Theoretical approach of complex DNA lesions : from formation to repair

Emmanuelle Bignon

N° d'ordre NNT : 2017LYSE1085

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de

l'Université Claude Bernard Lyon 1

École Doctorale de Chimie de Lyon ED206
Chimie, Procédés, Environnement

Spécialité de doctorat : Chimie

Soutenue publiquement le 08/06/2017, par :

Emmanuelle Bignon

Theoretical approach of complex DNA lesions : from formation to repair.

Devant le jury composé de :

Papaleo Elena, Senior Researcher, Danish Cancer Society Research Center	Rapportrice
Derreumaux Philippe, Professeur, Université Paris Diderot	Rapporteur
Antonczak Serge, Professeur, Université de Nice Sophia Antipolis	Examinateur
Douki Thierry, Directeur de Recherche, CEA de Grenoble	Examinateur
Jamet Hélène, Maitre de Conférences, Université Joseph Fourier Grenoble	Examinatrice
Lesage Anne, Ingénieure de Recherche, Université de Lyon	Examinatrice
Morell Christophe, Professeur, Université de Lyon	Directeur de thèse
Dumont Elise, Maitre de Conférences, ENS de Lyon	Co-Directrice de thèse

UNIVERSITE CLAUDE BERNARD-LYON 1

Président de l'Université	**M. le Professeur Frédéric FLEURY**
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLEE
Directrice Générale des Services	Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur G. RODE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. X. PERROT
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologie	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E.PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

You look at science (or at least talk of it) as some sort of demoralising invention of man, something apart from real life, and which must be cautiously guarded and kept separate from everyday existence. But science and everyday life cannot and should not be separated.

Rosalind Franklin

Dreams, indeed, are ambition; for the very substance of the ambitious is merely the shadow of a dream.

William Shakespeare, Hamlet

Remerciements

Je souhaite en premier lieu remercier chacun des membres du jury pour m'avoir fait l'honneur d'accepter d'évaluer ces travaux de thèse. En tant que rapporteurs, Dr. Elena Papaleo et Pr. Philippe Derreumaux, en tant que président du jury le Pr. Serge Antonczak, et en tant qu'examinateurs Dr. Thierry Douki, Dr. Hélène Jamet, et Dr. Anne Lesage.
Je remercie de tout coeur mes deux encadrants, Pr. Christophe Morell et Pr. Elise Dumont, pour m'avoir guidée durant ces trois années. Je suis entièrement reconnaissante de leur soutien et leur patience, et suis d'autant plus heureuse d'avoir rencontré là deux personnalités aussi agréables qu'expertes dans leur domaine. Merci de m'avoir fait confiance et envoyée aux quatre coins du monde pour vivre des expériences plus qu'enrichissantes !

Un grand merci aux chercheur(e)s avec qui j'ai pu collaborer tout au long de cette thèse, Jean-Luc, Antonio, François, Hugo, Filip... Ce fut un immense plaisir de travailler avec vous et de partager des moments plus que sympathiques au labo et en congrès.

A mes co-bureaux géniaux de l'ENS et l'ISA, Raymond, Martin, Chen, Manon, Walid, et Alexandra pour avoir mis de la bonne humeur et un surtout une bonne dose d'humour dans mon quotidien.

A mes amis du labo. Romain, pour les bonnes tranches de rire, les danses endiablées, la banane, l'accent écossais et touti quanti comme qui dirait muchachos. Fred, pour tes connaissances toujours des plus surprenantes, les soirées à la Migraine (le traquenard !!), ta collection de canards vivants, et ton aide précieuse. S'agissant d'aide précieuse, un grand merci également à Valérian, du premier jour à la soutenance tu m'as aidée sur tous les plans et je t'en suis plus que reconnaissante.

A mes stagiaires préférés, devrais-je dire à mes amis, Pierre et Victor. Pierre, merci pour ton enthousiasme sans borne, il est plus qu'agréable de te côtoyer, au labo comme en dehors. Ce fut un plaisir de t'avoir comme tout premier stagiaire, tu as mis la barre haute! Victor, should I call you my third year partner in crime? I am really glad that we met. You brought me the joy and crazyness that I really needed at the end of my PhD. Thank you for all of that, and go easy on the 'you have to', ok?

Un paragraphe tout spécial pour mes amis du premier jour ! Mathilde, c'est avec toi que j'ai pris mon premier apéro lyonnais, suivi de beaucoup d'autres. Les ragots, les fous rire, les soirées, un milliers de souvenirs avec toi, figés dans un très grand nombre de photos dossier. Thomas, merci pour tout ce temps passé ensemble, les têtes de débiles, le verre dans la bouche, les festochs, les discussions jusqu'à tard, et tant d'autres bons souvenirs. Kevin, pour ton rosé qui pique aux parties de Molkki, pour les soirées dans

ton appart au 5ème étage sans escalier, pour tous les bons délires. Chris, la tequila du Johnny's aura eu raison de nous pauvres pécheurs, et je suis intimement convaincue que c'est une sacrée prise...
Merci à vous quatre pour tous ces délires et tout ce temps passé ensemble au Johnny's et ailleurs.

To my dear friends of the TCCM summer and winter schools. I had so amazing time with you all, it will always be great to see you again. Ozge, you are just perfectly crazy.

Aux vieux et moins vieux de la Dem'. J'aurais aimé plus m'impliquer dans cette aventure géniale, les moments passés avec vous sont toujours au top. Je ne vois plus le fromage rapé du même oeil grâce à vous. Longue vie à DéMesures !

Aux membres de l'équipe modélisation de Valrose. C'est avec vous que tout a commencé et je ne vous en remercierai jamais assez. Vous êtes les plus cool.

A ma famille et amis pour leur soutien.

Contents

Introduction (French)

L'ADN est la molécule responsable du transport de l'information génétique des êtres vivants. Il est constitué de quatre briques élémentaires différentes, dénommées nucléotides. Ces derniers sont eux-mêmes composés d'un squelette phosphate/désoxyribose sur lequel est relié par liaison N-glycosidique l'une des quatre nucléobases canoniques : adenine (A), cytosine (C), thymine (T), ou guanine (G). Ces nucléobases peuvent être classées en deux sous familles : les pyrimidines, qui correspondent aux nucléobases monocycliques thymine et cytosine, et les purines qui représentent celles bicycliques, adénine et guanine. Leurs interactions mutuelles sont responsables de la structure d'ADN en hélice double-brin - voir Figure i. Cette géométrie particulière a été publiée pour la première fois par James Watson et Francis Crick en 1953, sur la base de clichés de diffraction aux rayons X obtenues par la crystallographe Rosalind Franklin[1]. Elle repose sur l'appariement de type Watson-Crick entre les différentes nucléobases des deux brins complémentaires, avec des interactions spécifiques A-T et G-C, et sur de fortes interactions π dûes à l'empilement des cycles conjugués des nucléobases. Toutefois, ce réseau d'interactions n'empêche en rien la séparation des brins, nécessaire durant la phase de réplication du matériel génétique. De façon intéressante, le ratio purines/pyrimidines dans les cellules des organismes vivants s'élève à 1:1 selon la règle de Chargaff[2–4].

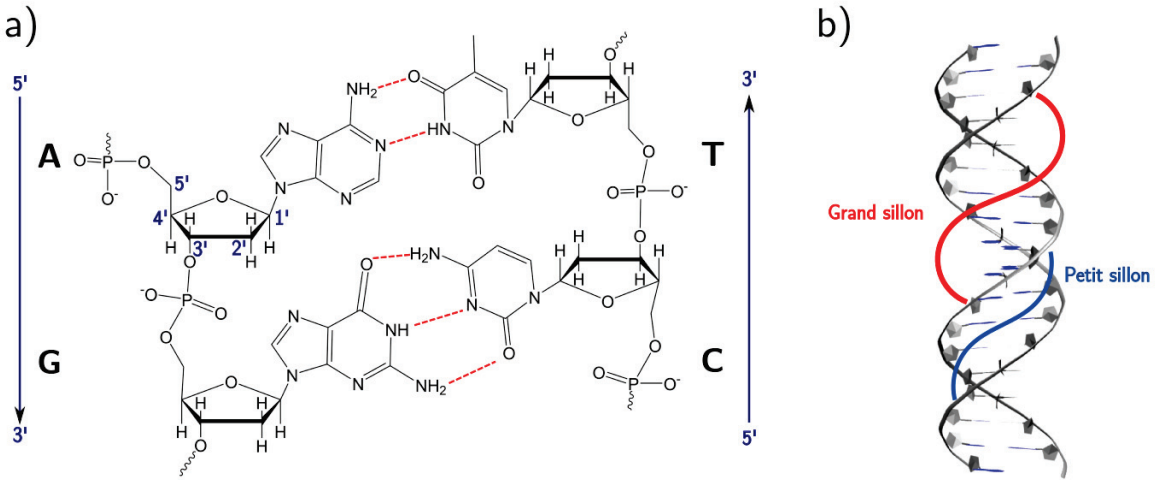


FIGURE I: a) Représentation schématique des quatres nucléotides canoniques de l'ADN dans un duplex d(AG):d(CT). Les liaisons hydrogène de type Watson-Crick sont matérialisées par des lignes pointillées rouges. L'adénine forme deux liaisons hydrogènes avec la thymine, tandis que la guanine interagit avec la cytosine par trois de ces mêmes liaisons. La numérotation classique des carbones du désoxyribose est spécifiée en bleu. Les brins d'ADN sont communément ordonnés dans le sens 5' $\rightarrow$ 3', représenté par des flèches bleues. b) Représentation 'ruban' d'une hélice d'ADN B classique. Le squelette phosphate/désoxyribose apparaît en gris tandis que les nucléobases sont représentées en bleu. Le petit et le grand sillons sont facilement distinguables l'un de l'autre par leur apparente différence de largeur.

La succession des $\sim$ 3 x 10^9 nucléotides présents dans les cellules humaines[5] constitue le code génétique. Ce dernier est la base nécessaire à la synthèse des protéines, celles-ci prenant part à l'intégralité des activités cellulaires indispensables à la survie de l'organisme. De fait, les triplets de nucléotides, dénommés codons, correspondent chacun à un acide aminé bien précis[6–8]. Ainsi, leur succession particulière détermine la séquence primaire des protéines synthétisées lors de l'expression génétique. De façon surprenante, environ la moitié du génome est constitué de séquences d'ADN répétitives, autrefois appelée "ADN poubelle", qui sont en réalité d'une importance capitale pour le contrôle de la synthèse des protéines[9,10].
En réalité, celles-ci ne sont pas générées directement à partir du code génétique. Premièrement, une molécule d'ARN messager (ARNm) est construite par complémentarité aux brins d'ADN. Les protéines sont ensuite synthétisées dans une partie spécifique de la cellule, le ribosome[11,12], lui-même sous-divisé en deux compartiments. Le plus petit de ces deux sous-unités sert à la lecture de l'ARNm, tandis que les acides aminés correspondant au code seront assemblés dans un plus grand compartiment. Ces derniers sont d'ailleurs transportés au sein du ribosome par l'ARN de transfert (ARNt). Ainsi, l'ARNt assure la correspondance entre l'information génétique délivrée par l'ARNm et la séquence de la protéine codée. Ce phénomène d'expression génétique constitue un processus fondamental pour les cellules. De ce fait, les erreurs qui peuvent survenir lors de son déroulement peuvent entraîner des conséquences délétères pour la cellule, donc l'organisme vivant. Notamment, les modifications du code génétique porté par l'ADN (mutations génétiques), le blocage de la réplication de l'ADN ou de l'expression génétique sont sources de cancérogenèse, apoptose (mort programmée des cellules), et de vieillissement. Or, les composants de l'ADN subissent les attaques incessantes d'agents mutagènes comme la lumière du soleil (notamment les UVB et UVA courts), les espèces réactives de l'oxygène engendrées par le stress oxydant ($^1O_2$,$HOO^{\cdot}$,H_2O_2,$HO^{\cdot}$...), etc. qui mettent en péril l'intégrité de l'expression génétique. Chaque jour, ce sont par cellule 10000 à 1 million de lésions qui sont générées dû à des facteurs environnementaux, et même au métabolisme cellulaire. Fort heureusement, les cellules contiennent une efficace machinerie de réparation de l'ADN, qui permet d'éliminer les dommages par des processus très complexes faisant intervenir de nombreuses enzymes[13]. Les processus de réparation sont initiés par la détection de la lésion, elle-même assurée par des enzymes bien spécifiques, capables de reconnaître les caractéristiques structurales et/ou chimiques des nucléotides endommagés. Ainsi, les modifications des propriétés structurales et chimiques induites par la présence d'un ou de plusieurs dommages est un point majeur dans la compréhension des mécanismes de leur réparation. De même, l'élucidation des mécanismes réactionnels qui gouvernent la formation de ces lésions est une question importante à laquelle les chercheurs s'efforcent de répondre.

Un large panel de techniques est mis en oeuvre pour étudier l'endommagement de l'ADN, expérimentales (spectroscopie UV, HPLC-MS, cristallographie à rayons X, etc.) aussi bien que théoriques - calculs de mécanique quantique, dynamique moléculaire, etc. L'étude de la formation, la structure, et la réparation de systèmes d'ADN endommagé est au coeur des travaux de cette thèse, et a été réalisée en utilisant des techniques de modélisation moléculaire. Les aspects de réactivité ont été traités par mécanique quantique (QM) et méthodes hybrides quantique/classique (QM/MM), tandis que les propriétés structurales et dynamiques ont été explorées par simulations de dynamique moléculaire classique (MM-MD). La pertinence et la versatilité de telles méthodes ont été démontrées pour l'étude de l'ADN canonique[14–18], et ce sont de nos jours des techniques très largement utilisées dans l'étude de systèmes biomoléculaires, notamment en catalyse enzymatique

et transport membranaire. Notamment, le projet Ascona B-DNA Consortium (ABC) regroupe neuf laboratoires internationaux dans le but de mener des études computationnelles intensives visant à effectuer des simulations de dynamique moléculaire incluant eau et contre-ions explicites sur des oligomères d'ADN B. Ils utilisent des protocoles bien définis et le programme AMBER, celui-ci étant le logiciel le plus standard pour l'étude de l'ADN[19].

Bien que l'endommagement de l'ADN soit matière de recherche depuis de nombreuses dizaines d'années, la complexité et la richesse des aspects chimiques sous-jacents de l'ADN fournissent toujours une impressionnante quantité de questionnements, et de nombreux effets très subtils restent inexpliqués. Dans ce travail de thèse, nous avons étudié la formation, la structure, et la réparation de lésions complexes, dans le but d'apporter des informations utiles à la compréhension des mécanismes régissant ce type de phénomènes. Ces travaux ont été mené à l'Ecole Normale Supérieure de Lyon (ENS de Lyon) et à l'Institut des Sciences Analytiques (ISA), et la grande majorité des calculs ont été réalisés grâce aux ressources computationnelles du Pôle Scientifique de Modélisation Numérique (PSMN). Les différents projets de cette étude sont nés de discussions avec de nombreux scientifiques français et internationaux, expérimentateurs (Dr. Jean-Luc Ravanat du CEA Grenoble, France ; Dr. Alexandros Georgakilas de NTU Athènes, Grèce) et théoriciens (Dr. Antonio Monari et Dr. François Dehez de l'Université de Nancy, France ; Pr. Iñaki Tuñón de l'Université de Valencia, Espagne ; Pr. Leif Eriksson de l'Université de Göteborg, Suède ; Dr. Filip Lankaš de l'Université de Prague, République Tchèque).

Ce manuscrit rassemble les projets réalisés dans le cadre de cette thèse. Ces études ont mené à 10 publications, 8 acceptées et 2 soumises, organisées par la suite en quatre grandes parties. Premièrement, les méthodes computationnelles utilisées seront décrites. Ensuite, un chapitre sera dédié aux projets concernant les mécanismes de formation de plusieurs lésions complexes. Dans le troisième chapitre, seront exposées les études sur les propriétés mécaniques et dynamiques de l'ADN endommagé, ayant pour but de comprendre dans quelle mesure la présence de lésions induit des distorsions de l'hélice et affecte le paysage conformationnel. Enfin, la dernière partie de ce manuscrit traitera de la réparation des dommages. Y seront décrits les principaux mécanismes de reconnaissance et les enzymes qui y jouent un rôle, deux desquelles ont fait objet de projets au sein de cette thèse : l'Ap-endonucléase APE1 et la glycosylase Fpg. Une liste des papiers publiés au long de cette thèse est disponible à la suite des conclusions générales.

Introduction

The DNA molecule is the gardian of living organisms genetic information. It is constituted by four main building blocks, the nucleotides. The latter are composed of a phosphate/deoxyribose backbone on which is bounded a nucleobase through a N-glycosidic linkage : adenine (A), cytosine (C), guanine (G), thymine (T). Nucleobases are divided in two subsets: on the one hand the pyrimidines that correspond to the monocyclic thymines and cytosines, and on the other hand the purines that concern the bicyclic adenines and guanines. Noteworthy, living organism cells exhibit a 1:1 ratio of purines and pyrimidines according to the Chargaff rule[2–4]. Interactions between the nucleobases ensure the stability of the DNA structure as a double stranded helix - see Figure i. This particular structure was first published by James Watson and Francis Crick in 1953, relying on X-ray images obtained by the crystallographer Rosalind Franklin[1]. The highlighted helix shape of DNA relies on the Watson-Crick pairing between the nucleobases of self-complementary strands, with A interacting specifically with T, and G with C. This structure stability is ensured as well by a strong π-stacking of the nucleobases' conjugated rings, yet all these interactions do not prevent strands separation during DNA replication.

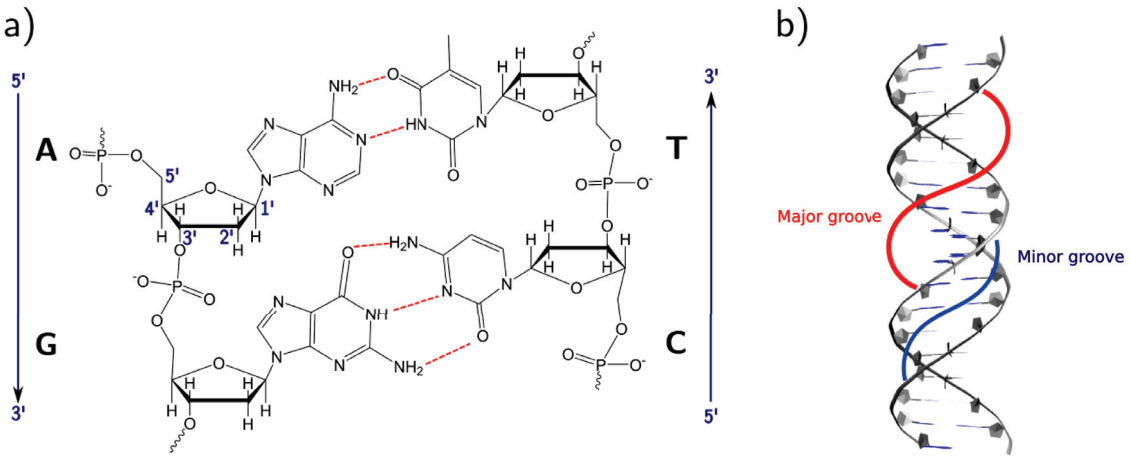


FIGURE I: a) Schematic representation of the four nucleotides in a d(AG):d(CT) 2-base pair (2-bp) duplex. Watson-Crick hydrogen bonding is depicted in red dashed lines. Adenine forms two hydrogen bonds with the thymine nucleobase, while the guanine interacts with the cytosine through three canonical hydrogen bonds. The usual numbering of the sugar ring atoms is denoted in blue. DNA strands are organized in a 5' $\rightarrow$ 3' way as represented by the blue arrows. b) New ribbons representation of a classical B-DNA helix. Backbone is depicted in grey while the nucleobases are depicted as flat blue shapes. Minor groove differs from the major one by its smaller width.

The succession of the $\sim$ 3 x 10^9 nucleotides in human cells[5] constitutes the genetic code. This latter is the base of protein synthesis, entities that ensure all the essential cellular works for living organisms survival. Nucleotide triplets, called codons, define which amino

acid (the building block of proteins) will be synthetized[6–8]. Thus, the succession of these triplets in DNA codes for the protein sequence. Noteworthy, about a half of the genome is constituted by repetitive DNA sequences, initially thought as "junk DNA", which are of major importance for gene expression controlling[9,10].
Actually, protein synthesis involves several protagonists that act stepwisely. First, a messenger RNA molecule (mRNA) is built by complementarity with DNA strands. Then, the protein will be synthetized in a specific part of the cell, the ribosome[11,12] which is divided in two compartments. A first small ribosomal subunit will read the mRNA, while a second large subunit will join amino acids to form the protein. Amino acids are carried by transfer RNA (tRNA), which serves as a physical link between mRNA and the amino acids, ensuring the correspondence of protein sequence with the genetic code. The gene expression phenomenon constitutes a fundamental process for cells. Thereby, when it is jeopardized, consequences are deleterious for the organism. Notably, a modification in the DNA code (genetic mutation) or a stalling during DNA replication and gene expression processes can induce cancerogenesis, apoptosis (programmed cell death), and aging.
Truth is that the cell machinery is constantly endangered. Indeed, DNA components undergo mutagenic agents attacks from sun light (UVB and early UVA), Reactive Oxygen Species (ROS) generated by oxidative stress, etc. Everyday, 10,000 to 1 million lesions per cell are induced by environmental factors and even by classical metabolic activities. Fortunately, cells contain an efficient DNA repair machinery, which consists in very complex processes which need intervention of a large number of enzymes[13]. The repair process is initiated by the recognition of the lesion, ensured by specific enzymes that are able to recognize changes in nucleotides properties. Thus, the structural and chemical modifications induced by one or several damages' presence is of utmost importance to understand their repair. On the other hand, the elucidation of mechanisms that governs their formation is also an important and timely issue for researchers.

DNA lesions study can be performed using a wide spectrum of methods, experimental (eg Ultraviolet-visible spectroscopy, High Performance Liquid Chromatography Mass Spectrometry, X-ray diffraction crystallography, etc.) as well as theoretical, both being complementary one to each other. In this thesis, molecular modelling techniques were used to led investigations about damaged DNA formation, structure and repair. Reactivity aspects were treated by quantum mechanics (QM) and hybrid quantum/molecular mechanics (QM/MM) methods, while mechanical properties were investigated using classical molecular dynamics (MM-MD). Such methods have proved their relevance and versatility on canonical DNA[14–18] , from nucleobases to genomic organizations, and are nowadays widely used in the study of biomolecules, especially for enzymatic catalysis. Noteworthy, a project called the Ascona B-DNA Consortium (ABC) gathers nine international research groups to lead a computationally intensive investigations aiming at carrying out molecular dynamics (MD) simulations including water and counterions on B-DNA oligomers. They used a well defined protocol and the AMBER suite of programs, the latter being the most common program for studying DNA[19].

Although investigations on the subject started several decades ago, the complexity and vastness of DNA damage chemical aspects still provide a huge matter of research, and lots of subtle effects remains unclear. In this thesis work, we investigated DNA damage, from the formation and structure of complex lesions to their repair, in order to understand mechanisms that rule such phenomena. This work was performed between the Institut des Sciences Analytiques (ISA) and the Ecole Normale Supérieure de Lyon, and the main part of the calculations were performed on the Pôle Scientifique de Modélisation Numérique

(PSMN) cluster. Projects were led in collaboration with several French and international scientists, experimentalists (Dr. Jean-Luc Ravanat from CEA Grenoble, France ; Dr. Alexandros Georgakilas from NTU Athens, Greece) as well as theoreticians (Dr. Antonio Monari et Dr. François Dehez from Nancy University, France ; Pr. Iñaki Tuñón from Valencia University, Spain ; Pr. Leif Eriksson from Gothenburg University, Sweden ; Dr. Filip Lankaš from Praha University, Czech Republic).
The following manuscript delineates the projects that were realized within the framework of this thesis. These studies led to 10 publications, 8 accepted and 2 submitted, sorted in four chapters. In the first one, the computational methods we used will be introduced. Then will come a chapter about investigations we led concerning DNA reactivity, precisely the chemistry behind a series of DNA lesions' formation. In a third chapter will be described the mechanical features of some complex damages within the double-helix, searching to understand to what extent they induce structural distorsions and reshape the conformational landscape. The last chapter will concern damaged DNA interactions with repair enzymes. Here will be delineated the different DNA repair recognition pathways through two enzymes, the Ap-endonuclease APE1 and the glycosylase Fpg, which fail to detect clustered abasic sites. A list of the papers published during this thesis is available after the general conclusions.

Chapter 1

Computational Methods

Molecular modelling is a field that provides a large variety of tools used for describing molecules in explicit solvents, periodic systems... notably biochemical systems. It offers the possibility to study reacting systems, excited states, dynamical behavior of large molecules, etc. at the atomic and electronic scales. In practice, theoretical chemistry is divided in two main paradigms : Quantum Mechanics (QM), in which one searches to describe the system's electrons, and Molecular Mechanics (MM) that is based on Newtonian mechanics, relying on the harmonic approximation. QM methods are relevant for describing electronic phenomena (reactions, spectroscopic features...), but are restricted to systems of few hundred of atoms since the computational time drastically increases with the system's size. On the contrary, molecular mechanics allows studying large molecules and is far less time-consuming than QM methods, as it does not require the calculation of electronic features. As a consequence, MM methods does not get access to reacting systems description.
In the seventies, scientists developed techniques to fuse these two complementary paradigms : the powerful QM/MM hybrid methods were born. These pionneers, named Martin Karplus, Arieh Warshel, and Michael Levitt were awarded by the Nobel Prize of Chemistry in 2013 for their unprecedent works. Combining QM and MM strengths, QM/MM methods allows to study reactivity of large systems such as enzymatic processes, giving impressive prospects for chemist' works, one "master piece" being the design of artificial enzymes.

Along this thesis work, the three methods (QM, MM, and QM/MM) were applied to carry out the several projects we studied. Thereby, these techniques' basics are described in the following chapter. In addition, a subsection within QM methods is dedicated to a methodologic project concerning conceptual DFT development, that was led within the framework of this thesis.

1.1 Quantum Mechanics

'Anyone who is not shocked by quantum theory has not understood it. '

Niels Bohr

Quantum Mechanics aims at derivating ground and excited states properties : molecular structures and energies, transition states, atomic charges, reaction pathways, spectro-

scopic features, etc. At any quantum system is associated a wavefunction Ψ that describes this system's state[20]. We are dealing with a collection of electrons, whose exact position and momentum cannot be known at the same time, according to the Heisenberg's uncertainty principle. The wavefunction itself does not have any physical significance, but the squared term $|\Psi|^2$ represents the probability density to find the electrons in certain volume when it is in the state described by Ψ.
In practice, this method is based on the Schrödinger equation (eq. 1.1) resolution, the nuclei being considered in a classical manner. For a given system, the time-dependent Schrödinger equation is the following :

$$H\Psi = i\hbar\frac{\delta\Psi}{\delta t} \tag{1.1}$$

For a conservative system, the chemical potential V is time independent and depends only on the space coordinates $\vec{r}$, meaning $V(r,t) = V(r)$. The Ψ function can then be written as $\Psi(r,t) = \psi(r)F(t)$, where the Hamiltonian only depends on the spatial term since $V(r,t)$ is time-independent.
Hence, one can demonstrate that

$$P = \int \Psi^*(r,t)\Psi(r,t)\delta\upsilon = \int \psi^*(r)\psi(r)\delta\upsilon \tag{1.2}$$

for a conservative system, which means that the probability density $|\Psi|^2$ calculation is time-independent.

Resolving the time-independent Schrödinger equation, the lowest energy eigenvalue E_0 is obtained for the ground-state wavefunction Ψ_0. Thus, any other wavefunction Ψ should be associated to an energy E higher than E_0 pursuant to the variational theorem. According to the Born-Oppenheimer approximation, nuclei's velocities are negligeable comparing to the electrons' ones considering that nuclei masses m_p are far heavier than the electrons ones m_e ($m_e \simeq \frac{m_p}{1850}$). Thus, the nuclei kinetic energy is considered as negligeable : $T_N = 0$. Thereby, for a molecule consisting of M nuclei A and N electrons i, the electronic Hamiltonian is expressed as follows :

$$H_{el} = \sum_{A=1}^{M}\sum_{i=1}^{N} -\frac{Z_A}{|R_A - r_i|} + \sum_{i=1}^{N} -\frac{1}{2}\nabla_r^2 + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{|r_i - r_j|} \tag{1.3}$$

$$= V_{en} + T_e + V_{ee} \tag{1.4}$$

Here, r_i is the coordinate of the electron i while Z_A corresponds to the charge of a nucleus A located at position R_A. The potential terms V_{en} and V_{ee} correspond to the electron-nuclei attraction and the electron-electron repulsion respectively, the electrons kinetic energy being the T_e contribution. Whereas the first two ones are monoelectronic operators and can be written down as sum of monoelectronic contributions, V_{ee} energy involves two electrons that cannot be decorrelated. Even with approximations, the Schrödinger equation cannot be solved analytically.
Nevertheless, numerous theories have been developed in order to circumvent these limits. They are based either on the use of the wavefunction (Hartree Fock and post-Hartree Fock), or on the use of the electronic density $\rho(r)$ (Density Functional Theory). In the following sections will be described the Hartree-Fock, DFT and conceptual DFT basics that have been used within the framework of this thesis.

1.1.1 The Hartree-Fock method

The Hartree-Fock theory is a variational method based on the wavefunction Ψ[21]. Each electron i is described by an orbital ϕ_i, which does not explicitly depend on the other electron coordinates. Under this approximation, the problematic electronic repulsion term V_{ee} in equation 1.4 is simplified to a mean field approach. Thus, this term will simply correspond to electronic repulsion between pairs of electrons placed in the mean field induced by the $N-1$ other electrons. Truth is that, considering their spin, electrons are associated to spin-orbitals, defined as the product of an orbital and a spin function. They are expressed as $\vec{x_i} = (\vec{r_i}, \omega)$, where the $\vec{r_i}$ term corresponds to the spatial positions and ω the spin coordinates (α or β).
Electrons cannot be discerned from each other, hence obey to the Pauli exclusion principle which states that two or more identical fermions (particles with half-integer spin) cannot occupy the same quantum state (ie the same spin-orbital). This implies that Ψ has to be antisymmetric : the wavefunction must change its sign upon the exchange of two electrons.
Within the framework of the Hartree-Fock approximation, Ψ can be written as a single Slater determinant :

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_N(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \dots & \phi_N(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_1(x_N) & \phi_2(x_N) & \dots & \phi_N(x_N) \end{vmatrix}, \tag{1.5}$$

Where the spin-orbitals $\phi_i(x_i)$ are solutions of the Hartree-Fock equations :

$$\hat{F}_i \phi_i(x_i) = \epsilon_i \phi_i(x_i) \tag{1.6}$$

Equation 1.6 brings out the one-electron Fock operator $\hat{F}$ which, applied on spin-orbital $\phi_i(x_i)$, leads to the corresponding energy ϵ_i.
It can also be written as :

$$\hat{F}_i \phi_i(x_i) = \sum_{i=1}^{N} h_i + \sum_{i=1}^{N} \sum_{j>i}^{N} \langle \phi_i | J_j(i) | \phi_i \rangle - \langle \phi_i | K_j(i) | \phi_i \rangle \tag{1.7}$$

where $\langle \phi_i | J_j(i) | \phi_i \rangle$ is the Coulomb term, which reflects the mean potential induced by the other electrons, and the $\langle \phi_i | K_j(i) | \phi_i \rangle$ term, called the exchange energy, which is a correction of the Coulomb term due to the antisymmetry. The h_i term simply corresponds to the sum of the kinetic energy of the electron i and the electrostatic potential between this electron and the nucleus (V_{eN}).
In order to solve the Schrödinger equation through HF approximation, the energy E is minimized according to the variational method. To do so, one uses the Self-Consistent Field (SCF) procedure. Starting from a set of guess orbitals, the latter are stepwisely adjusted until convergence is reached.

The Hartree-Fock method makes several assumptions concerning the Schödinger equations. In order to improve the definition of the electron correlation (approximated by the mean field in HF), several methods have been developped[21]. The latter, called post-Hartree-Fock methods, give access notably to a better description of excited states, and more accurate results than HF most often. However, they need more computational

ressources. Among these methods, we can cite the configuration interaction (CI)[22,23], the coupled cluster (CC)[24–26] and the Møller–Plesset perturbation (MP2, MP4, etc.)[27,28] theories, which are nowadays widely used.

1.1.2 Density Functional Theory

Within the framework of Density Functional Theory (DFT), it has been proven that the electron density $\rho(r)$ contains all the information about the system (molecules, solids, macromolecules) under study. Thus, the energy of a system can be expressed as a functional of its electron density as demonstrated by the Hohenberg-Kohn theorems[29]. The energy functional is partitioned as follows:

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{ext}[\rho] = F_{HK} + \int \rho(r)v(r)dr \tag{1.8}$$

where $T[\rho]$, $V_{ee}[\rho]$, $V_{ext}[\rho]$ and F_{HK} correspond to the kinetic, electronic repulsion, external potential, and Hohenberg-Kohn universal functionals respectively.

Theoretically, DFT can provide the exact system's energy, since all information about the system are included in the density expression. In practice, this is not the case, because the F_{HK} functional exact form is not known. In 1961, Kohn and Sham proposed a mapping with a fictitious system in order to circumvent this problem :

$$E[\rho] = T_s[\rho] + J[\rho] + V_{ext}[\rho] + E_{xc}[\rho] \tag{1.9}$$

where they introduced E_{xc}, the exchange-correlation functional :

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho] \tag{1.10}$$

T_s is the kinetic energy of a fictitious system with N non-interacting electrons and J the classical Coulomb interaction, which is the principal component of V_{ee}. Thus, E_{xc} represents the sum of the errors introduced by using a non-interacting kinetic energy and a classical definition of the electron-electron repulsion.
Hence, the Kohn-Sham ansatz reduces the Schrödinger equation resolution problem to the finding of an accurate form of the exchange-correlation functional. Since decades, numerous approximations of E_{xc} have been and are still developed. These functionals can be classified on what is called the Jacob's ladder of DFT - see Figure 1.1.1.
On the first rung lie the Local Density Approximation (LDA) functionals, that have been derived from homogeneous electron gas at the early stages of DFT development. Here, the charge density remains constant since the electrons are subject to a constant external potential. Nowadays such functionals are only used for solids since it offers poor description of molecular systems which present strong local variations of the density.
Then comes the Generalized Gradient Approximation (GGA) functionals. Here, the energy functional depends on both the density and its gradient $\nabla(\rho)$. This approximation improves significantly molecular systems description comparing to the LDA level. The next rung is dedicated to the meta-GGA functionals, which also includes the Laplacian of the electron density. Nowadays, the latter are not so much used, since they do not offer much more efficiency than GGA yet are more time-consuming.
On the fourth rung, one finds the hybrid exchange functionals, which balance the GGA exchange-correlation by a percentage of the abinitio Hartree-Fock exact exchange. They offer accurate results in reasonable computational times, so they are highly used - in this

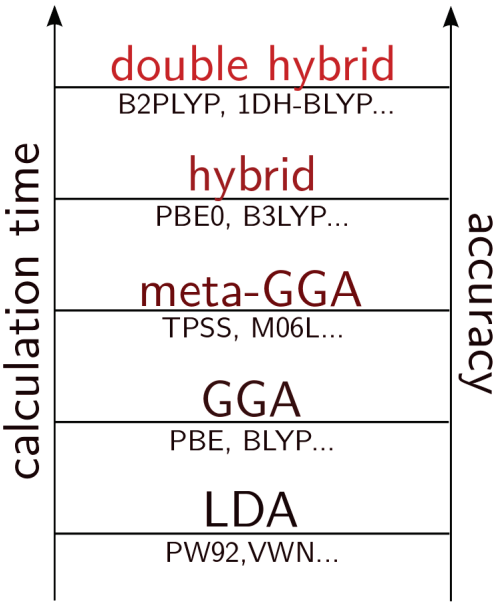


FIGURE 1.1.1: Schematic representation of Jacob's ladder of the DFT. Functionals are sorted by accuracy and calculation time, from LDA to double hybrid level of theory - noteworthy, the more accurate the more time-consuming. Examples of functionals are indicated for each rung.

thesis we most of the time used this family of functionals. We also used range-separated hybrid functionals (RSH) which incorporate the exact HF exchange not uniformly but as a function of the interelectronic distances.
Finally, the last rung of the Jacob's ladder is dedicated to the double hybrid functionals, which are more accurate than the hybrid ones since they include Hartree–Fock exchange and second-order Møller–Plesset (MP2) correlation. The main drawback of such functionals is that they are prohibitively time-consuming, as a consequence they are barely used.
Noteworthy, even though some ways to take into account the exchange-correlation term seem better than others, the efficiency of a functional highly depends on the considered system, as well as the properties one wants to compute. Thus, several types of systems have been benchmarked in order to provide guidance into the choice of a density functional.

DFT itself does not provide accurate description of time-dependent processes and excited states properties[30], hence its use is limited to the ground state particle density. However, in this thesis we needed to explore excited states in order to study triplet-triplet energy transfer[31]. Therefore, we used the Time-Dependent DFT (TDDFT). This method provides an extension of DFT beyond the ground-state quantum systems. It relies on the Runge-Gross theorem, which is the time-dependent analog of the Hohenberg-Kohn one[32].

1.1.3 Conceptual DFT

Density Functional Theory of Chemical Reactivity, the so-called Conceptual DFT[33], has emerged several decades ago through R.G. Parr and R.G. Pearson works[34–37]. The basic idea of this theory is that all concepts already known in chemistry (chemical potential, electronegativity...) can be mathematically defined by DFT. This theory offers the possibility to compute several chemical descriptors, directly from the electron density obtained by DFT calculations.

The first one of these descriptors was introduced by Hohenberg and Kohn seminal paper[29], and is the so-called chemical potential μ.

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{v(r)} . \tag{1.11}$$

This descriptor is the derivative of the energy E with respect to the number of electrons N, giving a real description of the way the density adapts under a variation of the electrons number, at fixed external potential v_r - ie fixed geometry. Noteworthy, μ is directly related to the electronegativity χ which is its exact opposite[38].
The power of CDFT theory relies on the description of chemical properties of entire molecules by simply applying an interplay of derivatives directly from results obtained by DFT calculations.
Indeed, if we continue to derivate the E with respect to N, ie computing its second order derivative which is equal to the first derivative of the chemical potential, it leads the hardness η. This descriptor captures the variation of the chemical potential with respect to a variation of the number of electrons - or not. Practically, it corresponds to the system's propensity to be stabilized by a variation in the number of electrons. Therefore, this descriptor defines hardness in the sense of Pearson's Hards and Soft Acids and Bases theory[39]. Hard species will exhibit a high value of η, and will undergo strong variations of μ upon variation of N, considering that they are not prone to share electrons and prefer to create ionic interations rather than covalently. On the contrary, soft molecules will be characterized by low values of η since they are more susceptible to accept variation of N.
Going through such interplay of derivatives, one get access to a large number of descriptors that are simply response functions of the energy to perturbations in N or $v(r)$[40] - see Figure 1.1.2.

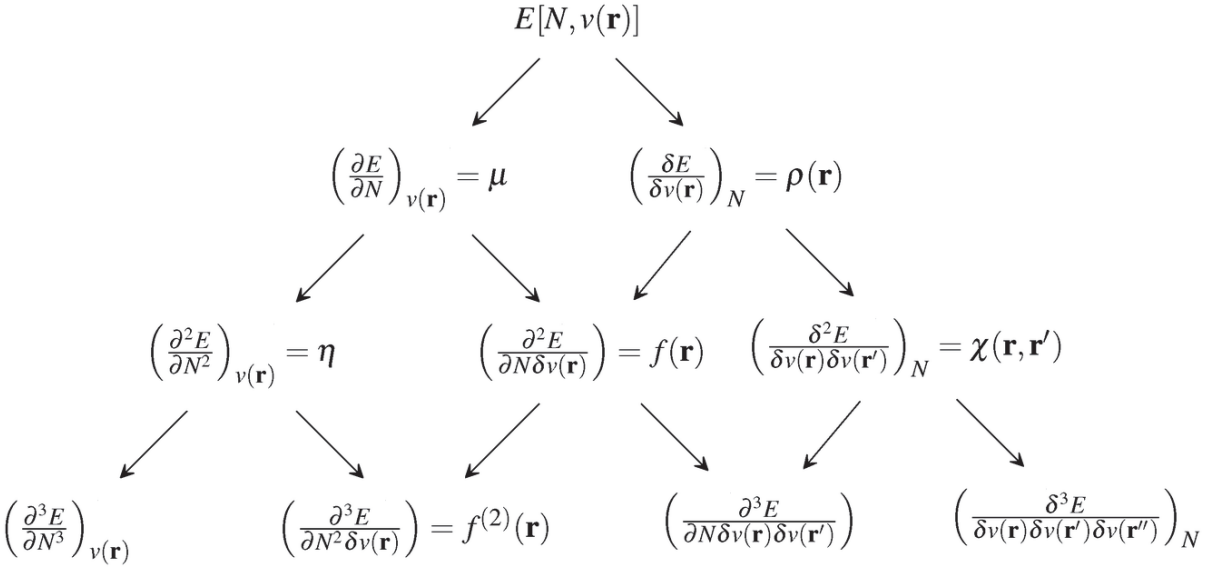


FIGURE 1.1.2: Tree of conceptual DFT descriptors. Top : energy E. Second raw : chemical potential μ[35] (left) and electronic density $\rho(r)$ (right). Third raw : chemical hardness η[39] (left), Fukui function $f(r)$[41] (middle), and linear response $\chi(r, r')$[42,43] (right). Bottom left to right : hyperhardness γ[44], dual descriptor $f^{(2)}(r)$[45,46], descriptor describing the response of the Fukui function at position r to a perturbation in the external potential at position r'[47], descriptor describing the variation of the linear response function upon perturbation of the external potential at a third point r"[48].

CDFT descriptors can also provide insights into molecular properties evolution along a reaction path. One of these indexes is the so-called Reaction Electronic Flux (REF) that is related to the spontaneity of chemical reactions[49–51]. This relatively recent descriptor is expressed as the opposite of the first derivative of the electronic chemical potential μ with respect to the intrinsic reaction coordinate ξ :

$$J(\xi) = -\frac{d\mu}{d\xi} \tag{1.12}$$

The chemical potential has been discussed above (equation 1.11), while the reaction coordinate ξ can be defined as the widely used Fukui's instrinsic reaction coordinate (IRC)[52,53]. The REF $J(\xi)$ describes the spontaneity of the considered reaction : positive values of REF reflects spontaneous processes while reactions exhibiting negative values of the REF can be considered as unspontaneous. REF values that tends to zero corresponds to equilibrium states, basically in the case of reactants and products.

Such insights into fluctuations of the electron density and the predictive power of this method allow a better understanding of chemical phenomena which is essential for research's advances in several fields of chemistry. Within this framework, we investigated the expression of the reaction electronic flux in the canonical and grand canonical ensembles[54]. This study led to a publication [P2] in *Theroretical Chemistry Accounts*, which can be found at the end of this chapter.

1.1.4 Basis set

A QM level of theory relies on the method used but also on how atomic orbitals (AO) are represented. In practice, AO are represented by basis functions whose linear combination leads to molecular orbitals, this set of functions being called the basis set. This latter can be more or less large, since one can choose to represent each AO by one, two (double zeta), three (triple-zeta) or more set of basis functions, giving results more and more accurate by improving the system description. The basis set can be composed of Slater-[55] or Gaussian-type[56] orbitals (STOs and GTOs respectively). Noteworthy, a Slater function can be approximated by a sum of Gaussian functions[57]. The use of GTOs provides very much simpler integrals than for STOs and, although more Gaussian than Slater functions are required to reach a given basis set quality, notably to recover the "cusp", the simplicity of Gaussian integrals more than compensates for this[58].
Polarization functions are most added, allowing AO to be polarized by mixing orbitals types, providing some angular flexibility. For instance, an s orbital can polarized if mixed with a p one, a p orbital is polarized if mixed with a d orbital, and so on. Moreover, the addition of diffuse functions is required in some cases such as for anions or very electronegative atoms. Indeed, when the system presents a high electron density, its description will need an enhanced radial flexibility, which can be provided by such diffuse functions. Basically, the latter permit a better description of electrons far from the nucleus.
In our work, we often used 6-31G(d) and 6-31+G(d) as basis sets, where core electrons are represented by a subset of 6 gaussian functions, while the valence ones are descripted by two different functions, the first one being a combination of 3 pre-adjusted gaussians and the second one a unique gaussian. The star indicates that polarization function is added on heavy atoms (all atoms except hydrogens), and the addition of a '+' means that the density radical flexibility is enhanced by diffuse functions on heavy atoms as well. These functions can also be extended to hydrogens, the notation being then double star and/or double '+'.

1.2 Molecular Mechanics

1.2.1 Basic principles

Molecular Mechanics relies on the harmonic approximation according to which atoms and bonds can be considered as balls and springs respectively. Atoms are thus described as point charges that exhibit a certain Van der Waals radius and bear a partial charge. As a consequence, the system energetics can be decomposed in bonded terms and non-bonded terms. Bonded terms represent energies of bonds, angles, and dihedral angles, while non-bonded terms reside in the electrostatic and Van der Waals contributions. Calculation of each component relies on force fields equations that depends of the force field considered. In our case, calculations are performed with the AMBER parm99 force field which is appropriate for biomolecules, with bsc0 and bsc1 corrections that refine DNA backbone torsion parameters and have been exhaustively tested and cross-validated[16,59]. In AMBER, force field equations are expressed as follows[60] :

$$E_r = k_b(r - r_0)^2 \tag{1.13}$$

$$E_\theta = k_\theta(\theta - \theta_0)^2 \tag{1.14}$$

$$E_\Phi = \frac{V_n}{2}[1 + \cos(n\Phi - \gamma)] \tag{1.15}$$

$$E_{VdW} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \tag{1.16}$$

$$E_{elec} = \frac{q_i q_j}{\epsilon r_{ij}} \tag{1.17}$$

Bonds r and angles θ contributions are defined by harmonic potentials (see eq. 1.13 and 1.14), with k_b and k_θ the constant forces driving the potential curvature, while r_0 and θ_0 being the equilibrium position parameter defined in the force field. For the dihedral Φ term (eq 1.15), V_n, n, and γ represent the torsion constant, the rotation periodicity and phase angle respectively. On the other hand, the Van der Waals contribution (eq 1.16) is defined by a Lennard-Jones potential that can be decomposed in two parts : the left one is a repulsive term, predominant at small distances, while the right one reflects the atoms attraction which is more important as r increases. A_{ij} and B_{ij} are given by combination rules, and depend on the atoms i and j types. Finally, the electrostatic energy brings into play the atomic charges q_i and q_j, the medium's permittivity ϵ, and the distance between the atoms r_{ij}. Using these equations, one can minimize energy with respect to the force field parameters, thus providing an optimized geometry prone to be used as a starting structure for classical Molecular Dynamics simulations.

1.2.2 Classical Molecular Dynamics

Exploring the dynamical behavior of biological macromolecules is of particular importance since they are not frozen structures, and undergo conformational changes under physiological conditions that can impact their activity. Molecular Dynamics (MD) simulations offer the possibility to gain insights into the dynamics of such systems, exploring their

conformational behavior and sampling the most probable structures, ie with the lowest energy, along simulation that may reach several μs. More specifically, classical MD is based on the MM principles and the Newton's equation of motion which is expressed as follows[61] :

$$\vec{F_i} = \frac{d\vec{p_i}}{dt} \tag{1.18}$$

With $\vec{F_i}$ the force exerced on the atom i, and $\vec{p_i}$ its momentum, which can also be written as $\vec{p_i} = m_i.\vec{v_i}$ with m_i the atom mass and $\vec{v_i}$ its velocity. Molecular Mechanics equations allow to compute the system's energy, which is related to the forces applied on each atom :

$$\vec{F_i} = -\vec{\nabla}E(r_i) \tag{1.19}$$

Here, it simply consists in deriving the total energy E with respect to the atom position $\vec{r_i}$. Thus, combining the equation 1.18 and 1.19, one can link a given atom energy with its acceleration $\vec{a_i} = \frac{d\vec{v_i}}{dt}$, hence its position :

$$-\vec{\nabla}E(r_i) = m_i.\vec{a_i} = m_i.\frac{d\vec{r}^2}{dt^2} \tag{1.20}$$

It is worth noting that this relation involves the time term dt^2 which is particularly important since we want to access the system's conformational behavior along time.
However, equation 1.20 is not analytically resolvable, hence one have to compute the atoms accelerations using eq.1.18, then integrate a first time to obtain their velocities, and a second time for their positions. Atoms coordinates at time $t+\Delta t$ are then computed using the positions at time t. Several algorithms can be used to perform this task, such as Verlet and its leap-frog alternative. In this latter for instance, position and velocities are defined from different times, $t-\frac{1}{2}\Delta t$ and $t+\Delta t$ respectively. Thus, velocities at $t+\frac{1}{2}t$ are computed according to the ones at $t-\frac{1}{2}t$ as follows :

$$v_i(t+\frac{\Delta t}{2}) = v_i(t-\frac{\Delta t}{2}) + a_i(t).\Delta t \tag{1.21}$$

From which one can deduce the atoms positions x_i :

$$x_i(t+\Delta t) = x_i(t) + v_i(t+\frac{\Delta t}{2}).\Delta t \tag{1.22}$$

Hence this algorithm goes on step by step, where positions and velocities 'leap over' each other like frog jumps, as one can easily see in Figure 1.2.1 representation.
Regardless of which algorithm is used, the time step Δt should not be chosen to be larger than 1 fs in order to describe the fastest molecular movements. Indeed, this value corresponds to the fastest interatomic vibration frequency, called Nyquist frequency, which is the one describing the vibration between heavy atoms and the hydrogen atom. However, one can choose to freeze X-H vibrations, thereby being free to increase this time step to the second fastest vibration value : the C-C bond with a frequency of 2 fs^{-1}. Thus, computational times are reduced by a factor 2. In our calculations, we use the SHAKE algorithm in order to freeze X-H vibrations, doubling the simulation time length without increasing the computational cost. On the other hand, long-range electrostatics are treated by the Particle Mesh Ewald summation[62,63].

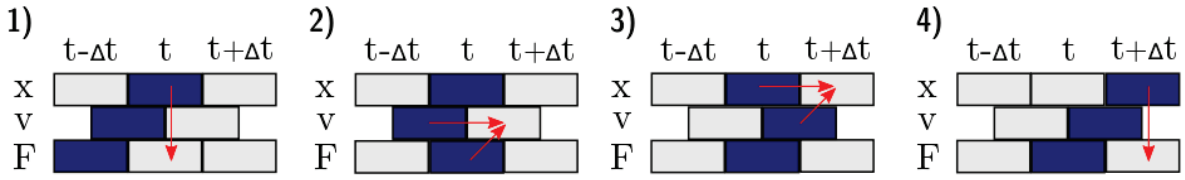


FIGURE 1.2.1: Schematic representation of leap-frog algorithm steps : 1) Forces $F_i(t)$ are computed at time t using positions $x_i(t)$ 2) Velocities $v_i(t+\frac{\Delta t}{2})$ are deduced from $F_i(t)$ and $v_i(t-\frac{\Delta t}{2})$ 3) Positions $x_i(t+\Delta t)$ are determined from $x_i(t)$ and $v_i(t+\frac{\Delta t}{2})$ 4) The same cycle is applied for the next step $t+\Delta t$.

In practice, classical MD is performed in three steps with the AMBER12 program[64] : heating, equilibration and production. All along the simulation, the system is in periodic boundary conditions : it is replicated in each directions of space. Thus, atoms at the system's extremities do not diffuse in the vacuum but are just appear at one side of the 'box' when they crossed the edge on the opposite side. Therefore, a cut-off must be applied for long-range interactions to avoid self-interactions between atoms' images from one box to the other.

Heating The starting geometry has been previously optimized at 0 K using force field equations - see section 1.2.1. However, when one wants to simulate molecular dynamics, temperature has to be higher than zero since it is directly related to kinetic energy, thus to atoms' velocities. Temperature has to be identical to real conditions, that is why the heating run is performed in order to increase the temperature to the value required. Several thermostats have been developed in order to control heating (in NVT conditions) and to ensure that the average temperature is the desired one during the remaining simulation steps (in NTP conditions). We usually use the Langevin thermostat[65], which consists in placing the system's particles in a continuum of fictitious much smaller atoms. The latter will give random kicks to large particles, hence inducing a damping force on them. These random collisions influence the system's dynamics, and as a consequence, its temperature. The collision fequency γln can be adjusted at one's convenience - 1ps^{-1} in our studies, chosen in agreement with other undammaged DNA MD simulations.

Equilibration Once the system has reached the desired temperature, it has to be equilibrated before to start the sampling of its representative conformations. Indeed, during heating the system's energy drastically increased, and it has to be relaxed, thus to decrease in energy, in order to ensure that one is not sampling conformations that are not supposed to be probable in the simulation conditions. Thus, a 1ns equilibration run is usually performed before the sampling.

Production The relaxed system can now be gently sampled, having access its to representative conformations. This step permits to explore the conformational surface corresponding to the system structural fluctuations, thereby giving informations about its dynamical behavior. Structures obtained by such simulations can be post-processed to provide mechanical properties values. For instance, we often use the Curves+ program[66] which is appropriate to DNA studies, providing inter- and intra-bases parameters, helix bending values etc. Along such dynamics, one cannot overcome relatively high barriers unless enhanced MD techniques such as Replica Exchange MD (REMD) or Umbrella Sampling are used.

Classical Molecular Dynamics is a really powerful method to probe the conformational behavior of large systems such as biological macromolecules, giving insights into their structural properties along simulation that can reach several *ms* long. We employed MM-MD in every project that we led during thesis' work. It is worth noting that such methods are nowadays very frequently used and have be proven to be able to reproduce experimental results, therefore opening perspectives to predictive models and a real computational design. Albeit we mostly use the AMBER program for our MM-MD simulations, several other simulation packages are available such as NAMD, CHARMM, GROMACS, GROMOS, LAMMPS, DYNAMO, etc.

1.2.3 Force field parameters generation for non-canonical nucleotides: LNAparm

For MM-MD simulations of DNA, we use the parm99 force field with bsc0 and bsc1 corrections[16,59]. However, this force field provides only parameters for canonical nucleotides. As a consequence, we need to generate lesions parameters when studying damaged DNA. To this extent, we use the following protocol.

First, the damaged nucleotide structure is optimized with DFT methods, the choice of the functional depending on benchmarks available for the chemical moiety it exhibits - most of the time B3LYP, but for example we used the range-separated functional LC-BLYP in our study on singlet oxygen attack onto guanine[67,68] (these papers can be found in Chapter 2), according to previous studies[69,70]. Then, Merz-Kollman charges are computed following the RESP protocol[71], the level of theory being HF/6-31G(d) to be consistent with the AMBER force field[64]. Frequencies are computed in order to ensure that the nucleotide geometry corresponds to a minimum on the potential energy surface. The nucleotide extremities are capped by a -OMe moiety at the 5' end and a methyl group at the 3' end in order to mimic the bonds with adjacent nucleotides. These 'capping atoms' are removed before generating the parameters file, and their point-charge values are distributed equally to the other atoms to ensure the conservation of overall polarization of the nucleobase. Then, the parameters are generated from the optimized geometry, using the antechamber procotol disponible in the AMBER suits of programs. Each atom is associated to an atom type depending on the force field used, and the corresponding parameters are directly taken from the force field data. For nucleotides, we use atom types from the same force field than for the rest of the DNA helix : parm99 with bsc0/bsc1 corrections[16,59]. For exogenous ligands such as in the study of ketoprofen interactions with DNA (see Chapter 2), the GAFF force field[72], which was developed for small organic molecules, was taken into account. In order to assess the relevance of the parameters generated for non-canonical nucleotides, one can try to reproduce experimental features, for instance spectroscopic or structural values (obtained by UV-vis spectroscopy, FRET...) that might be available in the literature.

Up to now, we generated parameters for: 8oxoG, its spiro degradation product and the endoperoxide and zwitterionic intermediates leading to its formation; interstrand crosslinks; abasic site; tetrahydrofurane nucleotide; C5'-C8 guanine and adenine intrastrand crosslinks; 6-4PP; Pyo; CPD; N4- and C5-methylcytosines; the thymine glycol. We are currently building a DNA lesions parameters database, called LNAparm, which will be available for the public.

1.3 Hybrid QM/MM methods

Hybrid QM/MM methods arise from the partitioning of a given system in a reactive core described in QM, embedded in a MM environment - see Figure 1.3.1. Thus, it has become a tool of choice for biomolecules' reactivity modeling. For instance, enzymatic activities can be modeled by describing atoms of the reaction site in QM, while the remaining of the enzyme and the solvent are represented in MM. This technique allows to study macromolecules reactivity without being hampered by prohibitive computational costs. Therefore, to compute the QM/MM energy of the system, one should determine : 1) the

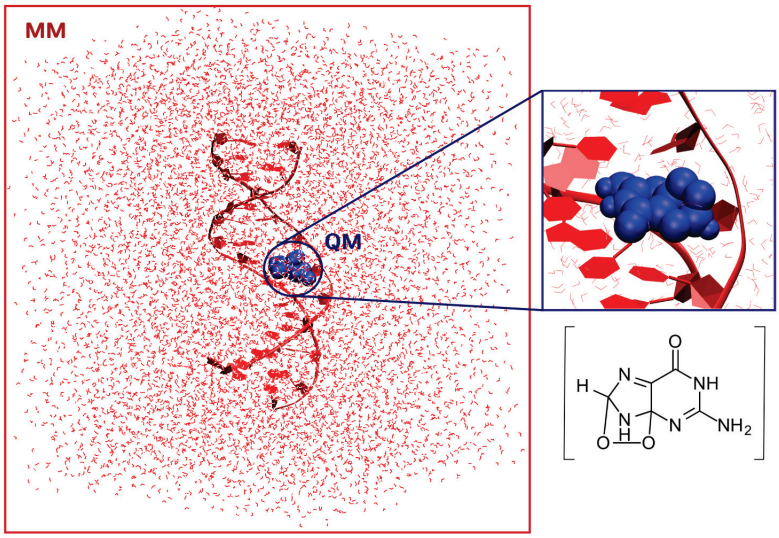


FIGURE 1.3.1: Example of QM/MM partitioning for a damaged oligonucleotide. In this case, one of the nucleobases is the inner site treated in QM (in blue, VdW representation), while the remaining of the system is described in MM (in red). The inset shows a zoom on the QM atoms, the damaged nucleobase being the endoperoxide intermediate leading to 8-oxo-7,8-desoxyguanosine which was the subject of one of this thesis reactivity projects - see Chapter 2.

MM energy of the outer system 2) the QM energy of the inner site 3) the interaction energy between both parts. Then, the total energy can be computed as follows:

$$E_{QM/MM} = E_{MM}^{outer} + E_{QM}^{inner} + E_{QM/MM}^{interaction} \tag{1.23}$$

where interactions between both parts include bonded, Van der Waals and electrostatic terms :

$$E_{QM/MM}^{interaction} = E_{QM/MM}^{bond} + E_{QM/MM}^{VdW} + E_{QM/MM}^{electrostatic} \tag{1.24}$$

The MM point-charges atoms are incorporated as one-electron terms in the QM Hamiltonian. Thus, the QM part can adapt according to the MM charge distribution, resulting in its automatic polarization. However, MM charges cannot be polarized in this scheme, that does not give a realistic representation of the charge distribution. Despite this weakness, carefully defined MM charges allows to obtain reasonable results, especially for widely used biomolecular force fields such as the ones from AMBER. The best way to get rid of this problem would be to use a polarizable force field, which would allow QM and MM mutual polarization. Unfortunately, accurate polarizable DNA force fields are not disponible so far, albeit some rare ones can be found for proteins - AMOEBA for instance[73].

The accurate definition of the boundary between the QM and MM parts ($E^{interaction}_{QM/MM}$) is challenging, particularly when it is placed on a covalent bond. Thereby, one of the atoms involved in this bond will be encompassed in the QM region (A^{QM}) while the other one will be defined in MM as a point-charge (A^{MM}). Several boundary schemes have been developed, the one commonly used being the Link Atom method[74]. Its principle is conceptually simple : A^{QM} is capped by a hydrogen to complete its valency. Any monovalent atom could work, but the hydrogen atom is the most commonly used. This method is widely used albeit some problems obviously arise from such an approximation. On the one hand, the link atom is chemically different from the initial group it replaces, inducing important approximations in mimicking the real mutual influence between QM and MM atoms. On the other hand, the link atom is spatially very close to the MM atom at the QM/MM boundary. Thus, the QM density lies closer to the MM atoms, hence is prone to be overpolarized by the point charge atom at the boundary. Another drawback of the Link Atom method is the fact that each link atom inserts three artificial degrees of freedom supplementary to the system, hence increasing the calculation costs. However, progress have realized since the first development of this method. Northeworthy, efforts have been made in order to remove the extra degrees of freedom using scaling factors[75–78] and to attenuate the QM density overpolarization[76,79–82].

In this thesis, QM/MM methods were used, on the one hand one-the-fly calculations in order to model $^1O_2$ attack onto guanine[68], and on the other hand static TDDFT calculations for the study of a triplet triplet energy transfer from an artificial base to a thymine[31] - see Chapter 2. Such calculations were performed on the one hand by an interface between the Tinker[83] and Gaussian09[84] packages developed by our collegues from Nancy University[85,86], and on the other hand by coupling of Gaussian09 within the fDynamo program[87].

1.4 In brief.

Molecular modelling tools are versatile and relevant for studying chemical systems. Biomolecules (DNA, proteins, lipids) have been widely investigated using computational methods, which have proved their worth and are essential in research, besides experimental chemistry. The diversity of tools that it provides allows researchers to study reactivity and structure of a large range of molecules, from small compounds to macromolecules. The 2013 Nobel Prize, attributed for investigations about QM/MM methods development, highlighted the constant inscrease of interest for computational methods, which are nowadays widely used for studying biomolecules among others.
Molecular Modelling tools are in constant development and improvement, and the exponential progress of computational performances promisses very interesting perspectives for such methods in the near future.

1.5 Publications

Theor Chem Acc (2015) 134:133
DOI 10.1007/s00214-015-1730-7

REGULAR ARTICLE

Insights into the chemical meanings of the reaction electronic flux

Christophe Morell[1] · Vincent Tognetti[2] · Emmanuelle Bignon[1,3] · Elise Dumont[3] · Noemi Hernandez-Haro[4] · Barbara Herrera[5] · André Grand[4] · Soledad Gutiérrez-Oliva[5] · Laurent Joubert[2] · Alejandro Toro-Labbé[5] · Henry Chermette[1]

Received: 13 July 2015 / Accepted: 13 September 2015 / Published online: 16 October 2015

Abstract The negative derivative of the chemical potential with respect to the reaction coordinate is called reaction electronic flux and has recently focused a wide interest to better understand chemical reactions at molecular level. After much consideration, it is now well accepted that positive REF values are associated with spontaneous processes, while negative REF ones translate unspontaneous phenomena. These characteristics of the REF are based on a thermodynamic analogy and have been shown right through computational results. In this paper, we develop two analytical expressions of the REF in both the canonical and the grand canonical ensembles. The connection between both equations is established. They are then analyzed, and some arguments are put forward to support the alleged characteristic of the REF and its ability to properly discriminate spontaneous from unspontaneous phenomena.

✉ Christophe Morell
christophe.morell@univ-lyon1.fr

✉ Vincent Tognetti
vincent.tognetti@univ-rouen.fr

[1] Université de Lyon, Institut des Sciences Analytiques, UMR 5280, CNRS, Université Claude Bernard Lyon 1, 5 rue de la Doua, 69100 Villeurbanne, France

[2] Normandy University, COBRA UMR 6014 & FR 3038, Université de Rouen, INSA Rouen, CNRS, 1 rue Tesniére, 76821 Mont St Aignan. Cedex, France

[3] Laboratoire de Chimie, UMR 5182, CNRS Ecole Normale Supérieure de Lyon, Lyon, France

[4] CEA Grenoble -INAC/SCIB/LAN (UMR-E n°3 CEA-UJF), CEA-Grenoble, 17, rue des Martyrs, 38054 Grenoble Cedex 9, France

[5] Nucleus Millenium Chemical Processes and Catalysis (CPC), Laboratorio de Quimica Teorica Computacional (QTC), Facultad de Quimica, Pontificia Universidad Catolica de Chile, Santiago, Chile

Keywords Conceptual DFT · Chemical potential · Reaction electronic flux · Spontaneous process

1 Introduction

In quantum chemistry, reaction mechanisms are often investigated by focusing on relevant points on the Born–Oppenheimer potential energy surface. Key intermediates and transition states along the so-called reaction path are of particular interest. Even though it can be defined in several nonequivalent fashions [1–5], the reaction path can always be parametrized using the so-called reaction coordinate (RC), Fukui's intrinsic RC (IRC) [2, 3] certainly being the most popular one. In this paper, it will be denoted ξ. Very often, monitoring the evolution of appropriate electronic properties with respect to the reaction coordinate sheds some light on the main forces that drive the studied process.

Among those properties, those that stem from conceptual DFT [6–9] have lately attracted a blooming interest. Thus, several papers have reported the monitoring of density-based descriptors, such as the chemical potential [10, 11], the chemical hardness [11–19], the electrophilicity index [18–23], chemical information [24] or polarizabilities [25], the conjunction [26] of such analyses affording a comprehensive overview of the physicochemical events occurring during a given chemical reaction. This approach has prompted the proposal of new descriptors. In this context, Toro-Labbé [10] defined the reaction force (RF) as the opposite of the derivative of the total molecular energy with respect to the intrinsic reaction coordinate.

$$F(\xi) = -\frac{\mathrm{d}E_{\mathrm{mol}}}{\mathrm{d}\xi} \tag{1}$$

It allows a formal decomposition of the chemical path into three regions, namely the reactant zone, the transition state one and finally the product zone [27–29]. This approach has been applied successfully to many chemical processes [30–33]. Another useful index has been defined as the opposite of the first derivative of the electronic chemical potential with respect to the intrinsic reaction coordinate and has been coined reaction electronic flux (REF) [34–39].

$$J(\xi) = -\frac{\mathrm{d}\mu}{\mathrm{d}\xi} \tag{2}$$

This latter descriptor has been used for several years as an indicator of electron density reorganization during a chemical process. Moreover, the sign of the REF is used to indicate whether an electronic process is allegedly spontaneous. Indeed, by analogy with thermodynamics, processes exhibiting positive values of REF have been surmised as spontaneous while processes exhibiting negative values of the REF have been considered as unspontaneous.

Another well-accepted feature is that processes exhibiting nil REF (constant chemical potential) are mainly associated with structural reorganization, such as geometrical rearrangements. In those cases, the electron density modification is supposedly minimal. For instance, for the reactants and products, the REF is often nil, $J(\xi) = 0$ indicating that electronic activity is at its basal and satisfies the equilibrium condition. These characteristics have been observed in lots of chemical reactions making this descriptor an excellent indicator of the existence and location along the chemical process of the electronic activity.

However, even though several tentative rationalizations have been undertaken, to our best knowledge, no definitive theories have been published so far. The ambition of this work is certainly not to lay the basis of such theory, but to put forward some arguments that unveil the physical basis behind the aforementioned characteristic of the REF.

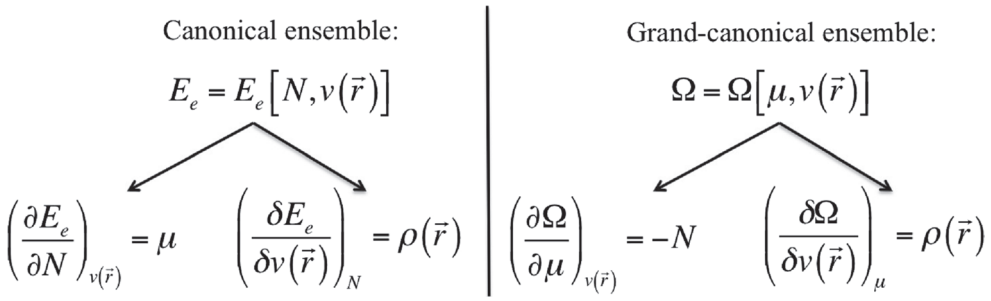
The article is therefore organized as follows. In Sect. 2, two analytical expressions of the reaction flux are derived, one in the canonical ensemble and the second in the grand canonical ensemble. Still in Sect. 2, the consistency of both approaches is shown. Then, Sect. 3 is dedicated to the interpretation of several equations supporting what is now well admitted about the chemical meaning of the REF. The paper final section summarizes the findings and provides some concluding remarks.

2 Analytical formulation of the reaction electronic flux

In conceptual DFT, two main ensembles (in analogy with statistical physics) are generally considered, namely the canonical and the grand canonical ensembles. Within the former, the electronic energy (E_e) is considered as a function of the number of electrons N and as a functional of the external potential $v(\vec{r})$. In the latter, the grand potential is described as a function of chemical potential μ and as a functional of the external potential $v(\vec{r})$. These two descriptions are linked by a simple Legendre transform so that they carry in principle the same physicochemical information, but encoded in different representations.

The canonical ensemble is more natural and more familiar to chemists since it deals with isolated molecular species (our main target). On the other hand, the grand canonical, more suited for species in interaction with their surroundings (like electron exchange with solvents), presents the advantage to directly consider the electronic chemical potential as a primary variable.

Once the ensemble is chosen, descriptors are then defined by inspecting the successive derivatives of the chosen "thermodynamic potential" with respect to their natural variables. The first derivatives in each ensemble are given by the equations gathered in Scheme 1:



Scheme 1 First derivatives in the canonical and grand canonical ensembles

2.1 Reaction electron flux in the canonical ensemble

In the canonical ensemble, the variation of the electronic energy for a chemical reaction proceeding at constant number of electrons reads:

$$\frac{\mathrm{d}E_e\left[N,v_\xi(\vec{r})\right]}{\mathrm{d}\xi}=\left(\frac{\partial E_e}{\partial N}\right)_{v(\vec{r})}\underbrace{\frac{\mathrm{d}N}{\mathrm{d}\xi}}_{0}+\int\left(\frac{\delta E_e}{\delta v_\xi(\vec{r})}\right)\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r=\int\rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r \tag{3}$$

$\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}$ itself can be expressed in terms of μ_ξ and $F[\rho_\xi]$ by using the fact that the electronic chemical potential can also be considered as the Lagrange multiplier associated with the normalization constraint in the Hohenberg–Kohn variational procedure [40] ($\mu_\xi=v_\xi(\vec{r})+\frac{\delta F[\rho_\xi]}{\delta\rho_\xi(\vec{r})}$), so that Eq. 3 becomes:

$$\frac{\mathrm{d}E_e\left[N,v_\xi(\vec{r})\right]}{\mathrm{d}\xi}=\int\rho_\xi(\vec{r})\frac{\mathrm{d}\mu_\xi}{\mathrm{d}\xi}\mathrm{d}^3r-\int\rho_\xi(\vec{r})\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{\delta F\left[\rho_\xi\right]}{\delta\rho_\xi(\vec{r})}\right)\mathrm{d}^3r. \tag{4}$$

The density normalization constraint now gives (since $\frac{\mathrm{d}\mu_\xi}{\mathrm{d}\xi}$ does not depend on real space coordinates):

$$\frac{\mathrm{d}E_e\left[N,v_\xi(\vec{r})\right]}{\mathrm{d}\xi}=N\frac{\mathrm{d}\mu_\xi}{\mathrm{d}\xi}-\int\rho_\xi(\vec{r})\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{\delta F\left[\rho_\xi\right]}{\delta\rho_\xi(\vec{r})}\right)\mathrm{d}^3r. \tag{5}$$

Finally, equating the right-hand side of Eq. 3 with that of Eq. 5 and identifying the REF straightforwardly provide:

$$J_{\text{canonical}}(\xi)=-\frac{1}{N}\left[\int\rho_\xi(\vec{r})\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{\delta F\left[\rho_\xi\right]}{\delta\rho_\xi(\vec{r})}\right)\mathrm{d}^3r+\int\rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r\right]. \tag{6}$$

2.2 Reaction electron flux in the grand canonical ensemble

In the grand canonical ensemble, the variation of the grand potential reads:

$$\frac{\mathrm{d}\Omega\left[\mu_\xi,v_\xi(\vec{r})\right]}{\mathrm{d}\xi}=\left(\frac{\partial\Omega}{\partial\mu_\xi}\right)_{v_\xi(\vec{r})}\frac{\mathrm{d}\mu_\xi}{\mathrm{d}\xi}+\int\left(\frac{\delta\Omega\left[\rho_\xi\right]}{\delta v_\xi(\vec{r})}\right)\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r=-N\frac{\mathrm{d}\mu_\xi}{\mathrm{d}\xi}+\int\rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r \tag{7}$$

As shown by Parr and Gàzquez [41], the grand potential can also be considered as a functional of the electron density through the hardness functional $\Omega=\Omega[\rho]=-H[\rho]$, so that:

$$N\frac{\mathrm{d}\mu_\xi}{\mathrm{d}\xi}=\frac{\mathrm{d}H\left[\rho_\xi\right]}{\mathrm{d}\xi}+\int\rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r. \tag{8}$$

As a result, the REF in the grand canonical ensemble reads:

$$J_{\text{grand-canonical}}(\xi)=-\frac{1}{N}\frac{\mathrm{d}H\left[\rho_\xi\right]}{\mathrm{d}\xi}-\frac{1}{N}\int\rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r \tag{9}$$

In the next section, the consistency of the two approaches is discussed.

2.3 Equivalence of the two formulations

The canonical and grand canonical ensembles thus provide the two following expressions for the reaction electronic flux:

$$\begin{cases}J_{\text{canonical}}(\xi)=-\frac{1}{N}\int\rho_\xi(\vec{r})\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{\delta F\left[\rho_\xi\right]}{\delta\rho_\xi(\vec{r})}\right)\mathrm{d}^3r-\frac{1}{N}\int\rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r\\ J_{\text{grand-canonical}}(\xi)=-\frac{1}{N}\frac{\mathrm{d}H\left[\rho_\xi\right]}{\mathrm{d}\xi}-\frac{1}{N}\int\rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r\end{cases} \tag{10}$$

Using the definition of the hardness kernel [41–44] $\eta\left(\vec{r},\vec{r}'\right)=\frac{\delta^2F[\rho]}{\delta\rho(\vec{r})\delta\rho(\vec{r}')}$, one has:

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{\delta F\left[\rho_\xi\right]}{\delta\rho_\xi(\vec{r})}\right)=\frac{1}{\mathrm{d}\xi}\left(\frac{\delta F}{\delta\rho_\xi(\vec{r})}\left[\rho_\xi+\mathrm{d}\rho_\xi\right]-\frac{\delta F\left[\rho_\xi\right]}{\delta\rho_\xi(\vec{r})}\right)=\frac{1}{d\xi}\int\eta\left(\vec{r},\vec{r}'\right)\mathrm{d}\rho_\xi\left(\vec{r}'\right)\mathrm{d}^3r', \tag{11}$$

Introducing then the so-called $\eta^D\left(\vec{r}'\right)$ density-based local hardness [42–46]:

$$\begin{aligned}\int\rho_\xi(\vec{r})\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{\delta F\left[\rho_\xi\right]}{\delta\rho_\xi(\vec{r})}\right)\mathrm{d}^3r&=\int\int\rho_\xi(\vec{r})\eta\left(\vec{r},\vec{r}'\right)\frac{\mathrm{d}\rho_\xi\left(\vec{r}'\right)}{\mathrm{d}\xi}\mathrm{d}^3r\mathrm{d}^3r'\\&=\int\mathrm{d}^3r'\frac{\mathrm{d}\rho_\xi\left(\vec{r}'\right)}{\mathrm{d}\xi}\underbrace{\int\rho_\xi(\vec{r})\eta\left(\vec{r},\vec{r}'\right)d^3r}_{N\eta_\xi^D\left(\vec{r}'\right)}\\&=N\int\eta_\xi^D\left(\vec{r}'\right)\frac{\mathrm{d}\rho_\xi\left(\vec{r}'\right)}{\mathrm{d}\xi}\mathrm{d}^3r'.\end{aligned} \tag{12}$$

As, derived by Parr and Gazquez [41], $\frac{\delta H[\rho]}{\delta\rho(\vec{r})}=N\eta^D(\vec{r})$, in which $\eta^D(\vec{r})$ is the local hardness based upon the electron density. Equation 12 then becomes:

$$\int\rho_\xi(\vec{r})\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{\delta F\left[\rho_\xi\right]}{\delta\rho_\xi(\vec{r})}\right)\mathrm{d}^3r=\int\frac{\delta H\left[\rho_\xi\right]}{\delta\rho_\xi\left(\vec{r}'\right)}\frac{\mathrm{d}\rho_\xi\left(\vec{r}'\right)}{\mathrm{d}\xi}\mathrm{d}^3r'=\frac{\mathrm{d}H\left[\rho_\xi\right]}{\mathrm{d}\xi}, \tag{13}$$

where the last equality comes from the chain rule for functional derivatives. Equation 13 thus establishes the equivalence of the two approaches. It will be now convenient for our purposes to use the following relationship:

$$J(\xi) = -\int \eta_\xi^D(\vec{r})\frac{\mathrm{d}\rho_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r - \frac{1}{N}\int \rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r \qquad (14)$$

Note that this equation is totally consistent with that obtained by Berkowitz and Parr [47].

3 Interpreting the expressions

In this section, the REF expressions obtained in the previous one are analyzed to lay the foundation of the common interpretation of the REF sign. To this aim, it is instructive to consider the variation of the interaction energy between the electrons and the nuclei E_{ne} along the reaction path:

$$\frac{\mathrm{d}E_{\mathrm{ne}}}{\mathrm{d}\xi} = \int \rho_\xi(\vec{r})\frac{\mathrm{d}v_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r + \int v_\xi(\vec{r})\frac{\mathrm{d}\rho_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r, \qquad (15)$$

so that:

$$J(\xi) = \frac{1}{N}\left(-\frac{\mathrm{d}E_{\mathrm{ne}}}{\mathrm{d}\xi} + \int \left(v_\xi(\vec{r}) - N\eta_\xi^D(\vec{r})\right)\frac{\mathrm{d}\rho_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r\right). \qquad (16)$$

The local hardness can be decomposed into a pure classical Coulombic contribution and that gathering (denoted here $\eta_\xi^{(1)}(\vec{r})$) all nonclassical effects (kinetic and exchange-correlation effects). Indeed, as the hardness kernel [42] can be decomposed according to $\eta(\vec{r},\vec{r}') = \frac{1}{\|\vec{r}-\vec{r}'\|} + \eta_\xi^{(1)}(\vec{r},\vec{r}')$, the local hardness can be expressed by:

$$\eta_\xi^D(\vec{r}) = \frac{1}{N}\int \frac{\rho(\vec{r}')}{\|\vec{r}-\vec{r}'\|}\mathrm{d}^3r' + \eta_\xi^{(1)}(\vec{r}). \qquad (17)$$

The first term in the right-hand side is actually proportional to the electronic part of the molecular electrostatic potential (MEP) which is defined by:

$$\mathrm{MEP}(\vec{r}) = v(\vec{r}) - \int \frac{\rho(\vec{r}')}{\|\vec{r}-\vec{r}'\|}\mathrm{d}^3r'. \qquad (18)$$

The REF can thus be expressed by:

$$J(\xi) = \frac{1}{N}\left(-\frac{\mathrm{d}E_{\mathrm{ne}}}{\mathrm{d}\xi} + \int \mathrm{MEP}(\vec{r})\frac{\mathrm{d}\rho_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r\right) + A\left[\eta_\xi^{(1)}, \rho_\xi\right]. \qquad (19)$$

If, as a first approximation, the $\eta_\xi^{(1)}(\vec{r})$ contribution (leading to the $A[\eta_\xi^{(1)}, \rho_\xi]$ one) is neglected, the REF simply reads:

$$J(\xi) \approx \frac{1}{N}\left(-\frac{\mathrm{d}E_{\mathrm{ne}}}{\mathrm{d}\xi} + \int \mathrm{MEP}_\xi(\vec{r})\frac{\mathrm{d}\rho_\xi(\vec{r})}{\mathrm{d}\xi}\mathrm{d}^3r\right). \qquad (20)$$

Interestingly, this expression only involves properties of the N electron species and does not require the explicit evaluation of the derivative with respect to the number of electrons (which is generally achieved using finite difference linearization, a procedure that can be troublesome for instance in some anionic systems).

Equation 20 also greatly facilitates the interpretation of the REF sign in terms of energies variations. Indeed, the REF is positive when either both right-hand side terms (the opposite of the variation of the nuclei–electrons interaction energy and the MEP-related contribution) are positive or when one of those terms is positive and higher in magnitude than the other. For a steady discussion, each term is analyzed separately. On the one hand, a positive value of the term $-\frac{1}{N}\frac{\mathrm{d}E_{ne}}{\mathrm{d}\xi}$ contributes to give the REF a positive value. Since the number of electrons is always positive, this term will be positive when the absolute value of the electron–nuclei interaction energy increases along the reaction coordinate, which means that the electron–nuclei interaction energy is decreasing.

To make a link with the mechanism of bonds formation, let us advocate the celebrated Slater's analysis of diatomic bond formation by means of the quantum virial theorem [48]. At any separation distance R (which exactly plays the role of the reaction coordinate):

$$E_{\mathrm{mol}}(R) = -\frac{1}{2}(E_{\mathrm{ne}} + E_{\mathrm{ee}}) - \frac{R}{2}\frac{\mathrm{d}E_{\mathrm{mol}}}{\mathrm{d}R}, \qquad (21)$$

which enlightens the fundamental role of E_{ne}. More specifically, covalent bonds are characterized by a strong electron concentration between the nuclei, so that the electrons in this zone are stabilized by the attraction of both nuclei. Such attraction induces more and more negative E_{ne} values, fully consistent with positive REF values partly driven by the sign of $-\frac{1}{N}\frac{\mathrm{d}E_{\mathrm{ne}}}{\mathrm{d}\xi}$.

As an integral, the sign of the other term in Eq. 20 is trickier to ascribe. Overall, the sign will be positive when, along the reaction coordinate, the electron density increases in molecular positions where the molecular electrostatic MEP is positive. Similarly, the sign will also be positive when, along the reaction coordinate, the electron density decreases in locations where the MEP is negative.

This means that the sign of the second term in the flux expression is positive when the electron density flees negative values of the MEP and flows toward positive values of the MEP. In such cases, we expect the electrons to be stabilized in such cases. Such phenomena can be illustrated in the particular case of the creation of a dative bond resulting from the interaction between a Lewis acid and a Lewis base. Indeed, the lone pair of the Lewis base is a region where the MEP is negative. When the lone pair coordinates, it will transfer some electrons to its partner (from an orbital point of view, this is achieved by the electron donation to an unoccupied orbital of the other fragment). In other words, in this lone pair region where $\mathrm{MEP}_\xi(\vec{r})$ is negative, $\frac{\mathrm{d}\rho_\xi(\vec{r})}{\mathrm{d}\xi}$ will be positive, so that this process gives

a positive contribution to the molecular reaction electronic flux. Of course, the very same interpretation can be done from the viewpoint of the Lewis acid.

This is of particular importance for noncovalent bonds such as halogen and pnicogen ones whose formation can be described by the paradigm of the so-called σ-hole [49–52]. We recall that this concept refers to a region of positive molecular electrostatic potential (MEP) on the outer side of an electronegative atom, which can form a bond by interacting with the available lone pair carried by another partner. Such an electrostatic description is certainly compatible with the physics embodied by Eq. 20.

It can be noticed that this whole analysis is also consistent with the recent formula [53] we recently derived for the chemical potential within the framework of Kohn–Sham DFT:

$$\mu = \alpha_{\mathrm{mol}} T_{\mathrm{s}} + \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i^{\mathrm{KS}}, \tag{22}$$

where α_{mol} is a molecule-dependent parameter, T_{s} is the kinetic energy of the fictitious system, and the $\varepsilon_i^{\mathrm{KS}}$ denote the energies of the (occupied) Kohn–Sham orbitals. The more the electrons are stabilized (by example by flowing toward positive MEP values), the more negative $\varepsilon_i^{\mathrm{KS}}$ will be, so that the chemical potential will decrease, in accordance with positive REF values.

All those processes can be said spontaneous from the electron perspective since they tend to decrease the total electron energy. However, the outcome of a chemical process is not only driven by electron energy but also depends on the variation of the nucleus–nucleus repulsion. A process that would be favorable from an electron perspective can require such an important energy to reorganize the nuclei location that it would be overall not spontaneous. As mentioned in all papers dealing with the REF, it can be stated that the sign of the REF only indicates whether a process is spontaneous from the electron perspective.

Besides, the values of the REF can be quite small in specific parts of the chemical path and it can even reach zero, at the stationary points of the reaction energy profile, namely the reactants and the product, and very likely at the transition state. Equation 23 provides an interesting relation for those cases.

$$J(\xi) = 0 \Rightarrow \frac{\mathrm{d}E_{\mathrm{ne}}}{\mathrm{d}\xi} = \int \mathrm{MEP}(\vec{r}) \frac{\mathrm{d}\rho_{\xi}(\vec{r})}{\mathrm{d}\xi} \mathrm{d}^3 r, \tag{23}$$

It is plain to see that for zero REF, the evolution of the electron–nuclei interaction energy is only due to the evolution of the electron density driven by the molecular electrostatic potential. This evolution of the electron density probably constitutes a baseline of the electronic activity.

4 Conclusion

In this paper, two analytical expressions of the reaction electron flux have been proposed. The first one stems from the canonical ensemble, while the second arises from the grand canonical ensemble. The two approaches are shown to be consistent, which underlines that the physics behind the two ensembles is the same. It is worth noticing that an expression describing the REF for optimum variation of the electron energy is derived. This particular expression shows that the REF is related to the electron density variation driven by the local hardness acting as a potential. Therefore, the REF can be seen as an indicator of electron density reorganization. Another finding is the general expression linking this descriptor to the evolution of the electron density under the influence of the molecular electrostatic potential and to the evolution of the electron–nuclei interaction energy. The analysis of all those expressions supports the well-accepted interpretation given to the sign of the REF. Indeed, it is shown that positive values of the REF are related to processes that tend to stabilize the electron density, while negative values of the REF can be ascribed to processes that destabilize the electron density. So from an electron-following perspective, the positive/negative sign of the REF is related to spontaneous/unspontaneous phenomena. However, it must be recalled that this interpretation is only a view from the electron perspective and does not account for the outcome of chemical reactions on their whole, the evolution of a chemical transformation being driven by both the electron energy evolution and the nuclei–nuclei repulsion.

Acknowledgments The research benefited from the support of Aviesan ITMO Cancer within the ‘‘Cancer Plan 2009–2013’’ and the application of Action 3.3. CM, AG, HC, ATL, BH and SG-O thank the joint program ECOS-CONICYT through action project ECOS No. C11E03. V.T. and L.J. gratefully acknowledge the Centre National de la Recherche Scientifique (CNRS) for a “Chaire d’Excellence” at the University of Rouen, the LABEX SynOrg for funding, and the CRIHAN computational centre for providing computational resources. HC acknowledges the GENCI/CINES for HPC resources/computer time (Project cpt2130). ATL, BH, SG-O acknowledge Project Nucleus Millenium CPC: ICM No. 120082 and Projects FONDECYT Nos. 1120093, 1130072 and 1141098.

References

1. Quapp W (2008) Chemical reaction paths and calculus of variations. Theor Chem Acc 121:227–237
2. Fukui K (1970) Formulation of the reaction coordinate. J Phys Chem 74:4161–4163
3. Fukui K (1981) The path of chemical reaction—the irc approach. Acc Chem Res 14:363–368
4. Basilevsky MV, Shamov AG (1981) The local definition of the optimum ascent path on a multi-dimensional potential energy

surface and its practical application for the location of saddle points. Chem Phys 60:347–358
5. Elber R, Karplus M (1987) A method for determining reaction paths in large molecules: application to myoglobin. Chem Phys Lett 139:375–380
6. Geerlings P, De Proft F, Langenaeker W (2003) Conceptual density functional theory. Chem Rev 103:1793–1873
7. Gázquez JL (2008) Perspectives on the density functional theory of chemical reactivity. J Mex Chem Soc 52:3–10
8. Chermette H (1998) Chemical reactivity indexes in density. J Comput Chem 20:129–154
9. Anderson JSM, Melin J, Ayers PW (2007) Conceptual density-functional theory for general chemical reactions, including those that are neither charge- nor frontier-orbital-controlled. 1. Theory and derivation of a general-purpose reactivity indicator. J Chem Theory Comput 3:358–374
10. Toro-Labbé A (1999) Characterization of chemical reactions from the profiles of energy, chemical potential, and hardness. J Phys Chem A 103:4398–4403
11. Bonnet M-L, Tognetti V (2011) The influence of density functional approximations on the description of $LiH + NH_3 \rightarrow LiNH_2 + H_2$ reaction. Chem Phys Lett 511:427–433
12. Datta D (1992) "Hardness profile" of a reaction path. J Phys Chem 96:2409–2410
13. De Proft F, Chattaraj PK, Ayers PW et al (2008) Initial hardness response and hardness profiles in the study of Woodward–Hoffmann rules for electrocyclizations. J Chem Theory Comput 4:595–602
14. Ess DH (2009) Distortion, interaction, and conceptual DFT perspectives of MO4–alkene (M = Os, Re, Tc, Mn) cycloadditions. J Org Chem 74:1498–1508
15. Mineva T, Russo N, Sicilia E, Toscano M (1997) Density functional potential-energy hypersurface and reactivity indices in the isomerization of X_3H^+ (X = O, S, Se, Te). J Chem Soc, Faraday Trans 93:3309–3312
16. Mineva T, Sicilia E, Russo N (1998) Density-functional approach to hardness evaluation and its use in the study of the maximum hardness principle. J Am Chem Soc 120:9053–9058
17. Ghanty TK, Gosh SK (1996) A Density functional approach to hardness, polarizability, and valency of molecules in chemical reactions. J Phys Chem 100:12295–12298
18. Labet V, Morell C, Toro-Labbé A, Grand A (2010) Is an elementary reaction step really elementary? Theoretical decomposition of asynchronous concerted mechanisms. Phys Chem Chem Phys 12:4142–4151
19. Pan S, Solà M, Chattaraj PK (2013) On the validity of the maximum hardness principle and the minimum electrophilicity principle during chemical reactions. J Phys Chem A 117:1843–1852
20. Parthasarathi R, Elango M, Subramanian V, Chattaraj PK (2005) Variation of electrophilicity during molecular vibrations and internal rotations. Theor Chem Acc 113:257–266
21. Morell C, Labet V, Grand A, Chermette H (2009) Minimum electrophilicity principle: an analysis based upon the variation of both chemical potential and absolute hardness. Phys Chem Chem Phys 11:3417–3423
22. Chamorro E, Chattaraj PK, Fuentealba P (2003) Variation of the electrophilicity index along the reaction path. J Phys Chem A 107:7068–7072
23. Morell C, Herrera B, Gutiérrez-Oliva S et al (2012) A relation between Different Scales of Electrophilicity: Are the scales consistent along a chemical reaction? J Phys Chem A 116:7074–7081
24. Borgoo A, Jaque P, Toro-Labbé A et al (2009) Analyzing Kullback–Leibler information profiles: an indication of their chemical relevance. Phys Chem Chem Phys 11:476–482
25. Jędrzejewski M, Ordon P, Komorowski L (2013) Variation of the electronic dipole polarizability on the reaction path. J Mol Model 19:4203–4207
26. Morales-Bayuelo A, Pan S, Caballero J, Chattaraj PK (2015) Analyzing torquoselectivity in electrocyclic ring opening reactions of trans-3,4-dimethylcyclobutene and 3-formylcyclobutene through electronic structure principles. Phys Chem Chem Phys 17:23104–23111
27. Herrera B, Toro-Labbé A (2004) The role of the reaction force to characterize local specific interactions that activate the intramolecular proton transfers in DNA basis. J Chem Phys 121:7096–7102
28. Toro-Labbé A, Gutiérrez-Oliva S, Murray JS, Politzer P (2009) The reaction force and the transition region of a reaction. J Mol Model 15:707–710
29. Politzer P, Toro-Labbé A, Gutiérrez-Oliva S et al (2005) The reaction force: three key points along an intrinsic reaction coordinate. J Chem Sci 117:467–472
30. Rincón E, Jaque P, Toro-Labbé A (2006) Reaction force analysis of the effect of Mg(II) on the 1,3 intramolecular hydrogen transfer in thymine. J Phys Chem A 110:9478–9485
31. Labet V, Morell C, Grand A, Toro-Labbé A (2008) Theoretical study of cytosine deamination from the perspective of the reaction force analysis. J Phys Chem A 112:11487–11494
32. Cerón ML, Herrera B, Araya P et al (2011) The mechanism of methanol decomposition by CuO. A theoretical study based on the reaction force and reaction electronic flux analysis. J Mol Model 17:1625–1633
33. Jaque P, Toro-Labbé A, Politzer P, Geerlings P (2008) Reaction force constant and projected force constants of vibrational modes along the path of an intramolecular proton transfer reaction. Chem Phys Lett 456:135–140
34. Echegaray E, Toro-Labbé A (2008) Reaction electronic flux: a new concept to get insights into reaction mechanisms. Study of model symmetric nucleophilic substitutions. J Phys Chem A 112:11801–11807
35. Vogt-Geisse S, Toro-Labbé A (2009) The mechanism of the interstellar isomerization reaction $HOC^+ \rightarrow HCO^+$ catalyzed by H_2: new Insights from the reaction electronic flux. J Chem Phys 130:244308
36. Flores-Morales P, Gutiérrez-Oliva S, Silva E, Toro-Labbé A (2010) The reaction electronic flux: a new descriptor of the electronic activity taking place during a chemical reaction. Application to the characterization of the mechanism of the Schiff's base formation in the Maillard reaction. J Mol Struct Theochem 943:121–126
37. Martínez-Araya JI, Toro-Labbé A (2015) Reaction electronic flux as a fluctuation of relative interatomic electronic populations. J Phys Chem C 119:3040–3049
38. Inostroza-Rivera R, Yahia-Ouahmed M, Tognetti V et al (2015) Atomic decomposition of conceptual DFT descriptors: application to proton transfer reactions. Phys Chem Chem Phys 17:17797–17808
39. Lamsabhi AM, Gutiéerrez-Oliva S, Mó O, Toro-Labbé A, Yáñez M (2015) Effects of the ionization in the tautomerism of uracil: a reaction electronic flux perspective. J Comput Chem. doi:10.1002/jcc.24054
40. Parr RG, Donnelly RA, Levy M, Palke WE (1978) Electronegativity: the density functional viewpoint. J Chem Phys 68:3801–3807
41. Parr RG, Gazquez JL (1993) Hardness functional. J Phys Chem 7:3939–3940
42. Liu S, De Proft F, Parr RG (1997) Simplified models for hardness kernel and calculations of global hardness. J Phys Chem A 101:6991–6997

43. Torrent-Sucarrat M, Duran M, Solà M (2002) Global hardness evaluation using simplified models for the hardness kernel. J Phys Chem A 106:4632–4638
44. Fuentealba P (1995) A local model for the hardness kernel and related reactivity parameters in density functional theory. J Chem Phys 103:6571
45. Ayers PW, Parr RG (2008) Beyond electronegativity and local hardness: higher-order equalization criteria for determination of a ground-state electron density. J Chem Phys 129:054111
46. Ayers PW, Parr RG (2008) Local hardness equalization: exploiting the ambiguity. J Chem Phys 128:184108
47. Berkowitz M, Ghosh SK, Parr RG (1985) On the concept of local hardness in chemistry. J Am Chem Soc 107:6811–6814
48. Slater JC (1933) The virial and the molecular structure. J Chem Phys 1:687–691
49. Clark T, Hennemann M, Murray JS, Politzer P (2007) Halogen bonding: the σ-hole. J Mol Model 13:291–296
50. Murray JS, Lane P, Politzer P (2009) Expansion of the σ-hole concept. J Mol Model 15:723–729
51. Clark T (2013) σ-Holes. WIRES Comput Mol Sci 3:13–20
52. Tognetti V, Joubert L (2015) Electron density Laplacian and halogen bonds. Theor Chem Acc 134:90
53. Tognetti V, Morell C, Joubert L (2015) Atomic electronegativities in molecules. Chem Phys Lett 635:111–115

Chapter 2

DNA damage formation

Everyday, each cell of living organisms undergoes 10,000 to 1 million attacks from mutagenic agents. The latter react with the nucleotides, resulting in the formation of damages. DNA lesions' chemical natures are diverse : strand breaks, DNA-DNA inter- and intrastrand cross-links, DNA-protein cross-linking, chemical modifications of nucleobases, abasic sites (absence of a nucleobase), etc - see Figure 2.0.1. Damaging can be induced by radiations (UV, γ), by reaction with radicals (ROS), or even by cell metabolism - erronous replication, release of a deleterious reaction intermediate, etc. Up to now, more than 70 modified nucleosides have been isolated and characterized[88]. Since few decades many efforts have been made in order to understand how these lesions are generated. Indeed, DNA lesions can be genotoxic and/or cytotoxic if not efficiently repaired, leading to cancerogenesis, cell death, aging, etc. Hence, it is of utmost importance to understand their formation mechanisms, in the perspective to develop solutions to prevent damaging (eg sun screens) and to fight againt its deleterious consequences - eg therapeutic agents design. Albeit the knowledge concerning DNA damaging processes has increased since decades, there still remains a lack of information concerning key aspects of the complex chemical phenomena they involve. In this context, quantum chemistry has proven its worth as a method of choice in order to get insights into energetics of lesions formation mechanisms, especially concerning UV-induced damages[89–92]. Indeed, reactions intermediates involved in these chemical phenomena can be high-lying in energy, hence unstable and very hard to isolate and characterize. In such cases, experimentalists can only make hypothesis about plausible mechanisms. Performing QM calculations gives access to the energetics of reaction paths proposed by experimentalists, assessing their feasability. However, DFT-based QM calculations are limited to systems containing few hundred atoms only ($\sim$ 10-bp nucleotides), limiting the calculations to reduced models albeit the macromolecular environment can reverse DNA damage inspection. Thereby, modelling of such systems requires to use hybrid QM/MM methods (see Chapter 1) in order to consider more accurate and realistic models. Nowadays, computational ressources improvement allow to use such hybrid methods on biomolecular systems reactivity including damaged DNA with explicit solvent[93–96].

In this thesis work, we led four projects on DNA damage formation, leading to the four papers enclosed at the end of this chapter - three published and one just submitted. These projects concerned the deleterious direct and indirect effects of radicals and UV light on DNA. Ultraviolet radiations are undoubtedly one of the most frequent causes of skin cancer, since cells are particularly exposed to these radiations which come directly from the sun light. UV light is able to induce DNA lesions directly or through photosensitization mechanisms, resulting in chemical and structural modifications of nucleobases, which

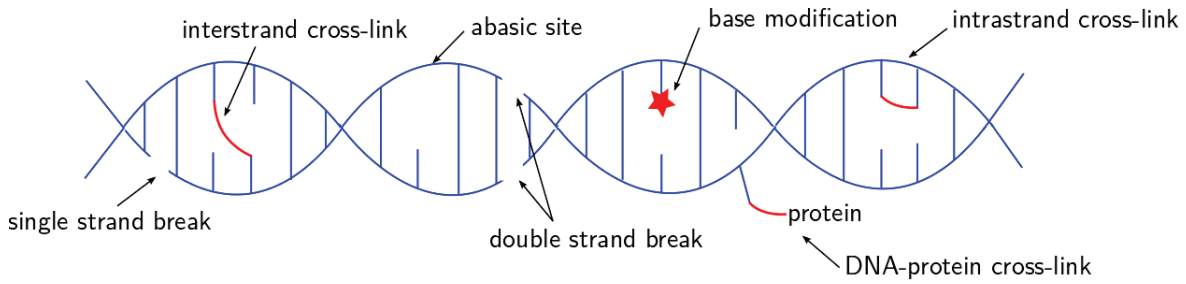


FIGURE 2.0.1: Schematic representation of some DNA damages types. The DNA helix is depicted in blue. Unsual covalent bondings and modified nucleobases are represented by red lines and stars respectively.

provoke geno- and/or cytotoxicity. Basics concerning these reactions are described in the following.

2.1 UV-induced DNA lesions

Albeit the atmosphere absorbs the high-energy UVC (<280nm) and a large part of longer wavelength UVB (280-315 nm)[97], DNA remains a key target for UV radiations. Indeed, DNA components absorb in the UVB, while UVA radiations can also induce photoproducts indirectly by action of an endo- or exogenous photosensitizer[98]. Photolesions generated by the reaction of these radiations on DNA mostly concern the pyrimidines nucleobases. The main UV-induced damages are the so-called cyclobutane pyrimidine dimers (CPDs) and the pyrimidine 6-4 pyrimidones (6-4PPs) moieties arising from the reaction between two adjacent pyrimidines - see Figure 2.1.1. CPDs are generated by a formal

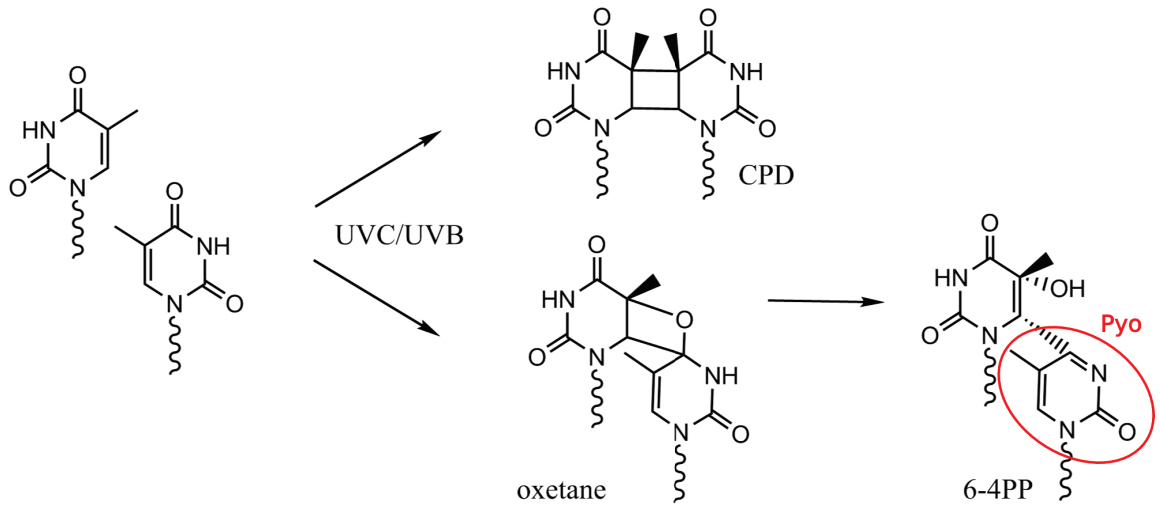


FIGURE 2.1.1: Schematic representation of the to main UV-induced DNA damages : CPD and 6-4PP. Here is the case of two adjacent thymines, but cytosines can also be involved in the reaction. Formation of the 6-4PP product involves the formation of an oxetane or an azetidine ring depending on the reacting 3' nucleobases (thymine or cytosine respectively). Noteworthy, this lesion exhibits a chromophore moiety, called Pyo (here surrounded by a red circle), which can act as an endogenous DNA photosensitizer[99]. One of this thesis projects concerned the impact of this Pyo moiety in double-stranded DNA - photophysics and triplet energy photoreactivity see publication [P1]. Glycosidic bonds are depicted in wavy lines.

[2+2] cycloaddition between on C5-C6 double bonds of the adjacent pyrimidines, while

6-4PPs arise from reaction between the ethylenic C5-C6 double bond of the 5' pyrimidine and the C4-N4 carbonyl (or imino) of the 3' one, following a Paternò–Büchi mechanism. Noteworthy, UVA favors the CPDs formation while UVB induces both photoproducts[100]. Furthermore, under UVA radiations, 6-4PPs can undergo photoisomerization, resulting in their Dewar valence isomer production[101]. They have been reported to also act as an endogenous photosensitizer[99]. However, no information exists yet to clearly explain which factors determine whether Dewar formation or photosensitization is favored.
CPDs and 6-4PPs are repaired by Transcription Coupled or Global Genome Nucleotide Excision Repair (see section 4.2.3), with a more efficient removal of 6-4PPs. This difference of repair rate is not yet fully understood, and we led a structural study on oligonucleotides containing these photoproducts in order to gain insights on features which may induce this different of removal efficiency - see Chapter 3. An other interesting fact is that the nucleosomal structure and the binding to transcriptional factors appear to prevent such damages formation, protecting DNA structure against UV radiations[102]. So far, UV-damages formation and consequences provide large matter of research, and one can hope that improvements of experimental and theoretical ressources might help to unravel pending questions on the subject.

2.2 Photosensitization

Endo- and exogenous photosensitizers are able to induce electron or energy transfer to DNA nucleobases after being excited by exposure to light. Two types of photosensitizaion reactions co-exist. One can simply describe these processes as follows : type I reactions induce one electron transfer from the excited photosensitizer (^3P) to a subtrate (nucleobase B), while type II consists in the production of singlet oxygen ($^1O_2$) or triplet thymine (^3T) by energy transfer from the photosensitizer[103].

$$TypeI : \qquad {}^3P + B \rightarrow P^+ + B^- \tag{2.1}$$

$$TypeII : \qquad {}^3P + {}^3O_2/T \rightarrow^? P + {}^1O_2/{}^3T \tag{2.2}$$

2.2.1 Type I photosensitizers

Type I photosensitizers are able to interact with a substrate and provoke one-electron transfer oxidation. These species can be exogenous (eg benzophenone) or endogenous (eg 6-4PP[99] and artificial bases[104]). Type I photosensitizers react directly on DNA nucleobases following a Dexter type mechanism. For instance, studies suggested that benzophenone efficiently provokes triplet triplet energy transfer (TTET) towards and adjacent thymine with which it interacts either by double insertion or minor groove binding[96,105].
On the other hand, type I photosensitizers also induce the formation of guanine chemical modifications by one electron oxidation. This reaction results in a guanine radical cation formation, which can react with several molecules such as water with which it produces 8-oxoG[106], or amino groups from proteins inducing DNA-protein cross-linking. Interestingly, albeit 8-oxoG can be produced by one-electron oxidation process, about 80% UV-photosensitized formation of 8-oxoG in DNA is induced by $^1O_2$ oxidation as the result of type II photosensitization mechanism[107].
Noteworthy, it is well established that in non-polar solvents, some type I photosensitizers such as ketoprofen and benzophenone can also induce type II reactions[108,109]. On the contrary, in aqueous medium it is not yet clear whether type I photosensitizers can also exhibit a singlet oxygen ($^1O_2$) scavenging activity[110,111]. Some studies suggest that this

is the case, especially in relatively recent studies about non-steroidal anti-inflammatory drugs (NSAID) such as ketoprofen and ibuprofen[108,112].

2.2.2 Type II photosensitizers : the singlet oxygen case

Type II photosensitizers mainly react with triplet molecular oxygen, promoting it to its singlet state. Singlet oxygen ($^1O_2$) is less stable hence much more reactive than its triplet ground state counterpart, and can be efficiently produced by photosensitization by a light-absorbing molecule in living organisms[113]. It belongs to the Reactive Oxygen Species (together with hydroxyl radical, hydrogen peroxide and superoxide anions), that are mainly responsible for oxidative stress in cells. They react with biomolecules, notably DNA components, leading to genotoxic and cytotoxic defects. Interestingly, $^1O_2$ lifetime τ has been evaluated to $\sim$4 μs in water, which provides enough time to be one of the main sources of damages for biomolecules[114]. When $^1O_2$ reacts with DNA components, its key target is the guanine, which exhibits the lowest oxidation potential among the four nucleobases hence being a hotspot for oxidation processes. The reaction between $^1O_2$ and

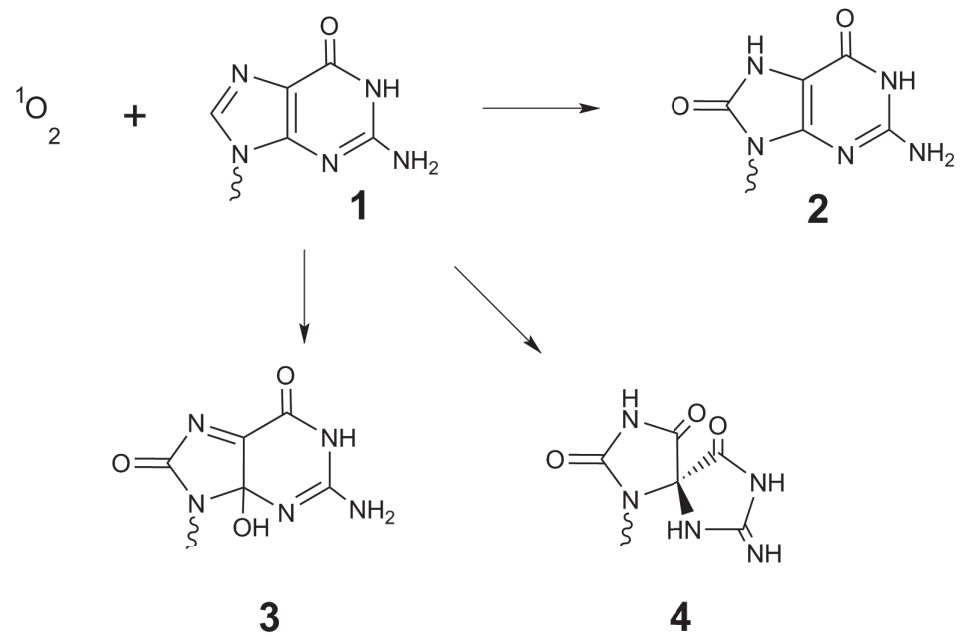


FIGURE 2.2.1: Main products resulting from singlet oxygen ($^1O_2$) attack onto guanine (**1**). -oxoG (**2**) is the major product, while spiroiminodihydantoin (**3**) and 4-hydroxy-8-oxo-7,8-dihydro-2'-deoxyguanosine (**4**) are generated with lower yields.

the guanine mainly leads to the formation of the very abundant 8-oxo-7,8-dihydroguanine (8oxoG) lesion in cells. Noteworthy, minor products of this reaction are the spiroiminodihydantoin and the 4-hydroxy-8-oxo-7,8-dihydro-2'-deoxyguanosine[106] (see Figure 2.2.1), and 8-oxoG itself can undergo degradation, leading to other defect moieties[115,116]. 8-oxoG is efficiently repaired by Base Excision Repair when detected by specific glycosylases - see Chapter 4. However, when passed across the DNA repair machinery, 8-oxoG can induce a mismatch with an adenine during DNA replication, hence leading to mutations.
Singlet oxygen also provokes single strand breaks formation at guanine sites[117], but this phenomenon is a relatively rare event since it occurs $\sim$17 time less than 8-oxoG formation[118].

2.3 Our projects in this framework

2.3.1 DNA photosensitization by an "insider": photophysics and triplet energy transfer of 5-methyl-2-pyrimidone deoxyribonucleoside [P1]

This project arised from a publication by Miranda and coworkers about 6-4PP propensity to act an endogenous photosensitizer[99]. Since experimental studies revealed the capacity of 6-4PP to induce triplet triplet energy transfer (TTET) to an adjacent thymine by its chromophore moiety (5-methyl-2-pyrimidone or Pyo), we assessed the deleterious action of such endogenous photosensitizer. Indeed, 6-4PP could induce the formation of triplet state thymine that might generate other lesions in its vicinity (eg CPD), resulting in tandem lesions that are very challenging for the DNA repair machinery. Thus, it was important to understand to what extent the 6-4PP chromophore, Pyo, could be an efficient photosensitizer - see Figure 2.1.1. We chose to study the Pyo artificial nucleobase as a first model.
First, classical MD simulations of the Pyo-containing oligonucleotide were performed in order to assess the structural impact of Pyo presence on the DNA helix. Then, we investigated the TTET from Pyo to the adjacent thymine by static QM/MM calculations along an interpolated reaction coordinate. This study has shown the ease of this transfer (barrier of 0.4 eV), highlighting the endogenous photosensitizer feature of Pyo, which may be relevent as cancer therapy agent but also reveal the deleterious effect of 6-4PP on DNA. Noteworthy, a study of the entire 6-4PP moiety reactivity would be interesting : a recent study shown that the photosensitizer feature of Pyo is enhanced when included in 6-4PP[119] - comparing to Pyo as an artificial nucleobase. We performed a first structural study of a 6-4PP containing 16-bp oligonucleotide (see Chapter 3). Modeling the TTET from 6-4PP to a thymine with structures extracted from this study would be the next step.
I realized the classical Molecular Dynamics simulations which necessitated non canonical parameters for the Pyo nucleobase. QM/MM calculations were performed by our colleagues from Nancy : Dr. Antonio Monari and his PhD student Hugo Gattuso. This study was published in *Chemistry - A European Journal*.

2.3.2 Probing the reactivity of singlet oxygen with purine [P3] and Singlet oxygen attack on guanine: reactivity and structural signature within the B-DNA Helix [P4]

This study was motivated by discussions with our collaborator from the CEA of Grenoble, Dr. Jean-Luc Ravanat. He was wondering about the mechanisms of formation of 8-oxoguanine (8-oxoG) which is a very abundant and genotoxic lesion, arising from singlet oxygen ($^1O_2$) reaction onto guanine - see Figure 2.2.1. A mechanistic hypothesis that consists in the formation of an endoperoxide intermediate is usually proposed. However, such intermediate was only characterized by low-temperature NMR for isolated guanine[120], thus no real proof of its existence was known. Thereby, we assessed the plausibility of this intermediate by QM calculations on isolated nucleobases in a preliminary study. We shown that the energetical pathway corresponding to 8-oxoG formation is lower that for adenine, in agreement with experimental studies. Then, we extended the project to the study of this intermediate in B-DNA by MM-MD and QM/MM-MD methods in order to assess the DNA helix environment effect on this reaction. Our cal-

culations highlighted the fact that the mechanism proposed by Dr. Ravanat is valide and pass through a charge-separated intermediate, exhibiting low activation energies for the endoperoxide formation. Results brought out the importance of the environment on $^1O_2$ addition, exhibiting a lower activation energy (∼6 kcal/mol) of the energetically decisive step comparing to study on isolated nucleobases. Thus, our results have reinforced the viability of our collaborator's hypothesis, palliating the lack of experimental data about the endoperoxide existence. This project led to two publications, the first one focused on QM calculations and MM-MD simulations published in *Nucleic Acids Research* ([P3]), the second one concerning the reactivity study in B-DNA by QM/MM-MD methods *Chemistry - A European Journal* ([P4]).
In this study, I performed the MM-MD simulations including endoperoxide and zwitterionic intermediates parameters generation, while QM and QM/MM-MD calculations were realized by Dr. Raymond Grüber in collaboration with Pr. Iñaki Tuñón from Valencia University, Spain.

2.3.3 Ibuprofen and ketoprofen photosensitization mechanism: how to enhance cells death under the influence of UVA light [P10]

This project was initiated during the one month Short Term Scientific Mission of COST CM1201 Biomimetic Radical Chemistry I have done in Gothenburg University. Pr. Leif Eriksson, our collaborator who hosted me, suggested to study interaction modes of a non-steroidal drug, the ketoprofen, with DNA. Indeed, this molecule is known to act as a type I photosensitizer on DNA, inducing damages and potentially mutations or cell death, although its first task is to cure. Thus, understanding of the molecular mechanisms of photosensitization by such drugs is essential to prevent the appearance of deleterious side effects. To this extent, I performed extensive MM-MD simulations on ketoprofen/DNA systems in order to unravel the preferential binding modes. We then decided to extend the project to ibuprofen, which is another very commonly used non-steroidal anti-inflammatory. This study was enriched by the implication of our collaborators from Nancy who initiated the addition of QM calculations about photodissociation and biological tests about cells viability in presence of the drugs under UV light. Hence, we led a broad investigation on ketoprofen and ibuprofen binding modes, reactivity, and aspects of biological relevance.
Among the various technics used in this investigation, I realized the classical Molecular Dynamics part in order to decipher the interactions of the drugs with DNA, including ketoprofen and ibuprofen parameters generation. In spite of a negative charge, they can transiently get close to the helix and the deleterious species formed upon photodissociation rapidly diffuse, being prone to attack B-DNA. Biological tests were performed by Pr. Stephanie Grandemange and Dr. Vanessa Besancenot and the QM calculations by Dr. Marco Marazzi and Dr. Antonio Monari, all from Nancy University. This study was recently submitted to *Scientific Reports*.

2.4 Publications

ChemPubSoc Europe

DOI: 10.1002/chem.201501212

CHEMISTRY A European Journal Full Paper

Photophysics

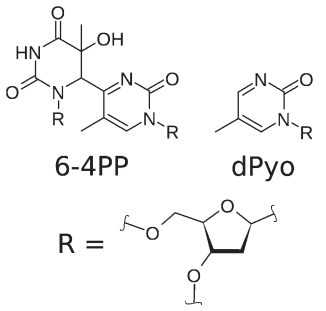
DNA Photosensitization by an "Insider": Photophysics and Triplet Energy Transfer of 5-Methyl-2-pyrimidone Deoxyribonucleoside

Emmanuelle Bignon,[a, b] Hugo Gattuso,[c, d] Christophe Morell,[b] Elise Dumont,*[a] and Antonio Monari*[c, d]

Abstract: The main chromophore of (6-4) photoproducts, namely, 5-methyl-2-pyrimidone (Pyo), is an artificial noncanonical nucleobase. This chromophore has recently been reported as a potential photosensitizer that induces triplet damage in thymine DNA. In this study, we investigate the spectroscopic properties of the Pyo unit embedded in DNA by means of explicit solvent molecular-dynamics simulations coupled to time-dependent DFT and quantum-mechanics/molecular-mechanics techniques. Triplet-state transfer from the Pyo to the thymine unit was monitored in B-DNA by probing the propensity of this photoactive pyrimidine analogue to induce a Dexter-type triplet photosensitization and subsequent DNA damage.

Introduction

DNA stores genetic code in living organisms and has the ability to deploy efficient electronic mechanisms to dissipate incident UV/Vis energy, hence opening channels for photostability.[1a] The most frequent DNA photodefects induced upon the direct absorption of UVB wavelengths are identified as pyrimidine dimers, such as pyrimidine(6-4)pyrimidone photoproducts ((6-4)PP; Scheme 1) or Dewar valence isomers.[1] Conversely, less energetic UVA wavelengths may induce DNA lesions only by an indirect mechanism known as photosensitization. In this case, the photon absorption excites a chromophore that interacts with DNA, which in turn triggers electron or energy transfer to the nucleobases or the generation of reactive oxygen species (ROS), such as singlet oxygen ($^1O_2$). Many different organic and inorganic molecular species can act as photosensitizers: simple aromatic ketones,[2a] porphyrins,[2b] or organometallic complexes.[2c] Most often, sensitization relies on noncovalent interactions between DNA and the drug and is nowadays extensively used for anticancer photodynamic therapy.[3]



Scheme 1. Molecular formula of the (6-4)PP and dPyo subunits.

Most photosensitizers are exogenous molecules, but UVA-endogenous photosensitizers are also known.[4] More recently, it has been demonstrated that even modified DNA nucleobases formed upon photodamage, such as (6-4)photoproducts[5a] or cytosine-containing cyclobutane dimers,[5b] can act as insider photosensitizers with a remarkable efficiency.

Most notably, the action of (6-4)PP as an internal photosensitizer is seen as a doorway, a Trojan horse to cite Miranda and co-workers,[5a] for the generation of additional pyrimidine dimers or oxidative DNA lesions in the vicinity of the original photodefect. The occurrence of several lesions within a few base pairs, known as "tandem" lesions, represents an extremely critical situation for DNA integrity, especially because of dramatically lowered repair rates.[6]

Photosensitizer (6-4)PP is formed by two cross-linked pyrimidine units derived by base dimerization (Scheme 1).[1] The highly delocalized 5-methyl-2-pyrimidone (5M2P; Pyo) moiety is responsible for the appearance of a new absorption band in the UVA region ($\lambda \approx 300$ nm) in damaged DNA.[7] Therefore, Pyo acts as a chromophore that extends the absorption region

[a] E. Bignon,[+] Dr. E. Dumont
Laboratoire de Chimie, UMR 5182
CNRS Ecole Normale Supérieure de Lyon, Lyon (France)
E-mail: elise.dumont@ens-lyon.fr

[b] E. Bignon,[+] Prof. C. Morell
Institut des Sciences Analytiques, UMR 5280
Université de Lyon1 (UCBL) CNRS, Lyon (France)

[c] H. Gattuso,[+] Dr. A. Monari
Université de Lorraine, Nancy Theory-Simulation-Modeling, SRSMC
Vandoeuvre-les-Nancy (France)
E-mail: antonio.monari@univ-lorraine.fr

[d] H. Gattuso,[+] Dr. A. Monari
CNRS, Nancy Theory-Simulation-Modeling
SRSMC, Vandoeuvre-les-Nancy (France)

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201501212.

of native DNA nucleobases. Furthermore, Miranda and co-workers[5a] elegantly demonstrated that the model system 1-(2-deoxy-β-D-erythropentofuranosyl)-5-methyl-2-pyrimidone (dPyo, which will be further noted P in one-letter codes) may efficiently induce triplet transfer to nearby thymine (T) nucleobases, thus opening the way for the so-called triplet photosensitization, which was postulated on the basis of laser flash photolysis studies of the Pyo triplet-state manifold (^{3}Pyo*). Notably, the ^{3}Pyo* energy was estimated to be compatible with the possibility of triplet transfer toward thymine.[5a,8] The role of Pyo in type-II photosensitization and in the production of ROS was also postulated and investigated.[5a] In particular, oxygen quenching of ^{3}Pyo* has been found to be responsible for the production of singlet oxygen ($^1O_2$) and hydroxy radicals, as confirmed by selective spin-trapping electron-paramagnetic resonance (EPR) analysis.[5a]

The knowledge of the triplet-transfer mechanism ^{3}Pyo$\rightarrow$^{3}T is decisive to evaluate the inherent propensity of photolesions such as (6-4)PP to induce further vicinal defects.

Far from being a simple model system, Pyo also behaves as an artificial nucleobase similar to the 6-thioguanine, which has been studied and patented.[9] In particular, following its inclusion in oligonucleotides, and also because of its proximity to π-stacked thymine units, dPyo may be prone to inducing a very efficient energy transfer, most notably following a Dexter-type mechanism. Indeed, the use of artificial nucleotides is a promising therapeutic strategy in cancer treatment.[10] Hence, stable dPyo-containing B-DNA structures combined with the application of UVA light may induce DNA lesions that can result in cell apoptosis, thus showing a potentially high therapeutic interest.

Modeling the structural and dynamic properties of dPyo embedded in B-DNA, which constitutes a strongly heterogeneous and flexible environment, is also of crucial importance to definitively unravel its behavior as an artificial nucleobase. Particular attention should be devoted to the deviation from the ideal B-DNA structures. By relying on state-of-the art molecular-modeling techniques, we report herein atomistic and electronic insight into the photochemistry of the Pyo chromophore. The role of this unit toward an efficient and facile triplet transfer is firmly evidenced, and its propensity as an insider photosensitizer can be compared on the same footing with the most common exogenous sensitizers. This investigation ultimately allows us obtain clear and unprecedented insight into Pyo-induced DNA photosensitization. To this end, we hereby model the structure of a B-DNA decamer in which one of the central bases (the sixth one) has been substituted by dPyo by using a combination of molecular dynamics (MD) and hybrid quantum-mechanics/molecular-mechanics (QM/MM) techniques[11]. The intrinsic Pyo absorption, fluorescence spectra, and phosphorescence maxima are analyzed.

Finally, the triplet-energy transfer to a vicinal thymine unit is fully modeled by following an approach recently proposed by some of us.[12] This challenge can be tackled by using our protocols and methods that have proven successful in providing a reliable atomistic and electronic description of similar DNA sensitizers.[12,13]

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 1) is partially destroyed. The mobility of

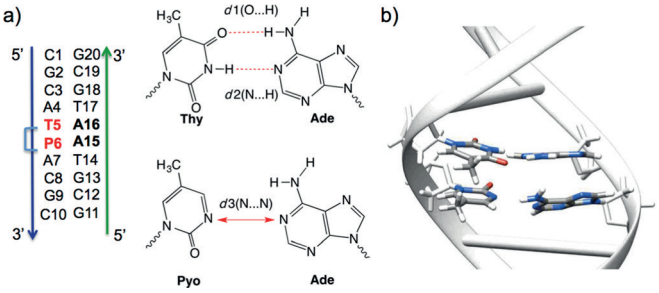


Figure 1. a) Sequence of the 10-mer in which dPyo is embedded as the sixth nucleobase. The Watson–Crick hydrogen-bond pairing between thymine (Thy) and adenine (Ade) no longer comes into play upon substitution by dPyo. Energy transfer toward the thymine T5 unit is investigated: b) representative structure 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.

these bases was quantified by the time evolution of the distance d_1, which is between the closest nitrogen atoms of the P6–A15 base pair (see the Supporting Information), which clearly shows the typical behavior of a dual regime characterized by the oscillation between two conformations. In the first conformation, the value of d_1 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 d_1 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.

The partial disruption of the Watson–Crick network remains local and does not affect the T5–A16 pairing, with an average distance d_2 between the hydrogen and nitrogen atoms in T5 and A16, respectively, of 2.09 Å. The distance d_3 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 2; as defined in ref. [14]) 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 small-

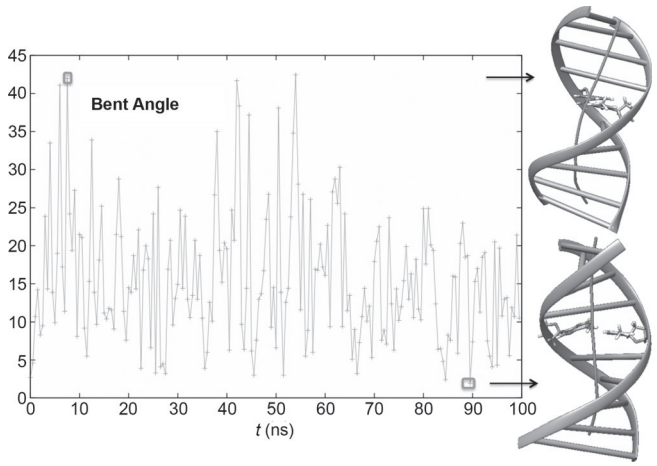


Figure 2. Time evolution of the global bending of the DNA strand. Two representations of the DNA double strand and the helical axis are given for the limiting conformations. The dPyo-containing base pair is displayed as a licorice representation.

er bending. Beyond this overall stability of the Pyo-containing sequence, a closer inspection of the intra- and interbase parameters[14] reveals more subtle influences of the T→Pyo mutation (Figure 3).

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 for the rotational intrabase parameters, with the particularly stunning case of the "propeller" that rises up to 100° deviation.

Although the general picture of the interbase parameters is similar to the intrabase parameters, we may notice that the deviations of the "rise" and "twist2 parameters are less pronounced for the base pair including Pyo. This effect could be related to the necessity of maintaining an optimal π-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 d_1 and hence can be regarded as an indicator of the transition between the two conformations (as reported in the Supporting

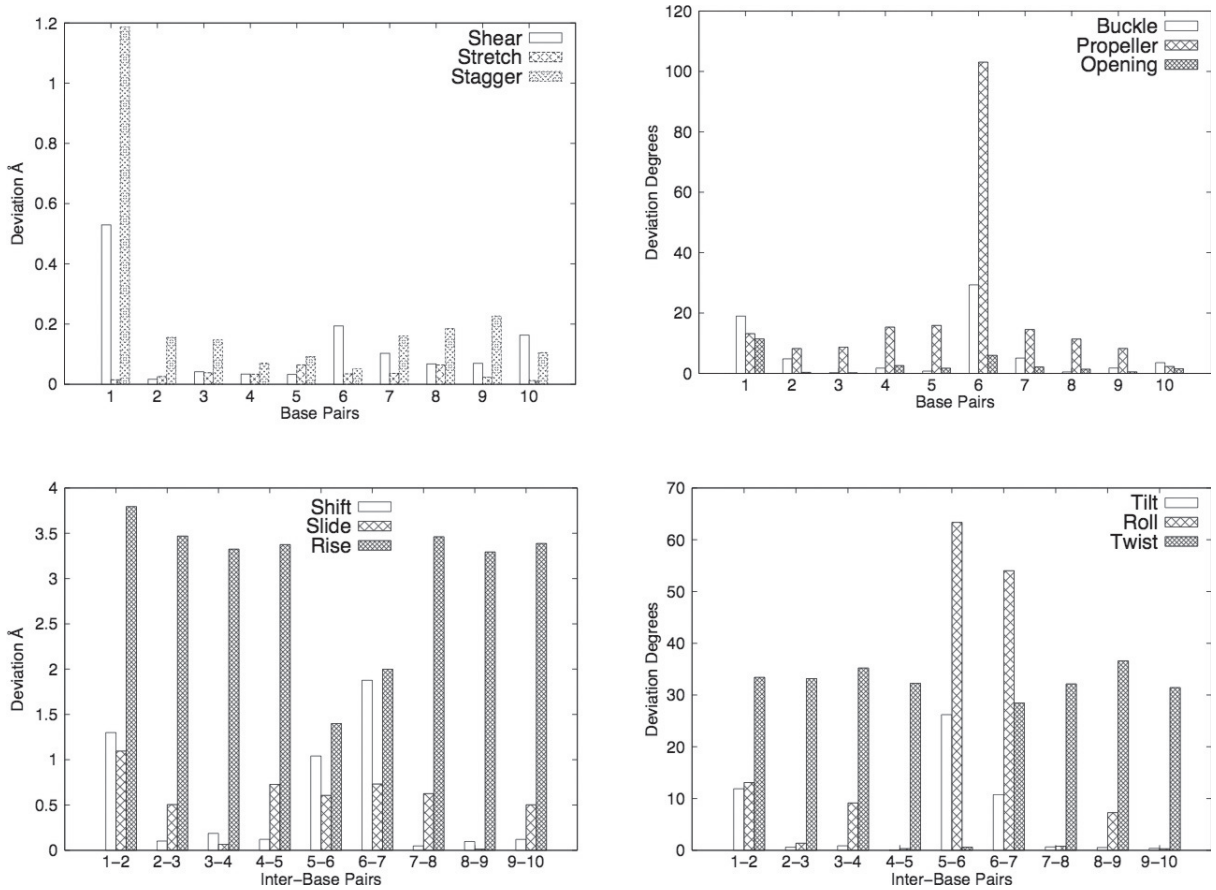


Figure 3. Average deviations from the ideal B-DNA structure for the intra- (top) and interstrand (bottom) parameters.

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[15] 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.[5a,16] The calculated TD-DFT absorption spectrum of Pyo embedded in a B-DNA environment is reported in Figure 4 (left panel). It appears evident that a rather intense band at about $\lambda =$ 310 nm dominates the UVA spectrum of the artificial nucleobase. The fluorescence spectrum shows a tail and shoulder at around $\lambda = 460$ 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 constrains.

On one hand, this outcome confirms the experimental results reasonably well[5a,16] for the mononucleoside and the DNA embedding. On the other hand, these findings 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 $\lambda = 290$ nm due to vibrational effects. Unsurprisingly, the first excited state is described as a π–π^* transition that extends over the entire conjugated ring. The circular dichroism (CD) spectrum (see the Supporting Information) matches the experimental results quite well[16] for the position and the shape of the bands. 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 4, right panel), and we found a maximum absorption at about $\lambda = 350$–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.[17] Note that the fluorescence also presents a rather asymmetric curve, with now an even more pronounced shoulder, which becomes almost a second maximum absorption, at about $\lambda = 420$ 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 $\lambda = 530$ nm, which once again compares satisfactorily with the experimental results.[16] 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 S1–T1 coupling in the Franck–Condon region has been estimated to be $\lambda = 28\ cm^{-1}$. Such a value coupled with the small singlet–triplet energy gap is sufficient to induce the population of T1 in the ultrafast regime. To sum up, the spectroscopic features of Pyo, consistent with the results of Miranda and co-workers for the mononucleoside,[5a] are all concordant in establishing the picture of a possible Pyo-driven DNA triplet photosensitization.

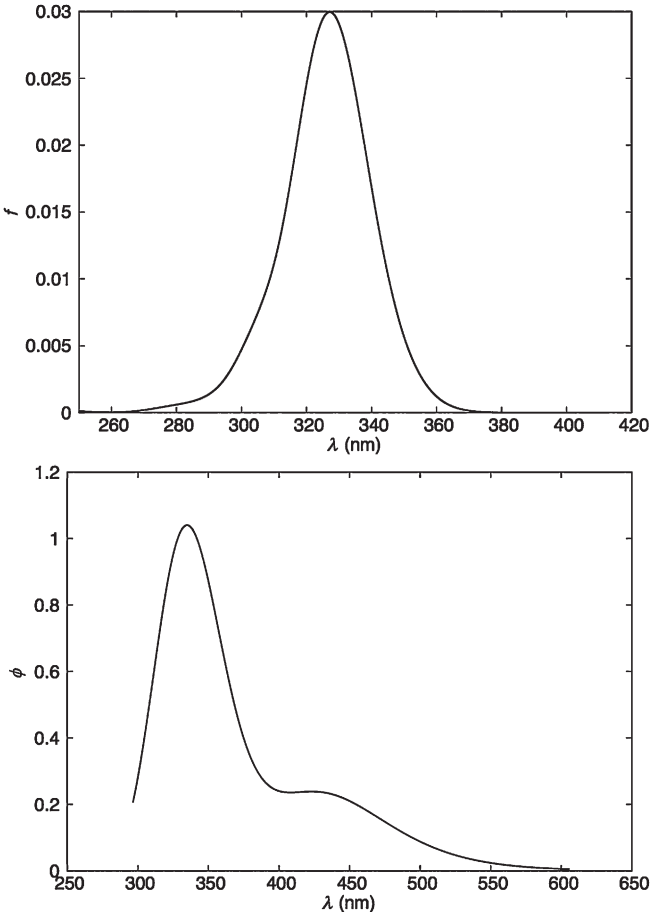


Figure 4. Calculated UVA absorption (top) and fluorescence (bottom) spectra of Pyo. The wavelengths are given in nm and the intensities in arbitrary units obtained from the oscillator strength.

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[11a] (see the Experimental Section 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.[11a] We delineate two representative cases from which some hints can be inferred (Figure 5). 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.[18] Nevertheless, the barrier is strongly dependent on the initial conditions, and in some cases it can reach values close to 1.0 eV.

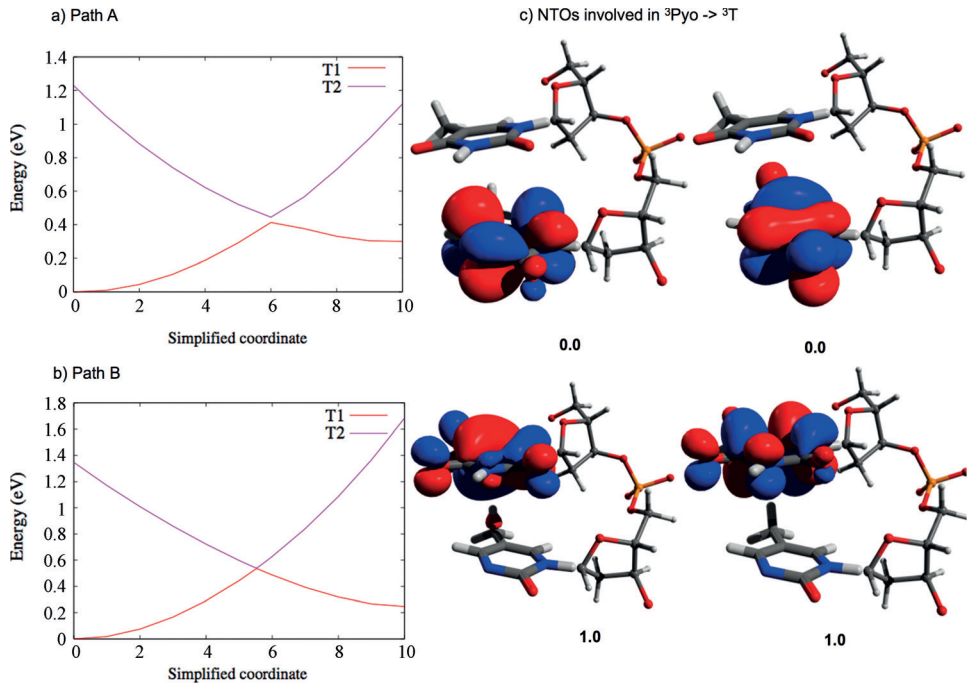


Figure 5. a, b) Energy profile along the $Q(\xi)$ approximate coordinate for two limiting paths (different initial conditions). c) Natural-transition orbitals (NTOs) for $\xi = 0.0$ (up) and $\xi = 1.0$ (bottom) show the transfer of the triplet state between the two interacting monomers.

Second, the relative stability of ^{3}Pyo and ^{3}T deserves a comment. From the simple point of view of potential energy, ^{3}Pyo appears to be about 0.2–0.3 eV more stable than ^{3}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}Pyo is long enough to allow the experimentally observed production of $^1O_2$.[5a] The energetic profile in B-DNA should also be compared to the one mimicking the nucleoside in aqueous solution, for which experimental results exist.[5a] To this end, we built a model system comprised of π-stacked 5M2P and thymine and we studied the triplet transfer in water and the gas phase, respectively (Figure 6). 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 crossing in some cases between the T1 and T2 potential-energy curves in a situation somehow reminiscent of a conical intersection (Figure 5b). 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;[5b,11a,19] namely, the higher the overlap, the better the transfer efficiency. In the case of the ^{3}Pyo→^{3}T energy transfer, we face a huge and extended mixing of the electronic densities of the two monomers, as clearly evidenced by the NTOs.[20] The electronic-density mixing is particularly evident when the T1 and T2 potential-energy curves actually cross (Figure 7).

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

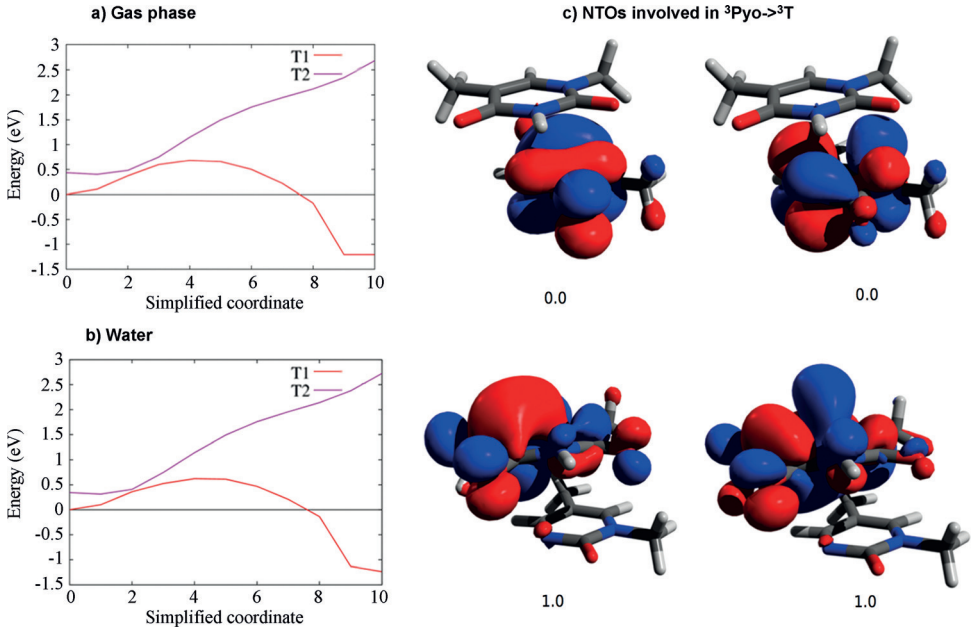


Figure 6. a,b) Energy profile along the $Q(\xi)$ approximate coordinate for π-stacked Pyo and thymine moieties in the gas phase and aqueous solution. c) NTOs for $\xi = 0.0$ (up) and $\xi = 1.0$ (bottom) shows the transfer of the triplet state between the two interacting monomers.

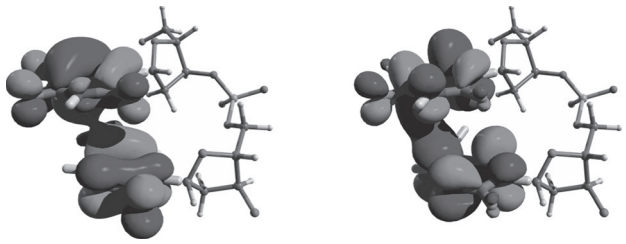


Figure 7. NTOs at the T1/T2 crossing point (Figure 5b; $\xi = 0.5$) show a total delocalization of the wavefunction between the two monomers, hence confirming the high efficiency of the energy transfer.

basis of the NTO overlap by following a protocol detailed elsewhere.[11a] As a comparison, the efficient energy transfer of benzophenone was driven by a wavefunction overlap of only about 0.1 eV.[11a] 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 ultraefficient; 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 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 (6-4)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 happen even in DNA without overcoming a too significantly high barrier. 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 π-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}Pyo could also extract hydrogen atoms from 2-deoxyribose units because previous studies have shown that excited pyrimidone derivatives have access to this photochemistry.[34] However, the exploration of this pathway needs different theoretical protocols and will be properly addressed in a forthcoming contribution. 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 carginogetic themselves and metabolitic pathways to incorporate Pyo in DNA are known. In the future, we plan to extend this study with a proper simulation of the triplet-energy transfer in the case of (6-4)PP lesions, which should see much larger deviations from the ideal helical structure. Also, the photophysical pathway that leads to Pyo triplet population and subsequent energy transfer in the B-DNA environment will be simulated with state-hopping molecular dynamics.

Experimental Section

Methods and computational details

The Amber14 molecular-dynamics software package was used for all the calculations.[21] The force-field parameters were taken from the GAFF[22] and ff99bsc0[23] sets. The atom-point charges for the Pyo nucleobase were generated (see the Supporting Information) by using the restrained electrostatic-potential[24] 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(CGCAT PACGC):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 K^+ ions. The modified DNA molecule with the counterions was placed in a box ($60.9 \times 60.4 \times 63.4$ Å^3) containing 5662 TIP3P water molecules.[25] The Particle-Mesh-Ewald method with a cutoff of 9 Å 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 simulated electronic circular dichroism (ECD) spectra (see the Supporting Information), and 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[12,26] between the Tinker[27] and Gaussian09[28] programs. The quantum partition consisted of the chromophoric unit 5M2P constituent of Pyo, a linking atom was placed to cap the covalent bond between 5MP2 and 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.[13] 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 M06-2X[30] functional. Polarizable embedding was modeled by taking into account the electronic response of the surrounding (ERS).[12] The geometry of the first singlet and triplet excited states, respectively, was optimized to model the fluorescence and phosphorescence properties. In the case of the ECD,[31] 20 snapshots were extracted from the dynamic runs and we chose a QM region constituted by 5M2P and the five surrounding canonical nucleobases. Linking atoms were also used to cap the dangling bonds. 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 π-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 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}Pyo–^{1}Thy) or on the thymine unit Q(^{1}Pyo–^{3}Thy). A simplified coordinate[11a] $Q(\xi)$ was used, with $\xi = 0$–1, according to the following formula:

$$Q(\xi) = \xi Q(\mathrm{3Pyo-1Thy}) + (1-\xi)Q(\mathrm{3Pyo-1Thy})$$

To compare the triplet transfer in the case of the isolated nucleotide in aqueous solution, we built a model system composed of π-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.[20]

Spin–orbit coupling was calculated at the TD-DFT level on the 5M2 ground-state equilibrium geometry by using the Dalton 2015 code[32,33] and the mean field approximation.

Acknowledgements

E.B. and H.G. are grateful for the funding of a Ph.D. fellowship by Université de Lyon and Université de Lorraine, respectively. Calculations were performed at the Lyon PSMN cluster and Nancy local computing service. This work was performed within the framework of the COST in Chemistry Action CM1201 'Biomimetic Radical Chemistry'. A.M. thanks CNRS for the funding of the "Chaire d'excellence" program.

Keywords: DNA · energy transfer · molecular dynamics · nucleobases · photophysics

[1] a) M. Barbatti, *J. Am. Chem. Soc.* **2014**, *136*, 10246 – 10249; b) P. H. Clingen, C. F. Arlett, L. Roza, T. Mori, O. Nikaido, M. H. L. Green, *Cancer Res.* **1995**, *55*, 2245 – 2248; c) M. Barbatti, A. J. A. Aquino, J. J. Szymczak, D. Natchigallova, P. Hobza, H. Lischka, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21453 – 21458.

[2] a) M. C. Cuquerella, V. Lhiaubet-Vallet, J. Cadet, M. A. Miranda, *Acc. Chem. Res.* **2012**, *45*, 1558 – 1570; b) R. J. Fiel, N. Datta-Gupta, E. H. Mark, J. C. Howard, *Cancer Res.* **1981**, *41*, 3543 – 3545; c) E. Gicquel, J. P. Souchard, F. Magnusson, J. Chemaly, P. Calsou, P. Vicendo, *Photochem. Photobiol. Sci.* **2013**, *12*, 1517 – 1526.

[3] a) R. Mensudar, *World. J. Med. Sc.* **2014**, *10*, 139 – 142; b) E. J. Dennis, G. J. Dolmans, D. Fukumura, R. K. Jain, *Nat. Rev. Cancer* **2003**, *3*, 380 – 387; c) P. Agostinis, K. Berg, K. A. Cengel, T. H. Foster, A. W. Girotti, S. O. Gollnick, S. M. Hahn, M. R. Hamblin, A. Juzeniene, D. Kessel, M. Korbelik, J. Moan, P. Mroz, D. Nowis, J. Piette, B. C. Wilson, J. Golab, *Ca-Cancer J. Clin.* **2011**, *61*, 250 – 281.

[4] a) B. Epe, *Photochem. Photobiol. Sci.* **2012**, *11*, 98 – 106; b) J. Cadet, R. Wagner, *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a012559; c) J. Cadet, S. Mouret, J.-L. Ravanat, T. Douki, *Photochem. Photobiol.* **2012**, *88*, 1048 – 1065; d) G. T. Wondrak, M. K. Jacobson, E. L. Jacobson, *Photochem. Photobiol. Sci.* **2006**, *5*, 215 – 237.

[5] a) V. Vendrell-Criado, G. M. Rodriguez-Muñiz, M. Consuelo Cuquerella, V. Lhiaubet-Vallet, M. A. Miranda, *Angew. Chem. Int. Ed.* **2013**, *52*, 6476 – 6479; *Angew. Chem.* **2013**, *125*, 6604 – 6607; b) T. Douki, I. Bérard, A. Wack, S. Andrä, *Chem. Eur. J.* **2014**, *20*, 5787 – 5794.

[6] a) H. Budworth, G. Matthewman, P. O'Neill, G. L. Dianov, *J. Mol. Biol.* **2005**, *351*, 1020 – 1029; b) D. T. Goodhead, *Int. J. Radiat. Biol.* **1994**, *65*, 7 – 17.

[7] K. Haiser, B. P. Fingeruth, K. Heil, Glas, T. T. Herzog, B. M. Piles, J. Schreider, W. Zinth, R. deVivie-Riedle, C. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 408 – 411; *Angew. Chem.* **2012**, *124*, 421 – 424.

[8] F. Bosca, V. Lhiaubet-Vallet, M. C. Cuquerella, J. V. Castell, M. A. Miranda, *J. Am. Chem. Soc.* **2006**, *128*, 6318 – 6319.

[9] H. P. Rappaport, *Nucleic Acids Res.* **1988**, *16*, 7253 – 7267.

[10] H. Wang, E. Rayburn, Y. Zhang, *Curr. Pharm. Des.* **2005**, *11*, 2889 – 2907.

[11] a) E. Dumont, M. Wibowo, D. Roca-Sanjuan, M. Garavelli, X. Assfeld, A. Monari, *J. Phys. Chem. Lett.* **2015**, *6*, 576 – 580; b) E. Dumont, A. Monari, *J. Phys. Chem. Lett.* **2013**, *4*, 4119 – 4124.

[12] A. Monari, J.-L. Rivail, X. Assfeld, *Acc. Chem. Res.* **2013**, *46*, 596 – 603.

[13] a) E. Dumont, A. Monari, *J. Phys. Chem. B* **2015**, *119*, 410 – 419; b) T. Etienne, T. Very, A. Perpète, A. Monari, X. Assfeld, *J. Phys. Chem. B* **2015**, *119*, 4973 – 4980; c) M. Huix-Rotllant, E. Dumont, N. Ferré, A. Monari, *Photochem. Photobiol.* **2015**, *91*, 323 – 330; d) T. Véry, D. Ambrosek, M. Otsuha, C. Gourlaouen, X. Assfeld, A. Monari, C. Daniel, *Chem. Eur. J.* **2014**, *20*, 12901 – 12909; e) T. Very, S. Despax, P. Hébraud, A. Monari, X. Assfeld, *Phys. Chem. Chem. Phys.* **2012**, *14*, 12496 – 12504.

[14] R. Lavéry, M. Moakher, J. H. Maddocks, D. Petkeviciute, K. Zakrzewska, *Nucleic Acids Res.* **2009**, *37*, 5917 – 5929.

[15] K. A. Kistler, S. Matsika, *Photochem. Photobiol.* **2007**, *83*, 611 – 624.

[16] a) B. A. Connolly, P. C. Newman, *Prog. Nucleic Acid Res. Mol. Biol.* **1989**, *17*, 4957 – 4974; b) B. Gildea, L. W. McLaughlin, *Nucleic Acids Res.* **1989**, *17*, 2261 – 2281.

[17] T. Etienne, H. Gattuso, A. Monari, X. Assfeld, *Comput. Theor. Chem.* **2014**, *1040 – 1041*, 367 – 372.

[18] J. J. Nogueira, M. Oppel, L. Gonzalez *Angew. Chem. Int. Ed.* **2015**, *54*, 4375 – 4378; *Angew. Chem.* **2015**, *127*, 4450 – 4453.

[19] P. G. Wu, L. Brand, *Anal. Biochem.* **1994**, *218*, 1 – 13.

[20] a) J.-D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615 – 6620; b) T. Etienne, X. Assfeld, A. Monari, *J. Chem. Theory Comput.* **2014**,

10, 3895 – 3905; c) T. Etienne, X. Assfeld, A. Monari, *J. Chem. Theory Comput.* **2014**, *10*, 3906 – 3914.

[21] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman (2014), AMBER 14, University of California, San Francisco.

[22] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157 – 1174.

[23] A. Pérez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatman, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817 – 3829.

[24] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollmann, *J. Phys. Chem.* **1993**, *97*, 10269 – 10280.

[25] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926 – 935.

[26] N. Ferré, X. Assfeld, J.-L. Rivail, *J. Chem. Phys.* **2002**, *117*, 4119 – 4125.

[27] Tinker Molecular Modeling Software. http://dasher.wustl.edu/ffe/Tinker.

[28] Gaussian 09, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, **2009**.

[29] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213 – 222.

[30] Y. Zhao, D. G. Thruhlar, *Theor. Chem. Acc.* **2008**, *120*, 215 – 241.

[31] H. Gattuso, X. Assfeld, A. Monari, *Theor. Chem. Acc.* **2015**, *134*, 36.

[32] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansik, H. J. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. Rybkin, P. Salek, C. C. M. Samson, A. Sánchez de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, H. Ågren, *WIREs Comput. Mol. Sci.* **2014**, *4*, 269 – 284.

[33] Dalton, a molecular electronic structure program, Release Dalton2015.0 (2015), see http://daltonprogram.org.

[34] a) G. Ryseck, T. Schmieier, K. Haiser, W. Schreier, W. Zinth, P. Gilch, *ChemPhysChem* **2011**, *12*, 1880 – 1888; b) G. Ryseck, T. Villnow, S. Hugenbruch, K. Shaper, P. Gilch, *Photochem. Photobiol. Sci.* **2013**, *12*, 1423 – 1430.

Received: March 27, 2015
Published online on July 1, 2015

56–62 *Nucleic Acids Research, 2016, Vol. 44, No. 1* *Published online 9 December 2015*
doi: 10.1093/nar/gkv1364

Probing the reactivity of singlet oxygen with purines

Elise Dumont[1,*], Raymond Grüber[1], Emmanuelle Bignon[1,2], Christophe Morell[2], Yohann Moreau[3], Antonio Monari[4,5] and Jean-Luc Ravanat[6,7,*]

[1]Laboratoire de Chimie, UMR 5182, Ecole Normale Supérieure de Lyon, Lyon France, [2]Institut des Sciences Analytiques, Université de Lyon 1 and CNRS, Villeurbanne, France, [3]iRTSV/CBM/MCT, CEA Grenoble, France, [4]Theory-Modeling-Simulation, SRSMC, Université de Lorraine Nancy, Vandoeuvre-lès-Nancy, France, [5]CNRS, Theory-Modeling-Simulation, SRSMC, Vandoeuvre-lès-Nancy, France, [6]INAC-SCIB, Université Grenoble Alpes, F-38000 Grenoble, France and [7]CEA, INAC-SCIB-LAN, F-38000 Grenoble, France

Received August 28, 2015; Revised November 20, 2015; Accepted November 24, 2015

ABSTRACT

The reaction of singlet molecular oxygen with purine DNA bases is investigated by computational means. We support the formation of a transient endoperoxide for guanine and by classical molecular dynamics simulations we demonstrate that the formation of this adduct does not affect the B-helicity. We thus identify the guanine endoperoxide as a key intermediate, confirming a low-temperature nuclear magnetic resonance proof of its existence, and we delineate its degradation pathway, tracing back the preferential formation of 8-oxoguanine versus spiro-derivates in B-DNA. Finally, the latter oxidized 8-oxodGuo product exhibits an almost barrierless reaction profile, and hence is found, coherently with experience, to be much more reactive than guanine itself. On the contrary, in agreement with experimental observations, singlet-oxygen reactivity onto adenine is kinetically blocked by a higher energy transition state.

INTRODUCTION

Singlet molecular oxygen ($^1O_2$) lies among the most deleterious reactive oxygen species (1). Indeed, $^1O_2$ is a common source of damage for many biomolecules: lipid (2), protein (3) and DNA (4,5) which can lead to genomic mutations. The reaction mechanism has been established on prototypical systems (ethylene (6), butadiene and benzene (7,8), thiophene (9)), which mimic lipids. Yet less is known concerning the mechanistic processes of singlet-oxygen-driven DNA damage. Experimentally, it has been shown that reaction of $^1O_2$ with DNA is limited to the guanine nucleobase, with induction of three main products (10), depicted in Figure 1, that have been identified and experimentally characterized over the years (nuclear magnetic resonance (NMR), mass spectrometry…) (11). At the nucleoside level, spironucleosides **6** were found to be mostly produced, together with 4-hydroxy-8-oxo-2′-deoxyguanosine (4-OH-8-oxodGuo **4**) and 8-oxo-7,8-dihydro-2′-deoxyguanosine (12) (8-oxodGuo **8**). In double-stranded DNA and at cellular level (13), 8-oxodGuo was found to be the major product. It has been shown that 8-oxodGuo is more reactive toward singlet oxygen than undamaged nucleoside dGuo **1** (by about 2 order of magnitude) and that spiro-nucleosides could be also generated through secondary oxidation of 8-oxodGuo (14). Thus, rationalizing the mechanism of formation or dGuo decomposition products is a very challenging task. It is supposed that all guanine decomposition products stem from a primary reaction path ($^1O_2$ addition), depicted in Figure 1. Three intermediates can be written down, which can eventually coexist: their relative ratio can be also strongly environment dependent, in the same way the final products are. In order to rationalize the formation of subproducts for which accurate measurements are available, the central question is probably to assess the feasibility of endoperoxide thermal formation. The initiation of $^1O_2$-induced degradation pathway (Figure 1) is the most critical since the so-formed intermediates have a relative short lifetime. Indeed, guanine endoperoxide is commonly accepted as the key intermediate by almost all the different proposed mechanistic pathways even if its isolation and characterization is extremely difficult due to its short lifetime. Interestingly, the endoperoxide has been characterized as the main reaction product of singlet oxygen with aromatic organic compounds, such as anthracene and naphthalene derivatives (15).

Furthermore, and more generally to clarify the reactivity of singlet oxygen toward nucleosides two fundamental questions remain answerless: $^1O_2$ strong selectivity toward guanine versus adenine should be rationalized; the environmentally controlled fragmentation leading to different products between B-DNA and mononucleoside solutions needs to be tackled. To address these questions, in this work, we report computational evidences that unambiguously establish the stability of the guanine endoperoxide **3** *versus* the open zwitterionic intermediate **2-a**, in contrast with previ-

*To whom correspondence should be addressed. Tel: +33 4 72 72 88 46; Email: elise.dumont@ens-lyon.fr
Correspondence may also be addressed to Jean-Luc Ravanat. Tel: +33 38 78 47 97; Email: jean-luc.ravanat@cea.fr

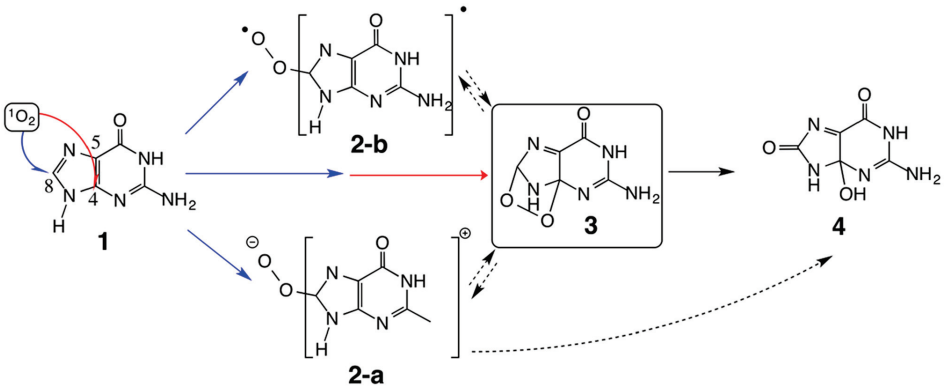


Figure 1. Initial attack of $^1O_2$ onto guanine, up to the endoperoxide intermediate **3**, evolving to the three main experimental products.

ous theoretical approaches (16). We situate an accessible, low-barrier pathway for its formation, which corroborates the existence of a transient guanine endoperoxide. The latter was proposed based on low-temperature NMR experiments performed on a guanosine derivative (17). Within B-DNA, the formation of a cyclic peroxide could eventually trigger an helical distortion energetically penalizing the intermediate **3**.

This hypothesis was tested performing explicit molecular dynamics (MD) simulations of the two intermediates **2-a** and **3** embedded in a 13-bp DNA fragment, and hence excluding any marked structural modifications of the DNA arising from the first attack. Finally, the possible degradation pathways leading respectively to spiro **6** or 8-oxodGuo **8** whether the reaction takes place at the nucleoside level or in B-DNA are explored to assess this environmental effect.

MATERIAL AND METHODS

Quantum mechanics

Quantum mechanics (QM) calculations were conducted starting from a structure of a methyl-capped guanine and a methyl-capped adenine, as a model system, respectively, for 2′-deoxyguanosine and 2′-deoxyadenine. Density functional theory was employed, according to a previous benchmark against the SCS-MP3/aug-cc-pVTZ//MP2/DZP++ (18) level of theory (19) that has stressed out the performance of the range-separated hybrid LC-BLYP. Its performance has been also very recently pinpointed for the singlet-oxygen attack onto trans-resveratrol (20). The Gaussian09 software Revision D.01 (21) was used for all the QM calculations. Since **2** is an intramolecular charge-separated entity and to better model experimental conditions, an implicit model accounting for solvation in water, polarizable continuum model ((22)), was used with a value of the dielectric constant $\epsilon_r = 78.3$. The inclusion of diffuse functions is also important to compare the electronic energies of **2** and **3** on the same footing. We selected the DZP++ basis set, proposed by Schaefer and including diffuse and polarization functions, as a valuable compromise between Pople basis set and the computationally demanding Dunning basis sets. All the stationary points have been characterized by calculating harmonic vibrational forces. Furthermore in the case of transition states intrinsic reaction coordinate profiles have been obtained to assure the transition state (TS) was indeed connecting reactants and products.

Molecular dynamics simulation

All classical MD simulations were performed with the Amber12 Molecular Dynamics software package (23). GAFF (24) and *ff99bsc0* (25) force fields parameters were used. For the oxidized nucleobase, charges and modified force field parameters have been used. Charges were generated using the Restrained Electrostatic Potential charge model of the Antechamber program (26) from QM calculations for the isolated fragment as prescribed by the usual protocol (see Supplementary Table S1): bonded force field parameters along the C-O-O bonds were assigned against high-level QM calculations at the LC-BLYP/DZP++ level of theory. For the open zwitterionic intermediate, the oxygen—oxygen distance is 1.504 Å with a harmonic constant force of 306.30 kcal.mol^{-1}.Å^{-2} (1.42 Å with a force constant of 343.60 kcal.mol^{-1}.Å^{-2} for the endoperoxide). Similarly, the CT…OS linkage for the zwitterionic moiety was taken as 1.439 Å (kr = 301.50 kcal.mol^{-1}.Å^{-2}). For the open structure, the covalent angle C8-O-O was described with an equilibrium value of 103.0 degrees, and an harmonic constant of 65.62 kcal.mol^{-1}.degree^{-2}. For the endoperoxide, the C-O-O angle was assigned to 105.01 degrees with a force constant of 65.89 kcal.mol^{-1}.degree^{-2} (same as c3-os-os in GAFF). The Amber topology input files for the DNA molecule were generated with the LEaP basic preparation program. Endoperoxide and zwitterionic guanine oxidized nucleobases (X) were incorporated into the standard B-DNA conformation of a 13-bp self-complementary alternate poly(dG)-poly(dC) sequence, namely d(GCGCGC**X**CGCGCG):d(CGCGCGCGCGCGC).
For both the endoperoxide and zwitterion intermediates, two different conformations are possible consisting in an attack of ^{1}O2 on guanine either directed toward the 5′ extremity (up) or the 3′ extremity (down). Consequently, a total of four starting structures, endoperoxide up/down and zwitterion up/down, have been built. Negative charges on DNA phosphate groups of each system were neutralized by potassium cations. Modified oligonucleotides with

their counterions were solvated in a 65.4 * 68.4 * 77.3 $Å^3$ parallelepiped box containing ∼8750 TIP3P water molecules (27) (total ∼27 100 atoms). The Particle Mesh Ewald method with a 9.0 Å cut-off was used for calculation of electrostatic interactions. Throughout the simulation, periodic boundary conditions were employed to eliminate undesirable edge effects. Each system was preliminary minimized in 5000 steps—-including 1500 conjugate gradient steps. Heating from the initial temperature of 0–300 K was performed in a thermalization run of 30 ps (NVT ensemble). The temperature was kept constant during MD simulations using the Langevin thermostat (28) with a collision frequency (γ ln) value of 2 ps^{-1}. To ensure an equilibrium sampling, thermalization was followed by an NPT equilibration of 300 ps with pressure of 1 atm., finally a production run of 100 ns was performed for each system. DNA structural parameters were analysed on-the-fly by using the Curves+ code (29); in particular the bending axis and the deviation of inter- and intrastrand parameters from ideal B-DNA's value were considered as an overall measure of the double helix deformation. For a full description of the different base parameters, the reader may refer to the original article by Lavery *et al.* (29). No significant difference concerning DNA distortion or stability is found between up and down conformers. In order to assess the structural stability in more biological relevant conditions MD trajectory for the different conformations were repeated at a temperature of 310 K following the same protocols both for the system set-up and for the analysis of the results.

RESULTS

We first consider the open intermediate **2**, which can a priori exist either as an intramolecular zwitterion **2-a** or as a biradical **2-b**. The energy gap between the two species is in favor of the closed-shell adduct **2-a** in implicit aqueous solvent by 9.2 kcal mol^{-1}: this will be reinforced within B-DNA since the peroxylate –O–O^- moiety is highly solvent exposed (see Supplementary Figure S4). Only the reactivity of the zwitterionic form **2-a** is considered hereafter. Its formation upon nucleophilic attack of $^1O_2$ onto guanine is revealed as an efficient, exothermic process (-10.2 kcal mol^{-1}), with an energy barrier for the first transition state structure $\Delta E^{\dagger(1\rightarrow 2a)}$ estimated to 20.1 kcal mol^{-1} (Figure 2, red solid line) for a distance C8…O of 1.82 Å. Our calculations confirm that the nucleophilic attack occurs on the C8 position, which is the most electrophilic position for guanine (30), in comparison to C4 and C5 (see Supplementary Figure S2). The open intermediate **2-a** coexists as two stable conformers, denoted ≪outer≫ and ≪inner≫ in Figure 3, and differing by the dihedral τ(H8C8OO). A transition state TS2 is located between the two conformers at 8.8 kcal mol^{-1} ($\tau = 111.0°$), although this rotation can be expected to proceed more easily and even be barrierless upon inclusion of dynamic and vibrational effects. The inner conformer of the zwitterion is very prone to evolve toward the endoperoxide **3** via an intramolecular cyclization onto the C4 position of guanine, characterized by a low barrier (2.8 kcal mol^{-1}, Figure 2): the gain in energy accounts for -9.4 kcal mol^{-1}, which corresponds to a charge transfer of 0.4 e between the peroxylate moiety and the guanine nucleobase fragments along the closure.

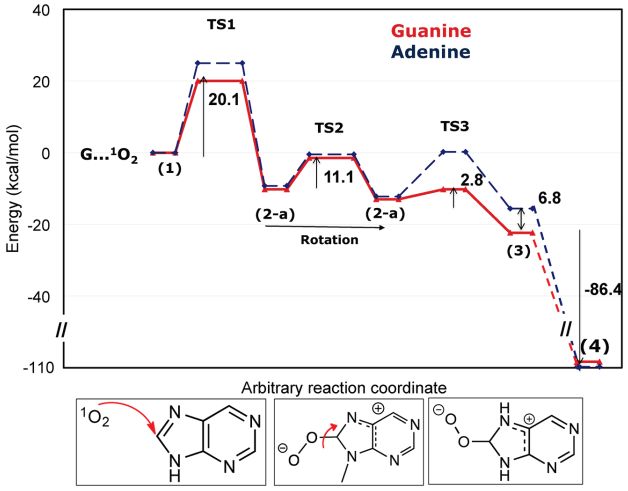


Figure 2. Reaction profile for the purine+$^1O_2$ system (free nucleosides).

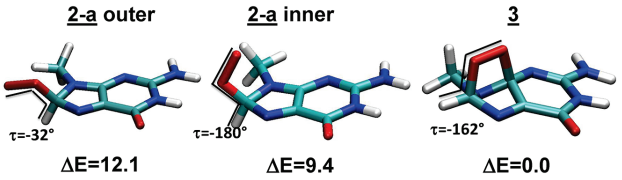


Figure 3. Representation of the guanine+$^1O_2$ intermediates.

To sum up, going from **1** to **3**, the G+$^1O_2$ damaging process is driven by a considerable thermodynamic force, the endoperoxide **3** being -22.4 kcal mol^{-1} more stable than the {G+$^1O_2$} adduct in its equilibrium geometry (at a distance $^1O_2$…C8 of 2.59 Å). The addition onto C8 clearly stands out as the limiting step. The existence of a second possible channel, the synchronous process with the [4+2] cycloaddition of $^1O_2$ across the 4,8-bond of the imidazole ring, leading to **3** directly from **1** (Figure 1) cannot be assessed due to its multireference character (7). But in case of a low barrier, this channel would reinforce our statement surmising the guanine endoperoxide as a key intermediate in guanine oxidation pathway as depicted in Figure 2.

The low ionization potential of guanine comes to stabilize the intramolecular adduct **2-a**, as revealed by comparison with the adenine profile: however the main difference for the profiles of the two bases is a higher transition state for the initial attack onto adenine ($\Delta E^{\dagger(1\rightarrow 2)} = 25.0$ kcal mol^{-1}, Figure 2 blue dashed line). The propensity to stabilize a partial positive charge on the guanine ring constitutes an additional driving force towards its oxidation. The selectivity of reaction of $^1O_2$ toward DNA constituents can therefore be explained on the basis of a kinetic blockage hampering the formation of the endoperoxide for adenine and more specifically the C8 addition. It is also noticeable that the critical O…C8 distance for adenine is found to be significantly lower (1.76 Å). Another feature is the less pronounced stability of the adenine endoperoxide structure ($\Delta\Delta E = 6.8$ kcal mol^{-1}), due to electronic factors since no difference of

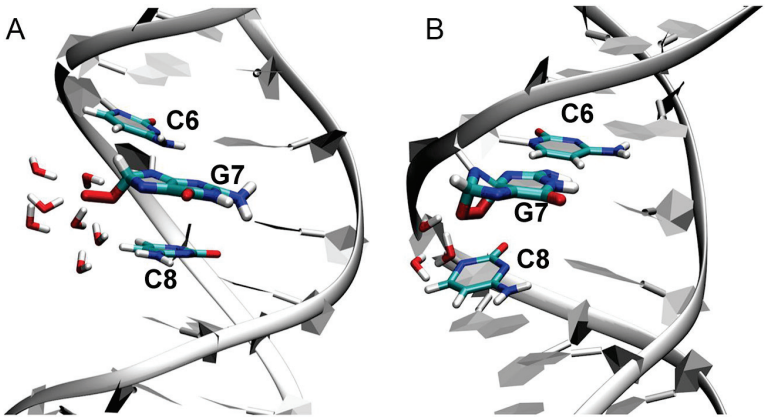


Figure 4. Representative structures for guanine intermediates **2-a** and **3**. Water molecules within 5 Å of O_2 are depicted.

geometries is predicted. Our interpretation is reinforced as no such energy difference is found for all the other intermediates that are planar. Also the closure of the endoperoxide is found to be more difficult with adenine ($TS^{(2\rightarrow3)}$ with a barrier of 10.5 kcal mol^{-1}) probably again due to a more strained structure. Globally these results put on a firm basis the existence of a kinetic blockage. We also note that the charge-separated intermediate **2-a** will be further stabilized by a 3′-purine, as known experimentally for reactions implying electron transfer, by ∼1 kcal mol^{-1} according to auxiliary calculations (Supplementary Figure S4).

The energetic profile in Figure 2 may differ as the singlet-oxygen reaction profile is considered for guanine embedded within a ds-DNA fragment. In particular large deformation of the DNA induced by the guanine oxidation could impose strong additional deformation-free energy, while the mechanical constrains could be seen as further factors producing potential kinetic blockages. To assess DNA structural stability, explicit-solvent MD were performed for an alternate poly(dG)-poly(dC)13-bp sequence where the seventh nucleobase is an oxidized guanine. Such simulations allow us to probe the local environment experienced by **2-a** or **3**, which can tune the relative energies. In turn, our simulations palliate the absence of experimental structures for these high-lying intermediates. The non-planar structure may jeopardize the endoperoxide stability within B-DNA by disrupting π-stacking with adjacent nucleobases C6 and C8 and possibly the Watson-Crick pairing with cytosine C20 on the representative 13-bp sequence depicted in Figure 4, weakening the three canonical hydrogen bonds. The global structural stability of the DNA double helix even in presence of the zwitterionic or endoperoxide intermediates is confirmed by the structural analysis performed along the MD trajectory. Indeed, the bending of the helix is quite modest assuming an average value of 13±6° at 300 K and of 20±10° at 310 K (Figure 5). Also the inter- and intrabasis parameters do not show any significant deviation from the ideal B-DNA values, and most notably the deviation from ideality experienced by the oxidized base-pair remains shy compared to the adjacent base-pair. The opening parameter is modest while a shift of 2 Å concomitant to a combined roll and tilt of ca. 15° are the most noticeable feature. However, those deformations are not accompanied by other significant intrastrand deviation (Supplementary Figure S5a). The same picture holds also for the higher temperature dynamics (see Supplementary Figure S5b). Such deformations are also small when compared to other DNA defects such as photodimers (31) and oxidative intrastrand cross-links (32). Besides G7, C6 is the most impacted nucleobase of the oligonucleotide, with a high 'stagger' to prevent a steric clash with the peroxo bridge. The zwitterionic intermediate is even more structurally innocent on the B-helicity (see Supplementary Figure S6). Hence, we can conclude that even if some strains are present, those differences are quite local and are not indicative of a strong enough structural deformation to preclude the embedding of the endoperoxide or zwitterion intermediates. We note that the solvation pattern and the main B-helical characteristic remain similar upon increase of the temperature to 310 K (see Supplementary Figure S7). Indeed apart from an increased thermal fluctuation, notably leading to a slightly larger bending, all the other main structural parameters are almost unaltered upon increase of the temperature.

From this analysis we conclude that no significant mechanical strain or deformation is induced on DNA structure by the oxidation, confirming the high probability of the existence of such intermediates. Note that as previously cited in the case of **2-a**, the terminal oxygen atom is strongly solvent exposed and hence the zwitterionic form is still favored compared to the diradical (see Supplementary Figure S3).

Downhill, our calculations confirm the experimental fact that the degradation cascade which starts from the endoperoxide **3** is most likely to continue the 4-hydroxy-8-oxo-2′-deoxyguanosine **4**. Stabilization energy accounts for 86.4 kcal mol^{-1} (Figure 2), a considerable exothermicity due to a release of the conformational strain of the endoperoxide (33), reflected in the lenghtening of the C8-O1 bond from 1.35 to 1.44 Å, and an increase of conjugation along the π-system. The rate-limiting step is found to be the opening of the endoperoxide (Figure 6) that implies a high-energy transition state situated at ca. 44 kcal mol^{-1} with at least two water molecules. The latter provides a shuttle for the proton transfer and eventually contributes to stabilize the partially negatively charged oxygen O_2 (see Supplementary Table S1). This channel can occur for solvated guanine ow-

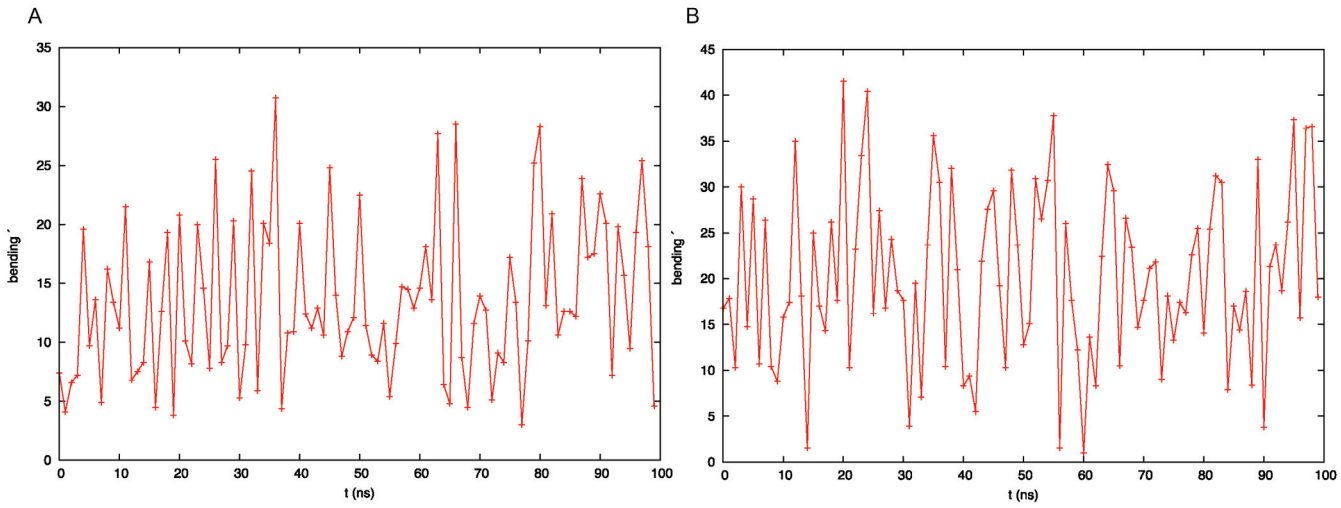


Figure 5. Evolution of the DNA double-strand axis bending all along the MD trajectory for the zwitterion (a) and the endoperoxide (b), respectively.

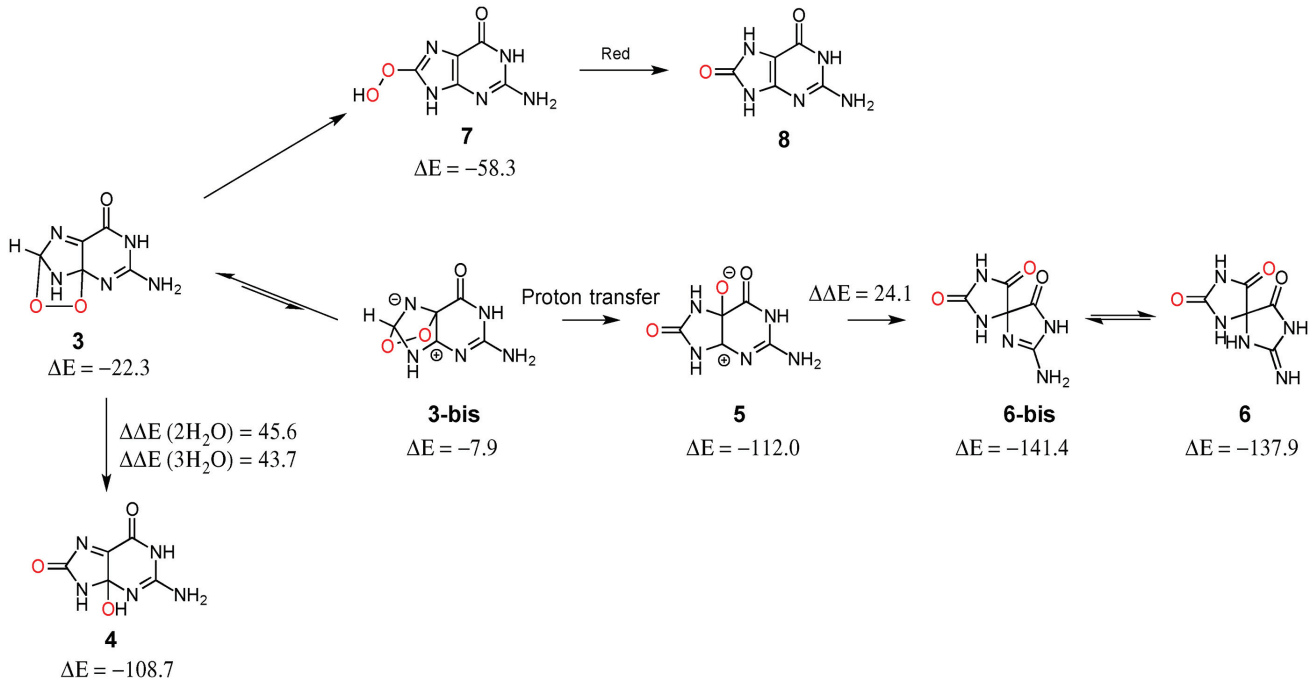


Figure 6. Relative energies (in kcal mol^{-1}) corresponding to three decomposition schemes for the endoperoxide **3**.

ing to dynamic effects that would further lower the barrier. The presence of a chemical species acting as a base would also assist the endoperoxide opening via a Kornblum-De Le Mare reaction (34). Another possible evolution of the endoperoxide **3** (Figure 6) will lead to the formation of the spironucleoside **6**, for which we identified a plausible reaction path via the migration of the oxygen O_2 onto the C5 position of the guanine ring, followed by a proton transfer. This last step constitutes the driving force with an energetic stabilization of -104.1 kcal mol^{-1}. Finally, it has been proposed that 8-oxodGuo can also further react with $^1O_2$ and undergo singlet-oxygen addition onto the C4 = C5 ethylenic bond (the C8 position is no longer available). We have computed the transition state structure for the attack of $^1O_2$ onto C4 position of the oxidized nucleobase and the barrier has been estimated to be lower than 2 kcal mol^{-1} (see ESI). This very small barrier is totally consistent with the experimental report of a 100-fold higher reactivity compared to guanine.

DISCUSSION

The oxidative attack of $^1O_2$ onto guanine has been rationalized both from a kinetic and an energetic points of view based on combined high-level quantum chemistry calculations and MD simulations. In particular mechanistic hy-

potheses explaining the peculiarity of this reaction have been presented. We report strong evidences for the singlet-oxygen attack onto guanine to proceed in two steps to yield an endoperoxide, via an open zwitterionic intermediate. Activation energies compatible with a thermally allowed reaction have been obtained in the case of guanine. The selectivity of $^1O_2$ for guanine compared to adenine has been rationalized too, as the attack on the guanine ring proceeds via lower activation energies, and leads to larger driving forces. On the other hand, the attack of singlet oxygen to 8-oxodGuo has been found to have only a very negligible barrier, rationalizing the experimentally observed larger reactivity than guanine itself. MD simulations, at room and body characteristic temperature, have also confirmed that B-DNA accommodates both endoperoxide and zwitterionic intermediates with no marked distortion. This observation is consistent with the fact that the endoperoxide has been isolated at low temperature, and strongly confirms its crucial role in determining DNA oxidation. For this first part of the singlet-oxygen attack, it would be desirable to further probe this reactivity within an explicitly solvated oligonucleotide, including the temperature and electrostatic embedding of a B-helix, resorting to hybrid quantum mechanics/molecular mechanics to evaluate free energies of activation. However, even if our isolated model for the reactivity is much simpler, we are confident that the same mechanistic picture will hold, even if the exact amount of the kinetic barrier could change. Hence, despite the limitations due to the simplified treatment of the environment, our computational scheme provides a chemically sound explanation of an ubiquitous reaction.

We modeled the strongly exoergonic further degradation of endoperoxide, which implies more complex reaction coordinates and the assistance of two water molecules probed along the classical MD trajectories. Indeed with this assistance endoperoxide can initiate a reactive cascade leading to spiro product, which happens to be the main reaction product in water solution. On the contrary in DNA this path is kinetically blocked, since MD proved water molecules can rarely reach the endoperoxide and the proximal phosphate group disfavors the occurrence of partially negatively-charged intermediates. The water-shielding effect of DNA helix was confirmed both for the 300 and 310 MD trajectories. Instead, an alternative degradation pattern is initiated that goes toward endoperoxide and subsequent fast production of 8-oxodGuo. Hence, the DNA tuning of reactivity has been rationalized and confirmed.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The Pole de Simulation Numérique et Modélisation (PSMN) is acknowledged for the computational resources. E.B. is grateful for a Ph.D. fellowship from the French Minister of Higher Education and Research.

FUNDING

COST action in Chemistry Action CM 1201 'Biomimetic Radical Chemistry'; Labex PRIMES [ANR-11-LABX-0063]. Funding for open access charge: Labex PRIMES [ANR-11-LABX-0063].
Conflict of interest statement. None declared.

REFERENCES

1. Triantaphylidès,C., Krischke,M., Hoeberichts,F.A., Ksas,B., Gresser,G., Havaux,M., Van Breusegem,F. and Mueller,M. J. (2008) Environmental stress and adaptation to stress. *Plant Phys.*, **148**, 960–968.
2. Stratton,S.P. and Liebler,D.C. (1997) Determination of singlet oxygen-specific versus radical-mediated lipid peroxidation in photosensitized oxidation of lipid bilayers: effect of β-Carotene and α-tocopherol. *Biochemistry*, **36**, 12911–12920.
3. Davies,M.J. (2003) Singlet oxygen-mediated damage to proteins and its consequences. *Biochem. Biophys. Res. Comm.*, **305**, 761–770.
4. Devasagayam,T.P.A., Steenken,S., Obendorf,M.S.W., Schulz,W.A. and Sies,H. (1991) Formation of 8-hydroxy(deoxy)guanosine and generation of strand breaks at guanine residues in DNA by singlet oxygen. *Biochemistry*, **30**, 6283–6289.
5. Piette,J. (1991) New trends in photobiology: biological consequences associated with DNA oxidation mediated by singlet oxygen. *J. Photochem. Photobiol. B*, **11**, 241–260.
6. Saito,T., Nishihara,S., Kataoka,Y., Nakanishi,Y., Kitagawa,Y., Kawakami,T., Yamanaka,S., Okumura,M. and Yamaguchi,K. (2010) Reinvestigation of the reaction of ethylene and singlet oxygen by the approximate spin projection method. Comparison with multireference coupled-cluster calculations. *J. Phys. Chem. A*, **114**, 7967–7974.
7. Bobrowski,M., Liwo,A., Oldziej,S., Jeziorek,D. and Ossowski,T. (2000) CAS MCSCF/CAS MCQDPT2 study of the mechanism of singlet oxygen ddition to 1, 3-butadiene d benzene. *J. Am. Chem. Soc.*, **122**, 8112–8119.
8. Leach,A.G. and Houk,K.N. (2002) Diels-Alder and ene reaction of singlet oxygen, nitroso compounds and triazolinediones: transition states and mechanism from contemporary theory. *Chem. Commun.*, **12**, 1243–1255.
9. Song,X., Fanelli,M.G., Cook,J.M., Bai,F. and Parish,C.A. (2012) Mechanism for the reaction of thiophene and methylthiophene with singlet and triplet molecular oxygen. *J. Phys. Chem. A*, **116**, 4934–4946.
10. Ravanat,J.-L. and Cadet,J. (1995) Reaction of singlet oxygen with 2′-deoxyguanosine and DNA. Isolation and characterization of the main oxidation products. *Chem. Res. Toxicol.*, **8**, 379–388.
11. Cadet,J., Ravanat,J.-L., Martinez,G., Medeiros,M. and Di Mascio,P. (2006) Singlet oxygen oxidation of isolated and cellular DNA: product formation and mechanistic insights. *Photochem. Photobiol.*, **82**, 1219–1225
12. Ravanat,J.-L., Martinez,G.R., Medeiros,M.H., Di Mascio,P. and Cadet,J. (2006) Singlet oxygen oxidation of 2′-deoxyguanosine. Formation and mechanistic insights. *Tetrahedron*, **62**, 10709–10715.
13. Ravanat,J.-L., Di Mascio,P., Martinez,G.R., Medeiros,M.H.G. and Cadet,J. (2000) Singlet oxygen induces oxidation of cellular DNA. *J. Biol. Chem.*, **275**, 40601–40604.
14. Martinez,G.R., Medeiros,M.H.G., Ravanat,J.-L., Cadet,J. and Di Mascio,P. (2002) 18O-labeled singlet oxygen as a tool or mechanistic studies of 8-oxo-7, 8-dihydroguanine oxidative damage: detection of spiroiminodihydantoin, imidazolone, and oxazolonederivatives. *Biol. Chem.*, **383**, 607–617.
15. Martinez,G.R., Ravanat,J.-L., Medeiros,M.H.G., Cadet,J. and Di Mascio,P. (2000) Synthesis of a naphthalene endoperoxide as a source of 18O-labeled singlet oxygen for mechanistic studies. *J. Am. Chem. Soc.*, **122**, 10212–10213.
16. Sheu,C. and Foote,C.S. (1993) Endoperoxide formation in guanosine derivatives. *J. Am. Chem. Soc.*, **115**, 10446–10447.
17. Kushwaha,P.S. and Mishra,P.C. (2005) Binding of singlet oxygen with a stacked guanine dimer. *Int. J. Quant. Chem.*, **102**, 435–442.

18. Gu,J., Wang,J. and Leszczynski,J. (2010) Electron attachment-induced DNA single-strand breaks at the pyrimidine sites. *Nucleic Acids Res.*, **38**, 5280–5290.
19. Grüber,R., Monari,A. and Dumont,E. (2014) Stability of the guanine endoperoxide intermediate: a computational challenge for density functional theory. *J. Phys. Chem. A*, **118**, 11612–11619.
20. Mazzone,G., Alberto,M.E., Russo,N. and Sicilai,E. (2014) Ab-initio calculation of the $^1O_2$ quenching mechanism by trans-resveratrol. *Phys. Chem. Chem. Phys.*, **16**, 12773–12781.
21. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Scalmani,G., Barone,V., Mennucci,B., Petersson,G.A. *et al.* (2009) *Gaussian09 Revision D.01*. Gaussian Inc., Wallingford C.
22. Tomasi,J., Mennucci,B. and Cammi,R. (2005) Quantum mechanical continuum solvation models. *Chem. Rev.*, **105**, 2999–3094.
23. Case,D.A., Berryman,J.T., Betz,R.M., Cerutti,D.S., Cheatham,T.E. III, Darden,T.A., Duke,R.E., Giese,T.J., Gohlke,H., Goetz,A.W. *et al.* (2015) *AMBER 2015*, University of California, San Francisco.
24. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Developing and testing a general amber force field. *J. Comp. Chem.*, **25**, 1157–1174.
25. Pérez,A., Marchan,I., Svozil,D., Sponer,J., Cheatham,T.E. III, Laughton,C.A. and Orozco,M. (2007) Refinement of the Amber force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
26. Cieplak,P., Cornell,W.D., Bayly,C. and Kollman,P.A. (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA and proteins. *J. Comp. Chem.*, **16**, 1357–1377.
27. Jorgensen,W.L., Chandraeskar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
28. Izaguirre,J.A., Catarello,D.P., Wozniak,J.M. and Skeel,R.D. (2001) Langevin stabilization of molecular dynamics. *J. Chem. Phys.*, **114**, 2090–2098.
29. Lavery,R., Moakher,M., Maddocks,J.H., Pekteviciute,D. and Zakrewska,K. (2009) Conformational analysis of nucleic acids revised: curves+. *Nucleic Acids Res.*, **37**, 5917–5929
30. Jena,N.R. (2012) DNA damage by reactive species: mechanism, mutation and repair. *J. Biosci.*, **37**, 503–517.
31. Park,H., Zhang,K., Ren,Y., Nadji,S., Singa,N., Taylor,J.-S. and Kang,C. (2002) Crystal structure of a DNA decamer containing a *cis-syn* thymine dimer. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 15965–15970
32. Garrec,J., Patel,C., Rothlisberger,U. and Dumont,E. (2012) Insights into intrastrand cross-link lesions of DNA from QM/MM molecular dynamics simulations. *J. Am. Chem. Soc.*, **134**, 2111–2119.
33. Castillo,A. and Greer,A. (2009) Theoretical studies of a singlet oxygen-releasing dioxapaddlane (1, 4-diicosanaphtalene-1, 4-endoperoxide). *Struct. Chem.*, **20**, 399–407.
34. Staben,S.T., Linghu,X. and Toste,F.D. (2006) Enantioselective synthesis of γ-hydroxyenones by chiral base-catalyzed Kornblum DeLaMare rearrangement. *J. Am. Chem. Soc.*, **128**, 12658–12569.

ChemPubSoc Europe

DOI: 10.1002/chem.201601287

CHEMISTRY
A European Journal
Full Paper

DNA Damage

Singlet Oxygen Attack on Guanine: Reactivity and Structural Signature within the B-DNA Helix

Elise Dumont,*[a] Raymond Grüber,[a] Emmanuelle Bignon,[a, b] Christophe Morell,[b] Juan Aranda,[c] Jean-Luc Ravanat,[d] and Iñaki Tuñón[c]

Abstract: Oxidatively generated DNA lesions are numerous and versatile, and have been the subject of intensive research since the discovery of 8-oxoguanine in 1984. Even for this prototypical lesion, the precise mechanism of formation remains elusive due to the inherent difficulties in characterizing high-energy intermediates. We have probed the stability of the guanine endoperoxide in B-DNA as a key intermediate and determined a unique activation free energy of around 6 kcal mol^{-1} for the formation of the first C–O covalent bond upon the attack of singlet molecular oxygen ($^1O_2$) on the central guanine of a solvated 13 base-pair poly(dG-dC), described by means of quantum mechanics/molecular mechanics (QM/MM) simulations. The B-helix remains stable upon oxidation in spite of the bulky character of the guanine endoperoxide. Our modeling study has revealed the nature of the versatile $^1O_2$ attack in terms of free energy and shows a sensitivity to electrostatics and solvation as it involves a charge-separated intermediate.

Introduction

DNA is formed of four aromatic nucleobases, sugar, and phosphate, and these nucleotides pair through hydrogen bonds and stack perpendicularly to the vertical axis. DNA integrity can be compromised by the action of sunlight or reactive oxygen species such as the hydroxyl radical, the most reactive oxygen species, followed by singlet oxygen, which is recognized as more mild.[1] It has an energy that is 22.4 kcal mol^{-1} above the ground triplet state, which is low enough to be associated with a lifetime usually in the microsecond time range[2] and long enough to diffuse[3] and trigger damage in situ in proximal biomolecules and provoke cellular damage. As it targets DNA, $^1O_2$ interacts selectively with guanine (G)[4,5] to form 8-oxoguanine as the main oxidation product.[6] Going from G to 8-oxoG formally corresponds to the oxidation at the C-8 position followed by the transfer of the 8-H to the N-7 atom. Guanine endoperoxide (see Figure 1), is commonly invoked as an intermediate in the singlet-oxygen-induced degradation pathway,[7] but only one experimental probe, by NMR spectroscopy at low temperature, has been reported.[8] To obtain a detailed mechanistic picture of 8-oxoG formation is a delicate matter because the first intermediates are high-lying, and all the more so as G is embedded within a macromolecular structure. We recently delineated the pathway for the reaction of guanine as an isolated entity,[9] which has also been independently reported based on mass spectrometric analyses and DFT calculations for protonated guanine models.[10] It remains unclear whether the suggested endoperoxide exists or not within B-DNA, as its bulky character could jeopardize the stacking interactions and trigger distortion of the helix. Also, the reactivity of isolated nucleobases can be revisited in B-DNA.[11]

[a] *Dr. E. Dumont, Dr. R. Grüber, E. Bignon*
Univ.Lyon, Ens de Lyon, CNRS UMR 5182
Université Claude Bernard Lyon 1, Laboratoire de Chimie
69342, Lyon (France)
E-mail: elise.dumont@ens-lyon.fr

[b] *E. Bignon, Prof. C. Morell*
Univ. Lyon, CNRS, Université Claude Bernard Lyon 1, Ens de Lyon
Institut des sciences analytiques, UMR 5280,
5 rue de la Doua, 69100, Villeurbanne (France)

[c] *Dr. J. Aranda, Prof. I. Tuñón*
Departament Quimica Fisica, Universidad de València
46100 Burjassot (Spain)

[d] *Dr. J.-L. Ravanat*
CEA and Université Grenoble Alpes, INAC-SyMMES
38000 Grenoble (France)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201601287.

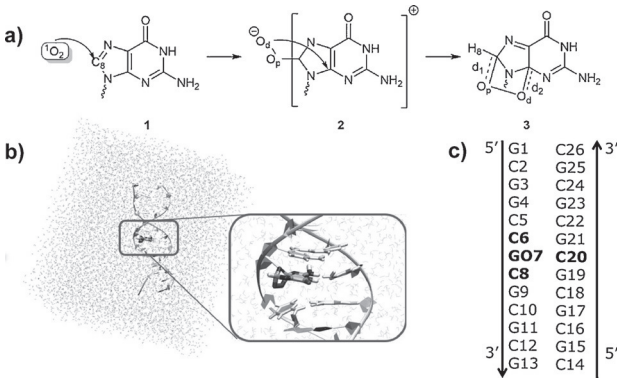


Figure 1. a) Two-step addition of $^1O_2$ to guanine leading to endoperoxide **3** and b) simulation and c) sequence of a 13 bp poly(dG-dC) featuring a central oxidized guanine (GO7)

In this article we report computational evidence for a most facile two-step formation of guanine endoperoxide **3** centered within a 13 base-pair (bp) poly(dG-dC) with an initial free energy of activation of around 6 kcal mol^{-1} and a marked exergonicity (ca. −60 kcal mol^{-1}). The reaction involves the formation of zwitterionic intermediate **2** in which the peroxylate moiety rotates towards the buried C-4 atom. The free-energy profile is compared with the reactivity of the intrinsic system we delineated in a previous study.[9] The contribution of the environment to the reaction profile is analyzed. We also probed the structure of the damaged oligonucleotide featuring an oxidized guanine at its center (GO7, Figure 1): The helical distortion induced upon formation of the endoperoxide was found to be limited. Thus, this study provides an invaluable insight because no experimental structure is available for this key postulated intermediate in oxidative damage.[12]

Results and Discussion

The free-energy profile describing the stepwise addition of singlet oxygen to guanine to form the endoperoxide at C-8 is given in Figure 2. The energy reference is the reactant **1**, with an approaching distance d(G/C-8···O_p) equal to 2.66 ± 0.03 Å (Figure 2a), which compares well with the equilibrium distance of 2.60 Å found for the isolated guanine + $^1O_2$ system.[9] At greater distance, the singlet oxygen experiences a slightly more repulsive interaction with the two closest phosphate moieties leading to a 1 kcal mol^{-1} increase in free energy for d_1 ≈ 3.5 Å. Such a shallow minimum in the free-energy surface is typical of highly diffusive species.

We note that the $^1O_2$ lies in between the two most proximal phosphate units, linking GO7 with C8 and C6, respectively: Average distances to the distal oxygen atom O_d, d(GO7/O_d···C8/OP2) and d(GO7/O_d···GO7/OP2), are around 3.8 and 4.5 Å (see Figure S3 in the Supporting Information). The closest potassium counter ion lies at around 7.4 Å, and this distance remains constant upon formation of **2**. The singlet oxygen attack on guanine exhibits an activation barrier $\Delta G^{\neq}$(**1**→**2**) of 6.2 kcal mol^{-1}, the transition-state structure being located at $d^{\neq}$(O_p···C-8) ≈ 1.95 Å. The same barrier for isolated guanine was found to be more than three times higher (20.4 kcal mol^{-1} in aqueous media, 28.6 kcal mol^{-1} in the gas phase) with a significantly shorter critical distance (1.82 Å). This evidences the critical role of the B-DNA environment and the dynamic contributions in easing the attack of $^1O_2$. Another important feature is the reinforced exergonicity: $\Delta_r G$(**1**→**2**) = −25 kcal mol^{-1} versus −10 kcal mol^{-1} at the nucleobase level. The driving force for oxidation damage to guanine is even stronger within B-DNA, with its environment preferentially stabilizing the charge-separated intermediate **2** over **1.** The zwitterionic character of **2** is almost total, as the oxygen atoms O_d and O_p bear an average negative charge of −0.74*e* and −0.16*e*, respectively; a charge of +0.90*e* is delocalized on the guanine ring. The strong increase in exergonicity as we consider $^1O_2$ attack within B-DNA is first attributed to more favorable electrostatic interactions for the charge-separated intermediate **2**[13] compared with for the DNA···$^1O_2$ reactant **1**. Indeed, for the "outer" conformer of

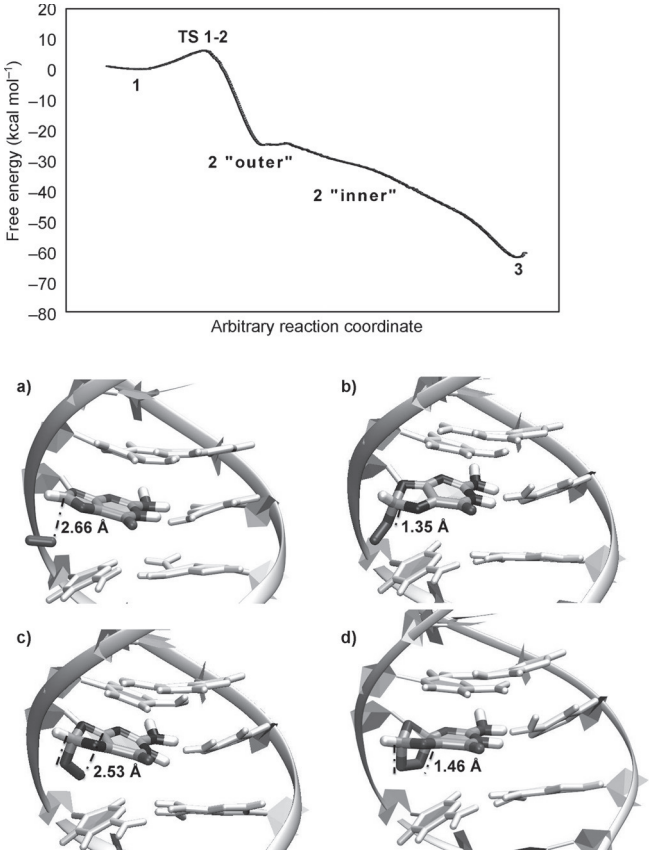


Figure 2. Reaction profile and carton representations for the stepwise addition of singlet oxygen to guanine with characteristic distances in Å (see movies S1 and S2 in the Supporting Information).

2, the C-8 position is directed towards the major groove, with a solvent-exposed -O_p—O_d^- moiety: The average shortest hydrogen bond to the solvent molecules is 1.83 Å, significantly shorter than for **1** (2.35 Å). Radial distribution functions calculated by a 100 ns classical MD simulation for intermediate **2** provides a similar insight[9] (see Figure S1 in the Supporting Information). In addition, $^1O_2$ attack at the C-8 position of guanine within B-DNA can be eased by π-cation interactions between the almost complete cationic character of the oxidized guanine ring and the adjacent nucleobases C6 and C8. Hence, almost total charge separation is favored in B-DNA and reinforces the ease of singlet-oxygen attack. Near the transition state, the charge separation already amounts to 0.5*e* and the B-DNA environment also contributes to lowering the activation barrier. We note that the QM/MM-MD description is by far superior in accounting for a more realistic description of the strongly heterogeneous B-DNA environment. The latter is found to play a decisive role in easing this mechanistic pathway, featuring a charge-separated intermediate with a very low free energy of activation and a strongly exergonic singlet oxygen attack.

Only minor structural changes take place along the reaction path, which is corroborated by monitoring the characteristic bend angle along the reaction profile (see Figure 3a and the Supporting Information). We also report the stability of the

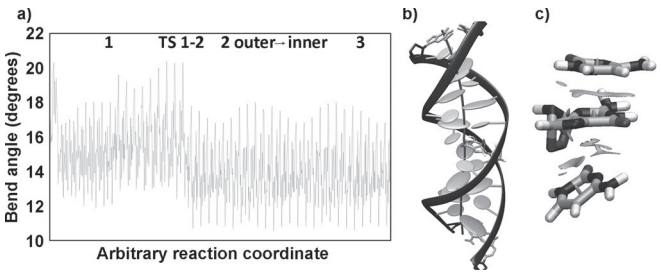


Figure 3. a) Bend angle along the reactive trajectory: The initial value is centered at around 17° and slightly decreases upon cyclization. b) Cartoon representation of the final GO-containing duplex. The helical axis is shown in red. c) NCI isodensities (isovalues 0.5 a.u., for a density cutoff of 0.1 a.u.) showing dispersive interactions (in green) between GO7 and C6 and C8.

Watson–Crick pairing between GO7 and C20 (see Figure S2 in the Supporting Information). The data confirm that the overall B-DNA fragment is only weakly impacted by the addition of singlet oxygen. Thus, there is no energetic penalty associated with a noticeable structural deformation. The formation of **2** is found to be strongly exergonic, by -25.0 kcal mol^{-1}, after passing a low-lying activation barrier, which characterizes a most facile damage process. In summary, the attack of $^1O_2$ is facilitated by a strong solvation pattern stabilizing the zwitterion **2** in the absence of any pronounced B-helical distortion.

We then examined the second part of the reaction profile, going from **2** to the endoperoxide **3** (intramolecular cyclization, blue line of the reaction profile given in Figure 2). This corresponds to a barrierless chemical process, driven by an even more marked exergonicity. The latter amounts to -37.0 kcal mol^{-1}, going from **2** to **3**, that is, a total stabilizing free energy, $\Delta_r G(\mathbf{1}\rightarrow\mathbf{3})$, of -62.0 kcal mol^{-1}. In the following, we will analyze the molecular reasons behind this drastic stabilization. Once linked to the guanine, the peroxylate moiety -O_p–O_d^- in intermediate **2** can rotate around the dihedral τ(O_d-O_p-C8-N7), which is significantly favored when the rotation is directed from the major groove towards the opposite strand. The gain in free energy is estimated to be around -10 kcal mol^{-1} between the "outer" and "inner" rotamers depicted in Figure 2b,c, and also no barrier is found along our QM/MM-MD reaction profile. In contrast, the two rotamers were found to be nearly isoenergetic at the nucleobase level, and are separated by a second transition state identified at 8.3 kcal mol^{-1}.[9] The rotation within the hydrophobic core of the B-DNA helix is accompanied by a nascent stabilizing interaction between the positively charged C-4 (partial charge 0.53*e*) and the terminal oxygen O_d ($-0.74e$), with a decrease of this interatomic distance (2.53 Å in Figure 2c). The K^+ counter ion, initially situated at around 7.7 Å from O_p, moves significantly closer, to a distance of 5.4 Å, following the cyclization and concomitantly the peroxylate moiety moves away from the phosphates (Figure S3 in the Supporting Information).

The charge annihilation in the formation of the non-zwitterionic endoperoxide **3** is the driving force for a most facile cyclization of **2** through the exergonic formation of the C-4···O_d linkage. The exergonicity of the reaction, $\Delta_r G(\mathbf{2}\rightarrow\mathbf{3})$, encompasses four additive components: The intrinsic difference in the free energy between **2** and **3**, which accounts for around -30 kcal mol^{-1} for gas-phase guanine (see Table S1 in the Supporting Information), the contribution from the electrostatic B-DNA environment, the change in the solvation pattern as the -O_p–O_d^- moiety moves from a highly solvent-exposed region to the hydrophobic core of the B-helix (including entropic contributions), and, eventually, the free energy for the concomitant distortion of B-DNA. We seek to assess its importance in the following paragraph. This way, we also gain insights into the structural "footprint" of the guanine endoperoxide within B-DNA, which is currently almost impossible to achieve by experimental means for a high-lying intermediate (as compared with 8-oxo-G[9]).

Figure 3 shows the overall stability of the B-DNA helix during the QM/MM-MD simulations, with a global bend centered at around $14\pm2°$ at the end of the reaction: A similar value ($13\pm6°$) was found in l00 ns classical MD simulations.[9] Thus, the duplex DNA re-adapts rapidly to accommodate the guanine endoperoxide **3** in its center. Other characteristic B-DNA intra- and inter-base-pair step angles and distances remain globally stable (see Figures S4 and S5 in the Supporting Information) with the exception of a stagger of around 1 Å for both C6 and C8 and a tilt angle of 20° for the GO7–C8 base pair: This reflects the fine re-positioning of the flanking pyrimidines to re-adjust to the bulky character of **3**. Hence the two-step peroxidation preserves the stacking interaction C6···GO7 and GO7···C8, as can be seen from the noncovalent interactions (NCI) analysis in Figure 3c. The global B-helical character of the endoperoxide **3** characterizes a very limited perturbation of the macromolecule. Also, there is no steric clash between GO7 and C8, but dispersive interactions between the peroxo bridge and the C8 ring contribute less than 1 kcal mol^{-1} (see the Supporting Information). This is possible due to the "external" position of the peroxo bridge on the five-membered ring of guanine. Meanwhile, the six-membered ring of GO7 maintains its Watson–Crick pairing with C20 (see Figure S2 in the Supporting Information). Overall, the local perturbation of noncovalent interactions featuring GO7 remains limited, which is further confirmed by longer classical simulations.[9]

The additional driving force in the B-DNA environment compared with in the isolated system mostly arises from the change from a solvated to a hydrophobic medium and from electrostatic embedding. The increment in $\Delta_r G(\mathbf{2}\rightarrow\mathbf{3})$ of -37 kcal mol^{-1} reflects to a large extent the intrinsic electronic preference for the endoperoxide: The latter ranges between -10.3 to -34.1 kcal mol^{-1} in aqueous media or in the gas phase (see Table S1 in the Supporting Information), but the decrease in energy is reinforced as -O_p–O_d^- migrates within the B-helix, entering into a hydrophobic cavity as the charge separation vanishes, with more distal backbone phosphates and a proximal K^+ counter ion found at around 5 Å. Its spatial position is well defined due to a dual interaction with the ketolic oxygen of C8 and the amino group of the guanine ring. To conclude, the rotation of -O_p–O_d^- is guided along an electrostatic funnel and the subsequent attack on C-4 is a most facile process within the hydrophobic core of B-DNA.

Experimental kinetic values for the addition of $^1O_2$ to guanine are not available to afford a direct comparison, the first

reason being that the reaction profile continues downhill as the endoperoxide undergoes fragmentation. Although **2** and **3** are high-lying intermediates that are not easily accessible experimentally and therefore debated, in this work we can situate them on an energy baseline within our computational procedure. We note that the free energy should most likely decrease, continuing further towards the intermediates surmised experimentally,[7] an evolution driven by a considerable gain in energy as aromaticity is restored (the fragmentation mechanisms have been sketched in ref. [9]). It is interesting to note that the unique activation barrier we found along our reaction profile is in perfect line with experimental measurements for the activation energy for the reaction of singlet oxygen with fatty acids or thiols, estimated to be between 0 and 6 kcal mol^{-1} and known to be rapid.[14] In such an asymmetric environment and due to the different solvent-accessible surface areas of the C-8 and C-4 positions, a concerted mechanism is less probable (a concerted reaction is hard to conceive in a complex environment with a buried C4 position), yet cannot be totally ruled out and may further ease the $^1O_2$-induced degradation of guanine.

Conclusion

Singlet oxygen is a ubiquitous geno- and cytotoxic agent[15] that can directly react with electron-rich double bonds without the formation of free-radical intermediates.[16,17] We evidence for a representative B-DNA fragment a reaction profile for the formation of the guanine endoperoxide, which we characterize as a very plausible intermediate in the $^1O_2$-driven degradation of guanine. The activation free energy for an attack on the solvent-exposed C-8 position of guanine is only 6 kcal mol^{-1}, which characterizes a very facile process. This is in perfect agreement with the experimental fact that $^1O_2$ is responsible for a large amount of oxidative damage to guanine. It is indeed eased by a higher solvation of the peroxylate moiety and by π-cation interactions stabilizing the oxidized guanine ring. The so-formed zwitterionic intermediate **2** G^+-[OO^-] is then driven towards the most facile and exergonic cyclization by the electrostatic B-DNA environment: The peroxylate moiety migrates from the major groove towards a hydrophobic cavity with a proximal K^+ counter anion. QM/MM-MD simulations allow a more realistic description of the strongly heterogeneous electrostatics of B-DNA and could enable us to investigate the $^1O_2$-induced reactivity at nonlocal topologies such as DNA quadruplexes[18] and more fine sequences effects.[19]

Computational Section

QM/MM simulations were performed starting with a representative structure of the endoperoxide product **3** as the best-defined, not diffusive system that was obtained after 100 ns of Amber 12[20] classical molecular dynamics simulations using the ff99bsc0 force field and deriving RESP[21] charges. The guanine endoperoxide is embedded in a 13 bp poly(dG-dC) solvated oligonucleotide. The QM/MM-MD simulations were performed by using Gaussian09 coupled to the fDynamo program[22] starting with a representative structure that was extracted after 100 ns of classical molecular dynamics simulations (see the Supporting Information). The high-level QM subsystem, consisting of the guanine endoperoxide nucleobase, was described at the restricted DFT/LC-BLYP/6-31+G* level of theory. This range-separated hybrid functional was recently calibrated to gain a balanced description between the zwitterion **2** and the endoperoxide **3**,[9] with electron correlation, a strained system, and an intramolecular charge delocalization, after performing a T1 diagnostic. Concerning the first step of the reaction, it was verified that DFT does not underestimate the intrinsic barrier (20.4 kcal mol^{-1}, see the Supporting Information). The *N*-glycosidic C–N dangling bond depicted with a wavy line in Figure 1a was saturated by using the link atom formalism, a standard choice for the DNA system.[23] After charge redistribution[24] and minimization, a 200 ps NVT simulation was performed by using the hybrid QM/MM potential.

Two distances C-8···Op and C-4···O_d (Figure 1a) were successively employed as reaction coordinates. We first explored the potential energy surface (PES) corresponding to the C-4···O_d bond cleavage from $d_2 = 1.35$–3.00 Å at the LC-BLYP/6-31+G*/MM level of theory for the QM part defined in Figure 1b. The reaction coordinate was then switched to describe the departure of $^1O_2$, with the disruption of the C8···O_p bond (d_1). From the optimized structures, 5 ps relaxation followed by 10 ps of production dynamics with a time step of 1.0 fs were performed for each simulation window (43 and 45, respectively). The collated free energy profile (PMF) was obtained by integration of the histograms using WHAM.[25] Atomic charges are given in atomic units (*e*) and were evaluated on-the-fly by means of the ChelpG procedure[26] within the fDynamo program.[22] The double-helix deformation (bend axis and inter- and intrastrand parameters) was assessed on-the-fly along the reaction profile by using the Curves+ code.[27] A noncovalent interactions (NCI) analysis[28] was conducted to visualize and identify the interactions taking place in the vicinity of the oxidized guanine (between GO7 and C6 and C8).

Acknowledgements

We acknowledge the Pole Scientifique de Modélisations Numériques (PSMN) for computational resources. We thank Dr. Antonio Monari for scientific discussions and assistance using Curves+. J.A. and I.T. acknowledge financial support from FEDER funds and the Ministerio de Economía y Competitividad (project CTQ2012-36253-C03-03) and the Generalitat Valenciana (ACOMP/2015/239). J.A. thanks the Ministerio de Economía y Competitividad for an FPI fellowship. This work was supported by the LABEX PRIMES (ANR-11-LABX-0063) of the Université de Lyon within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR) and by the COST action CM1201 "Biomimetic Radical Chemistry".

Keywords: DNA · DNA damage · DNA structure · molecular dynamics · molecular modeling · singlet oxygen

[1] J. Lee, N. Koo, D. B. Min, *Comp. Rev. in Food Sci. Food Saf.* **2004**, *3*, 21–33.
[2] K. K. Chin, C. C. Trevithick-Sutton, J. McCallum, S. Jockusch, N. J. Turro, J. C. Scaiano, C. S. Foote, M. A. Garcia-Garibay, *J. Am. Chem. Soc.* **2008**, *130*, 6912–6913.
[3] J. Moan, *J. Photochem. Photobiol. A* **1990**, 6, 343–344.

[4] H. Sies, C. F. Menck, *Mutat. Res.* **1992**, *275*, 367 – 375.
[5] J. Cadet, J.-L. Ravanat, G. R. Martinez, M. H. Medeiros, P. Di Mascio, *Photochem. Photobiol.* **2006**, *82*, 1219 – 1925.
[6] Y. Yu, J. G. Muller, W. Luo, C. L. Mayne, A. J. Shallop, R. A. Jones, C. J. Burrows, *J. Am. Chem. Soc.* **2003**, *125*, 13926 – 13927.
[7] *The Chemical Biology of DNA Damage* (Eds.: Wiley. N. E. Geacintov, S. Broyde), **2011**.
[8] C. Sheu, C. S. Foote, *J. Am. Chem. Soc.* **1993**, *115*, 10446 – 10447.
[9] E. Dumont, R. Grüber, E. Bignon, C. Morell, Y. Moreau, A. Monari, J.-L. Ravanat, *Nucleic Acids Res.* **2016**, *44*, 56 – 62.
[10] W. Lu, J. Liu, *Chem. Eur. J.* **2016**, *22*, 3127 – 3138.
[11] C. Patel, J. Garrec, C. Dupont, E. Dumont, *Biochemistry* **2015**, *54*, 425 – 431.
[12] J.-L. Ravanat, G. R. Martinez, M. H. Medeiros, P. Di Mascio, J. Cadet, *Tetrahedron* **2006**, *62*, 10709 – 10715.
[13] The biradical was found to be higher in energy by 9.2 kcal mol^{-1} in isolated moieties, as calculated at the same level of theory and by using an implicit solvation model: R. Grüber, A. Monari, E. Dumont, *J. Phys. Chem. A* **2014**, *118*, 11612 – 11619.
[14] W. T. Yang, D. B. Min, *Lipids in Food Flavors* **2009**, *2*, 15 – 29.
[15] L. F. Agnez-Lima, J. T. A. Melo, A. E. Silva, A. H. Oliveira, A. R. S. Timoteo, K. M. Lima-Bessa, G. R. Martinez, M. H. G. Medeiros, P. Di Mascio, R. S. Galhardo, C. F. M. Menck, *Mutat. Res.* **2012**, *751*, 15 – 28.
[16] F. Sevin, M. L. McKee, *J. Am. Chem. Soc.* **2001**, *123*, 4591 – 4600.
[17] M. Bobrowski, A. Liwo, S. Ołdziej, D. Jeziorek, T. Ossowski, *J. Am. Chem. Soc.* **2000**, *122*, 8112 – 8119.
[18] A. M. Fleming, C. J. Burrows, *Chem. Res. Toxicol.* **2013**, *26*, 593 – 607.
[19] L. F. Agnez-Lima, R. L. Napolitano, R. P. P. Fuchs, P. Di Mascio, A. R. Muotri, C. F. M. Menck, *Nucleic Acids Res.* **2001**, *29*, 2899 – 2903.
[20] D. A. Case, Amber12; University of California, CA, **2012**.
[21] P. Cieplak, W. D. Cornell, C. Bayly, P. A. Kollman, *J. Comput. Chem.* **1995**, *16*, 1357 – 1377.
[22] M. J. Field, M. Albe, C. Bret, F. Proust-De Martin, A. Thomas, *J. Comput. Chem.* **2000**, *21*, 1088 – 1100.
[23] J. Garrec, C. Patel, U. Rothlisberger, E. Dumont, *J. Am. Chem. Soc.* **2012**, *134*, 2111 – 2119.
[24] J. Aranda, K. Zinovjev, M. Roca, I. Tunon, *J. Am. Chem. Soc.* **2014**, *136*, 16227 – 16239.
[25] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, *J. Comput. Chem.* **1992**, *13*, 1011 – 1021.
[26] C. M. Breneman, K. B. Wiberg, *J. Comput. Chem.* **1990**, *11*, 361 – 373.
[27] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, K. Zakrzewska, *Nucleic Acids Res.* **2009**, *37*, 5917 – 5929.
[28] E. R. Johnson, S. Keinan, P. Mori-Sanchez, J. Contreras-Garcia, A. J. Cohen, W. Yang, *J. Am. Chem. Soc.* **2010**, *132*, 6498 – 6506.

Received: March 17, 2016
Published online on July 21, 2016

Ibuprofen and ketoprofen photosensitization mechanism: how to enhance cell death under the influence of UVA light

Emmanuelle Bignon,[1,2,†] Marco Marazzi,[3,4,†] Vanessa Besancenot,[5] Hugo Gattuso,[3,4] Christophe Morell,[1] Leif A. Eriksson,[6] Stephanie Grandemange,[5,*] Elise Dumont,[2,*] Antonio Monari[3,4,*]

[1]Institut des Sciences Analytiques, UMR 5280, Université de Lyon1 (UCBL) CNRS, ENS Lyon, Lyon, France

[2]Université de Lyon, ENS de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie, F69342, Lyon, France.

[3]Theory-Modeling-Simulation, Université de Lorraine − Nancy, SRSMC Boulevard des Aiguillettes, Vandoeuvre-lès-Nancy, Nancy, France

[4]Theory-Modeling-Simulation, CNRS, SRSMC Boulevard des Aiguillettes, Vandoeuvre-lès-Nancy, Nancy, France.

[5]Université de Lorraine Nancy and CNRS, CRAN, Vandoeuvre-lès-Nancy, France.

[6]Department of Chemistry & Molecular Biology, University of Gothenburg, Medicinaregatan 9 c, 40530 Göteborg, Sweden

ABSTRACT: Nonsteroidal 2-arylproprionic acids are widely used, over-the-counter, anti-inflammatory drugs. Photosensitivity is a commonly overlooked adverse effect of these drugs. Based on the combined use of cell viability assays and molecular modeling, we prove and rationalize the photochemical pathways triggering photosensitization for two drugs. Ketoprofen, upon triplet manifold population, produces singlet oxygen whereas ibuprofen photodissociates with subsequent production of two reactive radicals. The formation of metastable aggregates between the two drugs and B-DNA is also directly probed by molecular dynamics. Our approach characterizes the coupled influence of the drug's intrinsic photochemistry and the interaction pattern with DNA. The photosensitization activity of nonsteroidal 2-arylproprionic acids, being added to gels and creams for topical use, should be crucially analyzed and rationalized to enact the proper preventive measures.

INTRODUCTION

Ibuprofen[1–5] and ketoprofen[6–8] (Figure 1) are two common nonsteroidal anti-inflammatory drugs (NSAID) that are used since many years due to their anti-inflammatory,[9] analgesic,[10] and antipyretic[11,12] properties. As many other NSAIDs[13–15] they inhibit the cyclooxygenase enzyme[16] and act by decreasing the production of prostaglandine inflammatory precursors. Although efficient, both drugs present known side effects in particular related to the insurgence of stomach or intestinal bleeding[17–19] and of circulatory or cardiac deregulation.[20–22] Ibuprofen in particular was also associated to the possible appearance of skin blistering[23,24] and a moderate photosensitizing activity was recognized for both NSAIDs.[25–27] Both drugs are commonly sold by prescription or as components of other formulations also available over the counter, the sales volumes being impressively high. Most notably, both ketoprofen and ibuprofen can be found as key components of anti-inflammatory topical gels or creams to be applied on the skin,[28–31] and ibuprofen is also associated with acne treatment.[32]

Ketoprofen is a 2-arylpropionic acid derivative, featuring a benzophenone core (Figure 1). The latter compound, known for its photo-reactivity and photocatalysis efficiency, is usually regarded as a paradigmatic photosensitizer in particular acting against DNA.[33] A photosensitizer is a chromophore whose light absorption triggers photophysical or photochemical phenomena leading to the direct or indirect damage of biological macromolecules.[34–36] Photosensitization mechanisms may involve energy-[37,38] or electron-transfer,[39–42] or activation of molecular oxygen[43,44] via production of the reactive singlet oxygen ($^1O_2$) that ultimately will produce oxidative reactions. Benzophenone has been shown to be able to act via all the different mechanisms,[33,45,46] due to the relatively efficient population of its triplet manifold; in particular leading to triplet-triplet energy transfer or $^1O_2$ activation. However, the phototoxicity of NSAIDs is far less established and should be properly characterized. In this contribution, we rationalize the photosensitization efficiency of ketoprofen performing cell viability assays over different cell lines exposed to ketoprofen and UVA light. We also unveil the photochemical

mechanisms at electronic resolution, and clarify the interactions with biological macromolecules and DNA in particular.[34]

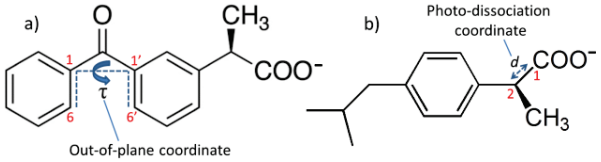


Figure 1) Molecular structures of ketoprofen a) and ibuprofen b). The two coordinates leading to the two different photophysical pathways are also depicted.

The same protocol and procedure are applied to ibuprofen, to also trace back its reported but non-clarified photosensitizing action,[25,26,47] For this drug, commonly used for topical skin applications,[48] the molecular similarity with known sensitizers is much less pronounced. Hence one cannot a priori surmise the photosensitizing efficiency nor hypothesize an evident photochemical pathway.

Combining cell biology techniques and molecular modeling, we confirm the sensitizing activity of both drugs under UVA absorption. The unraveling of the two very different sensitization pathways provides a rationale for the differential effects on cell viability. Furthermore, protective strategies to avoid unwanted sensitization may also be envisaged based on the results of the present communication.

MATERIAL AND METHODS

Cell culture and treatment

MCF7 and MDA-MB231 (ATCC) breast cancer cell lines were cultivated in RPMI 1640 (Gibco) supplemented with 10% of fetal calf serum (Sigma), 2mM of L-glutamine (Sigma), and 0.1mg/mL of gentamycin (Sigma) at 37°C, 5% CO_2.

Cell treatment: ibuprofen (K1751, Sigma) and ketoprofen (I4883, Sigma) were solubilized in ethanol at 200mM and 50mM concentrations, respectively. Cells were cultivated for 24h and treated with 500µM of ibuprofen or ketoprofen for 24h. Control cells were treated for 24h with only ethanol, after cultivation. After the treatments, cells were submitted to UV (0.5µJ/s/cm^2) irradiation during 15 or 30 seconds and re-incubated for 17h (MCF-7) or 6h (MDA-MB231) before analyses.

Cell viability, MTT assay

Cells were plated with 2.10^5 cells per well in 3.5cm plates for 24h with 2ml of culture medium and treated as mentioned above. Cell viability was evaluated using MTT staining. Following treatment, the medium was replaced by medium containing 0.5mg/mL of MTT and incubated 2h at 37°C. Tetrazolium is converted to an insoluble purple formazan that is then dissolved by adding 0.5mL of SDS 25%. Absorbance measurement was performed at 570 nm (plate reader, Victor^{TM} X3 Perkin Elmer). Results are expressed as percentage of cell viability compared to control cells (100%).

Apoptosis detection, western blot

After drug treatments, detached and attached cells were collected and lyzed with lysis buffer ((TrisHCl 10mM pH 7.4, EDTA 5mM, Triton X100 1%, protease inhibitor cocktail (Roche)) for 20min on ice. Protein concentration was determined by the Bradford method and 50 µg of protein were loaded on 10% SDS-PAGE gels and transferred to PVDF membrane. Membranes were blocked with 5% non-fat milk in TBST buffer (20mM Tris-HCl, 120mMNaCl, 0.1% Tween 20) for 1h and then incubated with a cleaved-PARP primary antibody at 1/1000 (552596, BD Biosciences) and tubulin at 1/2500 (AB52866, Abcam) overnight at 4°C. After washing, membranes were incubated with HRP-coupled secondary antibody for 1h and signals were revealed with chemiluminescence system (Biorad) and recorded with the Chemidoc Touch system (Biorad). Densitometric analysis was performed using Image J software.

Statistical Analysis

All results are represented as mean value ± standard error of mean (SEM). Statistical analyses were performed by using Student's *t*-test comparing untreated *versus* treated cells. Statistically significant results were represented as follows: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Quantum chemistry

The *ab initio* multiconfigurational CASPT2//CASSCF methodology[49,50] and time dependent-density functional theory (TD-DFT)[51,52] were applied, using the Molcas 8[53] and Gaussian 09[54] programs, respectively. In the CASPT2//CASSCF calculations, active spaces of 10 electrons-in-9 orbitals and 12 electrons-in-11 orbitals were considered for ibuprofen and ketoprofen, respectively, coupled to the ANO-L-VDZP

basis set. For both molecules, the CAM-B3LYP[55]/6-311+G(d,p)level of theory was applied in the TD-DFT calculations, after benchmarking different functionals (see Supplementary Information). In all cases, water solvent was modeled by the polarizable continuum model (PCM).[56] TD-DFT spin-orbit couplings[57] were calculated with the Dalton2016[58,59] suite of programs. At the TD-DFT level of theory, the electronic density reorganization of the excited states was analyzed in terms of natural transition orbitals (NTO)[60] obtained with the Nancy_EX code.[61]

Molecular Dynamics

All molecular dynamics simulations were carried out with the AMBER12 program.[62] The 10-bp poly(A-T) oligonucleotide was built using the NAB module. Force field parameters were taken from parm99[63] with bsc1 corrections.[64] Ibuprofen and Ketoprofen atomic point charges were computed following the RESP protocol, while parameters were generated using the GAFF force field.[63] The solvent was modeled by a TIP3P[65] water box (See Supplementary Information). Starting structures of the DNA/NSAID complexes were taken from a former study of benzophenone interactions with DNA.[66] Simulations were performed for several initial orientations of the drug interacting with DNA (8 for Ketoprofen and 4 for Ibuprofen) exhibiting either insertion or minor groove binding. After equilibration, trajectories of up to 300 ns were run at a temperature of 300 K (see Supplementary Information).

RESULTS

To determine the biological effect of UV exposure, being an important death inducer, on ibuprofen and ketoprofen pretreated cells, the viability was analyzed.

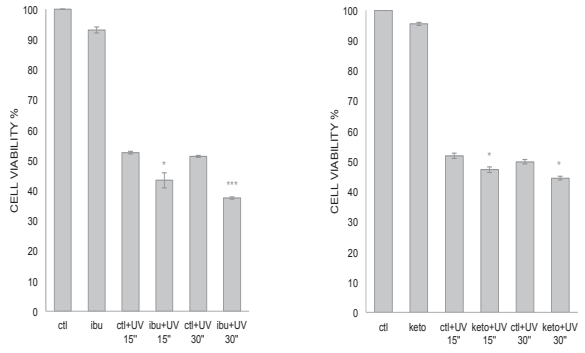


Figure 2) Ibuprofen (left) and ketoprofen (right) reduced cell viability upon UV radiation. Cell viability was assessed by MTT assays on MCF-7 cells untreated (ctl), or treated with 500µM of Ibuprofen / Ketoprofen (ibu / keto), and/or irradiated with UV 15s or 30s. Cells were pretreated with 500µM of ibuprofen or ketoprofen for 24h before UV irradiation and reincubated for 17h. The values are shown as the mean ± SEM from at least 3 independent experiments. * P<0.05; *** P<0.001.

As indicated in Figure 2, pretreatment of MCF-7 cells with either ibuprofen or ketoprofen decreased cell viability after UV irradiation.

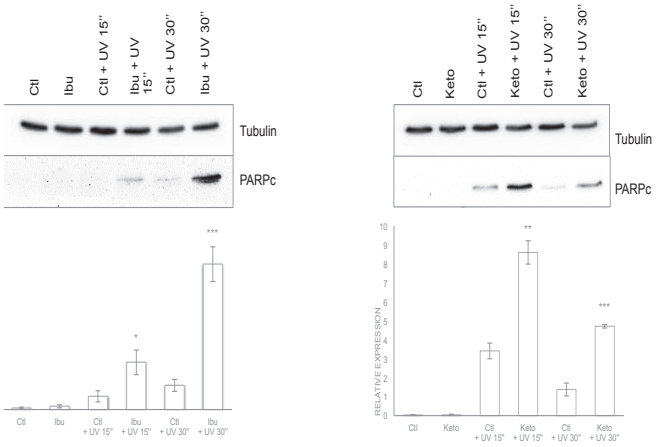


Figure 3) Ibuprofen (left) and ketoprofen (right) enhanced UV induced cell death. Cell death was evaluated by the detection of the cleaved PARP by western blotting on MCF-7 cells untreated (ctl) or treated with 500µM of ibuprofen / ketoprofen (ibu / keto)as described in Figure 2. 50µg of total cell lysate was used. Detection of tubulin corresponds to the loading control. The values are shown as the mean ± SEM from at least 3 independent experiments. * P<0.05; **P<0.01*** P<0.001.

To confirm this effect, analysis of cell death was performed by detecting the presence of cleaved PARP (Figure 3). Our data showed a significant increase of cleaved PARP in MCF-7 cells pretreated with ibuprofen or ketoprofen compared to control cells, after UV irradiation.

These results confirmed that ibuprofen and ketoprofen potentiate the induction of cell death. Our data were validated with the use of a second cell line. As shown in Figure 4, the MDA-MB231 cell viability was also decreased by both pretreatments. While the effects of the two molecules were similar to those observed in MCF-7 cells (no UV), the decrease in cell viability upon UV irradiation was significantly more pronounced after ketoprofen pretreatment compared to ibuprofen pretreatment.

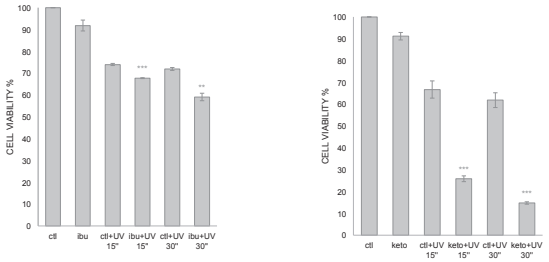


Figure 4) Ibuprofen and ketoprofen enhanced decrease of cell viability induced by UV light in MDA-MB231 cell lines (Cell viability was assessed by MTT assays as described in Figure 2, but with reincubation time 6h. The values are shown as the mean ± SEM from at least 3 independent experiments. ** P<0.01; *** P<0.001.

To rationalize these findings we analyze the properties of the two NAIDS at the electronic and macromolecular level by molecular modeling and simulation. Both ibuprofen and ketoprofen absorb in the UVA region with maxima at 230 and 270 nm, respectively, as confirmed by theory and experiment (see Supplementary Information). This involves direct population of the optically bright π–π^* S_1 state for ibuprofen while for ketoprofen two possible pathways are open: population either of the brightest π–π^* S_2 state followed by fast internal conversion to the n-π^* S_1 state, or direct population of the latter. The electronic density reorganization upon excitation is depicted in Supplementary Information. An important difference between the two NSAIDs is the direct participation of the carboxylate group observed only in the case of ibuprofen.

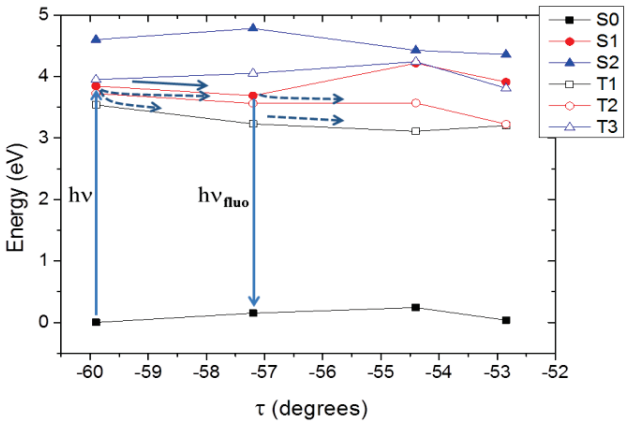


Figure 5) CAS-PT2 MEP of ketoprofen in the S1 state, along the out-of-plane coordinate τ. TD-DFT and CAS-SCF results are provided as Supplementary Information and give the same global picture. After absorption to S_1, different intersystem crossing pathways are open to populate the triplet manifold (dashed arrows), besides relaxation in S_1 followed by fluorescence (solid arrows).

In Figure 5 we report the ketoprofen minimum energy path (MEP) along the dihedral angle τ (see Figure 1a) in the photoexcited singlet S_1 state. As it was the case for benzophenone,[67] we can observe different relevant photophysical processes, mainly delayed fluorescence from the S_1 minimum and intersystem crossing.

Indeed, similar to the case of benzophenone, we recognize the presence of an extended region of quasi-degeneracy between S_1 and the triplet T_2 state, with the lowest-lying T_1 state also being energetically available. This fact coupled to the relatively high spin-orbit coupling (around 20 cm^{-1}) justifies the quasi-unitary population of the triplet manifold via a direct $S_1 \rightarrow T_1$ or indirect $S_1 \rightarrow T_2 \rightarrow T_1$ mechanism as observed for benzophenone and confirmed via static[67] and dynamic studies.[68,69]

Ibuprofen on the other hand proceeds via a completely different mechanism. Indeed, due to the direct participation of the carbonyl group in the electronic density reorganization in the excited state, a photodissociative pathway is open. In Figure 6a we report the MEP along the photodissociative C-C distance and we observe that the S_1 state proceeds to the dissociation of the chemical bond in a barrierless fashion. This in turn will produce two highly reactive radical species, i.e. $COO^{\cdot}$ and the residual of the core ibuprofen moiety as sketched in figure 6b.

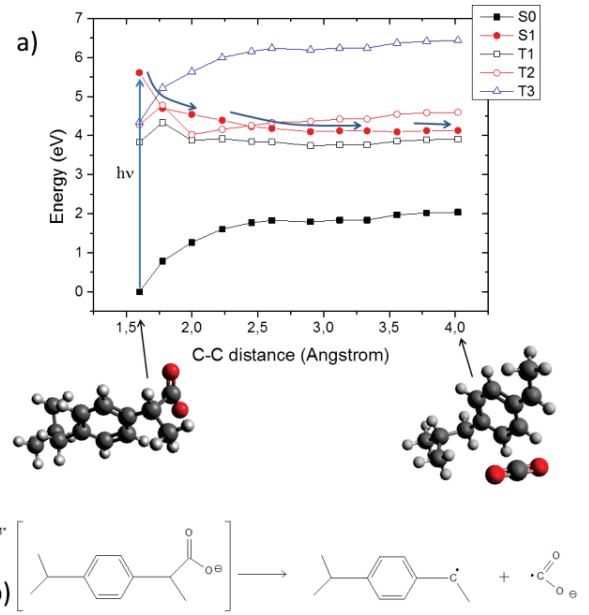


Figure 6) a) CAS-PT2 MEP of ibuprofen in the S_1 state, shown as a function of the photo-dissociation coordinate. TD-DFT and CAS-SCF results are provided in Supplementary Information and give the same global picture. b) Reactants and products are displayed, including the chemical formula.

On the other hand the triplet manifold population can be ruled out for two reasons: first the barrierless potential energy surface observed herein points toward a ultrafast photodissociation, secondly even though the triplet states are energetically accessible, in particular T_1, the spin-orbit coupling is extremely low all along the S_1 potential energy surface never even reaching 1 cm^{-1} (see Supplementary Information).

The photo-toxicity of ibuprofen and ketoprofen is not only a photochemical intrinsic feature, but is also modulated by the strength and mode of interaction with biomolecules. Benzophenone was notably found to lie close to DNA[66] during extended simulation times, while calculations of the binding free-energy[70] confirmed the existence of stable interaction modes. This fact assures the possibility, both from a spatial and temporal point of view, to trigger direct photodamage of the nearby macromolecule. In order to assess the interaction with biological relevant structures, and in particular with DNA that is a common target for photosensitization, we performed classical molecular dynamics to evaluate the formation of persistent drug/DNA aggregates.

For ketoprofen, our simulations characterize a transient double insertion mode represented in Figure 7, which persists a few nanoseconds. This lies in sharp contrast with the parent benzophenone whose DNA-bound stability exceeded 100ns. Interactions with the minor groove are strongly destabilized, probably due to the presence of the negatively charged carboxylate group and consequently of repulsive interactions with the negatively charged phosphates of the DNA backbone. Hence, the intercalation or insertion of the conjugated rings in the hydrophobic DNA core appears as the most favored channel leading to aggregate formation and stabilization, as already observed for similar sensitizers.[66,71–73] The DNA distortion upon interaction with the aryl ketone derivative remains quite localilzed, as evidenced by the analysis of the global parameters (see Supplementary Information). The more bulky character and the lower π-conjugation extent of ibuprofen, still bearing a negative charge, compromises the balance of non-covalent interactions. A consequence, and as opposed to previous docking based studies,[26] much less persistent aggregates are observed.

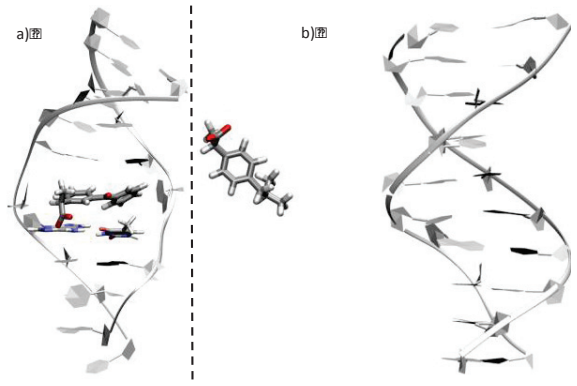


Figure 7) Representative structures extracted from the MD simulation of ketoprofen a) and ibuprofen b) interacting with a B-DNA duplex

DISCUSSION

The combined use of cellular biology assay and molecular simulations provides a global view of the photosensitization activity of two extremely common drugs.

It is evident that the combined exposure to UV light and ibuprofen or ketoprofen induces a marked decreased cell viability as well as an increase in cell death level. The results were confirmed on two different cell lines. However, ketoprofen and ibuprofen display markedly different behavior in the observed effects over the two lines, hence suggesting different modes of action at the molecular level.

Indeed two distinct pathways have been evidenced for the two drugs: while ketoprofen may easily and efficiently populate the triplet manifold, ibuprofen irradiation triggers photodissociation of the carboxylic moiety.

Once in the triplet manifold ketoprofen may provoke all the diverse sensitization effects characteristic of its parent compound benzophenone,[33] and in particular triplet-triplet energy transfer to DNA bases or singlet oxygen production. In addition, and coherently with our scenario, photodissociation of triplet state protonated ketoprofen,[74,75] following proton transfer from the carboxylic to the carbonyl group has been reported.[76] However, differently from benzophenone, and due to the influence of the negative charge of its salt form, ketoprofen only exhibits metastable interactions with DNA. Hence we may conclude that its predominant mode of action will be trough the production of singlet oxygen that does not necessitate a direct coupling with the DNA bases. Furthermore, the production of singlet oxygen may induce sensitization also of other cellular systems such as lipid membranes or proteins.

Ibuprofen on the other hand produces two reactive radical species that may induce considerable damages to biological macromolecules. In particular in the case of DNA one may hypothesize the production not only of nucleobase oxidation but also of strand breaks that are known to be extremely toxic to the cells. The photosensitization mechanism is mostly associated with the production of reactive species (here the two radicals), and has to be considered as non-selective such that it can affect not only DNA but also lipid membranes and proteins. Nevertheless, the presence of metastable interactions, coupled with the generation of highly reactive intermediates, would increase the propensity to induce DNA sensitization.

From a purely photochemical mechanistic point of view ibuprofen photophysics deserves to be regarded as most likely to induce considerable damages to DNA and other biological structures. Indeed its ultrafast dissociation should be more efficient than the triplet population of ketoprofen followed by the activation of singlet oxygen. However, one has to take into account that the ibuprofen absorption maximum is found at 230 nm and only the absorption tail covers the UVA window, *i.e.* the portion of the UV spectrum not filtered by the upper atmospheric ozone layers. Consequently, only a limited amount of ibuprofen molecules will indeed be excited by UVA light and hence the overall photosensitization rate will be less pronounced.

To summarize, our study has clearly shown, and firmly rationalized, the photosensitization efficiency of the two NSAIDs, correlating these to a strong decrease in cell viability and the consequent increase in cell death. However, UV light is efficiently screened by the skin and its penetration length is strongly reduced. Hence, oral intake of ibuprofen or ketoprofen drugs can be seen as safe from a photosensitization point of view. However, both drugs are also used in topical preparations, as creams, gels or patches, to be applied directly to the skin, *i.e.* in parts of the body directly exposed to UV light. Hence, care should be applied in the case of exposure to sunlight or phototherapy, in conjunction with the simultaneous application of topical ibuprofen or ketoprofen based formulations.

ASSOCIATED CONTENT

Supporting Information. Computational details. TD-DFT benchmark on ibuprofen. Experimental and theoretical absorption spectra pf ibuprofen and ketoprofen in water; minimum energy paths and molecular orbitals describing excited states reactivity at different levels of theory; analysis of DNA global parameters.

AUTHOR INFORMATION

[†]The two authors contributed equally and should be regarded as joint first author.

Corresponding Author

*E-mail: (ED) elise.dumont@ens-lyon.fr, (SG) stephanie.grandemange@univ-lorraine.fr, (AM) antonio.monari@univ-lorraine.fr

NOTES

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This research was supported by the COST action "MOLIM: Molecules in Motion". M. M. is thankful to the French and Austrian National Research Agencies (ANR and FWF, respectively) for a grant under the "DeNeTheor" project.

REFERENCES

1. Busson, M. Update on Ibuprofen: Review Article. *J. Int. Med. Res.* **14,** 53–62 (1986).
2. Doherty, M. *et al.* A randomised controlled trial of ibuprofen, paracetamol or a combination tablet of ibuprofen/paracetamol in community-derived people with knee pain. *Ann. Rheum. Dis.* **70,** 1534–41 (2011).
3. Bushra, R. & Aslam, N. An overview of clinical pharmacology of Ibuprofen. *Oman Med. J.* **25,** 155–1661 (2010).
4. Nagi, R., Yashoda Devi, B. K., Rakesh, N., Reddy, S. S. & Patil, D. J. *Clinical implications of prescribing nonsteroidal anti-inflammatory drugs in oral health care—a review*. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology* **119,** (2015).
5. Southey, E. R., Soares-Weiser, K. & Kleijnen, J. Systematic review and meta-analysis of the clinical safety and tolerability of ibuprofen compared with paracetamol in paediatric pain and fever. *Curr. Med. Res. Opin.* **25,** 2207–2222 (2009).
6. Kantor, T. G. Ketoprofen: a review of its pharmacologic and clinical properties. *Pharmacotherapy* **6,** 93–103
7. Seymour, R., Kelly, P. & Hawkesford, J. Pharmacokinetics and Efficacy of Low-Dose Ketoprofen in Postoperative Dental Pain. *Clin. Drug Investig.* **15,** 279–284 (1998).
8. Tuomilehto, H., Kokki, H. & Tuovinen, K. Comparison of intravenous and oral ketoprofen for postoperative pain after

adenoidectomy in children. *Br. J. Anaesth.* **85,** 224–227 (2000).

9. Jin, J. Nonsteroidal Anti-inflammatory Drugs. *JAMA* **314,** 1084 (2015).
10. Jones, P., Dalziel, S. R., Lamdin, R., Miles-Chan, J. L. & Frampton, C. in *Cochrane Database of Systematic Reviews* (ed. Jones, P.) (John Wiley & Sons, Ltd, 2015). doi:10.1002/14651858.CD007789.pub2
11. Davies, N. M. Clinical Pharmacokinetics of Ibuprofen. *Clin. Pharmacokinet.* **34,** 101–154 (1998).
12. Lindsley, C. B. Uses of Nonsteroidal Anti-inflammatory Drugs in Pediatrics. *Arch. Pediatr. Adolesc. Med.* **147,** 229 (1993).
13. Ong, C. K. S., Lirk, P., Tan, C. H. & Seymour, R. A. An evidence-based update on nonsteroidal anti-inflammatory drugs. *Clin. Med. Res.* **5,** 19–34 (2007).
14. Monteiro-Steagall, B., Steagall, P., Lascelles, B. & to rate, unable. Systematic Review of Nonsteroidal Anti-Inflammatory Drug-Induced Adverse Effects in Dogs. doi:10.1111/jvim.12127
15. Musa, K. A. K., Palwai, V. R. & Eriksson, L. A. New nonsteroidal anti-inflammatory molecules with reduced photodegradation side effects and enhanced COX-2 selectivity. *Int. J. Quantum Chem.* **111,** 1184–1195 (2011).
16. Duggan, K. C. *et al.* Molecular basis for cyclooxygenase inhibition by the non-steroidal anti-inflammatory drug naproxen. *J. Biol. Chem.* **285,** 34950–9 (2010).
17. Hernandez-Diaz, S. & Rodríguez, L. A. G. Steroids and Risk of Upper Gastrointestinal Complications. *Am. J. Epidemiol.* **153,** 1089–1093 (2001).
18. Sostres, C., Gargallo, C. J. & Lanas, A. Nonsteroidal anti-inflammatory drugs and upper and lower gastrointestinal mucosal damage. *Arthritis Res. Ther.* **15,** S3 (2013).
19. Matsui, H. *et al.* The pathophysiology of non-steroidal anti-inflammatory drug (NSAID)-induced mucosal injuries in stomach and small intestine. *J. Clin. Biochem. Nutr.* **48,** 107–11 (2011).
20. Antman, E. M., DeMets, D. & Loscalzo, J. Cyclooxygenase Inhibition and Cardiovascular Risk. *Circulation* **112,** (2005).
21. Patrono, C. & Baigent, C. Nonsteroidal Anti-Inflammatory Drugs and the Heart. *Circulation* **129,** (2014).
22. Park, K. & Bavry, A. A. Risk of stroke associated with nonsteroidal anti-inflammatory drugs. *Vasc. Health Risk Manag.* **10,** 25–32 (2014).
23. Lee, A. *Adverse drug reactions*. (Pharmaceutical Press, 2006).
24. Sánchez-Borges, M., Capriles-Hulett, A. & Caballero-Fonseca, F. Risk of skin reactions when using ibuprofen-based medicines. *Expert Opin. Drug Saf.* **4,** 837–848 (2005).
25. Bergner, T. & Przybilla, B. Photosensitization caused by ibuprofen. *J. Am. Acad. Dermatol.* **26,** 114–116 (1992).
26. Husain, M. A., Sarwar, T., Rehman, S. U., Ishqi, H. M. & Tabish, M. Ibuprofen causes photocleavage through ROS generation and intercalates with DNA: a combined biophysical and molecular docking approach. *Phys. Chem. Chem. Phys.* **17,** 13837–13850 (2015).
27. Boscá, F., Marín, M. L. & Miranda, M. A. Photoreactivity of the Nonsteroidal Anti-inflammatory 2-Arylpropionic Acids with Photosensitizing Side Effects¶. *Photochem. Photobiol.* **74,** 637–655 (2007).
28. Coaccioli, S. Ketoprofen 2.5% gel: a clinical overview. *Eur. Rev. Med. Pharmacol. Sci.* **15,** 943–9 (2011).
29. Mazières, B. Topical ketoprofen patch. *Drugs R. D.* **6,** 337–44 (2005).
30. Hyldhal, R. D., Keadle, J., Rouzier, P. A., Pearl, D. & Clarkson, P. M. Effects of Ibuprofen Topical Gel on Muscle Soreness. *Med. Sci. Sport. Exerc.* **42,** 614–621 (2010).
31. Hadgraft, J., Whitefield, M. & Rosher, P. H. Skin penetration of topical formulations of ibuprofen 5%: an in vitro comparative study. *Skin Pharmacol. Appl. Skin Physiol.* **16,** 137–42
32. Wong, R. C., Kang, S., Heezen, J. L., Voorhees, J. J. & Ellis, C. N. Oral ibuprofen and tetracycline for the treatment of acne vulgaris. *J. Am. Acad. Dermatol.* **11,** 1076–81 (1984).
33. Cuquerella, M. C., Lhiaubet-Vallet, V., Cadet, J. & Miranda, M. A. Benzophenone Photosensitized DNA Damage. *Acc. Chem. Res.* **45,** 1558–1570 (2012).
34. Epe, B. DNA damage spectra induced by photosensitization. *Photochem. Photobiol. Sc.*

11, 98–106 (2012).

35. Bonnett, R. Photosensitizers of the porphyrin and phthalocyanine series for photodynamic therapy. *Chem. Soc. Rev.* **24,** 19–33 (1995).
36. Cadet, J., Douki, T. & Ravanat, J.-L. Oxidatively generated damage to cellular DNA by UVB and UVA radiation. *Photochem. Photobiol.* **91,** 140–55
37. Zhao, J., Wu, W., Sun, J. & Guo, S. Triplet photosensitizers: from molecular design to applications. *Chem. Soc. Rev.* **42,** 5323 (2013).
38. Cadet, J., Mouret, S., Ravanat, J.-L. & Douki, T. Photoinduced Damage to Cellular DNA: Direct and Photosensitized Reactions†. *Photochem. Photobiol.* **88,** 1048–1065 (2012).
39. Mitra, R. K., Sinha, S. S., Maiti, S. & Pal, S. K. Sequence Dependent Ultrafast Electron Transfer of Nile Blue in Oligonucleotides. *J. Fluoresc.* **19,** 353–361 (2009).
40. Saito, I. *et al.* Photoinduced DNA Cleavage via Electron Transfer: Demonstration That Guanine Residues Located 5' to Guanine Are the Most Electron-Donating Sites. *J. Am. Chem. Soc.* **117,** 6406–6407 (1995).
41. Hirakawa, K., Ota, K., Hirayama, J., Oikawa, S. & Kawanishi, S. Nile blue can photosensitize DNA damage through electron transfer. *Chem. Res. Toxicol.* **27,** 649–655 (2014).
42. Marazzi, M., Gattuso, H. & Monari, A. Nile blue and Nile red optical properties predicted by TD-DFT and CASPT2 methods: static and dynamic solvent effects. *Theor. Chem. Acc.* **135,** 57 (2016).
43. Hirakawa, K., Hirano, T., Nishimura, Y., Arai, T. & Nosaka, Y. Dynamics of Singlet Oxygen Generation by DNA-Binding Photosensitizers. *J. Phys. Chem. B* **116,** 3037–3044 (2012).
44. Fernandez, J. M., Bilgin, M. D. & Grossweiner, L. I. Singlet oxygen generation by photodynamic agents. *J. Photochem. Photobiol. B Biol.* **37,** 131–140 (1997).
45. Dumont, E. *et al.* Resolving the Benzophenone DNA-Photosensitization Mechanism at QM/MM Level. *J. Phys. Chem. Lett.* 576–580 (2015).
46. Marazzi, M. *et al.* Hydrogen abstraction by photoexcited benzophenone: consequences for DNA photosensitization. *Phys. Chem. Chem. Phys.* **18,** 7829–7836 (2016).
47. Packer, J. L., Werner, J. J., Latch, D. E., McNeill, K. & Arnold, W. A. Photochemical fate of pharmaceuticals in the environment: Naproxen, diclofenac, clofibric acid, and ibuprofen. *Aquat. Sci. - Res. Across Boundaries* **65,** 342–351 (2003).
48. Jorge, L. L., Feres, C. C. & Teles, V. E. Topical preparations for pain relief: efficacy and patient adherence. *J. Pain Res.* **4,** 11–24 (2010).
49. Roos, B. O., Lindh, R., Malmqvist, P. Å., Veryazov, V. & Widmark, P. in *Multiconfigurational Quantum Chemistry* 157–219 (John Wiley & Sons, Inc., 2016). doi:10.1002/9781119126171.ch13
50. Roca-Sanjuán, D., Aquilante, F. & Lindh, R. Multiconfiguration second-order perturbation theory approach to strong electron correlation in chemistry and photochemistry. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2,** 585–603 (2012).
51. Dreuw*, A. & Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. (2005). doi:10.1021/CR0505627
52. Casida, M. E. & Huix-Rotllant, M. Progress in Time-Dependent Density-Functional Theory. *Annu. Rev. Phys. Chem.* **63,** 287–323 (2012).
53. Aquilante, F. *et al.* Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.* doi: 10.1002/jcc.24221 (2015).
54. Frisch, M. J. *et al.* Gaussian~09 Revision D.01.
55. Yanai, T., Tew, D. P. & Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **393,** 51–57 (2004).
56. Mennucci, B. Polarizable continuum model. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2,** 386–404 (2012).
57. Minaev, B. & Ågren, H. Theoretical DFT study of phosphorescence from porphyrins. *Chem. Phys.* **315,** 215–239 (2005).
58. Aidas, K. *et al.* The Dalton quantum chemistry program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4,** 269–284 (2014).
59. Dalton, a molecular electronic structure program. Release Dalton2016.A, see http://daltonprogram.org. *Dalton, a molecular electronic structure program. Release*

Dalton2016.A, see http://daltonprogram.org

60. Martin, R. L. Natural transition orbitals. *J. Chem. Phys.* **118,** 4775–4777 (2003).
61. Etienne, T., Assfeld, X. & Monari, A. Toward a quantitative assessment of electronic transitions" charge-transfer character. *J. Chem. Theory Comput.* **10,** 3896–3905 (2014).
62. Case, D. A. *et al.* (2015), AMBER 15, University of California, San Francisco.
63. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25,** 1157–74 (2004).
64. Ivani, I. *et al.* Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **13,** 55–8 (2016).
65. Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **105,** 9954–9960 (2001).
66. Dumont, E. & Monari, A. Benzophenone and DNA: Evidence for a Double Insertion Mode and Its Spectral Signature. *J. Phys. Chem. Lett.* **4,** 4119–4124 (2013).
67. Sergentu, D.-C., Maurice, R., Havenith, R. W. A., Broer, R. & Roca-Sanjuán, D. Computational determination of the dominant triplet population mechanism in photoexcited benzophenone. *Phys. Chem. Chem. Phys.* **16,** 25393–25403 (2014).
68. Marazzi, M. *et al.* Benzophenone Ultrafast Triplet Population: Revisiting the Kinetic Model by Surface-Hopping Dynamics. *J. Phys. Chem. Lett.* **7,** 622–626 (2016).
69. Favero, L., Granucci, G. & Persico, M. Surface hopping investigation of benzophenone excited state dynamics. *Phys. Chem. Chem. Phys.* **18,** 10499–506 (2016).
70. Gattuso, H., Dumont, E., Chipot, C., Monari, A. & Dehez, F. Thermodynamics of DNA: sensitizer recognition. Characterizing binding motifs with all-atom simulations. *Phys. Chem. Chem. Phys.* **18,** 33180–33186 (2016).
71. Galindo-Murillo, R., Garcia-Ramos, J. C., Ruiz-Azuara, L., Cheatham, T. E. & Cortes-Guzman, F. Intercalation processes of copper complexes in {DNA}. *Nucleic Acids Res.* doi:10.1093/nar/gkv467 (2015).
72. Dumont, E. & Monari, A. Interaction of Palmatine with {DNA}: An Environmentally Controlled Phototherapy Drug. *J. Phys. Chem. B* **119,** 410–419 (2015).
73. Hirakawa, K. & Hirano, T. The Microenvironment of DNA Switches the Activity of Singlet Oxygen Generation Photosensitized by Berberine and Palmatine. *Photochem. Photobiol.* **84,** 202–208 (2008).
74. Musa, K. A. K., Matxain, J. M. & Eriksson, L. A. Mechanism of Photoinduced Decomposition of Ketoprofen. *J. Med. Chem.* **50,** 1735–1743 (2007).
75. Musa, K. A. K. & Eriksson, L. A. in *Quantum Biochemistry* 805–834 (Wiley-VCH Verlag GmbH & Co. KGaA, 2010). doi:10.1002/9783527629213.ch30
76. Li, M.-D. *et al.* Phototriggered Release of a Leaving Group in Ketoprofen Derivatives via a Benzylic Carbanion Pathway, But not via a Biradical Pathway. *Chem. - A Eur. J.* **19,** 11241–11250 (2013).

Chapter 3

DNA structure

Among the possible tertiary structures of DNA, the most commonly encountered in cells is the so-called B-form[121,122]. However, under some conditions, the DNA helix can adopt other shapes, such as the left-handed Z-form[123] and the A-form. Noteworthy, this latter is more common for RNA duplexes due to the presence of a hydroxyl group on the 2' position of the ribose sugar[124]. DNA can also be found in triplex, G-quadruplex (in guanine-rich sequences), stem-loop (hairpin), or pseudoknot shapes - see Figure 3.0.2. These particular conformations exhibit important roles during cell replication and gene expression[125–129].
B-DNA structure exhibits an ideal equilibrium conformation that can be characterized by a series of geometrical parameters : helix bending and local rotations, base pair stacking descriptors, sugar conformation, etc.[130] However, DNA is a flexible macromolecule, whose stability is threatened upon interactions with other chemical entities (eg. proteins or drugs) and upon lesions formation[131]. Consequently, its free energy surface can be revisited. Formation of damages in the DNA B-helix induces more or less pronounced

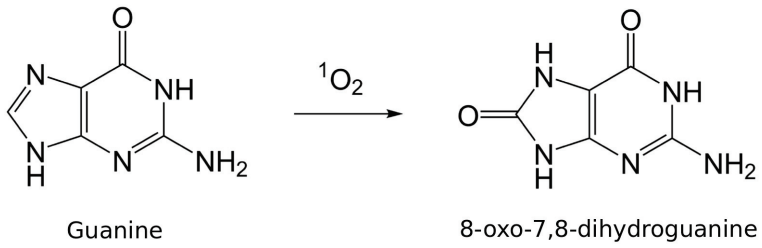


FIGURE 3.0.1: Scheme of the formation of 8-oxo-7,8-dihydroguanine (8-oxoG) from the reaction between a guanine and singlet oxygen ($^1O_2$). This reaction involves a stewise mechanism detailled in publications [P3] and [P4] at the end of Chapter 2.

structural deformations due to alterations of the Watson-Crick pairing and bases stacking. Interestingly, the DNA repair machinery uses these deformations to locate lesions and initiate their removal[93]. Recognition enzymes are able to detect damaged sites probing the changes in the Watson-Crick hydrogen bondings, the nucleobase shape, the charge distribution, or even probing the lesion's complementary nucleobases for bulky distorting lesions[132]. For instance, albeit the 8-oxoG damage differs by only two atoms from its canonical counterpart (see Figure 3.0.1), it is efficiently detected by the hOGG1 glycosylase through their opposite dipole moments[93].
Even if several repair processes coexist, the first stage of the repair always involves the damage recognition by specific enzymes - see Chapter 4. Without this recognition step, the

lesion could not be located, escaping its removal and provoking cyto- and/or genotoxicity. Thereby, the structural signature of a given DNA defect is the main information that permits the repair process. Hence, the understanding of deformations induced by the damage presence in the helix is one of the major current issues of DNA damage research [133].

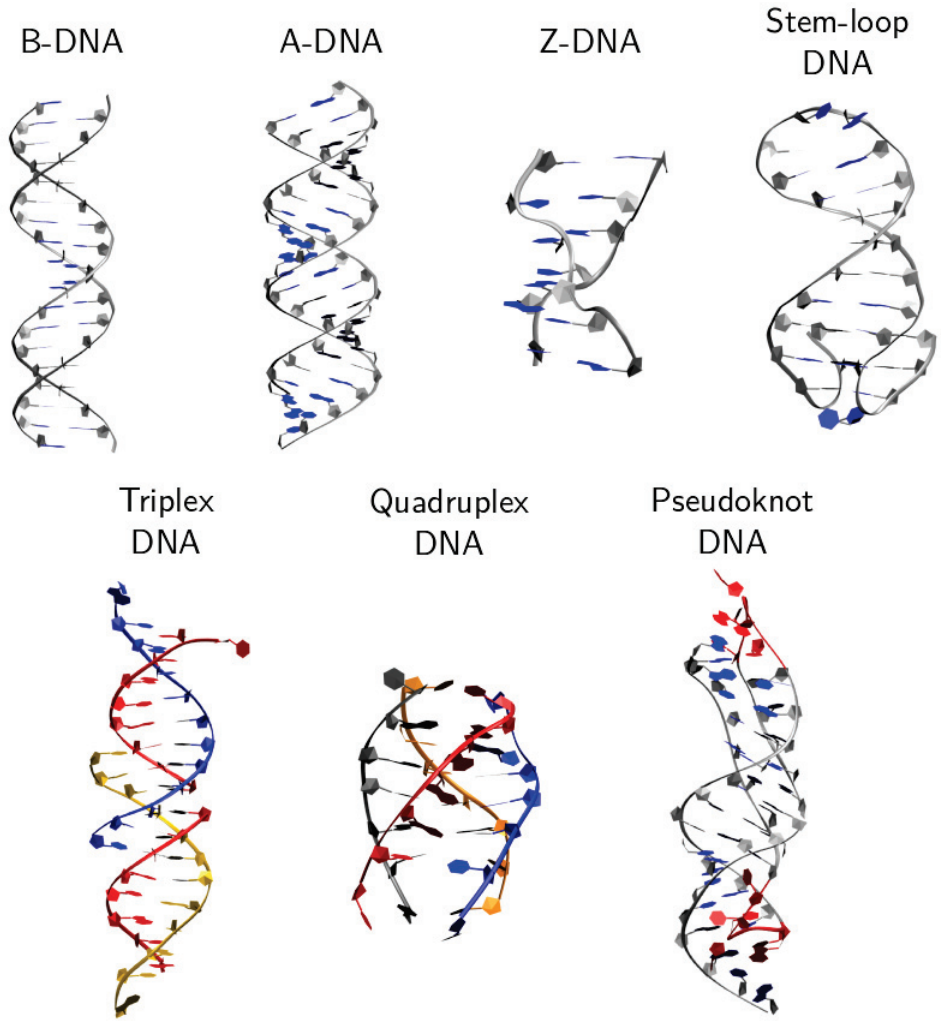


FIGURE 3.0.2: Non-exhautive bestiary of DNA conformations. Top (left to right) : The ideal B-DNA exhibits a major and a minor grooves with different width (structure built with the nab module of AMBER15) ; A-DNA is a shorter and more compact shape than its B analog. Here, base pairs are not perpendicular to the helix axis (structure built with the nab module of AMBER15) ; the Z-DNA is left-handed. This particularity induces the easily distinguishable Z-form of the backbone (PDB code 2IE1 [134]) ; Stem-loop DNA, which can be observed when DNA is single stranded. The hairpin shape is a very common form of single stranded RNA as well (PDB code 2M8Z [135]). Bottom (left to right) : Triplex DNA, the three strands being depicted in different colors. The HB network differs from canonical DNA: instead of the classical Watson-Crick interactions, triplexes exhibit the so-called Hoogsteen pairing (PDB code 1D3R [136] ; Quadruplex DNA of a guanine-rich sequence (G-quadruplex), the four strands are depicted in different colors. As well as for triplexes, quadruplex DNA nucleobases interact by Hoogsteen pairing (PDB code 2MS6) [137] ; Pseudoknot form of DNA. This shape is encountered when DNA is single stranded, and exhibits two hairpins, coloured here in red (PDB code 1YMO) [138].

Nowadays, only few NMR or X-ray structures of oxidatively-induced complex DNA lesions are available, since it is extremely difficult to produce damaged oligonucleotides

experimentally. Moreover, they are static geometries that do not always reflect the conformational fluctuations of the system, and can be affected by crystal packing. DNA is however a highly dynamical molecule whose flexibility is possibly enhanced upon damage, and its mechanical properties can play an important role in lesions formation, recognition and repair. Hence, one has to take into account such structural and dynamical informations when studying DNA damage[18].

Likewise, DNA binding to other molecules, such as drugs or proteins, can also impact its structure and stability[133]. The double helix provides several sites for π-stacking and hydrogen bonding by its nucleobases and negatively-charges phosphates. For instance, several anticancer drugs create stable interactions with DNA nucleobases in order to trigger cytotoxicity and to induce cell death[139]. In cells, DNA constantly interacts with other molecules, and the understanding of such phenomena is an important research field for therapeutics development and life mechanisms elucidation.

Investigations about DNA structure modifications provide real opportunities for cellular mechanisms understanding and biomedicine. In this framework, molecular modelling has proven to be an efficient and versatile tool, described by Schulten and coworkers as a real computational microscope[140]. All-atom simulations can provide unprecedent insights into DNA dynamical behavior, and structural parameters can be computed to determine to what extent the helix structure is impacted and (de)stabilized comparing to its ideal geometry[66,141].
Basics of complex DNA lesion structures and DNA interactions with small molecules are described in the following. A short section will be dedicated to the nucleosome, which offers particular structural features. Within this thesis work, we investigated the structural behaviour of four different systems that will be summarized in section 3.4. Three of them concerned damaged DNA structures (containing clustered abasic sites, interstrand crosslinks and photolesions), while the fourth one was focused on polyamines interactions with DNA. These projects lead to four papers, three published and one submitted, enclosed at the end of this chapter.

3.1 Structure of complex DNA lesions

Lesions formation can induce severe deformations on an initially ideal B-DNA helix, or conversely be structurally innocent - see Figure 3.1.1. Interestingly, such changes in DNA structure are used by the repair machinery to localize defects and initiate their repair[93,142]. However, some cases are challenging for the DNA repair machinery.
Indeed, the two main photolesions, pyrimidine 6-4 pyrimidones (6-4PPs) and cyclobutane pyrimidine dimers (CPDs), exhibit very different repair rates depending on irradiations conditions and cell type. In all cases, despite their apparent structural similarity (two pyrimidines covalently tethered, see Chapter 2), CPDs are poorly repaired compared to 6-4PPs[143,144]. Both photolesions are processed by NER enzymes (see Chapter 4), suggesting that the difference in their repair yields is commensurated to their structural impact on the DNA helix, that facilitate or not the repair process.
Likewise, sequence effects also have an impact on the repair machinery efficiency. Such subtle effects are strongly pronounced in the case of multiply damaged sites (MDS) for example. The latter, mostly induced by ionizing radiations, consist in the combination of two or more lesions on the same or opposite strands, close to each others (within 4-5 nm). Such clustered damages are biologically relevent : they can be lethal[145], and are

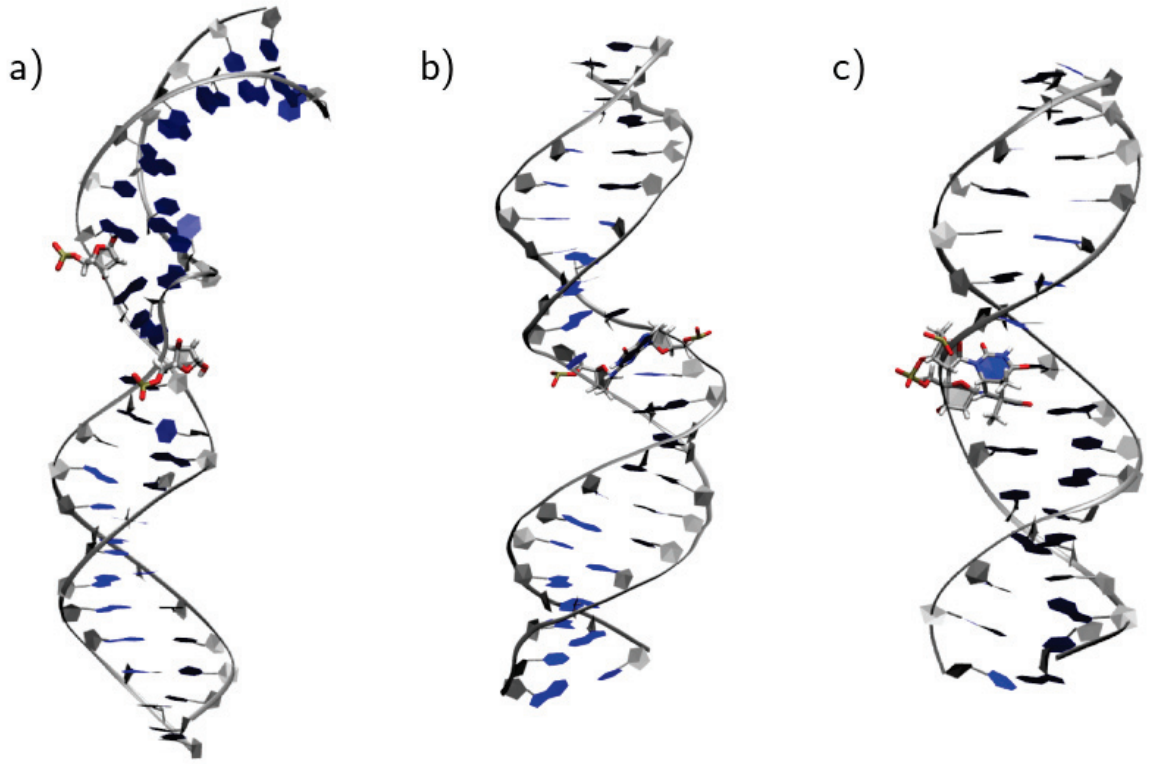


FIGURE 3.1.1: Different structural impact of some complex lesion (depicted in licorice) on the B-DNA shape. For comparisons, the ideal B-DNA is depicted in Figure 3.0.2. a) 23-bp duplex containing clustered abasic sites 3-bp distant. The helix undergoes very strong distortions due to the presence of a multiply damaged site. Such deformations are strongly sequence-dependent as we highlighted in the study discussed in section 3.4.1. b) Presence of an interstrand cross-link (ICL), here resulting from the condensation between an Ap site on one strand and a guanine on the other strand, involves strong local distortions. It was the subject of one of our investigations, described in section 3.4.2. c) 16-bp oligonucleotide containing a cyclobutane pydimidine dimer exhibits only moderate distortions, which can be the reason why it is so difficult to localize by Nucleotide Excision Repair enzymes. This structure comes from a study about photolesions structural impact on DNA, see section 3.4.3.

very challenging for the repair machinery, mainly resulting in the very deleterious double strand breaks[146–148]. Interestingly, several studies brought out the fact that MDS repair is strongly influenced by the nature and the relative position of the lesions within the helix[145,149,150]. Thus, the mechanical properties of such complex lesions are very important in view of their deleterious effects on the cell survival.

Structural informations about complex DNA damages are of major importance to understand the distortion they can induce, and their impact on processing by repair enzymes. Furthermore, modelling of these systems' dynamics provides essential structural insights, that often cannot be insured by experimental tools since such complex lesions are very hard to isolate and characterize. Indeed, radiations to DNA lead mainly to the formation of large amounts of 8-oxoG, hence the isolation of other damages is very complicated. Thereby, computations can palliate the abscence of experimental structures, that has not been resolved by NMR or X-ray yet. By probing of such systems conformational behavior by all-atom molecular dynamics, one can obtain predictive structures that can be further validated with experimental data such as spectroscopic features[18]. Computed mechanical

properties can also be compared to FRET measurements (fluorescence resonance energy transfer), which is relevent for the characterization of helix structural perturbations such as bulges, bend and kink[141,151].

3.2 DNA interactions with small molecules

DNA is prone to interact with numerous small molecules.Thereby, it is used as target for many anticancer drugs, and several DNA ligands are matter of research for development of cytotoxic agents used in cancer chemo- and radiotherapy. For example cisplatin, cyclophosphamide, and mustard agents act as a DNA-DNA crosslinker and alkylation agents, while bleomycin provokes strands cleavage[152].

Meanwhile, some drugs exhibiting photosensitizer features can be considered as potential radiotherapy agents. Noteworthy, some common drugs (eg ketoprofen and ibuprofen see section 2.3.3) contain a chromophore moiety. As a consequence, they could induce DNA lesions upon UV exposure. Hence it is important to delineate how strong could be their photosensitizer potential, in order to avoid undesirable side effects. The understanding of their mechanisms of action involves the in-depth comprehension of their interaction mode with DNA.

Three main DNA/ligand interaction modes have been characterized for B-DNA by Barton and coworkers[153] : minor or major groove binding, intercalation, and insertion - see Figure 3.2.1. Groove binding exhibits electrostatic interactions of the ligand with the backbone's negatively-charged phosphates and does not induce strong helix deformation. Intercalation is allowed when the ligand, which usually should possess planar aromatic rings, stacks between two adjacent base pairs. This binding mode is characterized by an increased viscosity of the medium, as well as a red shift and hypochromic effects in the absorption UV-spectra, induced by the helix elongation[154]. Such ligand can also bind by simple or double-insertion, by eversion of a nucleobase or a base pair respectively. For instance, planar aromatic chromophores, which are common cancer therapy agents, mainly bound to DNA by intercalation or double insertion[155].
Interaction modes can be distinguished by experimental means, using multi-spectroscopic methods such as absorption, fluorescence, NMR measurements, and Resonance Light Scattering technique[156,157].

Nowadays, the understanding of the interaction modes between drug candidates and DNA is a major issue for biomedicine and health research, in order to develop new efficient therapeutic molecules. Within this framework, molecular modelling is an efficient method in order to unravel DNA/drugs preferential binding modes by classical molecular dynamics simulations. In some case, such calculations can even highlight the co-existence of two competitive interaction modes, underlying the fact that such feature are not simple and cannot be reduced to only one major interaction type[158]. Moreover, such methods give access to quantitative and qualitative values that help to characterize the complex system. For instance, binding free energies estimations provide quantitative energetical values that allows to assess ligand affinity for DNA[159]. Likewise, the non-covalent interactions (NCI) method[160] provides insights into π-stacking and hydrogen bonds network strength, that can explain why some drugs are capable of (double-)insertion, providing stable interactions that compensate for the disruption of the initial Watson-Crick pairing and the ejection of nucleobase(s).

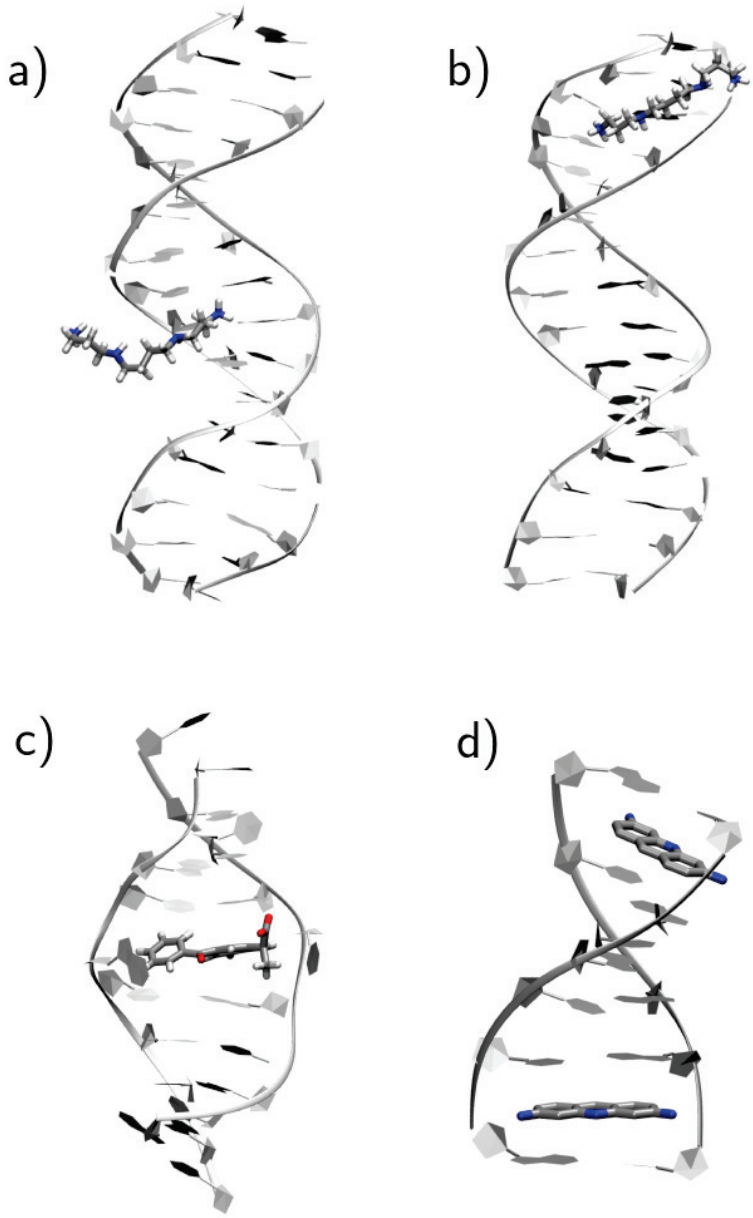


FIGURE 3.2.1: Taxonomy of some DNA/ligands interaction modes. DNA and ligands are depicted in new ribbons and licorice representations respectively. a) Major groove binding of the dicationic spermine. This structure was obtained in the framework of our study on polyamines interactions with DNA (see 3.4.4. b) Same than a) but here the spermine is lying in the minor groove. c) Double insertion mode of anionic ketoprofen within DNA. Eversion of the two nucleobases are easily distinguishable. This structure comes from the study we led on ketoprofen and ibuprofen interactions and photoreactivity with DNA, see section 2.3.3. d) Intercalation of two proflavine molecules within a 6-bp oligonucleotide. Here, there is no need of nucleobase(s) ejection, the ligand only intercalates between two base pairs. This crystal structure was obtained by X-ray diffraction (PDB code 3FT6)[161].

3.3 About nucleosomal DNA

In cells' nuclei, DNA is wrapped around a 8-histones core (two H2A/H2B dimers and a H3/H4 tetramer), the whole complex constituting the so-called nucleosome[162].

This particular structure allows the DNA macromolecule, which end to end reach around 2 meters, to be compacted in the cell, whose diameter lies within 5 to 10 μm. Furthermore, nucleosomal and naked DNAs do not exhibit the same behavior under damage formation and repair. Indeed, the nucleosomal structure protects DNA from damaging[163,164]. For example, it limits UV-induced damages formation for nucleotides facing the histones : the furthest from the histones surface, the most impacted by UV radiations[165]. Likewise, nucleosomal DNA containing clustered Ap sites are more inclined to undergo protein cross-linking and strand breaks because of the close presence of histone tails that favor their removal[166].

Moreover, the wrapping around histones can generate hindrance for the repair enzymes

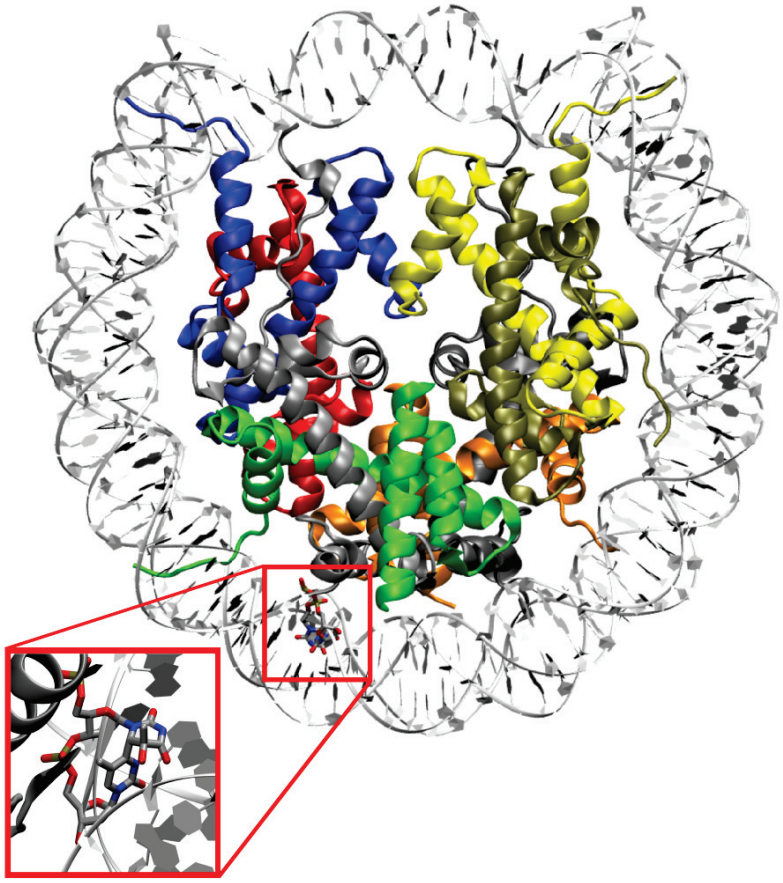


FIGURE 3.3.1: Nucleosome structure containing two 64PP lesions (crystal structure pdb code 4YM5). Histones are represented in new cartoon and colored by chains, the two H2A proteins are in silver and grey, H2B in green and orange, H3 in yellow and blue), and H4 in tan and red. The 146-bp DNA helix is depicted in new ribbons and in white. Only one of the two symmetric 64PP lesions is shown herein licorice representation at the bottom of the nucleosome. A zoom on the damage is depicted in the red box in order to have better appreciation of the lesion's structure. This nucleosomal structure has been resolved in 2015, which was a real ambitious goal.

processing[167], unless the highly compacted chromatin structure has been shown to stay accessible for these enzymes[164].

Hence, unravelling the dynamics of a lesion within nucleosomal DNA is important in order to understand its mechanical properties. The curvature of the helix and proximal residues

certainly revisite the conformational landscape, in line with experimental results. To this extent, molecular modelling is an efficient tool despite the fact that the high number of atoms drastically increases the calculation times comparing to naked DNA. At the end of this thesis work, we began to investigate 6-4PPs-containing nucleosomal DNA in order to assess their structural behavior within this particular environment - see Figure 3.3.1. However, calculations are still in progress since it is the project of our current Master 2 trainee. Results are preliminary and will not be described in this manuscript.

3.4 Our projects within this framework

3.4.1 Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion [P5]

Clustered abasic sites are highly cytotoxic lesions, rare under normal conditions but generated by ionizing radiations. Strong sequence effects impacting such multiply damaged sites repair rates have been highlighted by Georgakilas and coworkers[168,169]. Indeed, their investigations revealed that distance between the Ap sites highly impacts their processing by Ap-endonucleases Nfo and APE1. Thus, it was interesting to gain insights in structural properties of such systems using molecular modelling. In collaboration with Dr. Georgakilas (Athens NTU, Greece), we studied these sequences effects on the dynamical behaviour of clustered Ap sites containing 23-bp duplexes studied experimentally, in order to assess the structural differences that impact their repair. To this extent, together with our collaborators from Nancy University, we performed 500ns-long classical molecular dynamics simulations with non-canonical parameters generation, and a structural analysis of each system. This investigation revealed strong structural differences of the studied sequences, which were in good agreements with experimental results for the Nfo enzyme. To rationalize the sequence effect for APE1, it turns out that the DNA-enzyme complex should be studied - see chapter 4 and publication [P6].

3.4.2 Interstrand cross-linking implies contrasting structural consequences for DNA: insights from molecular dynamics [P7]

Interstrand cross-links (ICL) are biologically highly relevent damages. The covalent bond that is formed between the two complementary strands avoids their separation during DNA replication, provoking stalling of RNA-polymerase at the lesion site, hence cytotoxicity. Furthermore, their repair, which is so far not completely elucidated, may generate double strand breaks, which are very deleterious lesions prone to induce loss of genetic information - see section 4.2.5. Up to now, the structure of ICL are ill-defined in the literature, because they are hard to obtain by experimental means. A study of Price et al.[170] highlights the difference of formation yields between ICL arised from Ap sites with adenine vs with guanine, meanwhile underlying strong sequence effets. Relying on this experimental work, we led classical molecular dynamics simulations on sequences used by Price et al. in order to unravel the dynamical features that rule the formation yield of Ap-dA vs Ap-dG ICLs. Moreover, our calculations were able to palliate the abscence of experimental structures, giving access to conformational behaviour of such lesions in DNA. Our collaborators from the Institute of Organic Chemistry and Biochemistry of Praha (Czech Republic), Dr. Lankaš and Dr. Geršata, performed stiffness evaluation on

structures obtained by our simulations in order to gain insights into the oligonucleotide structural features.

3.4.3 Conformational polymorphism or structural invariance in DNA photoinduced lesions: implications for repair rates [P8]

Although experimental structures are available for pyrimidine 6-4 pyrimidone (6-4pp) and cyclobutane pyrimidine dimers (CPD) photoproducts, the fact that 6-4pp is repaired more efficiently than CPD is not clearly explained in the literature, with a lack of information about the damaged DNA dynamics features. In this study, we investigated the dynamical behaviour of a 6-4pp lesion within DNA in order to highlight different conformational features comparing to CPD, providing new informations that can be used to explain the repair rate difference between the two photolesions. Together with our collaborators from Nancy University, we performed molecular dynamics simulations on oligonucleotides containing 6-4pp on one hand and CPDs on the other hand. Thus, we highlighted differences regarding the B-helix distortion in the presence of 6-4pp or CPDs. This project brought out essential structural insights, underlining the fact that the recognition of the two photolesions by the NER enzymes might depend on their difference of structural signature into the DNA helix.

3.4.4 Molecular dynamics insights into polyamines-DNA binding modes: implications for cross-links selectivity [P9]

During this thesis, I had the chance to do a two weeks stay at the CEA of Grenoble, in the Nucleic Acids Laboratory. Through discussions with Dr. Ravanat, with whom we already worked on singlet oxygen attack onto guanine[67,68], we decided to study polyamines interaction modes with DNA. Indeed, it is well known that putrescine, spermidine and spermine interact with DNA, and are able to react with the guanine radical cation, leading to DNA-polyamine toxic cross-links. In order to decipher how such polyamines interact with DNA, we performed classical molecular dynamics simulations. Parameters were generated for polyamines. Our calculations brought evidences that according to their polyamine length (putrescine < spermidine < spermine) and their protonation state (from neutral to +4), polyamines interaction modes differ. Thus, major and minor grooves binding were evidenced, as well as interactions with negatively-charged phosphates, underlying the different possibilities of binding according to the charge and the length of the ligand. Moreover, this can be considered as a first model to study the DNA-protein cross-linking (specially with histones) through lysines, chemically similar to polyamines.

3.5 Publications

8588–8599 *Nucleic Acids Research, 2016, Vol. 44, No. 18* *Published online 1 September 2016*
doi: 10.1093/nar/gkw773

Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion

Emmanuelle Bignon[1,2,†], Hugo Gattuso[3,4,†], Christophe Morell[2], François Dehez[3,4], Alexandros G. Georgakilas[5,*], Antonio Monari[3,4,*] and Elise Dumont[1,*]

[1]Univ Lyon, Ens de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France, [2]Institut des Sciences Analytiques, Université de Lyon 1 and CNRS, F-69100, Villeurbanne France, [3]Université de Lorraine -Nancy, Theory-Modeling-Simulation SRSMC, F-54506, Vandoeuvre-lès-Nancy, France, [4]CNRS, Theory-Modeling-Simulation SRSMC, F-54506, Vandoeuvre-lès-Nancy, France and [5]DNA damage laboratory, Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou 15780, Athens, Greece

Received June 22, 2016; Revised August 22, 2016; Accepted August 23, 2016

ABSTRACT

Clustered apurinic/apyrimidinic (AP; abasic) DNA lesions produced by ionizing radiation are by far more cytotoxic than isolated AP lesion entities. The structure and dynamics of a series of seven 23-bp oligonucleotides featuring simple bistranded clustered damage sites, comprising of two AP sites, zero, one, three or five bases 3′ or 5′ apart from each other, were investigated through 400 ns explicit solvent molecular dynamics simulations. They provide representative structures of synthetically engineered multiply damage sites-containing oligonucleotides whose repair was investigated experimentally (Nucl. Acids Res. 2004, 32:5609-5620; Nucl. Acids Res. 2002, 30: 2800–2808). The inspection of extrahelical positioning of the AP sites, bulge and non Watson–Crick hydrogen bonding corroborates the experimental measurements of repair efficiencies by bacterial or human AP endonucleases Nfo and APE1, respectively. This study provides unprecedented knowledge into the structure and dynamics of clustered abasic DNA lesions, notably rationalizing the non-symmetry with respect to 3′ to 5′ position. In addition, it provides strong mechanistic insights and basis for future studies on the effects of clustered DNA damage on the recognition and processing of these lesions by bacterial or human DNA repair enzymes specialized in the processing of such lesions.

INTRODUCTION

Abasic DNA sites constitute one of the most common families of DNA lesions. An apurinic/apyrimidinic (AP) site is produced by the excision of one DNA constituent base leaving the uncapped deoxyribose moiety. Most notably AP lesions may be produced directly, as a consequence of exposure to ionizing radiation (IR) or as the products of monofunctional DNA N-glycosylases and base excision repair (BER) enzymes (1). In solution the AP sugar moiety experiences equilibrium between its cyclic and open forms (See Supplementary Data, Supplementary Figure S1). Although, the ring-opened form triggers dangerous interstrand cross-links (2) the equilibrium strongly favors the cyclic form that accounts for about 99% of the occurrence (3). NMR and molecular modeling agree in evidencing a strong local deformation of the DNA double-helix in response to the lesion (4). This distortion is also exacerbated by the fact that AP sites are prone to adopt an extra-helical position that may also correlate with the ease of repair. Indeed, the lesion is usually flipped out and inserted in a specific protein pocket during the excision reaction (5,6).

In the case of exposure to IR or even high oxidative stress, a specific pattern of highly localized ionizations and free radical attack is produced as also supported by various biophysical track structure models and Monte Carlo simulations. (7,8) The clustering of DNA damage may involve AP sites as well as double strand breaks (DSBs) and others non-DSB lesions like oxidized bases and it is currently accepted as one of the most important types of DNA damage with high biological significance due to its repair resistance as also reviewed in (9). Recently the repair resistance of clus-

*To whom correspondence should be addressed. Tel: +33 3 83 68 43 80; Fax: +33 3 83 68 43 71; Email: antonio.monari@univ-lorraine.fr
Correspondence may also be addressed to Elise Dumont. Tel: +33 4 72 72 88 46; Fax: +33 4 72 72 80 80; Email: elise.dumont@ens-lyon.fr
Correspondence may also be addressed to Alexandros G. Georgakilas. Tel: +30 210 772 4453; Fax: +30 210 772 3025; Email: alexg@mail.ntua.gr
†These authors contributed equally to the work as the first authors.

tered DNA damages comprising of bulky guanine adducts opposite to AP sites has also been considered(10). DNA lesion clustering induction, its role(s) in cell homeostasis, systemic effects and the pathways involved in its *in vivo* processing and repair, remain still an open question in current biology (11–13). Clustered abasic sites (14) (clustered AP) have been produced in oligonucleotides (15,16), plasmid, nucleosome (17–19) or found in irradiated DNA (20), cells and animals (21).

Repair processing of clustered AP has been investigated (17,21) and the available data show a clear dependence of the repair efficiency on the relative position of the clustered AP sites. Interestingly it appears that bacterial excision enzymes such as the Escherichia coli *endonuclease* IV (*Nfo*) behave quite differently than the corresponding human endonuclease APE1. However, the molecular bases of the observed specificity still need to be clarified. Earlier NMR and molecular dynamics studies have shown that when two bistranded AP sites are staggered 5′ to each other and 1–3 base pairs apart, they are located in the minor groove when adopting an extrahelical conformations. Therefore, major structural changes for APE1 incision are required. In contrast, when AP residues are staggered 3′ to each other and 1–3 bp apart they are extruded on the major groove, adopting a conformation that may favor enzyme binding by lowering kinetic energy barriers (4).

Numerical simulation techniques have nowadays reached a considerable maturity (22) and can complement experimental studies in tackling DNA damage (23) and repair problems, offering the possibility to visualize the interaction between DNA and enzymes at an atomistic resolution. Molecular modeling has been applied to damaged oligonucleotides, such as abasic sites (24–26), bulky adducts (10,27), cross-links (28) or photolesions, (23,29) and also to duplexes in interaction with repair enzymes (30). Molecular dynamics (MD) explorations of combined 8-oxoguanine and abasic site have been achieved independently (31,32), yet the combination of two AP to give clustered adducts is still missing. On the other hand the structural bases for the recognition of simple AP sites have been rationalized (33) and hence may offer a good comparison to understand the difference induced by the cluster lesion.

In this work, we explore by means of explicit solvent MD simulations, the structural evolution of the seven clustered AP oligonucleotides previously studied by Georgakilas *et al*. (17,21). They feature an AP site as the 35th nucleobase and a second one on the complementary strand. Their relative position is designed as seq ±0, 1, 3, 5 as shown in Scheme 1, in particular in seq0 the two AP sites are directly facing each other. Most notably a clear relative position dependence effect of cluster damages has been evidenced especially in the case of bacterial repair (21), as evidenced in Figure 1. MD simulations will allow us to unravel all the distinct structural deformations of clustered AP-containing strands as compared to the reference simple AP, providing a structural framework to rationalize the observed repair efficiency.

MATERIALS AND METHODS

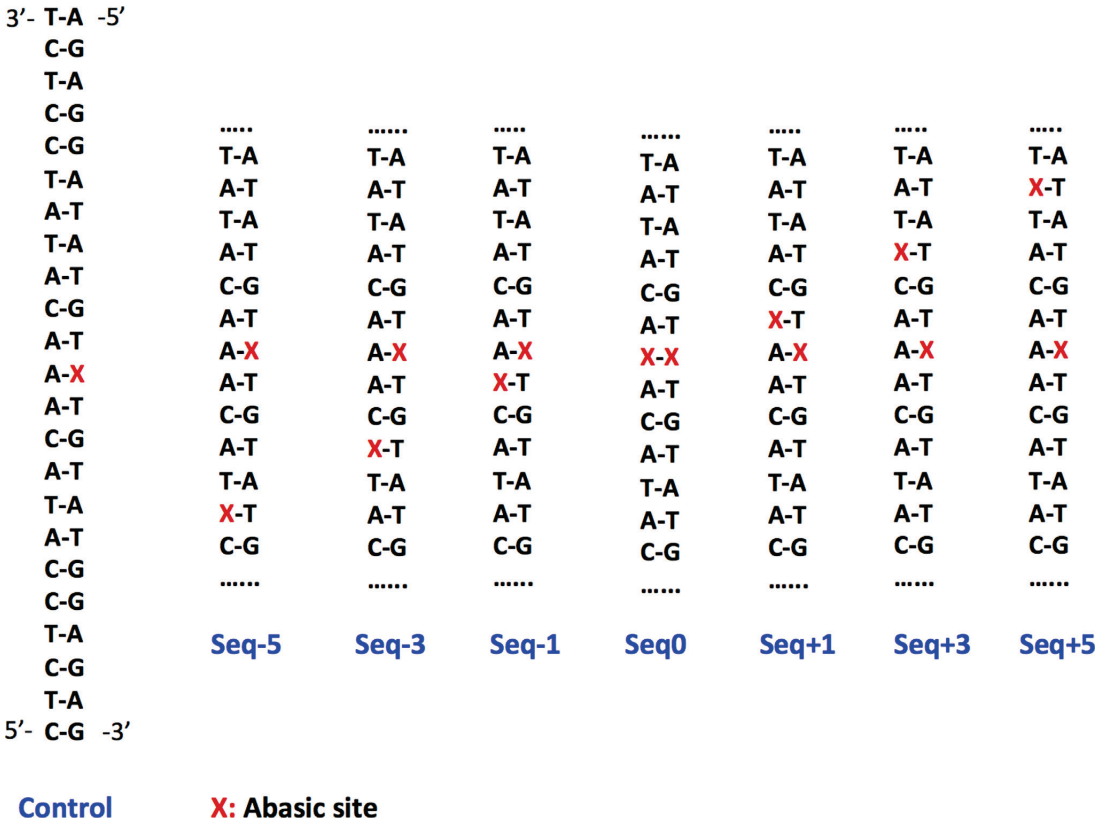
Oligonucleotide structure

The choice for the specific oligonucleotide sequence presented in Figure 1 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*. (17,21). 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 homology with human chromosomal regions (see Supplementary Data).

Computational protocol. The Amber12 software (34) package and the Amber *ff99* force field (35,36) including bsc0 corrections (37) 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 (34). Abasic sites were inserted at specific positions to generate the seven sequences corresponding to the synthetized oligonucleotides of reference (17), with the *xleap* module. The parameters for the AP site (see Supplementary Data) were generated with antechamber and *parmcheck* subprograms, and atom point charges were computed using the RESP protocol (38), see Supplementary Table S1 and Supplementary Figure S2. Sodium cations (Na^+) were added in order to neutralize the systems, which were immersed in a truncated octahedral TIP3P (39) 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 γ ln of 1 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 (40). To assess residency times for hydrogen bonds, we fix a cut-off on the H…O of H…N distances of 2.4 Å. We adopt the criterion that extrahelical positions correspond to a distance C1′…C1′ greater than 14 Å, as proposed by Lavery *et al*. (25). 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 Supplementary Data (Supplementary Table S2 and Supplementary Figure S3).

Note that due to the equilibrium between the closed and open form AP sites can exist as α- or β-anomers. However, as pointed out both theoretically (41) and via NMR determinations (32) the two isomers give rise to duplexes having



Scheme 1 Representation of the sequence of the studied oligonucleotides as taken from (17,21).

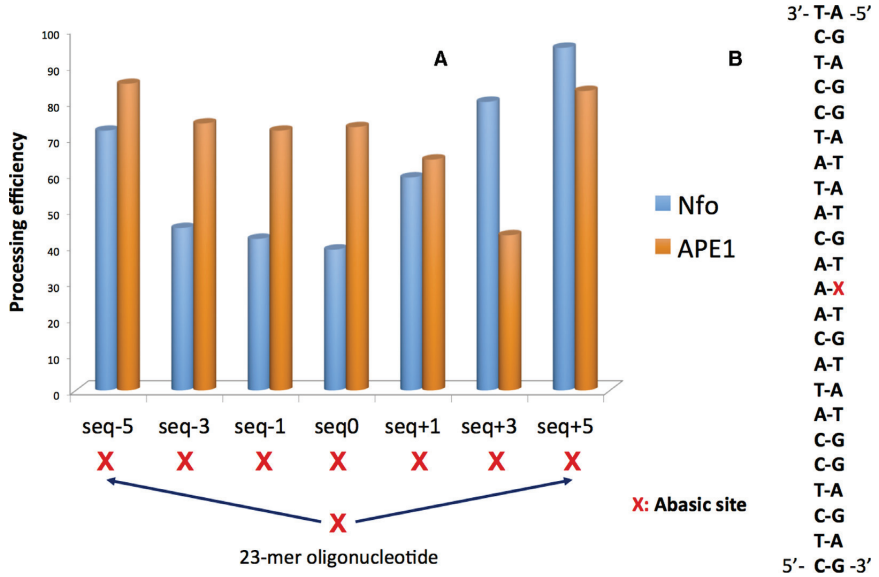


Figure 1. (**A**) Repair rate of the different clustered AP oligomer using Nfo (Escherichia coli) and APE1 (human) enzymes, respectively. (**B**) The original DNA double strand featuring the position of the AP marked as X. The clustered AP containing lesions have been produced inserting a secondary AP site in the complementary strand and up to five residues apart from the original lesion, see also Scheme 1. Data are adapted from references (17,21).

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 1 and Scheme 1. We described hereafter the main structural outcomes obtained for these oligonucleotides. Particular attention will also be given to the global bending and extrahelical position of the AP sites as reported in Table 1.

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 2C). 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 (42–45). 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…T13 and A34…A36 stacking (Figure 2), i.e. an energetic penalty corresponding to ∼10 kcal/mol. 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 ± 18.2° over the last 40 ns, which is only slightly more pronounced than the single-AP-containing control sequence (25.3 ± 11.8°).

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 3) with the two AP sites now occupying only scarcely extrahelical positions (∼3%, see Table 1). Figure 3C 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 ∼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 π-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′…C1′ distances of 11.2 ± 1.3 Å, versus 18.2 ± 0.5 Å in absence of lesions). The geometrical 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 (see Supplementary Figure S7). 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 4). 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 4). One identifies a stacking between A12, A34 and G14, 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 ± 14.4°, 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 5 (bottom). This implies a moderate constriction of the B-helix reflected by C1′…C1′ distances up to ∼16.3 ± 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 central base pairs are left unaltered in their Watson–Crick pairing. Interestingly the structure is slightly more bended than the reference sequence by more than 8°.

The structural evolution of the 23-bp duplex **seq-3** (Figure 5 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 be-

Table 1. Average structural descriptors for the damaged oligonucleotides over 400 ns and the latest 40 ns along MD simulations. Please note that the global parameter such has 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 ± 18.7°	25.2 ± 13.4°	AP7: 12.4% ; AP35: 12.5 %
seq-3	38.3 ± 19.3°	24.9 ± 19.3°	A1P12: 36.6% ; AP35: 27.7 %
seq-1	31.1 ± 16.5°	23.6 ± 14.1°	AP11: 2.5% ; AP35: 3.9 %
seq0	34.3 ± 17.8°	35.5 ± 18.2°	AP12 and AP35: 81.2 %
seq+1	27.0 ± 14.4°	27.4 ± 13.5°	AP13: 2.1% ; AP35: 4.2 %
seq+3	35.3 ± 19.4°	33.7 ± 18.1°	AP15: 13.2% ; AP35: 3.9 %
seq+5	31.5 ± 16.5°	27.3 ± 13.6°	AP17: 30.3% ; AP35: 28.0 %
Reference	32.3 ± 17.2°	25.3 ± 11.8°	AP35: 35.9 %

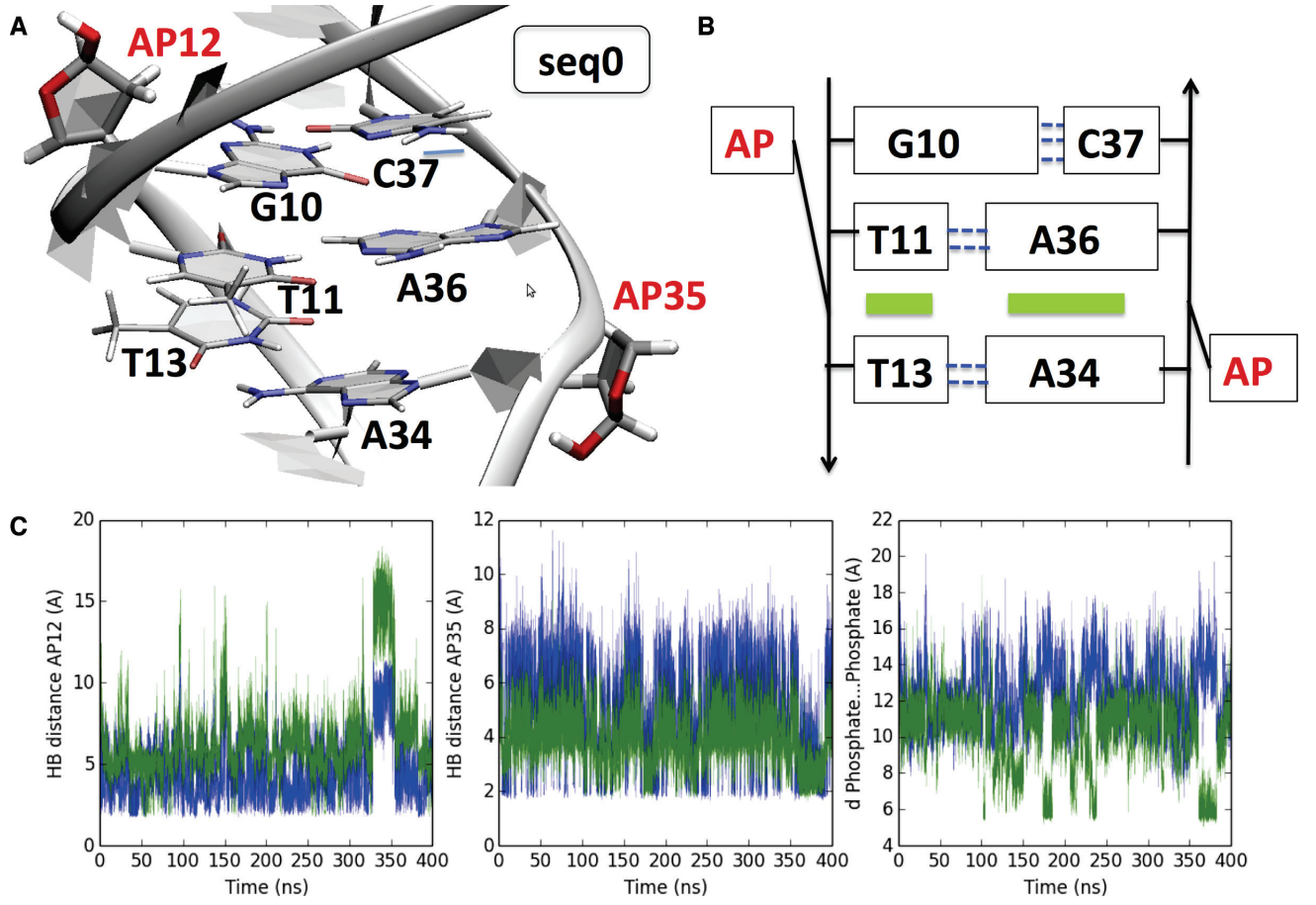


Figure 2. (**A**) Cartoon representation of the final structure obtained after 400 ns for seq0 and (**B**) scheme of the non-covalent interactions reshaping the oligonucleotide: green boxes indicate new π-stacking, and dashed blue lines indicate Watson–Crick hydrogen bonds. (**C**) Time evolution of 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; see also Supplementary Figure S4 for explicative diagrams.

tween the central non-damaged base pairs (T11…A36 and G10…C37) rapidly experience a disruption at ∼34 and 64 ns ending up to the exclusion of C37 at ∼280 ns and T13 at ∼320 ns. This conformation constitute a very distinctive feature of seq-3 and present 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 π-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°.

Further increasing the separation between the two AP sites leads to a situation where they begin to behave independently and partially loose the correlation between the AP sites. The structure of seq-5 (Figure 6 left panel) follows the rather well established behavior of isolated AP sites although we observe an extra-helical occupation of ∼12.5%, i.e. a noticeably lower extrahelicity than for isolated AP

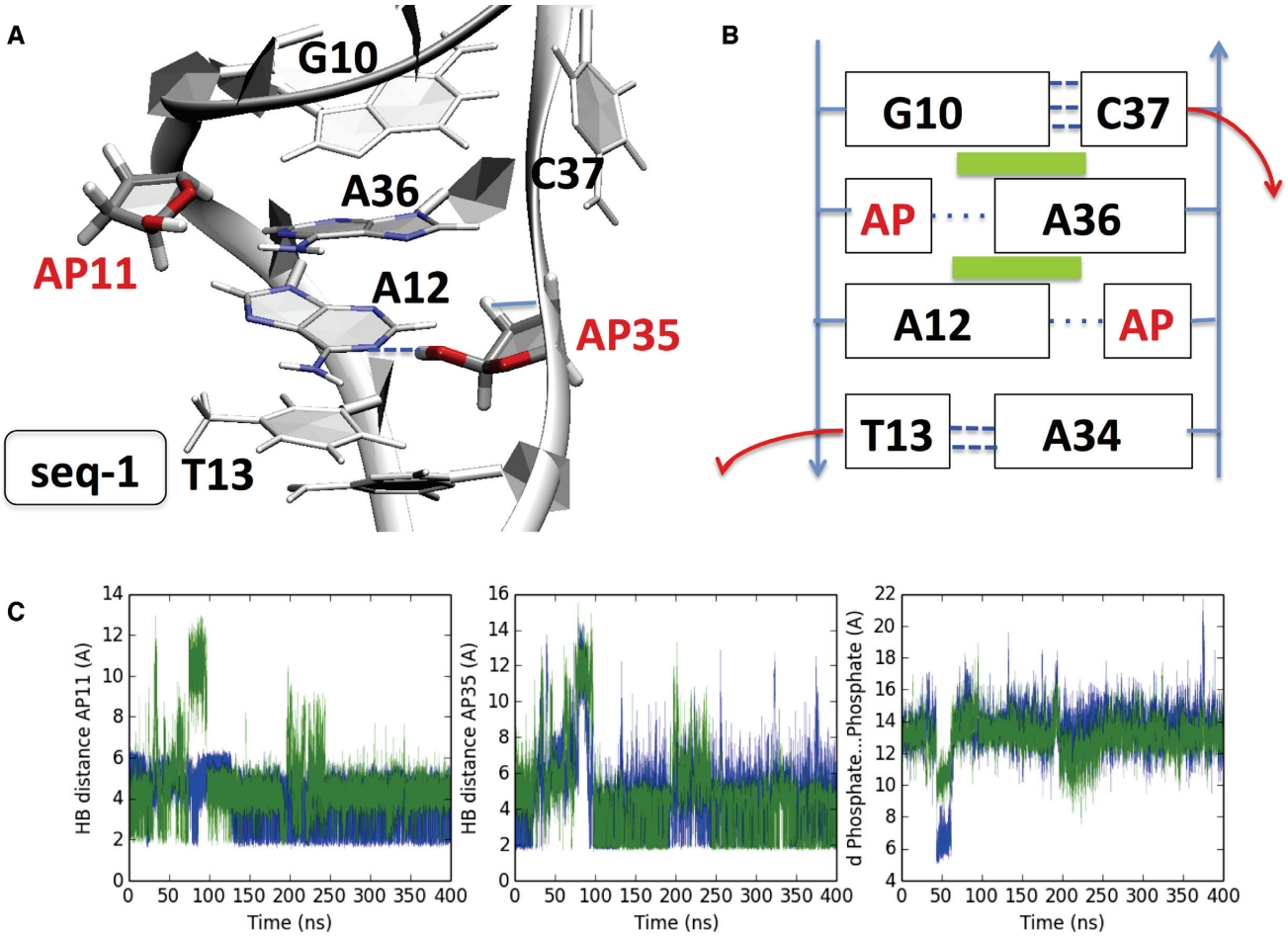


Figure 3. (**A**) Cartoon representation of the final structure obtained after 400 ns for **seq-1**. (**B**) Diagram showing the interaction pattern: Green rectangles indicate transverse π-stacking, and dashed blue lines new hydrogen bonds (thick ones for Watson–Crick pairing). (**C**) Left panel: blue line AP11H1…AP11O5′ distance, green line AP11H1…A36N1. Middle panel: blue line AP35H1…A12N1distance, green line AP35H1…A34N3 distance. Right panel: blue line A12P…C37P distance, green line T13P…A36P; see also Supplementary Figure S5 for explicative diagrams. 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.

sites. The AP sugar on the other hand explores a broad range of sugar puckers angles giving rise to a well established DNA bulge typical of AP strands (42). 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 ± 13.4°) and assumes values close to the ones experienced by single AP sites.

The symmetrical sequence seq+5 (Figure 6 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 ± 13.6°).

DISCUSSION

AP sites can occur either as a result of regular oxidative stress or IR, 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 (46,47). 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-

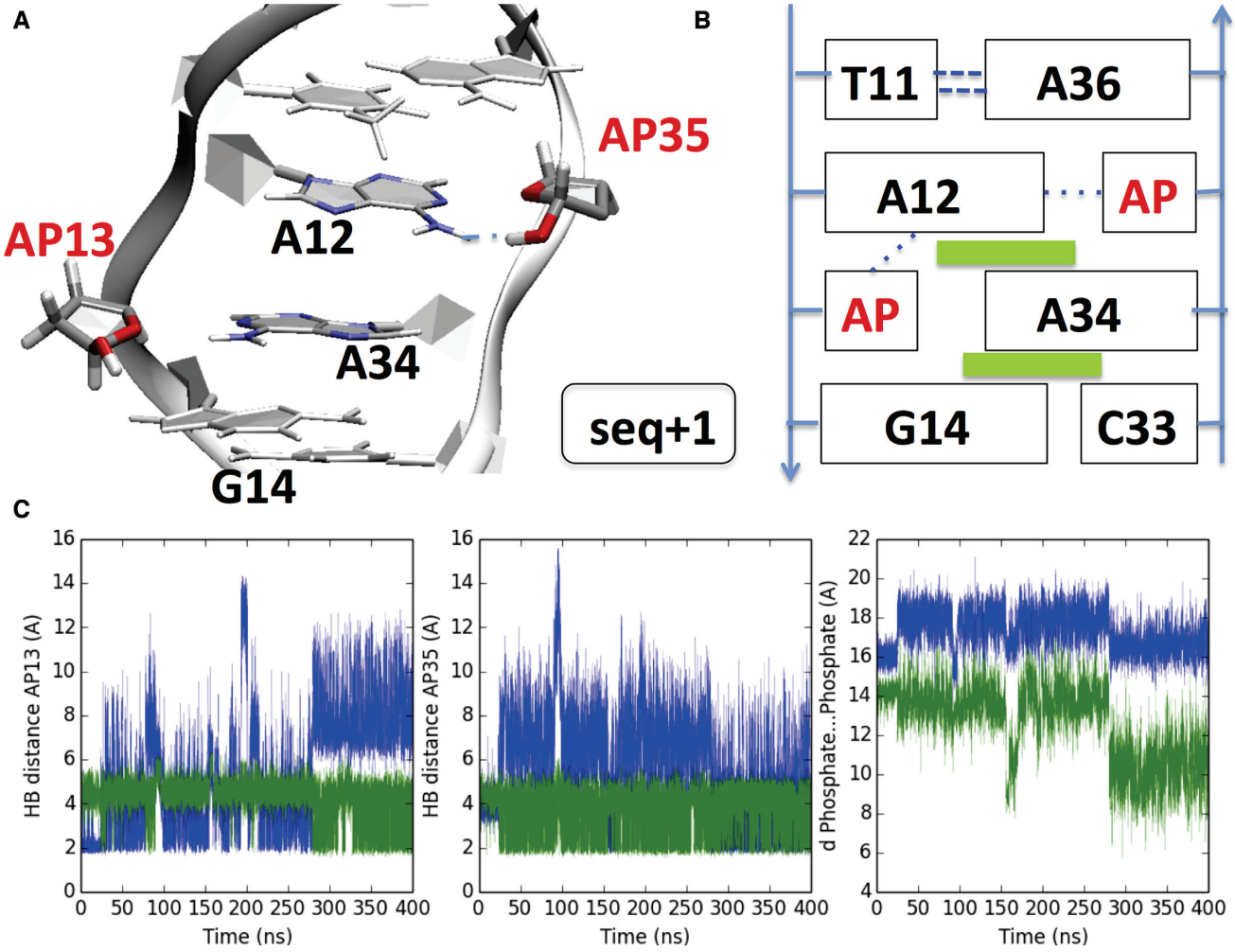


Figure 4. (**A**) Cartoon representation of the final structure obtained after 400 ns for seq+1 and (**B**) diagram illustrating the non-covalent interactions and the position of the bases. Dispersion interaction are represented as green rectangles, while dotted blue lines represent hydrogen bonds. (**C**) Left panel: blue line AP13H1…A12N3 distance, green line AP13H1…AP13:O5′. Middle panel: blue line A12N1…AP35H1 distance, green line AP35O5′…AP35H1distance. Right panel: blue line A12:P – A36:P distance, green line AP13 – AP35; see also Supplementary Figure S6 for explicative diagrams.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.

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. (48–50) 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 (13,46,51–53), thus leading to the so-called 'DSB avoidance' and repair retardation of the lesions belonging to a cluster (9). 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 (54,55). 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°, 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 (56). However

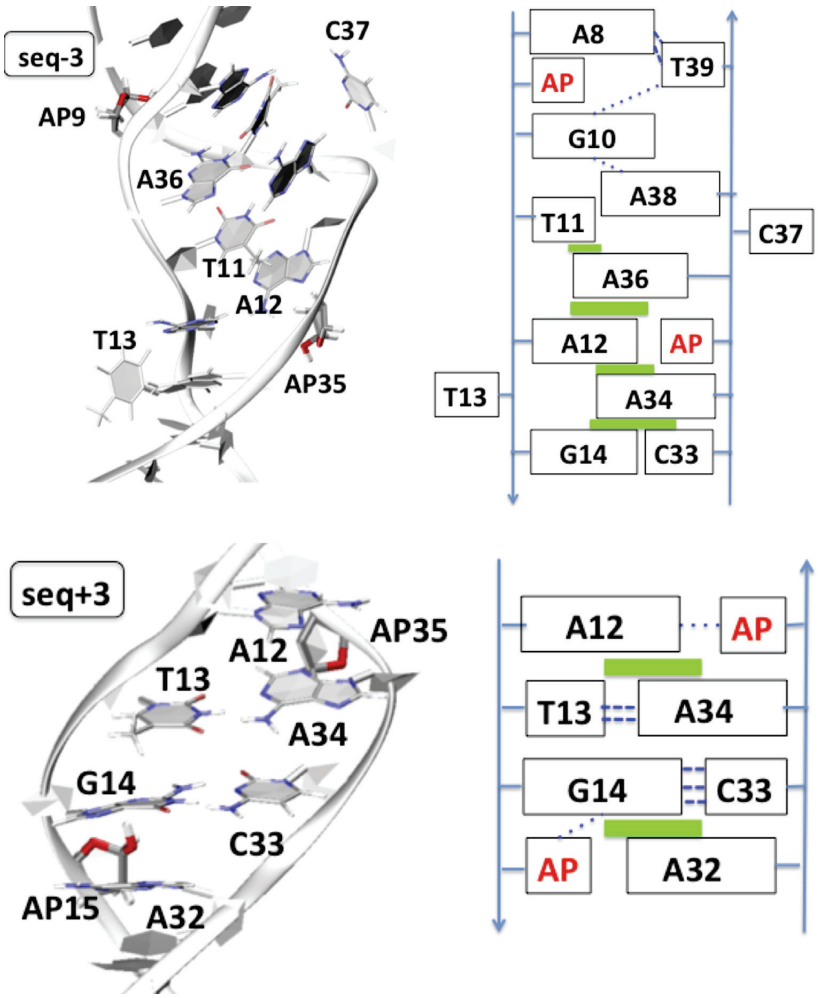


Figure 5. Cartoon representation of the final structure obtained after 400 ns for seq−3 (top) and seq+3 (bottom). A schematic representation is given at the right where green squares indicate π-stacking, and dashed blue lines new hydrogen bonds (thick ones for Watson–Crick pairing).

(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 (17,21,57). 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. Yet, the relation between the repair ratio and the structure is not entirely straightforward. For instance, formation of a complex with a repair enzyme could induce a constraint on DNA (57) that may trigger new conformations of the duplex, whose flexibility may be increased by the clustered AP. Furthermore, clustered AP sites embrace a combinatorial biochemistry, and their complex structural, dynamical and thermodynamical (58,59) 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 1). This suggests different evolutionary pathways (33), which are probably rooted in different interaction patterns between the oligonucleotides and the repair proteins. As revealed by the crystal structures (6), APE1 has an extended contact area between the enzyme and the DNA, furthermore strong electrostatic interactions between positively-charged residues

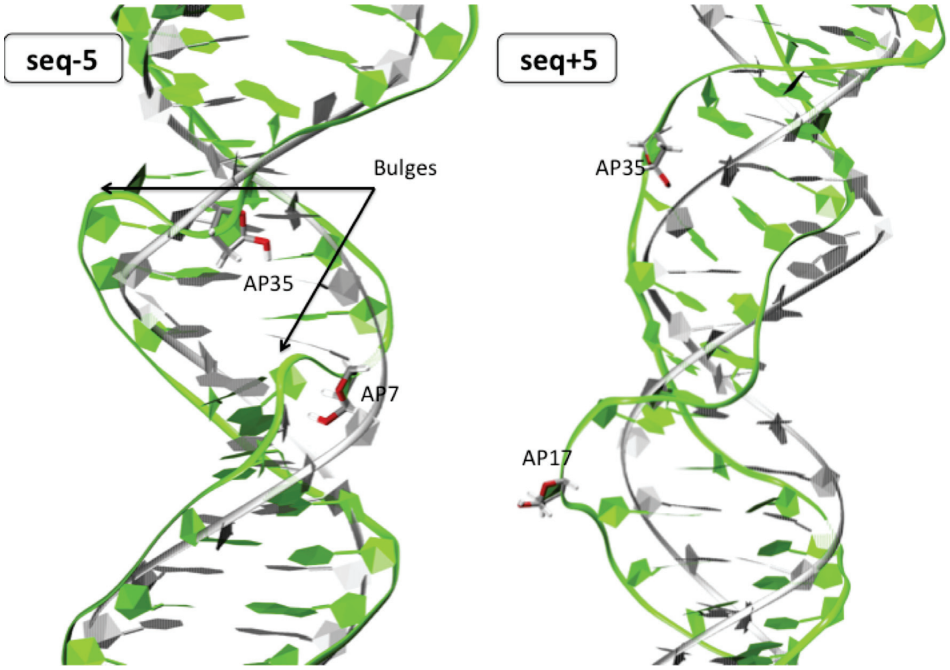


Figure 6. Cartoon representation of seq−5 (left) and seq+5 (right) after 400 ns of classical simulations. Both structures exhibit backbone deformation.

and the negatively-charged DNA backbone induces rather large structural deformations of the helix. This fact was also elegantly revealed by Beloglazova *et al.* (60) 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, this aspect and notably the surprisingly low repair of seq+3 by APE1 will be the object of a forthcoming contribution. On the other hand interaction with Nfo (61) 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 lesioned 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.

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 (28).

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 (4), 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 seq±1 and seq±5 the major groove occupancy predominates, seq+3 has a particular behavior with only one AP site preferentially occupying the minor groove. Consistently with the experimental observation (17, 21), 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 seq+5, since no peculiar structural distortion hamper the recognition by the repair enzyme (Figure 1). Even if a general conclusion can be drawn, it is however important to note, that some sequence effects still exist, and seq-5 has a slightly lower repair ratio, which we relate to a lower percentage of extrahelicity for both AP sites (∼12 %, versus ∼30% for seq+5 and ∼33% for single AP sites).

Seq0 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 π-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, seq0 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 seq±1. However, once again the seq-1 is somehow peculiar since non-damaged nucleobases too assume extrahelical position. This fact, also reported based on NMR structures (4,32,62), 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 seq+1.

However, the most striking position dependence is the one happening between seq+3 and seq−3. Indeed, seq+3 is almost perfectly repaired by Nfo and its efficiency is comparable with 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 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. The repair efficiency increases with the distance separating the residues bearing lesions. For AP sites distant by 5 bp apart to each other, i.e. at a distance larger than the enzyme recognition area (around 17 Å), repair is almost as efficient as for simple lesions. However, in some cases clustered AP position may severely alter the previous trend, or even almost reverse it, so that seq-3 has a dramatic low repair.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Calculations were performed using the local HPC resources of PSMN at ENS-Lyon and at Université de Lorraine – Nancy. The support by the COST Action CM1201 "Biomimetic Radical Chemistry" is gratefully acknowledged. This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX0007) operated by the French National Research Agency (ANR). Dr Georgakilas was supported by an EU grant MC-CIG-303514.

FUNDING

COST Action CM1201 "Biomimetic Radical Chemistry"; this work was performed within the framework of the LABEX PRIMES (ANR-11-LABX0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX0007) operated by the French National Research Agency (ANR); EU grant [MC-CIG-303514 to Dr Georgakilas]. Funding for open access charge: University of Lorraine and CNRS.

Conflict of interest statement. None declared.

REFERENCES

1. Demple,B. and Sung,J.-S. (2005) Molecular and biological roles of Ape1 protein in mammalian base excision repair. *DNA Repair (Amst)*, **4**, 1442–1449.
2. Clauson,C., Schärer,O.D. and Niedernhofer,L. (2013) Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb. Perspect. Biol.*, **5**, a012732.
3. Gates,K.S. (2009) An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol.*, **22**, 1747–1760.
4. Hazel,R.D., Tian,K. and de Los Santos,C. (2008) NMR solution structures of bistranded abasic site lesions in DNA. *Biochemistry*, **47**, 11909–11919.
5. Mol,C.D., Izumi,T., Mitra,S. and Tainer,J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature*, **403**, 451–456.
6. Freudenthal,B.D., Beard,W.A., Cuneo,M.J., Dyrkheeva,N.S. and Wilson,S.H. (2015) Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.*, **22**, 924–931.
7. Watanabe,R., Rahmanian,S. and Nikjoo,H. (2015) Spectrum of radiation-induced clustered non-DSB damage – A Monte Carlo track structure modeling and calculations. *Radiat. Res.*, **183**, 525–540.
8. Nikjoo,H., O'Neill,P., Terrissol,M. and Goodhead,D.T. (1999) Quantitative modelling of DNA damage using Monte Carlo track structure method. *Radiat. Environ. Biophys.*, **38**, 31–38.
9. Georgakilas,A.G., O'Neill,P. and Stewart,R.D. (2013) Induction and repair of clustered DNA lesions: What do we know so far? *Radiat. Res.*, **180**, 100–109.
10. Liu,Z., Ding,S., Kropachev,K., Jia,L., Amin,S., Broyde,S. and Geacintov,N.E. (2015) Resistance to nucleotide excision repair of bulky guanine adducts opposite abasic sites in DNA duplexes and relationships between structure and function. *PLoS One*, **10**, e0137124.
11. Nikitaki,Z., Hellweg,C.E., Georkakilas,A.G. and Ravanat,J.L. (2015) Stress-induced DNA damage biomarkers: Applications and limitations. *Front. Chem.*, **3**, 35.
12. Shikazono,N., Akamatsu,K., Takahashi,M., Noguchi,M., Urushibara,A., O'Neill,P. and Yokoya,A. (2013) Significance of DNA polymerase I in in vivo processing of clustered DNA damage. *Mutat. Res. Mol. Mech. Mutagen.*, **749**, 9–15.
13. Lomax,M.E., Cunniffe,S. and O'Neill,P. (2004) Efficiency of repair of an abasic site within DNA clustered damage sites by mammalian cell nuclear extracts. *Biochemistry*, **43**, 11017–11026.
14. Sutherland,B.M., Bennett,P.V., Sidorkina,O. and Laval,J. (2000) Clustered DNA damages induced in isolated DNA and in human cells by low doses of ionizing radiation. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 103–108.
15. Chastain,P.D., Nakamura,J., Rao,S., Chu,H., Ibrahim,J.G., Swenberg,J.A. and Kaufman,D.G. (2010) Abasic sites preferentially

form at regions undergoing DNA replication. *FASEB J.*, **24**, 3674–3680.
16. Chastain,P.D., Nakamura,J., Swenberg,J. and Kaufman,D. (2006) Nonrandom AP site distribution in highly proliferative cells. *FASEB J.*, **20**, 2612–2614.
17. Georgakilas,A.G., Bennett,P.V. and Sutherland,B.M. (2002) High efficiency detection of bi-stranded abasic clusters in gamma-irradiated DNA by putrescine. *Nucleic Acids Res.*, **30**, 2800–2808.
18. Singh,V. and Das,P. (2013) Condensation of DNA–a putative obstruction for repair process in abasic clustered DNA damage. *DNA Repair (Amst)*, **12**, 450–457.
19. Akamatsu,K., Shikazono,N. and Saito,T. (2015) Localization estimation of ionizing radiation-induced abasic sites in DNA in the solid state using fluorescence resonance energy transfer. *Radiat. Res.*, **183**, 105–113.
20. Eccles,L.J., Menoni,H., Angelov,D., Lomax,M.E. and O'Neill,P. (2015) Efficient cleavage of single and clustered AP site lesions within mono-nucleosome templates by CHO-K1 nuclear extract contrasts with retardation of incision by purified APE1. *DNA Repair (Amst)*, **35**, 27–36.
21. Georgakilas,A.G., Bennett,P.V., Wilson,D.M. 3rd and Sutherland,B.M. (2004) Processing of bistranded abasic DNA clusters in γ-irradiated human hematopoietic cells. *Nucleic Acids Res.*, **32**, 5609–5620.
22. Pérez,A., Luque,F.J. and Orozco,M. (2012) Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.*, **45**, 196–205.
23. Dumont,E. and Monari,A. (2015) Understanding DNA under oxidative stress and sensitization: The role of molecular modeling. *Front. Chem.*, **3**, 43.
24. Ayadi,L., Coulombeau,C. and Lavery,R. (2000) The impact of abasic sites on DNA flexibility. *J. Biomol. Struct. Dyn.*, **17**, 645–653.
25. Ayadi,L., Coulombeau,C. and Lavery,R. (1999) Abasic sites in duplex DNA: Molecular modeling of sequence-dependent effects on conformation. *Biophys J.*, **77**, 3218–3226.
26. Barsky,D., Foloppe,N., Ahmadia,S., Wilson,D.M. 3rd and MacKerell,A.D. Jr. (2000) New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.*, **28**, 2613–2626.
27. Mu,H., Kropachev,K., Wang,L., Zhang,L., Kolbanovskiy,A., Kolbanovskiy,M., Geacintov,N.E. and Broyde,S. (2012) Nucleotide excision repair of 2-acetylaminofluorene- and 2-aminofluorene-(C8)-guanine adducts: Molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies. *Nucleic Acids Res.*, **40**, 9675–9690.
28. Churchill,C.D.M., Eriksson,L.A. and Wetmore,S.D. (2016) DNA distortion caused by uracil-containing intrastrand cross-links. *J. Phys. Chem. B*, **120**, 1195–1204.
29. Bignon,E., Gattuso,H., Morell,C., Dumont,E. and Monari,A. (2015) DNA photosensitization by an insider: Photophysics and triplet energy transfer of 5-Methyl-2-pyrimidone deoxyribonucleoside. *Chemistry*, **21**, 11509–11516.
30. Faraji,S. and Dreuw,A. (2014) Physicochemical mechanism of light-driven DNA repair by (6-4) photolyases. *Annu. Rev. Phys. Chem.*, **65**, 275–292.
31. Fujimoto,H., Pinak,M., Nemoto,T., O'Neill,P., Kume,E., Saito,K. and Maekawa,H. (2005) Molecular dynamics simulation of clustered DNA damage sites containing 8-oxoguanine and abasic site. *J. Comput. Chem.*, **26**, 788–798.
32. Zálešák,J., Lourdin,M., Krejč,L., Constant,J.F. and Jourdan,M. (2014) Structure and dynamics of DNA duplexes containing a cluster of mutagenic 8-oxoguanine and abasic site lesions. *J. Mol. Biol.*, **426**, 1524–1538.
33. Lu,D., Silhan,J. and MacDonald,J.T. (2012) Structural basis for the recognition and cleavage of abasic DNA in Neisseria meningitidis. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16852–16857.
34. Case,D.A., Berryman,J.T., Betz,R.M., Cerutti,D.S., Chatham,T.E. III, Darden,T.A., Duke,R.E., Giese,T.J., Gohlke,H., Goetz,A.W. *et al.* (2015) AMBER.
35. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
36. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.
37. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E. 3rd, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma Conformers. *Biophys. J.*, **92**, 3817–3829.
38. Wang,J., Cieplak,P. and Kollman,P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049–1074.
39. Mark,P. and Nilsson,L. (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A*, **105**, 9954–9960.
40. Lavery,R., Moakher,M. and Maddocks,J.H. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
41. de los Santos,C., El-khateeb,M., Rege,P., Tian,K. and Johnson,F. (2004) Impact of the C1′ configuration of abasic sites on DNA duplex structure. *Biochemistry*, **43**, 15349–15357.
42. Zeglis,B.M., Boland,J.A. and Barton,J.K. (2009) Recognition of abasic sites and single base bulges in DNA by a metalloinsertor. *Biochemistry*, **48**, 839–849.
43. Shi,X., Beauchamp,K.A., Harbury,P.B. and Herschlag,D. (2014) From a structural average to the conformational ensemble of a DNA bulge. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E1473–E1480.
44. Li,M., Völker,J., Breslauer,K.J. and Wilson,D.M. 3rd (2014) APE1 incision activity at abasic sites in tandem repeat sequences. *J. Mol. Biol.*, **426**, 2183–2198.
45. Chiba,J., Aoki,S. and Yamamoto,J. (2014) Deformable nature of various damaged DNA duplexes estimated by an electrochemical analysis on electrodes. *Chem. Commun. (Camb)*, **50**, 11126–11128.
46. Cunniffe,S., O'Neill,P., Greenberg,M.M. and Lomax,M.E. (2014) Reduced repair capacity of a DNA clustered damage site comprised of 8-oxo-7,8-dihydro-2′-deoxyguanosine and 2-deoxyribonolactone results in an increased mutagenic potential of these lesions. *Mutat. Res. Mol. Mech. Mutagen.*, **762**, 32–39.
47. Kunkel,T.A., Schaaper,R.M. and Loeb,L.A. (1983) Depurination-induced infidelity of deoxyribonucleic acid synthesis with purified deoxyribonucleic acid replication proteins in vitro. *Biochemistry*, **22**, 2378–2384.
48. Lomax,M.E., Folkes,L.K. and O'Neill,P. (2013) Biological consequences of radiation-induced DNA damage: Relevance to radiotherapy. *Clin. Oncol. (R. Coll. Radiol.)*, **25**, 578–585.
49. Sage,E. and Harrison,L. (2011) Clustered DNA lesion repair in eukaryotes: Relevance to mutagenesis and cell survival. *Mutat. Res. Mol. Mech. Mutagen.*, **711**, 123–133.
50. Eccles,L.J., O'Neill,P. and Lomax,M.E. (2011) Delayed repair of radiation induced clustered DNA damage: Friend or foe? *Mutat. Res. Mol. Mech. Mutagen.*, **711**, 134–141.
51. Malyarchuk,S., Castore,R. and Harrison,L. (2009) Apex1 can cleave complex clustered DNA lesions in cells. *DNA Repair (Amst)*, **8**, 1343–1354.
52. Eccles,L.J., Lomax,M.E. and O'Neill,P. (2010) Hierarchy of lesion processing governs the repair, double-strand break formation and mutability of three-lesion clustered DNA damage. *Nucleic Acids Res.*, **38**, 1123–1134.
53. Eot-Houllier,G., Eon-Marchais,S., Gasparutto,D. and Sage,E. (2005) Processing of a complex multiply damaged DNA site by human cell extracts and purified repair proteins. *Nucleic Acids Res.*, **33**, 260–271.
54. Sedletska,Y., Radicella,J.P. and Sage,E. (2013) Replication fork collapse is a major cause of the high mutation frequency at three-base lesion clusters. *Nucleic Acids Res.*, **41**, 9339–9348.
55. Kozmin,S.G., Sedletska,Y., Reynaud-Angelin,A., Gasparutto,D. and Sage,E. (2009) The formation of double-strand breaks at multiply damaged sites is driven by the kinetics of excision/incision at base damage in eukaryotic cells. *Nucleic Acids Res.*, **37**, 1767–1777.
56. Zalesak,J., Constant,J.-F. and Jourdan,M. (2016) NMR solution structure of DNA featuring clustered 2′-deoxyribonolactone and 8-oxoguanine lesions. *Biochemistry*, **55**, 3899–3906.
57. Hosfield,D.J., Guan,Y., Haas,B.J., Cunningham,R.P. and Tainer,J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*, **98**, 397–408.

58. Ghosh,S. and Greenberg,M.M. (2015) Correlation of thermal stability and structural distortion of DNA interstrand cross-links produced from oxidized abasic sites with their selective formation and repair. *Biochemistry*, **54**, 6274–6283.
59. Gelfand,C.A., Plum,G.E., Grollman,A.P., Johnson,F. and Breslauer,K.J. (1998) Thermodynamic consequences of an abasic lesion in duplex DNA are strongly dependent on base sequence. *Biochemistry*, **37**, 7321–7327.
60. Beloglazova,N.G., Kirpota,O.O., Starostin,K.V., Ishchenko,A.A., Yamkovoy,V.I., Zharkov,D.O., Douglas,K.T. and Nevinsky,G.A. (2004) Thermodynamic, kinetic and structural basis for recognition and repair of abasic sites in DNA by apurinic/apyrimidinic endonuclease from human placenta. *Nucleic Acids Res.*, **32**, 5134–5146.
61. Mazouzi,A., Vigouroux,A., Aikeshev,B., Brooks,P.J., Saparbaev,M.K., Morera,S. and Ishchenko,A.A. (2013) Insight into mechanisms of 3′-5′ exonuclease activity and removal of bulky 8,5′-cyclopurine adducts by apurinic/apyrimidinic endonucleases. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E3071–E3080.
62. Lin,Z., Hung,K.N., Grollman,A.P. and de los Santos,C. (1998) Solution structure of duplex DNA containing an extrahelical abasic site analog determined by NMR spectroscopy and molecular dynamics. *Nucleic Acids Res.*, **26**, 2385–2391.

Nucleic Acids Research Advance Access published December 15, 2016

doi: 10.1093/nar/gkw1253

Interstrand cross-linking implies contrasting structural consequences for DNA: insights from molecular dynamics

Emmanuelle Bignon[1,2], Tomáš Dršata[3], Christophe Morell[1], Filip Lankaš[3,4,*] and Elise Dumont[2,*]

[1]Univ Lyon, CNRS, Université Claude Bernard Lyon 1, Ens de Lyon, Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, F-69100 Villeurbanne, France, [2]Univ Lyon, Ens de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F69342, Lyon, France, [3]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Praha 6, Czech Republic and [4]Laboratory of Informatics and Chemistry, University of Chemistry and Technology Prague, Technická 5, 166 28 Praha 6, Czech Republic

Received September 06, 2016; Revised November 18, 2016; Editorial Decision November 28, 2016; Accepted December 05, 2016

ABSTRACT

Oxidatively-generated interstrand cross-links rank among the most deleterious DNA lesions. They originate from abasic sites, whose aldehyde group can form a covalent adduct after condensation with the exocyclic amino group of purines, sometimes with remarkably high yields. We use explicit solvent molecular dynamics simulations to unravel the structures and mechanical properties of two DNA sequences containing an interstrand cross-link. Our simulations palliate the absence of experimental structural and stiffness information for such DNA lesions and provide an unprecedented insight into the DNA embedding of lesions that represent a major challenge for DNA replication, transcription and gene regulation by preventing strand separation. Our results based on quantum chemical calculations also suggest that the embedding of the ICL within the duplex can tune the reaction profile, and hence can be responsible for the high difference in yields of formation.

INTRODUCTION

The large variety of DNA defects that can be formed upon initial attack of reactive oxygen and nitrogen species is indicative of a rich and complex underlying chemistry (1). Amongst the ∼80 oxidatively-generated lesions that have been characterized so far, abasic (Ap) sites and 8-oxoguanine are the prominent single-nucleotide defects, but more complex lesions involving two nucleobases that become covalently tethered have also been identified (2) and are widely investigated (3–11). One possible degradation pathway implies a subsequent reactivity of Ap, which can condensate with the exocyclic amino group of purines. This generates interstrand cross-link (ICL) lesions, which have been firmly identified over the last decade (7,8,12,13): this proves in passing that the B-DNA helix is flexible enough to host the formation of a carbon–nitrogen linkage anchoring the two strands. A single ICL defect can trigger unyielding obstruction to DNA replication, because the strand separation is blocked, and can cause cellular apoptosis (14). Price *et al.* have recently reported (7,8) remarkably high yields (15–70%) for the formation of such defects upon attack of the N^6-amino group of adenine (Ap-dA or CL(A)). They also evidenced a more rare Ap-dG (or CL(G)) defect (ca. 2–3%), involving the exocyclic N^2-amino group of guanine (see Figure 1). It is particularly intriguing that such lesions were found to proceed selectively with a purine opposite a 3′-adjacent nucleotide, rather than directly with the orphan adenine (dA32). It was suggested by the authors that the formation of ICLs reflects the ease with which the two partners can approach under the mechanical constraint of a B-helical environment (7). But very recently Greenberg *et al*. have challenged the idea that thermal destabilization can determine the various yields of formation of these highly mutagenetic lesions (15). Covalently tethering the two strands inevitably implies a loss of B-helicity via an alteration of Watson–Crick hydrogen bond (HB) network, possibly ranging over several base pairs, and a redefinition of stacking interactions. The field suffers from the absence of structural data that are crucial to interpret the lack of repair.

Indeed, approaches using synthesized oligonucleotides and nucleotides (16) are not compatible with the produc-

*To whom correspondence should be addressed. Tel: +33 4 72 72 88 46; Fax: +33 4 72 72 88 60; Email: elise.dumont@ens-lyon.fr
Correspondence may also be addressed to Filip Lankaš. Tel: +420 220 44 4392; Fax: +420 220 410 320; Email: filip.lankas@vscht.cz

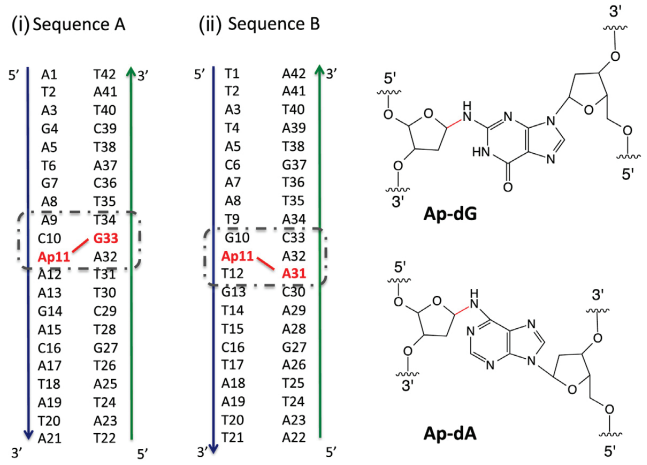


Figure 1. The 21-bp DNA sequences A and B featuring an interstrand cross-link between Ap and a purine (displayed with a red line). The yield of formation was experimentally reported to be of 2–3% versus 15–70% for sequences A and B, respectively (7).

tion of milligram quantities or high-quality crystals. Only the chemical structure for the isolated lesion, obtained after enzymatic digestion, can be proposed after careful inspection based on tandem mass spectrometry coupled to high pressure (or high performance) liquid chromatography (13) and/or nuclear magnetic resonance (8). One irremediably looses the structural information concerning the way the ICL redefines the oligonucleotide. A crystal structure of a nucleoside model for the Ap-dG cross-link was recently reported (17), but experimental structures of an ICL within B-DNA are not known.

Moreover, any experimental information concerning the mechanical flexibility or stiffness of these lesions is missing. The peculiar lesion mechanics may be decisive in recognizing the damage for repair. The complex repair mechanisms of ICLs are a subject of intense research (18–20). To palliate the lack of experimental information, molecular modeling can prove useful by analogy to what has been proposed for DNA-drug adducts (21,22) but also for covalently-tethered intrastrand lesions (23–25). In this work, relying on a computational approach, we resolve the structure of two 21-bp oligonucleotides featuring an ICL. We then derive the mechanical properties induced by the cross-link and we determine the stabilization or destabilization of the ICL within the B-DNA local environment. The stabilization energies and entropies can be related to the very different yields of formation of Ap-dA and Ap-dG as the ICL induction obeys a thermodynamic control.

MATERIALS AND METHODS

We focused on two 21 base pair (bp) duplex DNA sequences, A and B (Figure 1), investigated experimentally by Gates *et al*. (7). All classical molecular dynamics (MD) simulations were performed using the Amber 12 suite of programs (26). We started with reference DNA duplexes featuring an abasic site at the 11th nucleobase position (Ap11). Since no experimental structure has been reported for the ICL-containing duplex DNA, the two duplexes were built using the `nab` module of Amber, with conformational parameters characteristic of B-DNA. The parmbsc0 force field was used for DNA, the force field parameters for the Ap site were generated using Antechamber and atomic point charges were assigned using the RESP procedure. They are comparable to parameters similarly derived in the literature (27–29), and we verified that a similar behavior is obtained with the parmbsc1(30) force field, which is the current reference and most tested set of parameters for undamaged DNA (31).

Each duplex was neutralized with 40 potassium ions using Dang parameters, and solvated with TIP3P water molecules in an octahedric simulation box with a 10 Å buffer. After 10 000 steps of energy minimization including 5000 steps of steepest descent and 5000 steps of conjugate gradient, the temperature was increased to 300 K in a 30 ps thermalization run. During the rest of the simulation, the temperature was kept constant using the Langevin thermostat with a collision frequency γ_{ln} of 1 ps^{-1}. Then, each system was equilibrated during 1 ns in NPT conditions. Structures obtained after a 200 ns production run provide starting geometries bringing close together the Ap site and the offset purine forming the ICL identified by Gates *et al*. MD simulations were also performed with the open form of the abasic site over 200 ns, which albeit more rare is responsible for the ICL formation alongside a multi-step pathway (32).

As the C1'…N distance gets close to 4–5 Å, it becomes possible to enforce a covalent linkage between Ap11 and G33 or A31 by introducing a harmonic force constant between these two atoms (atom types were changed accordingly still using the parmbsc0 force field, see Supplementary Data). Hybrid quantum mechanics/molecular dynamics (QM/MD), where the ICL is described at the semi-empirical AM1 level, were performed over 1 ns to further validate the classical parameters that were assigned. No strong deviation was observed between the structures obtained by hybrid and unrestrained 500 ns MM-MD simulations (see Supplementary Data). Also the distribution of interatomic distances is comparable at the MM and semi-empirical level. From the final MM-MD structures, a fragment encompassing the ICL lesion and four vicinal nucleobases (as depicted in round dashed box in Figure 1) was extracted, and energies were evaluated within the framework of density functional theory (DFT) at the BLYP-D3BJ/6-31G(d) level of theory. This level of theory has been calibrated on DNA systems featuring non-covalent interactions (33) and in particular the use of a hybrid functional such as B3LYP leads to the formation of spurious hydrogen bonds for the isolated ICL-containing cluster (34). These DFT calculations were performed using the Gaussian09 suite of programs (35). As a control, two B-DNA duplexes analogous to sequences A and B but with no lesion were built, equilibrated and simulated for 200 ns as described above. Representative snapshots were found using the `cpptraj` module of AMBER, the rigid base and basepair coordinates were computed using the 3DNA program (36).

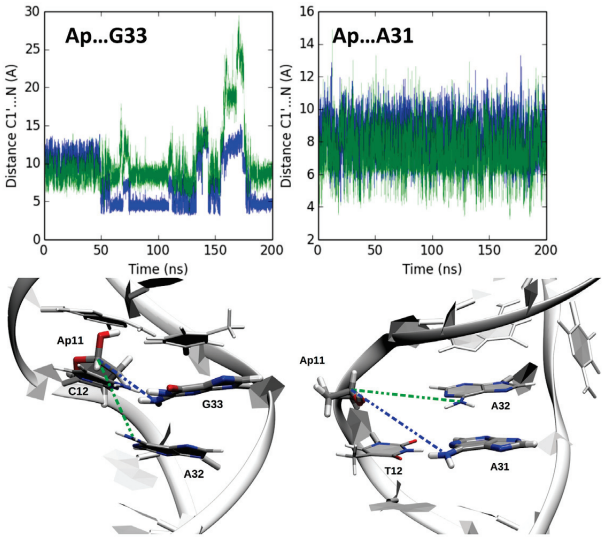


Figure 2. Cartoon representations for the Ap-containing duplexes before formation of the interstrand cross-link (ICL). For Sequence A, the structures are fully consistent with an attack with an offset, the approaching distances *d*(C1'...N) being close to 4–5 Å. The green line corresponds to the interatomic distance between Ap/C1' and A32/N6: the blue line refers to the distances between Ap/C1' and G33/N2 or A31/N6.

RESULTS

Simulations of the Ap-containing duplexes

The simulations of the two Ap-containing sequences allow to monitor the structural changes that take place when a covalent linkage is about to form between the Ap site and the base (G33 or A31, respectively, Figure 1), providing insight into the reactant structures. The structures are flexible enough to undergo structural adaptation, with formation of new hydrogen bond(s) replacing the disrupted Watson–Crick pairing and deviation of the backbone. Figure 2 shows cartoon representations of the damaged sites and time evolutions of atomic distances, indicating that the Ap site seeks to fill up the empty space left by the loss of thymine (27–29). For the oligonucleotide A (left), it can be seen that the distance between the carbon C1' of the AP site and the nitrogen of the amino group of G33 is 4.5 ± 0.3 Å that contrasts to the distance of 8.4 ± 0.5 Å to A32 (blue and green lines, respectively). Our simulations thus corroborate a higher propensity for the ICL to form with a nucleobase adjacent to the orphan base, as reported experimentally (7). For the oligonucleotide B (right), one can see that the Ap site approaches A32 and A31 up to 4–5 Å, yet more transiently, as the distance fluctuates on average around 7.4 and 8.0 Å (green and blue lines, respectively). This behavior holds when the transient open form of the Ap site is considered (see Supplementary Data). In what follows we carefully inspect the structural reorganization and flexibility of the ICL-containing oligomers.

Simulations of the ICL-containing duplexes

Figure 3 shows atomic structures of parts of Sequences A and B surrounding the lesion, with the covalent linkage already present. The structures were obtained as representative snapshots, in the sense of minimal sum of RMSD, from the MD ensembles of the ICL-containing duplexes (the pdb files are in Supplementary Data). Schematic representations and time evolution of selected hydrogen bonds are also shown.

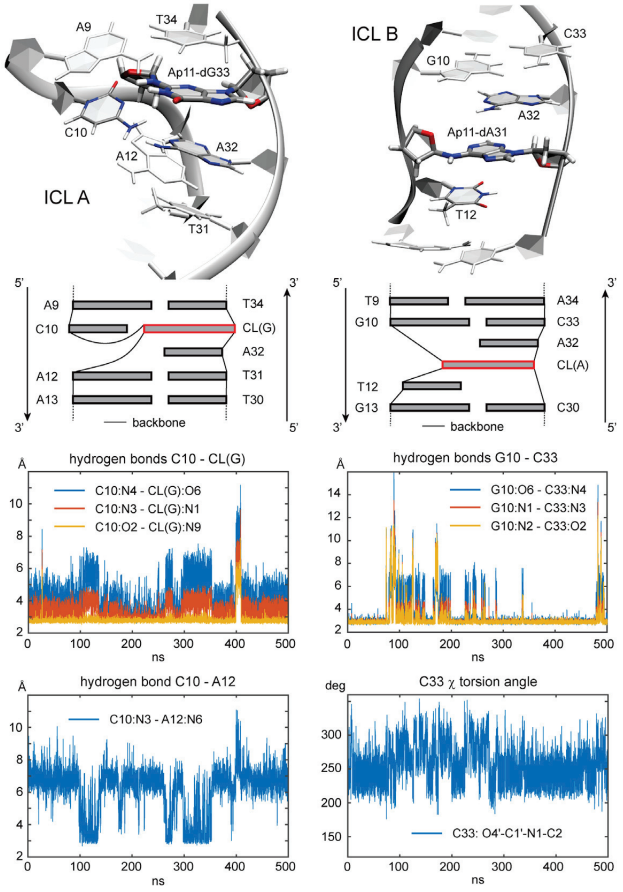


Figure 3. Cartoon representations and interatomic distances of ICL-containing duplexes A and B. The structures (top) are representative snapshots from the molecular dynamics (MD) conformational ensemble (the pdb files are in Supplementary Data). The figure exposes contrasting structural and dynamical features of the two lesions.

It is seen that for Sequence A (Figure 3 left), the cross-linked site CL(G) forms three Watson–Crick HBs to the C10 base in the opposite strand for most of the time. However, the system occassionally undergoes a conformational switch to an alternative substate in which the C10-CL(G) hydrogen bonding is lost and C10 forms one HB to the A12 base in the same strand instead (around 100 ns and 300 ns). Besides that, there is one short time interval in which C10 loses all hydrogen bonding and tends to flip out of the helix (around 400 ns). We exclude this major, but isolated and short lived perturbation from all subsequent analysis. As for Sequence B (Figure 3 right), there is only one, somewhat unstable Watson–Crick HB triplet, namely that between G10 and C33. Thus, the HB patterns in the two lesions are quite different. The same applies to the stacking patterns: while in Sequence A there is just one unpaired

base (A32) sandwiched between the C10-CL(G) and A12-T31 pairs, the CL(A) in Sequence B forms a stack together with the flanking unpaired bases T12 and A32.

These local differences are reflected in different overall structures and flexibilities of the two damaged sites. In order to quantify them, we define a virtual basepair step consisting of base pairs A9-T34 and A13-T30 for Sequence A, and the analogous pairs T9-A34 and G13-C30 for Sequence B (Figure 3). They are intact WC pairs flanking the lesion. The structure of the lesion as a whole is then defined by the six standard local basepair step coordinates (tilt, roll, twist, shift, slide and rise) associated with the virtual step. By analyzing each snapshot of the MD trajectory using the 3DNA software (36), we obtain time series of the step coordinates. The histograms of the coordinates (Figures S7 and S8 in Supplementary Data) indicate nearly Gaussian distributions. This may be surprising given the non-trivial local conformational dynamics involving two substates in Sequence A and an unstable base pair in Sequence B (Figure 3). A possible explanation is that the global structure results from multiple contributions, both harmonic and anharmonic that, thanks to the Central Limit Theorem, add up to an effective harmonic behavior.

The nearly harmonic probability distribution of the coordinates enables us to apply the model of equilibrium shape and harmonic stiffness (37) to the virtual step. The step coordinates, assembled in a six-dimensional vector $\mathbf{w}$, are associated with the effective harmonic deformation energy U of the form

$$U = \frac{1}{2}(\mathbf{w} - \widehat{\mathbf{w}}) \cdot \mathbf{K}(\mathbf{w} - \widehat{\mathbf{w}}). \quad (1)$$

The model contains two sets of parameters, namely the equilibrium coordinate values assembled in vector $\widehat{\mathbf{w}}$ and the symmetric, positive definite stiffness matrix $\mathbf{K}$. These parameters are related to the moments of the coordinate distribution as

$$\widehat{\mathbf{w}} = \langle \mathbf{w} \rangle, \ \mathbf{K} = k_B T \mathbf{C}^{-1} \quad (2)$$

where $\langle \mathbf{w} \rangle$ is the coordinate mean (first moment) and $\mathbf{C}$ is the coordinate covariance matrix (second moment). In our analysis, we estimate the moments using the coordinate time series obtained from the unrestrained MD simulations.

Table 1 shows the equilibrium coordinates of the virtual step in the two damaged sequences together with the same virtual step taken from the control, undamaged DNA duplexes. The structure of ICL A exhibits increased local bending (a tilt of ca. -21°) but is otherwise quite similar to the control. In contrast, ICL B affects bending only slightly but shows other important structural deviations from the control. In particular, it is undertwisted by 20°, its lateral displacement of the base pairs (shift and especially slide) change and the vertical displacement (rise) increases by >4 Å, reflecting the peculiar stacking pattern (Figure 3).

The effective mechanical flexibility as defined by the stiffness matrix $\mathbf{K}$ is also different for the two lesions. Each stiffness matrix contains 21 independent entries. A convenient way to capture the overall stiffness by one single number is to compute the entropy per one basepair step coordinate s_c

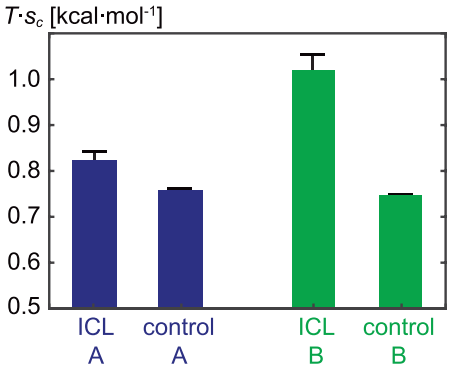


Figure 4. Entropy per coordinate of the virtual basepair step comprising the lesion, computed from the covariance of the rigid basepair coordinates (Equation (3)). Error bars are mean differences between values for the whole trajectory and for its halves. While the duplex involving the Ap-dG cross-link (ICL A) remains nearly as rigid as a regular B-DNA duplex, the Ap-dG cross-link (ICL B) introduces a flexible hotspot into the helix. More detailed entropy models (Table 3) fully support this finding. The simulation temperature $T = 300$ K.

(38), given by

$$s_c = \frac{1}{2N} k_B \ln[(2\pi e)^N \det \mathbf{C}] \quad (3)$$

where $N = 6$ is the number of coordinates, $e = 2.718\ldots$ is the base of the natural logarithm and det denotes the matrix determinant. The values of s_c are shown in Figure 4. While ICL A is only slightly more flexible (has higher entropy) than the control undamaged duplex, ICL B exhibits dramatically higher flexibility than the control. A more detailed insight can be obtained by inspecting the stiffness matrix diagonal. The diagonal entries are stiffness constants for deforming just that one coordinate, while the other coordinates remain unchanged. The values (not shown) reveal that ICL B is more flexible than the control in all the coordinates.

The virtual step model captures the overall flexibility of the ICL site. To obtain more insight, we also consider a more detailed model in which each base within the ICL site (Figure 3) is treated as a rigid body. The configuration is defined by a set of intra-basepair, basepair step and single-strand inter-base coordinates as introduced in the 3DNA program (36). The MD time series of the coordinate vector are then treated exactly as above to obtain the stiffness matrix and the entropy. Here, however, the coordinate distributions are significantly anharmonic in some cases (Supplementary Figure S9), meaning that the use of the harmonic model involves a certain coarse-graining of the probability distribution, namely replacing the actual (possibly complicated) distribution by an effective Gaussian with the same mean and covariance. Finally, we computed the entropy of the ICL sites (Figure 3) using the atomic-resolution quasi-harmonic scheme of Andricioaei and Karplus (39). Notice that this model is also based on the harmonic approximation.

We complement the structural description by measuring the overall bending of the oligomers, using the method described in (40). Briefly, three groups of bases are defined within the oligomer, one at each end and one in the middle.

Table 1. Mean basepair step coordinates of the virtual steps comprising the lesions (Figure 3) and the corresponding virtual steps of the control, intact B-DNA. Error estimates in parentheses are mean differences between values for the whole trajectory and for its halves

	tilt (°)	roll (°)	twist (°)	shift (Å)	slide (Å)	rise (Å)
ICL A	−20.7 (0.4)	12.1 (1.4)	125.2 (2.7)	−1.48 (0.14)	−2.47 (0.12)	12.84 (0.08)
control A	−2.6 (0.2)	13.2 (1.1)	125.7 (0.3)	−0.57 (0.13)	−3.11 (0.12)	13.13 (0.01)
ICL B	−1.5 (3.1)	23.6 (7.4)	107.6 (8.6)	−1.54 (0.11)	3.02 (0.24)	17.37 (0.08)
control B	1.5 (0.0)	16.4 (0.2)	127.6 (0.0)	0.12 (0.01)	−3.83 (0.08)	13.09 (0.00)

Base-fixed coordinate frames in each group are averaged to yield one mean frame per group. Bending coordinates in two perpendicular directions (global roll and global tilt) are averaged over the MD trajectory to define the mean bending. It is then expressed in terms of bending magnitude (angle between the z-axes of the end frames) and bending direction angle, defined as the angle between the bending direction and the x-axis of the middle frame pointing into the major groove. Thus, a direction of 0° indicates bending toward the major groove, 180° means bending toward the minor groove in the middle of the oligomer.

Here, we chose bases in pairs 4-5 and 17-18 as the end groups, and bases in the the two pairs defining the virtual step (see above) as the middle group. The results are summarized in Table 2. Both control sequences are slightly bent toward the minor groove. Introducing the lesion in Sequence A changes the bend direction toward the backbones, which is consistent with the negative tilt introduced by the lesion (Table 1). The lesion in Sequence B straightens the oligomer. It can be understood as the effect of increased roll (i.e. local bending towards the major groove, Table 1) at the lesion site that compensates for the minor groove bending in the rest of the oligomer.

Finally, we compute global bending and twist stiffness of the damaged and undamaged oligomers. The calculations are done exactly as in (38). Briefly, a quadratic model analogous to Equations (1) and (2) is used, with coordinates involving bending angles toward the grooves and toward the backbone together with total twist. The first two diagonal entries of the 3×3 stiffness matrix are stiffness constants for groove and backbone bending, respectively, from which the equivalent isotropic bending stiffness is computed as their harmonic average. The results in Table 2 again indicate exceptionally small stiffness of the oligomer containing ICL B.

Overall, our data indicate that the two lesions exhibit quite different hydrogen bonding and stacking patterns and affect the DNA properties in a different way. In particular, ICL A introduces moderate local bending toward the backbones and is only slightly more flexible than a piece of regular duplex DNA, while ICL B introduces undertwisting as well as a lateral displacement and creates a hotspot of high flexibility.

DISCUSSION

Approach between the abasic site and vicinal purines

Our simulation of Ap-contanining 21-bp duplexes provide a dynamic view of the approach between Ap11 and the vicinal purines A31 and G33. It differs significantly from the distances inferred from X-ray structures of undamaged DNA (e.g. PDB ID code 3BSE (7)) or on 'frozen' molecular models, which both neglect the rearrangement the duplex undergoes upon formation of an Ap site. Our simulations go beyond these inspections, as we can infer approaching distances between the Ap site and proximal purines, as we also reported for other interstrand cross-links such as dCyd341 (13,41). The two reactants exhibit different eases of approach: Ap11 gets closer to G33 spontaneously within ∼50 ns, prefiguring the cross-link Ap-dG. Interestingly, Ap and A31 in sequence B do not approach whereas the corresponding lesion Ap-dA was found experimentally to present a much higher yield of formation.

The selective formation of the ICL Ap-dA31 (15–70%) versus Ap-dA32 hence cannot be explained by the reactant structure. Furthermore, Ap approaches dG33 easily, but its yield of formation is reported to be only 2–3% (7). Our simulations demonstrate that not only the positioning of the two reactive partners within the oligonucleotide but also, and perhaps more importantly, the final structure of the product and its thermal stability determine the ease of formation of the ICL adduct: this also can be proposed since the reaction has been reported to be a reversible process (6).

Comparison of interstrand cross-links to other lesions

The two interstrand cross-links we characterize by computational means also single out compared to other intrastrand lesions implying a covalent linkage. For instance, we do not observe the exclusion of proximal pyrimidines, in spite of the marked narrowing of the minor groove. The value of the bend angle remains within the thermal fluctuations, again in contrast to what is known for photolesions such as (6-4)PP (42) and intrastrand cross-links (23–25). Other single-nucleotide lesions do not affect the B-helix so markedly. The most distinctive feature of this class of oxidatively-induced lesions is to prevent DNA strand separation: our simulations reveal more finely contrasted structural behavior. We evaluated notably the bending, which can be compared to interstrand lesions induced by cross-linking agents (43): diamminedichloroplatinum induces a bend by 35° to 45° and an unwinding by 79° for the *cis* conformer, but a less marked distortion for the *trans* (unwound by 12° and bent by 26°). Nitrogen mustard only weakly perturbates B-helicity, with a static bend of ca. 14° in DNA, whereas other agents induce no bending such as mitomycin and psoralens. It appears that each interstrand cross-link lesion can be associated to a structural signature, and our simulations for oxidatively-induced ICLs provide such data that are crucially missing.

Table 2. Bending magnitudes and directions of the duplexes, their bending stiffness and twist stiffness. The direction of 0° indicates bending toward the major groove in the oligomer centre, 180° means bending into the minor groove. Errors in parentheses were computed as in Table 1

	Bending magnitude (°)	Bending direction (°)	Bending stiffness (nm)	Twist stiffness (nm)
ICL A	15.2 (1.0)	129 (1)	68 (3)	64 (7)
control A	9.1 (1.7)	177 (3)	69 (2)	99 (2)
ICL B	2.6 (3.7)	190 (80)	45 (4)	39 (7)
control B	10.3 (0.5)	204 (2)	77 (3)	109 (2)

Energetic consequences of the formation of ICL

It is legitimate to ask whether or not the contrasting structure and flexibility of Sequences A and B may be related to the different yields of ICL formation measured experimentally. Indeed, within the hypothesis of a thermodynamic control, and the ICL formation being a reversible process (6), the ICLs thermal stability can be a decisive factor. We can provide an estimate of the reaction energies based on density functional theory (DFT, see Materials and Methods). To this end we compute the ICL formation (de)stabilization energies $\Delta E^{ICL} = E_{\text{prod}} - E_{\text{react}}$ between the reactants and the products for a fragment featuring the reactive system and the two vicinal base pairs. They account for +5.4 and −14.5 kcal/mol for Sequences A and B, whereas the values for the isolated Ap...A or Ap...G systems are ca. +6–7 kcal/mol (see Supplementary Figure S1). Hence the proximal nucleobases tune the thermal stability of the products, which in turn affects the energy profile, giving rise to contrasted yields. Indeed, from our unconstrained simulations of the Ap-containing oligonucleotides, the Ap sites spontaneously approach purines on the opposite strand at 4–5 Å (see Figure 2). Our DFT analysis suggests that the thermal stability of products is decisive for the selective formation of ICLs, whereas the reactant valley can be more shallow. Our results reinforce the postulate of a correlation between structural distortion and thermal stabilities of DNA interstrand cross-links (15). These overall energies reflect a subtle interplay between different non-covalent interactions. Ap-dA is characterized as a flexible hotspot with an intrinsic chemical stability, which suggests that this lesion in duplex DNA may have the power to block DNA-processing enzymes involved in transcription and replication (9). The ICL induction is more favored for Ap-dA (Sequence B) than for Ap-dG (Sequence A), as the former is associated with a more favorable reaction energy. Moreover, the Ap-dG ICL is further stabilized by the higher entropy of the lesion site (Table 3). To reach a more quantitive assessment, free energies and sequence effects should be tackled, which constitutes a perspective of this work.

Toward a cross-talk between simulations and experiments on ICLs

The structural and mechanical properties we derived agree fairly well with available experimental data, but an in-depth understanding and validation of our modellization findings can only arise from additional experiments. Our simulations hence call for an experimental proof: whereas the obtention of nuclear magnetic resonance or X-ray structures would be the most definitive probe and validate the calibration of the timescale of our simulations, it is a considerably difficult task. Another key information accessible experimentally is the yield of induction of oxidatively-induced ICLs, and also sometimes kinetic measurements (44). Hence, free energies should be determined with a high accuracy to reach a synergetic view with classical MD simulations: this would require accurate free energy calculations, since the scarce data available indicate differences that can be as low as 0.3–1.0 kcal/mol (44). Our DFT cluster analysis is probably only capable to characterize ICLs that differ importantly in energy. An equally important information that is more easily accessible consists of experimental measurements of oligonucleotides flexibility. The method of DNA cyclization can provide the magnitude and direction of bending as well as torsional and bending stiffness of a short, intact DNA sequence motif (45). Thus, if the lesion is incorporated in the middle of a short DNA sequence, then the same method should be able to measure the effect of the lesion on these properties. The measurements can then be directly compared to our computed values in Table 2. Only several copies of the lesion would be needed: for instance, no more than three copies of the sequence motif were used in (45). Measures of AP sites and interstrand cross-link half lifetime is also a valuable experimental information (46) characterizing the constrasted mechanical properties DNA lesions experience once in B-DNA or nucleosomal DNA.

CONCLUSION

Interstrand cross-links are particularly deleterious lesions, yet a deep understanding of their formation remains to be laid on a firm basis. In this work, explicit-solvent MD simulations combined with quantum chemical DFT calculations were performed to unravel the structure, dynamics and mechanical properties of two experimentally characterized ICL-containing oligonucleotides. The results refine the initial rigid model provided by experimentalists, who invoke a higher reactivity for the purine presenting the shortest proximal approaching distance to the electrophilic Ap site. This criterion implies the selectivity onto a purine adjacent to the orphan nucleobase rather than onto the orphan nucleobase itself, yet is not sufficient to explain the lower yield of Ap-dG versus Ap-dA.

Our simulations allow to build up oligonucleotides containing these two defects, and we evidence constrasted structural and mechanical properties that probably induce a different free energy profile. A natural perspective would be to determine a free energy profile for the multi-step reaction, or alternatively free energies of destabilization. This

Table 3. Entropy differences per coordinate with respect to the Control B simulation using the virtual step model (Figure 4), the rigid base model and the atomistic quasi-harmonic model. Errors in parentheses were obtained as in Table 1

	$T\Delta s_c$ (kcal/mol)		
	Virtual step	Rigid base	Atomistic
control A	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
ICL A	0.09 (0.02)	0.10 (0.01)	0.05 (0.00)
ICL B	0.29 (0.03)	0.28 (0.05)	0.13 (0.01)

would corroborate a nascent correlation between structure and thermal destabilization for interstrand cross-links (15).

Furthermore, the contrasting mechanical rigidity of the two lesion sites may affect the supercoiling properties of the damaged DNA and play a role in recognizing the ICLs during the initial stages of their repair. The replication-dependent repair requires one or possibly two replication forks to meet at the ICL and thus assumes a large-scale deformation well beyond the harmonic approximation adopted here. The same is true for the transcription-dependent repair that assumes a transcription bubble. Besides that, however, ICL repair in quiescent DNA also takes place, possibly to relieve the negative impact of the tethered strands on DNA topology and supercoiling (18). In this case, the peculiar stiffness of the lesion site may play a role in the ICL recognition, just as the sequence-dependent stiffness of undamaged DNA does for many DNA-binding proteins. The specific mechanisms of the ICL repair have only just started to be unravelled. A clear and systematic assessment of sequence effects, coupled to free energy calculations, would be helpful and will be a topic for future work.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program 'Investissements d'Avenir' (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). Calculations were performed using the local HPC resources of PSMN at ENS-Lyon. E.B. is grateful for a PhD fellowship from the French Ministry of Higher Education and Research. T.D. and F.L. were supported by the Grant Agency of the Czech Republic (14-21893S).

FUNDING

Université de Lyon [LABEX PRIMES (ANR-11-LABX-0063)]; French National Research Agency (ANR) ['Investissements d'Avenir' (ANR-11-IDEX-0007)]; Grant Agency of the Czech Republic [14-21893S]. Funding for open access charge: LABEX PRIMES [ANR-11-LABX0063 of Université de Lyon].
Conflict of interest statement. None declared.

REFERENCES

1. Cooke,M.S., Evans,M.D., Dizdaroglu,M. and Lunec,J. (2003) Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.*, **17**, 1195–1214.
2. Box,H.C., Budzinski,E.E., Dawidzik,J.B., Gobey,J.S. and Freund,H.G. (1997) Free radical-induced tandem base damage in DNA oligomers. *Free Rad. Biol. Med.*, **23**, 1021–1030.
3. Dutta,S., Chowdhury,G. and Gates,K.S. (2007) Interstrand cross-links generated by abasic sites in duplex DNA. *J. Am. Chem. Soc.*, **129**, 1852–1853.
4. Ding,H., Majumdar,A., Tolman,J.R. and Greenberg,M.M. (2008) Multinuclear NMR and kinetic analysis of DNA interstrand cross-link formation. *J. Am. Chem. Soc.*, **130**, 17981–17987.
5. Wang,J., Cao,H., You,C., Yuan,B., Bahde,R., Gupta,S., Nishigori,C., Niedernhofer,L.J., Brooks,P.J. and Wang,Y. (2012) Endogenous formation and repair of oxidatively induced G[8-5m]T intrastrand cross-link lesion. *Nucleic Acids Res.*, **40**, 7368–7374.
6. Johnson,K.M., Price,N.E., Wang,J., Fekry,M.I., Dutta,S., Seiner,D.R., Wang,Y. and Gates,K.S. (2013) On the formation and properties of interstrand DNA–DNA cross-links forged by reaction of an abasic site with the opposing guanine residue of 5'-CAp sequences in duplex DNA. *J. Am. Chem. Soc.*, **135**, 1015–1025.
7. Price,N.E., Johnson,K.M., Wang,J., Fekry,M.I., Wang,Y. and Gates,K.S. (2014) Interstrand DNA–DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *J. Am. Chem. Soc.*, **136**, 3483–3490.
8. Price,N.E., Catalano,M.J., Liu,S., Wang,Y. and Gates,K.S. (2015) Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA. *Nucleic Acids Res.*, **43**, 3434–3441.
9. Catalano,M.J., Liu,S., Andersen,N., Yang,Z., Johnson,K.M., Price,N.E., Wang,Y. and Gates,K.S. (2015) Chemical structure and properties of interstrand cross-links formed by reaction of guanine residues with abasic sites in duplex DNA. *J. Am. Chem. Soc.*, **137**, 3933–3945.
10. Yang,Z., Price,N.E., Johnson,K.M. and Gates,K.S. (2015) Characterization of interstrand DNA–DNA cross-links derived from abasic sites using bacteriophage ϕ29 DNA polymerase. *Biochemistry*, **54**, 4259–4266.
11. Zhang,X., Price,N.E., Fang,X., Yang,Z., Gu,L.-Q. and Gates,K.S. (2015) Characterization of interstrand DNA–DNA cross-links using the α-hemolysin protein nanopore. *ACS Nano*, **9**, 11812–11819.
12. Sczepanski,J.T., Jacobs,A.C., Majumdar,A. and Greenberg,M.M. (2009) Scope and mechanism of interstrand cross-link formation by the C4'-oxidized abasic site. *J. Am. Chem. Soc.*, **131**, 11132–11139.
13. Regulus,P., Duroux,B., Bayle,P.-A., Favier,A., Cadet,J. and Ravanat,J.-L. (2007) Oxidation of the sugar moiety of DNA by ionizing radiation or bleomycin could induce the formation of a cluster DNA lesion. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14032–14037.
14. Noll,D.M., Mason,T.M. and Miller,P.S. (2006) Formation and repair of interstrand cross-Links in DNA. *Chem. Rev.*, **106**, 277–301.
15. Ghosh,S. and Greenberg,M.M. (2015) Correlation of thermal stability and structural distortion of DNA interstrand cross-links produced from oxidized abasic sites with their selective formation and repair. *Biochemistry*, **54**, 6274–6283.
16. Kamiya,H. (2003) Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: approaches using synthetic oligonucleotides and nucleotides: survey and summary. *Nucleic Acids Res.*, **31**, 517–531.
17. Catalano,M.J., Ruddraraju,V., Barnes,C.L. and Gates,K.S. (2016) Crystal structure of a nucleoside model for the interstrand cross-link

formed by the reaction of 2'-deoxyguanosine and an abasic site in duplex DNA. *Acta Cryst.*, **E72**, 624–627.
18. Hlavin,E.M., Smeaton,M.B. and Miller,P.S. (2010) Initiation of DNA interstrand cross-link repair in mammalian cells. *Environ. Mol. Mutagen.*, **51**, 604–624.
19. Clauson,C., Schärer,O.D. and Niedernhofer,L. (2013) Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb. Perspect. Biol.*, **5**, a012732.
20. Williams,H.L., Gottesman,M.E. and Gautier,J. (2013) The differences between ICL repair during and outside of S phase. *Trends Biochem. Sci.*, **38**, 386–393.
21. Wilhelm,M., Mukherjee,A., Bouvier,B., Zakrzewska,K., Hynes,J.T. and Lavery,R. (2012) Multistep drug intercalation: molecular dynamics and free energy studies of the binding of daunomycin to DNA. *J. Am. Chem. Soc.*, **134**, 8588–8596.
22. Dumont,E. and Monari,A. (2013) Benzophenone and DNA: Evidence for a double insertion mode and its spectral signature. *J. Phys. Chem. Lett.*, **4**, 4119–4124.
23. Churchill,C. D.M., Eriksson,L.A. and Wetmore,S.D. (2011) Formation mechanism and structure of a guanine–uracil DNA intrastrand cross-link. *Chem. Res. Toxicol.*, **24**, 2189–2199.
24. Dumont,E., Oršata,T., Fonseca Guerra,C. and Lankaš,F. (2015) Insights into the structure of intrastrand cross-link DNA lesion-containing oligonucleotides: G[8–5m]T and G[8–5]C from molecular dynamics simulations. *Biochemistry*, **54**, 1259–1267.
25. Churchill,C. D.M., Eriksson,L.A. and Wetmore,S.D. (2016) DNA Distortion Caused by Uracil-Containing Intrastrand Cross-Links. *J. Phys. Chem. B*, **120**, 1195–1204.
26. Case,D.A., Darden,T.A., Cheatham,T.E. III, Simmerling,C.L., Wang,J., Duke,R.E., Luo,R., Walker,R.C., Zhang,W., Merz,K.M. *et al.* (2012) AMBER 12. University of California, San Francisco.
27. Ayadi,L., Coulombeau,C. and Lavery,R. (1999) Abasic sites in duplex DNA: molecular modeling of sequence-dependent effects on conformation. *Biophys. J.*, **77**, 3218–3226.
28. Barsky,D., Foloppe,N., Ahmadia,S., Wilson III,D.M. and MacKerell,A.D. Jr (2000) New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.*, **28**, 2613–2626.
29. Chen,J., Dupradeau,F.-Y., Case,D.A., Turner,C.J. and Stubbe,J. (2008) DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.*, **36**, 253–262.
30. Ivani,I., Dans,P.D., Noy,A., Perez,A., Faustino,I., Hospital,A., Walther,J., Andrio,P., Goñi,R., Balaceanu,A. *et al.* (2016) Parmbsc1: a refined force-field for DNA simulations. *Nat. Methods*, **13**, 55–58.
31. Galindo-Murillo,R., Robertson,J.C., Zgarbová,M., Šponer,J., Otyepka,M., Jureačka,P. and Cheatham,T.E. (2016) Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.*, **12**, 4114–4127.
32. Sviatenko,L., Gorb,L., Hovorun,D. and Leszczynski,J. (2012) Interaction of 2-Deoxyadenosine with cis-2-Butene-1,4-dial: Computational approach to analysis of multistep chemical reactions. *J. Phys. Chem. A*, **116**, 2333–2342.
33. Fonseca Guerra,C., van der Wijst,T., Poater,J., Swart,M. and Bickelhaupt,F.M. (2010) Adenine versus guanine quartets in aqueous solution: dispersion-corrected DFT study on the differences in π-stacking and hydrogen-bonding behavior. *Theor. Chem. Acc.*, **125**, 245–252.
34. Dupont,C., Patel,C. and Dumont,E. (2011) Improved DFT description of intrastrand cross-link formation by inclusion of london dispersion corrections. *J. Phys. Chem. B*, **115**, 15138–15144.
35. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Scalmani,G., Barone,V., Mennucci,B., Petersson,G. A. *et al.* (2009) *G.A.P. Gaussian~09 Revision D.01*. Gaussian Inc. Wallingford.
36. Lu,X.-J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
37. Drsata,T. and Lankas,F. (2013) Theoretical models of DNA flexibility. *WIREs Comput. Mol. Sci.*, **3**, 355–363.
38. Drsata,T., Spackova,N., Jurecka,P., Zgarbova,M., Sponer,J. and Lankas,F. (2014) Mechanical properties of symmetric and asymmetric DNA A-tracts: implications for looping and nucleosome positioning. *Nucleic Acids Res.*, **42**, 7383–7394.
39. Andricioaei,I. and Karplus,M. (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.*, **115**, 6289–6292.
40. Lankas,F., Spackova,N., Moakher,M., Purevjav,E. and Sponer,J. (2010) A measure of bending in nucleic acids structures applied to A-tract DNA. *Nucleic Acids Res.*, **38**, 3414–3422.
41. Patel,C., Garrec,J., Dupont,C. and Dumont,E. (2013) What singles out the G[8-5]C intrastrand DNA cross-link? mechanistic and structural insights from quantum mechanics/molecular mechanics simulations. *Biochemistry*, **52**, 425–431.
42. Kim,J.-K., Patel,D. and Choi,B.-S. (1995) Contrasting structural impacts induced by cis-syn cyclobutane dimer and (6-4) adduct in DNA duplex decamers: implication in mutagenesis and repair activity. *Photochem. Photobiol.*, **62**, 44–50.
43. Noll,D.M., Mason,T.M. and Miller,P.S. (2006) Formation and Repair of Interstrand Cross-Links in DNA. *Chem. Rev.*, **106**, 277–301.
44. Fridman,A.S., Brabec,V., Haroutiunian,S.G., Wartell,R.M. and Lando,D.Y. (2003) Melting of cross-linked DNA V. Cross-linking effect caused by local stabilization of the double helix. *J. Biomol. Struct. Dyn.*, **20**, 533–545.
45. Roychoudhury,M., Sitlani,A., Lapham,J. and Crothers,D.M. (2000) Global structure and mechanical properties of a 10-bp nucleosome positioning motif. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 13608–13613.
46. Catalano,M.J., Liu,S., Andersen,N., Yang,Z., Johnson,K.M., Price,N.E., Wang,Y. and Gates,K.S. (2015) Chemical structure and properties of interstrand cross-links formed by reaction of guanine residues with abasic sites in duplex DNA. *J. Am. Chem. Soc.*, **137**, 3933–3945.

Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx148

Conformational polymorphism or structural invariance in DNA photoinduced lesions: implications for repair rates

François Dehez[1,2,3,*], Hugo Gattuso[1,2], Emmanuelle Bignon[4,5], Christophe Morell[4], Elise Dumont[5,*] and Antonio Monari[1,2,*]

[1]CNRS, Theory-Modeling-Simulation, SRSMC F-54506 Vandoeuvre-lès-Nancy, France, [2]Université de Lorraine, Theory-Modeling-Simulation, SRSMC F-54506 Vandoeuvre-lès-Nancy, France, [3]Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana Champaign, [4]Institut des Sciences Analytiques, UMR 5280, Université de Lyon1 (UCBL) CNRS, ENS Lyon, Lyon, France and [5]Université de Lyon, ENS de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F69342 Lyon, France

Received December 20, 2016; Revised February 06, 2017; Editorial Decision February 20, 2017; Accepted February 21, 2017

ABSTRACT

DNA photolesions constitute a particularly deleterious class of molecular defects responsible for the insurgence of a vast majority of skin malignant tumors. Dimerization of two adjacent thymines or cytosines mostly gives rise to cyclobutane pyrimidine dimers (CPD) and pyrimidine(6-4)pyrimidone 64-PP as the most common defects. We perform all-atom classical simulations, up to 2 μs, of CPD and 64-PP embedded in a 16-bp duplex, which reveal the constrasted behavior of the two lesions. In particular we evidence a very limited structural deformation induced by CPD while 64-PP is characterized by a complex structural polymorphism. Our simulations also allow to unify the contrasting experimental structural results obtained by nuclear magnetic resonance or Förster Resonant Energy Transfer method, showing that both low and high bent structures are indeed accessible. These contrasting behaviors can also explain repair resistance or the different replication obstruction, and hence the genotoxicity of these two photolesions.

INTRODUCTION

DNA is well renown for its photostability, (1–3) which is a key feature to ensure genome stability and avoid mutations or carcinogenesis (4,5). Nucleobases and nucleic acids absorb in the UV spectrum, mostly in the UVB and slightly in the UVA region due to excitonic coupling. In spite of the existence of efficient photophysical channels (6) allowing non-reactive relaxation, UV absorption opens photochemical pathways leading to nucleobase modification (7). The most common DNA damages produced upon direct UV absorption are characterized by the dimerization of two adjacent, π-stacked pyrimidine rings in B-DNA (see Figure 1) (8–10). Ultimately, photoinduced DNA lesions may induce harmful effects at the cellular level that, in the case of eukaryotic organisms, may be translated into mutation, skin aging and carcinogenesis (11). Indeed, UV absorption, and unprotected sun exposure, are nowadays recognized as the main causes of malignant skin tumors. From a molecular point of view, tumor insurgence is mostly associated with the presence of cyclobutane pyrimidine dimers (CPD) and/or pyrimidine(6-4)pyrimidone 64-PP (8), both lesions being mutagenic and possibly inducing genome instability with differential biological effects (12,13).

UVB irradiation of cells (245 nm) generates CPD as the most abundant photolesions while 64-PP, resulting from the excitation to higher lying electronic states (14), accounts for ∼20% of the defects formed (15,16). Processing efficiency is especially low for CPD (17), and in particular way less efficient than for 64-PP (15). For the repair of both lesions the nucleotide excision repair (NER) mechanism is mobilized (18). However, NER is known to participate in case of strongly distorted DNA structures and in the presence of bulky lesions. Hence, since deformation plays a fundamental role in the lesion recognition (19) a detailed understanding of the structural and dynamical signatures of photolesions is necessary to allow for a rationalization of repair rates (20–23).

X-ray or nuclear magnetic resonance (NMR) structures are indeed available for 6-4PP complexed with repair enzymes (photolyase (17)), that lead to the understanding

*To whom correspondence should be addressed. Tel: +33 372745278; Fax: +33 372745271; Email: antonio.monari@univ-lorraine.fr
Correspondence may also be addressed to François Dehez. Tel: +33 372745076; Fax: +33 372745271; Email: francois.dehez@univ-lorraine.fr
Correspondence may also be addressed to Elise Dumont. Tel: +33 472728846; Fax: +33 4 72728080; Email: elise.dumont@ens-lyon.fr

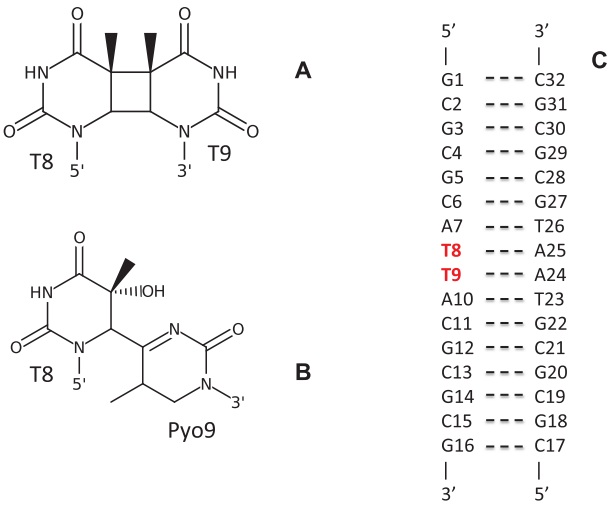


Figure 1. Molecular formula of CPD (**A**) and 64-PP (**B**). The 16 bp DNA double strand is also schematically reported with the bases' numbering (**C**). The two adjacent thymines at positions 8 and 9 (boldfaced in red) are the ones dimerizing to form the photolesions. In the case of 64-PP, the ninth nucleobase (T9) will be denoted Pyo9, being transformed into a pyrimidone unit.

of the enzymatic pathway toward repair at the molecular level (24), also thanks to molecular modeling and simulations (25,26). Yet structural resolution is far less often available for short oligonucleotides containing this photodamage. Two structures for solvated decamers containing 64-PP were reported (27–29), while most recently the structure of a 64-PP containing strand embedded within an histone (23) has been resolved. Other less direct measurements based on Förster resonant energy transfer (FRET) have also been published (30). Interestingly, the experimental determinations provided rather contrasting results: FRET measurements suggest a low DNA bending comparable to the one of the undamaged oligomers, as revealed by the large end-to-end distance of the strand, while X-ray and NMR agree on large bending going up to 44°. If the crystal packing or the interaction with histones can strongly modify the native structural properties of the DNA strands, still the experimental results in solution should be rationalized. Molecular dynamics (MD) simulations have been performed in the past on solvated oligonucleotides (31) or on oligonucleotides interacting with repair enzymes. (15,17,23). The pioneer simulation by Kollman's group (31) reported a very low bending for the 64-PP containing strand, however, the time-scales of the simulations were generally too small to capture their high structural flexibility. Also the structure of CPD containing oligomers has been resolved in the past concluding in favor of a general low bending of the DNA strand, (32) also in agreement with theoretical analysis (33). However, bending up to 30° have also been reported (34). While the repair mechanisms are now well understood owing to numerous and converging experimental (35) and modeling studies (36), insights are still scarce concerning the DNA-enzyme interaction leading to lesion recognition.

Structural deformation and strand flexibility have strong influence on the repair rates of DNA lesions. Indeed while bulky and important deformations are necessary to allow recognition by repair machinery (19), especially NER, flexibility has also been invoked as a factor enhancing the repair propensity of DNA lesions (37).

The use of well-calibrated (38,39), long-scale MD techniques allowing for an extended sampling of the conformational space of the DNA strand are needed to achieve a clear atomostic-scale characterization of the structural reorganization induced by photolesions. More importantly, sufficiently long MD will be able to capture the dynamical evolution of the structural deformation and hence will allow to properly assess the eventual polymorphisms exhibited by the damaged DNA oligonucleotides. MD simulations have been used to ultimately evidence the structural modification induced on DNA oligomers by oxidative lesions, intra- and inter-strand crosslinks, or via the interaction with exogenous and endogenous drugs (40–44). Recently, the structural deformations induced by clustered abasic sites have been correlated to the repair rate in human and bacterial organisms (21,22).

In this work we report an extensive all-atom MD simulation, up to the μs time scale, of DNA oligomers containing one CPD or one 64-PP damage, respectively; compared to the dynamics of the control undamaged B-DNA strand. The different structural behavior induced by the two lesions will be clearly assessed and compared to the evolution of the native B-DNA. The negligible deformations produced by CPD, and the extended polymorphism induced by 64-PP will be deeply analyzed and correlated with the low repair or the replication obstruction of the two photolesions.

MATERIALS AND METHODS

Computational protocol

For all the subsequent study we consider a 16-bp DNA strand (see Figure 1); 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 thymines 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 (46) codes for the three systems, B-DNA, CPD and 64-PP. DNA nucleobases were represented by the parmbsc0 force field (47,48). The performance of the parmbsc0 force field has also been tested against the recently developed bsc1 (38) 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 (49). We note that they are in very good agreement with the ones derived by Kollman *et al.* (31). Water molecules were represented by TIP3P model and 30 potassium cations were added to the simulation box to assure neutrality. A detailed description of the computational protocol and the force-field parameters for the damaged base can be found in Supplementary Data. 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 μs 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+ (50). The compactness of the structures was assessed by the calculation of the solvent accessible surface area (SASA) (51) obtained post-processing the MD trajectories with the measure command of VMD (52) and a probe radius of 1.4 Å. The local deformation was further explored by analyzing the extent of π-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 2, 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 2C), 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 2 rightside, globally show similar tendencies as for 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 3 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.

In order to assess the convergence of 64-PP simulation we also performed parallel tempered replica exchange (53) representing a total sampling time of about 5 μs. The distribution of bending and SASA obtained for 64-PP are reported in Figure 4. 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.

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 5 provides the time evolution of the distance between the nucleobases in the vicinity of the 64-PP lesion (see Supplementary Data for a proper definition). This parameter has been chosen since it allows to quantify the extent of the formation of π-stacking interactions. Along the 2 μs of the trajectory 64-PP explores a rather large conformational space with a complex equilibrium between different conformations, characterized by different π-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 6 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. Once again the same behavior is confirmed by the analysis of the distances distribution for the unbiased and parallel tempering simulation as reported in Supplementary Data, even though, as already evidenced for the SASA the relative population is slightly altered.

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 6, 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 1). Indeed, the large conformational landscape explored at the μs range arises from the fact the duplex can rearrange to favor interstrand dispersive interactions (green boxes in Figure 6) and hydrogen bonding with offset nucleobases (dashed blue line), which in some cases can lead to the exclusion of vicinal nuclebases.

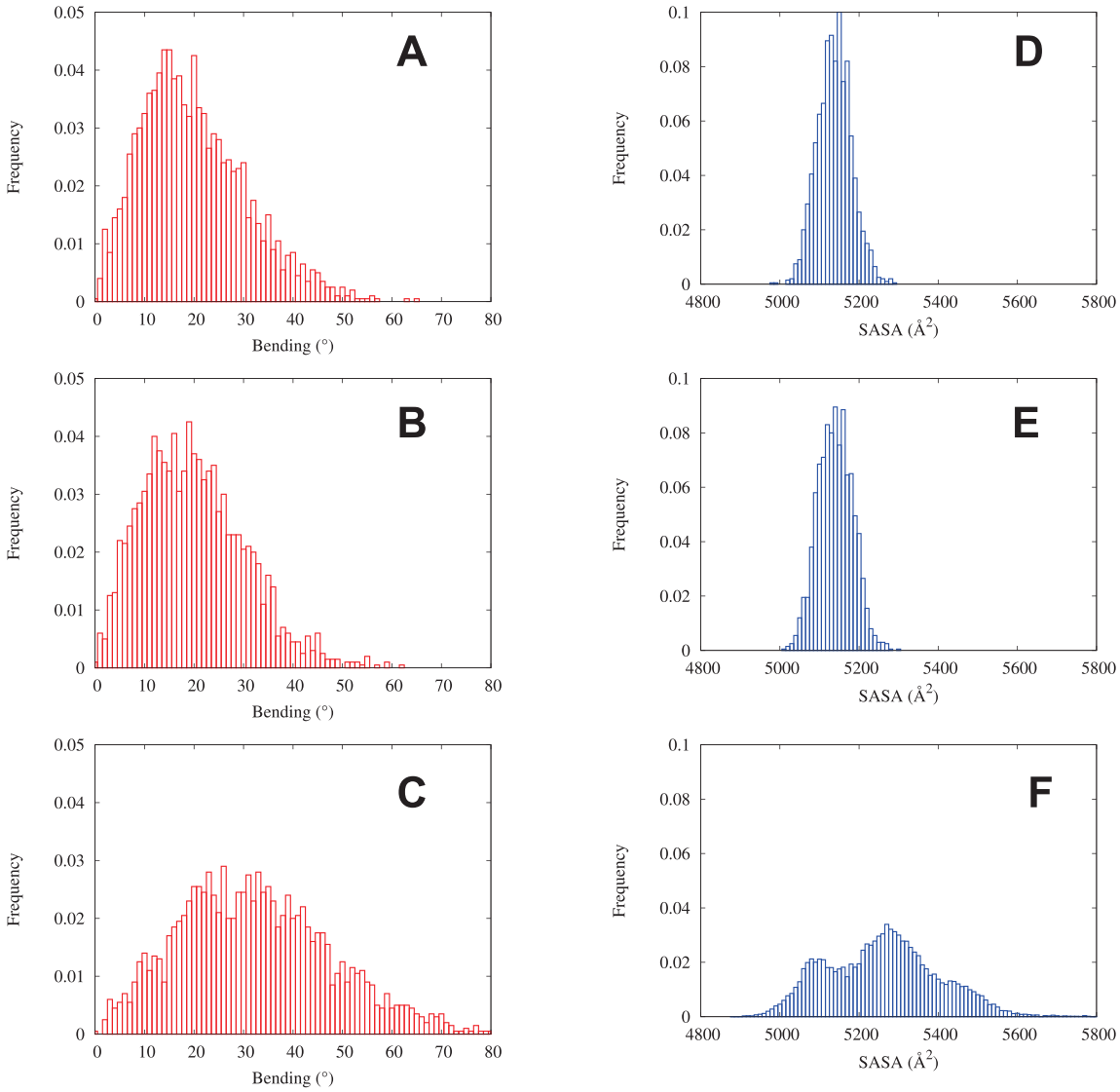


Figure 2. Distribution of the total DNA bending (in red) for B-DNA (**A**), CPD (**B**) and 64-PP (**C**); and of the SASA (in blue) for B-DNA (**D**), CPD (**E**) and 64-PP (**F**).

Conf-(1) maintains a spatial proximity between the damaged T8 and the originally complementary adenine A24, at a distance of ∼4Å which actually corresponds to the π-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 nucleobasis.

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°).

After 1 μs, the duplex rearranges to adopt a more straight and compact conformation Conf-(3) with interstrand π-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°), whereas no new hydrogen bonds comes into play to stabilize the damaged duplex.

At around 1.8 μs 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°). 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-

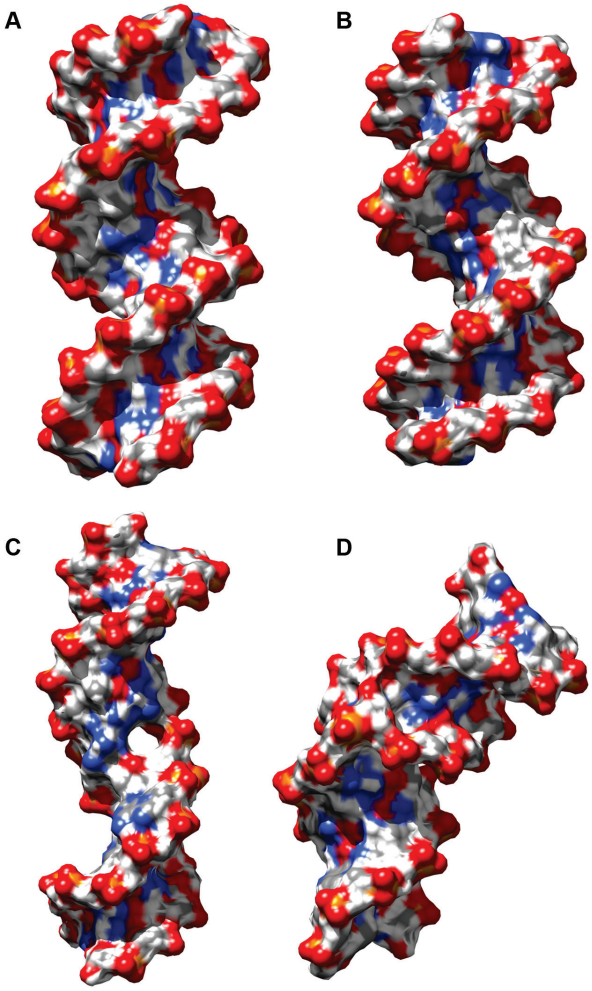


Figure 3. Representative snapshots rendered in surface mode for B-DNA (**A**), CPD (**B**) and 64-(PP) in a low bending high SASA conformation (**C**) and in a high bending low SASA one (**D**).

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 π-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+ (50), such as the minor and major groove width and depth (see Supplementary Data). 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 CONCLUSIONS

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° cannot be ruled out and can occur at room temperate. On the other hand, in the case of 64-PP our results show that both NMR (29) and FRET (30), i.e. high and low bending, are possible due to its polymorphisms and depending on the specific conformations. Indeed, we also report in Supplementary Data 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

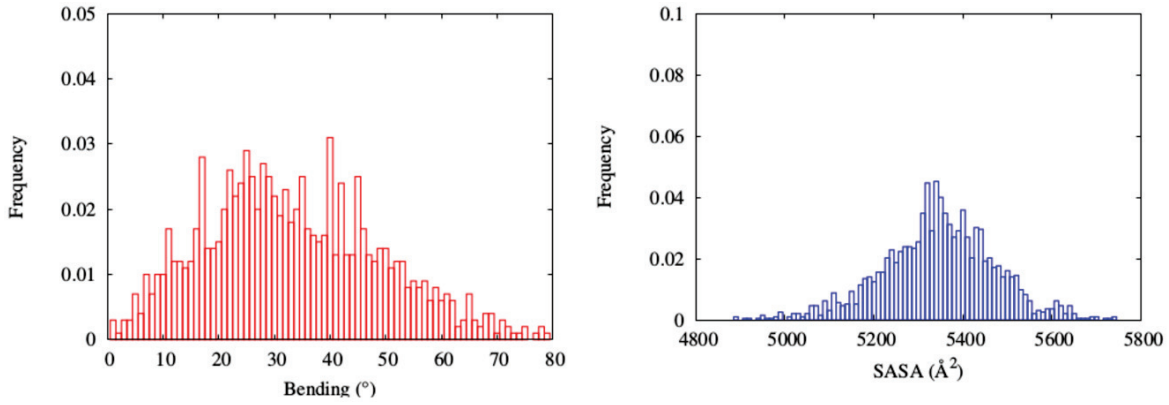


Figure 4. Distribution of the total DNA bending (red) and SASA (blue) obtained from replica exchange parallel tempering.

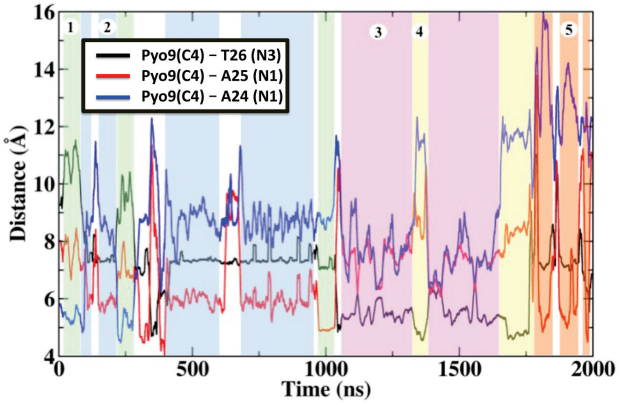


Figure 5. Time evolution of internucleobases distances measured between Pyo9 and A24, A25 and T26, characterizing the different π-stacking formations and hence the increased flexibility and polymorphism upon formation of 64-PP. The regions corresponding to different conformations are indicated by colored background.

values reported in (30) for strands having the same number of base pairs. The NMR predicted bending value of around 40° 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 (54) or by molecular simulation (55), it has also been correlated to the insurgence of different pathologies in particular when involving gene regulatory areas such as CpG islands (56) or telomeres (57,58).

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 (41,59). 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 (37). 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 (60).

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 (61). 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 (62) 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 (63), 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.* (23). In particular the constraints due to interacting proteins and the compact nucleosomal environment may restrict the accessibility of the explored conformational

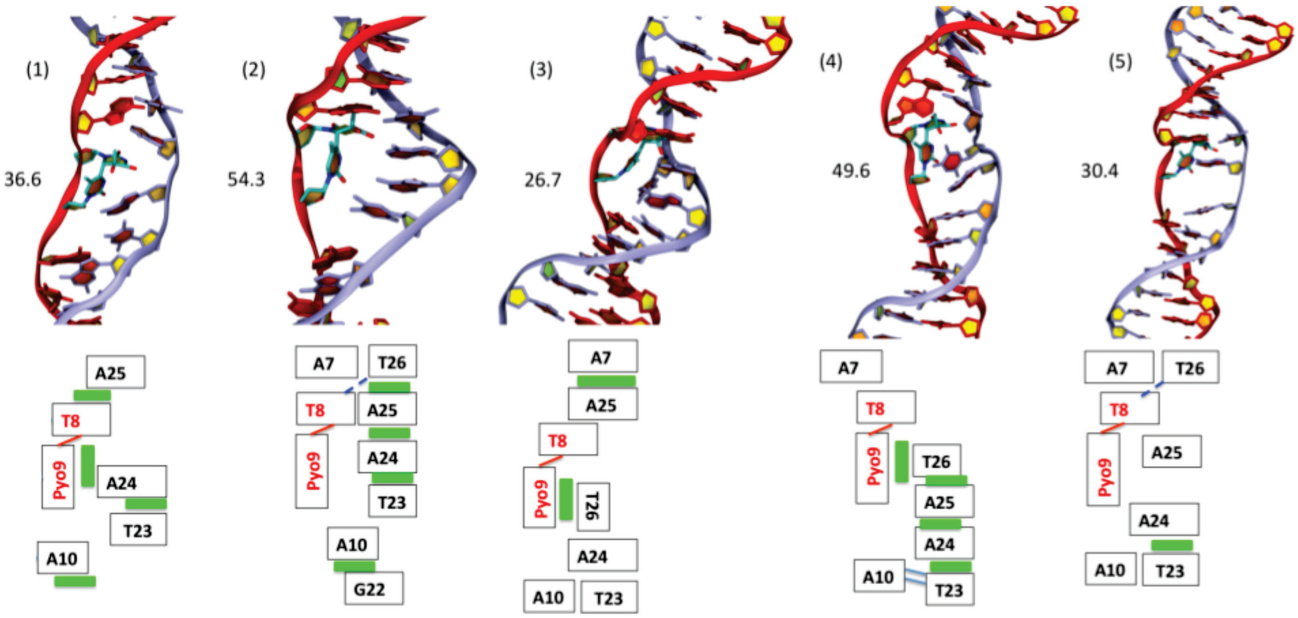


Figure 6. Cartoon representations of five representative conformation observed along the 2 μs trajectory. The schematic depiction of the interaction patterns is given in the bottom figures for each conformer, where green boxes represent π-stacking and blue lines hydrogen bonds. The value of the bending angle for each conformer is indicated in parenthesis.

landscape, still the influence of flexibility on the replication obstruction will hold.

Our simulations have allowed to shed a new light on the complex and dynamic structure of two very important lesions exploring complex free energy surfaces. They sketched two different and opposed behaviors: structural invariance (CPD) or conformational polymorphism (64-PP). In the following, and in order to draw a more precise correlation with repair efficiency, we preview to extend this study to the interaction between the different lesions' conformers and both nucleosomal environment and the repair enzymes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Supports from the University of Lorraine and ENS Lyon is gratefully acknowledged. AM thanks CNRS for support under the project 'Action pour l'interdisciplinarite 2016'. Support from the French Grand-Est region under the Idea project is acknowledged too. This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program 'Investissements d'Avenir' (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

FUNDING

University of Lorraine; ENS Lyon; CNRS under the project 'Action pour l'interdisciplinarite 2016' (to A.M.); French Grand-Est region under the Idea project; LABEX PRIMES of Université de Lyon [ANR-11-LABX-0063]; French National Research Agency (ANR) 'Investissements d'Avenir' program [ANR-11-IDEX-0007]. Funding for open access charge: University of Lorraine and Lyon.
Conflict of interest statement. None declared.

REFERENCES

1. Satzger,H., Townsend,D., Zgierski,M.Z., Patchkovskii,S., Ullrich,S. and Stolow,A. (2006) Primary processes underlying the photostability of isolated DNA bases: Adenine. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 10196–10201.
2. Barbatti,M., Aquino,A. J.A., Szymczak,J.J., Nachtigallova,D., Hozba,P. and Lischka,H. (2014) Relaxation mechanisms of UV-photoexcited DNA and RNA nucleobases. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21453–21458.
3. Buchner,F., Nakayama,A., Yamazaki,S., Ritze,H.-H. and Loebcke,A. (2015) Excited-state relaxation of hydrated thymine and thymidine measured by liquid-jet photoelectron spectroscopy: experiment and simulation. *J. Am. Chem. Soc.*, **137**, 2931–2938.
4. Deman,J. and Van Larebeke,N. (2001) Carcinogenesis: mutations and mutagens. *Tumor Biol.*, 191–202.
5. Morley,A.A. and Turner,D.R. (1999) The contribution of exogenous and endogenous mutagens to in vivo mutations. *Mutat. Res.*, **428**, 11–15.
6. Gustavsson,T., Improta,R. and Markovitsi,D. (2010) DNA/RNA: building blocks of life under UV irradiation. *J. Phys. Chem. Lett.*, **1**, 2025–2030.
7. Cadet,J., Mouret,S., Ravanat,J.-L. and Douki,T. (2012) Photoinduced damage to cellular DNA: direct and photosensitized reactions'. *Photochem. Photobiol.*, **88**, 1048–1065.
8. Mitchell,D.L. and Nairn,R.S. (1989) The biology of the (6-4) photoproduct. *Photochem. Photobiol.*, **49**, 805–819.
9. Gonzalez-Ramirez,I., Roca-Sanjuán,D., Climent,T., Serrano-Perez,J.J., Merchan,M. and Serrano-Andres,L. (2011) On the photoproduction of DNA/RNA cyclobutane pyrimidine dimers. *Theor. Chem. Acc.*, **128**, 705–711.
10. Rauer,C., Nogueira,J.J., Marquetand,P. and Gonzalez,L. (2016) Cyclobutane thymine photodimerization mechanism revealed by nonadiabatic molecular dynamics. *J. Am. Chem. Soc.*, **138**, 15911–15916.

11. D'Orazio,J., Jarrett,S., Amaro-Ortiz,A. and Scott,T. (2013) UV radiation and the skin. *Int. J. Mol. Sci.*, **14**, 12222–12248.
12. Lo,H.-L., Nakajima,S., Ma,L., Walter,B., Yasui,A., Ethell,D.W. and Owen,L.B. (2005) Differential biologic effects of CPD and 6-4PP UV-induced DNA damage on the induction of apoptosis and cell-cycle arrest. *BMC Cancer*, **5**, 135.
13. Pfeifer,G.P. and Besaratinia,A. (2012) UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. *Photochem. Photobiol. Sci.*, **11**, 90–97.
14. Banyasz,A., Douki,T., Improta,R., Gustavsson,T., Onidas,D., Vayá,I., Perron,M. and Markovitsi,D. (2012) Electronic excited states responsible for dimer formation upon UV absorption directly by thymine strands: joint experimental and theoretical study. *J. Am. Chem. Soc.*, **134**, 14834–14845.
15. Balajee,A.S., May,A. and Bohr,V.A. (1999) DNA repair of pyrimidine dimers and 6-4 photoproducts in the ribosomal DNA. *Nucleic Acids Res.*, **27**, 2511–2520.
16. Perdiz,D., Grof,P., Mezzina,M., Nikaido,O., Moustacchi,E. and Sage,E. (2000) Distribution and repair of bipyrimidine photoproducts in solar UV-irradiated mammalian cells: possible role of dewar photoproducts in solar mutagenesis. *J. Biol. Chem.*, **275**, 26732–26742.
17. Li,J., Liu,Z., Tan,C., Guo,X., Wang,L., Sancar,A. and Zhong,D. (2010) Dynamics and mechanism of repair of ultraviolet-induced (6-4) photoproduct by photolyase. *Nature*, **466**, 887–890.
18. Marteijn,J.A., Lans,H., Vermeulen,W. and Hoeijmakers,J. H.J. (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Biol.*, **15**, 465–481.
19. Scharer,D.O. (2013) Nucleotide excision repair in eukaryotes. *Cold Spring Harb. Perspect. Biol.*, **5**, a012609.
20. Han,C., Srivastava,A.K., Cui,T., Wang,Q.-E. and Wani,A.A. (2016) Differential DNA lesion formation and repair in heterochromatin and euchromatin. *Carcinogenesis*, **37**, 129–138.
21. Bignon,E., Gattuso,H., Morell,C., Dehez,F., Georgakilas,A.G., Monari,A. and Dumont,E. (2016) Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion. *Nucleic Acids Res.*, **44**, 8588–8599.
22. Gattuso,H., Durand,E., Bignon,E., Morell,C., Georgakilas,A.G., Dumont,E., Chipot,C., Dehez,F. and Monari,A. (2016) Repair rate of clustered abasic DNA lesions by human endonuclease: molecular bases of sequence specificity. *J. Phys. Chem. Lett.*, **7**, 3760–3765.
23. Akihisa,O., Hiroaki,T., Wataru,K., Naoki,H., Syota,M., Mayu,H., Naoyuki,M., Tatsuya,T., Junpei,Y., Fumio,H. *et al.* (2015) Structural basis of pyrimidine-pyrimidone (64) photoproduct recognition by UV-DDB in the nucleosome. *Sci. Rep.*, **5**, 16330–16330.
24. Yamamoto,J., Martin,R., Iwai,S., Plaza,P. and Brettel,K. (2013) Repair of the (64) photoproduct by DNA photolyase requires two photons. *Angew. Chem. Int. Ed.*, **52**, 7432–7436.
25. Dokainish,H.M. and Kitao,A. (2016) Computational assignment of the histidine protonation state in (6-4) photolyase enzyme and its effect on the protonation step. *ACS Catal.*, **6**, 5500–5507.
26. Knips,A. and Zacharias,M. (2017) μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Sci. Rep.*, **7**, 41324.
27. Kim,J.-K., Patel,D. and Choi,B.-S. (1995) Contrasting structural impacts induced by cis-syn cyclobutane dimers and (6-4) adduct in DNA duplex decamers: Implication in mutagenesis and repair activity. *Photochem. Photobiol.*, **62**, 44–50.
28. Kim,J.-K. and Choi,B.-S. (1995) The solution structure of DNA duplex-decamer containing the (6-4) photoproduct of Thymidylyl(3′–>5′)Thymidine by NMR and relaxation matrix refinement. *Euro. J. Biochem.*, **228**, 849–854.
29. Lee,J.-H., Hwang,G.-S. and Choi,B.-S. (1999) Solution structure of a DNA decamer duplex containing the stable 3′ T.G base pair of the pyrimidine(64)pyrimidone photoproduct [(64) adduct]: Implications for the highly specific 3′ T –> C transition of the (64) adduct. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6632–6636.
30. Mizukoshi,T., Kodama,T.S., Fujiwara,Y., Furuno,T., Nakanishi,M. and Iwai,S. (2001) Structural study of DNA duplexes containing the (64) photoproduct by fluorescence resonance energy transfer. *Nucleic Acids Res.*, **29**, 4948–4954.
31. Spector,T.I., Cheatham,T.E. and Kollman,P.A. (1997) Unrestrained molecular dynamics of photodamaged DNA in aqueous solution. *J. Am. Chem. Soc.*, **119**, 7095–7104.
32. Wang,C.I. and Taylor,J.S. (1991) Site-specific effect of thymine dimer formation on dAn.dTn tract bending and its biological implications. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 9072–9076.
33. Cooney,M.G. and Miller,J.H. (1997) Calculated distortions of duplex DNA by a Cis, Syn cyclobutane thymine dimer are unaffected by a 3′ TpA step. *Nucleic Acids Res.*, **25**, 1432–1436.
34. Park,H., Zhang,K., Ren,Y., Nadji,S., Sinha,N., Taylor,J.-S. and Kang,C. (2002) Crystal structure of a DNA decamer containing a cis-syn thymine dimer. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 15965–15970.
35. Ando,H., Fingerhut,B.P., Dorfman,K.E., Biggs,J.D. and Mukamel,S. (2014) Femtosecond stimulated raman spectroscopy of the cyclobutane thymine dimer repair mechanism: a computational study. *J. Am. Chem. Soc.*, **136**, 14801–14810.
36. Masson,F., Laino,T., Tavernelli,I., Rothlisberger,U. and Hutter,J. (2008) Computational study of thymine dimer radical anion splitting in the self-repair process of duplex DNA. *J. Am. Chem. Soc.*, **130**, 3443–3450.
37. Yang,W. (2011) Surviving the Sun: repair and bypass of DNA UV lesions. *Protein Sci.*, **20**, 1781–1789.
38. Ivani,I., Dans,P.D., Noy,A., Perez,A., Faustino,I., Hospital,A., Walther,J., Andrio,P., Goni,R., Balaceanu,A. *et al.* (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **38**, 55–58.
39. Perez,A., Luque,F.J. and Orozco,M. (2012) Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.*, **45**, 196–205.
40. Dumont,E. and Monari,A. (2015) Understanding DNA under oxidative stress and sensitization: the role of molecular modeling. *Front. Chem.*, **3**, 43.
41. Dumont,E., Drsata,T., Fonseca Guerra,C. and Lankas,F. (2015) Insights into the structure of intrastrand cross-link DNA lesion-containing oligonucleotides: G[85m]T and G[85]C from molecular dynamics simulations. *Biochemistry*, **54**, 1259–1267.
42. Dumont,E. and Monari,A. (2013) Benzophenone and DNA: Evidence for a double insertion mode and its spectral signature. *J. Phys. Chem. Lett.*, **4**, 4119–4124.
43. Nogueira,J.J. and Gonzalez,L. (2014) Molecular dynamics simulations of binding modes between methylene blue and DNA with alternating GC and AT sequences. *Biochemistry*, **53**, 2391–2412.
44. Spiegel,K., Rothlisberger,U. and Carloni,P. (2004) Cisplatin binding to DNA oligomers from hybrid car-parrinello/molecular dynamics simulations. *J. Phys. Chem. B*, **108**, 2699–2707.
45. Case,D., Betz,R., Botello-Smith,W., Cerutti,D., III,T.C., Darden,T., Duke,R., Giese,T., Gohlke,H., Goetz,A. *et al.* (2016) *AMBER 16*, University of California, San Francisco, CA.
46. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kale,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
47. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
48. Perez,A., Marchan,I., Svozil,D., Sponer,J., III,T.E.C., Laughton,C.A. and Orozco,M.2007 Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma Conformers. *Biophysical.*, **92**, 3817–3829.
49. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.
50. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
51. Shrake,A. and Rupley,J.A. (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.*, **79**, 351–371.
52. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD—visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
53. Sugita,Y. and Okamoto,Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, **314**, 141–151.
54. Kypr,J., Kejnovska,I., Renciuk,D. and Vorlickova,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.

55. Pasi,M., Maddocks,J.H., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T. III, Dans,P.D., Jayaram,B., Lankas,F., Laughton,C. *et al.* (2014) μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.
56. Dans,P.D., Faustino,I., Battistini,F., Zakrzewska,K., Lavery,R. and Orozco,M. (2014) Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res.*, **42**, 11304–11320.
57. Cogoi,S., Paramasivam,M., Spolaore,B. and Xodo,L.E. (2008) Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Res.*, **36**, 3765–3780.
58. Gaynutdinov,T.I., Neumann,R.D. and Panyutin,I.G. (2008) Structural polymorphism of intramolecular quadruplex of human telomeric DNA: effect of cations, quadruplex-binding drugs and flanking sequences. *Nucleic Acids Res.*, **36**, 4079–4087.
59. Churchill,C.D., Eriksson,L.A. and Wetmore,S.D. (2016) DNA distortion caused by uracil-containing intrastrand cross-links. *J Phys. Chem. B*, **25**, 1195–1204.
60. Sale,J.E., Lehmann,A.R. and Woodgate,R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.*, **13**, 141–152.
61. Courdavaultas,S., Baudouinb,C., Charveronb,M., Canguilhemb,B., Faviera,A., Cadeta,J. and Douki,T. (2005) Repair of the three main types of bipyrimidine DNA photoproducts in human keratinocytes exposed to UVB and UVA radiations. *DNA Repair*, **4**, 836–844.
62. Kuang,Y., Sun,H., Blain,J.C. and Peng,X. (2012) Hypoxia-selective DNA interstrand cross-link formation by two modified nucleosides. *Chemistry*, **18**, 12609–12613.
63. Gentil,A., Page,F.L., Margot,A., Lawrence,C.W., Borden,A. and Sarasin,A. (1996) Mutagenicity of a unique thymine-thymine dimer or thymine-thymine pyrimidine pyrimidone (64) photoproduct in mammalian cells. *Nucleic Acids Res.*, **24**, 1837–1840.

Molecular dynamics insights into polyamines-DNA binding modes: implications for cross-links selectivity.

Emmanuelle Bignon,[*,‡,§] Chen-Hui Chan,[§] Christophe Morell,[‡] Antonio Monari,[‖,⊥] Jean-Luc Ravanat,[¶] and Elise Dumont[*,§]

Institut des Sciences Analytiques, UMR 5280, Université de Lyon1 (UCBL) CNRS, ENS Lyon, Lyon, France, Laboratoire de Chimie, Univ Lyon, Ecole Normale Supérieure de Lyon, CNRS UMR 5182, Université Lyon 1, Laboratoire de Chimie, 46 allée d'Italie, F-69364, LYON, France, Université de Lorraine Nancy, Theory-Modeling-Simulation, SRSMC F-54506 Vandœuvre-lès-Nancy, France, CNRS, UMR 7565, SRSMC, F-54506 Vandœuvre-lès-Nancy, France, and CEA and Université Grenoble Alpes, INAC-SyMMES, F-38000 Grenoble, France

E-mail: emmanuelle.bignon@univ-lyon1.fr; elise.dumont@ens-lyon.fr

Phone: +33 (0)4 72 72 88 46. Fax: +33 (0)4 72 72 88 60

[*]To whom correspondence should be addressed

[‡]Institut des Sciences Analytiques, UMR 5280, Université de Lyon1 (UCBL) CNRS, ENS Lyon, Lyon, France

[§]Laboratoire de Chimie, Univ Lyon, Ecole Normale Supérieure de Lyon, CNRS UMR 5182, Université Lyon 1, Laboratoire de Chimie, 46 allée d'Italie, F-69364, LYON, France

[‖]Université de Lorraine Nancy, Theory-Modeling-Simulation, SRSMC F-54506 Vandœuvre-lès-Nancy, France

[⊥]CNRS, UMR 7565, SRSMC, F-54506 Vandœuvre-lès-Nancy, France

[¶]CEA and Université Grenoble Alpes, INAC-SyMMES, F-38000 Grenoble, France

Running header

Molecular dynamics insights into polyamines-DNA binding modes: implications for cross-links selectivity.

Abstract

Biogenic polyamines are ubiquitous, and are found at millimolar concentration in the nucleus of eukaryotic, playing a role in DNA condensation and stabilisation. The interaction modes of three polyamines: putrescine (Put), spermine (Spm), and spermidine (Spd) with a self-complementary 16-bp duplex is investigated owing to all-atom explicit-solvent molecular dynamics. The length of the amine aliphatic chain leads to a change of the interaction mode from minor groove binding to major groove binding. Through all-atom dynamics, we unravel non-covalent interactions that stabilize the polyamine-DNA complex and prefigure the reactivity leading to the low-barrier formation of deleterious DNA–polyamines cross-links, after one-electron oxidation of a guanine nucleobase. The binding strength is quantified from the obtained trajectories by Molecular Mechanics Generalized Born Surface Area post-processing (MM-GBSA). The values of binding free energies provide the same affinity order Put < Spm < Spd than recent isothermal calorimetry measurements [Kabir et al., PLOS ONE (2013) 8:1–13], with a satisfactory correlation, which validates our structural predictions. The binding modes and carbon–nitrogen distances along the series of polyamine sketch the selectivity towards deleterious DNA-polyamine cross-link formation as we extract average approaching distances between the C8 atom of guanines and the ammonium. Our results imply that the formation of DNA-polyamines cross-links presumably implies a deprotonation of the guanine radical cation to attack onto polyamines, which must be positively-charged to lie in the vicinity of the B-helix.

Introduction

Polyamines are natural small cations that interact strongly with DNA (*1*) and can act as radical scavengers (*2*). Recent evidences have put forward the idea that this protective role (*3*) is somewhat counterbalanced as these nucleophilic species can induce cross–linking reactions with the guanine radical cation obtained after one-electron oxidation (*4*, *5*). This gives rise to toxic DNA–polyamines cross-links (*6*), and it is known that changes in level

of polyamines are associated to aging and related diseases (*7*). The chemical mechanisms underlying condensation and cross-linking has been delineated (*8*) and recently validated by density functional theory on model systems (methylamine) (*9*, *10*). Yet difference of reactivities has been reported between the three well-investigated naturally-occuring polyamines (*4*): putrescine (Put), spermidine (Spd), spermine (Spm) (*11*) represented in Figure 1.

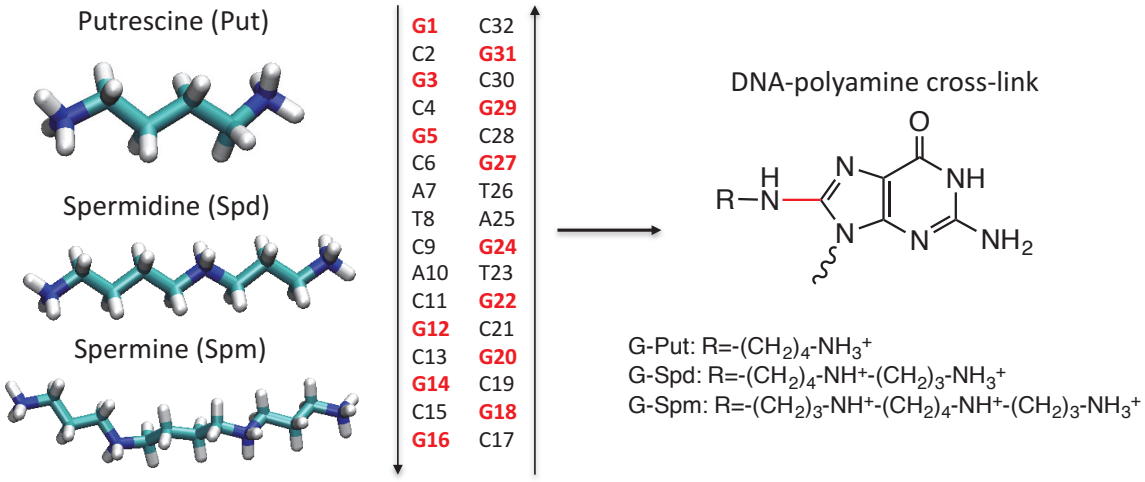


Figure 1: (leftside) Neutral polyamines structures. Simulations were performed on neutral and charged polyamines: putrescine 0,+1,+2 ; spermidine 0,+1,+2,+3 ; spermine 0,+1,+2,+3,+4. The sequence of the 16-bp DNA duplex is given, with guanines being boldfaced. The amino termini were considered as being protonated before the central secondary amines. (rightside) Chemical structure of a DNA–polyamine cross-link: the C8–N covalent bond is displayed in red.

Given the small activation energy needed to induce the formation of cross-links (*9*, *10*), the origin of selectivity may arise from differences in the binding modes strength and topology. In this work, all-atom, explicit-solvent classical molecular dynamics are performed for a 16-bp oligonucleotide d(GCGCGCATCACGCGCG):d(CGCGCGTGATGCGCGC) interacting with the three polyamines: Put, Spd and Spm to identify preferred binding mode(s) and assess the carbon-nitrogen distances between the amino- or ammonium group of the polyamine and the guanine's C8 atoms. Interaction modes between polyamines and DNA have been addressed experimentally by electronic circular dichroism spectroscopy (*12*, *13*), Raman

spectroscopy (*14*, *15*) and NMR (*16*). More recently thermodynamics properties were obtained by isothermal titration calorimetry (ITC) (*17*). All-atom classical molecular dynamics are able to precisely unravel, with an atomic resolution, the peculiar interaction modes between small compounds and biological polymers, and nucleic acids in particular, as recently evidenced for a number of different DNA drugs (*18–22*). Much attention has been paid towards polyamines–DNA interactions with previous MD studies as early as 1989 (*11*), from decamers (*11*, *23–25*) to infinitely-long (*23*, *26*) DNA double-stranded sequences.

In this work, we investigated polyamines binding towards a 16-bp sequence with a longer sampling (150 ns) compared to previous studies, and extract binding free energies based on a MM-GBSA post-processing. Furthermore, we have been able to observe and characterize transitions between competitive binding modes occurring in a scale of tens of nanoseconds. Our results support that a deprotonation of the radical cation of guanine (pKa=3.9 (*27*)) should occur prior to the attack on a positively-charged primary ammonium group (see Figure S1 for the possible mechanism). Indeed, since only the positively-charged ammonium is able to persistently interact with the DNA double strand, the presence of the guanine radical cation will induce instantaneous electrostatic repulsion that will prevent the two fragments to reach sufficiently close distance to allow significant reactivity.

Computational methods

All calculations have been performed using the AMBER12 package (*28*). Parameters for the MD simulations were taken from GAFF and parm99bsc1 force field (*29*). The oligonucleotide was chosen arbitrarily to be a 16-bp B-helix, with a self-complementary d(GCGCGCATCA-CGCGCG):d(CGCGCGTGATGCGCGC) sequence. This choice differs from previous MD studies where shorter oligonucleotides were considered (*24*). Putrescine, spermidine and spermine force field parameters were generated using antechamber and RESP (*30*) protocols, with atom types assigned from GAFF (*31*). For each system, the ligand molecule was initially

placed near the oligonucleotide's major groove. Then, 30 potassium counter ions were added in order to ensure neutrality. The system was placed in an orthorombic TIP3P (*32*) water box. The resulting systems is composed of ca. 29000 atoms including ca. 9200 explicit water molecules. Each structure was minimized in 10000 steps decomposed in 5000 steepest descent steps followed by 5000 conjugate gradient steps. Then, the temperature was increased from 0K to 300K during a 30 ps run, followed by a 1 ns equilibration performed in the NTP ensemble. Finally, the conformational surface of the system was sampled along a 150 ns production run. A cutoff of 9 Å was used for long-range electrostatic effects in order to avoid self-interactions.

Binding free energies were computed using the MM-GBSA module of the Ambertools. 500 snapshots were extracted from the MD simulation trajectory for each system, and energies were determined with the values of external and internal dielectric constants fixed at 80.0 and 1.0 respectively, and the salt concentration set to 0.1 M. The entropy was computed via the quasi-harmonic approximation. We note that MM-GBSA is popular especially to estimate DNA-drug binding (*33–35*), although its complementary MM-PBSA approach is usually slightly favored for proteic systems (*36*).

Meanwhile, a cluster analysis was performed on each system's MD simulation. In order to reduce the calculation time, the interval between considered frames was set to 3. Thus, 5000 structures extracted from the simulation were taken into account in the clustering that was based on the RSMD of the DNA helix and the polyamine, the previous procedure resulted in the production of 30 representative structures. For each of them a Curves+ (*37*) analysis was performed to characterize the B-helix distortion upon polyamine binding. No significant bend or structural signature was identified for most of the systems.

Table 1: Binding free energies $\Delta G^{\text{binding}}$ (kcal/mol), main interaction mode and percentage of prevalence along the 150 ns simulations for each polyamine, according to their protonation state. Experimental free energies are reported for CT-DNA at 293.15 K (*17*).The most probable protonation state at physiological pH is evidenced in bold

	Putrescine	Spermidine	Spermine
0	— No stable interaction	— No stable interaction	— No stable interaction
+1	-5.0±4.1 Major groove binding 63%	-4.2±4.2 Major groove binding 50%	-5.2±6.2 Major/minor groove binding 23/13%
+2	**-7.1±3.7** **Major groove binding** **75%**	-11.7±9.5 Major groove binding 89%	-18.6±9.5 Minor groove binding 74%
+3	/	**-11.1±6.4** **Minor groove binding** **56%**	-18.2±8.9 Minor groove binding 66%
+4	/	/	**-13.7±6.5** **DNA backbone** **62%**
Exp. ()	-6.897±0.011	-7.602±0.019	-8.411±0.013

Results

The results of a series of all-atom, unbiaised 150 ns simulations to investigate the binding mode of Put, Spd and Spm are summarized in Table 1, in which predominant interaction modes are listed alongside with percentage of occurrence. In particular the corresponding binding free energies, evaluated according to a MM-GBSA procedure (see ESI), are reported, too. In order to properly access the DNA-polyamine interactions, we chose to exhaustively test the protonation states of these three polyamines, listed in Figure 1. Obviously and considering their pKa values, polyamines should be protonated at all sites at physiological pH, and we will make this assumption when comparing our computed binding free energies to the experimental values reported by Kabir et al. (*17*). However, when considering the subsequent reactivity leading to the DNA-protein cross-link adducts, it was usually assumed an interaction between the neutral amine and the guanine radical cation. This assumption is also justified considering the complex electrostatic environment the polyamine experiences

as it gets close to the B-DNA helix, that could eventually induce deprotonation along the mechanistic pathway given in Figure S2.

Figures 2, 3 and 4 provide cartoon representations for the significant interaction modes between the 16-bp sequence and Put, Sdp and Spm, that we identified based on a cluster analysis of our dynamics. First of all, none of the uncharged species gave rise to a well-identified stable interaction mode. Only opportunistic interactions take place at the extremities of the 16-bp oligonucleotide, more often for spermidine and spermine than for putrescine – respectively 41% and 49% vs. 13%: cartoon representations are given in Figure S2. The stability of this positioning arises from CH...π bonding that putrescine cannot offer due to its short C4 chain. The driving force for this edge interactions is probably due to the fact that aromatic nucleobases have less affinity for water, which justifies the end-to-end association: however, we cannot assess binding energies as values are probably lower than -1 kcal.mol^{-1}. In the following paragraph, the binding mode of the (poly)cationic polyamines with the 16-bp sequence we chose, is carefully analyzed.

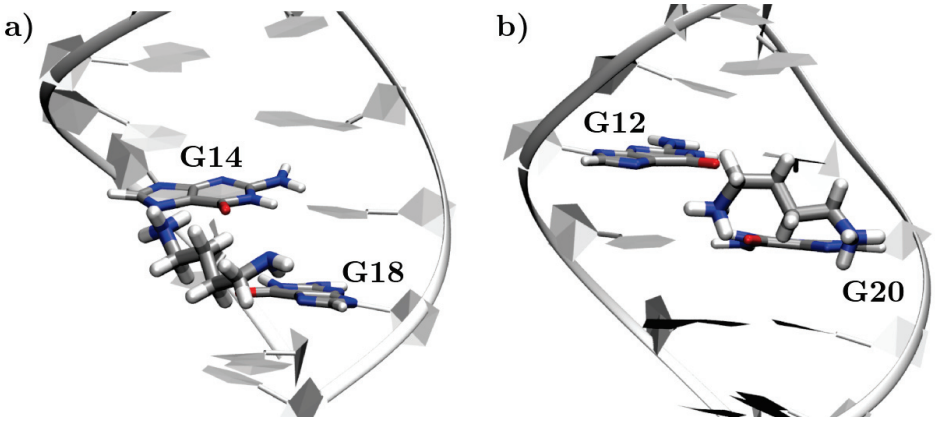


Figure 2: Representative structures of a) +1 and b) +2 putrescine interacting in the oligonucleotide's major groove. These structures were generated by cluster analysis of the trajectories and correspond the representative geometry of the predominant cluster. Putrescine and the nucleic acids with which it interacts are depicted in licorice representation.

Putrescine During the first 30 ns, no stable interaction comes into play between the monocationic putrescine (Put+1) and the ds-DNA duplex. Put+1 remains within the bulk, only

interacting very shortly within the minor groove, getting close to the negatively-charged backbone (see Figure S3). After this regime and until the end of the simulation, the charged ammonium moeity of Put mostly comes to interact with O6 and N7 atoms of G14 and G18 – see Figure 2-a. The neutral amino terminus of Put+1 sometimes lies in the vicinity of G14/C8 and G18/C8, yet without being very stable in time. At ∼65ns of simulation, Put+1 moves to the minor groove in the G14 region for ca. 10 ns, where it interacts with nucleobases by its NH_2 extremity while the NH_3^+ extremity makes hydrogen bonds with the negatively-charged phosphates of the backbone (see Figure S1). Between 98 and 135 ns, Put+1 slides in the major groove, to finally be stabilized back in the G14 and G18 region. In contrast with the previous results, putrescine in the dicationic fully protonated state (Put+2) interacts more firmly and behaves like a more rigid tweezers, although it does not significantly constraint the B-helix (total bend 13.2°). Indeed, comparison between Figures S3-a and S3-b evidences more stable interactions mode as the drug is double protonated. However, in all the cases cationic putrescines Put+1 and Put+2 interact mainly in the major groove (63 and 75% respectively) – see Figure 2. Oppositely, the dicationic species (Put+2) maintains hydrogen bonds that interexchange between the two ammonium termini and guanines' N7/C6 (mainly G12 and G20) more or less all along the simulation. Albeit interactions with G12 and G20 are more stable than for Put+1, with respective binding energies of -5.0 and -7.1 kcal.mol^{-1}, Put+2 goes away from these nucleobases and slides along the DNA backbone through hydrogen bonding to phosphates during ∼30 ns as evidenced in the central region of the simulation (65–95 ns), coming back to its initial location after this period.

Spermidine The monocationic spermidine (Spd+1) develops hydrogen bonds through its positively-charged ammonium moiety, thereby interacting with amino and carbonyl groups of guanines along the 16-bp duplex. Meanwhile the $-NH_2$ extremity interacts transiently with its surroundings such as phosphates as depicted in Figure 3–a. This fact has the consequence to maintain relatively short distances between nitrogen atom and several nucleobases

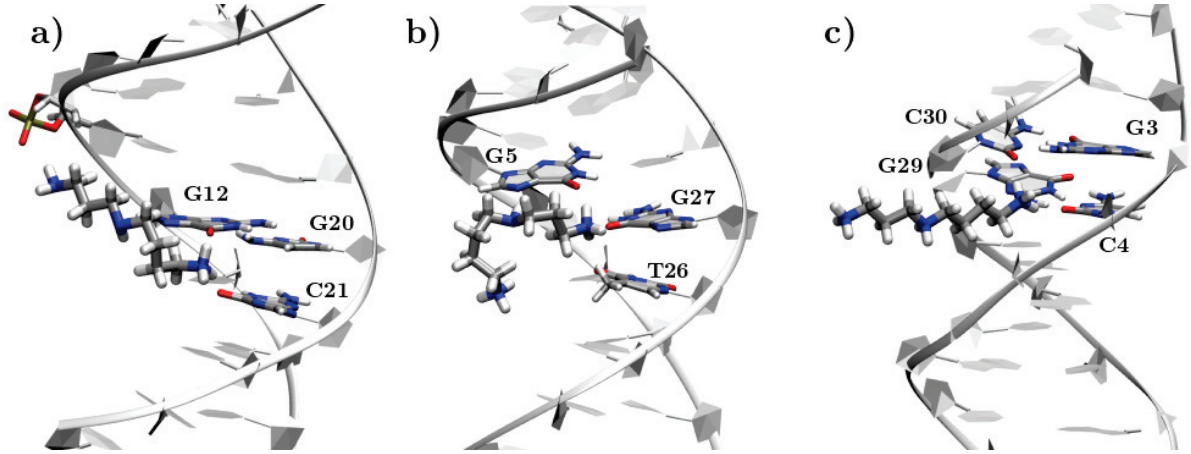


Figure 3: Representative structures of a) +1, b) +2 and c) +3 spermidine interacting in the major groove for +1 and +2 and in the minor groove for +3. These structures were generated by cluster analysis of the trajectories and correspond the representative geometry of the predominant cluster. Spermidine and the nucleic acids with which it interacts are depicted in licorice representation.

C8 position, for instance G12 and G20 are close to Spd+1 during 50% of our simulations according to the cluster analysis results – see Table 1. On the other hand, dicationic, Spd+2 interacts via one of the ammonium group with N7 and O6 atoms of G5 and G27, almost all along the simulation (150 ns, Figure S3-d). In this sense, it acts as a molecular tweezer lying in the major groove. Meantime, the other extremity forms transient hydrogen bonds with the surroundings (backbone and nucleobases), or points towards the bulk, as illustrated in Figure 3–b. Spd+2 lies in the major groove during 89% of the simulation time, giving several opportunities for its amino groups to get close to a guanine's C8 (G5 or G7, 4.8±0.5 $\mathring{A}$).

Spd+3 is prone to form hydrogen bonds with many partners along the double-stranded duplex. It first binds in the major groove, in the G22 region, the three positively-charged amino groups interacting strongly with the backbone and the nucleobases. After ∼25 ns, Spd+3 begins to slide along the double-helix, guided by transient interactions especially along the backbone. At 65 ns, it stabilizes in the minor groove while interacting with G29 and the surrounding nucleobases - see Figure 3. From 105 ns to the end of the simulation,

it moves along the helix, interacting in an opportunistic manner with DNA components. Comparatively to Spd+2, the binding strength is nearly equivalent ($\sim$ -11 kcal.mol^{-1}) .

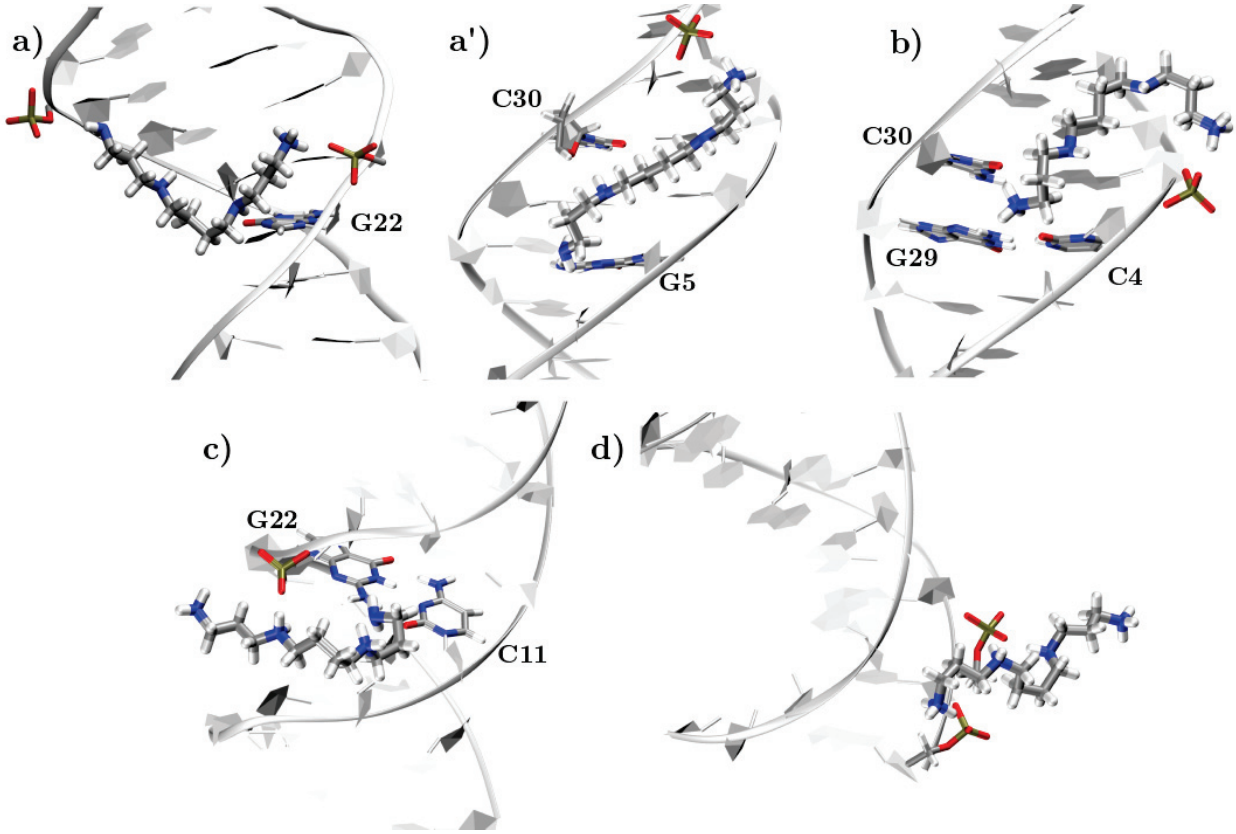


Figure 4: Representative structures of a) +1 in the major groove, a') +1 in the minor groove, b) +2, c) +3, and d) +4 spermine interacting with DNA components. These structures were generated by cluster analysis of the trajectories and correspond the representative geometry of the predominant cluster. Spermine and the nucleic acids with which it interacts are depicted in licorice representation.

Spermine Monocationic spermine (Spm+1) first lies within the minor groove where its –NH_2 extremity is situated in the vicinity of G5, one of the -NH- groups developing hydrogen bonds with C30 and G29, while the ammonium moiety prefers to interact with the negatively-charged backbone. Spm+1 stays in this region for $\sim$15 ns, then leaving to the bulk. It interacts transiently with DNA until it stabilizes in the major groove at 108 ns, in the C23 region (Figure 4–a), where it interacts mostly with the backbone through its amino

extremities, creating a tweezers between the two strands. It can be seen from the cartoon representation that Spm experiences non negligible constraints from the B-helix as it bends to achieve a correct positioning. Consistently, the structural constrains imposed to the DNA by the drug is reflected in the total bend of the helix, which is more marked in that case (18.5°). Probably as a consequence of the strong structural constraints interplay, this position is not enough stable for Spm+1 to stay more than 20 ns: afterwards, at 128 ns, it leaves to the bulk and interacts transiently with the oligonucleotide extremities – see Figure S1.

Spm+2 behaves as a major groove binder during ∼20 ns at the beginning of the simulation. One of its $–NH_3^+$ group induces hydrogen bonds with several guanines while the other one interacts with the surrounding backbone phosphates. Then, it slides along the helix and stabilizes in the minor groove where one of its extremities strongly interacts with C4 and C30 carbonyl groups. The other $–NH_3^+$ weakly interacts with the surrounding phosphates. The two central -NH- moieties interact with the exocyclic amino groups from the proximal guanines. Spm+2 maintains the position until 115 ns, when it slides to the edge of the oligonucleotide, still interacting with the surrounding nucleobases, then slides in the minor groove until the remaining of the simulation (74% minor groove binder in total). The cooperative non-covalent interactions from the two central -NH- moieties leads to a change in the interaction mode (minor vs. major groove for Put+2 and Spd+2 , respectively), and a strong increase of the binding free energy from -5.2 to -18.6 kcal.mol^{-1}. The latter is consistent with the much more persistent interaction mode.

Spm+3 slides along the DNA backbone and quickly stabilizes in the minor groove, interacting with G22. Interestingly, the terminal nitrogen prone to induce cross-linking lies rather close to G12:C8, within ∼3.5 Å during two intervals of 20 ns during the simulation. At 50 ns, Spm+3 moves from this position and slides along the minor groove until the end of the trajectory at 150 ns. During this time, amino groups continue to interact transiently with nucleobases and the backbone. Interestingly, whereas Spm+3 is prone to initiate relatively stable interactions with nucleobases, Spm+4 is not able to do so. On the contrary, it only

slides on the DNA backbone, interacting with the negatively-charged phosphates by its NH_3^+ and NH_2 moieties – see Figure 4. Interestingly, the binding free energy accounts for -18.2 kcal.mol^{-1}, i.e. almost equivalent to the one of -18.6 kcal.mol^{-1} estimated for Spm+2. The association pattern between a polycationic polyamine and DNA is not ruled only by the number of positive charges but is modulated more finely by the tuning of several non-covalent interactions.

Spm+4 is found to interact mostly with the DNA backbone (residency time 62%). Three protonated ammonium groups lie in between the phosphate groups, whereas the remaining–NH_3^+ extremity is directed toward the bulk (Figure 4-d). This is the second case where a significant B-helix bend is predicted (22.4°), possibly because this molecule is long enough to be able to impose a slight structural constraint on DNA. In turn this structural observation are linked to a noticeable decrease of the binding free energies, from -18.2 (Spm+3) to -13.7 kcal.mol^{-1} (Spm+4). Once again this aspect reflects that the affinity towards DNA is not a only a purely electrostatically-driven phenomena, but reflect a more subtle structural interplay. A cross-analysis is led in the following section.

Discussion

Versatility of binding modes and common features DNA–polyamine binding modes reflect a complex pattern of non-covalent interactions: CH...π, electrostatic, hydrogen bonds, $\cdots$ Both the length of the polyamine and its protonation state determine a preference between the major *vs.* minor groove binding, as can be seen from the data collected in Table 1. Unlike other aromatics drugs (*22*), the DNA–drug interaction mode is rather ill-defined and does not present a marked stability because of the lack of π-stacking. For the same reason, insertion or intercalation can be safely ruled out and only groove binding is relevant. As a consequence each of the given polyamines do adopt several binding modes along our 150 ns classical molecular dynamics simulations, in many cases switching from minor to

major groove. Yet the polyamines present interactions towards DNA, which are compatible with a subsequent reactive pathway leading to the formation of DNA–polyamine cross-links. A first protonation is necessary to observe a transient DNA–polyamine binding along our simulations. The mode of interaction can be identified mainly as a major groove binding for Put and Spd, with an occurrence of 63% and 50%, respectively. Spm presents a longer aliphatic chain, and though it can act as a major groove binder (23%), it needs to adopt a tweezers conformation allowing the simultaneous interaction with the two phosphate groups. Spm interacts also as a minor groove binder as can be seen in Figures 4-a and 4-a': however this accounts only for 13%. The rest of the time, the drug lies at a higher distance to DNA, and hence can be considering as externally bound or even non-interacting. As seen in Figure 3-a the Spm $–NH_3^+$ terminus lies in between the two carbonyl groups of the guanine G12 and cytosine C21. Interestingly, both $–NH_2$ and $–NH_3^+$ termini present affinities for guanine, which is confirmed as the prefered interaction hotspot. This is a critical feature since a proximity of the polyamine to guanine is a prerequisite to the eventual DNA cross-linking. Hence, we may hypothesize that the guanine radical cation should be deprotonated upon its formation leaving a sufficiently long-lived neutral radical that will promote interaction with the charged polyamines. Furthermore, the observed selectivity, and in particular the absence of any interactions in the adenine/thymine area of our duplex, is also in agreement with previous studies concluding that GC regions are targeted preferentially over AT ones (*14*).

Similarly a second protonation contributes to further reinforce the interaction and the residency time of polyamines towards DNA. The second $–NH_3^+$ moiety seeks a new partner to develop hydrogen bond. Hence, it may interact with the oxygen O6 of a guanine. In the case of Put+2, this is realized with the oxygen of G20 (See Figure 2-b), while Spd+2 (Figure 3-b) and Spm+2 (Figure 4-b) interact preferentially with G27 and G29, respectively. Especially in the case of the latter compound, the aliphatic chain is long enough to also afford transient interactions with proximal phosphate(s). More generally, it appears clearly that the length

of the aliphatic chain is a key feature determining the specific interaction mode, together with its interplay with the global charge of the drug. Indeed, minor groove binding mode appears to be favored for the longest aliphatic side chain (Spm) and/or for highly charged polyamine (Spd+3). However, one has to take into account that our simulation have been conducted at minimal salt concentration, physiologically ionic strength can indeed modify the binding affinity for the groove. However, our results globally indicate that the main interactions taking place are between the positively charged amine and the negative backbone posphates. This is also in agreement with recent experimental evidence that have shown a decrease in cross-link adducts formation when operating in a posphate buffer hence pointing to a competition between the interaction with DNA backbone.

Binding free energies To be more quantitative and to directly compare our results with recent experimental data (*17*), we extracted binding free energies relying on the MM-GBSA post-analysis (see Table 1). Globally, our ranking is in very good agreement with recent isothermal titration calorimetry measurements (*17*) which report the affinity order to be: Put $<$ Spm $<$ Spd for long B-DNA oligonucleotides. Interestingly the same order is also found for tRNA sequences (*38*). If we assume, based on pKa arguments, that all the polyamines should be fully protonated, our values lie in good agreement with the experimental measurements at 293.15 K for CT-DNA, as reported in Figure 5.

An overestimation of the calculated free energies can be observed for Spm and Spd. Indeed, while our values are predicted to range from -5 and -20 kcal.mol^{-1} the experimental ones span a much more limited range experimentally: -6.9 to -8.4 kcal.mol^{-1} (see last line of Table 1 and Figure 5). However, one has to take into account also the fact that our simulations predict the polyamines to span different binding modes. Hence, this polymorphism results in relatively large error bars, as reported in Figure 5. If we take into account the incertitude related to the difficult sampling of the system, the experimental values fall well in the predicted range. This aspect is in sharp contrast with the more usual case of protein/drug interactions in which the

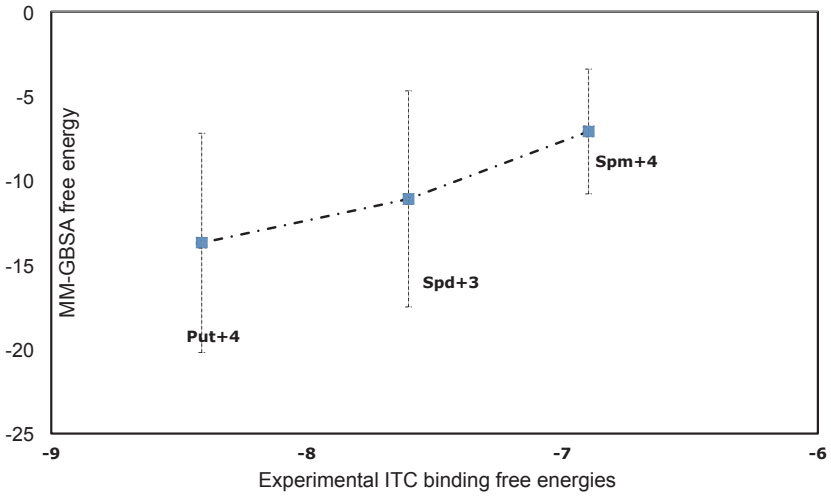


Figure 5: Relation between experimental binding free energies (*17*) and computed in this work based on MM-GBSA. Error bars are indicated for the values computed.

presence of a well defined active site or interaction pocket strongly limit the flexibility and the degrees of freedom of the system, this aspect was also addressed for DNA/benzophenone using more elaborated methods (*39*). Furthermore, one has also to take into account the fact that we are addressing here the binding free energies of increasingly charged ligands, and hence approaching the limits of the rather simple MM-GBSA approach, that is, most notably, not taking into account the solvent explicitly. Indeed, even if possible in principle resorting to more sophisticated schemes such as explicit-solvent, biased techniques, like free energy perturbation and alchemical transformations useful for DNA–drug systems (*20*, *39*). They would indeed be here complicated by the definition of the restraints and the complex equilibrium between different binding modes.

Implications for DNA–polyamine cross-link formation The initial motivation for this work was to assess the binding modes and provide a clear-cut picture of the ease of approach between the C8 atom of guanine embedded in a 16-bp duplex and the terminal nitrogen atom of a polyamine. Indeed, the reactivity of polyamines has been studied with methylamine as a model for amino-groups and compared to water (*9*): Burrows and cowork-

ers have reported low barriers values for the attack. The most common mechanisms of attack are recalled in Figure S1, and may proceed either via the guanine radical cation or its deprotonated form.

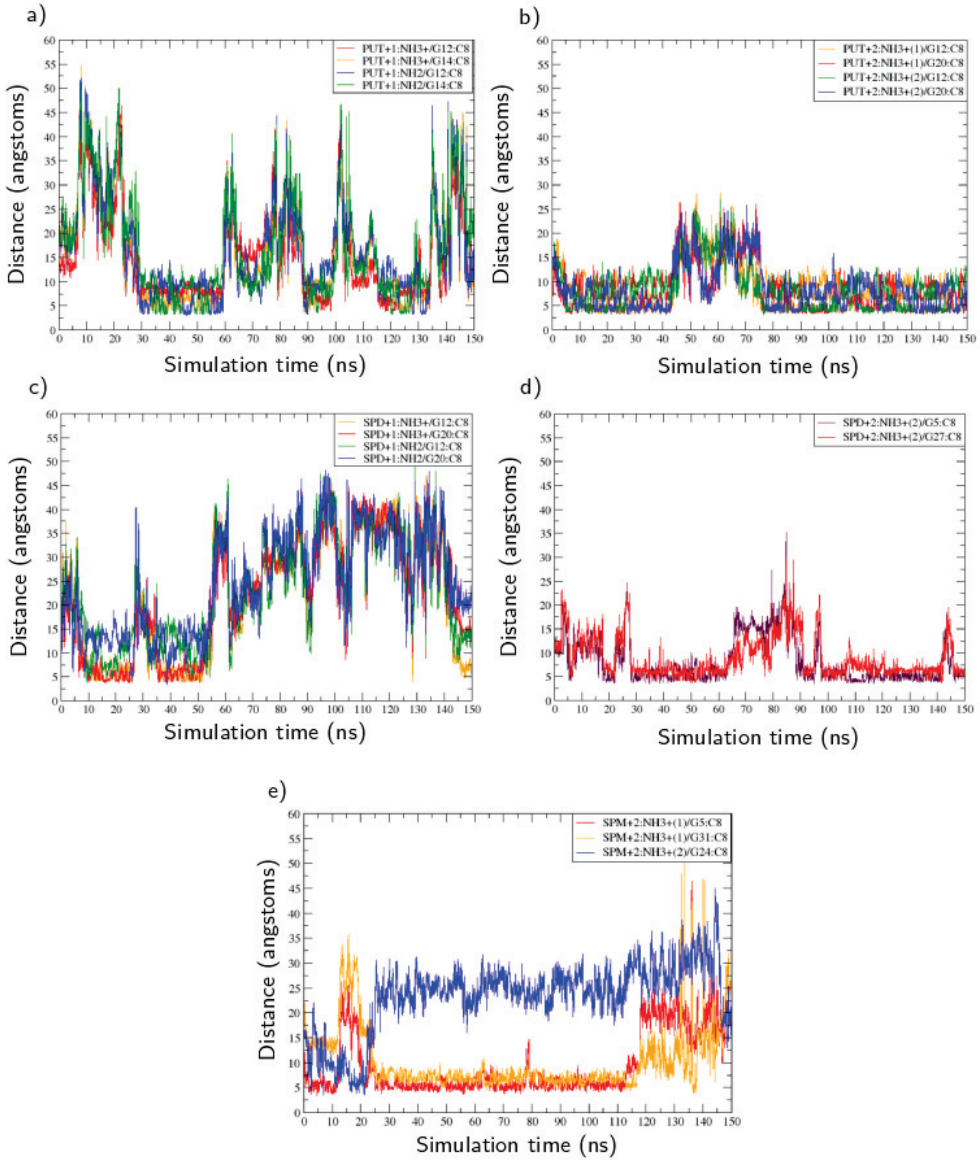


Figure 6: Time evolutions of interatomic distances between polyamine terminal nitrogens and the C8 atoms of proximal guanines for Put+1/+2 (a)/(b), Spd+1/+2 (c/d), and Spm+2 (e).

The low activation energies imply that the binding mode is most probably a decisive feature determining the reactivity and the selectivity. Indeed, the postulate is that the shorter

the approach between the guanine and the polyamine, the more likely the formation of a DNA–polyamine cross-link. In Figure 6, we report relevant distances that we monitored between a given positively-charged polyamine amino groups and the C8 atom of the proximal guanines along the simulation. It is evident that in all the cases the dication produces much closer aggregates than the corresponding monocationic species. Indeed, while Put+1 and Spd+1 exhibit a large deviation of the nitrogen–C8 distance with values as large as 50Å, the corresponding +2 species have distances close to 5 Å over a large segment of the trajectory. This more pronounced stability is certainly due to the formation of more stable hydrogen bonds. On the other hand the tricationic species Spm+3 is less prone to form persistent interactions with the guanines, and indeed the two groups stay in close proximity only during ∼20ns. On the other hand, as already evidenced in the previous sections, Spm is rather peculiar and gives rise to less stable and persistent aggregates; more particularly Spm+1, +3 and +4 do not stay longer than 4 ns next to a guanine C8 atom. The dicationic Spm+2 species (Figure 6-e) follows globally the same tendencies as sketched out for the other polyamines with at least two nitrogens persistently residing in very close contact with guanines.

Globally we may evidence that the protonation state, and hence the positive charge, of polyamines is not only dictated by thermodynamic pKa considerations but is also a requirement to induce both stable interactions with the DNA duplex and a close proximity between the guanine C8 and the nitrogen that is necessary to insure the following reactivity. As a consequence, and considering that the approach of two positively-charged ions will be energetically unfavorable, we may suggest, that at least in DNA environment, the cross-linking should proceed via a two step mechanisms: First a deprotonation of the guanine radical cation, followed by the interaction and reaction with the polyamine. The latter scenario is also comforted by the value of the pKa for the $G^{+\bullet}$ that is estimated to be of 3.9. (*27*)

Conclusion

The rationalization of the inherently complex interaction pattern between DNA and polyamines is an important subject of study (*40*), where clear-cut thermodynamical insights are sought. This is particularly true since DNA-polyamine cross-links occur in biological environment and is known to generate extremely harmful adducts. Furthermore, this rather simple phenomenon can also be regarded as a model for the study of the complex DNA-protein cross-link phenomena. We report the first MD investigation beyond 100 ns, where free energies are estimated: the correlation with experimental data is satisfactory and correctly reproduce the observed affinity order. We have proven that the interaction mode is highly complex and usually results in a fine equilibrium between minor and major groove binding. Furthermore, although the presence of positive charges is essential we also evidenced the complex interplay with the aliphatic chain length in tuning the interaction strength and selectivity. It has to be reminded here that we only considered one polyamine, i.e. we neglected aggregation phenomena. The latter is however known to induce crystallization as proven by other studies (*41*). The analysis of the distances between guanine and nitrogen atoms suggests that polycationic drugs are required to assure a sufficiently high encounter probability of the cross-link reactants. As a consequence we suggest that in the ds-DNA environment the reaction should proceed after a preliminary deprotonation of the guanine radical cation necessary to avoid electrostatic clash between two positively-charged moiety. In future studies we plan to tackle reactivity aspects where the formation of DNA–polyamine cross-links will be challenged owing to hybrid QM/MM-MD simulations for DNA–polyamine systems and nucleosomal DNA, where the compact structure may induce structural constraints limiting the accessibility of the polyamine. However in the case of nucleosomal DNA the reactivity of polyamine may also be in competition with the formation of DNA/protein cross-links in particular due the high density of charged lysines in the nucleosomal core and tails. (*42*, *43*)

References

1. Balasundaram, D., and Tyagi, A. K. (1991) Polyamine — DNA nexus: structural ramifications and biological implications. *Molecular and Cellular Biochemistry 100*, 129–140.

2. Ha, H. C., Sirisoma, N. S., Kuppusamy, P., Zweier, J. L., Woster, P. M., and Casero, R. A. (1998) The natural polyamine spermine functions directly as a free radical scavenger. *Proceedings of the National Academy of Sciences 95*, 11140–11145.

3. Thierry Douki, J. C., Yann Bretonniere (2000) Protection against Radiation-Induced Degradation of DNA Bases by Polyamines. *Radiation Research 153*, 29–35.

4. Silerme, S., Bobyk, L., Taverna-Porro, M., Cuier, C., Saint-Pierre, C., and Ravanat, J.-L. (2014) DNA-Polyamine Cross-Links Generated upon One Electron Oxidation of DNA. *Chemical Research in Toxicology 27*, 1011–1018, PMID: 24798911.

5. Gomez-Mendoza, M., Banyasz, A., Douki, T., Markovitsi, D., and Ravanat, J.-L. (2016) Direct Oxidative Damage of Naked DNA Generated upon Absorption of UV Radiation by Nucleobases. *The Journal of Physical Chemistry Letters 7*, 3945–3948, PMID: 27642654.

6. Pegg, A. E. (2013) Toxicity of Polyamines and Their Metabolic Products. *Chemical Research in Toxicology 26*, 1782–1800, PMID: 24224555.

7. Nadège Minois, F. M., Didac Carmona-Gutierrez (2011) Polyamines in aging and disease. *Aging 3*, 716—732.

8. Xu, X., Muller, J. G., Ye, Y., and Burrows, C. J. (2008) DNA-Protein Cross-links between Guanine and Lysine Depend on the Mechanism of Oxidation for Formation of C5 Vs C8 Guanosine Adducts. *Journal of the American Chemical Society 130*, 703–709, PMID: 18081286.

9. Thapa, B., Munk, B. H., Burrows, C. J., and Schlegel, H. B. (2016) Computational Study of the Radical Mediated Mechanism of the Formation of C8, C5, and C4 Guanine:Lysine Adducts in the Presence of the Benzophenone Photosensitizer. *Chemical Research in Toxicology 29*, 1396–1409, PMID: 27479718.

10. Thapa, B., Munk, B. H., Burrows, C. J., and Schlegel, H. B. (2017) Theoretical Study of Oxidation of Guanine by Singlet Oxygen (1Δg) and Formation of Guanine:Lysine Cross-Links. *Chemistry – A European Journal* n/a–n/a.

11. Burt G.Feuerstein, N. P., and Marton, L. J. (1989) Molecular dynamics of spermine-DNA interactions : sequence specificity and DNA bending for a simple ligand. *Nucl. Acids. Res. 17*.

12. Hilde Damaschun, M.-E. B. R.-M., G.Damaschun, and D.Zirwer, (1978) Study of DNA-spermine interactions by use of small-angle and wide-angle X-ray scattering and circular dichroism. *Nucleic Acids Research 5*.

13. Ruiz-Chica, A., Medina, M., Sánchez-Jiménez, F., and Ramírez, F. (2005) Study by electronic circular dichroism spectroscopy of the interaction between aminooxy analogues of biogenic polyamines and selected oligonucleotides. *Journal of Molecular Structure 744–747*, 691 – 698, {MOLECULAR} {SPECTROSCOPY} {AND} {MOLECULAR} {STRUCTURE} 2004A Collection of Papers Presented at the {XXVIIth} European Congress on Molecular Spectroscopy, Krakow, Poland, September 5-10, 2004.

14. Deng, H., Bloomfield, V. A., Benevides, J. M., and Jr, G. J. T. (2000) Structural basis of polyamine–DNA recognition: spermidine and spermine interactions with genomic B-DNAs of different GC content probed by Raman spectroscopy. *Nucleic Acids Research 28*, 3379–3385.

15. Ruiz-Chica, J., Medina, M., Sánchez-Jiménez, F., and Ramírez, F. (2001) Fourier Trans-

form Raman Study of the Structural Specificities on the Interaction between {DNA} and Biogenic Polyamines. *Biophysical Journal 80*, 443 – 454.

16. van Dam, L., Korolev, N., and Nordenskiöld, L. (2002) Polyamine–nucleic acid interactions and the effects on structure in oriented DNA fibers. *Nucleic Acids Research 30*, 419–428.

17. Kabir, A., and Suresh Kumar, G. (2013) Binding of the Biogenic Polyamines to Deoxyribonucleic Acids of Varying Base Composition: Base Specificity and the Associated Energetics of the Interaction. *PLOS ONE 8*, 1–13.

18. Špačková, N., Cheatham, T. E., Ryjáček, F., Lankaš, F., van Meervelt, L., Hobza, P., and Šponer, J. (2003) Molecular Dynamics Simulations and Thermodynamics Analysis of DNA-Drug Complexes. Minor Groove Binding between 4',6-Diamidino-2-phenylindole and DNA Duplexes in Solution. *Journal of the American Chemical Society 125*, 1759–1769, PMID: 12580601.

19. Ziebarth, J., and Wang, Y. (2009) Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophysical Journal 97*, 1971 – 1983.

20. Wilhelm, M., Mukherjee, A., Bouvier, B., Zakrzewska, K., Hynes, J. T., and Lavery, R. (2012) Multistep Drug Intercalation: Molecular Dynamics and Free Energy Studies of the Binding of Daunomycin to DNA. *Journal of the American Chemical Society 134*, 8588–8596, PMID: 22548344.

21. Dumont, E., and Monari, A. (2013) Benzophenone and DNA: Evidence for a Double Insertion Mode and Its Spectral Signature. *The Journal of Physical Chemistry Letters 4*, 4119–4124.

22. Dumont, E., and Monari, A. (2015) Interaction of Palmatine with DNA: An Environmentally Controlled Phototherapy Drug. *The Journal of Physical Chemistry B 119*, 410–419, PMID: 25526561.

23. Korolev, N., Lyubartsev, A. P., Laaksonen, A., and Nordenskiöld, L. (2003) A molecular dynamics simulation study of oriented DNA with polyamine and sodium counterions: diffusion and averaged binding of water and cations. *Nucleic Acids Research 31*, 5971.

24. Brysont, K., and Greenall, R. J. (2000) Binding Sites of the Polyamines Putrescine, Cadaverine, Spermidine and Spermine on A- and B-DNA Located by Simulated Annealing. *Journal of Biomolecular Structure and Dynamics 18*, 393–412, PMID: 11149516.

25. Dai, L., Mu, Y., Nordenskiöld, L., and van der Maarel, J. R. C. (2008) Molecular Dynamics Simulation of Multivalent-Ion Mediated Attraction between DNA Molecules. *Phys. Rev. Lett. 100*, 118301.

26. Korolev, N., Lyubartsev, A. P., Laaksonen, A., and Nordenskiöld, L. (2004) A molecular dynamics simulation study of polyamine– and sodium–DNA. Interplay between polyamine binding and DNA structure. *European Biophysics Journal 33*, 671–682.

27. Rokhlenko, Y., Cadet, J., Geacintov, N. E., and Shafirovich, V. (2014) Mechanistic Aspects of Hydration of Guanine Radical Cations in DNA. *Journal of the American Chemical Society 136*, 5956–5962, PMID: 24689701.

28. Case, D. et al. (2012), AMBER 12, University of California, San Francisco.

29. Ivan Ivani, A. N.-A. P. I. F. A. H. J. W. P. A. R. G. A. B. G. P. F. B. J. L. G. C. G. M. V. C. A. L. S. A. H. D. A. C., Pablo D Dans, and Orozco, M. (2016) Parmbsc1: a refined force field for DNA simulations. *Nature Methods 38*, 55–58.

30. Wang, J. M., Cieplak, P., and Kollman, P. A. (2000) How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem. 21*, 1049–1074.

31. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem. 25*, 1157–1174.

32. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys. 79*, 926–935.

33. Beuerle, M. G., Dufton, N. P., Randi, A. M., and Gould, I. R. (2016) Molecular dynamics studies on the DNA-binding process of ERG. *Mol. BioSyst. 12*, 3600–3610.

34. Srivastava, H. K., and Sastry, G. N. (2013) Efficient estimation of MMGBSA-based BEs for DNA and aromatic furan amidino derivatives. *Journal of Biomolecular Structure and Dynamics 31*, 522–537, PMID: 22877232.

35. Niu, Y., Pan, D., Shi, D., Bai, Q., Liu, H., and Yao, X. (2015) Influence of Chirality of Crizotinib on Its MTH1 Protein Inhibitory Activity: Insight from Molecular Dynamics Simulations and Binding Free Energy Calculations. *PLOS ONE 10*, 1–15.

36. Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling 51*, 69–82, PMID: 21117705.

37. Lavery, R. et al. (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Research 38*, 299–313.

38. Ouameur, A. A., Bourassa, P., and Tajmir-Riahi, H.-A. (2010) Probing tRNA interaction with biogenic polyamines. *RNA 16*, 1968–1979.

39. Gattuso, H., Dumont, E., Chipot, C., Monari, A., and Dehez, F. (2016) Thermodynamics of DNA: sensitizer recognition. Characterizing binding motifs with all-atom simulations. *Phys. Chem. Chem. Phys. 18*, 33180–33186.

40. Lindemose, S., Nielsen, P. E., and Møllegaard, N. E. (2005) Polyamines preferentially interact with bent adenine tracts in double-stranded DNA. *Nucleic Acids Research 33*, 1790.

41. Iacomino, G., Picariello, G., and D'Agostino, L. (2012) DNA and nuclear aggregates of polyamines. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1823*, 1745 – 1755.

42. Angelov, D., Charra, M., Müller, C. W., Cadet, J., and Dimitrov, S. (2003) Solution Study of the NF-kB p50-DNA Complex by UV Laser Protein–DNA Cross-linking. *Photochemistry and Photobiology 77*, 592–596.

43. Tretyakova, N. Y., Groehler, A., and Ji, S. (2015) DNA–Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. *Accounts of Chemical Research 48*, 1631–1644, PMID: 26032357.

Chapter 4

DNA repair

DNA damage repair is essential to maintain the genomic integrity, thus the biological system's perennity. This field has been intensively investigated since several decades, and nowadays we are able to globally understand the processes that are involved in such a complex machinery. Pioneers in this research field have been rewarded by the Nobel Price of chemistry in 2015 : Tomas Lindahl, Paul Modrich, and Aziz Sancar for their work on base excision repair, mismatch repair and chemically complex lesions repair respectively. These processes are highly complex : according to the case, several different repair pathways can come into play, in which numerous enzymes are recruited depending on the lesion type and the organism's nature - eukaryote or bacteria. Up to now, six different pathways have been identified : Single Strand Break Repair (SSBR), Base Excision Repair (BER), Nucleotide Excision Repair (NER), Mismatch Repair (MMR), Double Strand Break Repair (DSBR), and Interstrand CrossLinks Repair (ICLR)[13]. The large amount of work that has been produced in order to understand theses processes' mechanisms led to the understanding of their principal features, especially thanks to the resolution of enzymes NMR and X-ray structures. However, not all the subtle aspects driving them are not yet elucidated.

Albeit DNA damage repair has been intensively investigated since few decades, a lot of questions still remain pending. More especially, the reaction mechanisms, the enzymes-DNA interplay, as well as the way subtle effects (eg. sequence modifications) impacts damaged DNA processing by the repair machinery are still matter of investigations. The main picture has been delineated but details are missing. Biologists and chemists have provided and still provide each year an unprecedent amount of data about DNA repair mechanisms, but there remains a lack of information about mechanisms and structures of DNA complex lesions at the molecular resolution. Complementary to experimental chemistry, molecular modelling has been proved to be a powerful and versatile tool in order to gain insights into structural and reactivity features of such processes at the atomic scale, using molecular dynamics and QM/MM computations. The cross-talk between theoreticians and experimentalists is of major importance to reveal each details of the wide and complex picture that is DNA repair.

In order to unravel some of these pending questions, we focused on two enzymes that are involved in the BER pathway - see section 4.2.1 : a DNA-glycosylase and an Ap-endonuclease. Indeed, we investigated the structural behavior of the Fpg glycosylase and the APE1 Ap-endonuclease enzymes in interaction with complex oxidatively-induced DNA lesions. Results of these investigations gave us as the two papers available at the end of this chapter - one published, one under redaction whereas results will be discussed.

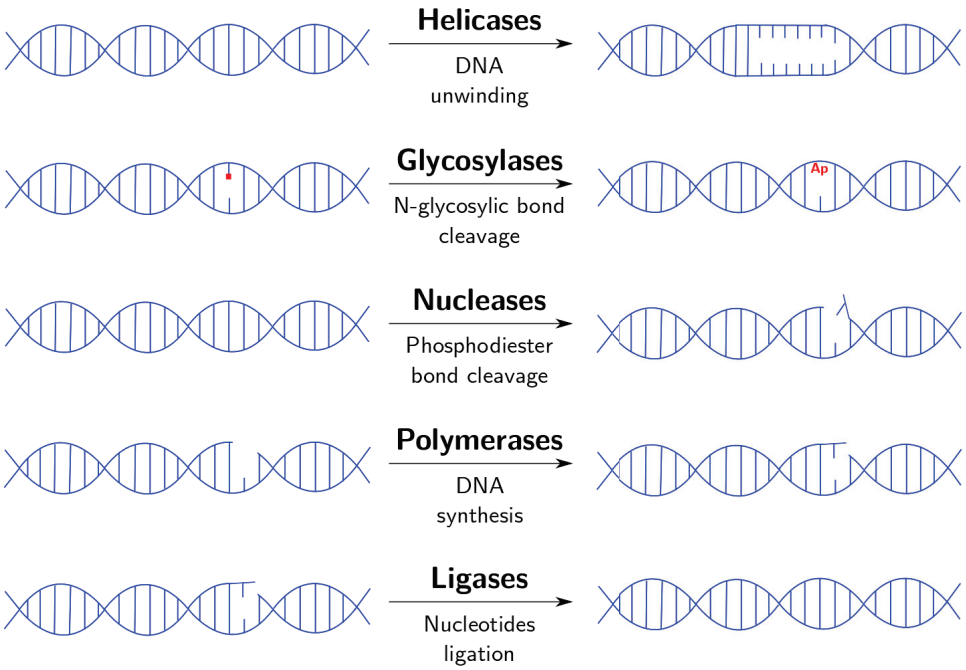


FIGURE 4.1.1: Principal enzymatic activities involved in DDR. First line : helicases separate DNA strands. Second line : glycosylases remove damaged nucleobases (represented by the red square) resulting in an Ap site formation. Third line : nucleases are able to cleave DNA backbone 5' or 3' of a damaged site. Fourth line : polymerases are responsible for DNA synthesis by complementarity with the opposite strand - here a single-nucleotide gap. Fifth line : ligases perform the 5'-phosphate / 3'-OH ligation.

In the following is summed up on one hand the two investigations we led in this framework, and on the other hand a description of the background knowledge concerning DNA repair enzymes and pathways in order to delianeate our work's context.

4.1 Enzymatic functions involved in DNA repair

4.1.1 Helicases : DNA unwinding

Helicases catalyze double stranded DNA unwinding (see Figure 4.1.1), which occurs in an adenosine triphosphate (ATP) -dependent manner. Such enzymes are able to translocate along one of the DNA (or RNA) strand in a specific direction (5' to 3' or the reverse)[171], allowing to unwind the double helix while sliding on it.
The number of DNA-helicases was estimated in 2011 to 31 in the human genome[172], and exhibit a large variety of natures and functions. For instance, some are specialized to unwind G-quadruplex (G4) DNA, telomeric displacement loops (T-loops), or other forms of DNA structures[173,174]. Mechanisms that describe this process are not completely understood, but it appears that ATP binding to the helicase favors its translocation activity, while ATP hydrolysis drives the dsDNA unwinding[175].

4.1.2 Glycosylases : N-glycosylic bonds cleavage

Nowadays, glycosylases are the best understood DNA repair enzymes, because their structural/mechanistic features and functions have been widely investigated since the 70's. Moreover, this strong knowledge comes from the important number of glycosylases crys-

tal structures that have been resolved, with or without their respective substrates[176–181]. Glycosylases are responsible for N-glycosylic bonds cleavage (bond between the nucleobase and the desoxyribose), resulting in an abasic site (Ap) formation - see Figure 4.1.1. Norteworthy, these enzymes are the first protagonists acting during the Base Excision Repair pathway which is commonly used for DDR of modified nucleobases[182].

DNA glycosylases are damage-specific. Thus, the nature of the enzyme that will intervene during DDR is driven by the nature of the lesion. For instance, the hOGG1 and its bacterial homolog Fpg repair 8-oxo-7,8-desoxyguanosine[183,184], while uracil glycosylase removes mismatched U:G base pair uracil produced by deamination of a cytosine[185]. Damage recognition is ensured by random diffusion of the glycosylase along DNA and amino acids insertion into the double-helix[186]. Glycosylases action mechanisms depend on their mono- or bifunctional feature - see section 4.2.1.

4.1.3 Nucleases : phosphodiester bonds cleavage

Nucleases are responsible for DNA-backbone cleavage at the damage site. More precisely, they break the phosphodiester bond between two adjacent nucleotides, resulting in a single strand break - see Figure 4.1.1. Two subclasses of nucleases can be distinguished : endo- and exo-nucleases. The first ones are able to cleave bonds within the polynucleotide chain, whereas the second ones can only process from the extremity of a strand. Among all the nuclease enzymes, the Ap-endonucleases are one of the cornerstones of DDR since they are able to repair the very abundant abasic sites which are, inter alia, one of the Base Excision Repair intermediates - see section 4.2.1. These enzymes cleave the phosphodiester bond at the 5' end of Ap sites, resulting in a free 3'-OH end. They induce bending and kinking of the damaged DNA double-helix, that favors Ap site flipping towards the enzyme's active site, accompanied by the insertion of two loops into the gap generated by the nucleobase absence. The small size of the active site prevents from binding to normal nucleobases. The mechanisms of these enzymes necessitate the intervention of a metal cation, a Mg^{2+} for APE1 that intervenes to stabilize the negatively-charged O3' produced by the cleavage.

4.1.4 Polymerases : DNA synthesis

DNA-polymerases' role is to incorporate correct nucleobases in replacement of damaged ones during DNA repair - see Figure 4.1.1. For this purpose, they possess two enzymatic activities : DNA synthesis and deoxyribose phosphate (dRP) lyase. The lyase activity is used to modify 5'-dRP moieties in 5'-phosphates in order to allow the damaged base replacement. Thus, the system exhibits a 5'-phosphate end and a gap, which length varies according to the case (eg a single-nucleotide gap in BER and a 22-30 nucleotides one in NER), the polymerase coming to fill this gap with canonical nucleobases - see sections 4.2.1 and 4.2.3.

For this process to take place, the polymerase binds to DNA in a first step, then to a nucleotide triphosphate (dNTP) which matches with the coding base, preserving Watson-Crick pairing[187]. The enzymatic mechanisms that take place involve two Mg^{2+} ions : the first one coordinates dNTP and a second one comes to stabilize the system. The polymerase's product is a nicked DNA with a 5'-phosphate and a 3'-hydroxyl group, prone to be processed by a DNA ligase to complete the damage repair. Among the several types of polymerases, only the polymerases β (pol β), δ and ϵ are involved in DDR.

4.1.5 Ligases : nucleotides ligation

Ligases come into play in the ultimate step of DNA damage repair, recombination and replication. It seals the nick between a 3'-hydroxyl and a 5'-phosphate in a three steps enzymatic process that require the intervention of a ATP (or NAD^+) cofactor[188] - see Figure 4.1.1. Several eukariotic and bacterial ligases structures are available[188–191], that bring informations about structural features of the ligation process.

4.2 DNA repair pathways

4.2.1 Base Excision Repair

Base excision repair is a simple multistep process involving several enzymes. It is used for the repair of chemically modified nucleobases and abasic sites, whose structural impact on DNA duplex involves only minor distortions[192]. This process can be defined in two main stages that depend on the mono- or bi-functional feature of the glycosylase, and is concluded by two steps of single strand break (SSB) repair - see Figure 4.2.1.

On the one hand, the monofunctional glycosylases such as Fpg exhibit a hydrolase activity, that is essential to cleave the N-glycosylic bond of a damaged nucleobase with the help of a water molecule, resulting in an abasic site[193] that will be processed by a DNA nuclease.
On the other hand, bifunctional glycosylases such as Endonuclease III[194] offers, in addition to glycosylase properties, an Ap-lyase activity. It uses an amino group rather than a water molecule during the N-glycosylic bond cleavage, resulting in a covalent imino intermediate[195]. This latter undergoes a β, δ-elimination and a Schiff base hydrolysis, resulting in a single strand break at the 3' end of the damaged site, exhibiting an α,β-unsaturated aldehyde moiety. This product is then processed by a 3'-phosphodiesterase in order to modify the 3'-phosphate in a 3'-hydroxyl - see Figure 4.2.1. Once the 3'-OH end is obtained, the system is processed as a SSB (see section 4.2.2) : gap-filling by a DNA-polymerase and ligation by a DNA-ligase.

4.2.2 Single Strand Break Repair

Single Strand Breaks (SSBs) can be generated by several causes : ROS attack, as an intermediate in BER (see section4.2.1), and even by erroneous activity of cellular enzymes[196]. They are usually repaired in four steps : SSB detection, DNA end processing, gap filling and ligation that are depicted in Figure 4.2.2.
SSBs presence in the cell activates the PARP1 enzyme, that binds to the damaged site and recuits repair proteins[197]. In some cases, the detected SSB exhibits 5'- and 3'-ends degradation, and the nature of the damaged termini will determine which enzyme will be recruited for processing - polynucleotide kinase 3'-phosphatase (PKNP) for 3'-end phosphorylation, Aprataxin (APTX) that processes the 5'-AMP ends, Pol β that cleaves the 5'-dRP end, etc. It is noteworthy that XRCC1, which is also recruited by PARP1, stimulates these repair enzymes' activity by interacting with them. Processing of the degrated DNA ends provides 5'-phosphate and 3'-OH ends that are essential for the next step : gap filling. This stage is performed by a polymerase (β, δ or ϵ), and can occur in two ways. First, the short-patch repair where the missing nucleotide is basically inserted. Second, the long-patch repair in which the gap-filling concerns multiple nucleotides with assitance of the proliferating cell nuclear antigen (PCNA), a sliding clamp loaded onto

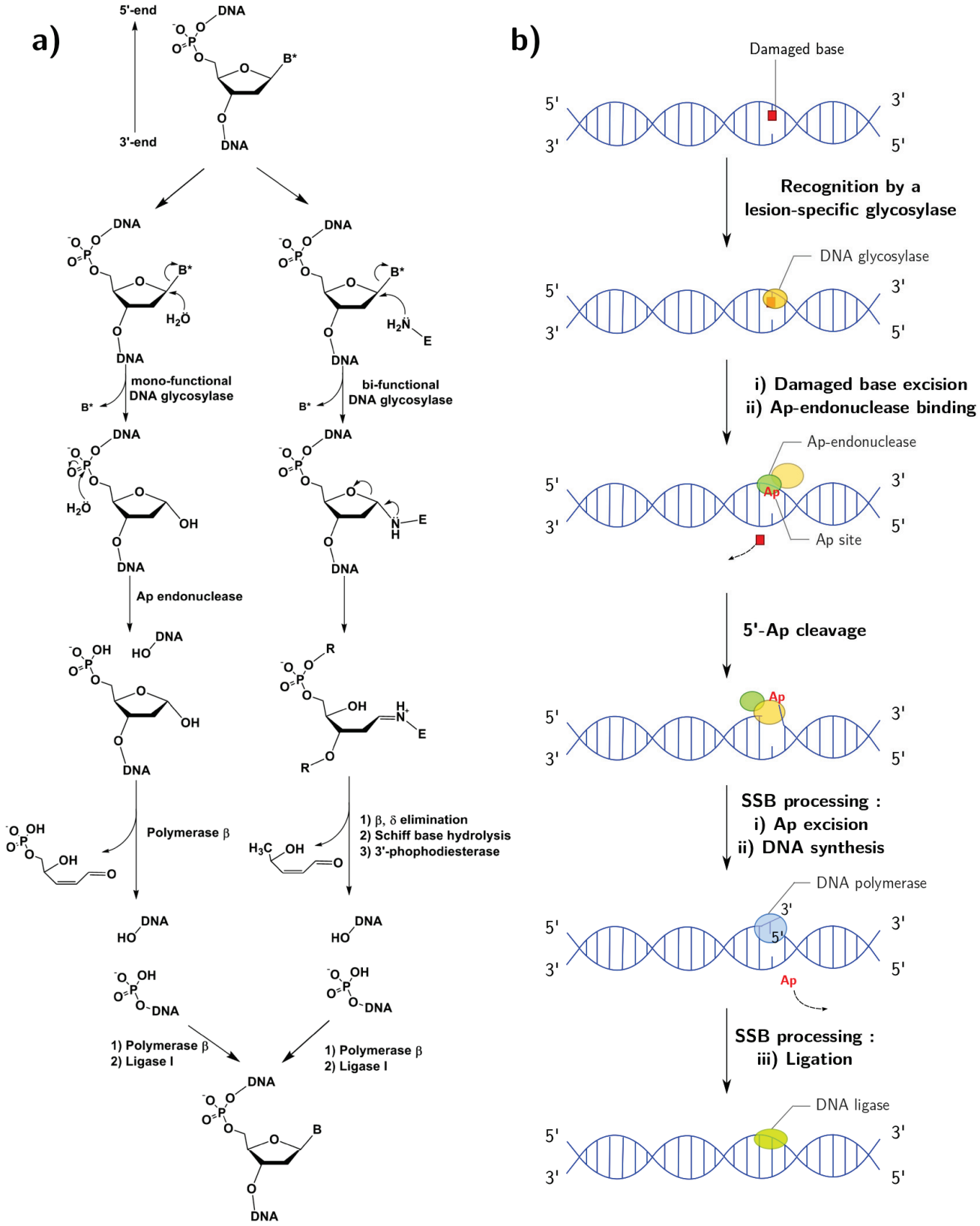


FIGURE 4.2.1: a)Base Excision Repair mechanisms with mono-functional glycosylase (left) or bi-functional glycosylase (right). B = base ; B* = damaged base. b) Schematic representation of BER process steps.

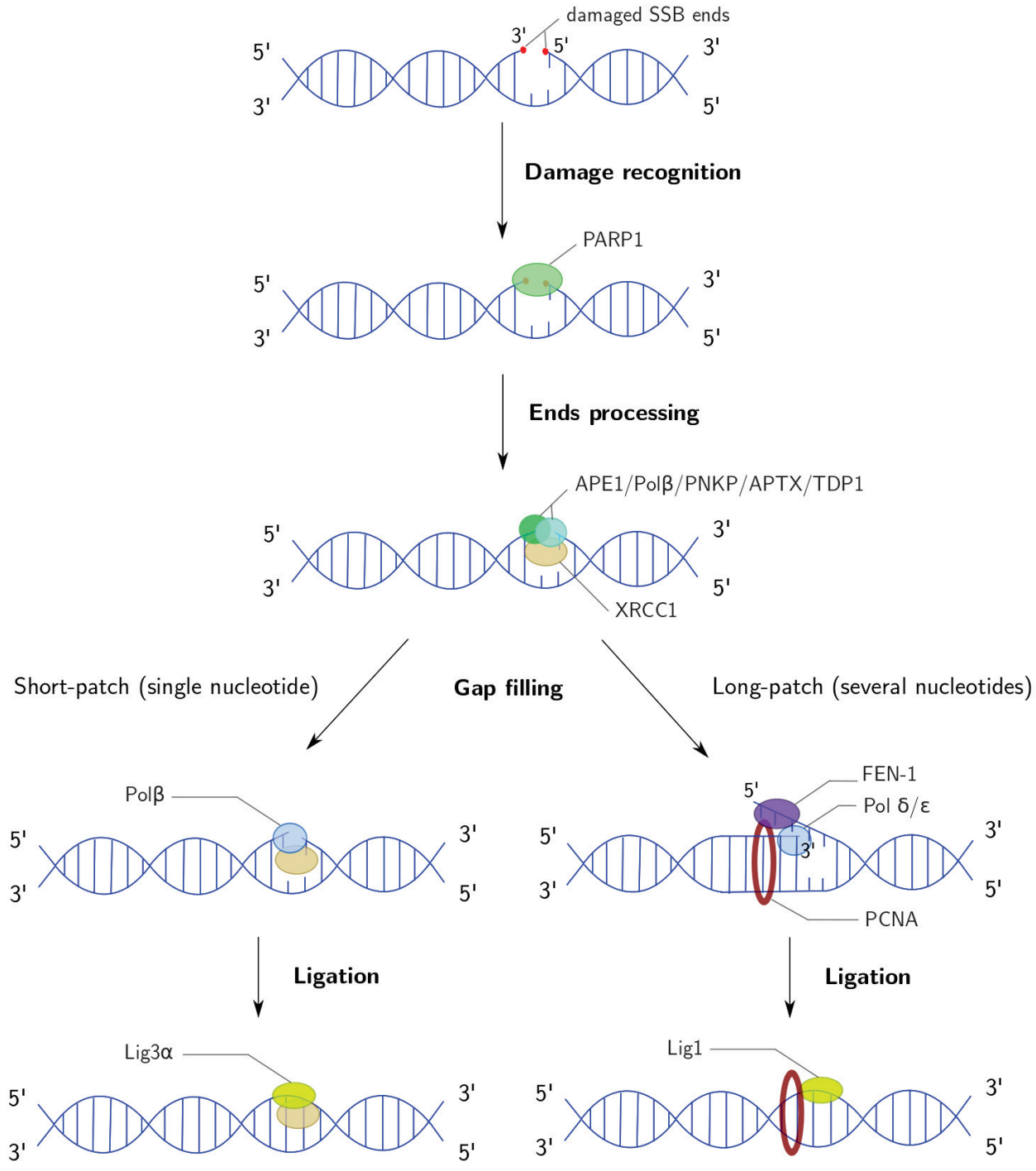


FIGURE 4.2.2: Schematic representation of Single Strand Break Repair process steps. Short- and long-patch gap filling stages are depicted on the left and the right respectively.

DNA by the replication factor C (RFC). During this process, the FEN-1 protein cuts the displaced 5'-nucleotides. Finally, ligation of the 5' and 3' ends is ensured by the Lig1 or Lig3α/XRCC1 ligases for long-patch and short-patch repair respectively.
Noteworthy, some studies highlighted the importance of DNA end-processing in DDR since in hereditary neurodegenerative disease several enzymes involved in this process are mutated[196].

4.2.3 Nucleotide Excision Repair

Nucleotide Excision Repair is a powerful pathway that processes a large range of unrelated lesion. Notably, it is involved in CPDs, 6-4PPs, bulky adducts, and intrastrand crosslinks repair. Two subpathways have been identified : global genome NER (GG-NER) and transcription-coupled NER (TC-NER). TC-NER is activated only when a lesion is detected by stalling of the RNA polymerase II activity during DNA transcription, while in the GG-NER subpathway the entire genome is probed in order to localize helix distortions induced by damages[198].
For an easier comprehension of this processes description, on can rely on the Figure 4.2.3 while reading the following.

Global Genome NER The entire genome is probed by the XPC protein, which is supported by the UV excision repair protein RAD23B and centrin (CETN2). XPC binds to the single strand opposite to the lesion by inserting its carboxy-terminal double βhairpin in the gap induced by disrupted base pairing[199]. CPDs are particular cases because they induce only few helix deformations, hence are poor substrates for XPC[200,201]. The recognition of the lesion is enabled by the UV-DNA damage binding protein complex (UV-DDB complex) that directly binds to UV-induced damage site and stimulates the binding of XPC on the opposite strand[202,203].
After XPC and UV-DDB binding to the damage site, the lesion is verified by association of the TFIIH (transcription initiation factor IIH) complex, which is composed of ten subunits[204] including two DNA helicases XPB and XPD extend the gap between DNA strands. The lesion verification process consists in the localization of helicase-blocking lesions by TFIIH scanning of the helix in the orientation 5'$\rightarrow$3' . Thereby, if the damage is not verified , the repair process can abort[205,206].
The next step is the damaged site removal by 5' and 3' double incision performed by the structure-specific endonucleases XPF-ERCC1 and XPG. These enzymes cleave irreversibly DNA at short distances of the damaged site in 5' and 3' respectively, resulting in a 22-30 nucleotides long SSB[207]. All along this process, other essential proteins intervene to interact with NER proteins for lesion-verification stimulation, TFIIH binding to DNA and XPF-ERCC1 recruitment (by the XPA protein[208,209]), and for protection of the non-damaged DNA strand as well as the coordination of the gap-filling synthesis and the DNA double incision - by the RPA protein[210].
The final step, consisting in DNA gap-filling synthesis and ligation is executed by polymerases and ligases enzymes, assisted by the proliferating cell nuclear antigen (PCNA), a sliding clamp loaded onto DNA by the replication factor C (RFC). The nature polymerase and ligase that intervene is ruled by the cell status - see section 4.2.7.

Transcription-coupled NER TC-NER mechanisms are far less understood than GG-NER. During DNA transcription, RNA polymerase II (RNA Pol II) synthetizes the mRNA precursor by complementarity of a DNA single strand. In presence of a lesion site that has not been repaired by GG-NER, RNA Pol II can be stalled. For example, CPDs repair by

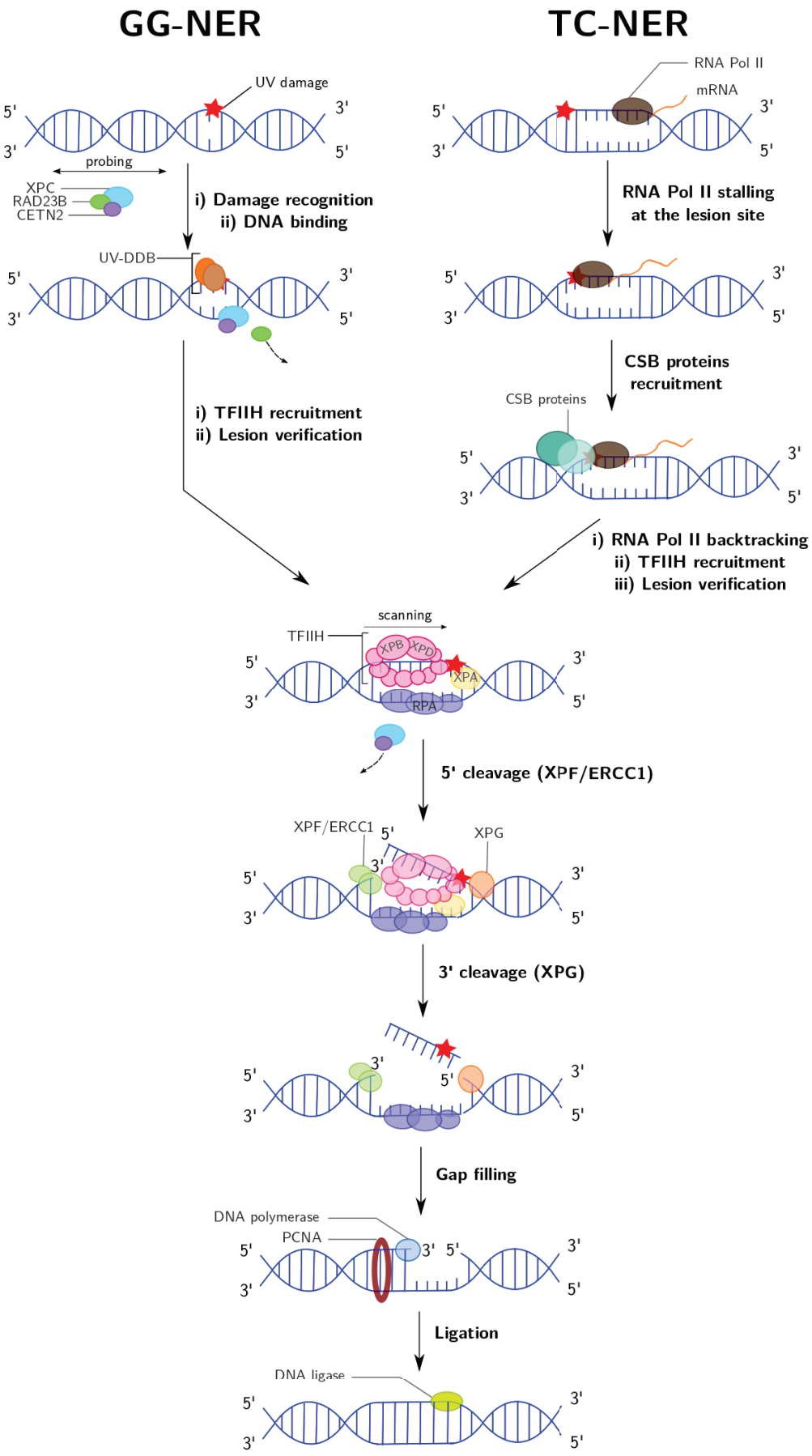


FIGURE 4.2.3: Schematic representation of Nucleotide Excision Repair process steps. Damage detection following the GG-NER and TC-NER steps are depicted on the left and the right respectively.

this GG-NER is not so efficient, even in the presence of the UV-DDB complex. Thereby, it can block the RNA Pol II when transcripting the damaged strand.
In theory, TC-NER would repair lesions that sliped through the GG-NER and BER pathways. The lesion-stalled RNA Pol II recruits the Cockayne syndrome proteins, CSA and CSB, essential for the TC-NER proteins assembly[211]. Then, RNA Pol II has to be dissociated from the DNA since it covers ∼35 nucleobases of the damaged strand, avoiding NER enzymes access, hence the repair[212]. Up to now, the mechanisms and factors implied in the RNA Pol II backtracking and the TT-NER complex assembly remain elusive, albeit the CSB proteins might play a role in the backtracking process[213,214].
The difference between GG- and TC-NER lies in the lesion's recognition, both pathways sharing the same action mechanisms once TFIIH complex binding is done.

4.2.4 Double Strand Break Repair

Double strand breaks (DSBs) are particularly deleterious for cells, and their repair is of utmost importance to ensure the cell's viability as well as the genome's integrity. DSB repair (DSBR) can take place in four different ways depending on the cell-cycle and the nature of the DSB ends. The two main pathways are the classical non-homologous end joining (C-NHEJ) and the homologous recombination (HR) that will be described below. Moreover, HC can end in more error-prone secondary mechanisms : alternative End Joining (alt-EJ) and Single-Strand Annealing (SSA)[215].
As for the NER description, one should rely on the Figure 4.2.4 while reading the DSBR steps details in the following.

Classical Non-Homologous End Joining C-NHEJ consists of the nicked strands ends direct ligation, requiring the intervention of several factors : Ku70/80 (Ku heterodimer), DNA-PKcs (DNA-dependent protein kinase catalytic subunit, which catalyzes phosphorylation), and XCCR4/DNA ligase IV (Lig4) heterodimer. First, the Ku complex binds to DNA ends with which it has a strong affinity[216], protecting the ends from degradation since they have to be compatible for an efficient and high fidelity joining. When bound to DNA ends, Ku also serves as a dock for enzymes that intervene in the remaining NHEJ steps[217]. Thereby, DNA-PKcs is able to form complex with Ku only when the latter is bound to DNA[218]. DNA-PKcs binds to DNA ends and stimulates the NHEJ enzymes activity. Then, the XRCC4/Lig4 heterodimer binds to the Ku/DNA complex and rejoins the DNA ends. If needed, degradated ends processing can be performed by the Artemis endonuclease, which interacts with DNA-PKcs[219,220].

Homologous Recombination Homologous recombination pathway involves the use of a homologous undamaged strand from sister chromatid to repair DSB. The sister chromatid has to be accessible so HR is the main DSBR pathway only during the S and G2 cell-cycle phases[221] - see section 4.2.7. HR is initiated by DSB binding to the heterotrimeric MRE11–RAD50–NBS1 (MRN) complex. MRE11, together with the CtBP-interacting protein (CtIP), induce the DSB 5'-3' resection resulting in a single stranded DNA that is necessary for the homology search. The ssDNA binds to the replication protein A (RPA) that prevents from the formation of secondary structures and activates the loading of the recombinase Rad51, whose activity is mediated by several proteins such as the breast cancer type 2 susceptibility (BRCA2) and the RAD52 complex. Rad51 produces a presynaptic nucleoprotein filament on the ssDNA, which will be paired to its homologous undamaged duplex if found during the homologous search on the sister chromatin. This pairing permits the gap-filling by DNA synthesis , hence concluding the

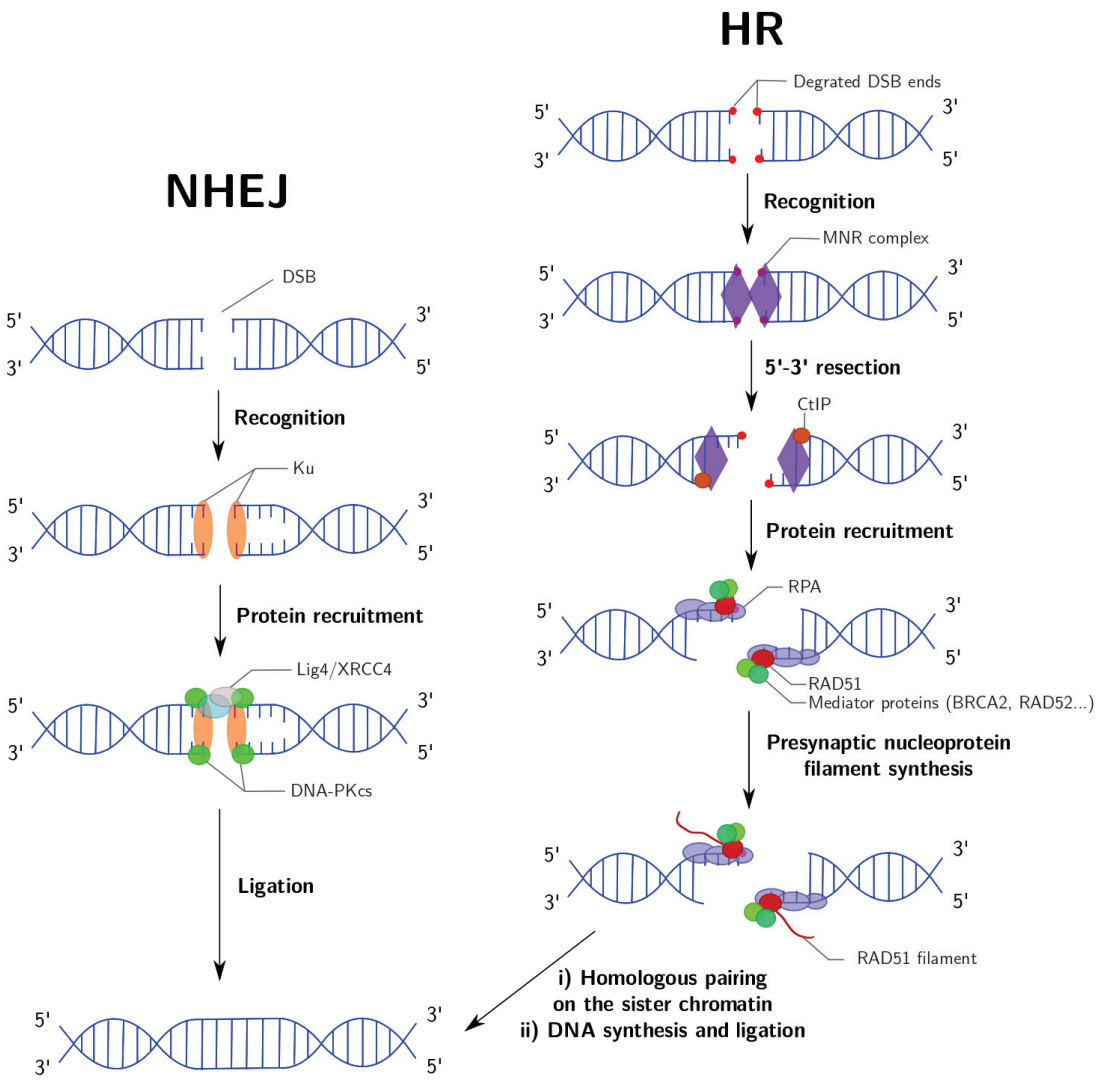


FIGURE 4.2.4: Schematic representation of Double Strand Break Repair (SSBR). On the left: Non-Homologous End Joining process. On the right: Homologous Recombination.

repair by the the DNA ends ligation. This final step can occur in several manner that will be not described here, but it is worth noting that this is a very challenging process that is prone to induce error.

4.2.5 Interstrand Cross-Links Repair

ICL are processed during the replication phase (S phase) where they stall helicases and polymerases activities, hence provoking apoptosis[222]. As a consequence, this process cannot occur during the G1 phase and in non-cycling cells since a replication fork is necessary to recognize the lesion. ICL recognition is ensured by structure-specific DNA-binding enzymes, the FANCM endonucleases, when the damage is localized by stalling of a replication fork[223]. The Fanconi anaemia (FA) core complex then comes to interact with FANCM on the damage site, involving the recruitment of several other enzymes. The latter nick the DNA backbone on both sides of the ICL on one strand, resulting in a single-strand break that will be processed by Homologous Recombination since the complementary strand cannot be used as an homologous template[224]. The flipped-out ICL that remains on the untreated strand could be removed by NER but it is still matter of research.
Several aspects of ICL repair mechanisms are not elucidated yet and still remains as hypothesis. Thereby, this has to be intensively investigated, specially because ICL are highly cytotoxic but also cause their repair generate the very deleterious double strand breaks whose repair is strongly challenging and prone to induce genotoxicity.

4.2.6 Mismatch Repair

DNA mistmatch repair (MMR) is particularly important and challenging since it consists in the repair of base-base mispairs that has to be localized in several thousands of nucleotides. Its mechanisms are widely studied using *E. Coli* bacteria as a model[225].
Recognition of the damaged site happens just after replication, because daughter strand that contains the mismatch can be distinguished from its parent strand by the abscence of nucleobases methylation. DNA methylation is an epigenetic modification that controls the silencing of genes that are not needed in a particular cell. This way, enzymes involved in this process can discriminate between the daughter (that is not methylated yet) and the parent strands and ensure the genome integrity. First, the MutS ATPase protein recognizes and binds to the mismatch site. MutS contains an adenine diphosphate (ADP) that is exchanged for and ATP when recognizing the mismatch[226]. ATP binding induces conformational changes that result in the formation of a sliding clamp on DNA[227,228], with which the MutL protein will interact specifically[229–231]. MutL activates the endonuclease MutH, which will cleave the daughter strand[232]. Discrimination of the daughter strand by MutH is facilitated by analyzing the adenine methylation at d(GATC) sites. MutL also serves for the helicase UvrD activation[233], which processes the daughter strand from the nick whilst exonucleases cleave the unwound strand to remove the mismatch site. Polymerases and ligases then conclude the process by DNA synthesis and ligation of the nicked strand.
In human, MutS homolog is MutSα, a heterodimer formed by the MSH2 and MSH6 proteins, and MutL counterpart is MutLα, a MLH1/PMS2 complex. Numerous crystal structures have been reported, allowing the comprehension of isolated MMR enzymes mechanisms, but the interconnections between them still remain unclear and are matter of intensive work.

4.2.7 Cell cycle influence

The cell cycle is divided in four phases : Mitosis phase (M), Gap 1 phase (G1), Synthesis phase (S), and Gap 2 phase (G2) - see Figure 4.2.5. First, cell begins to grow and produces proteins in preparation for the DNA replication stage : it is the G1 phase. Then comes the S phase, during which each chromosome is replicated in an identical sister chromatid. The second gap phase G2 occurs next, during which the cell continues to gain in size and to produce proteins in prevision of cell mitosis. Finally, the mitosis phase starts and the cell is divided in two identical daughter cells sharing the same chromosomes. Cells can also exit from the replication cycle find themselves in a non-replicating phase (G0), called quiescence[234]. The cell phase is very important in DNA damage repair. Checkpoints are present before replication (G1/S) and mitosis (G2/M) in order to ensure the integrity of the replicated cells. The cell cycle progression is driven by the level of cyclin-dependent

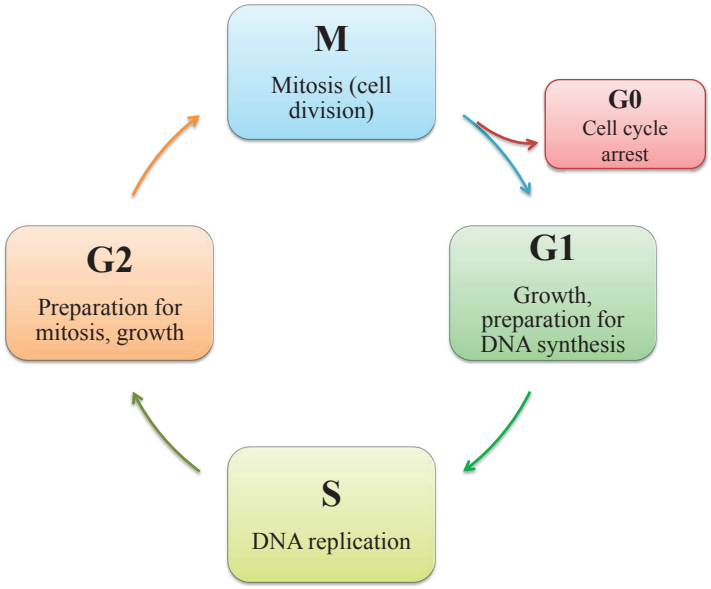


FIGURE 4.2.5: Cell cycle four steps : Mitosis, G1, S, and G2 phases. The cell can exit from this cycle to reach a G0 non-replicating cell state.

Ser/Thr kinases (CDKs) enzymes, which plays an important role on several DDR enzymes activation since it is responsible for their phosphorylation catalysis. For instance, during DSBR CDK promotes the DNA 5'-3' resection by phosphorylating the CtIP protein[235] - see section 4.2.4. Thus, the homologous recombination that requires the 5'-3' resection, can work only at high CDK concentrations, ie not during the G1 or G0 phases.
Likewise, in NER final steps (DNA gap-filling and ligation see section 4.2.3), proteins that come into play depends on the status of the cell[236]. Thereby, while in replicating cells the DNA polymerase ϵ and ligase 1 enzymes intervene in these steps, in non-replicating cells (G0 status) the DNA synthesis is executed by Pol δ and Pol κ. It also happens that the low concentration of ligase 1 in non-cycling cells induces its replacement by the XRCC1-ligase 3 complex[237] which also intervene in SSBR short-patch repair - see section 4.2.2.

Cell cycle definitely rules DNA damage repair processes, and that opens a large perspectives on repair regulation by cell cycle control.

4.3 Our projects in this framework

4.3.1 Tandem 8-oxoGs and Ap/8-oxoG influence on 8-oxoG eversion ease by Fpg enzyme.

In perspective of our work on 8-oxoG mechanisms of formation (see section 2.3.2), we studied DNA interactions with the Fpg glycosylase, which is responsible for this lesion cleavage during the BER process (see section 4.2.1) in Escherichia Coli. This investigation is currently in progress (Chen Hui Chan PhD project), thus only preliminary results are described in the following. I co-supervised him as he was running simulations.
8-oxoG is well-known to mismatch with adenine, hence needs to be efficienly repaired: this task is devoted to Formamidopyrimidine DNA glycosylase (Fpg, bacterial hOGG1) and then an Ap-endonuclease. During the base excision repair (BER)[238,239], Fpg recognizes 8oxoG among other oxidatively-induced lesions and then proceed to eversion[240,241]. The mechanisms of recognition[242] and extrusion[240,243,244] of 8oxoG have been scrutinized and are now well characterized, through a series of techniques including molecular modelling. It is experimentally established that the combination of two 8-oxoGs dramatically lowers the repair rate[245]. However an interpretation at the molecular level is currently missing, since X-ray or NMR structures have not been resolved. In this work we investigate owing to molecular mechanics simulations the interaction between Fpg and an oligonucleotide harbouring one or two 8-oxoguanine defects. We rely on explicit-solvent molecular dynamics[246,247] to test/probe in silico the Fpg:DNA interaction as an adjacent guanine to the 8oxoG present in the X-ray structure (PDB code 3GO8) is mutated in silico to a second 8oxoG - see Figure 4.3.1.
Our simulations afford an unprecedent atomic view of the change in the recognition pattern of a tandem lesion. They shed light on an unambigously major structural change as the catalytic arginine no longer stays close to 8oxoG: instead the phenylalanine F111 partially inserts into the DNA helix. This might be made possible by a structural flexibility/instability of the duplex in presence of the tandem lesion. Furthermore, Potential Mean Force (PMF) calculations were carried out in order to assess the energetics corresponding to the 8-oxoG eversion towards Fpg active site in the case of isolated and tandem 8-oxoG. Similar calculations were performed on oligonucleotides containing tandem Ap site/8-oxoG. Indeed, several experimental studies about clustered 8-oxoG/Ap are available, but the exact influence between both lesions is not yet clearly delineate. Notably, the impact of an Ap site presence on 8-oxoG mutagenic potential appears to be controversial in the literature[248–250]. Thus, we wanted to bring out structural informations as well about this tandem lesion. PMF calculations shed on light the energetics signature of 8-oxoG eversion in these three cases, as depicted in Figure 4.3.2. This study is mainly led by Chen Hui Chan, a first year PhD student, in close collaboration with our colleagues from Nancy University, and may lead to two papers, one about our results, the other concerning the methodology used for the PMF calculations, which differ from papers published recently on the subject[251,252]. I co-supervised the first part of this study - MM-MD calculations.

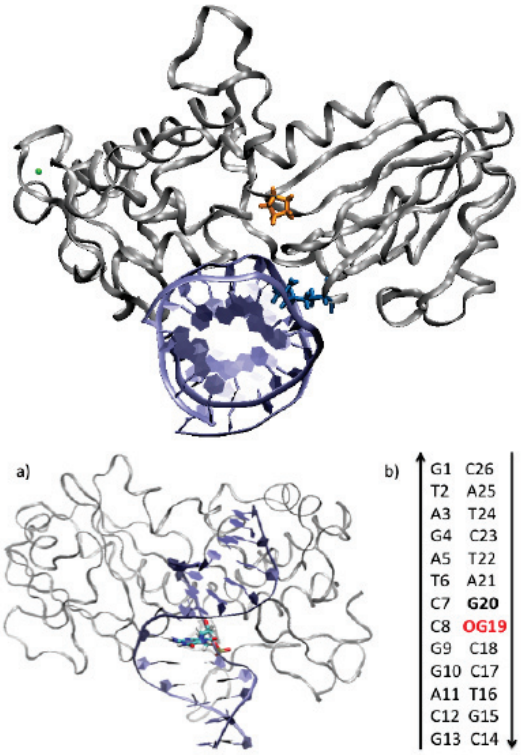


FIGURE 4.3.1: Cartoon representation of bacterial Fpg in interation with a DNA double-strand harbouring 8-oxoG as the 19th nucleobase - PDB code 3GO8. For our simulations, G20 was mutated into 8oxoG (or Ap) to investigate the structural aftermaths. On the top picture, F111 is depicted in yellow and the catalytic arginine in blue, this latter lying in the proximity of the single 8-oxoG.

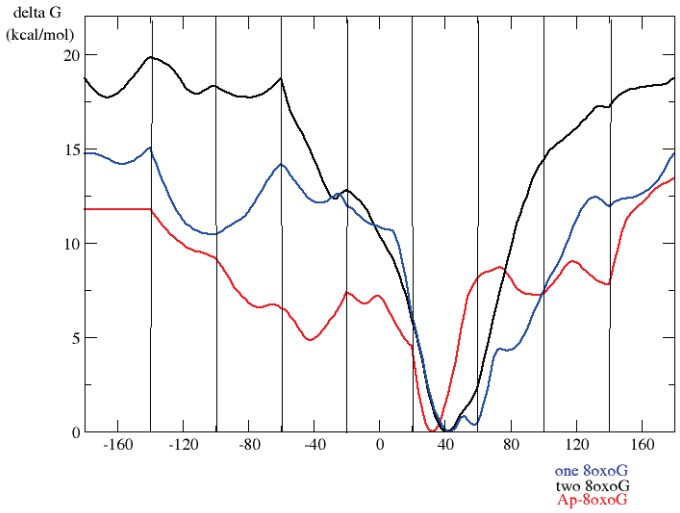


FIGURE 4.3.2: PMF of 8-oxoG exclusion towards Fpg active site in the three cases : single 8-oxoG (blue), tandem 8-oxoG (black), tandem Ap/8-oxoG (red). The presence of another 8-oxoG at position 20 clearly disfavors the 8-oxoG19 eversion, while tandem with an Ap site seems to ease this process. This might deal with the absence of π-stacking induced by the presence of the Ap site, which could facilitate the 8-oxoG19 flipping.

4.3.2 Repair rate of clustered abasic DNA lesions by human endonuclease: molecular bases of sequence specificity [P6]

The APE1 study was in perspective of the work we performed on clustered Ap containing B-DNA (see Chapter 3), work that came from discussion with Alexandros Georgakilas (NTU, Athens, Greece COST) who had made experiments on such systems and was questionning about sequence effect on DNA structure, that have a strong impact on enzyme recognition and hence, repair. Together with our colleagues from University of Lorraine, we investigated the effects of these same sequence effects in B-DNA interacting with APE1, an Ap-endonuclease responsible for Ap sites cleavage in human. This study gave insights into structural properties of such complex which are of major importance for DDR since Ap site are abondant cytotoxic damages in cells. Noteworthy, we highlighted the sequence effect on two amino acids of APE1 which act as a stabilizing 'pincer' by interactions in the gap within the helix, induced by the Ap site flipping. Results obtained were in good agreement with the experimental data, and were published in *Journal of Physical Chemistry Letters*.

4.4 Publications

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

Repair Rate of Clustered Abasic DNA Lesions by Human Endonuclease: Molecular Bases of Sequence Specificity

Hugo Gattuso,[†,‡] Elodie Durand,[†,‡] Emmanuelle Bignon,[§,∥] Christophe Morell,[∥] Alexandros G. Georgakilas,[⊥] Elise Dumont,[*,§] Christophe Chipot,[†,‡,▽,⊗] François Dehez,[*,†,‡,⊗] and Antonio Monari[*,†,‡]

[†]Université de Lorraine—Nancy, Theory-Modeling-Simulation SRSMC, 54000 Vandoeuvre-lès-Nancy, France

[‡]CNRS, Theory-Modeling-Simulation SRSMC, 54000 Vandoeuvre-lès-Nancy, France

[§]Univ Lyon, Ens de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F-69342 Lyon, France

[∥]Université de Lyon, Institut des Sciences Analytiques UMR 5280, CNRS, Université de Lyon 1, ENS Lyon 5 rue de la Doua, F-69100 Villeurbanne, France

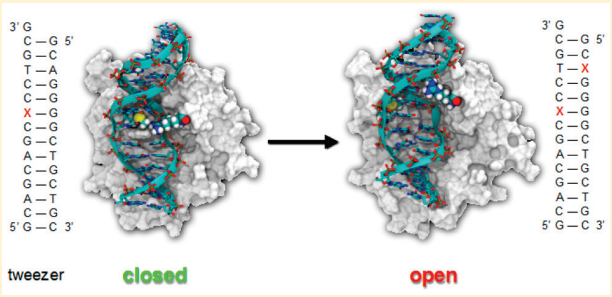
[⊥]DNA damage laboratory, Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou, Athens 15780, Greece

[▽]Department of Physics, University of Illinois at Urbana—Champaign, 1110 West Green Street, Urbana, Illinois 61801, United States

[⊗]Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana−Champaign

Supporting Information

ABSTRACT: In the present contribution, the interaction between damaged DNA and repair enzymes is examined by means of molecular dynamics simulations. More specifically, we consider clustered abasic DNA lesions processed by the primary human apurinic/apyrimidinic (AP) endonuclease, APE1. Our results show that, in stark contrast with the corresponding bacterial endonucleases, human APE1 imposes strong geometrical constraints on the DNA duplex. As a consequence, the level of recognition and, hence, the repair rate is higher. Important features that guide the DNA/protein interactions are the presence of an extended positively charged region and of a molecular tweezers that strongly constrains DNA. Our results are on very good agreement with the experimentally determined repair rate of clustered abasic lesions. The lack of repair for one particular arrangement of the two abasic sites is also explained considering the peculiar destabilizing interaction between the recognition region and the second lesion, resulting in a partial opening of the molecular tweezers and, thus, a less stable complex. This contribution cogently establishes the molecular bases for the recognition and repair of clustered DNA lesions by means of human endonucleases.



Some of the most common, and nonetheless toxic, DNA lesions are the abasic or apurinic/apyrimidic sites (AP). AP sites are produced upon cleavage of the DNA glycosidic bond, thus leaving only the uncapped deoxyribose moiety in the DNA strand. Noncoding abasic (AP) sites can form in the DNA either spontaneously or as a result to the exposure to reactive oxygen/nitrogen species (ROS/RNS), ionizing radiations or chemotherapeutic drugs like bleomycin.[1] Although the residual AP moiety is in equilibrium between a cyclic sugar and an open form (see Figure S1 in Supporting Information (SI)), the cyclic conformer is absolutely predominant, accounting for about 99% of the occurrence[2,3] and hence it is the only one considered in the following simulations. Molecular modeling[4] and NMR structural determination[5] agree in underlying a strong local deformation of the DNA helix as a response to the presence of the AP site. In particular, one can pinpoint a marked preference of the AP site to adopt an extrahelical position that correlates with the presence of a bulge in the DNA backbone. Interestingly, the latter structural deformation can be seen as extremely relevant for lesion repair.[6−8]

AP intermediates can be the most common toxic or mutagenic DNA lesions in vivo. The human APE1[9,10] and bacterial Nfo[11] repair enzymes epitomize the two well-conserved 5′-P endonuclease families but are structurally nonhomologous. Moreover, they do not present similar interaction patterns with DNA and do not require the same metal ion.[12] The processing of AP lesions[13] involves recognition by the enzyme followed by a cleavage of the

Received: July 30, 2016
Accepted: September 9, 2016
Published: September 9, 2016

DOI: 10.1021/acs.jpclett.6b01692
J. Phys. Chem. Lett. 2016, 7, 3760−3765

phosphodiester bond mediated by a Mg^{2+} ion in the case of APE1, Zn^{2+} for Nfo and a water molecule. In the case of human APE1[14] the active site is embedded in the enzyme and exhibits an extended contact region between the positively charged residues and the negatively charged DNA backbone. On the other hand, in the case of bacterial *Escherichia coli* endonuclease IV (Nfo), the contact interface with DNA is much smaller, and resumes essentially to the few residues composing the AP recognition site (Figure 1). As a consequence, the interaction with APE1 induces much larger structural deformations to the damaged DNA fragment, whereas its native structure is globally preserved by Nfo.

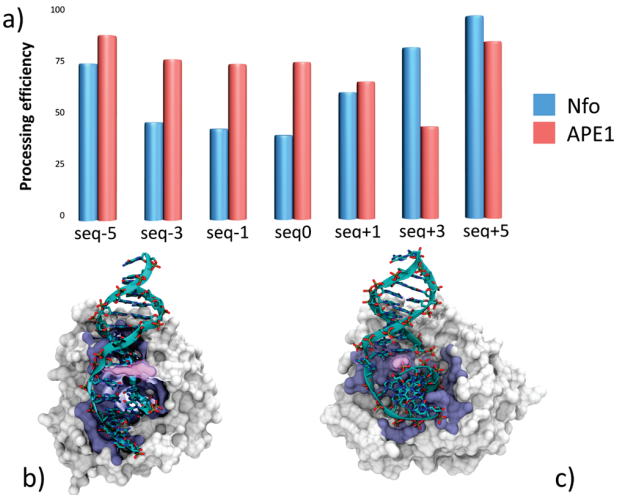


Figure 1. (a) Repair efficiency (cleavage) of different clustered AP lesions in different DNA sequences by human and bacterial endonucleases based on previously published data.[7,8] (b) Representation of the crystal structure of bacterial Nfo interacting with a DNA sequence (PDB code 4K1G).[11] (c) Representation of the crystal structure of human APE1 interacting with a DNA sequence (PDB code 1DEW).[10] Recognition area is highlighted in mauve and the contact area close to the AP site are further highlighted, notice the presence of the tweezers in APE1 holding the lesioned DNA, whereas for Nfo, only a small finger interacting with the lesioned DNA is visible. See Figure S2 in SI for a detailed definition of the oligomers' sequence nomenclature.

The chemical mechanism of recognition and repair of AP sites is interesting per se, also due to the high occurrence of this lesion (~5000−10000 lesions per cell daily).[15] Moreover, it is nowadays established that exposure to ionizing radiations (γ- and X-rays), like in the case of radiotherapy, may lead to the accumulation of two or more AP sites, as well as oxidized bases, double-strand breaks (DSBs) and non-DSB lesions within 2 or 3 DNA helical turns.[16,17] Two closely spaced DNA lesions constitute indeed the minimal size cluster damage.[18] Clustering of DNA lesions may become extremely threatening for the cells[19,7,20,21] because of their well-established repair resistance and the high possibility of non-DSB clustered lesions to be readily converted into DSBs during base excision repair (BER).[22]

Furthermore, a strong interplay between DSBs and cluster sites may also enhance their biological harmfulness; indeed, the processing of DSB can be strongly limited, or totally inhibited, by the presence of neighboring AP sites.[23]

Numerical simulation techniques have steadily become an invaluable tool to characterize with an atomistic and molecular resolution the outcome of damages on the structural and dynamical properties of different DNA sequences as well as their interaction with repair enzymes.[24−29] Molecular modeling has been applied to damaged oligonucleotides, such as abasic sites,[4,30,31] bulky adducts,[32,33] cross-links,[34] or photolesions,[20,35] and to duplexes in interaction with repair enzymes.[36] MD explorations of clustered 8-oxoguanine and abasic sites have been recently published.[37,38] We recently reported[39] the molecular modeling and simulation of clustered AP sites that may be induced in DNA duplexes by low-dose radiation.[40] The latter class of lesions have been produced in plasmid or phage DNA,[7,21,41] and isolated DNA fibers.[42,43] By exploiting the DNA repair rate measurements[44,45] clustered AP sites have been confirmed up to nucleosomal DNA[46] and human cells.[8]

The repair rate of clustered AP sites has been studied[7,8] in the case of both human APE1 and bacterial (*Escherichia coli*) Nfo (Figure 1, see also SI for the definition of the sequences and the nomenclature). Nfo features a global lower repair rate and strong dependence on the AP position as similarly suggested by independent works.[47] We have correlated the effect of the relative position of the two AP sites with the structural deformation of the DNA structure induced by the lesions, remarkably our finding also provide a rationale for the observed processing rate of bacterial Nfo. The low processing rate for seq0 and especially seq-3 (see SI and Figure 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 nonlesioned purines for seq-3.[39] This inference is also justified by the observation that the bacterial enzyme contact with DNA is quite local (Figure 1) and, as a consequence, is not able to induce strong structural deformation and mechanical constraints on the DNA duplex.

The behavior of APE1 enzyme is singularly different. First, the processing rate is globally higher compared to that of the bacterial enzyme. Second, the ease of repair in humans, and, in general in all eukaryotic organisms, is consistent with a different evolutionary pathway and with the necessity to preserve genome integrity in complex organisms. The sequence dependence is also far less pronounced, all the lesions being repaired at a rate going from 70 to 80%; the only exception is the seq+3 lesion for which a mediocre repair rate of about 45% is observed.[8] Interestingly, the latter was also one of the best-repaired lesions by the bacterial Nfo. One, therefore, can immediately surmise very different recognition patterns between the two organisms and, as a consequence, the correlation between DNA structural deformation and repair ought to be revised. Toward this end, we explore in present work the behavior of clustered AP sites containing a DNA double strand in solution and in interaction with the human APE1 enzyme. The structural evolution of the DNA-enzyme complex has been simulated using all-atom molecular dynamics (MD) with the parm99bsc1 force field.[48,29,49] The simulations have been carried out in cubic water boxes with the NAMD package.[50] Production runs of 100 ns have been carried out in the NPT ensemble (see SI for a full description of the methodology). As will be illustrated in the following 100 ns appear sufficient to explore the conformational space of the DNA/APE1 complex in particular due to the strong interactions and high rigidity of the aggregates. Cluster AP lesions have been produced by specific mutations of the crystallographic DNA strands.[10]

DOI: 10.1021/acs.jpclett.6b01692
J. Phys. Chem. Lett. 2016, 7, 3760−3765

A molecular tweezers, capped at its end respectively by one arginine and one methionine residues (see also SI), stabilizes further the DNA−protein complex. In the case of the isolated enzyme (Figure 2a), this molecular tweezers is open and the

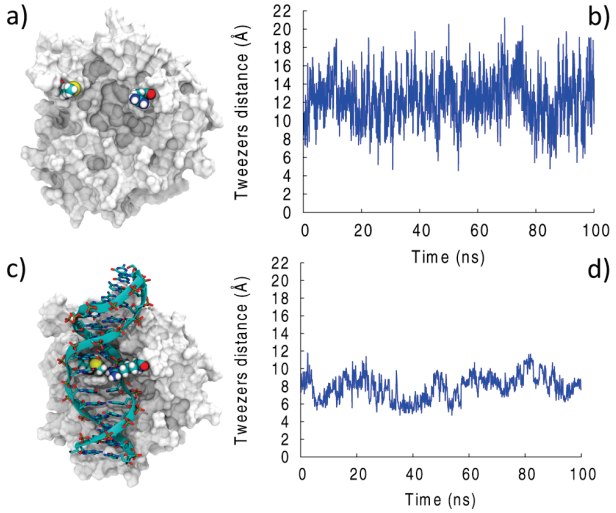


Figure 2. Representative snapshot and time evolution of the molecular tweezers width for the free APE1 (a,b) and for the complex between APE1 and the single AP containing DNA strand (c,d).

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.

Interestingly, although the global DNA/APE1 structure is quite compact, we have observed three channels (see SI, Figure S4) allowing ingress of cations that may easily reach the active site, where the presence of Mg^{2+} is required for the cleavage reaction to occur. Note, however, that the simulations were only performed in the presence of Na^{+}; however, the fast and efficient ion entrance should not influence DNA recognition because it follows complexation with the DNA oligomers.

The behavior of APE1 enzyme free and in interaction with the single AP-containing DNA strand is depicted in Figure 2. The DNA/APE1 contact region is quite extended (see Figure 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 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 SI and Figure 3) spanning all

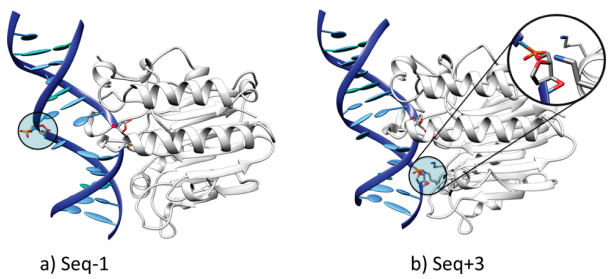


Figure 3. Representative snapshots for the seq-1 (a) and seq+3 (b) DNA/APE1 complex. DNA and protein are represented in carton. AP sites are evidenced in licorice representation. The interaction between the secondary AP site of seq+3 and the recognition region of APE1 is evidenced in the zoom of panel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

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 4). As already evidenced in the case of isolated nucleotides,[39] 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 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 3b), 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 5), 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.

DOI: 10.1021/acs.jpclett.6b01692
J. Phys. Chem. Lett. 2016, 7, 3760−3765

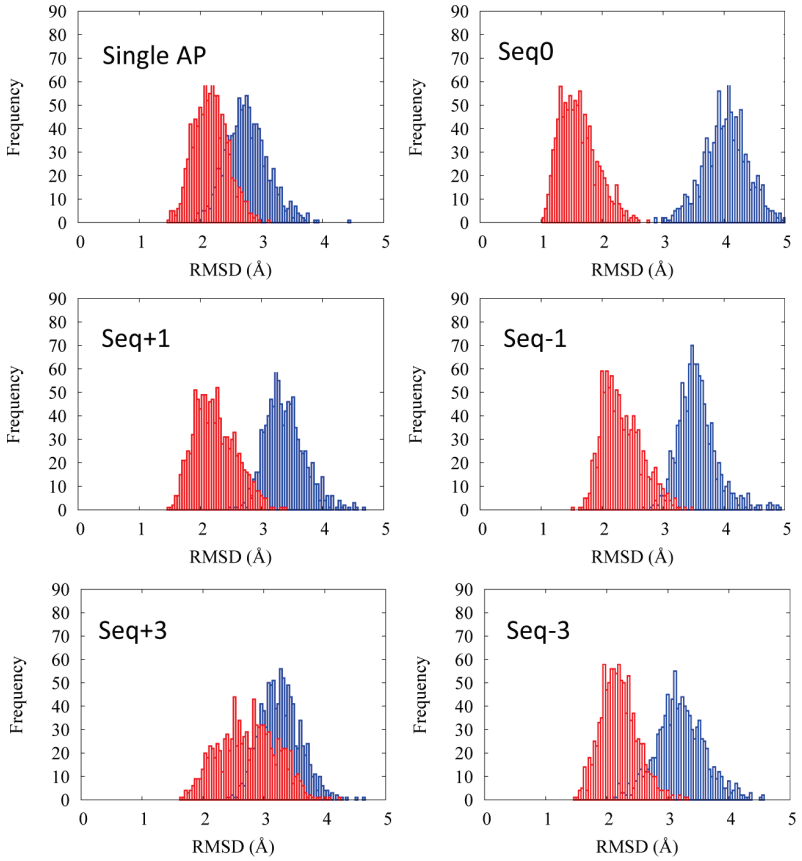


Figure 4. Distribution of the DNA root-mean-square deviation (RMSD) with respect to the crystal structure for the single containing and clustered AP lesions, inside the APE1 complex (red) and in solution (blue). The corresponding repair rates are given in Figure 1, whereas time series are reported in SI.

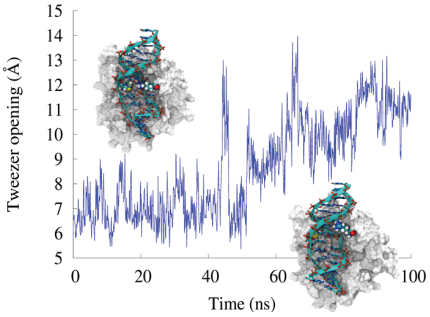


Figure 5. Representative conformations and time evolution of the molecular tweezers width for seq+3 showing the partial opening and the loss of stabilizing interactions.

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, 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,[51−53] ultimately being able to enhance radiotherapy efficiency.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.6b01692.

Description of the molecular dynamic simulation computational details and parameters for the AP site.

DOI: 10.1021/acs.jpclett.6b01692
J. Phys. Chem. Lett. 2016, 7, 3760−3765

Chemical formulas of the AP sites, definition of the clustered DNA sequences nomenclature, definition of the tweezers and the tweezers' distance, representative snapshots of all the stable DNA/protein complexes, cations entrance paths, time series of the RMSD evolution for solvated and APE1 complexed oligomers. (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: antonio.monari@univ-lorraine.fr.
*E-mail: elise.dumont@ens-lyon.fr.
*E-mail: francois.dehez@univ-lorraine.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Calculations were performed using the local HPC resources of PSMN at ENS-Lyon and at Université de Lorraine—Nancy. This work was supported by the COST Action CM1201 "Biomimetic Radical Chemistry" is gratefully acknowledged. This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX0007) operated by the French National Research Agency (ANR). Dr. Georgakilas was supported by an EU grant MC−CIG-303514.

REFERENCES

(1) Fung, H.; Demple, B. Distinct Roles of Ape1 Protein in the Repair of DNA Damage Induced by Ionizing Radiation or Bleomycin. *J. Biol. Chem.* **2011**, *286* (7), 4968−4977.

(2) Gates, K. S. An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals. *Chem. Res. Toxicol.* **2009**, *22* (11), 1747−1760.

(3) Clauson, C.; Schärer, O. D.; Niedernhofer, L. Advances in Understanding the Complex Mechanisms of DNA Interstrand Cross-Link Repair. *Cold Spring Harbor Perspect. Biol.* **2013**, *5* (10), a012732.

(4) Ayadi, L.; Coulombeau, C.; Lavery, R.; Arnott, S.; Chandrasekaran, R.; Birdsall, D. L.; Leslie, A. G. W.; Ratliff, R. L.; Ayadi, L.; Coulombeau, C.; et al. Abasic Sites in Duplex DNA: Molecular Modeling of Sequence-Dependent Effects on Conformation. *Biophys. J.* **1999**, *77* (6), 3218−3226.

(5) Hazel, R. D.; Tian, K.; de Los Santos, C. NMR Solution Structures of Bistranded Abasic Site Lesions in DNA. *Biochemistry* **2008**, *47* (46), 11909−11919.

(6) Wilson, D. M.; Barsky, D. The Major Human Abasic Endonuclease: Formation, Consequences and Repair of Abasic Lesions in DNA. *Mutat. Res., DNA Repair* **2001**, *485* (4), 283−307.

(7) Georgakilas, A. G.; Bennett, P. V.; Sutherland, B. M. High Efficiency Detection of Bi-Stranded Abasic Clusters in Gamma-Irradiated DNA by Putrescine. *Nucleic Acids Res.* **2002**, *30* (13), 2800−2808.

(8) Georgakilas, A. G.; Bennett, P. V.; Wilson, D. M.; Sutherland, B. M. Processing of Bistranded Abasic DNA Clusters in γ-Irradiated Human Hematopoietic Cells. *Nucleic Acids Res.* **2004**, *32* (18), 5609−5620.

(9) Freudenthal, B. D.; Beard, W. A.; Cuneo, M. J.; Dyrkheeva, N. S.; Wilson, S. H. Capturing Snapshots of APE1 Processing DNA Damage. *Nat. Struct. Mol. Biol.* **2015**, *22* (11), 924−931.

(10) Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. A. DNA-Bound Structures and Mutants Reveal Abasic DNA Binding by APE1 and DNA Repair Coordination [Corrected]. *Nature* **2000**, *403* (6768), 451−456.

(11) Mazouzi, A.; Vigouroux, A.; Aikeshev, B.; Brooks, P. J.; Saparbaev, M. K.; Morera, S.; Ishchenko, A. A. Insight into Mechanisms of 3′-5′ Exonuclease Activity and Removal of Bulky 8,5′-cyclopurine Adducts by Apurinic/apyrimidinic Endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (33), E3071−E3080.

(12) Tsutakawa, S. E.; Shin, D. S.; Mol, C. D.; Izumi, T.; Arvai, A. S.; Mantha, A. K.; Szczesny, B.; Ivanov, I. N.; Hosfield, D. J.; Maiti, B.; et al. Conserved Structural Chemistry for Incision Activity in Structurally Non-Homologous Apurinic/apyrimidinic Endonuclease APE1 and Endonuclease IV DNA Repair Enzymes. *J. Biol. Chem.* **2013**, *288* (12), 8445−8455.

(13) Molina, R.; Stella, S.; Redondo, P.; Gomez, H.; Marcaida, M. J.; Orozco, M.; Prieto, J.; Montoya, G. Visualizing Phosphodiester-Bond Hydrolysis by an Endonuclease. *Nat. Struct. Mol. Biol.* **2014**, *22* (1), 65−72.

(14) Demple, B.; Sung, J.-S. Molecular and Biological Roles of Ape1 Protein in Mammalian Base Excision Repair. *DNA Repair* **2005**, *4* (12), 1442−1449.

(15) Kryston, T. B.; Georgiev, A. B.; Pissis, P.; Georgakilas, A. G. Role of Oxidative Stress and DNA Damage in Human Carcinogenesis. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2011**, *711* (1), 193−201.

(16) Nikjoo, H.; O'Neill, P.; Terrissol, M.; Goodhead, D. T. Quantitative Modelling of DNA Damage Using Monte Carlo Track Structure Method. *Radiat. Environ. Biophys.* **1999**, *38* (1), 31−38.

(17) Watanabe, R.; Rahmanian, S.; Nikjoo, H. Spectrum of Radiation-Induced Clustered Non-DSB Damage - A Monte Carlo Track Structure Modeling and Calculations. *Radiat. Res.* **2015**, *183* (5), 525−540.

(18) Shikazono, N.; Noguchi, M.; Fujii, K.; Urushibara, A.; Yokoya, A. The Yield, Processing, and Biological Consequences of Clustered DNA Damage Induced by Ionizing Radiation. *J. Radiat. Res.* **2009**, *50* (1), 27−36.

(19) Nikitaki, Z.; Hellweg, C. E.; Georgakilas, A. G.; Ravanat, J.-L. Stress-Induced DNA Damage Biomarkers: Applications and Limitations. *Front. Chem.* **2015**, *3*, 35.

(20) Dumont, E.; Monari, A. Understanding DNA under Oxidative Stress and Sensitization: The Role of Molecular Modeling. *Front. Chem.* **2015**, *3*, 43.

(21) Singh, V.; Das, P. Condensation of DNA–a Putative Obstruction for Repair Process in Abasic Clustered DNA Damage. *DNA Repair* **2013**, *12* (6), 450−457.

(22) Georgakilas, A. G.; O'Neill, P.; Stewart, R. D. Induction and Repair of Clustered DNA Lesions: What Do We Know so Far? *Radiat. Res.* **2013**, *180* (1), 100−109.

(23) Lomax, M. E.; Folkes, L. K.; O'Neill, P. Biological Consequences of Radiation-Induced DNA Damage: Relevance to Radiotherapy. *Clin. Oncol. (R. Coll. Radiol).* **2013**, *25* (10), 578−585.

(24) Qi, Y.; Spong, M. C.; Nam, K.; Banerjee, A.; Jiralerspong, S.; Karplus, M.; Verdine, G. L. Encounter and Extrusion of an Intrahelical Lesion by a DNA Repair Enzyme. *Nature* **2009**, *462* (7274), 762−766.

(25) Ivanov, I.; Tainer, J. A.; McCammon, J. A. Unraveling the Three-Metal-Ion Catalytic Mechanism of the DNA Repair Enzyme Endonuclease IV. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (5), 1465−1470.

(26) Maiti, A.; Noon, M. S.; MacKerell, A. D.; Pozharski, E.; Drohat, A. C. Lesion Processing by a Repair Enzyme Is Severely Curtailed by Residues Needed to Prevent Aberrant Activity on Undamaged DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (21), 8091−8096.

(27) Cai, Y.; Geacintov, N. E.; Broyde, S. Nucleotide Excision Repair Efficiencies of Bulky Carcinogen-DNA Adducts Are Governed by a Balance between Stabilizing and Destabilizing Interactions. *Biochemistry* **2012**, *51* (7), 1486−1499.

(28) Radzimanowski, J.; Dehez, F.; Round, A.; Bidon-Chanal, A.; McSweeney, S.; Timmins, J. An "Open" Structure of the RecOR Complex Supports ssDNA Binding within the Core of the Complex. *Nucleic Acids Res.* **2013**, *41* (16), 7972−7986.

(29) Perez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alphaγConformers. *Biophys. J.* **2007**, *92* (11), 3817−3829.

DOI: 10.1021/acs.jpclett.6b01692
J. Phys. Chem. Lett. 2016, 7, 3760−3765

(30) Ayadi, L.; Coulombeau, C.; Lavery, R. The Impact of Abasic Sites on DNA Flexibility. *J. Biomol. Struct. Dyn.* **2000**, *17* (4), 645−653.
(31) Barsky, D.; Foloppe, N.; Ahmadia, S.; Wilson, D. M., III; MacKerell, A. D. MacKerell, a D. New Insights into the Structure of Abasic DNA from Molecular Dynamics Simulations. *Nucleic Acids Res.* **2000**, *28* (13), 2613−2626.
(32) Liu, Z.; Ding, S.; Kropachev, K.; Lei, J.; Amin, S.; Broyde, S.; Geacintov, N. E. Resistance to Nucleotide Excision Repair of Bulky Guanine Adducts Opposite Abasic Sites in DNA Duplexes and Relationships between Structure and Function. *PLoS One* **2015**, *10* (9), e0137124.
(33) Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S. Nucleotide Excision Repair of 2-Acetylaminofluorene- and 2-Aminofluorene-(C8)-Guanine Adducts: Molecular Dynamics Simulations Elucidate How Lesion Structure and Base Sequence Context Impact Repair Efficiencies. *Nucleic Acids Res.* **2012**, *40* (19), 9675−9690.
(34) Churchill, C. D. M.; Eriksson, L. A.; Wetmore, S. D. DNA Distortion Caused by Uracil-Containing Intrastrand Cross-Links. *J. Phys. Chem. B* **2016**, *120* (7), 1195−1204.
(35) Bignon, E.; Gattuso, H.; Morell, C.; Dumont, E.; Monari, A. DNA Photosensitization by an Insider: Photophysics and Triplet Energy Transfer of 5-Methyl-2-Pyrimidone Deoxyribonucleoside. *Chem. - Eur. J.* **2015**, *21* (32), 11509−11516.
(36) Faraji, S.; Dreuw, A. Physicochemical Mechanism of Light-Driven DNA Repair by (6−4) Photolyases. *Annu. Rev. Phys. Chem.* **2014**, *65*, 275−292.
(37) Fujimoto, H.; Pinak, M.; Nemoto, T.; O'Neill, P.; Kume, E.; Saito, K.; Maekawa, H. Molecular Dynamics Simulation of Clustered DNA Damage Sites Containing 8-Oxoguanine and Abasic Site. *J. Comput. Chem.* **2005**, *26* (8), 788−798.
(38) Zálešák, J.; Lourdin, M.; Krejčí, L.; Constant, J.-F.; Jourdan, M. Structure and Dynamics of DNA Duplexes Containing a Cluster of Mutagenic 8-Oxoguanine and Abasic Site Lesions. *J. Mol. Biol.* **2014**, *426* (7), 1524−1538.
(39) Bignon, E.; Gattuso, H.; Morell, C.; Dehez, F.; Georgakilas, A. G.; Monari, A.; Dumont, E. Correlation of Bistranded Clustered Abasic DNA Lesion Processing with Structural and Dynamic DNA Helix Distortion. *Nucleic Acids Res.* **2016**, gkw773.
(40) Sutherland, B. M.; Bennett, P. V.; Sidorkina, O.; Laval, J. Clustered DNA Damages Induced in Isolated DNA and in Human Cells by Low Doses of Ionizing Radiation. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (1), 103−108.
(41) Akamatsu, K.; Shikazono, N.; Saito, T. Localization Estimation of Ionizing Radiation-Induced Abasic Sites in DNA in the Solid State Using Fluorescence Resonance Energy Transfer. *Radiat. Res.* **2015**, *183* (1), 105−113.
(42) Chastain, P. D.; Nakamura, J.; Rao, S.; Chu, H.; Ibrahim, J. G.; Swenberg, J. A.; Kaufman, D. G. Abasic Sites Preferentially Form at Regions Undergoing DNA Replication. *FASEB J.* **2010**, *24* (10), 3674−3680.
(43) Chastain, P. D.; Nakamura, J.; Swenberg, J.; Kaufman, D. Nonrandom AP Site Distribution in Highly Proliferative Cells. *FASEB J.* **2006**, *20* (14), 2612−2614.
(44) Lomax, M. E.; Cunniffe, S.; O'Neill, P. Efficiency of Repair of an Abasic Site within DNA Clustered Damage Sites by Mammalian Cell Nuclear Extracts †. *Biochemistry* **2004**, *43* (34), 11017−11026.
(45) Singh, V.; Kumari, B.; Das, P. Repair Efficiency of Clustered Abasic Sites by APE1 in Nucleosome Core Particles Is Sequence and Position Dependent. *RSC Adv.* **2015**, *5*, 23691−23698.
(46) Eccles, L. J.; Menoni, H.; Angelov, D.; Lomax, M. E.; O'Neill, P. Efficient Cleavage of Single and Clustered AP Site Lesions within Mono-Nucleosome Templates by CHO-K1 Nuclear Extract Contrasts with Retardation of Incision by Purified APE1. *DNA Repair* **2015**, *35*, 27−36.
(47) David-Cordonnier, M.-H.; Cunniffe, S. M. T.; Hickson, I. D.; O'Neill, P. Efficiency of Incision of an AP Site within Clustered DNA Damage by the Major Human AP Endonuclease. *Biochemistry* **2002**, *41* (2), 634−642.
(48) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(49) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2015**, *13* (1), 55−58.
(50) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.
(51) Al-Safi, R. I.; Odde, S.; Shabaik, Y.; Neamati, N. Small-Molecule Inhibitors of APE1 DNA Repair Function: An Overview. *Curr. Mol. Pharmacol.* **2012**, *5* (1), 14−35.
(52) Luo, M.; Kelley, M. R. Inhibition of the Human Apurinic/apyrimidinic Endonuclease (APE1) Repair Activity and Sensitization of Breast Cancer Cells to DNA Alkylating Agents with Lucanthone. *Anticancer Res.* **2004**, *24* (4), 2127−2134.
(53) Poletto, M.; Malfatti, M. C.; Dorjsuren, D.; Scognamiglio, P. L.; Marasco, D.; Vascotto, C.; Jadhav, A.; Maloney, D. J.; Wilson, D. M.; Simeonov, A.; et al. Inhibitors of the Apurinic/apyrimidinic Endonuclease 1 (APE1)/nucleophosmin (NPM1) Interaction That Display Anti-Tumor Properties. *Mol. Carcinog.* **2016**, *55* (5), 688−704.

DOI: 10.1021/acs.jpclett.6b01692
J. Phys. Chem. Lett. 2016, 7, 3760−3765

General conclusions

Through this thesis work, we investigated DNA damage formation, structure, and repair by molecular modelling. The main goal of this research project was to provide new clear-cut insights about DNA reactivity, mechanical properties and interactions with repair enzymes by means of QM, MM-MD and QM/MM(-MD) calculations. Such investigations are real issues in current biochemistry, that improve our understanding of the living world, and provide perspective for biomedicine.

DNA reactivity was investigated through three projects. On the one hand, we studied the mechanisms of formation of the 8-oxo-7,8-dihydroguanine. We assessed the feasability of some intermediates that were missing experimental proofs of their existence. To do so, we performed QM, MM-MD and QM/MM-MD calculations, in collaboration with Dr. Jean-Luc Ravanat from CEA Grenoble, France, and Pr. Iñaki Tuñon and Dr. Juan Aranda from the University of Valencia, Spain. A preliminary study on isolated purines was published *Nucleic Acids Research*, and a second one in which we probed the helix environment effect on the reaction was released in *Chemistry - A European Journal*. Throughout these two papers, we shown that $^1O_2$ attack onto guanine is largely favored by the DNA environment, due to the negatively-charged phosphates that promote the zwiterionic intermediate closure to form an endoperoxide moiety. Our calculations also brought out the feasability of the latter intermediates proposed by Dr. Ravanat, which existence have not been firmly established yet by experimental means.
On the other hand, we assessed the photosensitizing potential of the pyrimidine 6-4 pyrimidone product, one of the main UV-induced lesions with the cyclobutane pyrimidine dimers. The chromophore of this photolesion, called Pyo, absorbs in the UVA wavelenghts, and can induce a triplet triplet energy transfer (TTET) towards an adjacent thymine. This phenomenon results in the formation of very deleterious tandem lesions. We investigated the mechanical properties of Pyo as an artificial nucleobase in dsDNA, and its photosensitizer potential by MM-MD and QM-MM calculations. This project was led in collaboration with Dr. Antonio Monari and Hugo Gattuso from Nancy University, France, and was published in *Chemistry - A European Journal*. These calculations highlighted the effencient DNA photosensitization by the Pyo chromophore. The barrier of the TTET has been estimated to 0.4 eV, which underlined the ease of this process within B-DNA. The evaluation of Pyo mechanical properties in dsDNA revealed that this moiety does not involve strong deformation of the helix, offering good opportunities for its use as potentiel radiotherapy agent.
The Short Term Scientific Mission that I have done in Pr. Leif Eriksson's lab (Gothenburg University, Sweden, COST action CM1201 grant) led to a project on DNA photosensitization by two nonsteroidal anti-inflammatory drugs, ketoprofen and ibuprofen. As well in collaboration with theoreticians from Nancy, we performed MM-MD and QM calculations in order to unravel interaction modes and photosentizing properties of these two very common drugs on DNA. These results were recently submitted in *Scientific Re-*

ports. Through this project we delineated ketoprofen and ibuprofen interaction modes with DNA. Furthermore, we shown that these two drugs may effectively act as DNA photosensitizers due to middle-range photodissociation, underlying the dangerous effects that anti-inflammatory creams might induce on skin cells when exposed to UV-light.

On the other hand, we studied DNA mechanical properties in four different systems. The first one concerned sequence effects of clustered abasic sites, in collaboration with Nancy coworkers and Dr. Alexandros Georgakilas from NTU Athens, Greece, through the COST action CM1201. Experiments of Dr. Georgakilas shown that clustered abasic sites repair efficiency highly depends on the relative position of the two lesions in the oligonucleotide, thus highlighting strong sequence effects. In order to obtain structural informations of the seven sequences used in the experimental study, we performed MM-MD calculations. Our results nicely fitted to the experiments, with very strong distorsions were obtained for the less repaired sequences, while only local deformations corresponded to the sequences showing the highest repair rates. Thus, we managed to delineate subtle sequence effects, which are often hard to tackle since it is a huge work to take into account so many sequences. These results were published in *Nucleic Acids Research*.
Likewise, we investigated the structural impact involved by interstrand cross-links (ICL), which result from the reaction between an abasic site and a nucleobase on the opposite strand. These damages are highly cytotoxic, since they prevent strands separation hence the DNA replication. Albeit their highly deleterious effects, ICL are ill-understood. Indeed, experimental structures of such complex lesions are very hard to obtain. Noteworthy, ICL repair mechanisms are still far from being completely unraveled. Therefore, there is a lack of information concerning structural properties, to which we searched to palliate with our MM-MD calculations. To do so, we relied on a recent experimental study which concerned two oligonucleotides with different sequences. One of these favored the formation of Ap-dA cross-link while the other one promoted Ap-dG generation, with higher formation yields for Ap-dA. We first brought out difference in structural behavior of the two sequences. The Ap-dA oligonucleotide exhibits one stable conformation, whereas the other one (favoring Ap-dG) exhibited two co-existing structures. Furthermore, we brought energetics informations through QM computations that explained the difference of formation yields between two sequences, showing that reaction is thermodynamically driven. This study was led in collaboration with Dr. Filip Lankaš and Dr. Tomas Dršata from the Praha University, Czech Repuclic and was published in *Nucleic Acids Research*.
In the context of the reactivity study we led about the 6-4PP chromophore Pyo, we performed MM-MD calculation in order to assess the dynamical behaviour of the entire lesion, and the other main UV-induced product, the cyclo butane pyrimidine dimer (CPD). Indeed, experimental studies revealed that the two lesions exhibit very different repair rates : CPDs are far less efficiently repaired than 6-4PPs. Crystal structures and FRET measurements are available for both lesions, but their mechanical properties remain controversial. In order to shed light on their dynamical behavior, we performed extensive classical molecular dynamics (up to 2 μs). Our results brought out an evident conformational polymorphism of the 6-4PP which contrasts with the structural invarience of the CPD-containing oligomer. This project was led in collaboration with our common coworkers from Nancy University (notably Dr. François Dehez) and associated results were released in a publication in *Nucleic Acids Research*.
The fourth system we investigated in this context concerned the binding mode of three polyamines to dsDNA. These small molecules can provide a protective role for DNA as radical scavengers, but they can also induce cross-links with the guanine radical cation. Reactivity of polyamines is well known, but so far no investigation about their interaction

mode with DNA have been reported. To palliate this lack of structural informations, we performed all-atom MD simulations on three poylamines, putrescine, spermidine and spermine, taking into account their possible protonation state (up to +4 for spermine). These calculations revealed that the binding mode of polyamines depends on their protonation state and their length. Thus, the small putrescine will prefer to bind in the major groove whereas the longer spermine will bind in the minor groove (and in the major groove for the monocationic species). Furthermore, we assessed the proximity of amino groups with guanines distribution of distances. It allowed to highlight the fact that mono- and dicationic polyamines are often close to guanine C8 (with which they can form cross-link), whereas +3 and +4 charged ones prefer to slide on the backbone, interacting with the negatively-charged phosphates. This study was performed in collaboration with Dr. Ravanat and was submitted to *Journal of Physical Chemistry B*.

The third part of this thesis work concerned DNA repair. In this context, we led two projects. The first one was performed in perspective of clustered abasic sites structural study. Indeed, we investigated the interaction of the APE1 enzyme with the seven sequences we already investigated as isolated oligonucleotides. APE1 is the Ap-endonuclease responsible for Ap sites cleavage in human, and experiments of our collaborator Dr. Georgakilas shown that sequence effects induced by multiply damaged sites highly impact their removal by this enzyme. This study gave insights into structural properties of the DNA:APE1 complex, which are of major importance for DDR since Ap site are abondant cytotoxic damages in cells. We notably shed light on two amino acids that behave as a pincer in the gap let by the Ap site when rotated towards the enzyme's active site. This project was performed in collaboration with our coworkers from Nancy University and led to one publication in *Journal of Physical Chemistry Letters*.
The second project concerning DNA repair was about interactions between the Fpg protein and oligonucleotide containing tandem 8-oxoG - two adjacent lesions on the same strand. The Fpg protein is responsible for the specific N-glycosydic bond cleavage of 8-oxoG in human cells. Using MM-MD calculations, we shown that the flipping of 8-oxoG is disfavored in the case of tandem lesions compared to single ones. Thus, such tandem damages are particularly challenging for glycosylases as they prevent damaged base eversion to active site of the enzyme.

To conclude, during this thesis work we investigated the reactivity, structure and repair of several damaged DNA systems by means of molecular modelling. The majority of our projects successed and this open door to layer-scale, predictive computational investigations upstream to experiments. These projects were performed in close collaboration with French and international researchers. They brought out unprecedent insights concerning mechanisms of formation and structure of complex lesions. Noteworthy, they provide investigations about subtle sequence effects of multiply damaged sites in naked DNA and in the presence of a repair enzyme. Such results highlighted the robustness of our methodology, providing large perspectives for damaged DNA studies. Moreover, this thesis' projects provided a contribution to the understanding of the complex DNA-protein machinery, which has been and will always be one of the major issues of researchers.
The next step will be to apply these protocols to nucleosomal DNA. Few crystal structures of such system containing damages (eg 6-4PP and Ap sites) are already available, and offer good starting structures for our calculations. Thus, such studies would provide unprecedent insights into damaged nucleosome mechanical properties albeit induce an obvious drastic increase of calculation times.

Hopefully, we nowadays assist to an exponential improvement of computational technologies power, and simulations times that are now accessible (up to several μs) could have barely considered several decades ago. Such advances could provide considerably enhanced computational ressources, offering large perspectives in our field. This combined with advances of experimental techniques could let one dream about systems reachable for calculations in few years.
A damaged nucleosome in interaction with a repair enzyme ? A damaged oligonucleotide in interaction with repair multienzymatic complex ?
In all cases, the amazing complexity of the living world will always let us matters of research and the exploration of the living, its history and its functioning is before us.

L'avancée de la science découle de nouvelles techniques, de découvertes et de nouvelles idées, probablement dans cet ordre.

Sydney Brenner, Nobel Prize 2002

List of publications

[P10]. Bignon, E.; Marazzi, M.; Besancenot, V.; Gattuso, H.; Morell, C.; Eriksson, L.; Grandemange, S.; Dumont, E.; Monari, A. Ibuprofen and ketoprofen photosensitization mechanism: how to enhance cells death under the influence of UVA light, accepted in *Sci. Rep.*

[P9]. Bignon, E.; Chan, C-H.; Morell, C.; Ravant, J-L.; Dumont, E. Molecular dynamics insights into polyamines-DNA interaction: implications for cross-links selectivity, accepted in *Chem. Eur. J.*

[P8]. Dehez, F; Gattuso, H; Bignon, E; Morell, C; Dumont, E; Monari, A. Conformational polymorphism or structural invariance in DNA photoinduced lesions: implications for repair rates. *Nucl. Acids Res.*, **2017**, published online.

[P7]. Bignon, E.; Dršata, T.; Morell, C.; Lankaš, F.; Dumont, E. Interstrand cross-linking implies contrasting structural consequences for DNA: insights from molecular dynamics, *Nucl. Acids Res.*, **2016**, published online, doi : 10.1093/nar/gkw1253.

[P6]. Gattuso, H.; Durand, E.; Bignon, E.; Morell, C.; Georgakilas, A. G.; Dumont, E.; Chipot, C.; Dehez, F.; Monari, A. Repair Rate of Clustered Abasic DNA Lesions by Human Endonuclease: Molecular Bases of Sequence Specificity, *J. Phys. Chem. Lett.*, **2016**, 7, 3760-3765. doi : 10.1021/acs.jpclett.6b01692.

[P5]. Bignon, E.; Gattuso, H.; Morell, C.; Dehez, F.; Georgakilas, A. G.; Monari, A.; Dumont, E. Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion, *Nucl. Acids Res.*, **2016**, published online, doi : 10.1093/nar/gkw773.

[P4]. Dumont, E.; Grüber, R.; Bignon, E.; Morell, C.; Aranda, J.; Tunon, I. Singlet Oxygen Attack on Guanine: Reactivity and Structural Signature within the B-DNA Helix, *Chem. Eur. J.*, **2016**, 22, 12358-12362. doi : 10.1002/chem.201601287.

[P3]. Dumont, E.; Grüber, R.; Bignon, E.; Morell, C.; Moreau, Y.; Monari, A.; Ravanat, JL. Probing the reactivity of singlet oxygen with purines, *Nucl. Acids Res.*, **2016**, 44, 56-62. doi : 10.1093/nar/gkv1364.

[P2]. Morell, C.; Tognetti, V.; Bignon, E.; Hernandez-Haro, N.; Herrera, B.; Grand, A.; Gutierrez-Oliva, S.; Joubert, L.; Toro-Labbé, A.; Chermette, H. Insights into the chemical meanings of the reaction electronic flux, *Theor. Chem. Acc.*, **2015**, 134, 133. doi : 10.1007/s00214-015-1730-7.

[P1]. Bignon, E.; Gattuso, H.; Morell, C.; Dumont, E.; Monari, A. DNA photosensitization by an "insider". Photophysics and triplet energy-transfer of 5-methyl-2-pyrimidone, *Chem. Eur. J.*, **2015**, 32, 11509-11516. doi : 10.1002/chem.201501212.

Bibliography

[1] James D Watson and Francis HC Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–8, 1953.

[2] Rivka Rudner, John D Karkas, and Erwin Chargaff. Separation of b. subtilis DNA into complementary strands, I. biological properties. *Proceedings of the National Academy of Sciences*, 60(2):630–635, 1968.

[3] John D Karkas, Rivka Rudner, and Erwin Chargaff. Separation of b. subtilis DNA into complementary strands. II. template functions and composition as determined by transcription with RNA polymerase. *Proceedings of the National Academy of Sciences*, 60(3):915–920, 1968.

[4] Rivka Rudner, John D Karkas, and Erwin Chargaff. Separation of b. subtilis DNA into complementary strands. 3. direct analysis. *Proceedings of the National Academy of Sciences*, 60(3):921–922, 1968.

[5] International Human Genome Sequencing Consortium et al. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, 2004.

[6] Francis Crick, Leslie Barnett, Sydney Brenner, and Richard J Watts-Tobin. *General nature of the genetic code for proteins*. Macmillan Journals Limited, 1961.

[7] Marshall W Nirenberg and J Heinrich Matthaei. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences*, 47(10):1588–1602, 1961.

[8] Earl Ubell. The code of life finally cracked. *New Horizons in Science*, 1961.

[9] Carl W Schmid and Prescott L Deininger. Sequence organization of the human genome. *Cell*, 6(3):345–358, 1975.

[10] Jerzy Jurka, Vladimir V Kapitonov, Oleksiy Kohany, and Michael V Jurka. Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.*, 8:241–259, 2007.

[11] George E Palade. A small particulate component of the cytoplasm. *The Journal of biophysical and biochemical cytology*, 1(1):59, 1955.

[12] Richard B Roberts. *Microsomal particles and protein synthesis*. Pergamon: London, 1958.

[13] Orlando D Schärer. Chemistry and biology of DNA repair. *Angewandte Chemie International Edition*, 42(26):2946–2974, 2003.

[14] Philippe Cluzel, Anne Lebrun, Christoph Heller, Richard Lavery, Jean-Louis Viovy, Didier Chatenay, and François Caron. DNA: an extensible molecule. *Science*, pages 792–794, 1996.

[15] Marc Poncin, Brigitte Hartmann, and Richard Lavery. Conformational sub-states in B-DNA. *Journal of molecular biology*, 226(3):775–794, 1992.

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

[17] Alberto Pérez, F Javier Luque, and Modesto Orozco. Dynamics of B-DNA on the microsecond time scale. *Journal of the American Chemical Society*, 129(47):14739–14745, 2007.

[18] Elise Dumont and Antonio Monari. Understanding DNA under oxidative stress and sensitization: the role of molecular modeling. *Front. Chem.*, 3, 2015.

[19] David L Beveridge, Thomas E Cheatham, and Mihaly Mezei. The ABCs of molecular dynamics simulations on B-DNA, circa 2012. *Journal of biosciences*, 37(3):379–397, 2012.

[20] Jean-Louis Basdevant, Jean Dalibard, and Manuel Joffre. *Mécanique quantique*. Editions Ecole Polytechnique, 2002.

[21] Christopher J Cramer. *Essentials of computational chemistry: theories and models*. John Wiley & Sons, 2013.

[22] David Maurice and Martin Head-Gordon. Analytical second derivatives for excited electronic states using the single excitation configuration interaction method: theory and application to benzo [a] pyrene and chalcone. *Molecular Physics*, 96(10):1533–1541, 1999.

[23] Martin Head-Gordon, Rudolph J Rico, Manabu Oumi, and Timothy J Lee. A doubles correction to electronic excited states from configuration interaction in the space of single substitutions. *Chemical Physics Letters*, 219(1-2):21–29, 1994.

[24] George D Purvis III and Rodney J Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics*, 76(4):1910–1918, 1982.

[25] Krishnan Raghavachari, Gary W Trucks, John A Pople, and Martin Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, 157(6):479–483, 1989.

[26] Troy Van Voorhis and Martin Head-Gordon. Two-body coupled cluster expansions. *The Journal of Chemical Physics*, 115(11):5033–5040, 2001.

[27] Chr Møller and Milton S Plesset. Note on an approximation treatment for many-electron systems. *Physical Review*, 46(7):618, 1934.

[28] R Krishnan and John A Pople. Approximate fourth-order perturbation theory of the electron correlation energy. *International Journal of Quantum Chemistry*, 14(1):91–100, 1978.

[29] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[30] Carsten A Ullrich. *Time-dependent density-functional theory: concepts and applications*. OUP Oxford, 2011.

[31] Emmanuelle Bignon, Hugo Gattuso, Christophe Morell, Elise Dumont, and Antonio Monari. DNA photosensitization by an "insider": Photophysics and triplet energy transfer of 5-methyl-2-pyrimidone deoxyribonucleoside. *Chemistry–A European Journal*, 21(32):11509–11516, 2015.

[32] Erich Runge and Eberhard KU Gross. Density-functional theory for time-dependent systems. *Physical Review Letters*, 52(12):997, 1984.

[33] Kenichi Fukui. Theory of orientation and stereoselection. In *Orientation and Stereoselection*, pages 1–85. Springer, 1970.

[34] Robert G Parr. Density functional theory of atoms and molecules. In *Horizons of Quantum Chemistry*, pages 5–15. Springer, 1980.

[35] Robert G Parr, Robert A Donnelly, Mel Levy, and William E Palke. Electronegativity: the density functional viewpoint. *The Journal of Chemical Physics*, 68(8):3801–3807, 1978.

[36] Ralph G Pearson. *Chemical hardness*. Wiley-VCH, 1997.

[37] Robert G. Parr and Weitao Yang. *Density Functional Theory of atoms and molecules*. Oxford University Press, 1989.

[38] Raymond P Iczkowski and John L Margrave. Electronegativity. *Journal of the American Chemical Society*, 83(17):3547–3551, 1961.

[39] Robert G. Parr and Ralph G. Pearson. Absolute hardness: companion parameter to absolute electronegativity. *J. Am. Chem. Soc.*, 105:7512–7516, 1983.

[40] Paul Geerlings, Stijn Fias, Zino Boisdenghien, and Frank De Proft. Conceptual DFT: chemistry from the linear response function. *Chemical Society Reviews*, 43(14):4989–5008, 2014.

[41] Robert G. Parr and Weitao Yang. Density functional approach to the frontier-electron theory of chemical reactivity. *Journal of the American Chemical Society*, 106(14):4049–4050, 1984.

[42] RG Parr and W Yang. Density functional theory of atoms and molecules oxford univ. *Press, New York*, 1989.

[43] EKU Gross and Walter Kohn. Local density-functional theory of frequency-dependent linear response. *Physical review letters*, 55(26):2850, 1985.

[44] Christophe Morell, André Grand, Alejandro Toro-Labbé, and Henry Chermette. Is hyper-hardness more chemically relevant than expected? *Journal of molecular modeling*, 19(7):2893–2900, 2013.

[45] Christophe Morell, André Grand, and Alejandro Toro-Labbé. Theoretical support for using the δf (r) descriptor. *Chemical physics letters*, 425(4):342–346, 2006.

[46] Christophe Morell, André Grand, and Alejandro Toro-Labbé. New dual descriptor for chemical reactivity. *The Journal of Physical Chemistry A*, 109(1):205–212, 2005.

[47] Weitao Yang and Robert G Parr. Hardness, softness, and the fukui function in the electronic theory of metals and catalysis. *Proceedings of the National Academy of Sciences*, 82(20):6723–6726, 1985.

[48] P Geerlings and F De Proft. Conceptual DFT: the chemical relevance of higher response functions. *Physical Chemistry Chemical Physics*, 10(21):3028–3042, 2008.

[49] Eleonora Echegaray and Alejandro Toro-Labbe. Reaction electronic flux: a new concept to get insights into reaction mechanisms. study of model symmetric nucleophilic substitutions. *The Journal of Physical Chemistry A*, 112(46):11801–11807, 2008.

[50] Patricio Flores-Morales, Soledad Gutiérrez-Oliva, Eduardo Silva, and Alejandro Toro-Labbé. The reaction electronic flux: A new descriptor of the electronic activity taking place during a chemical reaction. application to the characterization of the mechanism of the schiff's base formation in the maillard reaction. *Journal of Molecular Structure: THEOCHEM*, 943(1):121–126, 2010.

[51] Martìnez Jorge Ignacio and Alejandro Toro-Labbé. Reaction electronic flux as a fluctuation of relative interatomic electronic populations. *The Journal of Physical Chemistry C*, 119(6):3040–3049, 2015.

[52] Kenichi Fukui. Formulation of the reaction coordinate. *The Journal of Physical Chemistry*, 74(23):4161–4163, 1970.

[53] Kenichi Fukui. The path of chemical reactions-the IRC approach. *Accounts of chemical research*, 14(12):363–368, 1981.

[54] Christophe Morell, Vincent Tognetti, Emmanuelle Bignon, Elise Dumont, Noemi Hernandez-Haro, Barbara Herrera, André Grand, Soledad Gutiérrez-Oliva, Laurent Joubert, Alejandro Toro-Labbé, et al. Insights into the chemical meanings of the reaction electronic flux. *Theoretical Chemistry Accounts*, 134(11):133, 2015.

[55] John C Slater. Atomic shielding constants. *Physical Review*, 36(1):57, 1930.

[56] S Francis Boys. Electronic wave functions. i. a general method of calculation for the stationary states of any molecular system. 200(1063):542–554, 1950.

[57] Peter MW Gill. Molecular integrals over gaussian basis functions. *Advances in quantum chemistry*, 25:141–205, 1994.

[58] Ernest R Davidson and David Feller. Basis set selection for molecular calculations. *chem. Rev*, 86(4):681–696, 1986.

[59] Ivan Ivani, Pablo D Dans, Agnes Noy, Alberto Pérez, Ignacio Faustino, Adam Hospital, Jürgen Walther, Pau Andrio, Ramon Goñi, Alexandra Balaceanu, et al. Parmbsc1: a refined force field for DNA simulations. *Nature methods*, 13(1):55–58, 2016.

[60] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[61] Joel A Shapiro. Classical mechanics, 2003.

[62] Abdulnour Toukmaji, Celeste Sagui, John Board, and Tom Darden. Efficient particle-mesh ewald based approach to fixed and induced dipolar interactions. *The Journal of chemical physics*, 113(24):10913–10927, 2000.

[63] Celeste Sagui, Lee G Pedersen, and Thomas A Darden. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *The Journal of chemical physics*, 120(1):73–87, 2004.

[64] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. Amber 12, 2012.

[65] Jesus A Izaguirre, Daniel P Catarello, Justin M Wozniak, and Robert D Skeel. Langevin stabilization of molecular dynamics. *The Journal of chemical physics*, 114(5):2090–2098, 2001.

[66] Richard Lavery, Krystyna Zakrzewska, David Beveridge, Thomas C. Bishop, David A. Case, Thomas Cheatham, Surjit Dixit, B. Jayaram, Filip Lankas, Charles Laughton, John H. Maddocks, Alexis Michon, Roman Osman, Modesto Orozco, Alberto Perez, Tanya Singh, Nada Spackova, and Jiri Sponer. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Research*, 38(1):299–313, 2010.

[67] Elise Dumont, Raymond Grüber, Emmanuelle Bignon, Christophe Morell, Yohann Moreau, Antonio Monari, and Jean-Luc Ravanat. Probing the reactivity of singlet oxygen with purines. *Nucleic acids research*, 44(1):56–62, 2015.

[68] Elise Dumont, Raymond Grüber, Emmanuelle Bignon, Christophe Morell, Juan Aranda, Jean-Luc Ravanat, and Iñaki Tuñón. Singlet oxygen attack on guanine: Reactivity and structural signature within the B-DNA helix. *Chemistry-A European Journal*, 22(35):12358–12362, 2016.

[69] Jiande Gu, Jing Wang, and Jerzy Leszczynski. Electron attachment-induced DNA single-strand breaks at the pyrimidine sites. *Nucleic acids research*, page gkq304, 2010.

[70] Raymond Gruüber, Antonio Monari, and Elise Dumont. Stability of the guanine endoperoxide intermediate: a computational challenge for density functional theory. *The Journal of Physical Chemistry A*, 118(49):11612–11619, 2014.

[71] Piotr Cieplak, Wendy D Cornell, Christopher Bayly, and Peter A Kollman. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry*, 16(11):1357–1377, 1995.

[72] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174, 2004.

[73] Pengyu Ren and Jay W Ponder. Consistent treatment of inter-and intramolecular polarization in molecular mechanics calculations. *Journal of computational chemistry*, 23(16):1497–1506, 2002.

[74] Hai Lin and Donald G Truhlar. Qm/mm: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts*, 117(2):185, 2007.

[75] Debananda Das, Kirsten P Eurenius, Eric M Billings, Paul Sherwood, David C Chatfield, Milan Hodošček, and Bernard R Brooks. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *The Journal of chemical physics*, 117(23):10534–10547, 2002.

[76] Alex H De Vries, Paul Sherwood, Simon J Collins, Anthony M Rigby, Marcello Rigutto, and Gert Jan Kramer. Zeolite structure and reactivity by combined quantum-chemical- classical calculations. *The Journal of Physical Chemistry B*, 103(29):6133–6141, 1999.

[77] Tom K Woo, Luigi Cavallo, and Tom Ziegler. Implementation of the IMOMM methodology for performing combined QM/MM molecular dynamics simulations and frequency calculations. *Theoretical Chemistry Accounts*, 100(5-6):307–313, 1998.

[78] Stefan Dapprich, István Komáromi, K Suzie Byun, Keiji Morokuma, and Michael J Frisch. A new ONIOM implementation in gaussian98. part I. the calculation of energies, gradients, vibrational frequencies and electric field derivatives. *Journal of Molecular Structure: THEOCHEM*, 461:1–21, 1999.

[79] Dirk Bakowies and Walter Thiel. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *The Journal of Physical Chemistry*, 100(25):10580–10594, 1996.

[80] U Chandra Singh and Peter A Kollman. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl+ Cl- exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry*, 7(6):718–730, 1986.

[81] Hai Lin and Donald G Truhlar. Redistributed charge and dipole schemes for combined quantum mechanical and molecular mechanical calculations. *The Journal of Physical Chemistry A*, 109(17):3991–4004, 2005.

[82] Johannes Grotendorst. Modern methods and algorithms of quantum chemistry. *John von Neumann Institute for Computing*, 2000.

[83] Tinker molecular modeling software. https://dasher.wustl.edu/tinker/.

[84] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G Scalmani, V Barone, B Mennucci, GA Petersson, et al. Gaussian 09, revision D. 01, 2009.

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

[86] Nicolas Ferré and Xavier Assfeld. Application of the local self-consistent-field method to core-ionized and core-excited molecules, polymers, and proteins: True orthogonality between ground and excited states. *The Journal of chemical physics*, 117(9):4119–4125, 2002.

[87] Martin J Field, Marc Albe, Céline Bret, Proust-De Martin, Aline Thomas, et al. The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *Journal of Computational Chemistry*, 21(12):1088–1100, 2000.

[88] Jean Cadet, Thierry Douki, Carine Badouard, Alain Favier, and Jean-Luc Ravanat. Oxidatively generated damage to cellular DNA: mechanistic aspects. In *Oxidative Damage to Nucleic Acids*, pages 1–13. Springer, 2007.

[89] Bo Durbeej and Leif A Eriksson. Reaction mechanism of thymine dimer formation in DNA induced by UV light. *Journal of Photochemistry and Photobiology A: Chemistry*, 152(1):95–101, 2002.

[90] Juan José Serrano-Pérez, Israel Gonzalez-Ramirez, Pedro B Coto, Manuela Merchán, and Luis Serrano-Andrés. Theoretical insight into the intrinsic ultrafast formation of cyclobutane pyrimidine dimers in UV-irradiated DNA: Thymine versus cytosine. *The Journal of Physical Chemistry B*, 112(45):14096–14098, 2008.

[91] Xifeng Li, Michael D Sevilla, and Léon Sanche. Density functional theory studies of electron interaction with DNA: can zero ev electrons induce strand breaks? *Journal of the American Chemical Society*, 125(45):13668–13669, 2003.

[92] Jerzy Leszczynski. *Computational chemistry: reviews of current trends*, volume 8. World Scientific, 2003.

[93] Anirban Banerjee, Wei Yang, Martin Karplus, and Gregory L Verdine. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *nature*, 434(7033):612, 2005.

[94] Julian Garrec, Chandan Patel, Ursula Rothlisberger, and Elise Dumont. Insights into intrastrand cross-link lesions of DNA from QM/MM molecular dynamics simulations. *Journal of the American Chemical Society*, 134(4):2111–2119, 2012.

[95] Arturo Robertazzi and James A Platts. A QM/MM study of cisplatin–DNA oligonucleotides: from simple models to realistic systems. *Chemistry–A European Journal*, 12(22):5747–5756, 2006.

[96] Elise Dumont, Meilani Wibowo, Daniel Roca-Sanjuán, Marco Garavelli, Xavier Assfeld, and Antonio Monari. Resolving the benzophenone DNA-photosensitization mechanism at QM/MM level. *The journal of physical chemistry letters*, 6(4):576–580, 2015.

[97] S. Madronich, Robert L. McKenzie, L. O. Björn, and Phillip Mack Caldwell. Changes in biologically active ultraviolet radiation reaching the earth's surface. *Journal of Photochemistry and Photobiology B*, 46(1-3):5–19, 1998.

[98] Rajeshwar P. Sinha and Donat-P. Häder. UV-induced DNA damage and repair: a review. *Photochemical and Photobiological Sciences*, 1:225–236, 2002.

[99] Victoria Vendrell-Criado, Gemma M Rodríguez-Muñiz, M Consuelo Cuquerella, Virginie Lhiaubet-Vallet, and Miguel A Miranda. Photosensitization of DNA by 5-methyl-2-pyrimidone deoxyribonucleoside:(6-4) photoproduct as a possible trojan horse. *Angewandte Chemie International Edition*, 52(25):6476–6479, 2013.

[100] Stéphane Mouret, Caroline Baudouin, Marie Charveron, Alain Favier, Jean Cadet, and Thierry Douki. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proceedings of the National Academy of Sciences*, 103(37):13765–13770, 2006.

[101] T. Douki and E. Sage. Dewar valence isomers, the third type of environmentally relevant DNA photoproducts induced by solar radiation. *Photochem. Photobiol. Sci.*, 15:24–30, 2016.

[102] Peng Mao, Michael J Smerdon, Steven A Roberts, and John J Wyrick. Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proceedings of the National Academy of Sciences*, page 201606667, 2016.

[103] Jean-Luc Ravanat, Thierry Douki, and Jean Cadet. Direct and indirect effects of uv radiation on DNA and its components. *Journal of Photochemistry and Photobiology B: Biology*, 63(1):88–102, 2001.

[104] Linda Antusch, Nadine Gaß, and Hans-Achim Wagenknecht. Elucidation of the dexter-type energy transfer in DNA by thymine–thymine dimer formation using photosensitizers as artificial nucleosides. *Angewandte Chemie International Edition*, 2016.

[105] M Consuelo Cuquerella, Virginie Lhiaubet-Vallet, Jean Cadet, and Miguel A Miranda. Benzophenone photosensitized DNA damage. *Accounts of chemical research*, 45(9):1558–1570, 2012.

[106] Jean Cadet, Thierry Douki, and Jean-Luc Ravanat. Oxidatively generated damage to the guanine moiety of DNA: mechanistic aspects and formation in cells. *Accounts of chemical research*, 41(8):1075–1083, 2008.

[107] J-P Pouget, T Douki, M-J Richard, and J Cadet. DNA damage induced in cells by γ and UVA radiation as measured by HPLC/GC- MS and HPLC- EC and comet assay. *Chemical research in toxicology*, 13(7):541–549, 2000.

[108] Virginie Lhiaubet, Nicole Paillous, and Nadia Chouini-Lalanne. Comparison of DNA damage photoinduced by ketoprofen, fenofibric acid and benzophenone via electron and energy transfer. *Photochemistry and photobiology*, 74(5):670–678, 2001.

[109] Alexander P Darmanyan and Christopher S Foote. Solvent effects on singlet oxygen yield from n, π^* and π, π^* triplet carbonyl compounds. *Journal of physical chemistry*, 97(19):5032–5035, 1993.

[110] Daniel Peña, Cristina Martí, Santi Noneil, Luis A Martínez, and Miguel A Miranda. Time-resolved near infrared studies on singlet oxygen production by the photosensitizing 2-arylpropionic acids. *Photochemistry and photobiology*, 65(5):828–832, 1997.

[111] Sandra Monti, Salvatore Sortino, Guido De Guidi, and Giancarlo Marconi. Photochemistry of 2-(3-benzoylphenyl) propionic acid (ketoprofen) part 1a picosecond and nanosecond time resolved study in aqueous solution. *Journal of the Chemical Society, Faraday Transactions*, 93(13):2269–2275, 1997.

[112] David Costa, Ana Gomes, José LFC Lima, and Eduarda Fernandes. Singlet oxygen scavenging activity of non-steroidal anti-inflammatory drugs. *Redox Report*, 13(4):153–160, 2008.

[113] Christian Triantaphylidès and Michel Havaux. Singlet oxygen in plants: production, detoxification and signaling. *Trends in plant science*, 14(4):219–228, 2009.

[114] Francis Wilkinson, W Phillip Helman, and Alberta B Ross. Rate constants for the decay and reactions of the lowest electronically excited singlet state of molecular oxygen in solution. an expanded and revised compilation. *Journal of Physical and Chemical Reference Data*, 24(2):663–677, 1995.

[115] Jean-Luc Ravanat, Christine Saint-Pierre, and Jean Cadet. One-electron oxidation of the guanine moiety of 2'-deoxyguanosine: influence of 8-oxo-7, 8-dihydro-2'-deoxyguanosine. *Journal of the American Chemical Society*, 125(8):2030–2031, 2003.

[116] Jean Cadet, Thierry Douki, Jean-Luc Ravanat, and Paolo Di Mascio. Reactions of singlet oxygen with nucleic acids. *Singlet Oxygen: Applications in Biosciences and Nanosciences Volume 1*, page 393, 2016.

[117] TP Devasagayam, Steen Steenken, MS Obendorf, Wolfgang A Schulz, and Helmut Sies. Formation of 8-hydroxy (deoxy) guanosine and generation of strand breaks at guanine residues in DNA by singlet oxygen. *Biochemistry*, 30(25):6283–6289, 1991.

[118] Helmut Sies and Carlos FM Menck. Singlet oxygen induced DNA damage. *Mutation Research/DNAging*, 275(3-6):367–375, 1992.

[119] Victoria Vendrell-Criado, Gemma M Rodríguez-Muñiz, Virginie Lhiaubet-Vallet, M Consuelo Cuquerella, and Miguel A Miranda. The (6–4) dimeric lesion as a DNA photosensitizer. *ChemPhysChem*, 17(13):1979–1982, 2016.

[120] Chimin Sheu and Christopher S Foote. Endoperoxide formation in a guanosine derivative. *Journal of the American Chemical Society*, 115(22):10446–10447, 1993.

[121] Anirban Ghosh and Manju Bansal. A glossary of DNA structures from A to Z. *Acta Crystallographica Section D: Biological Crystallography*, 59(4):620–626, 2003.

[122] Richard E Dickerson. [5] DNA structure from A to Z. *Methods in enzymology*, 211:67–111, 1992.

[123] Struther Arnott, Rengaswami Chandrasekaran, DL Birdsall, AGW Leslie, and RL Ratliff. Left-handed DNA helices. *Nature*, 283(5749):743–745, 1980.

[124] Struther Arnott, DWL Hukins, and SD Dover. Optimised parameters for RNA double-helices. *Biochemical and biophysical research communications*, 48(6):1392–1399, 1972.

[125] Richard Lavery and Krystyna Zakrzewska. Base and base pair morphologies, helical parameters, and definitions. *Oxford Handbook of Nucleic Acid Structure. Oxford University Press, NY*, pages 39–76, 1999.

[126] Maxim D Frank-Kamenetskii and Sergei M Mirkin. Triplex DNA structures. *Annual review of biochemistry*, 64(1):65–95, 1995.

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

[128] Calliste Reiling and Luis A Marky. Contributions of the loops on the stability and targeting of DNA pseudoknots. *Biochemical Compounds*, 2(1):3, 2014.

[129] David Bikard, Céline Loot, Zeynep Baharoglu, and Didier Mazel. Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiology and Molecular Biology Reviews*, 74(4):570–588, 2010.

[130] Richard E Dickerson and Horace R Drew. Structure of a B-DNA dodecamer: II. influence of base sequence on helix structure. *Journal of molecular biology*, 149(4):761–786, 1981.

[131] Tomas Lindahl. Instability and decay of the primary structure of DNA. *Nature*, 362(6422):709–715, 1993.

[132] Wei Yang. Structure and mechanism for DNA lesion recognition. *Cell research*, 18(1):184–197, 2008.

[133] Richard R Sinden. *DNA structure and function*. Elsevier, 2012.

[134] Hirofumi Ohishi, Mamiko Odoko, Kazimierz Grzeskowiak, Yoichi Hiyama, Koji Tsukamoto, Naoyoshi Maezaki, Toshimasa Ishida, Tetsuaki Tanaka, Nobuo Okabe, Keiichi Fukuyama, et al. Polyamines stabilize left-handed Z-DNA: Using x-ray

crystallographic analysis, we have found a new type of polyamine (PA) that stabilizes left-handed Z-DNA. *Biochemical and biophysical research communications*, 366(2):275–280, 2008.

[135] Kah Wai Lim and Anh Tuân Phan. Structural basis of DNA quadruplex–duplex junction formation. *Angewandte Chemie International Edition*, 52(33):8566–8569, 2013.

[136] Sangkee Rhee, Zong-jin Han, Keliang Liu, H Todd Miles, and David R Davies. Structure of a triple helical DNA with a triplex- duplex junction. *Biochemistry*, 38(51):16810–16815, 1999.

[137] Arpita Tawani and Amit Kumar. Structural insight into the interaction of flavonoids with human telomeric sequence. *Scientific reports*, 5:17574–17574, 2014.

[138] Carla A Theimer, Craig A Blois, and Juli Feigon. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Molecular cell*, 17(5):671–682, 2005.

[139] Jonathan B Chaires. Drug-DNA interactions. *Current opinion in structural biology*, 8(3):314–320, 1998.

[140] Eric H Lee, Jen Hsin, Marcos Sotomayor, Gemma Comellas, and Klaus Schulten. Discovery through the computational microscope. *Structure*, 17(10):1295–1306, 2009.

[141] Anna K Woźniak, Gunnar F Schröder, Helmut Grubmüller, Claus AM Seidel, and Filipp Oesterhelt. Single-molecule FRET measures bends and kinks in DNA. *Proceedings of the National Academy of Sciences*, 105(47):18337–18342, 2008.

[142] Errol C Friedberg. DNA damage and repair. *Nature*, 421(6921):436–440, 2003.

[143] Sophie Courdavault, Caroline Baudouin, Marie Charveron, Bruno Canguilhem, Alain Favier, Jean Cadet, and Thierry Douki. Repair of the three main types of bipyrimidine DNA photoproducts in human keratinocytes exposed to UVB and UVA radiations. *DNA Repair*, 4(7):836–844, 2005.

[144] Evelyne Sage. Distribution and repair of photolesions in DNA: genetic consequences and the role of sequence context. *Photochemistry and photobiology*, 57(1):163–174, 1993.

[145] Lynn Harrison, Zafer Hatahet, and Susan S Wallace. In vitro repair of synthetic ionizing radiation-induced multiply damaged DNA sites. *Journal of molecular biology*, 290(3):667–684, 1999.

[146] JF Ward. The complexity of DNA damage: relevance to biological consequences. *International journal of radiation biology*, 66(5):427–432, 1994.

[147] JF Ward. Biochemistry of DNA lesions. *Radiation research*, 104(2s):S103–S111, 1985.

[148] Stanislav G Kozmin, Yuliya Sedletska, Anne Reynaud-Angelin, Didier Gasparutto, and Evelyne Sage. The formation of double-strand breaks at multiply damaged sites is driven by the kinetics of excision/incision at base damage in eukaryotic cells. *Nucleic acids research*, 37(6):1767–1777, 2009.

[149] Lynn Harrison, Zafer Hatahet, Andrei A Purmal, and Susan S Wallace. Multiply damaged sites in DNA: interactions with escherichia coli endonucleases III and VIII. *Nucleic acids research*, 26(4):932–941, 1998.

[150] Alexandros G Georgakilas, Peter O'Neill, and Robert D Stewart. Induction and repair of clustered DNA lesions: what do we know so far? *Radiation research*, 180(1):100–109, 2013.

[151] David P Millar. Fluorescence studies of DNA and RNA structure and dynamics. *Current opinion in structural biology*, 6(3):322–326, 1996.

[152] Laurence H Hurley. DNA and its associated processes as targets for cancer therapy. *Nature Reviews Cancer*, 2(3):188–200, 2002.

[153] Brian M Zeglis, Valerie C Pierre, and Jacqueline K Barton. Metallo-intercalators and metallo-insertors. *Chemical Communications*, (44):4565–4579, 2007.

[154] An-Xin Hou, Zhi Xue, Yi Liu, Song-Sheng Qu, and Wai-Kwok Wong. Microcalorimetric and spectroscopic investigation of the antibacterial properties of cationic ytterbium (III)–porphyrin complexes lacking charged peripheral groups. *Chemistry & biodiversity*, 4(12):2889–2899, 2007.

[155] Nial J Wheate, Craig R Brodie, J Grant Collins, Sharon Kemp, and Janice R Aldrich-Wright. DNA intercalators in cancer therapy: organic and inorganic drugs and their spectroscopic tools of analysis. *Mini reviews in medicinal chemistry*, 7(6):627–648, 2007.

[156] Hongqin Guo, Changqun Cai, Hang Gong, and Xiaoming Chen. Multi-spectroscopic method study the interaction of anti-inflammatory drug ketoprofen and calf thymus DNA and its analytical application. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 79(1):92–96, 2011.

[157] Cheng Zhi Huang and Yuan Fang Li. Resonance light scattering technique used for biochemical and pharmaceutical analysis. *Analytica chimica acta*, 500(1):105–117, 2003.

[158] Elise Dumont and Antonio Monari. Benzophenone and DNA: evidence for a double insertion mode and its spectral signature. *The Journal of Physical Chemistry Letters*, 4(23):4119–4124, 2013.

[159] Hugo Gattuso, Elise Dumont, Christophe Chipot, Antonio Monari, and François Dehez. Thermodynamics of DNA: sensitizer recognition. characterizing binding motifs with all-atom simulations. *Physical Chemistry Chemical Physics*, 18(48):33180–33186, 2016.

[160] Erin R Johnson, Shahar Keinan, Paula Mori-Sanchez, Julia Contreras-Garcia, Aron J Cohen, and Weitao Yang. Revealing noncovalent interactions. *Journal of the American Chemical Society*, 132(18):6498–6506, 2010.

[161] T Maehigashi, O Persil, NV Hud, and LD Williams. Crystal structure of proflavine in complex with a DNA hexamer duplex. *Doi: http://dx. doi. org/10*, 2210, 2010.

[162] Amber R Cutter and Jeffrey J Hayes. A brief review of nucleosome structure. *FEBS letters*, 589(20PartA):2914–2922, 2015.

[163] Brian D Thrall, David B Mann, Michael J Smerdon, and David L Springer. Nucleosome structure modulates benzo [a] pyrenediol epoxide adduct formation. *Biochemistry*, 33(8):2210–2216, 1994.

[164] David B Mann, David L Springer, and Michael J Smerdon. DNA damage can alter the stability of nucleosomes: effects are dependent on damage type. *Proceedings of the National Academy of Sciences*, 94(6):2215–2220, 1997.

[165] James M Gale, Karen A Nissen, and Michael J Smerdon. Uv-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases. *Proceedings of the National Academy of Sciences*, 84(19):6644–6648, 1987.

[166] Jonathan T Sczepanski, Remus S Wong, Jeffrey N McKnight, Gregory D Bowman, and Marc M Greenberg. Rapid DNA–protein cross-linking and strand scission by an abasic site in a nucleosome core particle. *Proceedings of the National Academy of Sciences*, 107(52):22475–22480, 2010.

[167] Ryujiro Hara, Jinyao Mo, and Aziz Sancar. DNA damage in the nucleosome core is refractory to repair by human excision nuclease. *Molecular and cellular biology*, 20(24):9173–9181, 2000.

[168] Alexandros G Georgakilas, Paula V Bennett, David M Wilson, and Betsy M Sutherland. Processing of bistranded abasic DNA clusters in γ-irradiated human hematopoietic cells. *Nucleic acids research*, 32(18):5609–5620, 2004.

[169] Alexandros G Georgakilas, Paula V Bennett, and Betsy M Sutherland. High efficiency detection of bi-stranded abasic clusters in γ-irradiated DNA by putrescine. *Nucleic acids research*, 30(13):2800–2808, 2002.

[170] Nathan E Price, Kevin M Johnson, Jin Wang, Mostafa I Fekry, Yinsheng Wang, and Kent S Gates. Interstrand DNA–DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *Journal of the American Chemical Society*, 136(9):3483–3490, 2014.

[171] Martin R Singleton, Mark S Dillingham, and Dale B Wigley. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, 76:23–50, 2007.

[172] Pavan Umate, Narendra Tuteja, and Renu Tuteja. Genome-wide comprehensive analysis of human helicases. *Communicative & integrative biology*, 4(1):118–137, 2011.

[173] Robert M Brosh and Vilhelm A Bohr. Human premature aging, DNA repair and RecQ helicases. *Nucleic acids research*, 35(22):7527–7544, 2007.

[174] Katrin Paeschke, John A Capra, and Virginia A Zakian. DNA replication through G-quadruplex motifs is promoted by the saccharomyces cerevisiae pif1 DNA helicase. *Cell*, 145(5):678–691, 2011.

[175] Isaac Wong and Timothy M Lohman. Allosteric effects of nucleotide cofactors on escherichia coli rep helicase–DNA binding. *Science*, 256(5055):350, 1992.

[176] Ulrich Zander, Guillaume Hoffmann, Irina Cornaciu, J-P Marquette, Gergely Papp, Christophe Landret, Gaël Seroul, Jérémy Sinoir, Martin Röwer, Frank Felisaz, et al. Automated harvesting and processing of protein crystals through laser photoablation. *Acta Crystallographica Section D: Structural Biology*, 72(4):454–466, 2016.

[177] Tracey E Barrett, Renos Savva, George Panayotou, Tom Barlow, Tom Brown, Josef Jiricny, and Laurence H Pearl. Crystal structure of a G: T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell*, 92(1):117–129, 1998.

[178] Clifford D Mol, Andrew S Arvai, Geir Slupphaug, Bodil Kavli, Ingrun Alseth, Hans E Krokan, and John A Tainer. Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*, 80(6):869–878, 1995.

[179] Steven D Bruner, Derek PG Norman, and Gregory L Verdine. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, 403(6772):859–866, 2000.

[180] Albert Y Lau, Orlando D Schärer, Leona Samson, Gregory L Verdine, and Tom Ellenberger. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell*, 95(2):249–258, 1998.

[181] Renos Savva, Katherine McAuley-Hecht, Tom Brown, and Laurence Pearl. The structural basis of specific base-excision repair by uracil–DNA glycosylase. *Nature*, 373:487 – 493, 1995.

[182] Tomas Lindahl. New class of enzymes acting on damaged DNA. *Nature*, 259:64–66, 1976.

[183] Serge Boiteux, Franck Coste, and Bertrand Castaing. Repair of 8-oxo-7, 8-dihydroguanine in prokaryotic and eukaryotic cells: Properties and biological roles of the fpg and ogg1 DNA n-glycosylases. *Free Radical Biology and Medicine*, 2016.

[184] Giuseppe La Rosa and Martin Zacharias. Global deformation facilitates flipping of damaged 8-oxo-guanine and guanine in DNA. *Nucleic Acids Research*, 44(20):9591–9599, 2016.

[185] Tomas Lindahl, Siv Ljungquist, Wolfgang Siegert, Barbro Nyberg, and BDNA Sperens. DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *Journal of Biological Chemistry*, 252(10):3286–3294, 1977.

[186] Andrea J. Lee and Susan S. Wallace. Hide and seek: How do DNA glycosylases locate oxidatively damaged DNA bases amidst a sea of undamaged bases? *Free Radical Biology and Medicine*, pages –, 2016.

[187] K Tanabe, EW Bohn, and SH Wilson. Steady-state kinetics of mouse DNA polymerase. beta. *Biochemistry*, 18(15):3401–3406, 1979.

[188] John M Pascal, Patrick J O'Brien, Alan E Tomkinson, and Tom Ellenberger. Human DNA ligase i completely encircles and partially unwinds nicked DNA. *Nature*, 432(7016):473–478, 2004.

[189] Hosahalli S Subramanya, Aidan J Doherty, Stephen R Ashford, and Dale B Wigley. Crystal structure of an atp-dependent DNA ligase from bacteriophage T7. *Cell*, 85(4):607–615, 1996.

[190] Martin R Singleton, Kjell Håkansson, David J Timson, and Dale B Wigley. Structure of the adenylation domain of an NAD+-dependent DNA ligase. *Structure*, 7(1):35–42, 1999.

[191] Mark Odell, Verl Sriskanda, Stewart Shuman, and Dimitar B Nikolov. Crystal structure of eukaryotic DNA ligase–adenylate illuminates the mechanism of nick sensing and strand joining. *Molecular cell*, 6(5):1183–1193, 2000.

[192] Thomas Carell. DNA repair. *Angewandte Chemie International Edition*, 54(51):15330–15333, 2015.

[193] James T Stivers and Yu Lin Jiang. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chemical reviews*, 103(7):2729–2760, 2003.

[194] J Christopher Fromme and Gregory L Verdine. Structure of a trapped endonuclease III–DNA covalent intermediate. *The EMBO journal*, 22(13):3461–3471, 2003.

[195] Amanda K McCullough, ML Dodson, and R Stephen Lloyd. Initiation of base excision repair: glycosylase mechanisms and structures. *Annual review of biochemistry*, 68(1):255–285, 1999.

[196] Keith W Caldecott. DNA single-strand break repair. *Exp Cell Res*, 329(1):2–8, 2014.

[197] KW Caldecott. Protein ADP-ribosylation and the cellular response to DNA strand breaks. *DNA repair*, 19:108–113, 2014.

[198] Jurgen A Marteijn, Hannes Lans, Wim Vermeulen, and Jan HJ Hoeijmakers. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature reviews Molecular cell biology*, 15(7):465–481, 2014.

[199] Jung-Hyun Min and Nikola P Pavletich. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, 449(7162):570–575, 2007.

[200] Kaoru Sugasawa, Tomoko Okamoto, Yuichiro Shimizu, Chikahide Masutani, Shigenori Iwai, and Fumio Hanaoka. A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes & development*, 15(5):507–521, 2001.

[201] Joyce T Reardon and Aziz Sancar. Recognition and repair of the cyclobutane thymine dimer, a major cause of skin cancers, by the human excision nuclease. *Genes & development*, 17(20):2539–2551, 2003.

[202] Andrea Scrima, Renata Koníčková, Bryan K Czyzewski, Yusuke Kawasaki, Philip D Jeffrey, Regina Groisman, Yoshihiro Nakatani, Shigenori Iwai, Nikola P Pavletich, and Nicolas H Thomä. Structural basis of UV DNA-damage recognition by the DDB1–DDB2 complex. *Cell*, 135(7):1213–1223, 2008.

[203] Mitsuo Wakasugi, Aki Kawashima, Hiroshi Morioka, Stuart Linn, Aziz Sancar, Toshio Mori, Osamu Nikaido, and Tsukasa Matsunaga. DDB accumulates at DNA damage sites immediately after uv irradiation and directly stimulates nucleotide excision repair. *Journal of Biological Chemistry*, 277(3):1637–1640, 2002.

[204] Marcel Volker, Martijn J Moné, Parimal Karmakar, Anneke van Hoffen, Wouter Schul, Wim Vermeulen, Jan HJ Hoeijmakers, Roel van Driel, Albert A van Zeeland, and Leon HF Mullenders. Sequential assembly of the nucleotide excision repair factors in vivo. *Molecular cell*, 8(1):213–224, 2001.

[205] Li Fan, Jill O Fuss, Quen J Cheng, Andrew S Arvai, Michal Hammel, Victoria A Roberts, Priscilla K Cooper, and John A Tainer. XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell*, 133(5):789–800, 2008.

[206] Nadine Mathieu, Nina Kaczmarek, Peter Rüthemann, Andreas Luch, and Hanspeter Naegeli. DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. *Current Biology*, 23(3):204–212, 2013.

[207] Adebanke F Fagbemi, Barbara Orelli, and Orlando D Schärer. Regulation of endonuclease activity in human nucleotide excision repair. *DNA repair*, 10(7):722–729, 2011.

[208] Orlando D Schärer. Nucleotide excision repair in eukaryotes. *Cold Spring Harbor perspectives in biology*, 5(10):a012609, 2013.

[209] Barbara Orelli, T Brooke McClendon, Oleg V Tsodikov, Tom Ellenberger, Laura J Niedernhofer, and Orlando D Schärer. The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways. *Journal of Biological Chemistry*, 285(6):3705–3712, 2010.

[210] René M Overmeer, Jill Moser, Marcel Volker, Hanneke Kool, Alan E Tomkinson, Albert A van Zeeland, Leon HF Mullenders, and Maria Fousteri. Replication protein a safeguards genome integrity by controlling NER incision events. *The Journal of cell biology*, 192(3):401–415, 2011.

[211] Maria Fousteri, Wim Vermeulen, Albert A van Zeeland, and Leon HF Mullenders. Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Molecular cell*, 23(4):471–482, 2006.

[212] Stefan Sigurdsson, A Barbara Dirac-Svejstrup, and Jesper Q Svejstrup. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Molecular cell*, 38(2):202–210, 2010.

[213] Elisabetta Citterio, Vincent Van Den Boom, Gavin Schnitzler, Roland Kanaar, Edgar Bonte, Robert E Kingston, Jan HJ Hoeijmakers, and Wim Vermeulen. ATP-dependent chromatin remodeling by the cockayne syndrome B DNA repair-transcription-coupling factor. *Molecular and cellular biology*, 20(20):7643–7653, 2000.

[214] Nancy Beerens, Jan HJ Hoeijmakers, Roland Kanaar, Wim Vermeulen, and Claire Wyman. The CSB protein actively wraps DNA. *Journal of Biological Chemistry*, 280(6):4722–4729, 2005.

[215] Wolf-Dietrich Heyer, Kirk T. Ehmsen, and Jie Liu. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet.*, 44:113–139, 2010.

[216] William S Dynan and Sunghan Yoo. Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic acids research*, 26(7):1551–1559, 1998.

[217] Michael R. Lieber. The mechanism of human nonhomologous DNA end joining. *Journal of Biological Chemistry*, 283(1):1–5, 2008.

[218] Mariana Yaneva, Tomasz Kowalewski, and Michael R. Lieber. Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. *The EMBO Journal*, 16(16):5098–5112, 1997.

[219] Tomas Aparicio, Richard Baer, and Jean Gautier. DNA double-strand break repair pathway choice and cancer. *DNA repair*, 19:169–175, 2014.

[220] Susovan Mohapatra, Steven M Yannone, Suk-Hee Lee, Robert A Hromas, Konstantin Akopiants, Vijay Menon, Dale A Ramsden, and Lawrence F Povirk. Trimming of damaged 3' overhangs of DNA double-strand breaks by the metnase and artemis endonucleases. *DNA repair*, 12(6):422–432, 2013.

[221] Lorraine S Symington and Jean Gautier. Double-strand break end resection and repair pathway choice. *Annual review of genetics*, 45:247–271, 2011.

[222] Upasana Roy and Orlando D. Schärer. Involvement of translesion synthesis {DNA} polymerases in {DNA} interstrand crosslink repair. *{DNA} Repair*, 44:33 – 41, 2016. Cutting-edge Perspectives in Genomic Maintenance IIICutting-edge Perspectives in Genomic Maintenance {III}.

[223] Laura J Niedernhofer. The fanconi anemia signalosome anchor. *Molecular cell*, 25(4):487–490, 2007.

[224] Andrew J Deans and Stephen C West. DNA interstrand crosslink repair and cancer. *Nature Reviews Cancer*, 11(7):467–480, 2011.

[225] Flora S Groothuizen and Titia K Sixma. The conserved molecular machinery in DNA mismatch repair enzyme structures. *DNA repair*, 38:14–23, 2016.

[226] Emily Jacobs-Palmer and Manju M. Hingorani. The effects of nucleotides on MutS-DNA binding kinetics clarify the role of MutS ATPase activity in mismatch repair. *Journal of Molecular Biology*, 366(4):1087 – 1098, 2007.

[227] Dwayne J. Allen, Alexander Makhov, Michelle Grilley, John Taylor, Randy Thresher, Paul Modrich, and Jack D. Griffith. MutS mediates heteroduplex loop formation by a translocation mechanism. *The EMBO Journal*, 16(14):4467–4476, 1997.

[228] Cherlhyun Jeong, Won-Ki Cho, Kyung-Mi Song, Christopher Cook, Tae-Young Yoon, Changill Ban, Richard Fishel, and Jong-Bong Lee. MutS switches between two fundamentally distinct clamps during mismatch repair. *Nature structural and molecular biology*, 18(3):379–385, 2011.

[229] Karin Drotschmann, Hans-Joachim Fritz, Alexander Aronshtam, and M. G. Marinus. The Escherichia coli MutL protein stimulates binding of Vsr and MutS to heteroduplex DNA. *Nucleic Acids Research*, 26(4):948–953, 1998.

[230] M Grilley, K M Welsh, S S Su, and P Modrich. Isolation and characterization of the Escherichia coli MutL gene product. *Journal of Biological Chemistry*, 264(2):1000–1004, 1989.

[231] Samir Acharya, Patricia L. Foster, Peter Brooks, and Richard Fishel. The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. *Molecular Cell*, 12(1):233–246, 2003.

[232] K G Au, K Welsh, and P Modrich. Initiation of methyl-directed mismatch repair. *Journal of Biological Chemistry*, 267(17):12142–12148, 1992.

[233] Miyuki Yamaguchi, Vivian Dao, and Paul Modrich. MutS and MutL activate DNA helicase II in a mismatch-dependent manner. *Journal of Biological Chemistry*, 273(15):9197–9201, 1998.

[234] Nicole Hustedt and Daniel Durocher. The control of DNA repair by the cell cycle. *Nature Cell Biology*, 19(1):1–9, 2017.

[235] Yael Aylon, Batia Liefshitz, and Martin Kupiec. The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *The EMBO journal*, 23(24):4868–4875, 2004.

[236] Tomoo Ogi, Siripan Limsirichaikul, René M Overmeer, Marcel Volker, Katsuya Takenaka, Ross Cloney, Yuka Nakazawa, Atsuko Niimi, Yoshio Miki, Nicolaas G Jaspers, et al. Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Molecular cell*, 37(5):714–727, 2010.

[237] Jill Moser, Hanneke Kool, Ioannis Giakzidis, Keith Caldecott, Leon HF Mullenders, and Maria I Fousteri. Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase IIIα in a cell-cycle-specific manner. *Molecular cell*, 27(2):311–323, 2007.

[238] Sheila S David, Valerie L O'shea, and Sucharita Kundu. Base-excision repair of oxidative DNA damage. *Nature*, 447(7147):941–950, 2007.

[239] P Fortini, B Pascucci, E Parlanti, M D'errico, V Simonelli, and E Dogliotti. 8-oxoguanine DNA damage: at the crossroad of alternative repair pathways. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 531(1):127–139, 2003.

[240] Yan Qi, Marie C Spong, Kwangho Nam, Martin Karplus, and Gregory L Verdine. Entrapment and structure of an extrahelical guanine attempting to enter the active site of a bacterial DNA glycosylase, MutM. *Journal of Biological Chemistry*, 285(2):1468–1478, 2010.

[241] Mark Leo Michaels, Lucie Pham, Christina Cruz, and Jeffrey H Miller. MutM, a protein that prevents GC→ TA transversions, is formamidopyrimidine-DNA glycosylase. *Nucleic Acids Research*, 19(13):3629–3632, 1991.

[242] J Christopher Fromme and Gregory L Verdine. DNA lesion recognition by the bacterial repair enzyme MutM. *Journal of Biological Chemistry*, 278(51):51543–51548, 2003.

[243] J Christopher Fromme and Gregory L Verdine. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature Structural & Molecular Biology*, 9(7):544–552, 2002.

[244] Yan Qi, Marie C Spong, Kwangho Nam, Anirban Banerjee, Sao Jiralerspong, Martin Karplus, and Gregory L Verdine. Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature*, 462(7274):762–766, 2009.

[245] François Bergeron, Frédéric Auvré, J Pablo Radicella, and Jean-Luc Ravanat. HO • radicals induce an unexpected high proportion of tandem base lesions refractory to repair by DNA glycosylases. *Proceedings of the National Academy of Sciences*, 107(12):5528–5533, 2010.

[246] Alberto Pérez, F Javier Luque, and Modesto Orozco. Frontiers in molecular dynamics simulations of DNA. *Accounts of Chemical Research*, 45(2):196–205, 2011.

[247] Hugo Gattuso, Elodie Durand, Emmanuelle Bignon, Christophe Morell, Alexandros G Georgakilas, Elise Dumont, Christophe Chipot, François Dehez, and Antonio Monari. Repair rate of clustered abasic DNA lesions by human endonuclease: Molecular bases of sequence specificity. *The Journal of Physical Chemistry Letters*, 7(19):3760–3765, 2016.

[248] M Abul Kalam and Ashis K Basu. Mutagenesis of 8-oxoguanine adjacent to an abasic site in simian kidney cells: Tandem mutations and enhancement of G→ T transversions. *Chemical research in toxicology*, 18(8):1187–1192, 2005.

[249] Siobhan MT Cunniffe, Martine E Lomax, and Peter O'Neill. An AP site can protect against the mutagenic potential of 8-oxog when present within a tandem clustered site in E. coli. *DNA repair*, 6(12):1839–1849, 2007.

[250] Savas T Tsikis. Mutagenesis of 8-oxoguanine adjacent to an abasic site in escherichia coli cells proficient or deficient in DNA polymerase IV. *Honors Scholar Theses*, 344, 2014.

[251] Haoquan Li, Anton V Endutkin, Christina Bergonzo, Lin Fu, Arthur Grollman, Dmitry O Zharkov, and Carlos Simmerling. DNA deformation-coupled recognition of 8-oxoguanine: conformational kinetic gating in human DNA glycosylase. *Journal of the American Chemical Society*, 139(7):2682–2692, 2017.

[252] Giuseppe La Rosa and Martin Zacharias. Global deformation facilitates flipping of damaged 8-oxo-guanine and guanine in DNA. *Nucleic Acids Research*, 44(20):9591–9599, 2016.

Appendix A

Supporting Informations : DNA photosensitization by an "insider": photophysics and triplet energy transfer of 5-methyl-2-pyrimidone deoxyribonucleoside.

CHEMISTRY
A European Journal

Supporting Information

DNA Photosensitization by an "Insider": Photophysics and Triplet Energy Transfer of 5-Methyl-2-pyrimidone Deoxyribonucleoside

Emmanuelle Bignon,[a, b] Hugo Gattuso,[c, d] Christophe Morell,[b] Elise Dumont,*[a] and Antonio Monari*[c, d]

chem_201501212_sm_miscellaneous_information.pdf

Force Field Parameters for the Pyo unit

The standard Antechamber procedure recommended together with AMBER MD simulations was applied. Atom point charges modifications for the Pyo nucleobase, based on the RESP charges calculated with Gaussian09 (LC-BLYP/6-31G**). The bond force field parameters were taken from the Generalized Amber Force Field (GAFF) parameter and parm99 parameter set.

Atom number	Atom name	Atom type	Atom charge
1	N1	N*	-0.2838
2	C6	CM	-0.0325
3	H6	HC	0.1465
4	C5	CM	-0.0871
5	C7	CT	-0.3784
6	H71	HC	0.1023
7	H72	HC	0.1023
8	H73	HC	0.1023
9	C4	CM	0.3350
10	H4	HC	0.0638
11	N3	NC	-0.7279
12	C2	C	0.9809
13	O2	O	-0.6052

Table S1 Pyo nucleobase atoms list, with their number, name, type and modified point charges.

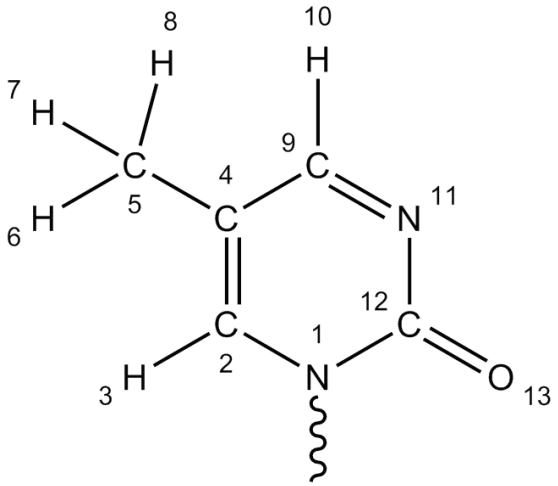


Figure S1 Pyo nucleobase drawing with atom numbers.

q

Time Evolution of the Key Intra-Base Distances P6-A15 and T5-A16

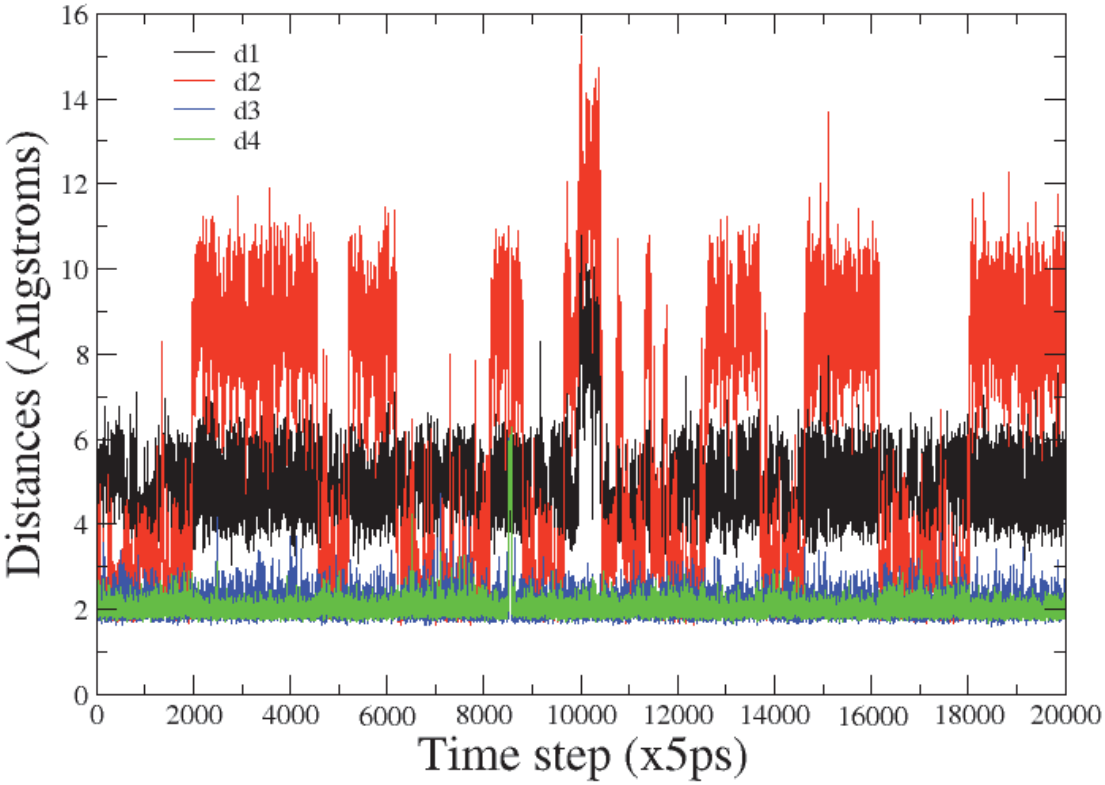


Figure S2 Time evolution of the distances d_1 between the closest nitrogen atoms of P6 and A15, and d_2 corresponding to the distance between O on P6 and amino H on A16. For comparison, the time evolution of the distance between O on T5 and H of A16 (d_3) and between H on T5 and N on A16 (d_4) are reported too.

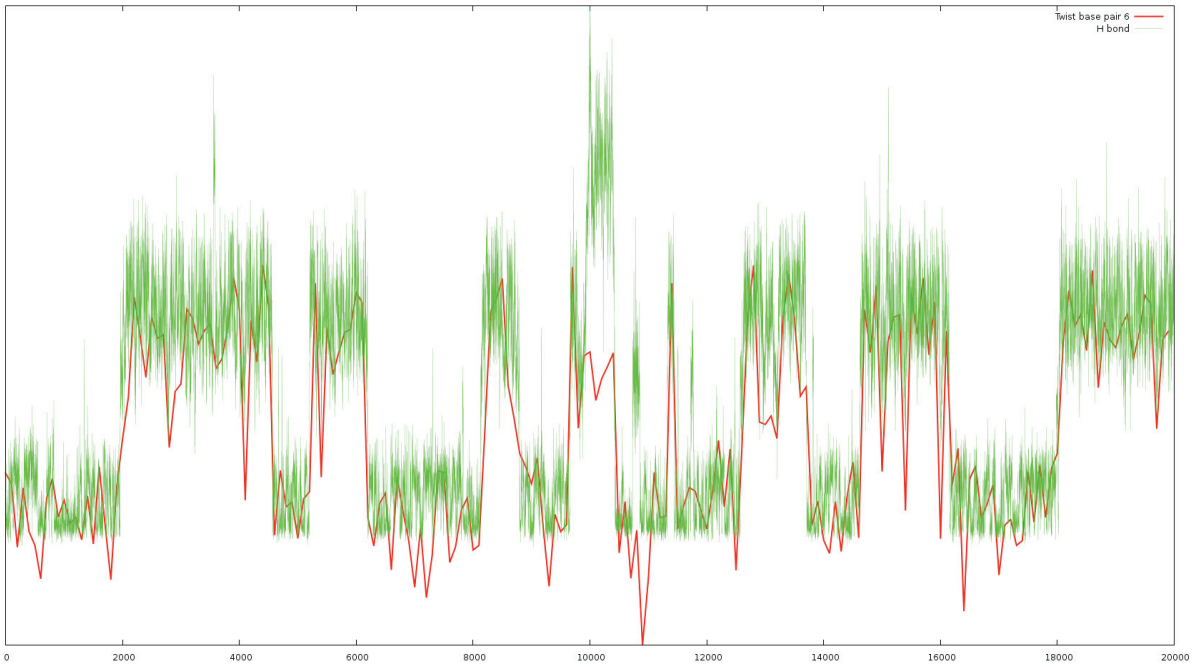


Figure S3 Comparison of the time evolution of d1 and the deviation from ideal value of the twist interbase parameter for base pair couple 6-7.

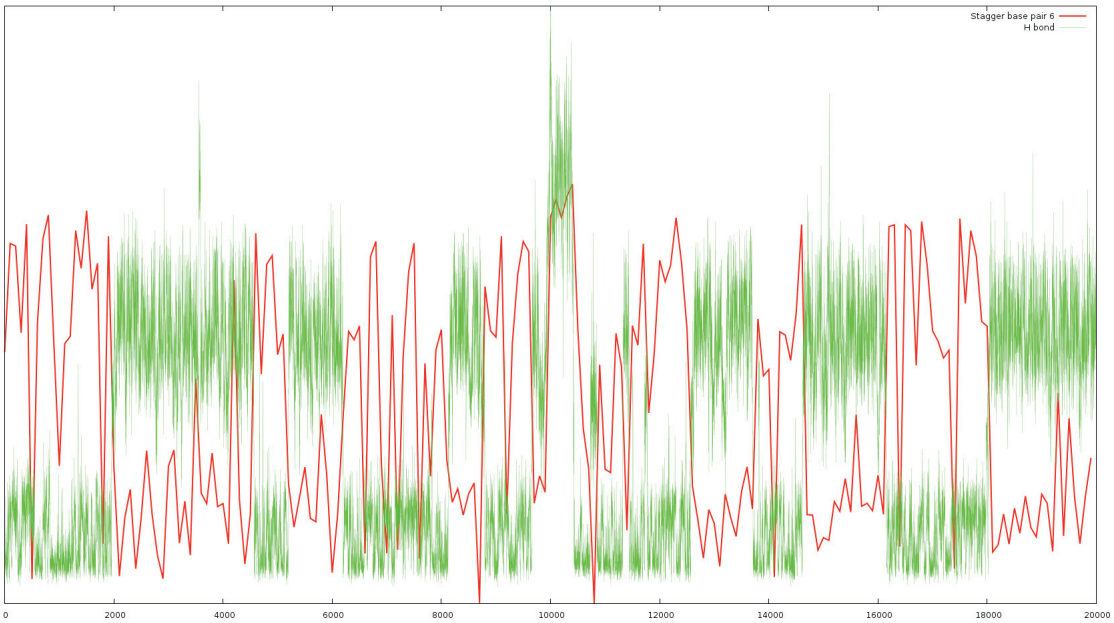


Figure S4 Comparison of the time evolution of d1 and the deviation from ideal value of the stagger intrabase parameter for base pair couple 6.

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.

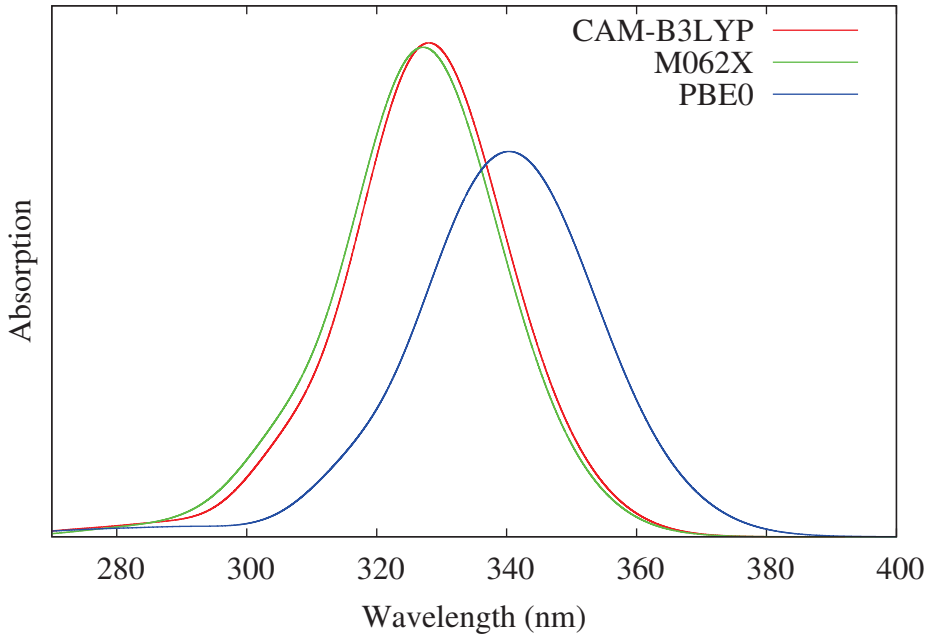


Figure S5 Calculated absorption spectra for the Pyo nucleobase in B-DNA. TD-DFT with different functionals has been used. The protocols to obtain the spectra are the same as the one described in the main text.

Electronic Circular Dichroism 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.

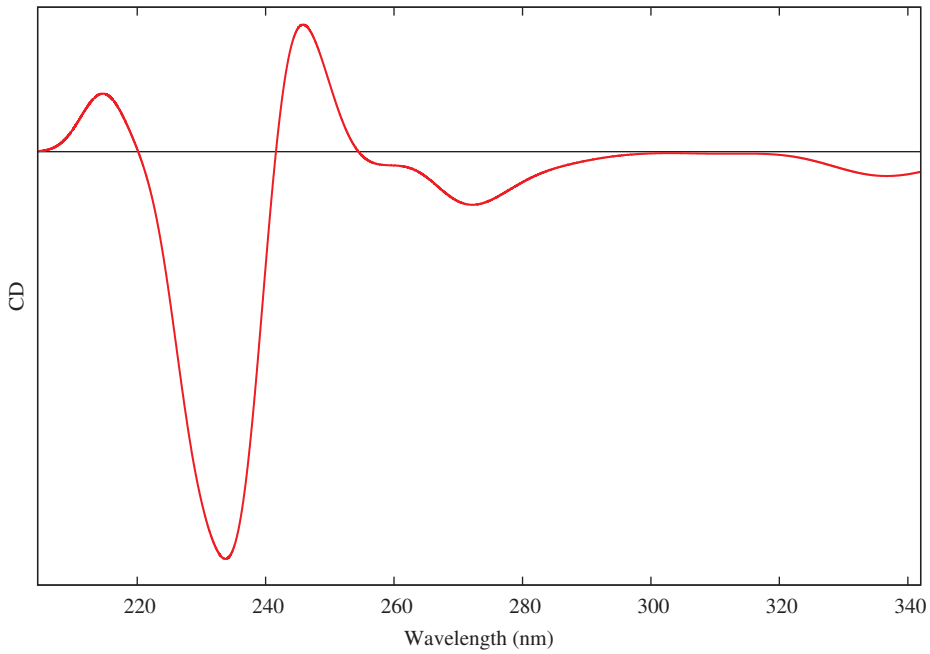


Figure S6. Calculated Electronic Circular Dicroism Spectrum for the Pyo containing DNA double strand.

Electronic Response of the Environment Methodology.

In order to take into account polarizable embedding we propose an alternative solution developed in our laboratory and called Electronic Response of the Environment (ERS) technique. For the reader convenience we recall here the bases of the technique: to simulate the environment polarization induced by the change in the chromophore electronic density determined by the excitation the QM partition is embedded in a polarizable continuum characterized by the value of its dielectric constant extrapolated to infinite frequencies (ε_∞). The latter quantity indeed takes into account the fast electron cloud reorganization (polarization). Note that the continuum interacts with the QM wavefunction but not with the MM point charges to avoid discontinuities. Furthermore ε_∞ for insulator has an almost universal value comprised between 1.5 and 2.0 a.u., hence the value of 1.776 (water) has been used. This technique has already been used in a number of application concerning biological and chemical relevant materials and has been shown to catch most of the polarizable embedding phenomena.

Appendix B

Supporting Informations : Singlet oxygen attack on guanine: reactivity and structural signature within the B-DNA helix.

CHEMISTRY
A European Journal

Supporting Information

Singlet Oxygen Attack on Guanine: Reactivity and Structural Signature within the B-DNA Helix

Elise Dumont,*[a] Raymond Grüber,[a] Emmanuelle Bignon,[a, b] Christophe Morell,[b] Juan Aranda,[c] Jean-Luc Ravanat,[d] and Iñaki Tuñón[c]

chem_201601287_sm_miscellaneous_information.pdf

Singlet oxygen attack onto guanine: reactivity and structural signature within the helix

Supporting Information

Elise Dumont[1,*], Raymond Grüber[1], Emmanuelle Bignon[2], Christophe Morell[2], Juan Aranda[3,†], Jean-Luc Ravanat[4,5], Iñaki Tuñón[3]

[1] *Ecole Normale Supérieure de Lyon, CNRS UMR 5182, Université Lyon 1, Laboratoire de Chimie, 46 allée d'Italie, F-69364, Lyon, France*

[2]*Université de Lyon, Institut des Sciences Analytiques, UMR 5280 CNRS, Université Lyon 1, ENS Lyon, 5 rue de la Doua, F-69100 Villeurbanne, France*

[3]*Departament Química Física Universidad de València 46100 Burjassot, Spain*

[4]*Université Grenoble Alpes, INAC-SCIB, F-38000 Grenoble, France*

[5]*CEA, INAC-SCIB-LAN, F-38000 Grenoble, France*

([†]) Present adress: Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona, Spain

Reference static profile

The static profile was investigated at the LC-BLYP/6-31+G* level of theory to provide an intrinsic reference for the G + $^1O_2$ reaction profile, i.e. using the same level of theory than the one used for the QM subsystem in our QM/MM-MD simulations. This level of theory was calibrated, both for the density functional and the basis set, in an earlier study (Reference 9 of the main text). Since **2** is an

intramolecular charge-separated entity and to better model experimental conditions, an implicit model accounting for solvation in water, polarizable continuum model (PCM), was used with a value of the dielectric constant ε_r=78.3. We also report the values in gas phase (reoptimized) in the second block for the sake of comparison. We estimate the free energies by performing frequencies calculations, using the rigid rotor model: the difference with the electronic energies is found to be limited, notably for the transition state corresponding to the attack of $^1O_2$ onto the C8 position of guanine. We stress out that the triplet-singlet gap is estimated to 18.1 and 22.2 kcal/mol respectively for 1 and TS1-2. Hence the need for a multireference treatment is alleviated, in contrast to studies of singlet oxygen reactivity on concerted mechanisms with *symmetric* alkenes (ethylene, benzene, butadiene). The oxygen-oxygen distance spans a very limited range of values 1.25~1.27 Å going from 1 to TS1-2, to be compared to the value of 1.19 Å in the triplet state optimized geometry.

LC-BLYP/6-31+G*(IEFPCM)		LC-BLYP/DZP++(IEFPCM) [9]	
1	0.0 (0.0)	1	0.0 (0.0)
TS^{1-2}	20.4 (20.1)	TS^{1-2}	20.1 (19.4)
2 “outer”	-9.9 (-8.0)	2 “outer”	-10.2 (-8.7)
2 “inner”	-12.9 (-10.6)	2 “inner”	-13.0 (-10.9)
3	-20.2 (-15.8)	3	-22.3 (-18.2)
LC-BLYP/6-31+G*(gas)		LC-BLYP/DZP++(gas)	
1	0.0 (0.0)	1	0.0 (0.0)
TS^{1-2}	28.6 (28.5)	TS^{1-2}	28.0 (28.0)
2 “outer”	12.1 (13.6)	2 “outer”	11.0 (12.9)
2 “inner” (*)	n/a	2 “inner” (*)	n/a
3	-22.0 (-17.5)	3	-24.4 (-19.8)

Table S1. Relative electronic energies estimated for the isolated guanine (in kcal/mol). Values in parenthesis correspond to free energy estimates based on frequency calculations (rigid rotor model). (*) The "inner" conformer of **2** gives rise to **3** in gas phase, such that no energy can be proposed.

Classical MD simulations protocol

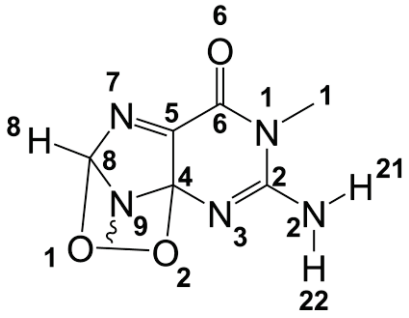
In order to generate the non-canonical nucleobase parameters, the isolated endoperoxide structure was first optimized at the LC-BLYP/6-31+G* with the Gaussian09 series of program. The Restrained Electrostatic Potential charge model was used to calculate the atom-point charges of the nucleobase (see Table S2). Parameters were computed with the parmcheck Ambertools program.[2] potassium counterions were added to neutralize the system, which was placed in an orthorhombic solvent box counting ~8750 TIP3P water molecules measuring 65.4 * 68.4 * 77.3 A^3. The *ff12* force field was employed.

A 5000 steps minimization with 1500 conjugate gradient steps was first performed. Then, the system was heated from 0 to 300K in a 30 ps thermalization run, using the Langevin thermostat with a collision frequency (γln) of 2ps^{-1}. The latter was kept during equilibration and production steps in order to maintain the temperature at 300K. A 300 ps NTP equilibration was performed in order to relax the system before the 100 ns production.

Along the QM/MM-MD profile, a structural analysis was performed using the Curves+ program.[3] The stacking interactions between GO7 and C6 and C8 were visualized using the NCI[4] method.

Table S2. Atom-point charges and type assigned for the oxidized guanine.

Atom name	Atom type	Atom charge
N1	NA	-0.6881
H1	H	0.3783
C2	CA	0.8288
N2	N2	-0.9382
H21	H	0.4272
H22	H	0.4272
N3	NC	-0.6323
C4	CT	0.5247
C5	CB	0.2507
C6	CA	0.6476
O6	O	-0.4864
N7	NB	-0.5963
C8	CT	0.8207
H8	H1	0.0405
O1	OS	-0.2522
O2	OS	-0.2603
N9	N*	-0.5807



Assessment of dispersive interaction between –O-O- and C8

Cartesian coordinates of the GO7 and C8 nucleobases were extracted from the final snapshot and we estimated the dispersive forces using the only gCP-d3 webservice. On the same snapshot, we extracted the cartesian coordinates of the C6, GO7 and C8 on which we performed a NCI calculation at the LC-

BLYP/6-31+G* level of theory. The dispersive interaction between the peroxo bridge and the π-ring of cytosine C8 amounts to 0.7 kcal/mol.

Figure S1. Radial distribution functions for **2**, obtained along a 100ns classical MD trajectory, highlighting a high solvation for the zwitterionic intermediate, which is consequently stabilized in the reaction profile. The profile for **1** cannot be obtained by classical MD simulations and hence is not reported here.

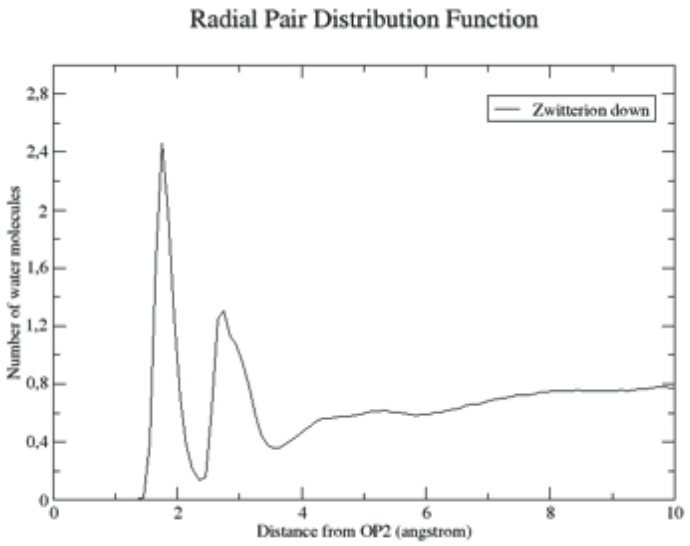


Figure S2. Numbering scheme for the three-HB Watson-Crick association between GO7 and C20 and corresponding time evolutions along the reaction profile. The red line displays the running average.

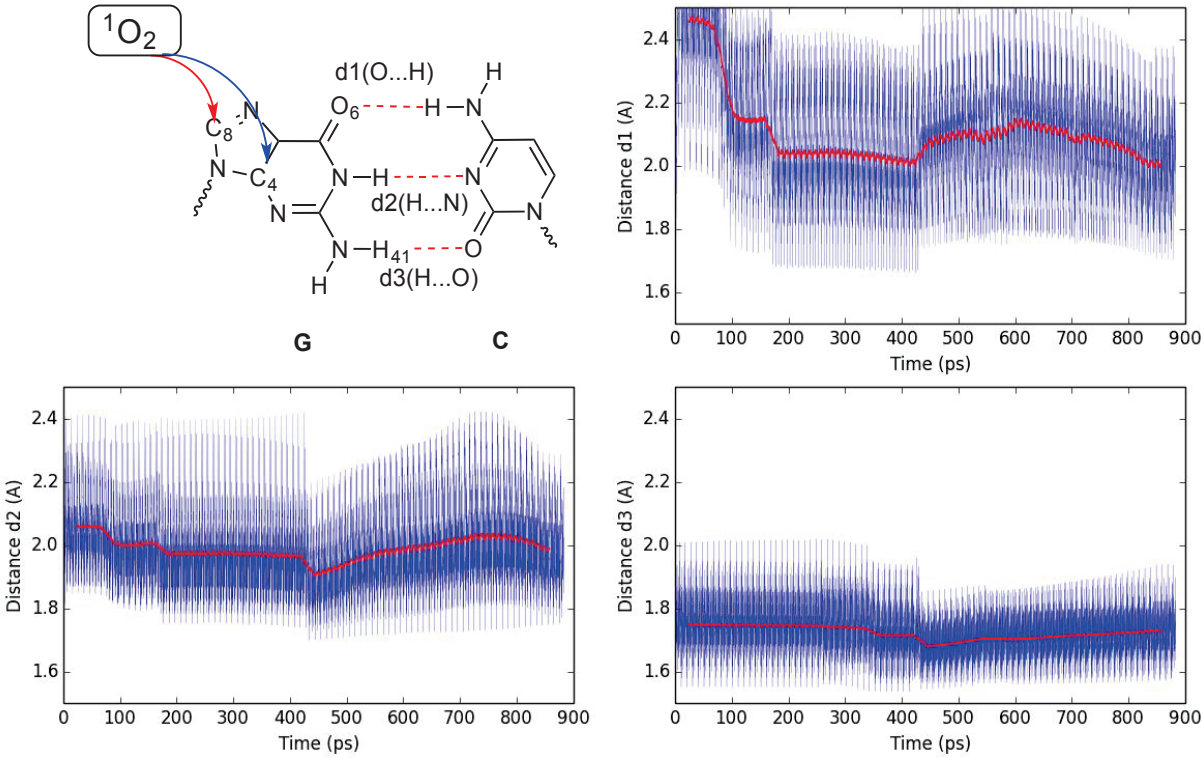


Figure S3. Evolution of distances (in blue, with running averages in red) of –O-O- to the phosphates and the closest K^+ cation along the reaction profile: representative cartoon representations are given in Figure 4. The x axis corresponds to the total time of the QM/MM-MD simulation, going from the endoperoxide **3** to the DNA+$^1O_2$ adduct **1**.

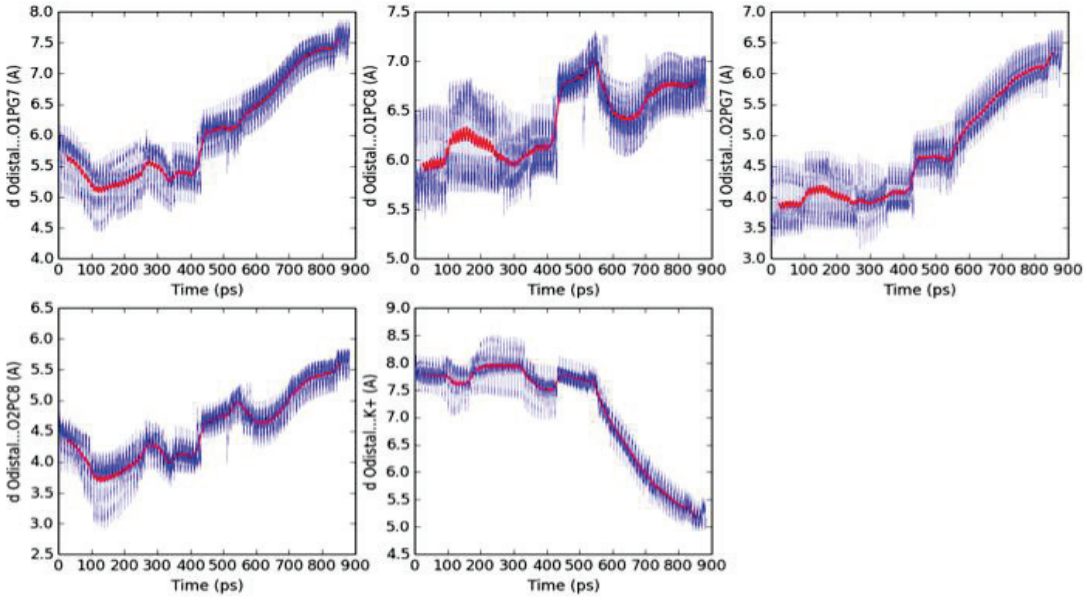


Figure S4. Averaged distances and angles characterizing the B-helical distortion of the 13-bp duplex featuring a central guanine endoperoxide. The more pronounced variations for the terminal base pairs are characteristic of their higher flexibility.

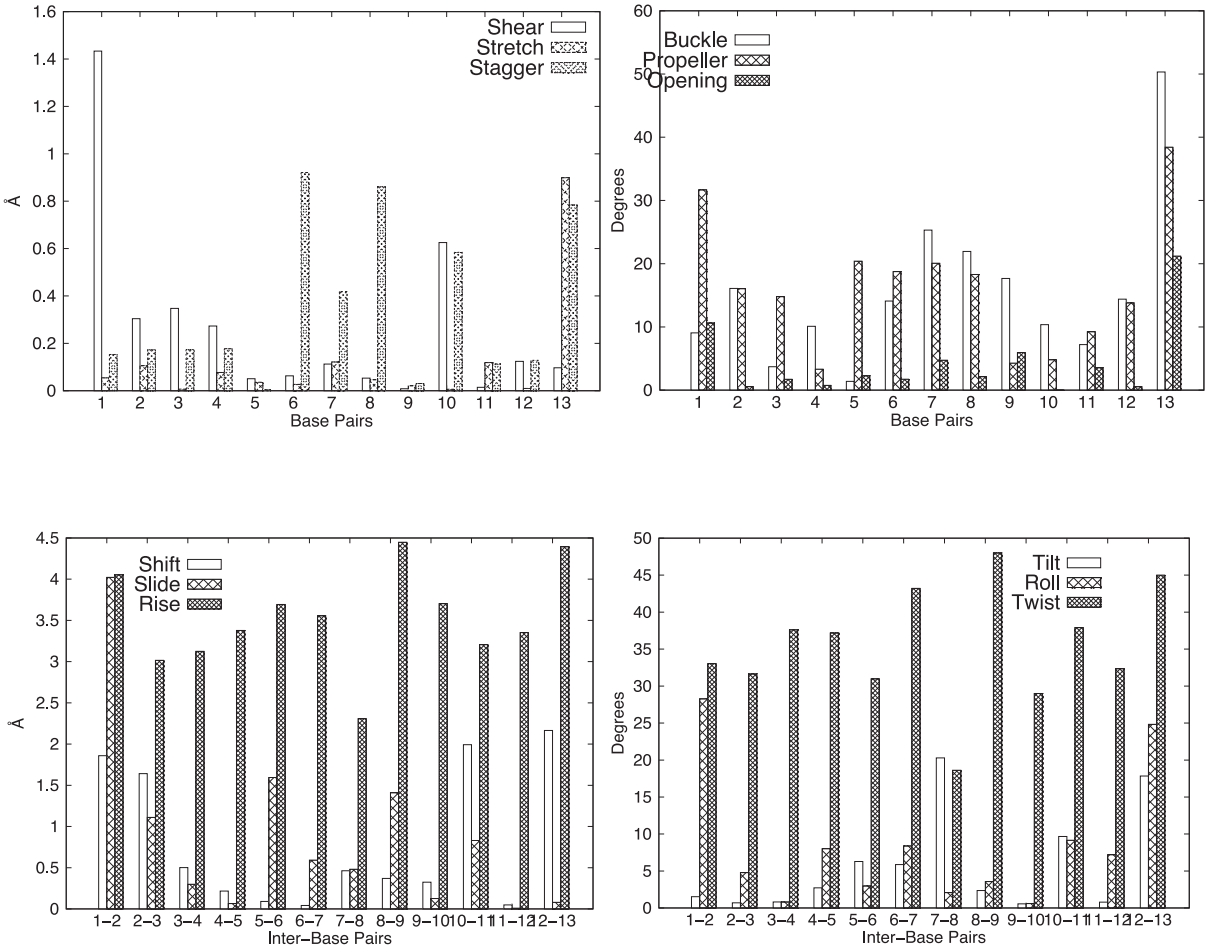
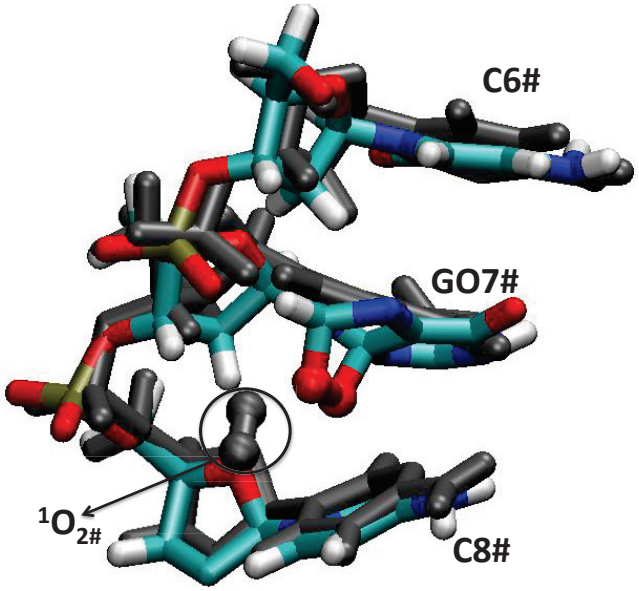


Figure S5. Cartoon representation for **1** (in grey sticks) and **3** (in colors sticks), illustrating the modest structural deviations of C6 and C8 upon formation of the endoperoxide. The singlet-oxygen moiety in **1** and the peroxo moiety in **3** are displayed in balls and sticks.



References

(1) Case, D. A. Amber12 ; University of California, CA, **2012**.

(2) Cieplak, P.; Cornell, W.D.; Bayly, C.; and Kollman, P.A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comp. Chem.* **1995**, *16*, 1357–1377, DOI: 10.1002/jcc.540161106.

(3) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska K. Conformational Analysis of Nucleic Acids Revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917-5929, DOI: 10.1093/nar/gkp608.

(4) Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-Garcia, J.; Cohen, Aron J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498-6506, DOI: 10.1021/ja100936w.

Appendix C

Supporting Informations : Ibuprofen and ketoprofen photosensitization mechanism: how to enhance cells death under the influence of UVA light.

Ibuprofen and ketoprofen photosensitization mechanism: how to enhance cell death under the influence of UVA light

Emmanuelle Bignon,[1,2,†] Marco Marazzi,[3,4,†] Vanessa Besancenot,[5] Hugo Gattuso,[3,4] Christophe Morell,[1] Leif A. Eriksson,[6] Stephanie Grandemange,[5,*] Elise Dumont,[2,*] Antonio Monari[3,4,*]

[1]Institut des Sciences Analytiques, UMR 5280, Université de Lyon1 (UCBL) CNRS, ENS Lyon, Lyon, France

[2]Université de Lyon, ENS de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie, F69342, Lyon, France.

[3]Theory-Modeling-Simulation, Université de Lorraine– Nancy, SRSMC Boulevard des Aiguillettes, Vandoeuvre -lès-Nancy, Nancy, France

[4]Theory-Modeling-Simulation, CNRS, SRSMC Boulevard des Aiguillettes, Vandoeuvre-lès-Nancy, Nancy, France.
[5]Université de Lorraine Nancy and CNRS, CRAN, Vandeouvre-lès-Nancy, France.

[6]Department of Chemistry&MolecularBiology, University of Gothenburg, Medicinaregatan 9 c, 40530 Göteborg, Sweden

Content

1. Computational details

Concerning molecular dynamics, for each system, 19 potassium cations and an orthorhombic TIP3P water box were added, resulting in ~25000 atoms in total. Then, the geometry was minimized using 5000 steps with the steepest descent algorithm followed by 5000 using the conjugate gradient. The temperature was raised from 0 to 300K in a 30 ps heating run. The 300 K temperature value was kept constant during the remaining of the simulation using the Langevin thermostat with a collision frequency γln of 1 ps^{-1}. The system was then relaxed in NTP conditions during a 1 ns equilibration step. Finally, a production stage was performed to sample the conformational space of each system during 300 ns for inserted starting structures and only 100 ns for minor groove binding initial positions since the latter were not stable due to the repulsion between negatively charged DNA phosphates and the drugs' carboxylate.

Concerning quantum chemistry calculations, for both ibuprofen and ketoprofen, Multi State (MS)-CASPT2//State Average (SA)-CASSCF calculations were performed, including the three lowest-energy singlet (S_0, S_1, S_2) and triplet (T_1, T_2, T_3) states. When applying the CASPT2 energy correction on top of the CASSCF calculation, an imaginary shift of 0.2 and no IPEA shift were included.

2. TD-DFT benchmark

A benchmark study to select the appropriate DFT functional was performed on ibuprofen. Especially, the static absorption spectrum – *i.e.* the absorption spectrum calculated for the ground state (S_0) optimized molecule – is shown in Table S1. Solvent (water) effects are taken into account by PCM (see main text).

Table S1. Ibuprofen absorption spectra calculated at TD-DFT level for different functionals. The 6-311+G** basis set is used for all cases. Energy values are given in eV and (nm). In the last row, the CASPT2/ANO-L-VDZP results are shown for comparison.

Functional/method	S_0–T_1	S_0–T_2	S_0–T_3	S_0–S_1
CAM-B3LYP	3.44 (361)	4.39 (282)	4.57 (272)	5.60 (222)
B3LYP	3.64 (340)	4.22 (294)	4.43 (280)	4.96 (250)
BLYP	3.65 (339)	3.68 (337)	3.80 (326)	4.03 (308)
PBE0	3.48 (356)	4.25 (292)	4.45 (279)	5.11 (243)
PBE	3.59 (346)	3.61 (343)	3.71 (334)	3.96 (313)
BP86	3.61 (343)	3.64 (341)	3.75 (330)	4.01 (309)
CASPT2	3.83 (324)	4.28 (290)	4.34 (286)	5.61 (221)

As it can be seen, the relative energy of the bright singlet state that is populated after irradiation (S_0–S_1) is can be highly dependent on the selected functional. As compared to CASPT2/ANO-L-VDZP calculations, we decided to select the CAM-B3LYP functional, being only 0.01 eV blue shifted. For all functionals, three triplet states lie below S_1, as suggested also by the CASPT2 reference calculation.

3. Experimental and theoretical absorption spectra in water

Absorption spectra were calculated for ibuprofen and ketoprofen, also including the dynamical effect: after ground state (S_0) optimization, the phase space around the S_0 minimum structure was sampled through a Wigner distribution.[S1,S2] As input, the vibrational frequencies and the corresponding normal mode vectors were provided by a frequencycalculation at the CAM-B3LYP/6-311+G** level of theory. Further, an excited states calculation of the lowest-energy five singlet states (S_0 to S_4) was performed for twenty random structures obtained from the Wigner distribution. The absorption spectrum has been calculated as the Gaussian convolution (FWHM = 0.4) of the linear transitions for all the structures. Solvent (water) effects are taken into account by PCM (see main text).

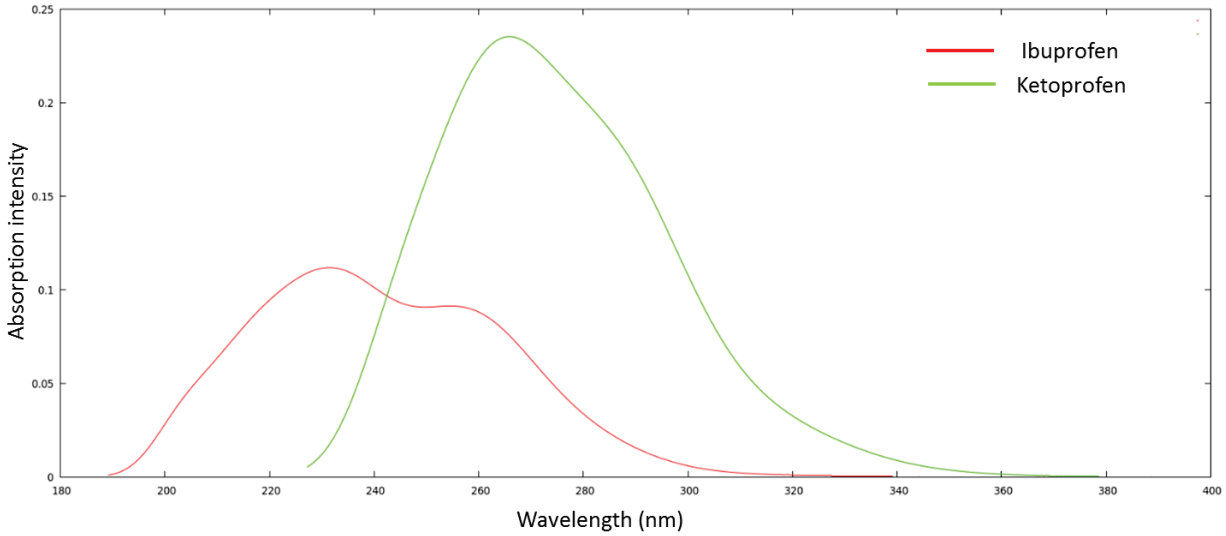


Figure S1. Ibuprofen (red) and ketoprofen (green) calculated absorption spectra in water.

When compared to the available experimental data (ibuprofen: S3, S4; ketoprofen: S5), we can observe how the simulated spectra are in almost quantitative agreement with the experience, also reproducing the overall spectrum shape.

4. Molecular orbitals and minimum energy paths at different levels of theory

In Figure S2 the minimum energy paths of ibuprofen and ketoprofen at TD-DFT and CASSCF levels are shown.

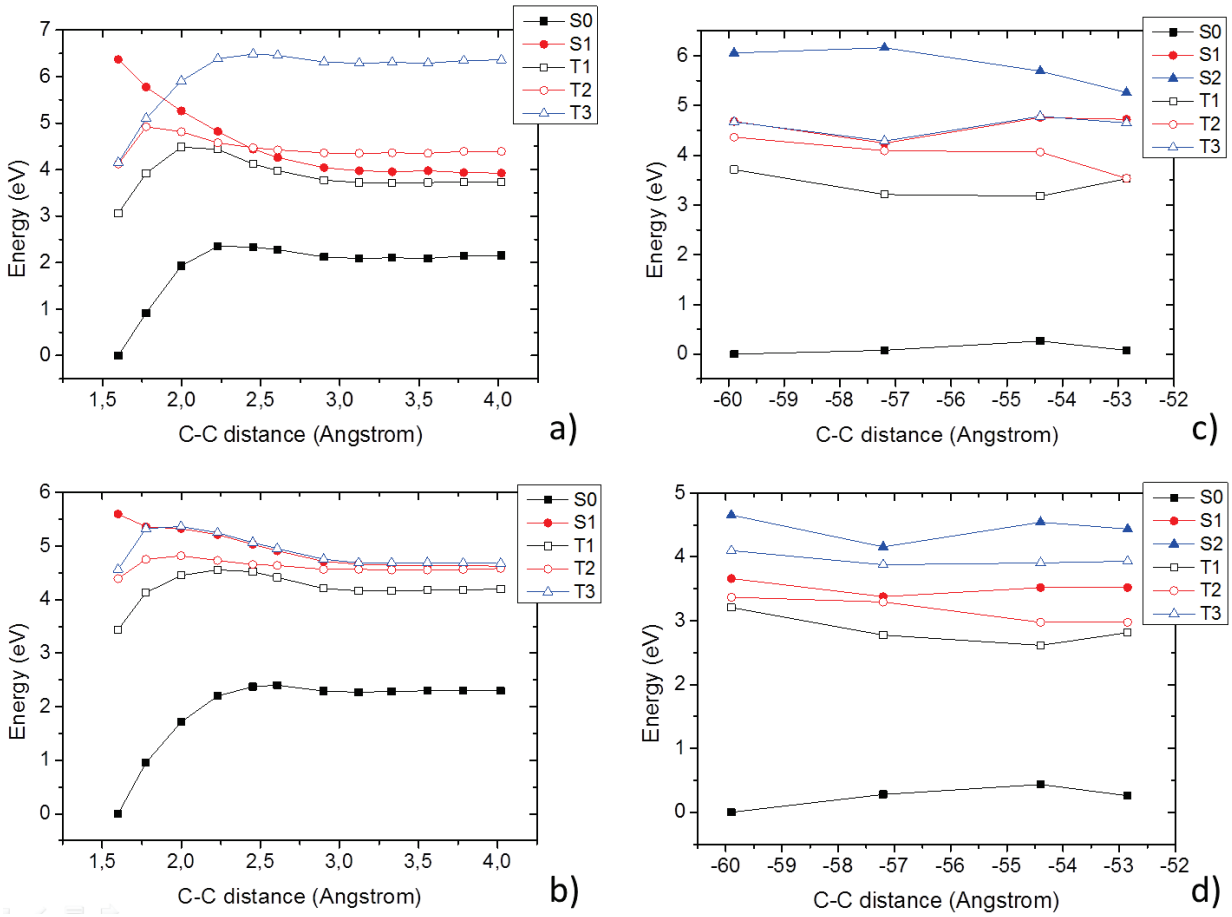


Figure S2. Ibuprofen (a,b) and ketoprofen (c,d) minimum energy paths at CASSCF/ANO-L-VDZP (a,c) and CAM-B3LYP/6-311+G** (b,d) levels of theory

In Figure S3 it is shown the spin-orbit coupling calculated for ibuprofen along the photo-dissociation coordinate, between S_1 and the three lowest-energy triplet states (T_1, T_2 and T_3). As it can be seen, the values are really low in all cases, never reaching 1 cm^{-1}. Therefore, singlet-to-triplet intersystem crossing mechanisms can be discarded. On the other hand, the spin-orbit coupling calculated for ketoprofen (in Franck–Condon and S_1 minimum) confirmed its benzophenon-like behavior, being constantly of *ca.* 20 cm^{-1}.

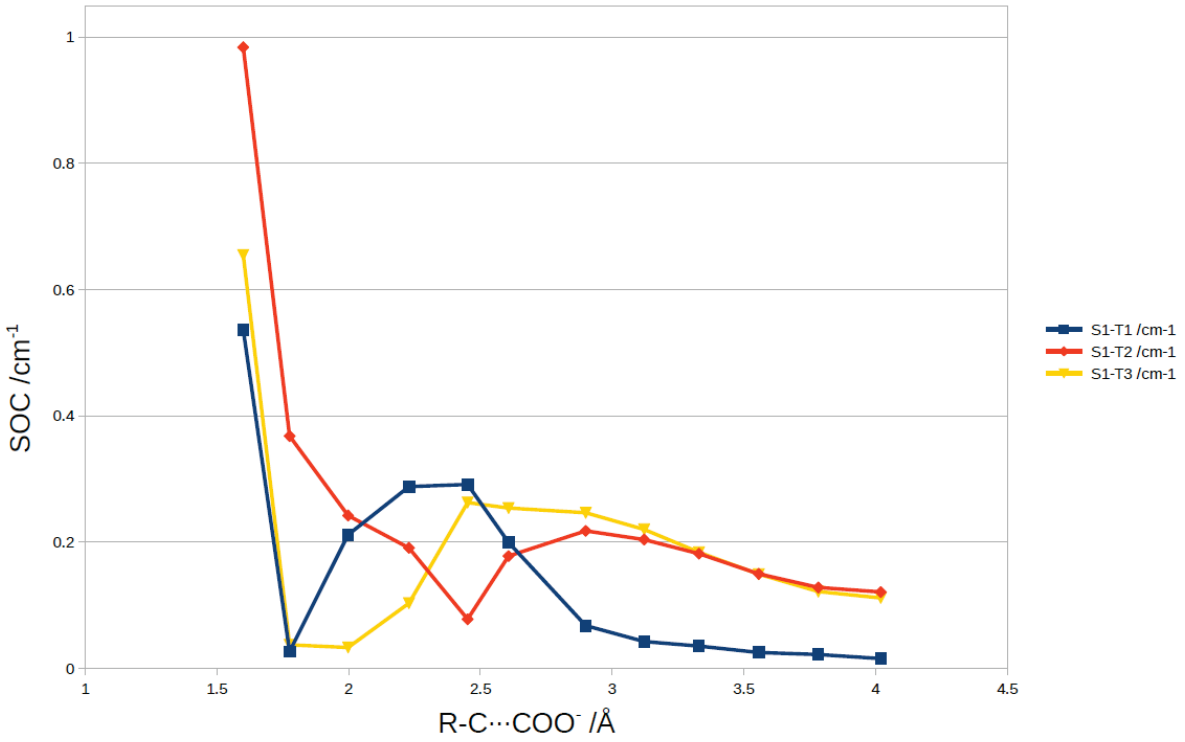


Figure S3. Ibuprofen spin-orbit coupling (SOC) along the mimimum energy path, shown as a function of the photodissociation coordinate.

Appendix D

Supporting Informations : Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distorsion.

Supporting Information for:

Correlation of clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion

Emmanuelle Bignon[1,2], Hugo Gattuso[3,4], Christophe Morell[2], Francois Dehez[4,] Alexandros G. Georgakilas[5,*], Antonio Monari[3,4,*], Elise Dumont[1,*]

[1]Univ Lyon, Ens de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France, [2]Institut des Sciences Analytiques, Université de Lyon 1 and CNRS, Villeurbanne France, [3]Université de Lorraine -Nancy, Theory-Modeling-Simulation SRSMC, Vandoeuvre-lès-Nancy, France, [4]CNRS, Theory-Modeling-Simulation SRSMC, Vandoeuvre-lès-Nancy, [5]Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou 15780, Athens, Greece

* To whom correspondence should be addressed: elise.dumont@ens-lyon.fr
*Correspondence may also be addressed to: antonio.monari@univ-lorraine.fr
*Correspondence may also be addressed to: alexg@mail.ntua.gr

The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors

Tautomerism of the abasic site

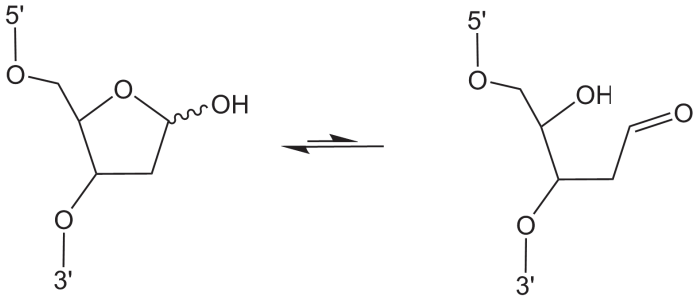


Figure S1 Illustration of the 99:1 equilibrium between the open and closed form

Hydrogen bond between the ejected base and DNA backbone

Molecular dynamics setup

All classical molecular dynamics simulations were performed using the Amber12 suite of programs [1], using ff99bsc0 [2] and gaff [3] force fields. Since no experimental structure is available for sequences used by Georgakilas et al. [4], we built the corresponding 23bp oligonucleotides with the nab module. RESP charges [5] and parameters were generated for the AP site according to the standard Antechamber protocol (see table S1). Then, the system was placed in an octahedral TIP3P [6] water box of 12A buffer, with 44 sodium cations to ensure the medium's neutrality. For each sequence, 400ns classical molecular dynamics were performed. A 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 300K in a 20ps themalization run. Langevin thermostat with a 1.0 ps^{-1} collision frequency γln was used to keep it constant during the remaining of the simulation. Then, a first 100ps equilibration run in NPT was performed, followed by a second one in NVT conditions. Finally, a 400ns production was done in the NPT thermodynamic ensemble.

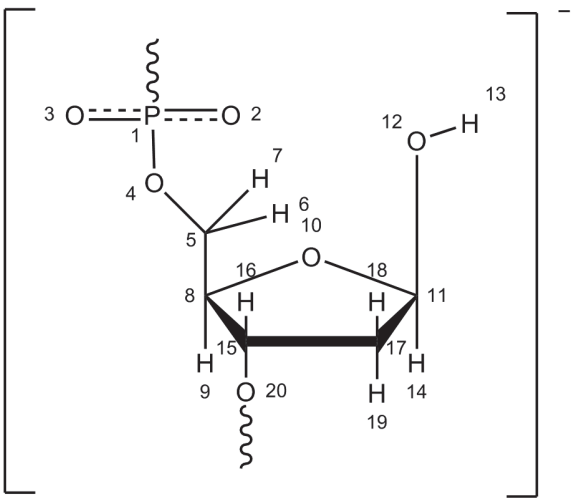


Figure S2: Numbering convention for the constituting atoms of the AP site

	Name	Type	Charge
1	P	P	1.1659
2	OP2	O2	-0.7761
3	OP1	O2	-0.7761
4	O5’	OS	-0.4954
5	C5’	CI	-0.0069
6	H5’	H1	0.0754
7	H5”	H1	0.0754
8	C4’	CT	0.2098
9	H4’	H1	0.0046
10	O4’	OS	-0.4058
11	C1’	CT	0.3679
12	O1	OH	-0.5959
13	H1	HO	0.3761
14	H1’	HC	0.02
15	C3’	CT	0.3259
11	H3’	HC	0.0562
12	C2’	CT	-0.286
13	H2’	HC	0.0941
14	H2”	HC	0.0941
15	O3’	OS	-0.5232

Table S1: Atomic charges (in a.u.) and type specification for the AP site see Figure S3 for the numbering

Minor groove occupation of the etrahelical AP site

The minor groove occupation was estimated considering the diehedral angle between the C4'-C5' (on the base at n+2) and O4'-C1' (on the AP site) see Figure S4. A positive value for the angle indicates minor groove occupancy.

		% minor groove occupancy
seq0	Ap12	92.9
	Ap35	99.9
seq+1	Ap13	17.4
	Ap35	14.4
seq-1	Ap11	18.1
	Ap35	7.9
seq+3	Ap15	76.4
	Ap35	16.0
seq-3	Ap9	92.5
	Ap35	73.4
seq+5	Ap17	28.0
	Ap35	21.5
seq-5	Ap7	39.1
	Ap35	41.1
seqC	Ap12	6.4

Table S2 Minor groove occupancy of the extrahelical AP sites

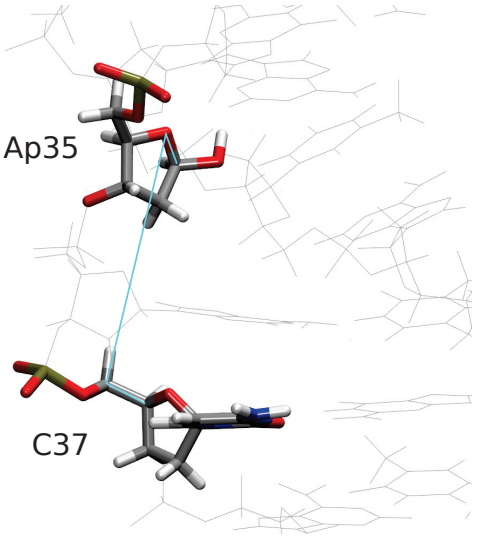


Figure S3 representation of the atoms used to calculated the dihedral angle and estimate minor groove occupancy.

Crucial distances reported in Figure 2-4 in Main text.

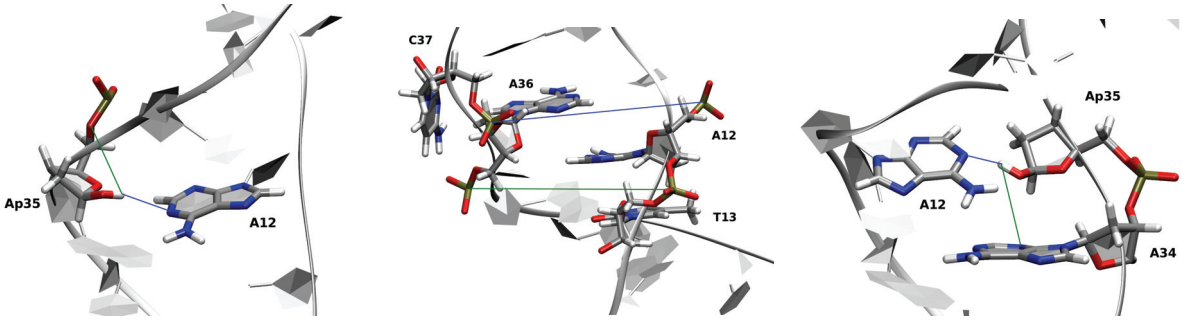


Figure S4 Cartoon representation reported in the tree panels of Figure 2 c) in the main text

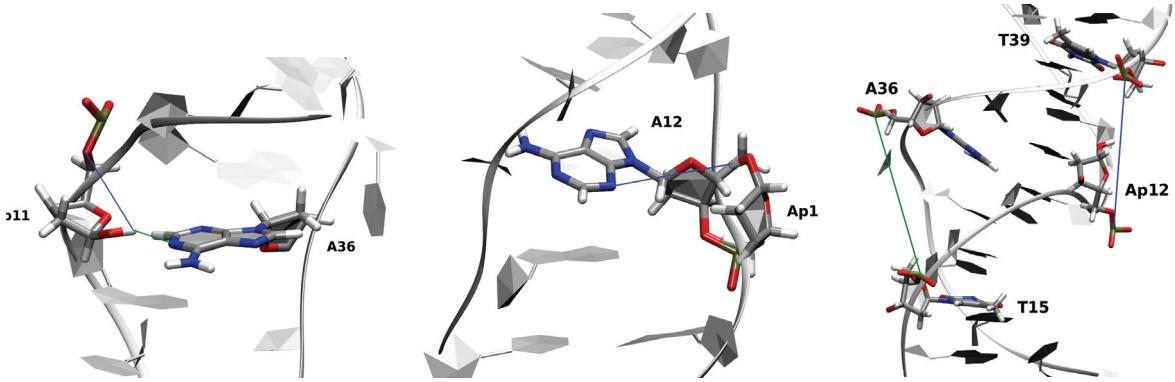


Figure S5 Cartoon representation reported in the tree panels of Figure 3 c) in the main text

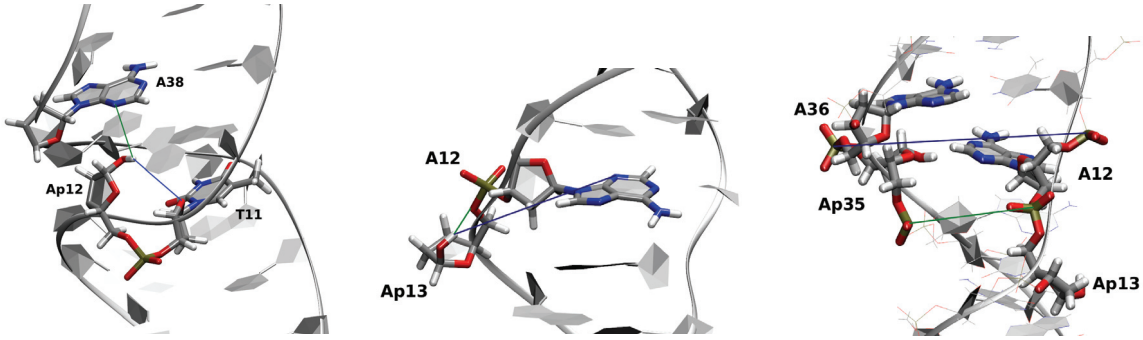


Figure S6 Cartoon representation reported in the tree panels of Figure 4 c) in the main text

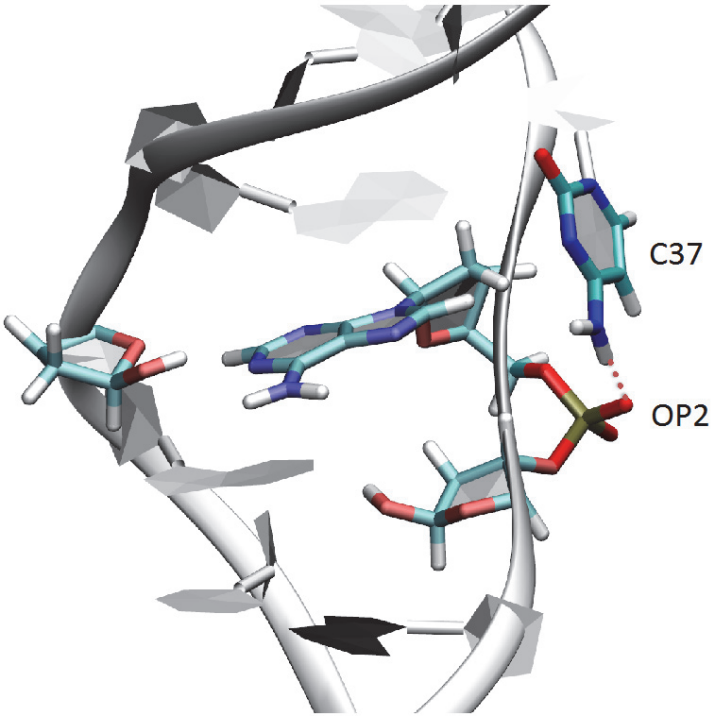


Figure S7 Representative MD snapshot showing the occurence of an HB between the ejected C37 pyrimidine and the backbone oxygen. The latter locks C37 in an extrahelical position.

Choice of the specific oligonucleotide sequence

The choice for the specific oligonucleotide sequence was based on the existent results for DNA AP cluster processing for both enzymes Nfo and APE1 as described in the original publications. Additional reasons for this was that the specific sequence as searched initially by Dr. Georgakilas did not form any hairpin loops and it was quite stable in solution even after the induction of the AP sites. For the control DNA sequence, current calculations agreeing with the older ones, gave an approximate melting temperature of ~51.7 ^{o}C rising to ~59.2 ^{o}C for 50 Mm Na^{+} salt adjusted [1]. For more details on Tm calculations please visit :
http://biotools.nubic.northwestern.edu/OligoCalc.html

In addition using BLAST, a sequence alignment analysis revealed similarity for this oligonucleotide sequence with several primates like Macaca and Human chromosomes 1, 6, 14. Some details are shown below:

BLASTN 2.4.0+ Sequence alignment
http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&RID=TM4RT4FT015

5′-AGAGGATATGTXTGTATGGAGAG-3′
3′- TCTCCTATACATACATACCTCTC-5’

TM4RT4FT015 (Expires on 07-29 22:13 pm)
Query ID **lcl|Query_129577**
Molecule type: nucleic acid
Query Length: 23

Homo sapiens mitochondrial pyruvate carrier 1 (MPC1), RefSeqGene on chromosome 6

Sequence ID: ref|NG_032888.1|Length: 25095Number of Matches: 1

Related Information

Map Viewer-aligned genomic context
Range 1: 16249 to 16268GenBankGraphics

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Minus

```
Query  1      TCTCCTATACATACATACCT  20
              ||||||||||||||||||||
Sbjct  16268  TCTCCTATACATACATACCT  16249
```

Homo sapiens AT-rich interaction domain 4B (ARID4B), RefSeqGene on chromosome 1

Sequence ID: ref|NG_029484.1|Length: 168323Number of Matches: 1

Related Information

Map Viewer-aligned genomic context
Range 1: 115262 to 115281GenBankGraphicsNext MatchPrevious Match

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Minus

```
Query  3       TCCTATACATACATACCTCT  22
               ||||||||||||||||||||
Sbjct  115281  TCCTATACATACATACCTCT  115262
```

Homo sapiens FOSMID clone ABC8-42662600P1 from chromosome 4, complete sequence

Sequence ID: gb|AC234090.4|Length: 39078Number of Matches: 1

Related Information

Map Viewer-aligned genomic context
Range 1: 18162 to 18181GenBankGraphicsNext MatchPrevious Match

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Plus

```
Query  1      TCTCCTATACATACATACCT  20
              ||||||||||||||||||||
Sbjct  18162  TCTCCTATACATACATACCT  18181
```

Homo sapiens FOSMID clone ABC11-49614700O23 from chromosome 14, complete sequence

Sequence ID: gb|AC231261.2|Length: 39862Number of Matches: 1

Related Information

Map Viewer-aligned genomic context

Range 1: 19814 to 19833GenBankGraphics

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Plus

```
Query  1      TCTCCTATACATACATACCT  20
              ||||||||||||||||||||
Sbjct  19814  TCTCCTATACATACATACCT  19833
```

Homo sapiens interleukin 1 receptor accessory protein like 1 (IL1RAPL1), RefSeqGene on chromosome X

Sequence ID: ref|NG_008292.1|Length: 1375337Number of Matches: 1

Related Information

Map Viewer-aligned genomic context

Range 1: 561228 to 561247GenBankGraphics

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Minus

```
Query  1       TCTCCTATACATACATACCT  20
               ||||||||||||||||||||
Sbjct  561247  TCTCCTATACATACATACCT  561228
```

References

1. Case DA, Berryman JT, Betz RM *et al.* AMBER 2015.
2. Pérez A, Marchán I, Svozil D *et al.* Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alphaγConformers. *Biophys J* 2007;**92**:3817–29.
3. Wang J, Wolf RM, Caldwell JW *et al.* Development and testing of a general amber force field. *J Comput Chem* 2004;**25**:1157–74.
4. Georgakilas AG, Bennett P V, Sutherland BM. High efficiency detection of bi-stranded abasic clusters in gamma-irradiated DNA by putrescine. *Nucleic Acids Res* 2002;**30**:2800–8.
5. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 2000;**21**:1049–74.
6. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem A* 2001;**105**:9954–60.

Appendix E

Supporting Informations : Interstrand cross-linking implies contrasting structural consequences for DNA: insights from molecular dynamics.

Supporting information for: Interstrand cross-linking implies contrasting structural consequences for DNA: insights from molecular dynamics

Emmanuelle Bignon,[†,†] Tomáš Dršata,[‡] Christophe Morell,[†] Filip Lankaš,[*,‡,¶] and Elise Dumont[*,§]

Univ Lyon, CNRS, Université Claude Bernard Lyon 1, Ens de Lyon, Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, F-69100 Villeurbanne, France, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Praha 6, Czech Republic, Laboratory of Informatics and Chemistry, University of Chemistry and Technology Prague, Technická 5, 166 28 Praha 6, Czech Republic, and Univ Lyon, Ens de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France

E-mail: filip.lankas@vscht.cz; elise.dumont@ens-lyon.fr

Phone: +420 220 44 4392; +33 (0)4 72 72 88 46. Fax: +420 220 410 320; +33 (0)4 72 72 88 60

[*]To whom correspondence should be addressed

[†]Univ Lyon, CNRS, Université Claude Bernard Lyon 1, Ens de Lyon, Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, F-69100 Villeurbanne, France

[‡]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Praha 6, Czech Republic

[¶]Laboratory of Informatics and Chemistry, University of Chemistry and Technology Prague, Technická 5, 166 28 Praha 6, Czech Republic

[§]Univ Lyon, Ens de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France

List of Figures

List of Tables

Molecular dynamics setup

All classical molecular dynamics (MD) simulations were performed using the Amber12 suite of programs,[S1] using parmbsc0 and gaff[S2] force fields. Our MD simulations have been conducted following a rather standard and well-established protocol, whereas a more critical step was to derive reliable force fields parameters for the Ap and interstrand cross-links lesions. This derivation is described in the next section.

For each oligonucleotide, we built corresponding 21-bp oligonucleotides with the `nab` (Nucleic Acid Builder) module. 40 potassium counterions (Dang parameters) were added to the system in order to neutralize the negatively charged atomic groups of the macromolecule. It was then placed in a truncated octahedral TIP3P water box with a 10 Å buffer. In order to smoothly readapt our starting system with parameter of the force field, the system's energy was minimized by optimizing its geometry along 10000 steps including 5000 steps of conjugated gradient and 5000 steps of steepest descent. Then, the system was stepwisely heated from 0 to 300K during 60ps. During the remaining of the simulation, the temperature was kept at 300K using the Langevin thermostat with a collision frequency γln equal to $1ps^{-1}$. A 1ns equilibration step was then performed, followed by the production step of 200ns for the Ap site containing systems and 500ns for the ICL sequences. A 9 Å cutoff was imposed for electrostatic potential in order to avoid interaction between DNA helix images due to the periodic boundary conditions.

These simulations of 200ns were performed to unravel the structure of the two Ap-containing oligonucleotides, before formation of the cross-link. This provides starting structures that we can use for bringing close together the Ap site and the offset purine forming the ICL identified by Wang.

Structure of the AP-containing oligonucleotides before ICL formation

Our study first implies a first series of MD simulations for the two sequences A and B containing an abasic site as the 11th nucleobase (Ap11), i.e. at the center of the 21-bp duplexes (see Figure 1 of the main text). The behavior of such Ap, as they seek to "fill up the empty space" left upon basis loss has been studied in earlier simulations works.[S3–S5] Indeed along our simulations both structures are found to fill the initial "hole" left by the abasic site in front of A32. The nature of the orphan base, here A32, is known to play a non-innocent role as a purine. Also our simulations confirm that the Ap site is prone to interactions through its hydroxyl group.

We aim more particularly here at unravelling the positionning of Ap11 towards A32, and also towards the purine undergoing the nucleophilic attack, i.e. either G33 or A31. Along 200 ns of unconstrained MD, the sequence A was found to spontaneously evolve towards a more proximal Ap...G, as depicted in Figure 2–a. The approaching distance Ap...G measured between C1′(Ap) and N2(G33) [1] is of 4.84 Å for the final snapshot. In sharp contrast, the sequence B does not spontaneously evolve to bring Ap11 and A31 close together. It was hence required to introduce a biais in a second series of molecular dynamics via an umbrella sampling procedure to enforce the rearrangment within the DNA macromolecule: the distance d(N6/A31$\cdots$C1′/Ap11) was chosen as the reaction coordinate from a near-equilibrium value of 7.0 Å to 4.0 Å. The two structures can be compared in Figures 2–b and 2–c, and differ in free energy by +6 kcal/mol as estimated along our free energy profile. We stress out that DNA is flexible such that over the simulation windows the oligonucleotide has enough time to readapt to the constraint. Such a difference of free energy can be easily accessible at ambient temperature and could also be overestimated by our umbrella sampling procedure.

Hence our simulations provide representative conformations of the prereactants with respect to the formation of an ICL. Our approach allows to monitor the structural rearrangement of the macro-

[1] N2/N6 denotes the nitrogen of the exocyclic amine of guanine/adenine.

molecule as Ap11 and a purine coming close together. This is not necessarily a constraint but can occur spontaneously, as for Sequence A where the ICL Ap-dG is prefigured. The ICL formation corresponds to a multi-step reaction, that could not be described with QM/MM-MD scheme, yet our simulations show that the conformation of the reactant is different from the structures obtained relying on "frozen" geometry of molecular models[S6] where this structural relaxation of DNA is not taken into account.

Building ICL from the Ap containing 21-bp oligonucleotide: adequate structures and derivation of force fields

In order to obtain ICL starting structure, we first performed 200 ns classical MD simulations on the two oligonucleotides containing an abasic site at position 11 of the 21-bp oligonucleotide. Abasic site force field parameters were obtained by antechamber and RESP protocols (see Table S1 and Figure S4). RESP values were derived from quantum chemical calculations at the B3LYP/6-31G(d,p) level of theory. The bond force field parameters were taken from the Generalized Amber Force Field (GAFF)[S2] parameter and parm99bsc0 parameter set.

For Ap-dG, the system rapidly adopts a stable conformation which exhibits a real proximity between the reacting atoms (abasic site C1′ and nitrogen from the purine's exocyclic amino group), whereas for seqB there are strong fluctuations of the distance between these atoms with no really stable conformation. For Ap-dA, we imposed a constraint corresponding to the reaction coordinate for the formation of the ICL, which are to be brought close to each other. An umbrella sampling procedure was applied. We sampled 16 windows of 50ps (hence for a total time simulation of 8ns) with a 20 kcal/mol constraint force from 7.0 Å to 4.0 Å with a 0.2 Å step. At the end of the simulation, the two reacting atoms lie at 4.16 Å from each other, offering a suitable structure for the ICL building.

ICL force field parameters for this latter were computed in the same maneer than for the abasic site (see Tables S2 and S3). RESP charges were computed for the complete ICL, and we generated

two libraries for both fragments (DL1 and DL2) as depicted in Figure S5. The constant force associated to the C1′-N bond was taken at 481.0 kcal/mol/Å^2 with an equilibrium distance of 1.340 Å as listed in the parm99bsc0 force field for a typical CA-N2 bond. Additional parameters were needed for angles and improper dihedral angle in order to describe at best the covalent bond between the abasic site and the purine (see Table S4). Using the `xleap` utilitary, we managed to obtain reasonably relaxed starting structures for ICL containing systems.

Hybrid QM/MM molecular dynamics simulation

The classical force field description was checked against 1 ns hybrid QM/MM-MD simulations that we performed within Amber. The ICL was described using the AM1 semi-empirical approach. The two moieties belonging to the ICL were encompassed in the QM part (Ap11 nucleoside and A31/G33 nucleobase). The starting structure was chosen as the last snapshot of the 500ns full classical MD simulation. Superposition of the final ICL structures and RMSD obtained by MM-MD and QM/MM-MD shows no drastical deviation (see Figures S6, S7 and S8). The average Ap(C1′)-G33(N2) distance monitored for sequence A lies at 1.46 ± 0.12 Å for the MM-MD production and 1.51 ± 0.14 Å for the QM/MM-MD simulation. For sequence B, the distance is of 1.46 ± 0.13 Å in purely classical MM-MD simualtions and 1.49 ± 0.11 Å in QM/MM-MD.

DFT-based evaluation of the energy difference

From the representative snapshots extracted based on a minimal RMSD by cluster analysis, we extracted a cluster featuring the ICL and four vicinal nucleobases: the two clusters are represented in Figure 1 of the main text. Dangling bonds were satured by hydrogen and the corresponding structures were optimized at the BLYP-D3BJ/6-31(d) level of theory, using the Gaussian09 suite of programs. We compute the ICL formation (de)stabilization energies $\Delta E^{ICL} = E_{prod} - E_{react}$ between the reactants and the products for a fragment featuring the reactive system and the two vicinal base pairs. This provides a semi-quantitative estimate of the energy cost or gain upon ICL

formation, which turns out to be almost identical for the Ap...Purine pairs, but strigently different for covalently-linked adducts. This characterized out two lesions with contrasted energy profile, as a consequence of the different reorganization the B-helix undergoes and can be related to the experimentally different yields of formation.

Figure S1: Electronic energy differences evaluated at the BLYP-D3BJ/6-31(d) level of theory on representative snapshots. The QM part encompasses the capped nucleobase (A31 or G33) and the Ap site, whereas the rest of the environment is ommitted.

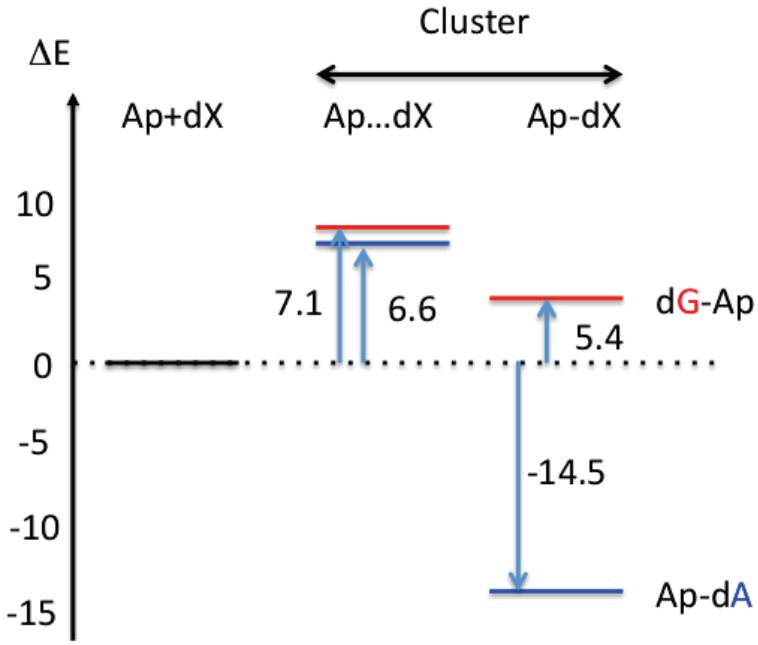
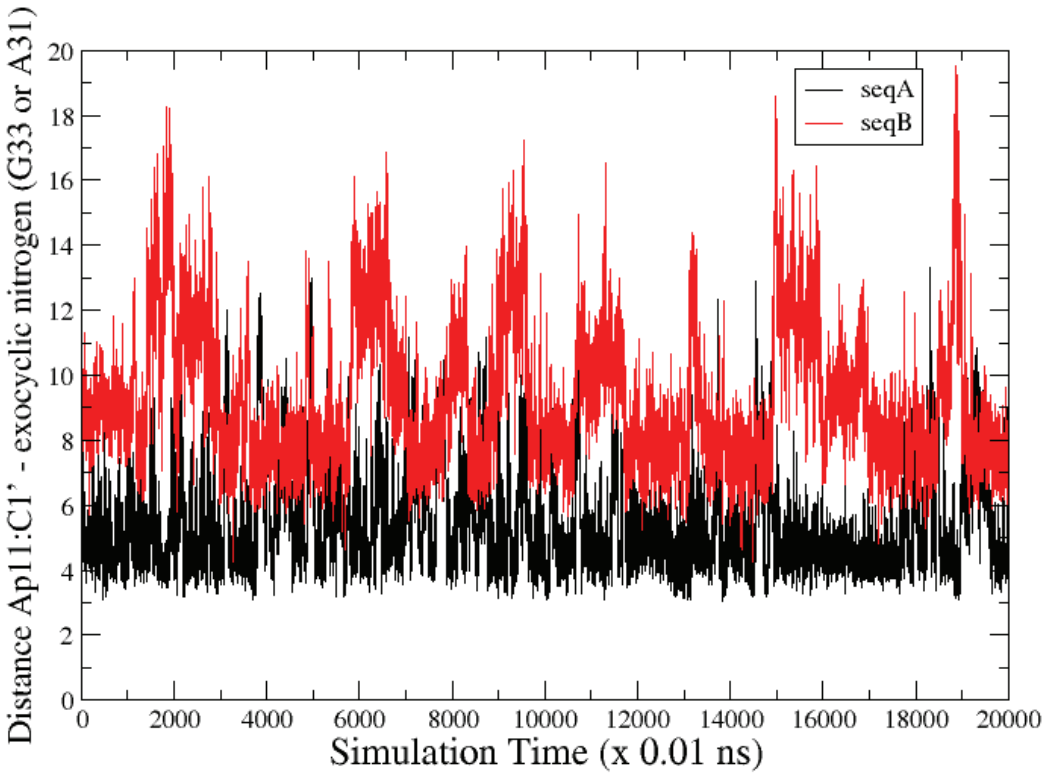


Figure S2: Approaching distances d(C1′...N) along simulations performed with the parmbsc1 force field.



Molecular dynamics with the opened form of the Ap site

An abasic lesion is not a chemically unique species but exists as an equilibrium mixture of the ring-closed cyclic hemiacetals and open-chain aldehyde: the latter is transient, accounting for 1% of the time, yet triggers the multisteps pathway degradation towards ICL formation.[?] We thus decided to test whether or not the approach of the Ap is affected by its chemical nature.

For sequence A, C1′ is directed inside the B-helix at 3–4 Å from the N2 atom of G33, which is favorable for the formation the ICL, 60% of the time and the rest of the time toward the bulk. The d(C1′...N) is increased and fluctuates around 10–15 Å. One swaps between the two forms with a small free energy barrier, and with only a very slight perturbation of the DNA helix. Oppositely, for the duplex B, the aldehyde never approaches the adenine A31, and it appears along our simulations that its position downstream of the helix (5') is less accessible. Consequently, the distance d(C1′...N) strongly oscillated around 8 Å.

For both Ap-containing sequences, the hydroxyle interacts very strongly with the oxygen of the phosphate moiety (O1P or O2P). The C2'–C1' linkage points toward the 3'of the complemmentary strand. Du coup pour atteindre la A31 il faudrait dÃl'former un peu quand mÃłme alors que G33 est vraiment juste Ãă cÃt'tÃl'. Je lance un US pour voir si Ãğa donne la mÃłme chose qu'avec le Ap fermÃl'.

Overall, the behavior of the Ap site is similar whether it is cyclic or open, privileging an attack towards G33. But it is not the decisive feature that govern the ICL selective formation, as ...

Figure S3: Approaching distances d(C1′...N) for the open form of the abasic site

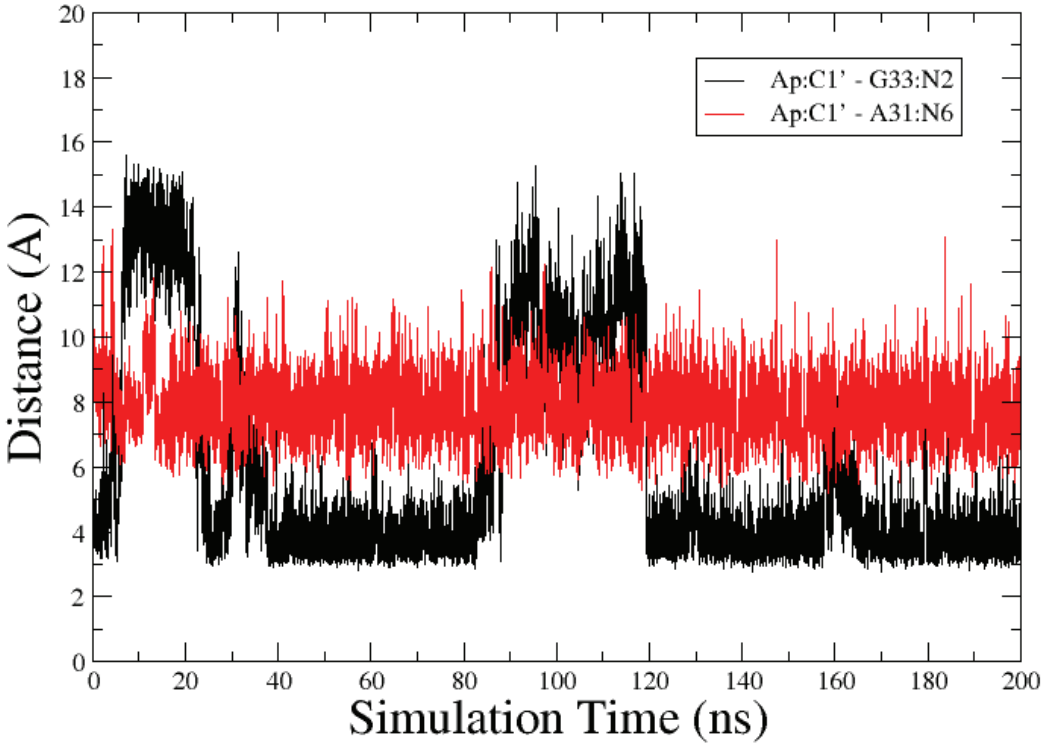


Table S1: Atoms specifications and charges of Ap site. For atom numbers, see Figure S4.

Number	Name	Type	Charge
1	P	P	1.1659
2	OP2	O2	-0.7761
3	OP1	O2	-0.7761
4	O5’	OS	-0.4954
5	C5’	CI	-0.0069
6	H5’	H1	0.0754
7	H5”	H1	0.0754
8	C4’	CT	0.2098
9	H4’	H1	0.0046
10	O4’	OS	-0.4058
11	C1’	CT	0.3679
12	O1	OH	-0.5959
13	H1	HO	0.3761
14	H1’	HC	0.02
15	C3’	CT	0.3259
16	H3’	HC	0.0562
17	C2’	CT	-0.286
18	H2’	HC	0.0941
19	H2”	HC	0.0941
20	O3’	OS	-0.5232

Figure S4: Ap site scheme with atom numbers.

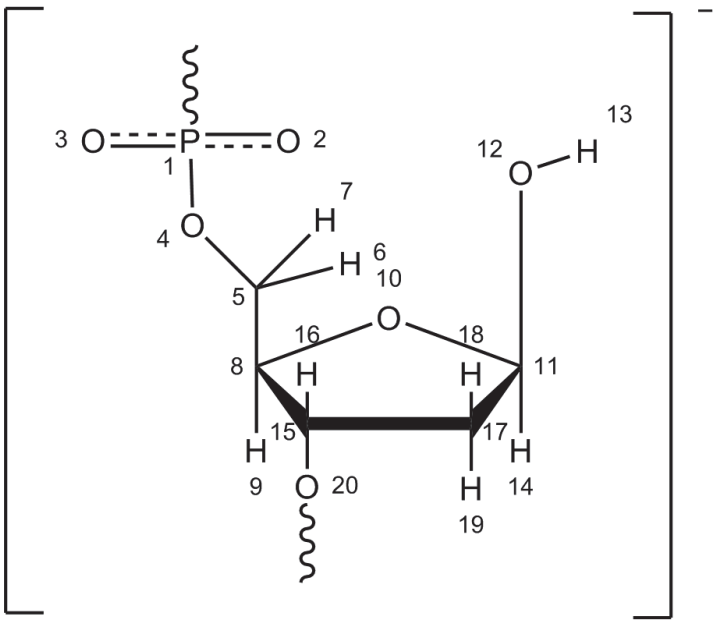


Figure S5: Scheme of a) Ap-dG and b) Ap-dA with atom numbers. DL1 residue is depicted in black and DL2 in red.

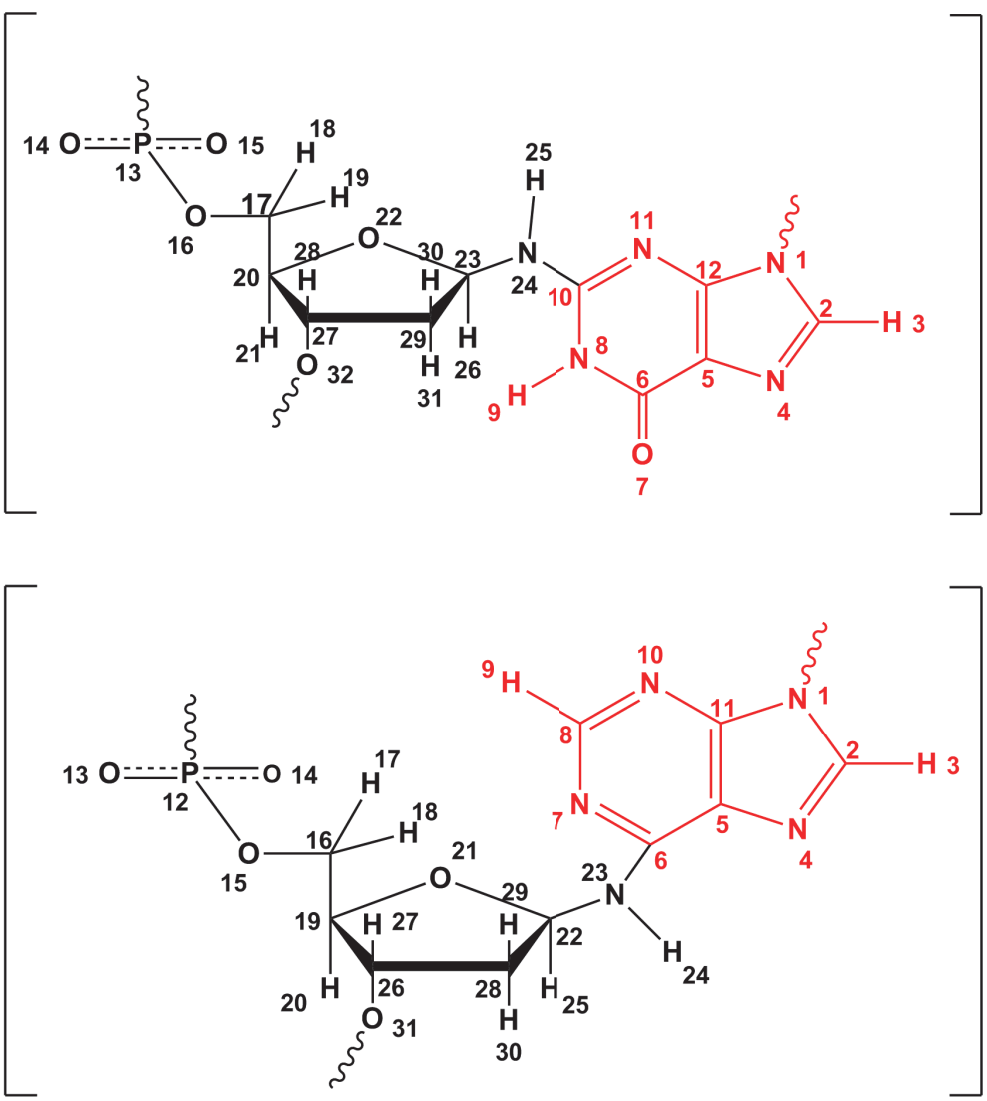


Table S2: Atoms specifications and charges of ICL in sequence A. For atom numbers, see figure S5.

Number	Name	Type	Charge
1	N9	N*	-0.0027
2	C8	CK	0.1609
3	H8	H5	0.1433
4	N7	NB	-0.5391
5	C5	CB	0.0963
6	C6	C	0.523
7	O6	O	-0.5376
8	N1	NA	-0.5735
9	H1	H	0.3509
10	C2	CA	0.759
11	N3	NC	-0.5876
12	C4	CB	0.1565
13	P	P	1.1659
14	OP1	O2	-0.7761
15	OP1	O2	-0.7761
16	O5’	OS	-0.4954
17	C5’	CI	-0.0069
18	H5’	H1	0.0754
19	H5”	H1	0.0754
20	C4’	CT	0.1629
21	H4’	H1	0.1176
22	O4’	OS	-0.3691
23	C1’	CT	0.0431
24	N10	N2	-0.6391
25	H10	H	0.4956
26	H1’	H2	0.1838
27	C3’	CT	0.0713
28	H3’	H1	0.0985
29	C2’	CT	-0.0854
30	H2’	HC	0.0718
31	H2”	HC	0.0718
32	O3’	OS	-0.5232

Table S3: Atoms specifications and charges of ICL in sequence B. For atom numbers, see figure S5.

Number	Name	Type	Charge
1	N9	N*	-0.1992
2	C8	CK	0.2822
3	H8	H5	0.1247
4	N7	NB	-0.5005
5	C5	CB	-0.0819
6	C6	CA	0.7149
7	N1	NC	-0.7061
8	C2	CQ	0.5058
9	H2	H5	0.0275
10	N3	NC	-0.7194
11	C4	CB	0.5099
12	P	P	1.1659
13	OP1	O2	-0.7761
14	OP1	O2	-0.7761
15	O5’	OS	-0.4954
16	C5’	CI	-0.0069
17	H5’	H1	0.0754
18	H5”	H1	0.0754
19	C4’	CT	0.1629
20	H4’	H1	0.1176
21	O4’	OS	-0.3691
22	C1’	CT	0.0431
23	N10	N2	-0.6404
24	H10	H	0.4719
25	H1’	H2	0.1838
26	C3’	CT	0.0713
27	H3’	H1	0.0985
28	C2’	CT	-0.0854
29	H2’	HC	0.0718
30	H2”	HC	0.0718
31	O3’	OS	-0.5232

Table S4: Additional parameters for the ICL non canonical residue

Angles				
OS-CT-N2	70.04	111.23		Same as n2-c3-os in gaff
H2-CT-N2	49.60	110.33		Same as h2-c3-nh in gaff
Improper dihedral angles for seqA				
H5-N*-CK-NB	1.1	180.0	2.0	General improper torsional angle (2 general atom types)
CB-NA-C -O	10.5	180.0	2.0	General improper torsional angle (2 general atom types)
C -CA-NA-H	1.0	180.0	2.0	General improper torsional angle (2 general atom types)
Improper dihedral angles for seqB				
H5-N*-CK-NB	1.1	180.0	2.0	Using the default value
CA-CB-CB-NB	1.1	180.0	2.0	Using the default value
H5-NC-CQ-NC	1.1	180.0	2.0	Using the default value
CB-N*-CB-NC	1.1	180.0	2.0	Using the default value

Figure S6: Superposition of the Ap-dA structures obtained after 500ns MM-MD (orange) and 1ns semi-empirical QM/MM-MD simulations. The rest of the oligonucleotide and the environment are omitted.

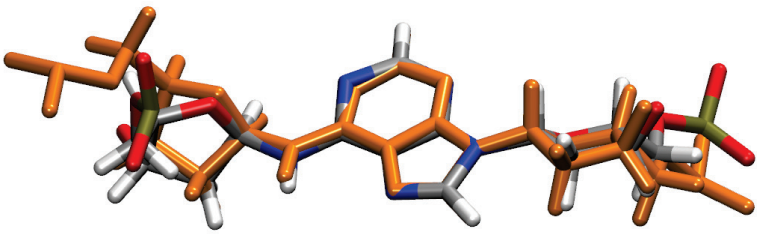


Figure S7: Superposition of the Ap-dG structures obtained after 500ns MM-MD (orange) and 1ns semi-empirical QM/MM-MD simulations. The rest of the oligonucleotide and the environment are omitted.

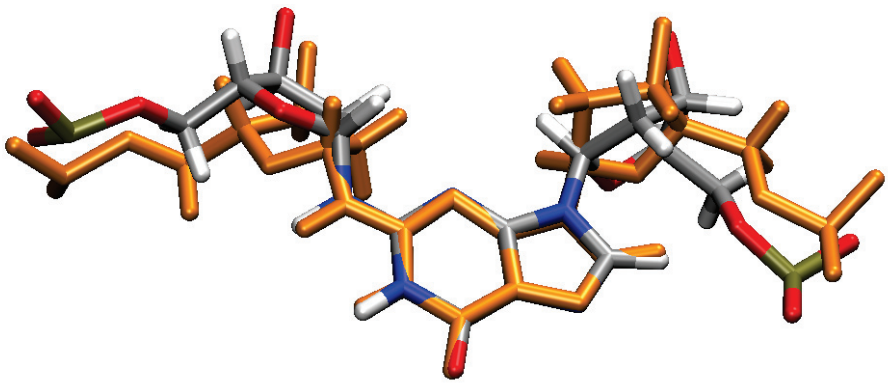


Figure S8: Root Mean Square Deviation of Ap-dG and Ap-dA along the 1ns QM/MM-MD simulation.

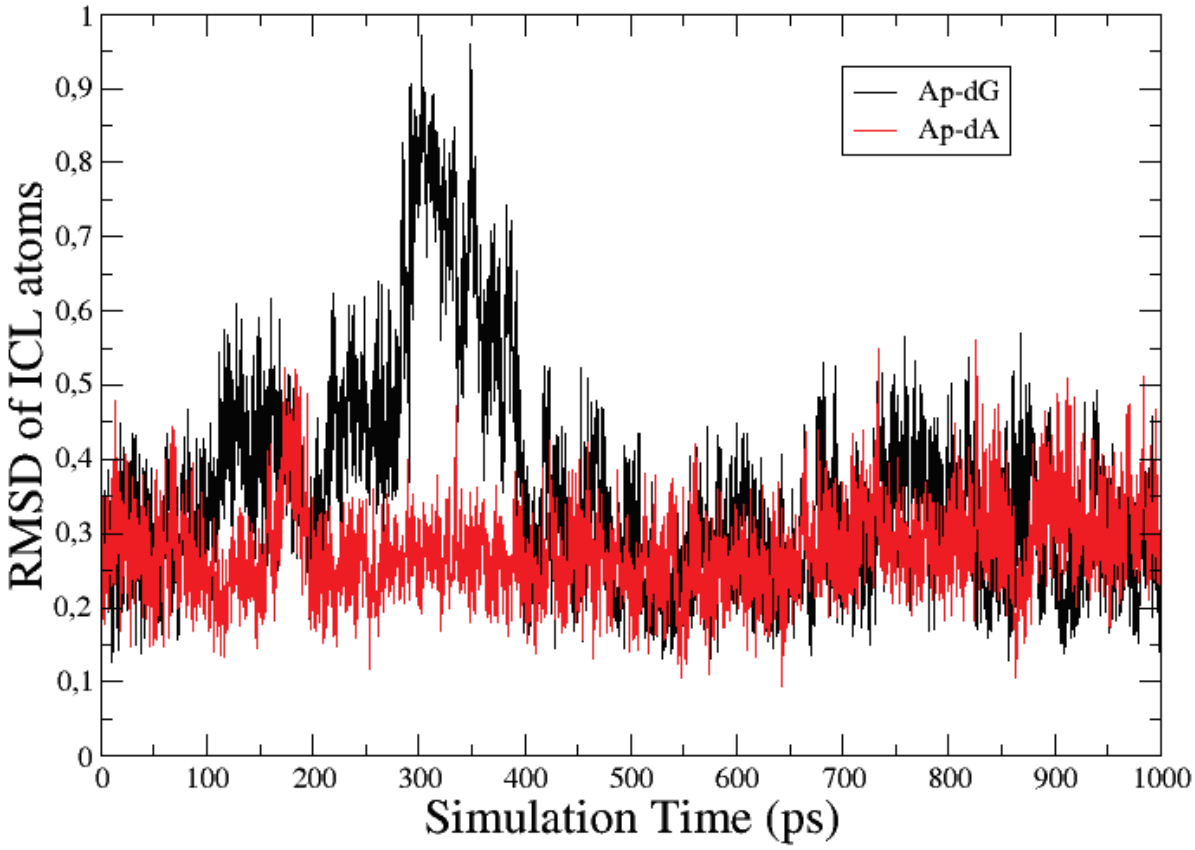


Figure S9: Probability density functions of the basepair step coordinates for the virtual steps in damaged and undamaged sequence A. Red curves indicate the Gaussian with the same mean and standard deviation as the MD data (blue).

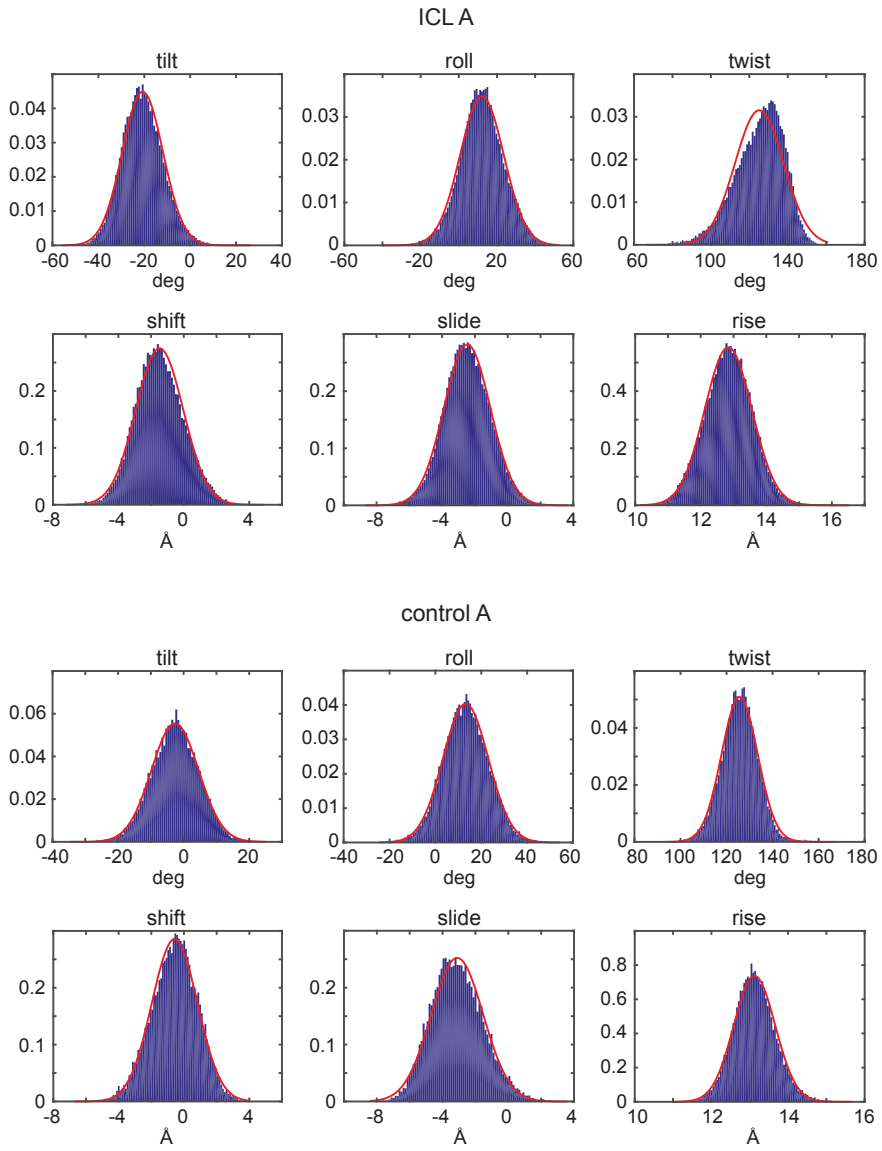


Figure S10: Probability density functions of the basepair step coordinated for the virtual steps in damaged and undamaged sequence B. Red curves indicate the Gaussian with the same mean and standard deviation as the MD data (blue).

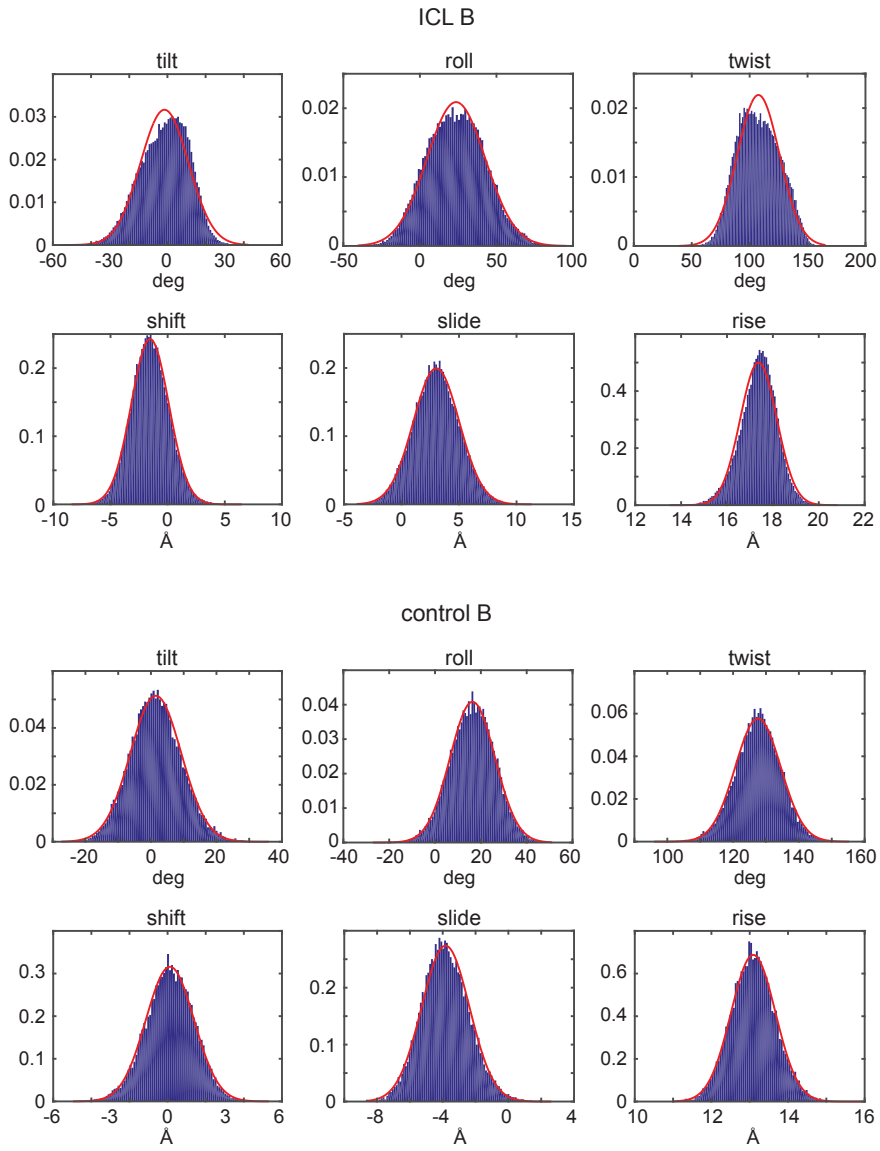
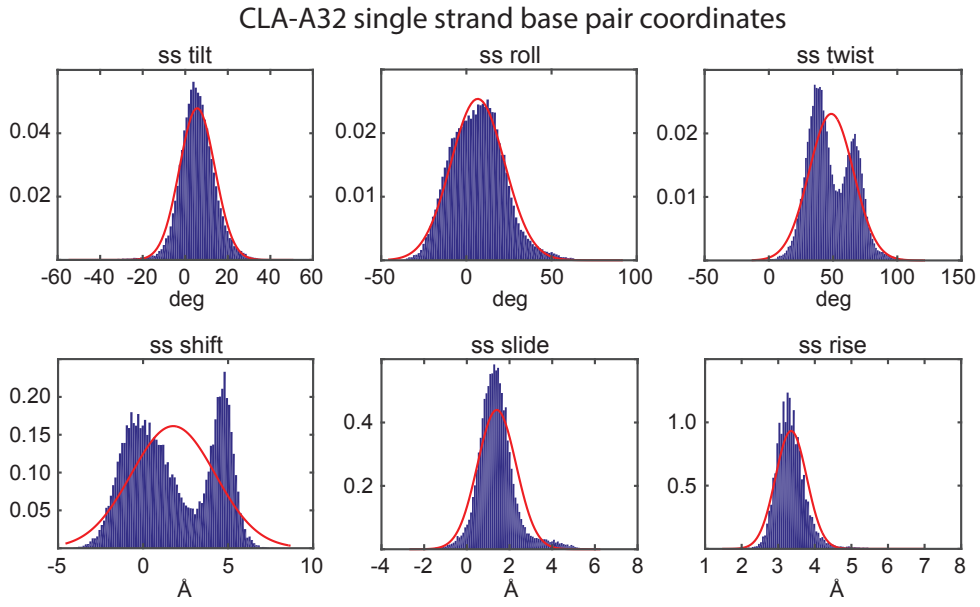


Figure S11: Probability density functions of single-strand inter-base coordinates between the CL(A) lesion and the adenine A32. Red curves indicate the Gaussian with the same mean and standard deviation as the MD data (blue).



References

(S1) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. AMBER 12. 2012; `http://ambermd.org/`.

(S2) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(S3) Ayadi, L.; Coulombeau, C.; Lavery, R. Abasic sites in duplex DNA: molecular modeling of sequence-dependent effects on conformation. *Biophys. J.* **1999**, *77*, 3218–3226.

(S4) Barsky, D.; Foloppe, N.; Ahmadia, S.; Wilson III, D. M.; MacKerell Jr, A. D. New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.* **2000**, *28*, 2613–2626.

(S5) Chen, J.; Dupradeau, F.-Y.; Case, D. A.; Turner, C. J.; Stubbe, J. DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.* **2008**, *36*, 253–262.

(S6) Price, N. E.; Johnson, K. M.; Wang, J.; Fekry, M. I.; Wang, Y.; Gates, K. S. Interstrand DNAâĂŞDNA Cross-Link Formation Between Adenine Residues and Abasic Sites in Duplex DNA. *Journal of the American Chemical Society* **2014**, *136*, 3483–3490, PMID: 24506784.

Appendix F

Supporting Informations : Conformational polymorphism or structural invariance in DNA photoinduced lesions: implications for repair rates

Supplementary Information for

Conformational polymorphism or structural invariance in DNA photoinduced lesions: implications for repair rates.

François Dehez[1,2,3,*], Hugo Gattuso[1,2], Emmanuelle Bignon[4,5], Christophe Morell[4], Elise Dumont[5,*] and Antonio Monari[1,2,*]

[1]CNRS, Theory-Modeling-Simulation, SRSMC F-54506 Vandoeuvre-lès-Nancy, France [2]Universitéd e Lorraine, Theory-Modeling-Simulation, SRSMC F-54506 Vandoeuvre-lès-Nancy, France [3]Laboratoire International Associe Centre National de la Recherche Scientifique et University of Illinois at UrbanaChampaign [4]Institut des Sciences Analytiques, UMR 5280, Université de Lyon1 (UCBL) CNRS, ENS Lyon, Lyon, France and [5]Université de Lyon, ENS de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie, F69342, Lyon, France

Molecular Dynamics 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 64-PP and CPD lesions' sites according to the standard Antechamber protocol (see table S1). Then, the system was placed in an octahedral TIP3P water box of 12Å buffer, with 30 potassium (K^+) cations to ensure the medium's neutrality. For CPD and B-DNA containing sequences, 100ns classical molecular dynamics were performed after an equilibration procedure we described hereafter. On the contrary, for the 64-PP sequence the production run reached 2μs. 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 64-PP 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 300K in a 20ps themalization run. Langevin

thermostat with a 1.0 ps^{-1} collision frequency γln was used to keep it constant during the remaining of the simulation. Then, a first 100ps 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 (1) 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 64PP lesion was evolved via replica exchange MD for 75 ns, representing a total sampling time of 4.8 μs. Target temperatures of the 64 replicas ranged from 290 K to 418 K, a temperature step of 2K was chosen to ensure an exchange success rate of about 30%. Exchange attempts between adjacent replicas were made every 200 fs.

Force-Field Point Charges for non-standard residues

DTO (T8) 64-PP		
Name	Atom Type	Charge
P	P	1.1659
O1P	O2	-0.7761
O2P	O2	-0.7761
O5'	OS	-0.4954
C5'	CI	-0.0069
H5'1	H1	0.0754
H5'2	H1	0.0754
C4'	CT	0.1629
H4'	H1	0.1176
O4'	OS	-0.3691
C1'	CT	0.068
H1'	H2	0.1804
N1	N*	-0.2499
C2	C	0.6751
N3	NA	-0.6212
C4	C	0.6737
C5	CT	0.1976
C6	CT	0.0097
O2	O	-0.5288
H3	H	0.3698
O4	O	-0.5546
O5	OH	-0.5681
HO5	HO	0.3847
C7	CT	-0.4732
H71	HC	0.1469
H72	HC	0.1469
H73	HC	0.1469
H6	H1	0.1837
C3'	CT	0.0713
H3'	H1	0.0985
C2'	CT	-0.0854
H2'1	HC	0.0718
H2'2	HC	0.0718
O3'	OS	-0.5232

Table S1: Atom Types and Charges for the DTO residue T8 in 64-PP

DPR (Pyo9) 64-PP		
Name	Atom Type	Charge
P	P	1.1659
O1P	O2	-0.7761
O2P	O2	-0.7761
O5'	OS	-0.4954
C5'	CI	-0.0069
H5'1	H1	0.0754
H5'2	H1	0.0754
C4'	CT	0.1629
H4'	H1	0.1176
O4'	OS	-0.3691
C1'	CT	-0.0116
H1'	H2	0.1963
N1	N*	-0.2312
C2	C	0.8357
N3	NC	-0.6173
C4	CA	0.3679
C5	CM	-0.0619
C6	CM	-0.0547
O2	O	-0.5815
C7	CT	-0.2735
H71	HC	0.1012
H72	HC	0.1012
H73	HC	0.1012
H6	H4	0.184
C3'	CT	0.0713
H3'	H1	0.0985
C2'	CT	-0.0854
H2'1	HC	0.0718
H2'2	HC	0.0718
O3'	OS	-0.5232

Table S2: Atom Types and Charges for the DPR residue Pyo9 in 64-PP

T8 & T9 CPD		
Name	Atom Type	Charge
P	P	1.1659
O1P	O2	-0.7761
O2P	O2	-0.7761
O5'	OS	-0.4954
C5'	CI	-0.0069
H5'	H1	0.0754
H5''	H1	0.0754
C4'	CT	0.1629
H4'	H1	0.1176
O4'	OS	-0.3691
C1'	CT	0.0431
H1'	H2	0.1838
N9	NS	-0.0268
C8	CK	0.1607
H8	H5	0.1877
N7	NB	-0.6175
C5	CB	0.0725
C6	CA	0.6897
N6	N2	-0.9123
H61	H	0.4167
H62	H	0.4167
N1	NC	-0.7624
C2	CQ	0.5716
H2	H5	0.0598
N3	NC	-0.7417
C4	CB	0.38
C3'	CT	0.0713
H3'	H1	0.0985
C2'	CT	-0.0854
H2'	HC	0.0718
H2''	HC	0.0718
O3'	OS	-0.5232
C6	CT	-0.012
N1	NS	-0.302
C2	C	0.77
N3	NA	-0.583
C4	C	0.52
C5	CT	0.16
O2	O	-0.583
H3	H	0.37
O4	O	-0.483
H71	HC	0.12
H72	HC	0.12
H73	HC	0.12
H6	H1	0.07
C7	CT	-0.413
P	P	1.1659
O1P	O2	-0.7761
O2P	O2	-0.7761
O5'	OS	-0.4954
C5'	CI	-0.0069
H5'	H1	0.0754
H5''	H1	0.0754
C4'	CT	0.1629
H4'	H1	0.1176
O4'	OS	-0.3691
C1'	CT	0.068
H1'	H2	0.1804
C3'	CT	0.0713
H3'	H1	0.0985
C2'	CT	-0.0854
H2'	HC	0.0718
H2''	HC	0.0718
O3'	OS	-0.5232

Table S3: Atom Types and Charges for the T8 and T9 residues in CPD

Definition of the π-stacking extent for the different 64-PP conformers

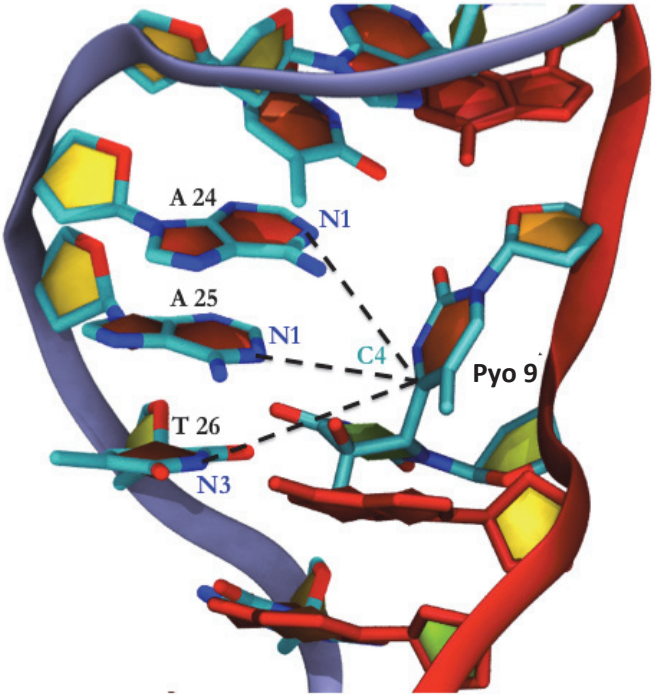


Figure S1: Definition of the distances used to identify the extent of p-stacking between 64-PP and the nearby bases. In particular the stacking with the pyrimidine (Pyo9) unit is considered for the three closest nucleobases belonging to the opposite strand (A24, A25, T26). See Figure 4 in the main text for the results.

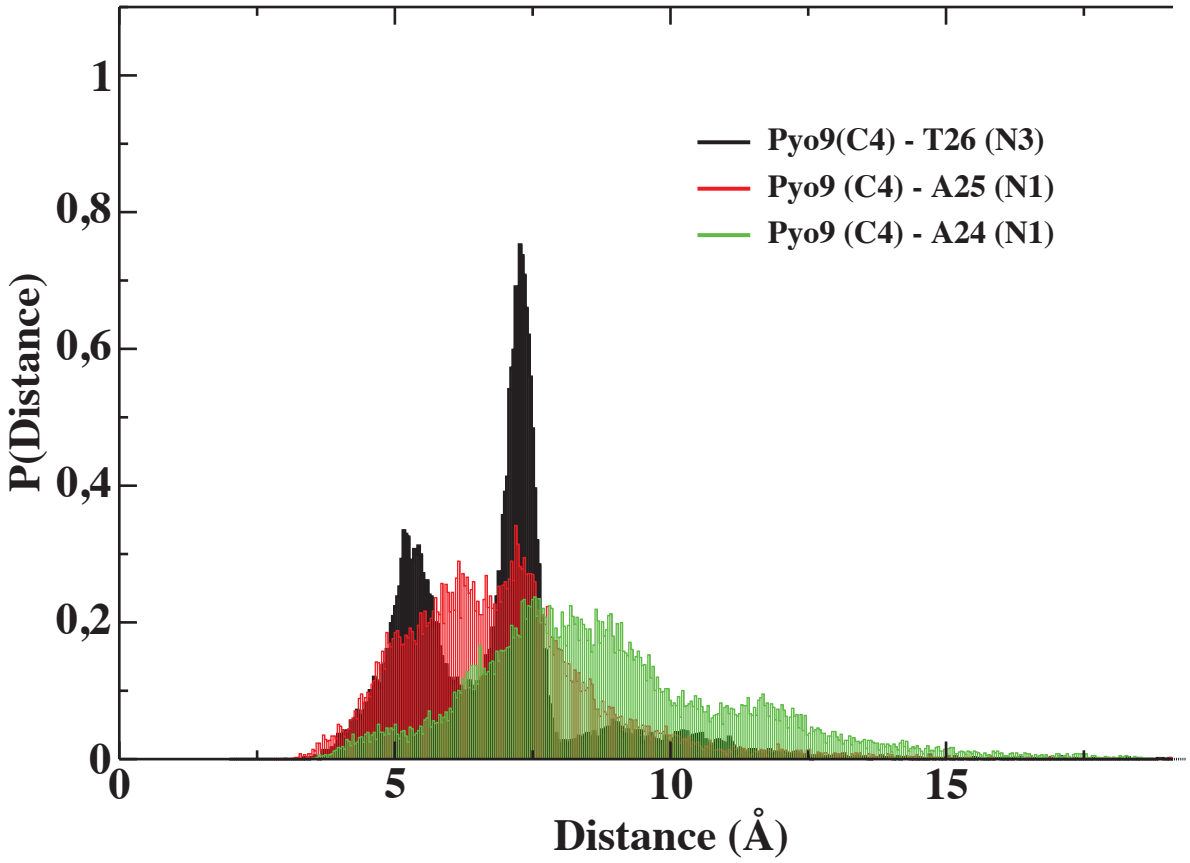


Figure S2: Probability distribution functions of internucleobase distances between T9 and A24, A24 and A26 measured along the 2 μs trajectory.

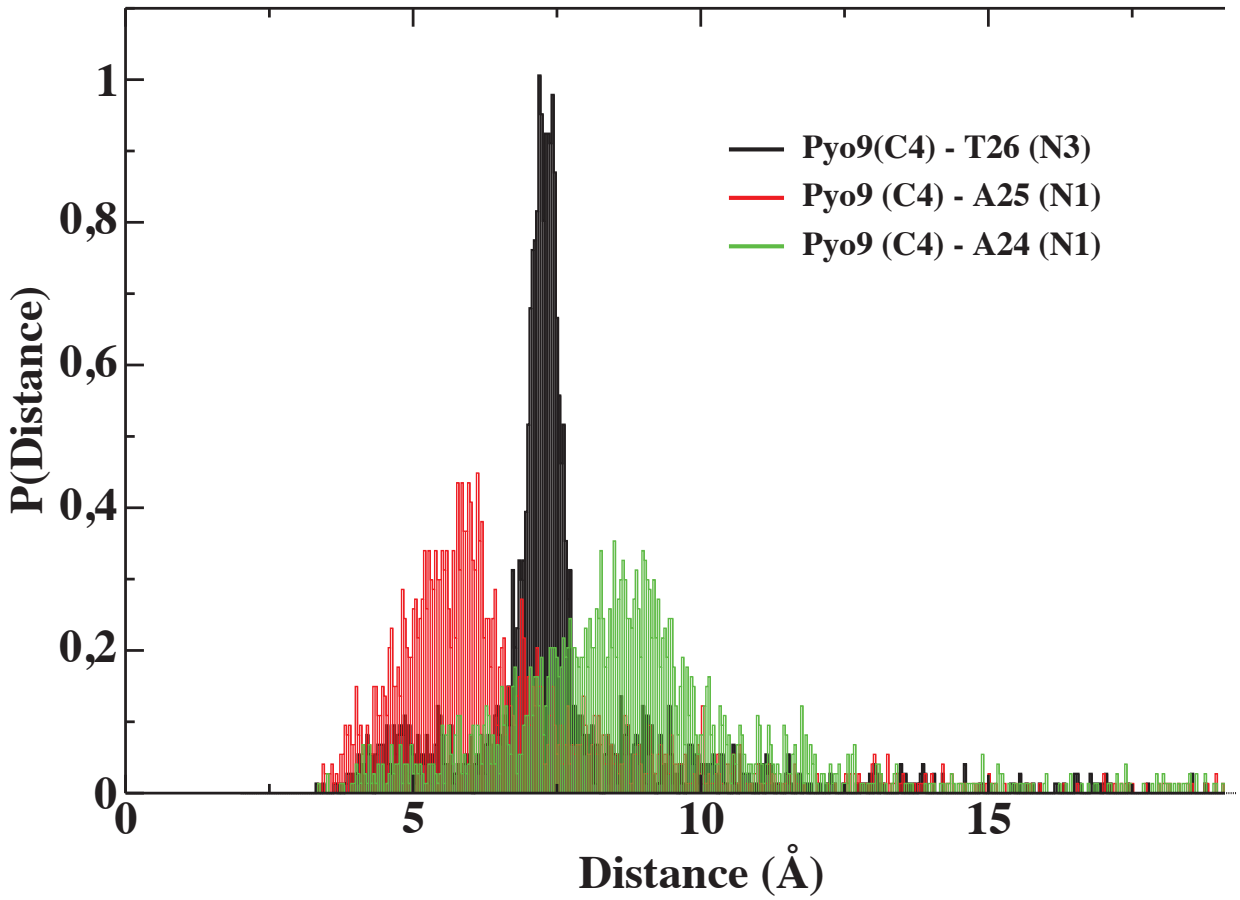


Figure S3: Probability distribution functions of internucleobase distances between T9 and A24, A24 and A26 measured from configurations at 300K extracted from temperature-exchange simulations.

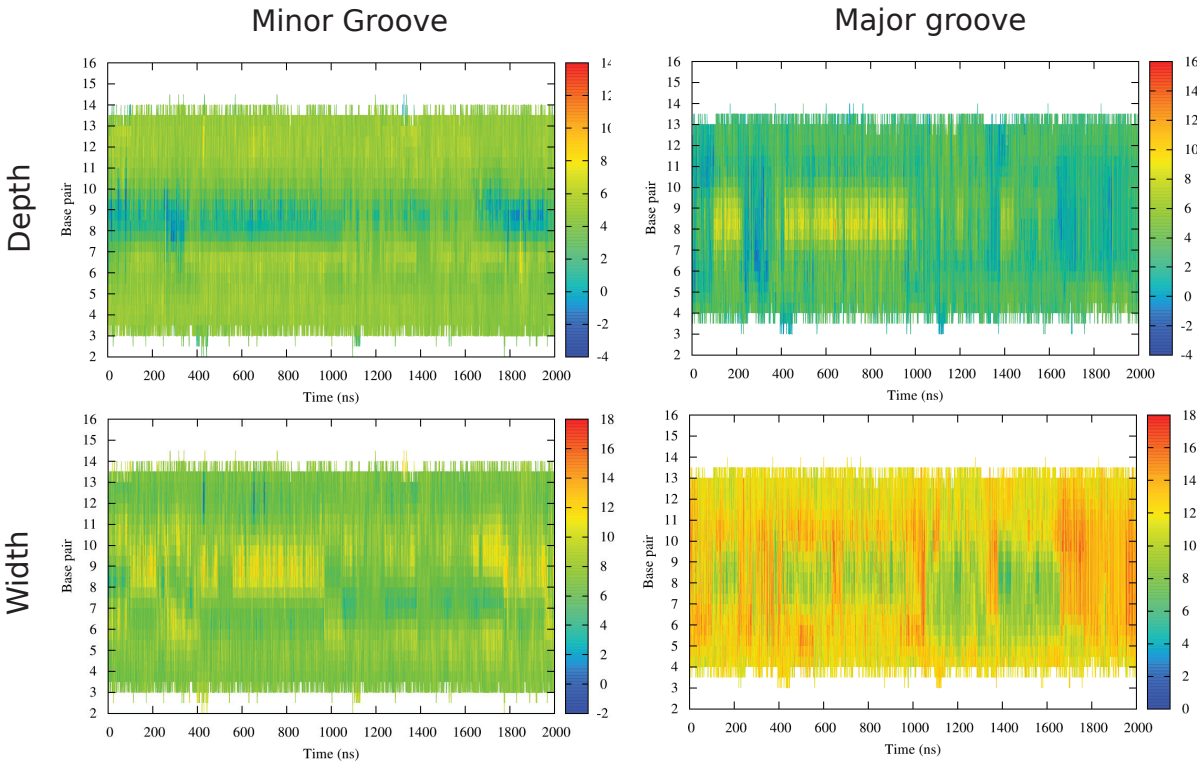


Figure S4: Time series maps of the evolution of the minor (left) and major (right) groove width and depth

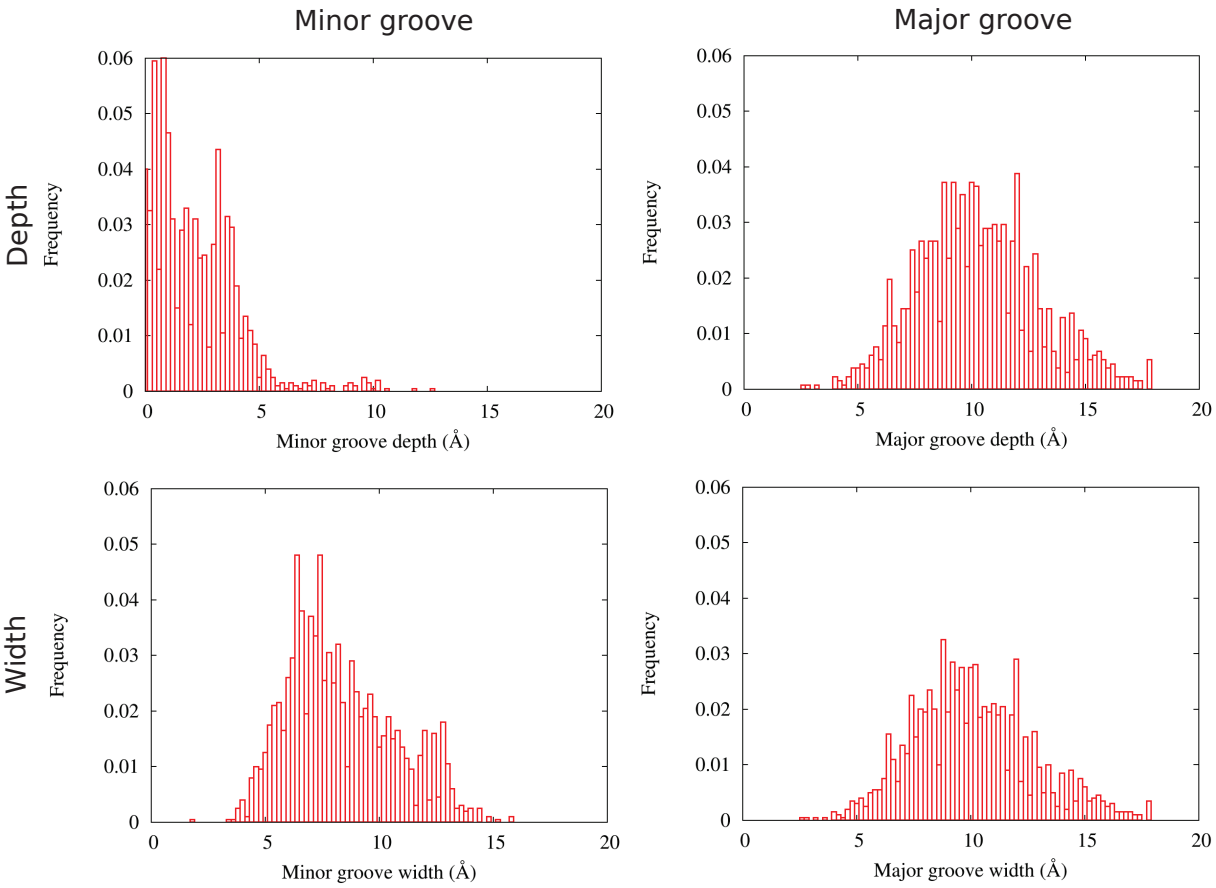


Figure S5: Distribution of the minor (left) and major (right) groove width and depth in correspondence of the 64-PP lesion

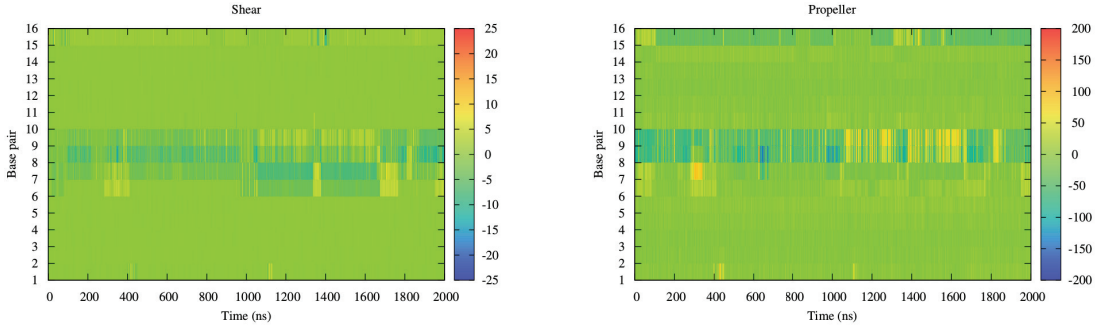


Figure S6: Time series maps of the evolution of the shear (left) and propeller (right) parameters

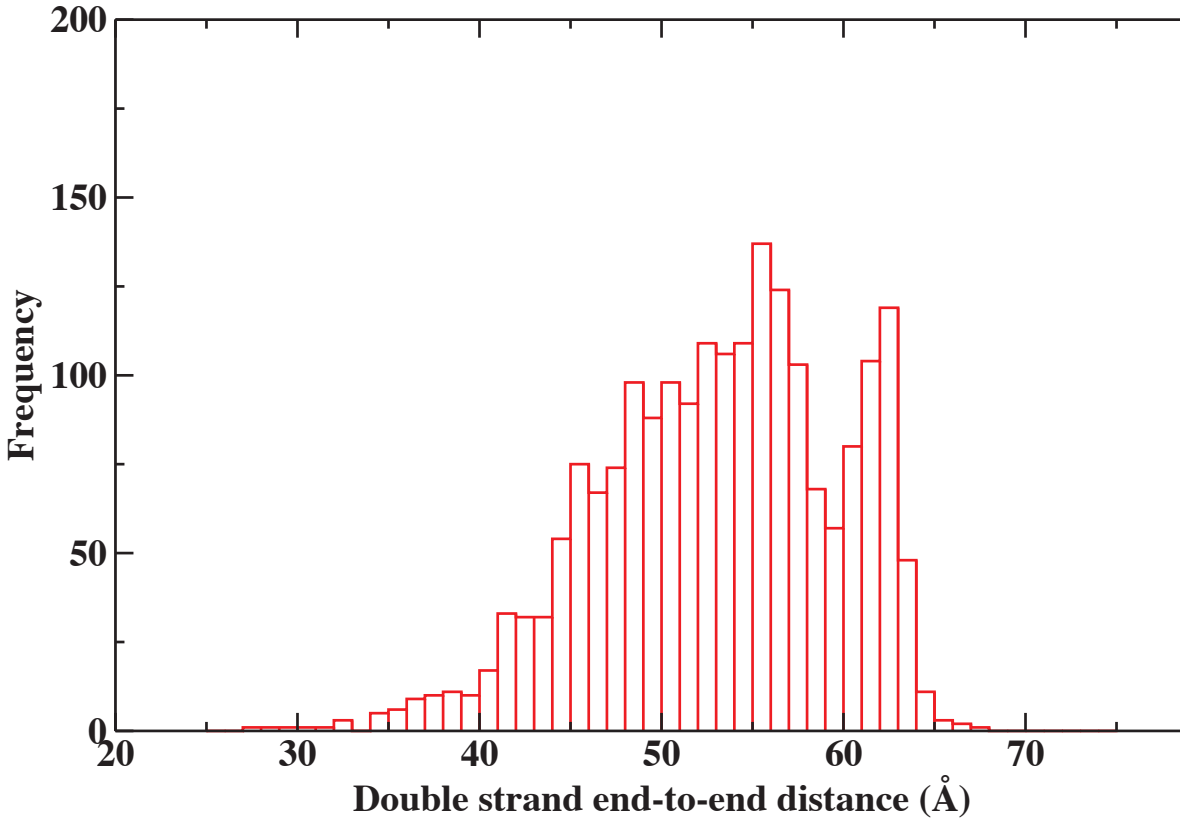


Figure S7: Distribution of the end-to-end distance of the double strand

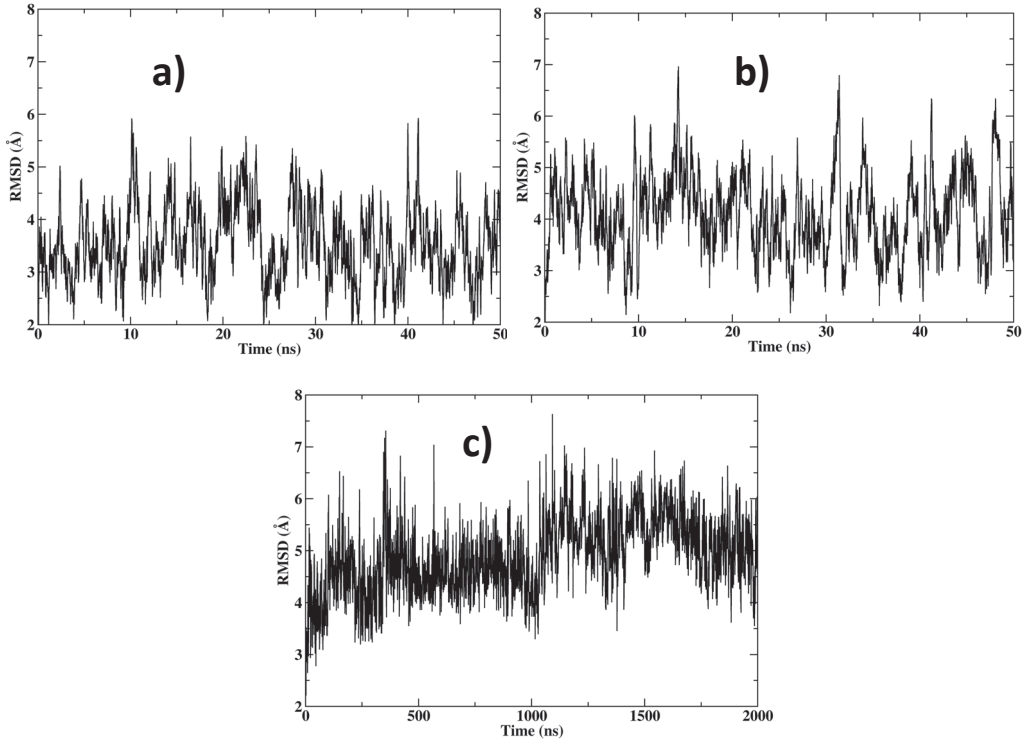


Figure S8: Time series of the RMSD with respect to the the initial structure for B-DNA (a), CPD (b) and 64-PP (c).

A representative movie of the 64-PP trajectory illustrating the polymorphisms exhibited by the double strand is also available.

Appendix G

Supporting Informations : Molecular dynamics insights into polyamines-DNA binding modes: implications for cross-links selectivity.

Supporting Information for: Molecular dynamics insights into polyamines-DNA binding modes: implications for cross-links selectivity.

Emmanuelle Bignon,[*,‡,§] Chen-Hui Chan,[§] Christophe Morell,[‡] Antonio Monari,[‖,⊥] Jean-Luc Ravanat,[¶] and Elise Dumont[*,§]

‡*Institut des Sciences Analytiques, UMR 5280, Université de Lyon1 (UCBL) CNRS, ENS Lyon, Lyon, France*

§*Laboratoire de Chimie, Univ Lyon, Ecole Normale Supérieure de Lyon, CNRS UMR 5182, Université Lyon 1, Laboratoire de Chimie, 46 allée d'Italie, F-69364, LYON, France*

‖*Université de Lorraine Nancy, Theory-Modeling-Simulation, SRSMC F-54506 Vandœuvre-lès-Nancy, France*

⊥*CNRS, UMR 7565, SRSMC, F-54506 Vandœuvre-lès-Nancy, France*

¶*CEA and Université Grenoble Alpes, INAC-SyMMES, F-38000 Grenoble, France*

E-mail: emmanuelle.bignon@univ-lyon1.fr; elise.dumont@ens-lyon.fr

Phone: +33 (0)4 72 72 88 46. Fax: +33 (0)4 72 72 88 60

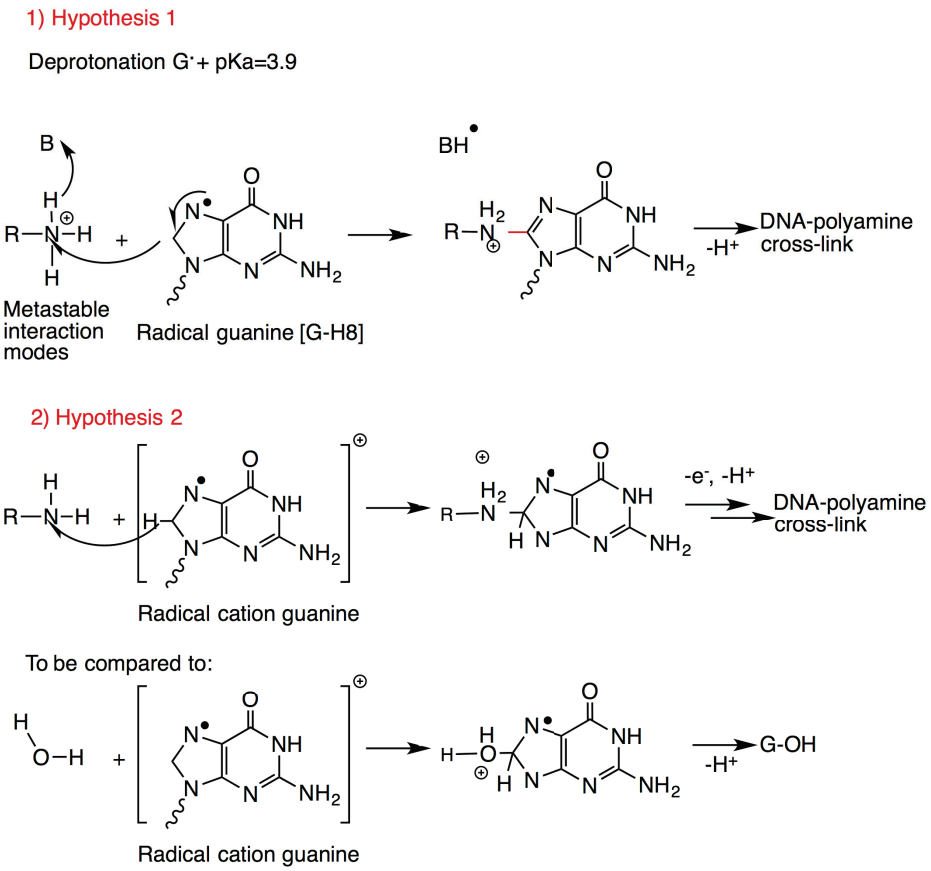


Figure 1: Two scenarii for the formation of guanine–polyamine cross-links for a generic amine, depending on the protonation state. The addition towards water is recalled.

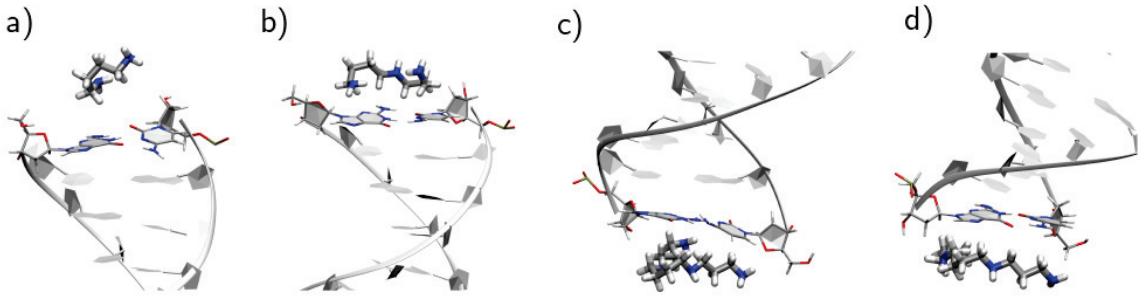


Figure 2: Representative structures of neutral a) putrescine, b) spermidine, c) spermine, and d) +1 charged spermine interacting with the extremities of the oligonucleotide. Such interactions appear during a) 13%, b) 41%, c) 49%, and d) 30% of the simulation time.

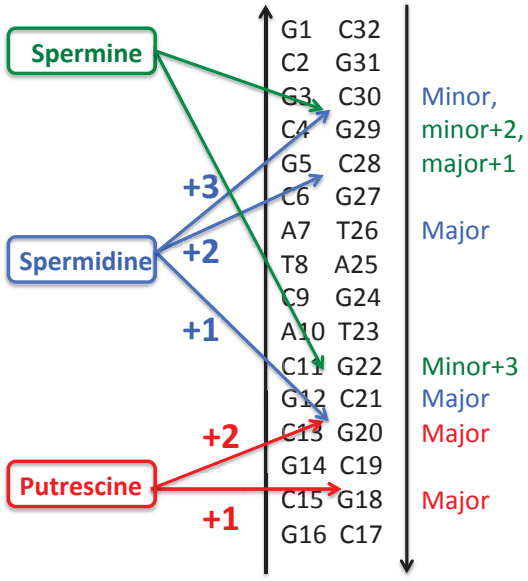


Figure 3: Scheme of the sites of interaction identified on the 16-bp duplex for the three polyamines..

.

Appendix H

Supporting Informations : Repair rate of clustered abasic DNA lesions by human endonuclease: molecular bases of sequence specificity.

Supporting Information for: Repair Rate of Clustered Abasic DNA Lesions by Human Endonuclease: Molecular Bases of Sequence Specificity

Hugo Gattuso[1,2], Elodie Durand[1,2], Emmanuelle Bignon[3,4], Christophe Morell[4], Alexandros G. Georgakilas[5,], Elise Dumont[3], Christophe Chipot[1,2,6], François Dehez[1,2], Antonio Monari[1,2]*

[1]Université de Lorraine -Nancy, Theory-Modeling-Simulation SRSMC, Vandoeuvre-lès-Nancy, France, [2]CNRS, Theory-Modeling-Simulation SRSMC, Vandoeuvre-lès-Nancy,[3]Univ Lyon, Ens de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F-69342, Lyon, France, [4]Université de Lyon, Institut des Sciences Analytiques UMR 5280, CNRS, Université de Lyon 1, ENS Lyon 5 rue de la Doua F-69100 Villeurbanne France, [5]Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou 15780, Athens, Greece [6] Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, USA

Computational details:

The structure of the APE1 protein bound to an AP containing DNA double strand was extracted from the X-ray crystal structure obtained by Mol et al.[1] (pdb databank code: 1DEW). It is composed of the 278 amino-acids and a 14 bases pairs B-DNA double strand 3'-(G)CGTCCXCGACGACG-5', with the first guanine being a dangling base without its facing partner and X the AP site. Clustered AP double strands were manually created from the starting sequence by removing one nucleobase from the complementary strand at the -3, -1, 0, +1 and +3 distance from the original AP site. Consequently, the sequences have been named seqX-APE1. Single, and clustered AP containing DNA strand were simulated in presence of the protein and as isolated duplexes in solution. The isolated APE1 in absence of any DNA strand was also simulated.

All the systems where placed in TIP3P water boxes[2] with a buffer of 7 Å, charges have been neutralized with Na+ or Cl- atoms (24 Na^+ ions for isolated DNA, 27 Na^+ ions APE1/DNA complexes, and 2 Cl^- for APE1 alone in water).

The AP site was parametrized using amber parmf99bsc0 parameters[3], atomic charges were obtained using the RESP procedure[4].

After an energy minimization of 1000 steps using the conjugate gradient algorithm, the system (excluding the water molecules and ions) was equilibrated by restraining it via a harmonic potential and slowly releasing the restraints through 4 distinct steps of 1 ns. After equilibration a production run of 100 ns has been performed for all the systems. This time scale appears sufficient to explore the conformational space of the DNA/APE1 complex in particular due to the strong interactions and high rigidity of the aggregates. As concerns the more dynamic solvated

DNA oligomers the high value of the RMSD is already witnessing important structural deformation reflecting the flexibility of the solvated oligomers. All molecular mechanics and molecular dynamics were conducted using the NAMD program[5]. The amber99 force field[6] with the recent bsc1 correction[7] for DNA backbone dihedral angles were used to model the systems.

The molecular dynamics simulations, integration time step of 2 fs in the NPT ensemble, were carried out using the Langevin dynamics and the Langevin piston[8] algorithms at a temperature of 300 K and a pressure of 1atm.

Long-range electrostatic interactions, represented by the Particle-Mesh-Ewald method[9] were truncated, as well as Van der Waals ones, for distances greater than 9 Å. DNA and DNA-APE1 rendered representations were obtained using VMD[10] and Chimera[11] rendering software. Moreover trajectory analysis where also performed using VMD.

The tweezers (composed of a methionine (MET) and an arginine (ARG)) edges distance, were obtained as the distance between the sulfur atom of MET and the γ-carbon atom of ARG (Figure 3). Moreover, the 10 central base pairs of the DNA double strands were taken into account for the evaluation of the RMSD. This is to avoid a deviation caused by the highly flexible external bases.

	Name	Type	Charge
1	P	P	1.1659
2	OP2	O2	-0.7761
3	OP1	O2	-0.7761
4	O5’	OS	-0.4954
5	C5’	CI	-0.0069
6	H5’	H1	0.0754
7	H5”	H1	0.0754
8	C4’	CT	0.2098
9	H4’	H1	0.0046
10	O4’	OS	-0.4058
11	C1’	CT	0.3679
12	O1	OH	-0.5959
13	H1	HO	0.3761
14	H1’	HC	0.02
15	C3’	CT	0.3259
11	H3’	HC	0.0562
12	C2’	CT	-0.286
13	H2’	HC	0.0941
14	H2”	HC	0.0941
15	O3’	OS	-0.5232

Table S1) Atomic charges (in a.u.) and type specification for the AP site.

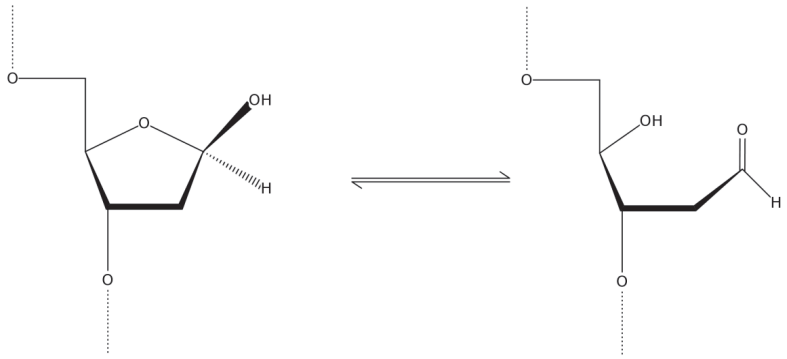


Figure S1) Equilibrium between the open and cyclic conformations of the AP lesion. Note that the equilibrium may produce the two R and S enantiomer. Here we represent only the R enantiomer that has been used in this study. Preliminary calculations have shown however the general global structural behavior of duplexes belonging R or S AP enantiomers.

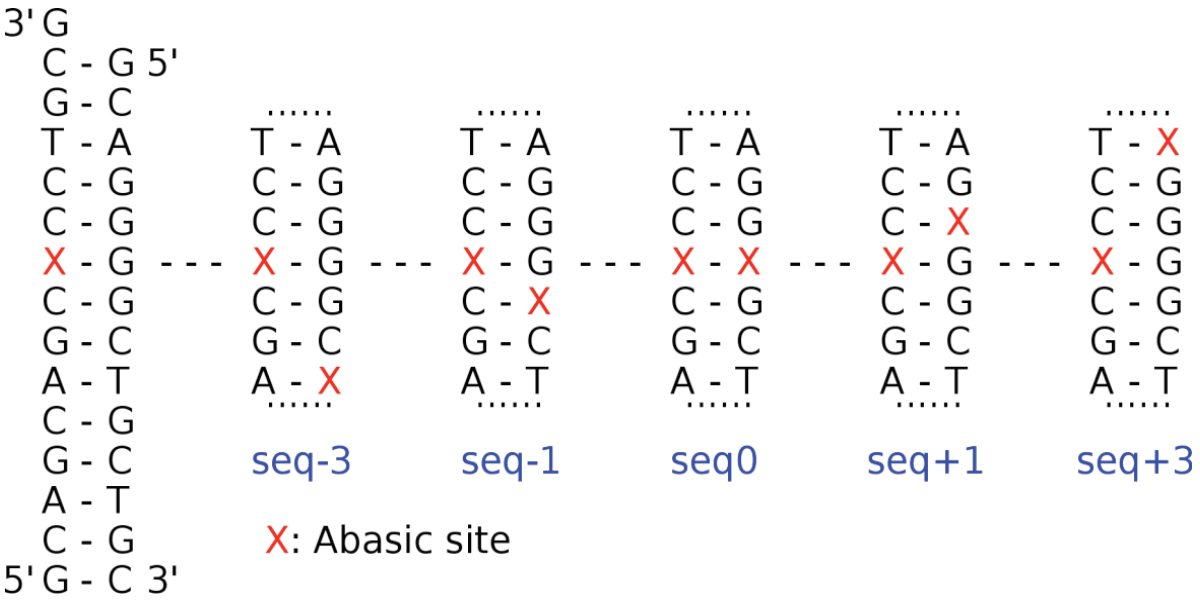


Figure S2) Representation of the investigated DNA sequences.

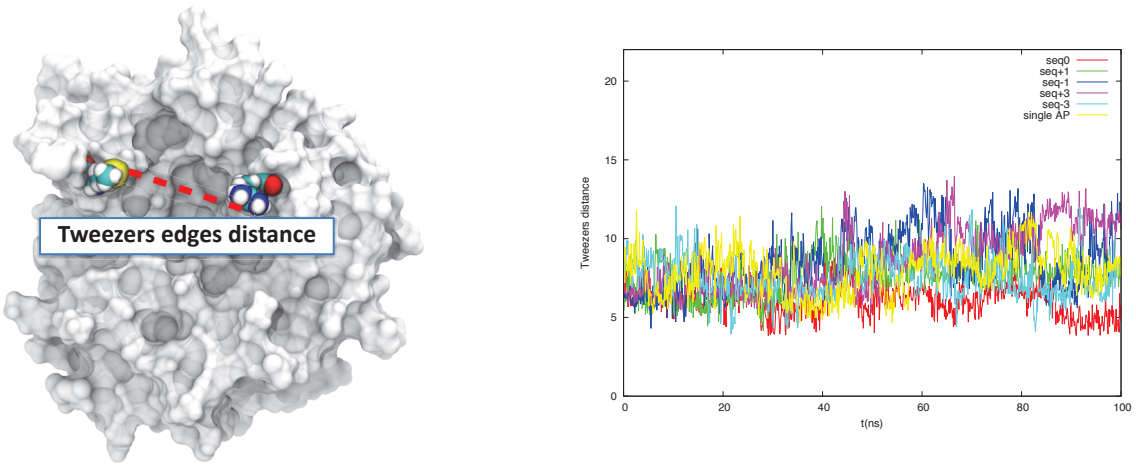


Figure S3) Definition of the tweezers' opening, taken as the distance between the sulfur atoms of the Methionine and the γ-Carbon atom of the Arginine and time series of the tweezers' edges distance for all the for all the DNA/APE1 complexes simulated in this work.

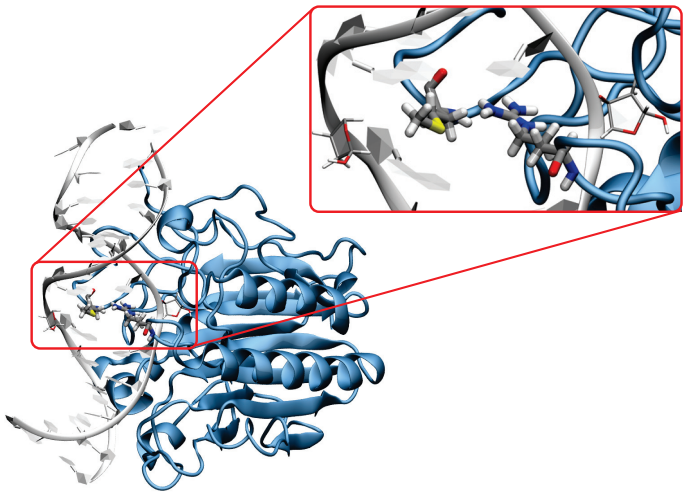


Figure S4) A representative snapshot of the DNA APE1 complex showing the tweezers and its edges.

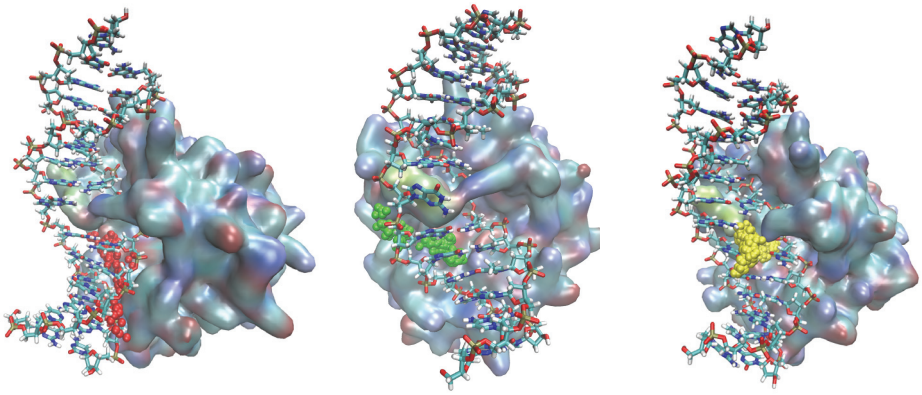


Figure S5) Visualization of the different entrance pathways of the Na^+ cations in the complex between AP containing DNA strand and APE1.

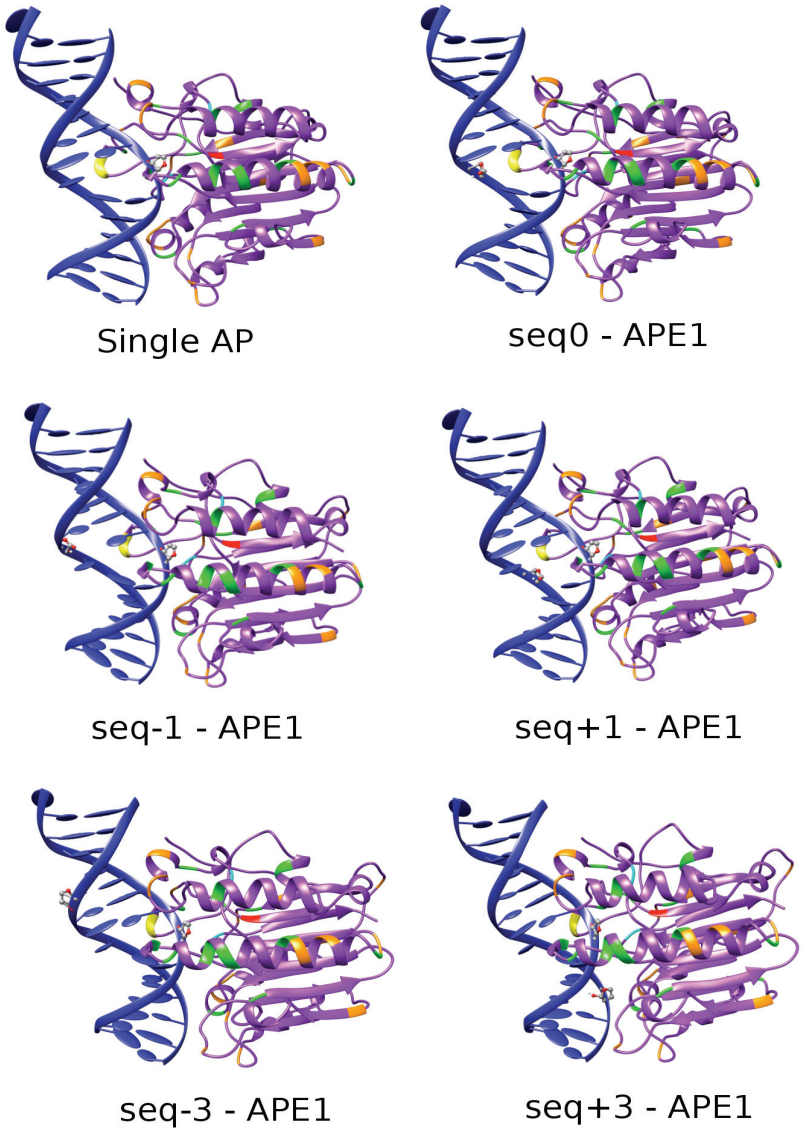


Figure S6) Representative snapshots of DNA/APE1 complexes. DNA and protein are represented in carton. AP sites are evidenced in cpk representation.

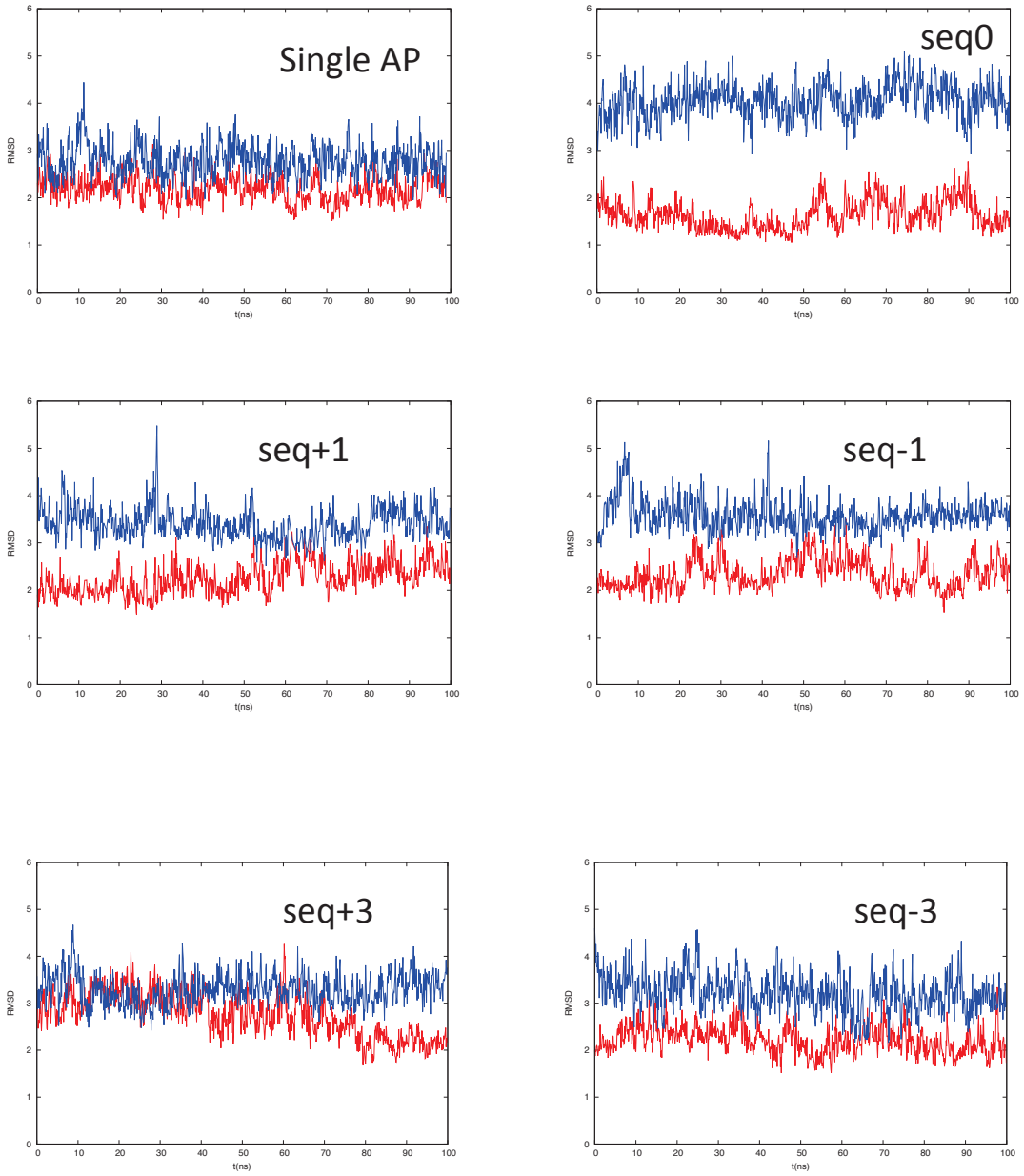


Figure S7) Time series of the Root mean Square Deviation (RMSD) in Å with respect to the crystal structure for the single containing and clustered AP lesions, inside the APE1 complex (red) and in solution (blue), see Figure 4 of the main text for the corresponding distribution.

References:

1. Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. A. DNA-Bound Structures and Mutants Reveal Abasic DNA Binding by APE1 and DNA Repair Coordination [Corrected]. *Nature* **2000**, *403* (6768), 451-456.

2. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein M. L. Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 1993, *79*, 926-935.

3. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

4. Bayly, C. I.; Piotr, C.; Cornell, W. D.; Kollman, P. A. Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

5. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comp. Chem.* **2005**, *26*, 1781-1802.

6. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

7. Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: A

Refined Force Field for DNA Simulations. *Nat. Methods,* **2015**, *13*, 55-58.

8. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

9. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

10. Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, **1996**, *14*, 33-38.

11. Pettersen E. F.; Goddard T. D.; Huang C. C.; Couch G. S.; Greenblatt D. M.; Meng E. C.; Ferrin T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25(13)*, 1605-1612.

Theoretical Approach of complex DNA lesions : from formation to repair.

Résumé : Ce travail de thèse vise à étudier l'endommagement de l'ADN, de la formation de lésions à leur réparation par des méthodes de modélisation moléculaire. Plusieurs projets ont pris forme dans ce contexte, lesquels peuvent être classés en trois grandes catégories. D'un côté, nous nous sommes intéressés la formation de lésions induites par des agents mutagènes. Nous avons étudié les mécanismes de formation de la 8-oxo-7,8-dihydroguanine (8oxoG), mais aussi le caractère de photosensibilisateur endogène de la pyrimidine 6-4 pyrimidone (6-4PP), et la photosensibilisation de l'ADN par deux anti-inflammatoires : le kétoprofène et l'ibuprofène. D'un autre côté, les propriétés mécaniques de l'ADN endommagé ont été simulées. La structure de lésions complexes est d'une importance capitale pour comprendre la manière dont elles sont réparées. Malheureusement, seulement peu de structures RMN et cristallographiques sont disponibles à ce jour. Pour pallier à ce manque et obtenir des informations sur leur dynamique, nous avons étudié un panel de lésions complexes : les clusters de sites abasiques, les pontages inter-brins, et la photolésion 6-4PP. De même, nous nous sommes penchés sur les modes d'interaction de certaines polyamines avec l'ADN, ces molécules étant connues pour interagir avec la double hélice. Enfin, la troisième partie de cette thèse concerne les interactions ADN-enzyme de réparation. En perspective avec l'étude de clusters d'abasiques, nous avons étudié le comportement dynamique du même système, cette fois-ci en interaction avec l'endonucléase APE1. Nous nous sommes également penchés sur les interactions entre la glycosylase Fpg avec un oligonucléotide contenant un tandem de lésions 8-oxoG d'un côté, et un cluster de lésions 8-oxoG - site abasique de l'autre. Ces multiples projets ont permis l'accumulation de nouvelles connaissances à propos des lésions complexes de l'ADN, et ont également apporté un appui computationnel aux expérimentations, qui peuvent se révéler très délicates dans ce domaine. Nos résultats ouvrent de larges perspectives dans le domaine de la pharmacologie, la cosmétique et plus généralement la compréhension du vivant.
Mots-clés: Chimie computationelle, biochimie, endommagement de l'ADN, dynamique moléculaire.

Abstract: This thesis work is focused on the theoretical modelling of DNA damages, from formation to repair. Several projects have been led in this framework, which can be sorted into three different parts. One on hand, we studied complex DNA reactivity. It included a study about 8-oxo-7,8-dihydroguanine (8oxoG) mechanisms of formation, a project concerning the UV-induced pyrimidine 6-4 pyrimidone (6-4PP) endogenous photosensitizer features, and an other one about DNA photosensitization by nonsteroidal anti-inflammatory drugs (ie ketoprofen and ibuprofen). On the other hand, we investigated mechanical properties of damaged DNA. The structural signature of a DNA lesion is of major importance for their repair, unfortunately only few NMR and X-ray structures of such systems are available. In order to gain insights into their dynamical structure, we investigated a series of complex damages : clustered abasic sites, interstrand cross-links, and the 6-4PP photolesion. Likewise, we studied the interaction modes DNA with several polyamines, which are well known to interact with the double helix, but also with the perspective to model DNA-protein cross-linking. The third part concerned the study of DNA interactions with repair enzymes. In line with the structural study about clustered abasic sites, we investigated the dynamics of the same system, but this time interacting with the APE1 endonuclease. We also studied interactions between the Fpg glycosylase with an oligonucleotides containing tandem 8-oxoG on one hand and 8-oxoG - abasic site as multiply damaged sites. Thus, we shed new lights on damaged DNA reactivity, structure and repair, which provides perspectives for biomedicine and life's mechanisms understanding as we begin to describe nucleosomal DNA.
Keywords: Computational chemistry, biochemistry, damaged DNA, molecular dynamics.

INSTITUT DES SCIENCES ANALYTIQUES
ENS DE LYON
UNIVERSITÉ DE LYON
Lyon 1