reviews* (2007); (107)6: 2725–2756.
- [200] T. Bach and J. P. Hehn. *Angewandte Chemie - International Edition* (2011); (50)5: 1000–1045.
- [201] J. B. Xia, C. Zhu, and C. Chen. *Journal of the American Chemical Society* (2013); (135)46: 17494–17500.
- [202] Z. Li and K. B. Grant. *RSC Adv.* (2016); (6)29: 24617–24634.
- [203] J. M. Brown. In: (435) 2007, Elsevier.
- [204] P. Greenspan and S. D. Fowler. *Journal of Lipid Research* (1985); (26)7: 781–789.
- [205] H. Ju, Y. Ye, and Y. Zhu. *Electrochimica Acta* (2005); (50)6: 1361–1367.
- [206] K. Hirakawa et al. *Chemical Research in Toxicology* (2014); (27)4: 649–655.
- [207] S. Fukuzumi et al. *Journal of Physical Chemistry A* (2005); (109)15: 3285–3294.
- [208] I. Saito et al. *Journal of the American Chemical Society* (1995); (117)26: 6406–6407.
- [209] Y. Zaho and D. Truhlar. *Theoretical Chemistry Accounts* (2008); (120) 215–241.

- [210] T. Etienne, X. Assfeld, and A. Monari. *Journal of Chemical Theory and Computation* (2014); (10)9: 3906–3914.
- [211] R. K. Mitra et al. *Journal of Fluorescence* (2009); (19)2: 353–361.
- [212] A. Kumar et al. *Journal of Physical Chemistry B* (2016); (120)9: 2115–2123.
- [213] M. McAllister et al. *Journal of Physical Chemistry Letters* (2015); (6)15: 3091–3097.
- [214] K. R. Siefermann et al. *Nature Chemistry* (2010); (2)4: 274–279.
- [215] P. Salek et al. *Journal of Chemical Physics* (2002); (117)21: 9630–9645.
- [216] K. Aidas et al. *Wiley Interdisciplinary Reviews: Computational Molecular Science* (2014); (4)3: 269–284.
- [217] R. L. Martin. *Journal of Chemical Physics* (2003); (118)11: 4775–4777.
- [218] T. Etienne. *Journal of Chemical Theory and Computation* (2015); (11)4: 1692–1699.
- [219] T. Etienne, X. Assfeld, and A. Monari. *Journal of Chemical Theory and Computation* (2014); (10)9: 3896–3905.
- [220] A. H. Steindal et al. *Physical Chemistry Chemical Physics* (2012); (14)16: 5440–5451.
- [221] F. A. Tanious et al. *Biochemistry* (1997); (36)49: 15315–15325.
- [222] E. De Meulenaere et al. *Chemical Science* (2012); (3)4: 984.
- [223] F. A. Tanious et al. *European Journal of Biochemistry* (2001); (268)12: 3455–3464.
- [224] L. Chomicz et al. *Physical Chemistry Chemical Physics* (2014); (16)14: 6568–6574.
- [225] R. Barrios, P. Skurski, and J. Simons. *Journal of Physical Chemistry B* (2002); (106)33: 7991–7994.
- [226] N. Lin et al. *ChemPhysChem* (2011); (12)17: 3392–3403.
- [227] V. Vendrell-Criado et al. *Journal of the American Chemical Society* (2013); (135)44: 16714–16719.
- [228] T. Douki et al. *Organic Letters* (2015); (17)2: 246–249.
- [229] T. Douki et al. *Chemistry* (2014); (20)19: 5787–5794.
- [230] F. Bosca et al. *Journal of the American Chemical Society* (2006); (128)19: 6318–6319.
- [231] H. Wang, E. Rayburn, and R. Zhang. *Current Pharmaceutical Design* (2005); (11)22: 2889–2907.

- [232] H. Rappaport. *Nucleic Acids Research* (1988); (16)15: 7253–7267.
- [233] B. A. Connolly and P. C. Newman. *Nucleic Acids Research* (1989); (17)13: 4957–4974.
- [234] B. Gildea and L. McLaughlin. *Nucleic Acids Research* (1989); (17)6: 2261–2281.
- [235] P. G. Wu and L. Brand. *Analytical Biochemistry* (1994); (218)1: 1–13.
- [236] G. Ryseck et al. *ChemPhysChem* (2011); (12)10: 1880–1888.
- [237] G. Ryseck et al. *Photochemical and Photobiological Sciences* (2013); (12)8: 1423–1430.
- [238] B. Demple and J. S. Sung. *DNA Repair* (2005); (4)12: 1442–1449.
- [239] K. S. Gates. *Chemical Research in Toxicology* (2009); (22)11: 1747–1760.
- [240] C. Clauson, O. D. Schärer, and L. Niedernhofer. *Cold Spring Harbor Perspectives in Biology* (2013); (5)10:
- [241] A. G. Georgakilas, P. O’Neill, and R. D. Stewart. *Radiation Research* (2013); (180)1: 100–109.
- [242] Z. Liu et al. *PLoS ONE* (2015); (10)9: e0137124.
- [243] Z. Nikitaki et al. *Frontiers in Chemistry* (2015); (3) 35.
- [244] N. Shikazono et al. *Mutation Research* (2013); (749)1-2: 9–15.
- [245] M. E. Lomax, S. Cuniffe, and P. O’Neill. *Biochemistry* (2004); (43)34: 11017–11026.
- [246] A. G. Georgakilas, P. V. Bennett, and B. M. Sutherland. *Nucleic Acids Research* (2002); (30)13: 2800–2808.
- [247] A. G. Georgakilas et al. *Nucleic Acids Research* (2004); (32)18: 5609–5620.
- [248] R. D. Hazel, K. Tian, and C. de Los Santos. *Biochemistry* (2008); (47)46: 11909–11919.
- [249] L. Ayadi, C. Coulombeau, and R. Lavery. *Journal of Biomolecular Structure and Dynamic* (2000); (17)4: 645–653.
- [250] L. Ayadi, C. Coulombeau, and R. Lavery. *Biophysical Journal* (1999); (77)6: 3218–3826.
- [251] D. Barsky et al. *Nucleic Acids Research* (2000); (28)13: 2613–2626.
- [252] H. Mu et al. *Nucleic Acids Research* (2012); (40)19: 9675–9690.
- [253] C. D. M. Churchill, L. A. Eriksson, and S. D. Wetmore. *Journal of Physical Chemistry B* (2016); (120)7: 1195–1204.



- [254] E. Bignon et al. *Chemistry* (2015); (21)32: 11509–11516.
- [255] P. Mark and L. Nilsson. *Journal of Physical Chemistry A* (2001); (105)43: 9954–9960.
- [256] C. De Los Santos et al. *Biochemistry* (2004); (43)49: 15349–15357.
- [257] J. Zalesak et al. *Journal of Molecular Biology* (2014); (426)7: 1524–1538.
- [258] B. M. Zeglis, J. A. Boland, and J. K. Barton. *Biochemistry* (2009); (48)5: 839–849.
- [259] X. Shi et al. *Proceedings of the National Academy of Sciences of the United States of America* (2014); (111)15: 1473–1480.
- [260] M. Li et al. *Journal of Molecular Biology* (2014); (426)11: 2183–2198.
- [261] J. Chiba et al. *Chemical Communications* (2014); (50)76: 11126–11128.
- [262] S. Cunniffe et al. *Mutation Research* (2014); (762)1: 32–39.
- [263] T. A. Kunkel, R. M. Schaaper, and L. A. Loeb. *Biochemistry* (1983); (22)10: 2378–84.
- [264] M. E. Lomax, L. K. Folkes, and P. O'Neill. *Clinical Oncology* (2013); (25)10: 578–585.
- [265] E. Sage and L. Harrison. *Mutation Research* (2011); (711)1-2: 123–133.
- [266] L. J. Eccles, P. O'Neill, and M. E. Lomax. *Mutation Research* (2011); (711)1-2: 134–141.
- [267] S. Malyarchuk, R. Castore, and L. Harrison. *DNA Repair* (2009); (8)12: 1343–1354.
- [268] G. Eot-Houllier et al. *Nucleic Acids Research* (2005); (33)1: 260–271.
- [269] Y. Sedletska, J. P. Radicella, and E. Sage. *Nucleic Acids Research* (2013); (41)20: 9339–9348.
- [270] S. G. Kozmin et al. *Nucleic Acids Research* (2009); (37)6: 1767–1777.
- [271] J. Zalesak, J. F. Constant, and M. Jourdan. *Biochemistry* (2016); (55)28: 3899–3906.
- [272] D. J. Hosfield et al. *Cell* (1999); (98)3: 397–408.
- [273] S. Ghosh and M. M. Greenberg. *Biochemistry* (2015); (54)40: 6274–6283.
- [274] C. A. Gelfand et al. *Biochemistry* (1998); (37)20: 7321–7327.
- [275] D. Lu et al. *Proceedings of the National Academy of Sciences* (2012); (109)42: 16852–16857.

- [276] B. D. Freudenthal et al. *Nature Structural and Molecular Biology* (2015); (11)22: 924–931.
- [277] N. G. Beloglazova et al. *Nucleic Acids Research* (2004); (32)17: 5134–5146.
- [278] A. Mazouzi et al. *Proceedings of the National Academy of Sciences of the United States of America* (2013); (110)33: 3071–3080.
- [279] Z. Lin et al. *Nucleic Acids Research* (1998); (26)10: 2385–2391.
- [280] E. Bignon et al. *Nucleic Acids Research* (2016); (44)18: 8588–8599.
- [281] R. Al-Safi et al. *Current Molecular Pharmacology* (2012); (5)1: 14–35.
- [282] M. Poletto et al. *Molecular Carcinogenesis* (2016); (55)5: 688–704.
- [283] D. L. Mitchell and R. S. Nairn. *Photochemistry and Photobiology* (1989); (49)6: 805–819.
- [284] A. S. Balajee, A. May, and V. A. Bohr. *Nucleic Acids Research* (1999); (27)12: 2511–2520.
- [285] D. Perdiz et al. *Journal of Biological Chemistry* (2000); (275)35: 26732–26742.
- [286] J. A. Martejn et al. *Nature Reviews Molecular Cell Biology* (2014); (15)7: 465–481.
- [287] O. D. Schärer. *Cold Spring Harbor Perspectives in Biology* (2013); (5)10:
- [288] C. Han et al. *Carcinogenesis* (2015); (37)2: 129–138.
- [289] H. Gattuso et al. *Physical Chemistry Chemical Physics* (2016); (18)48: 33180–33186.
- [290] A. Osakabe et al. *Scientific Reports* (2015); (5) 16330.
- [291] J. Li et al. *Nature* (2010); (466)7308: 887–890.
- [292] J. K. Kim and B. S. Choi. *European Journal of Biochemistry* (1995); (228)3: 849–854.
- [293] J. K. Kim, D. Patel, and B. S. Choi. *Photochemistry and Photobiology* (1995); (62)1: 44–50.
- [294] J. H. Lee, G. S. Hwang, and B. S. Choi. *Proceedings of the National Academy of Sciences of the United States of America* (1999); (96)12: 6632–6636.
- [295] T. Mizukoshi et al. *Nucleic Acids Research* (2001); (29)24: 4948–4954.
- [296] T. I. Spector, T. E. Cheatham, and P. A. Kollman. *Journal of the American Chemical Society* (1997); (119)30: 7095–7104.
- [297] C. Wang and J. Taylor. *Proceedings of the National Academy of Sciences of the United States of America* (1991); (88) 9072–9076.

- [298] M. G. Cooney and J. H. Miller. *Nucleic Acids Research* (1997); (25)7: 1432–1436.
- [299] W. Yang. *Protein Science* (2011); (20)11: 1781–1789.
- [300] I. Ivani et al. *Nature Methods* (2016); (13)1: 55–58.
- [301] A. Shrake and J. A. Rupley. *Journal of Molecular Biology* (1973); (79)2: 351–371.
- [302] Y. Sugita<sup>1</sup> and Y. Okamoto<sup>a</sup>. *Chemical Physics Letters* (1999); (314)1-2: 141–151.
- [303] M. Pasi et al. *Nucleic Acids Research* (2014); (42)19: 12272–12283.
- [304] S. Cogoi et al. *Nucleic Acids Research* (2008); (36)11: 3765–3780.
- [305] T. I. Gaynutdinov, R. D. Neumann, and I. G. Panyutin. *Nucleic Acids Research* (2008); (36)12: 4079–4087.
- [306] J. E. Sale, A. R. Lehmann, and R. Woodgate. *Nature Reviews: Molecular cell biology* (2012); (13)3: 141–152.
- [307] S. Courdavault et al. *DNA Repair* (2005); (4)7: 836–844.
- [308] Y. Kuang et al. *Chemistry - A European Journal* (2012); (18)40: 12609–12613.
- [309] A. Gentil et al. *Nucleic Acids Research* (1996); (24)10: 1837–1840.

**Résumé** Le travail présenté dans ce manuscrit est basé sur l'utilisation de la modélisation moléculaire, de la simulation et de la chimie théorique pour l'étude de la **photosensibilisation de l'ADN** ; c'est à dire l'augmentation de la sensibilité de l'ADN vis-à-vis de la lumière au travers de l'action d'agents photosensibilisants. Premièrement, les voies photophysiques et photochimiques de plusieurs molécules connues telles que nile bleu, nile rouge, BMEMC ou une base modifiée endogène, Pyo, ont été étudiés dans le but de comprendre leurs mécanismes de photosensibilisation. Les phénomènes associés qui ont été mis en évidence sont des transferts d'électrons et d'énergie, la production d'électrons solvatés et l'activation à deux photons. De plus, deux outils pour l'étude des interactions entre les molécules et l'ADN ont été développés; i) un protocole calculatoire capable de fournir l'énergie libre d'interaction de drogues dans leurs poches; ii) un outil basé sur l'hamiltonien semi-empirique de Frenkel qui permet de modéliser le spectre de dichroïsme circulaire électronique de biomacromolécules. Ensuite, les effets de photolésions sur la structure et la flexibilité de l'ADN ont été étudiés ; i.e. les dimères de pyrimidines, la pyrimidine(6-4)pyrimidone (6-4PP) et les clusters de sites abasiques. Finalement, la reconnaissance de brins d'ADN lésés par des protéines de réparation et le rapport avec leurs activités enzymatiques a été analysé. Le lecteur peut se référer à la première partie de ce manuscrit pour une présentation vulgarisée du contexte de ce projet.

**Abstract** The work presented in this manuscript is based on the use of molecular modeling, simulation and theoretical chemistry in order to study the **photosensitization of DNA**; i.e. the enhancement of the sensitivity of DNA to light through the action of a photosensitizing agent. A first aspect has been to study the photophysical and photochemical pathways of several known sensitizers such as nileblue, nilered, BMEMC or an endogenous modified nucleobase, Pyo, in order to understand their mechanisms of photosensitization. The related phenomena that have been observed are electron transfers, triplet-triplet energy transfers, production of solvated electrons and two-photons activations. Moreover, two tools have been developed to study the interaction between photosensitizing agents and DNA; i) a protocol able to provide the binding free energy of drugs in their interaction pockets; ii) a tool based on the semi-empirical Frenkel Hamiltonian to model the electronic circular dichroism of biomacromolecular systems in a straightforward way. Then the effects of photoinduced lesions on the DNA structure and flexibility have been investigated; i.e. cyclopurine dimers (CPD), pyrimidine(6-4)pyrimidone (6-4PP) and cluster abasic sites. Finally the recognition of damaged DNA strands by repair enzymes is presented and the implication on enzymatic activities has been highlighted. The reader can refer to the first section of the manuscript for a popularized presentation of the project context.