Towards a better understanding of the relationships between biomolecules interactions and their chemical properties via multiscale numerical simulations

Natacha Gillet

UNIVERSITÉ DE LYON

ENS DE LYON

HABILITATION À DIRIGER DES RECHERCHES

délivrée par

l'École Normale Supérieure de Lyon

Discipline : Chimie

Soutenue publiquement le 27 Novembre 2024, par :

Natacha Gillet

Towards a better understanding of the relationships between biomolecules interactions and their chemical properties via multiscale numerical simulations

---

Devant le jury composé de :

Dr. François Dehez	Directeur de Recherche	CNRS/Université de Lorraine
Dr. Adèle Laurent	Directrice de Recherche	CNRS/Nantes Université
Pr. Anne Milet	Professeure	Université Grenoble Alpes
Pr. Elise Dumont	Professeure	Université Côte d'Azur
Pr. Marcus Elstner	Professeur	Karlsruhe Institute of Technology
Dr Olivier Maury	Directeur de Recherche	CNRS/ENS de Lyon

Acknowledgements

Dix ans après la rédaction et la soutenance de ma thèse, cette habilitation à diriger des recherches représente une nouvelle occasion d'exprimer ma profonde reconnaissance à toutes les personnes qui m'ont accompagnée et m'accompagnent encore dans ma carrière de chercheuse, à celles que j'ai rencontrées depuis et m'ont accordé leur confiance, à toutes celles qui m'ont offert l'opportunité d'échanges riches et fertiles en nouvelles découvertes...

Je commencerais par remercier l'ensemble des membres de mon jury pour avoir accepté de lire et rapporter ce manuscrit. J'ai pu échanger avec plaisir avec chacun d'entre eux lors de différents évènements de la vie de chercheuse: en conférence, lors d'invitation, en comité de sélection... ou tout simplement parce que j'ai la chance de travailler avec eux.

De la chance, j'estime en avoir eu énormément tout au long de mon parcours, tout d'abord par le soutien incroyable de mes encadrants et directeur/directrice de thèse. Encore merci à Isabelle Demachy, Aurélien de la Lande, Vicent Moliner et Javier Ruiz Pernia pour leur confiance tout au long de ces trois années de doctorats merveilleuses, un peu anciennes maintenant... Ensuite, la chance a été d'intégrer l'équipe de Marcus Elstner que je remercie pour les innombrables opportunités qu'il m'a offerte, dans la liberté de mener à bien ma propre recherche, d'encadrer des stagiaires et doctorants, de m'impliquer dans la rédaction de projets, ainsi que pour sa confiance et la poursuite de nos collaborations. Bien que courte, ma deuxième expérience post-doctorale avec Annick Dejaegere et Roland Stote représente également pour moi une très belle rencontre scientifique et humaine.

Enfin, depuis 2019, j'ai eu la chance d'intégrer le Laboratoire de Chimie de l'ENS de Lyon. Je tiens à souligner l'immense rôle joué par Elise Dumont dans ce recrutement, son implication à chaque étape de l'élaboration du projet et du concours . Elle m'a ouvert les portes de ses collaborations, m'a aidé à m'intégrer, à comprendre un peu de politique de laboratoire... Je la remercie d'être venue me proposer de candidater lors des RCTF de Toulouse et surtout pour l'ensemble de nos interactions depuis qui ont fait qu'au delà d'une collègue de travail, elle est devenue une amie. Par ailleurs, je remercie les différents directeurs et directrices du Laboratoire de Chimie qui se sont succédés depuis mon arrivée, Chantal Andraud, Stéphane Parola et Carine Michel, pour m'avoir acceptée dans cette unité et me permettre de mener à bien mes recherches dans un cadre bienveillant (même dans les situations exceptionnelles...une pensée particulière pour mars 2020 et les trop nombreux mois de confinements et couvres-feux qui ont suivis). Je remercie aussi Tangui Le Bahers et Stephan Steinmman, formidables animateurs d'axes, pour tous nos échanges au quotidien, sur nos soucis de chercheurs et plus encore. Il y a de nombreuses autres personnes du laboratoire que j'aimerais remercier pour les collaborations, les discussions scientifiques ou moments partagés en salle café, et j'espère qu'elles se reconnaîtront. Un merci tout particulier à Christian Melkonian et les membres du pôle gestion pour votre aide au quotidien, et votre patience face à mes soucis avec l'administratif... Enfin, je souhaiterais remercier l'ensemble des étudiants, des doctorants et post-doctorants que j'ai eu la chance d'encadrer, de

conseiller et d'accompagner. Sans eux, ce manuscrit n'existerait tout simplement pas, il est le reflet de leur travail, de nos échanges scientifiques et de belles rencontres humaines. A tous et toutes, merci.

Enfin plus personnellement, je remercie les personnes qui m'accompagnent et me soutiennent, ou plus précisément nous supportent, moi, mes doutes et mes angoisses, depuis de nombreuses années (ou un peu moins)... Je pense à mes amies, mes frères, mes parents... et aussi à Sélina, qui même avant sa naissance a subi, au plus proche, mon stress de fin de thèse, et qui m'offre depuis dix ans tellement de belles découvertes.

Table of Contents

Natacha GILLET
Date of birth March, 2nd 1988
Professional address Laboratoire de Chimie
ENS Lyon
46, allée d'Italie
69364 LYON CEDEX 7, FRANCE
Phone number +334 72 72 81 44
E-mail natacha.gillet@ens-lyon.fr

Keywords: Structure-function relationship in biomolecules, enzymatic process, charge transfers in biomolecules, classical and multiscales (QM/MM) molecular dynamics.
Communications: 28 articles, 2 book chapters, 17 oral communications in conferences
Other experience: Master (9) and PhD (5) supervision, teaching

RESEARCH EXPERIENCE

CNRS Position

October 2019-now Laboratoire de Chimie ENS Lyon.

Post-Doctoral Positions

January 2019-September 2019 Post-doctoral position (ARC Fellow) Institut de Génétique et de Biologie Moléculaire et cellulaire (Strasbourg, France) *Understand the link between mutation and structural dynamics in th PPARγ/RXRα nuclear receptor complex*

June 2015-December 2018 Post-doctoral position (2 years with Humboldt fellowship)– Institute of Physical Chemistry Karlsruhe Institute of Technology, (Karlsruhe, Germany) *Electron transfers in Protein and QM/MM simulations*.

Thesis

October 2011 – July 2014 Thesis – Laboratoire de Chimie Physique Université Paris Sud (Orsay, France) and Departamento de Quimica Fisica y Analitica Universitat Jaume I (Castellon, Spain): *Numerical Simulations of Intertwined Electrons and Protons Transfers in Proteins*

DIPLOMAS

2014 PhD in Theoretical Chemistry - Laboratoire de Chimie Physique Université Paris Sud (Orsay, France) et Departamento de Quimica Fisica y Analitica Universitat Jaume I (Castellon, Spain)

2011 Master degree in Molecular Chemistry – UPMC, ENS (Paris) –
ENS diploma – ENS (Paris)

2009 Bachelor's degree in Chemistry – ENS (Paris) –

GRANTS AWARDS AND TROPHIES

2023	PCSI Proof of concept BERNUMOL (PI with H. Menoni) 60 k€
2020	ANR JCJC NucleoMAP ~160 K€
2019	ARC Retour en France Post-doctoral Fellowship
2016	Alexander von Humboldt Foundation Fellowship
2013	French l'Oréal-UNESCO *for women in science* fellowship

OTHER EXPERIENCES

Supervision

March 2024-September 2024	Supervision of a master 2 student on charge transfer in nucleosomal DNA
March 2024-April 2024	Supervision of a bachelor student on protein-DNA interaction
January 2024-July 2024	Supervision of a master 2 student on PCSI BERNUMOL project
January 2024-March 2024	Supervision of a master student on machine learning for ionization potential calculation
October 2023-now	Co-supervision of a PhD student on Nucleosome damages simulation with Dr. A. de la Lande (Université Paris Saclay)
September 2023-February 2024	Supervision of a master student in the IISER-ENS partnership framework on protein-crystallophore interaction
May 2023-September 2023	Supervision of a master 2 student on protein-crystallophore interaction
October 2022-now	Co-supervision of a PhD student on PCET in proteins with Pr. M. Elstner (KIT, Germany) and Dr. C. Michel (LCh, Lyon)
October 2021-now	Co-supervision of a PhD student in the ANR NucleoMAP context with Pr. E. Dumont (LCh, Lyon)
June 2021-November 2021	Supervision of an internship student (BA) on charge transfer in cyrochromeP-CPR complex
June 2021	Supervision of an internship student (BA) on protein-crystallophore interaction
January 2021-June 2021	Supervision of master 2 student in the ANR NucleoMAP context
October 2020-now	Co-supervision of a PhD student on nucleosome simulations and oxidative damages with Pr. E. Dumont (LCh, Lyon)
May 2020-April 2021	Supervision of a master student in the IISER-ENS partnership framework: charge transfer in G-quadruplexes
October 2019-January 2021	Co-supervision of a Post-Doc on protein-molecular glue interaction with Pr. E. Dumont (LCh, Lyon)
March 2018-October 2018	Supervision of a master student on PCET in RNR
October 2017-April 2018	Co-supervision of a master student on excitons transfers in proteins
May 2017-July 2017	Supervision of an internship student (BA) on charge transfer in Azurins
July 2016 – September 2016	Supervision of an internship student (BA) Computational study of the impacts of mutations on charge transfer in PhrA.
January 2016-December 2018	Supervision of a master-PhD student, charge transfer and protonation mechanisms in cryptochrome and photolyases

June 2014 – July 2014	Supervision of an internship student (BA) on Flavocytochrome b2 reactivity

Teaching

2022-2023	ENS de Lyon/Université de Dijon: Lecture and Practicals in computational Chemistry (Master 2)
February March 2023	ENS de Lyon: Lecture and Practicals in biochemistry (L3)
March 2022	ENS de Lyon: Practicals in computational biochemistry (L3)
2011-2013	Université Paris Sud: - lectures and tutorials in atomistic (L1) - practicals in solution chemistry (L2) - tutorials in inorganic chemistry (Master 1) - Practicals in statistical thermodynamics (Master 1) - Practicals in computational biochemistry (Master 2)

Administrative tasks

2023-Now	GENCI CT7 expert
2021-Now	Animation of the Chemistry for Life axis
2021-Now	Equality correspondent at LCh ENS de Lyon
2021-Now	Elected member at the Laboratory Council

RESEARCH COMMUNICATIONS

Publications

What tunes guanines ionization potential in a nucleosome? An all-in-one systematic QM/MM assessment
Kermarrec, Maxime; Dumont, Elise; Gillet, Natacha*; *Biophys J.*, (Accepted)

Influence of Chemical Modifications of the Crystallophore on Protein Nucleating Properties and Supramolecular Interactions Network.
Roux, Amandine; Alsalman, Zaynab; Jiang, Tao; Mulatier, Jean.-Christophe; Pitrat, Delphine; Dumont, Elise; Riobé, François; Gillet, Natacha; Girard, Eric; Maury, Olivier *Chemistry – A European Journal*, **30**: e202400900 (2024)

One touch is all it takes: the supramolecular interaction between ubiquitin and lanthanide complexes revisited by paramagnetic NMR and molecular dynamics.
Dos Santos, Karen; Bartocci, Alessio; Gillet, Natacha; Denis-Quanquin, Sandrine; Roux, Amandine; Lin, Eugene; Xu, Zeren; Finizola, Raphael; Chedozeau, Pauline; Chen, Xi; Caradeuc, Cédric; Baudin, Mathieu; Bertho, Gildas; Riobé, François; Maury, Olivier; Dumont, Elise;Giraud, Nicolas *Physical Chemistry Chemical Physics*, **26**: 14573 (2024)

DNA damage and repair in the nucleosome: insights from computational methods.
Gillet, Natacha; Dumont, Elise; Bignon, Emmanuelle, *Biophysical Reviews* , **16**: 265 (2024)

DNA-Histone Cross-link Formation Via Hole Trapping in Nucleosome Core Particles.
Wen, Tingyu; Kermarrec, Maxime; Dumont, Elise; Gillet, Natacha*; Greenberg, Marc* *JACS*, **145:** *237* (2023)

A new G-quadruplex-specific photosensitizer inducing genome instability in cancer cells by triggering oxidative DNA damage and impeding replication fork progression
Deiana, Marco*; Andrés Castán, José María; Josse, Pierre; Kahsay, Abraha; Sánchez, Darío Puchán; Morice, Korentin; Gillet, Natacha; Ravindranath, Ranjitha; Patel, Ankit Kumar; Sengupta, Pallabi; Obi, Ikenna; Rodriguez-Marquez, Eva; Khrouz, Lhoussain; Dumont, Elise; Abad Galán, Laura; Allain, Magali; Walker, Bright; Ahn, Hyun Seo; Maury, Olivier; Blanchard, Philippe; Le Bahers, Tangui; Öhlund, Daniel; von Hofsten, Jonas; Monnereau, Cyrille*; Cabanetos, Clément*; Sabouri, Nasim* *Nucleic Acid Research* **51**: 6264-6285 (2023)
Dynamics and energetics of PCBP1 binding to severely oxidized RNA
Gillet, Natacha*; Dumont, Elise *Frontier in Molecular Bioscience* **9**: 994915 (2022)
Radical cation transfer in a guanine pair: An insight to the G-quadruplex structure role using constrained DFT/MM
R, Ranjitha; Mondal, Padmabati; Gillet Natacha* *Theoretical Chemistry Accounts* **140:** 1-8 (2021)
A Dynamic View of the Interaction of Histone Tails with Clustered Abasic Sites in a Nucleosome Core Particle
Bignon, Emmanuelle; Gillet, Natacha; Jiang, Tao; Morell, Christophe; Dumont, Elise* *Journal of Physical Chemistry Letters* **12:** 6014-6019 (2021)
Assessing the sequence dependence of pyrimidine–pyrimidone (6–4) photoproduct in a duplex double-stranded DNA: A pitfall for microsecond range simulation
Gillet, Natacha*; Bartocci, Alessio; Dumont, Elise* *Journal of Chemical Physics* **154**:135103 (2021)
Recognition of a tandem lesion by DNA bacterial formamidopyrimidine glycosylases explored combining molecular dynamics and machine learning
Bignon, Emmanuelle*; Gillet, Natacha; Chan, Chen-Hui; Jiang, Tao; Monari, Antonio; Dumont, Elise *Computational and structural biotechnology journal* **19**: 2861-2869 (2021)
Molecular Dynamics Approach for Capturing Calixarene–Protein Interactions: The Case of Cytochrome C
Alessio Bartocci, Natacha Gillet, Tao Jiang, Florence Szczepaniak, and Elise Dumont* *J. Phys. Chem. B* **124:** 11371 (2020)
Nucleosomal embedding reshapes the dynamics of abasic sites
Emmanuelle Bignon,* Victor E. P. Claerbout, Tao Jiang, Christophe Morell, Natacha Gillet & Elise Dumont* *Scientific Reports* **10**: 17314 (2020)
Impact of the Nucleosome Histone Core on the Structure and Dynamics of DNA-Containing Pyrimidine–Pyrimidone (6–4) Photoproduct
Eva Matoušková, Emmanuelle Bignon, Victor E. P. Claerbout, Tomáš Dršata, Natacha Gillet, Antonio Monari*, Elise Dumont*, and Filip Lankaš*
J. Chem. Theory Comput., **16**: 5972 (2020)
Biological Relevance of Charge Transfer Branching Pathways in Photolyases
Holub, Daniel; Lamparter, Tilman; Elstner, Marcus; Gillet, Natacha* *PCCP* **21**: 17072 (2019).
What accounts for the different functions in Photolyases and Cryptochromes: a computational study of proton transfers to FAD
Holub, Daniel; Kubař, Tomáš; Mast, Thilo ; Elstner, Marcus ; Gillet, Natacha* *PCCP* **21**: 11956 (2019).
Two aspartate residues close to the lesion binding site of Agrobacterium (6-4) photolyase are required for Mg2+ stimulation of DNA repair
Ma, Hongju; Holub, Daniel; Gillet, Natacha; Kaeser, Gero; Thoulass, Katharina; Elstner, Marcus; Krauß, Norbert; Lamparter, Tilman*, *FEBS Journal* **286**: 1765 (2019).
Coupled-perturbed DFTB-QM/MM metadynamics: Application to proton-coupled electron transfer

Gillet Natacha, Marcus Elstner et Tomáš Kubař * *J. Chem. Phys* **149**: 072328 (2018)

Functional Role of an Unusual Tyrosine Residue in the Electron Transfer Chain of a Prokaryotic (6-4) Photolyases

Holub, Daniel; Ma, Hongju; Krauß, Norbert; Lamparter, Tilman ; Elstner, Marcus; Gillet, Natacha* *Chemical Science* **9**: 1259-1272 (2018)

Theoretical Estimation of Redox Potential of Biological Quinone Cofactors

Gillet, Natacha*; Lévy, Bernard; Moliner, Vicent; Demachy, Isabelle; de la Lande, Aurélien *J. Comput. Chem.* **38**:1612-1621 (2017)

Electronic Coupling Calculations for Bridge-Mediated Charge Transfer Using CDFT and Effective Hamiltonian approaches at DFT and FODFTB level

Gillet, Natacha; Berstis, Laura*; Wu, Xiaojing; Gajdos, Fruzsina; Heck, Alexander; de la Lande, Aurélien*; Blumberger, Jochen*; Elstner, Marcus* *J. Chem. Theory Comput.* **12**:4793-4805 (2016).

QM/MM study of L-lactate oxidation by Flavocytochrome *b2*

Gillet, Natacha*; Ruiz-Pernia, Javier; Demachy, Isabelle; Lévy, Bernard; Lederer Florence; de la Lande, Aurélien; Moliner, Vicent* *Phys Chem Chem Phys* **18**:15609 - 15618 (2016)

Progress and Challenges in Simulating and Understanding Electron Transfer in Proteins

de la Lande, Aurélien*; Gillet, Natacha; Chen Shufeng; Salahub Denis R.* *Arch Biochem Biophys.* **582**:28-41 (2015).

Electron transfer, decoherence and protein dynamics. Insights from atomistic simulations

Narth, Christophe; Gillet, Natacha; Cailliez, Fabien; Lévy, Bernard; de la Lande, Aurélien* *Acc. Chem. Res.* **48**:1090-1097 (2015).

Entasis Through Hook-and-Loop Fastening in a Glycoligand with Cumulative Weak Forces Stabilizing CuI

Garcia, Ludivine; Cisnetti, Federico; Gillet, Natacha; Guillot, Regis; Aumont-Nicaise, Magali; Piquemal, Jean-Philip; Desmadril, Michel; Lambert, François; Policar, Clotilde* *JACS* **137**:1141-1146 (2015).

Electron and hydrogen atom transfers in the hydride carrier protein EmoB

Gillet, Natacha*; Lévy, Bernard; Moliner, Vicent; Demachy, Isabelle; de la Lande, Aurélien* *J. Chem. Theory Comput.* **10**:5036–5046 (2014).

Investigation of the Molecular Mechanism of Electronic Decoherence within a Quinone Cofactor.

Narth, Christophe; Gillet, Natacha; Lévy, Bernard; Demachy, Isabelle; de la Lande, Aurélien* *Canadian Journal of Chemistry* **91**:628-636 (2013).

Coupling Quantum Interpretative Techniques: Another Look at Chemical Mechanisms in Organic Reactions.

Gillet, Natacha; Chaudret, Robin*; Contreras-García, Julia; Yang, Weitao; Silvi, Bernard; Piquemal, Jean-Philip* *J. Chem. Theory Comput.* **8**:3993-3997 (2012)

Book Chapter

Influence of DNA Structure on Lesion Formation and Repair: Role of Modelling and Simulations

Gillet, Natacha; Bignon, Emmanuelle ; Dumont, Elise ; Monari, Antonio**.** ; Roberto Improta, Thierry Douki (eds) DNA Photodamage: From Light Absorption to Cellular Responses and Skin Cancer, 2021, 105-132, RSC

DNA Photodamage and Repair: Computational Photobiology in Action

Antonio Francés-Monerris, Natacha Gillet, Elise Dumont, Antonio Monari* In: Andruniów T., Olivucci M. (eds) QM/MM Studies of Light-responsive Biological Systems. Challenges and Advances in Computational Chemistry and Physics, vol 31. Springer, Cham.

Oral communication in congress (*: presented by)

A new G-quadruplex specific photosensitizer inducing genome instability in cancer cells for photodynamics

Natacha Gillet*, Olivier Maury, Tangui le Bahers, Marco Deiana, Clément Cabanetos, Nasim Sabouri, Cyrille Monnereau Journée scientifique du médicament Grenoble, France (July 2023)

Mapping Guanine Oxidation in Nucleosomal DNA using Multiscale Simulations

Gillet Natacha*; Kermarrec Maxime; Dumont Elise *ABC2023*, Ascona, Switzerland (April 2023)

How crystallophores enhance protein interactions for nucleation: some clues from all-atom molecular dynamics simulations

Gillet Natacha *; Maury, Olivier, Dumont, Elise *Journée Scientifiques CenTRA 2023*, Palaiseau, France (March 2023)

Mapping Guanine Oxidation in Nucleosomal DNA using Multiscale Simulations

Gillet Natacha*; Allahkaram Laleh; Kermarrec Maxime; Dumont Elise *ISQBP President's meeting 2022*, Innsbruck, Austria (July 2022)

Mapping Guanine Oxidation in Nucleosomal DNA using Multiscale Simulations

Gillet Natacha*; Allahkaram Laleh; Kermarrec Maxime; Dumont Elise *ESPA*, Vigo, Spain (June 2022)

Understanding Charge transfer in Cryptochromes and Photolyases via QM/MM studies: Application to PhrB protein

Gillet, Natacha*; Holub, Daniel; Lüdemann Gesa; Elstner Marcus *ISQBP President's Meeting* Bergen, Norway (June 2016).

QM/MM study of L-lactate oxidation by Flavocytochrome *b2*

Gillet, Natacha*; Ruiz-Pernia, Javier; Demachy, Isabelle; Lévy, Bernard; Lederer Florence; de la Lande, Aurélien; Moliner, Vicent *Theory Workshop 'Computer Simulation and Theory of Macromolecules'*, Huenfeld Monastery, Germany (May 2016)

Computational Chemistry Tools for Biological Electron Transfers

Gillet, Natacha*; Elstner, Marcus *Network Meeting of the Alexander von Humboldt Foundation*¸ Düsseldorf Germany (March 2016)

Description d'un transfert d'hydrure en transferts d'électron et d'hydrogène dans la protéine EmoB

Gillet, Natacha*; Demachy, Isabelle; Moliner, Vicent; Lévy, Bernard; de la Lande, Aurélien *RCTF 2014*, Paris France (July 2014)

Electron and Hydrogen Atom Transfers in the Hydride Carrier Protein EmoB

Gillet, Natacha*; Demachy, Isabelle; Moliner, Vicent; Lévy, Bernard; de la Lande, Aurélien *iCHAT 2014*, Monteporzio Catone Italy (June 2014)

Mechanism of hydride transfer between flavins.

Gillet, Natacha*; Demachy, Isabelle; Moliner, Vicent; Lévy, Bernard; Piquemal, Jean-Philip; de la Lande, Aurélien *Theobio 2013*, Götheborg Sweden (June 2013)

Invited contribution

Non-covalent interactions in protein nucleation mechanism assisted by a molecular glue

Gillet, Natacha* Rencontre GDR Sigma Hole, Rennes, France (September 2024)

DNA-Histone cross-link formation via Hole Trapping in Nucleosome core particle

Gillet, Natacha* Moths, Pont-à-Moussons, France (May 2024)

Damaged DNA from Double Strands to Nucleosome using Molecular Dynamics simulations

Gillet Natacha* VIII SeedMol Online (November-December 2020)

Recent development of DFTB/MM for transfer mechanisms in proteins
Gillet, Natacha*, 19th DeMon developers Workshop, Fréjus, France (May 2019)
Développements Récents de Méthodes DFTB/MM pour l'Etude de Différents Transfert dans des Protéines
Gillet, Natacha* ; Holub, Daniel ; Bold, Beatrix ; Elstner, Marcus *RCTF 2018*, Toulouse, France (October 2018)
Functional Role of different Tyrosine Residues of a Prokaryotic (6-4) Photolyase: Ping-Pong between Theory and Experiment
Gillet Natacha*; Holub, Daniel; Elstner, Marcus *Theobio*, San Sebastian, Spain (June 2017)

Seminar
Mapping Guanine Oxidation in Nucleosomal DNA using Multiscale Simulations
Gillet Natacha*; Kermarrec Maxime; Dumont Elise Nancy, France (April 2023)
Mapping Guanine Oxidation in Nucleosomal DNA using Multiscale Simulations
Gillet Natacha*; Kermarrec Maxime; Dumont Elise Nantes, France (March 2023)
How Crystallophores enhance proteins interactions for nucleation: some clues from all-atoms molecular dynamics simulations
Gillet Natacha* Dijon, France (January 2023)
Mapping DNA oxidation hot spots in nucleosome using all atoms simulations: insight to the role of histone tails
Gillet Natacha* Physical Genomics meeting *ENS de Lyon* Lyon, France (April 2022)
Damaged DNA A combinatorial Chemistry
Gillet Natacha* Physical Genomics meeting *ENS de Lyon* Lyon, France (October 2020)
Evolution in the cryptochrome-photolyase family: a computational point of view
Gillet, Natacha*; Holub, Daniel; Elstner Marcus, Valencia, Spain (January 2020)
Evolution in the cryptochrome-photolyase family: a computational point of view
Gillet, Natacha*; Holub, Daniel; Elstner Marcus *IBPC*, Paris, France (February 2019)
Charge transfer simulations in Photolyases and RNR using DFTB/MM calculations
Gillet, Natacha*; Holub, Daniel; Elstner Marcus *Laboratoire de Chimie* ENS LYON, Lyon, France (October 2018)
Charge transfer simulations in Photolyases and RNR using DFTB/MM calculations
Gillet, Natacha*; Holub, Daniel; Elstner Marcus *Department of Chemistry* Boston University, Boston USA (August 2018)
Effects of protein environment on charge transfers: QM/MM studies
Gillet, Natacha*; Demachy, Isabelle; Moliner, Vicent; Lévy, Bernard; de la Lande, Aurélien *Department of Chemistry* University of Basel Switzerland (October 2014)
Hydride transfer in EmoB and Fcb2 Proteins: Two different QMMM studies
Gillet, Natacha*; Demachy, Isabelle; Moliner, Vicent; Lederer, Florence; Lévy, Bernard; Ruiz Pernia, Javier; de la Lande, Aurélien *Department of Physics* Freie Universität Berlin Germany (January 2014)

Chapter 1

General Introduction

1.1 Non-covalent interactions in biology

Protein and nucleic acids consist in monomer assembly, covalently bounded by their backbone part. Each monomer differs by its lateral chain or nucleobase respectively. Beyond this primary structure or sequence, the three dimensional aspect is obviously controlled and tuned by a framework of non-covalent bonds: hydrogen bond, crucial for the Watson-Crick structure of double strand DNA and at the heart of genetic information storage or for the formation of α-helices or β-sheet in protein, ionic bridges, π-stacking between nucleobases or aromatic residues, hydrophobic interaction...(Figure 1.1) The structure-function relationship of biomolecule underlines the prominent role of non-covalent interactions in biology. They also trigger the interactions between biomolecules, such as between an enzyme and its substrate, in protein-protein or protein-DNA complexes or in the receptor-ligand recognition and insure biological specificity. In addition to the structural role of these interactions, their dynamical fluctuations must also be considered to understand many biological processes: for instance, a regulator binding can affect the active or inactive state of a protein by a propagation of subtle conformational modifications along several Å by the so-called allosteric effect.

Moreover, the non-covalent interactions are strongly related to any chemical change in the structures: they create a specific environment for the biomolecule to fulfill its role, by playing on the electrostatic field, the steric hindrance or the stabilization of a given chemical assembly. On the same way, they can modulate chemical mechanism by modifying the close environment, for instance by triggering a charge transfer through biomolecules by favouring a tunneling transfer and tuning the redox properties of the cofactors.

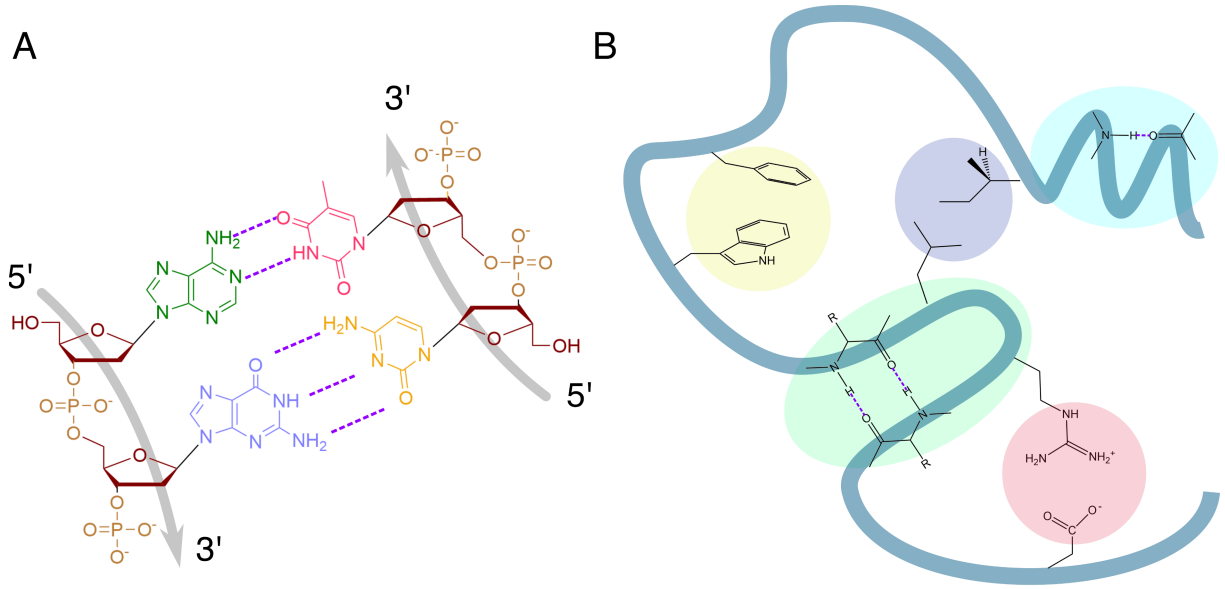

Figure 1.1: **A**. Structure of DNA: A(green)-T(pink) and G(blue)-C(yellow) base pairs in Watson-Crick hydrogen bonds (purple), sugar (red) and phosphate (ochre). **B**. Schematic representation of structural non-covalent interactions: hydrogen bonds involving the protein backbone in α helix (cyan) or β sheet (green), π-stacking (yellow), apolar interactions (grey) and ionic bridges (red).

1.2 Nucleosome: a wonderful toy

In the nucleus of eukaryotic cells, DNA is compacted in the chromatin which state regulates the replication, transcription, repair etc. of the genetic material.[1,2] The primary unit of the chromatin is the nucleosome (Figure 1.2): a fragment of double strand DNA wrapped around a protein cylindrical core of 8 histones, two of each kind, namely H3, H4, H2B and H2A, the free linker DNA and possibly the linker histone H1.[3,4] The nucleosome core particle (NCP) is a nucleosome restricted to the histones core and the 146-147 base pairs wrapped around. Each nucleobase can be situated according to the central position of the DNA strand, called dyad, according to a superhelicoidal coordinate. The wrapping supposes both a mechanical constrain on the B-DNA bend angle and a heterogeneous environment around the nucleobases. Indeed, the histone core obviously decreases the solvent accessibility of the DNA for the inward part (close to the histone core, opposite "outward" solvated nucleobases), but also creates a heterogeneous electrostatic field. Notably, its highly positively charged surface is able to attach the DNA backbone thanks to ionic interactions. Besides, each histone has a flexible N-terminal tail, which is intrinsically disordered and also rich in positively charged amino-acids arginines and lysines (the C-terminal tail of H2A can also be considered in these flexible parts of histones). Moreover, numerous residues of these tails are prone to post-translational modification (PTM) that can play on their total charge, structure and flexibility.[5–7] Due to these properties, histone tails are able to form transient interactions with DNA[8] and insure communication between two

nucleosomes,[9,10] mediate chromatin structure and the nucleosome breathing,[11–17] and within DNA bounded protein.[18–20]

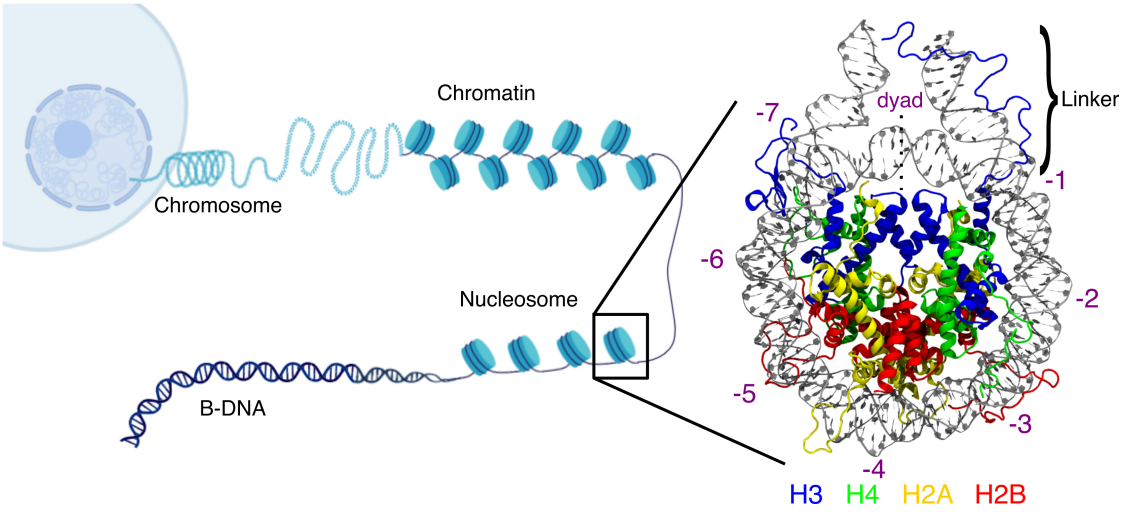

Figure 1.2: Schematic representation of DNA compaction in eukaryotic cells (created with Biorender.com) with a focus on the nucleosome structure from simulations starting with modified 5NL0 PDB structure:[3] DNA (grey), histone H3 (blue), H4 (green), H2A (yellow), H2B (red), superhelicoidal position (purple numbers, dyad is 0).

Because of its complexity, the nucleosome has been less studied than free DNA experimentally or computationally. For two decades, the improvement of experimental devices and protocol followed by the increasing computational power has open the door to nucleosome studies "which are coming of age".[21,22] Even though the number of available DNA sequences that are able to from nucleosome *in vitro* is limited, crystrallographic and cryoEM structure of NCP or one or several nucleosomes, including damaged DNA sequences, different histone variants, or other proteins partners have been deposited on the Protein Data Bank. In addition to NMR[23,24] or FRET experiments,[25] computational chemistry can figure out the dynamical aspect on these structures, and provide a molecular insight to the nucleosome properties, as exemplify by the dramatic increase in nucleosome microsecond timescale all-atom simulations since 2015.

Consequently, dealing with nucleosome or NCP to study the chemical properties of DNA means taking into a large variety of parameters such as sequence, superhelicoidal position, solvent accessibility, interactions with histone core and histone tails etc. The dynamical aspect of some of these parameters can be of paramount importance in nucleosome chemistry.

1.3 Multiscales molecular dynamics simulations

All-atom molecular dynamics simulations (MDs) represent now a compulsory part of most of biochemical studies:[26] they have gain quality of structural and dynamical description of biomolecules and even in prediction.[27–29] Computational chemists can notice the increasing number

of chemists and structural biologists who ask for simulations to improve the significance of their experimental studies. In that respect, classical MDs remain the most common tool: based on the representation of atoms by charged and rigid balls linked by strings representing the covalent bonds, they use a combination of parameters, called force fields, to calculate intramolecular and intermolecular energies using classical physics (harmonic potential, Lennard-Jones potential, Coulomb electrostatic model); these energy terms will be derived to get forces and propagate the atoms motions according to the second Newton's law. The possibility to easily reach micro-second simulation timescale in combination with the increasing quality of force fields make these approaches ideal for a large conformational sampling of biomolecules and of their non-covalent interaction landscape between each other. A current drawback can be the accumulation of data and their management.[30] Besides, the recent development of machine learning algorithms dedicated to improve the force fields, MD power or trajectories analyses, is pushing the boundaries of the field.[31–36]

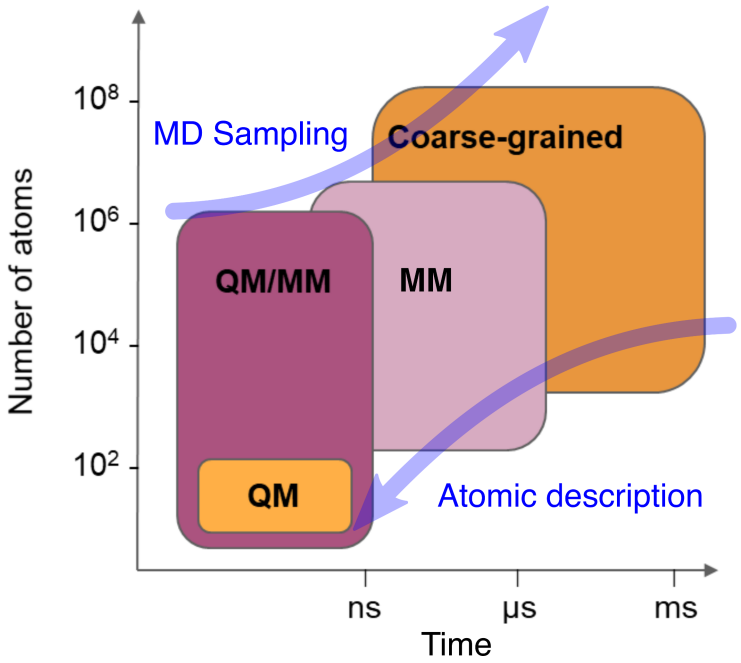

Figure 1.3: Different computational chemistry approaches for biomolecular MD simulations classified according to the affordable number of atoms and simulation timescale.

However, the classical level of theory is not able to describe "chemistry", that is the electron behaviour at the origin of bond formation and breaking, excitation, electron and proton transfers... To do so, a quantum description of the electronic density is mandatory but rely on methods that are much more expensive than classical MD (Figure 1.3). A solution I have adopted since my PhD work consists in multiscale QM/MM approaches where only a part of the system is treated at a quantum level (the QM part where the chemical reactivity takes place) while its environment is treated classically (the MM part). The chosen QM level remains the limiting aspect of such methods with respect to the simulation time scale, but the reachable size of the

system is equivalent to pure classical systems. To circumvent the time issue, one can use bias approaches or, when the problematic allows it, like for redox properties, trust the classical force field and perform QM/MM calculations on a large amount of classical geometries. In my point of view, the choice of methodology must fit the best balance between accuracy and conformational sampling so adapted to each system and each problematic.

1.4 Objectives

After this short introduction, I hope that I have convinced the readers of the importance of a description of biochemical properties including the structural and dynamical aspects of biological macromolecules and their non-covalent interactions landscape. In the following, I will describe how I try to apply computational chemistry to decipher the chemical properties induced by these interactions for three main issues:

- the relationship between DNA dynamics and DNA lesions (UV-photoproduct or oxidative damages) at different stages: before and after formation of a damage and for recognition by repair proteins. Both free B-DNA and nucleosomal DNA are considered.
- the charge transfer properties of nucleosomal DNA, with a focus on guanines and the role of the histones tails.
- the interaction between protein and a lanthanide complex playing the role of a molecular glue for crystallization process.

For each topic, I will present the key points of the methodology and the main results obtained in my group and/or in a collaborative research team.

Bibliography

[1] Bickmore, W. A. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet* **2013**, *14*, 67–84.

[2] Hubner, M. R.; Eckersley-Maslin, M. A.; Spector, D. L. Chromatin organization and transcriptional regulation. *Curr Opin Genet Dev* **2013**, *23*, 89–95.

[3] Bednar, J. et al. Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell* **2017**, *66*, 384–397 e8.

[4] Garcia-Saez, I.; Menoni, H.; Boopathi, R.; Shukla, M. S.; Soueidan, L.; Noirclerc-Savoye, M.; Le Roy, A.; Skoufias, D. A.; Bednar, J.; Hamiche, A.; Angelov, D.; Petosa, C.;

Dimitrov, S. Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. *Mol Cell* **2018**, *72*, 902–915 e7.

[5] Bowman, G. D.; Poirier, M. G. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chemical Reviews* **2015**, *115*, 2274–2295.

[6] Winogradoff, D.; Echeverria, I.; Potoyan, D. A.; Papoian, G. A. The Acetylation Landscape of the H4 Histone Tail: Disentangling the Interplay between the Specific and Cumulative Effects. *Journal of the American Chemical Society* **2015**, *137*, 6245–6253.

[7] Ikebe, J.; Sakuraba, S.; Kono, H. H3 Histone Tail Conformation within the Nucleosome and the Impact of K14 Acetylation Studied Using Enhanced Sampling Simulation. *PLOS Computational Biology* **2016**, *12*, e1004788.

[8] Ghoneim, M.; Fuchs, H. A.; Musselman, C. A. Histone Tail Conformations: A Fuzzy Affair with DNA. *Trends in Biochemical Sciences* **2021**, *46*, 564–578.

[9] Arya, G.; Schlick, T. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proceedings of the National Academy of Sciences* **2006**, *103*.

[10] Wakamori, M.; Fujii, Y.; Suka, N.; Shirouzu, M.; Sakamoto, K.; Umehara, T.; Yokoyama, S. Intra- and inter-nucleosomal interactions of the histone H4 tail revealed with a human nucleosome core particle with genetically-incorporated H4 tetra-acetylation. *Scientific Reports* **2015**, *5*, 17204.

[11] Papamokos, G. V.; Tziatzos, G.; Papageorgiou, D. G.; Georgatos, S. D.; Politou, A. S.; Kaxiras, E. Structural Role of RKS Motifs in Chromatin Interactions: A Molecular Dynamics Study of HP1 Bound to a Variably Modified Histone Tail. *Biophysical Journal* **2012**, *102*, 1926–1933.

[12] Iwasaki, W.; Miya, Y.; Horikoshi, N.; Osakabe, A.; Taguchi, H.; Tachiwana, H.; Shibata, T.; Kagawa, W.; Kurumizaka, H. Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* **2013**, *3*, 363–369.

[13] Chakraborty, K.; Kang, M.; Loverde, S. M. Molecular Mechanism for the Role of the H2A and H2B Histone Tails in Nucleosome Repositioning. *The Journal of Physical Chemistry B* **2018**, *122*, 11827–11840.

[14] Kameda, T.; Awazu, A.; Togashi, Y. Histone Tail Dynamics in Partially Disassembled Nucleosomes During Chromatin Remodeling. *Frontiers in Molecular Biosciences* **2019**, *6*.

[15] Huertas, J.; Schöler, H. R.; Cojocaru, V. Histone tails cooperate to control the breathing of genomic nucleosomes. *PLOS Computational Biology* **2021**, *17*, e1009013.

[16] Armeev, G. A.; Kniazeva, A. S.; Komarova, G. A.; Kirpichnikov, M. P.; Shaytan, A. K. Histone dynamics mediate DNA unwrapping and sliding in nucleosomes. *Nature Communications* **2021**, *12*.

[17] Furukawa, A.; Wakamori, M.; Arimura, Y.; Ohtomo, H.; Tsunaka, Y.; Kurumizaka, H.; Umehara, T.; Nishimura, Y. Characteristic H3 N-tail dynamics in the nucleosome core particle, nucleosome, and chromatosome. *iScience* **2022**, *25*, 103937.

[18] Bilotti, K.; Kennedy, E. E.; Li, C.; Delaney, S. Human OGG1 activity in nucleosomes is facilitated by transient unwrapping of DNA and is influenced by the local histone environment. *DNA Repair* **2017**, *59*, 1–8.

[19] Bhat, J. A.; Balliano, A. J.; Hayes, J. J. Histone protein surface accessibility dictates direction of RSC-dependent nucleosome mobilization. *Nucleic Acids Research* **2022**, *50*.

[20] Ekaterina, S.; Emmanuelle, B.; Patrick, S.; Gabor, P.; Adam, B.-S. Binding to nucleosome poises SIRT6 for histone H3 de-acetylation. *eLife* **2023**, *12*.

[21] Zhou, K.; Gaullier, G.; Luger, K. Nucleosome structure and dynamics are coming of age. *Nature Structural & Molecular Biology* **2019**, *26*.

[22] Fedulova, A. S.; Armeev, G. A.; Romanova, T. A.; Singh-Palchevskaia, L.; Kosarim, N. A.; Motorin, N. A.; Komarova, G. A.; Shaytan, A. K. Molecular dynamics simulations of nucleosomes are coming of age. *WIREs Computational Molecular Science* **2024**, *14*, e1728.

[23] Meyer, S.; Becker, N. B.; Syed, S. H.; Goutte-Gattat, D.; Shukla, M. S.; Hayes, J. J.; Angelov, D.; Bednar, J.; Dimitrov, S.; Everaers, R. From crystal and NMR structures, footprints and cryo-electron-micrographs to large and soft structures: nanoscale modeling of the nucleosomal stem. *Nucleic Acids Research* **2011**, *39*, 9139–9154.

[24] Lebedenko, O. O.; Salikov, V. A.; Izmailov, S. A.; Podkorytov, I. S.; Skrynnikov, N. R. Using NMR diffusion data to validate MD models of disordered proteins: Test case of N-terminal tail of histone H4. *Biophysical Journal* **2024**, *123*, 80–100.

[25] Lehmann, K.; Felekyan, S.; Kühnemuth, R.; Dimura, M.; Tóth, K.; Seidel, C. A. M.; Langowski, J. Dynamics of the nucleosomal histone H3 N-terminal tail revealed by high precision single-molecule FRET. *Nucleic Acids Research* **2020**, *48*, 1551–1571.

[26] Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *NATURE STRUCTURAL BIOLOGY* **2002**, *9*.

[27] Dans, P. D.; Gallego, D.; Balaceanu, A.; Darré, L.; Gómez, H.; Orozco, M. Modeling, Simulations, and Bioinformatics at the Service of RNA Structure. *Chem* **2019**, *5*, 51–73.

[28] Šponer, J.; Islam, B.; Stadlbauer, P.; Haider, S. In *Annual Reports in Medicinal Chemistry*; Neidle, S., Ed.; Quadruplex Nucleic Acids As Targets For Medicinal Chemistry; Academic Press, 2020; Vol. 54; pp 197–241.

[29] Nam, K.; Shao, Y.; Major, D. T.; Wolf-Watz, M. Perspectives on Computational Enzyme Modeling: From Mechanisms to Design and Drug Development. *ACS Omega* **2024**, *9*, 7393–7412.

[30] Hospital, A.; Battistini, F.; Soliva, R.; Gelpí, J. L.; Orozco, M. Surviving the deluge of biosimulation data. *WIREs Computational Molecular Science* **2020**, *10*, e1449.

[31] Ceriotti, M. Unsupervised machine learning in atomistic simulations, between predictions and understanding. *The Journal of Chemical Physics* **2019**, *150*, 150901.

[32] Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications* **2018**, *9*.

[33] Fleetwood, O.; Kasimova, M. A.; Westerlund, A. M.; Delemotte, L. Molecular Insights from Conformational Ensembles via Machine Learning. *Biophysical Journal* **2019**, S0006349519344017.

[34] Grazioli, G.; Martin, R. W.; Butts, C. T. Comparative Exploratory Analysis of Intrinsically Disordered Protein Dynamics Using Machine Learning and Network Analytic Methods. *Frontiers in Molecular Biosciences* **2019**, *6*, 42.

[35] Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *CURRENT OPINION IN STRUCTURAL BIOLOGY* **2020**, *61*, 139–145.

[36] Zhou, J.; Huang, M. Navigating the landscape of enzyme design: from molecular simulations to machine learning. *Chemical Society Reviews* **2024**, *53*, 8202–8239.

Chapter 2

Formation and Repair of DNA damages

Involved People

Alessio Bartocci (Post-Doc), Laleh Allahkaram (PhD), Maxime Kermarrec (Master and PhD), Thomas Boukéké-Lesplulier (PhD), Valentin Bransolle (Master), Viktoria Georgieva (L3), Zineb Elftmaoui (Master)

Related Article

-E. Matoušková, E. Bignon, V. E. P. Claerbout, T. Dršata, N. Gillet, A. Monari, E. Dumont, Filip Lankaš, *JCTC*, **2020**, *16*, 5972

- N. Gillet, A. Bartocci, E. Dumont, *Theor. Chem. Acc.*, **2021**,*140*, 89

- E. Bignon, V. E. P. Claerbout, T. Jiand, C. Morell., N. Gillet, E. Dumont, *Sci. Rep.* **2020**, *10*, 1

- E. Bignon, N. Gillet, T. Jiang, C. Morell. E. Dumont *J. Phys. Chem. Lett*, **2021**, *12*, 6014

- T. Wen, M. Kermarrec, E. Dumont, N. Gillet, M. M. Greenberg, *JACS*, **2023**, *145*, 23702

2.1 Introduction

As a guarantee of the genetic integrity of an organism and its progeny, the DNA molecular structure is highly stable and well conserved in the whole living realm. However, as organic molecules, DNA components are sensitive to various endo or heterogeneous stresses such as radiations (from UV-light to *gamma*-ray) or oxidative stress.[1–4] The molecular consequences range from slight nucleobase modifications to double strand breaks as summarized in Figure 2.1. At larger scale, these chemical modifications can lead to mutations, and their accumulation to hinder cells functions: aging, cell apoptosis or cancer figure among the repercussion of the DNA exposure to stresses.[5–9]

To counterbalance the occurrence of these deleterious reactions, proteins have evolved to detect damages and catalyzes their repairs.[10–14] In humans, we can distinguish two majors mech-

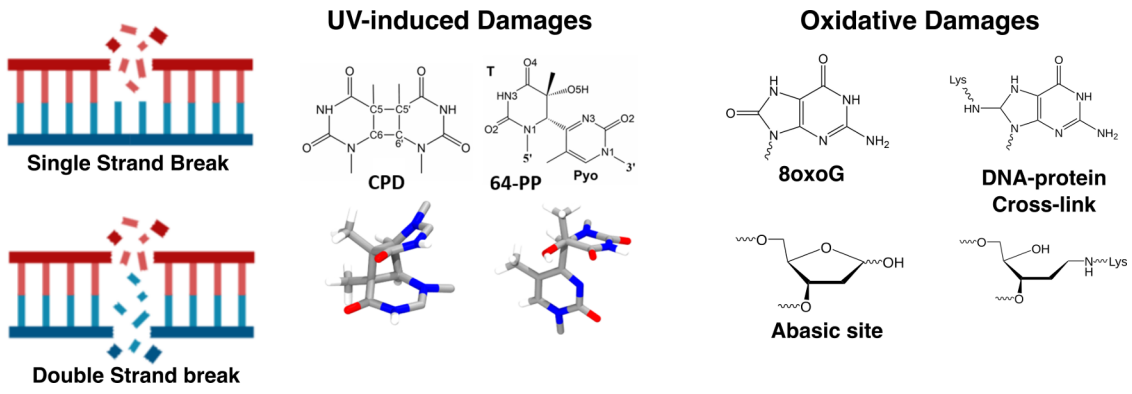

Figure 2.1: Some DNA damages with a focus on the ones described in this chapter.

anisms called Base excision repair (BER) and Nucleotide excision repair (NER) which differ not only on excision target but also on the proteins partners involved and their preferential substrates (Figure 2.2). Despite this specificity, an interplay between protein players from both repair mechanisms has been considered, especially for oxidative damages.[15,16] This hypothesis is at the heart of the PCSI project BERNUMOL, which I will detail below. In addition to this common and versatile mechanisms, most organisms, from archea-bacteria to some marsupials - but not placental mammals like us-, possess photolyases, proteins which are able to recognize and repair UV-induced damages with a high efficiency using blue light.[17] The study of the mechanism of these proteins constitutes the main topic of my post-doctoral stay in Pr. M. Elstner group.[18–21]

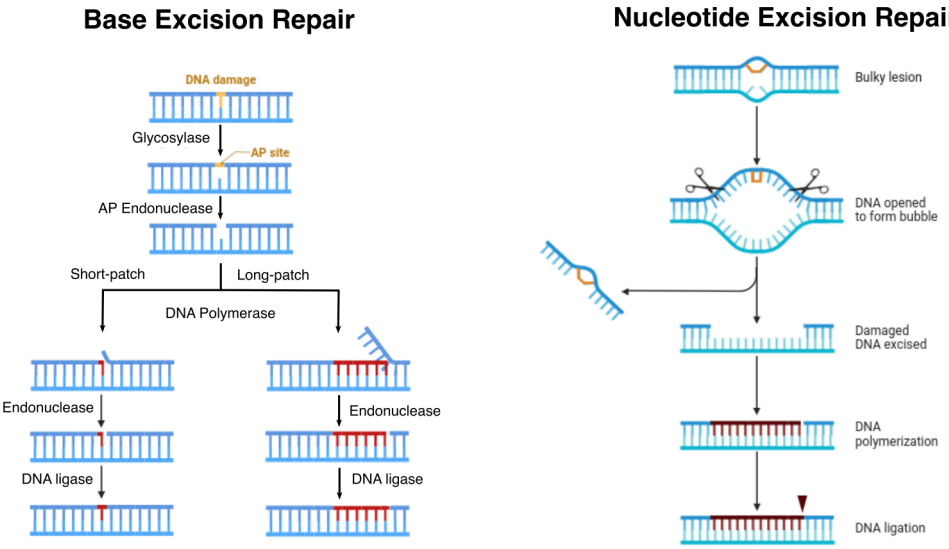

Figure 2.2: Schematic representation of NER and BER mechanisms. Created with Biorender.com.

Several questions are associated to the DNA damages: their formations, their consequences on the DNA conformational and dynamical behaviour, the detection by repair proteins and their repair mechanism. Many computational studies have already explored the behavior of the dam-

ages in free DNA..[22,23] One can notice the emergence of simulations of damaged nucleosomes,[24] in combination with the determination of the corresponding crystallographic or cryo-EM structures.[25–29] Several studies underline the impact of the histones on the lesions occurrence and repair: the DNA wrapping induces a periodic distribution of damages and/or mutations while histones can play a protective role or be directly involved in lesions *via* DNA-protein cross-link.[30–32]

In this chapter, I will detail my contributions to this field, mostly by mean of classical MD simulations, including biased and/or enhanced sampling, to explore different items presented in this chapter:

- conformational and dynamical behavior of 6-4 photoproduct in B-DNA and nucleosomal DNA
- conformational and dynamical behavior of abasic sites on nucleosomal DNA
- conformational and dynamical behavior of 8-oxoGuanine on nucleosomal DNA
- histone-DNA interactions at the origin of oxidative DNA-protein cross-link formations

2.2 Methods for an extended conformational sampling

Thanks to the dramatic increase of the computational power of classical MD software, microseconds timescale simulations of highly complex biochemical complexes have become affordable for the last decade. Applied to DNA, they allow to get a better understanding of the conformational and dynamical impact of the lesions in free or nucleosomal DNA.[33–37] In addition to the classical molecular dynamics, several approaches can be employed to improve the conformational sampling of a damage. In this chapter, I will detail the methods we used to capture rare events.

Firstly, one can want to improve a conformational sampling of a system without any preconceived notion of the interesting event to capture. Running long simulations on several replica remains a solution, but its performance fluctuates depending on the system flexibility. Replica exchanges approaches allow a substantial increase of conformational sampling by running parallel replica in different conditions (for instance, at different temperatures) on the same time and allowing exchange between them on a system energy criterion as described in Figure 2.3 **A**. Then, the free energy barrier between two conformations can be overcome even though the corresponding mechanism remains unveiled. The final MD simulation of the replica at room temperature (300 K in our simulations) consists in a succession of possible conformations in this condition, but without any temporal continuity. A cluster analysis of the conformational landscape can help to analyse the final distribution of the different structures. However, the number

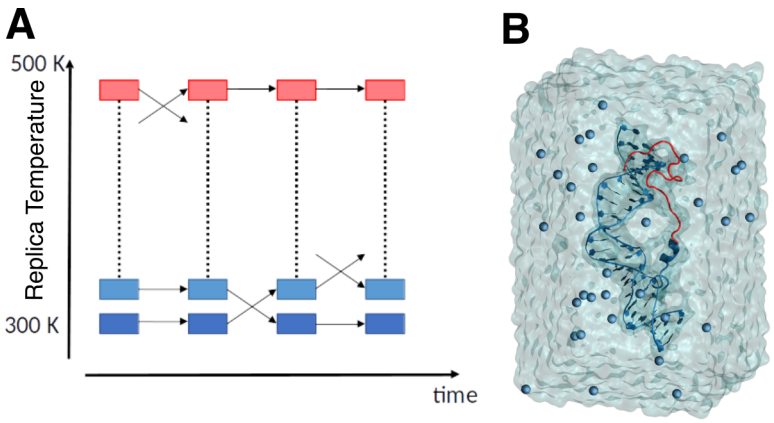

Figure 2.3: **A**. Scheme of replica exchange approach with different temperatures between replica. **B**. An example for REST2 system with a DNA strand and a histone tail; only the part of the tail in red is heated between the replica.

of run replicas must determined to provide a sufficiently high exchange probability, which dramatically depend on the size of the system in term of number of atoms. As solvent molecules are in general not interesting to sample, one can focus on only a part of the system, the solute of interest, to be modified between parallel replicas. Replica exchange with Solute Tempering (REST2) approach is based on this idea (2.3 **B**).[38] Only a part of the system is thus concerned by the Hamiltonian modification, decreasing the difference between two replicas and increasing the exchange probability. Consequently, such approaches require less replicas with no loss in sampling relevancy. This approach has been widely used in RNA folding studies.[39,40]

Secondly, one's interest can be a specific reaction, such as a flipping-out of a DNA damage. This is of high interest for the damage repair as the lesion must flip in the repair protein active site. It also consists in a measure of the DNA flexibility and structure destabilization induced by the presence of a damage. In term of computational approach, such an issue supposes to define first a (set of) reaction coordinate(s) along which the reaction occurs. This step can be tricky as many variables can drive the reaction. However, the system must stay sufficiently free to evolve in a physical way. In the literature, some dihedral angles involving the center of mass of nucleobases and sugar have been used as flipping reaction coordinate.[41–43] Then, several methods allow to obtain a free energy profile for the given reaction such as umbrella sampling, metadynamics, string methods... In the present work, I present only some preliminary results using umbrella sampling.[44] In this methods, MD simulations are carried out with a harmonic restraint on a given value of the reaction coordinate. Using weighted histogram analysis method,[45] the free energy profile can be obtained from the sampling of the reaction coordinate along the different simulations. This approach can be combined to a string method algorithm to draw the minimum free energy path starting from a guess between the reactants and the products valleys.[46] Using MDs around the different points of the path to explore the free energy surface and keeping the distance between these points, the path progressively converges to a less expens-

ive one with respect to free energy. Finally, we can also use metadynamics simulations where gaussian functions are progressively added to push the system to explore the free energy surface along the reaction coordinate.[47] It can be coupled to extended adaptive biasing force (eABF).[48] In this approach, biasing forces are added along the reaction coordinate, instead of a restrain on the energy, in a way that accelerates the free energy surface sampling.

2.3 6-4 photoproduct dynamics

6-4 photoproduct, induced by UV-radiation, is a bulky lesion obtained by the formation of a covalent bond between the C6 of a thymine to the C4 of the 3' adjacent thymine, concomitant with an oxygen and a proton transfer. It results in a bulky intrastrand DNA cross-link, less frequent but much more mutagenic than cyclobutane pyrimidine dimer photolesion. Indeed, the covalent bond formation supposes the breaking of Watson-Crick hydrogen bond for a least one the thymine due to their almost perpendicular conformation. Then the consequences of the overall DNA structure can cover a wide range of conformations with a dependence on the DNA sequence and the impact of the nucleosome core mechanical constraint can be high. Two computational studies have been carried out to explore these questions.

On one hand, we tried to determine the impact of the sequence on the 6-4 photoproduct in the free DNA context enhancing the conformational sampling by the use of the REST2 approach.[38] Different tetrameric sequences have been selected within a common 25 mer B-DNA strand (structure from[49]): **AAAG**, **TAAC**, **TAAT**, **AAAA**, **CAAT**, **GAAA**. They represent different situations as the G-C pair is stronger to the A-T ones thanks to the additional hydrogen bond. In this work, we observe that the 5' base pair before the damage is much less flexible than the 3' ones when a purine stands before the damage. On the contrary, a sequence where the damage is surrounded by two pyrimidines is much more flexible at both ends and can even adopt some extrahelical conformations that are essential for the DNA repair (Figure 2.4 **A**). Even though this work does not cover all the possible sequences, it validates the use of REST2 approach to enlarge our conformational sampling and represents one of the few computational research on the sequence effect on the 6-4 photoproduct DNA landscape.[50]

On the other hand, I also participated to the analysis of the microsecond timescale trajectories of nucleosomes containing 6-4 photoproducts.[36] This study belongs to a collaborative work with Pr. F. Lankaš. We focused on the comparison of the structural behavior for free or nucleosomal DNA and the different protein-DNA interactions and their possible impact on the observed local flexibility of the damage. It appears that the mechanical constraint due to the DNA wrapping around the histone core modifies the DNA behaviour around the damage. Despite the truncated character of the tail, we were able to demonstrate that different protein residues can interact with the damage and the opposite adenine (see, for instance, Figure 2.4 **B**). They enhance their conformational flexibility, possibly triggering a recognition by repair proteins if the damage is

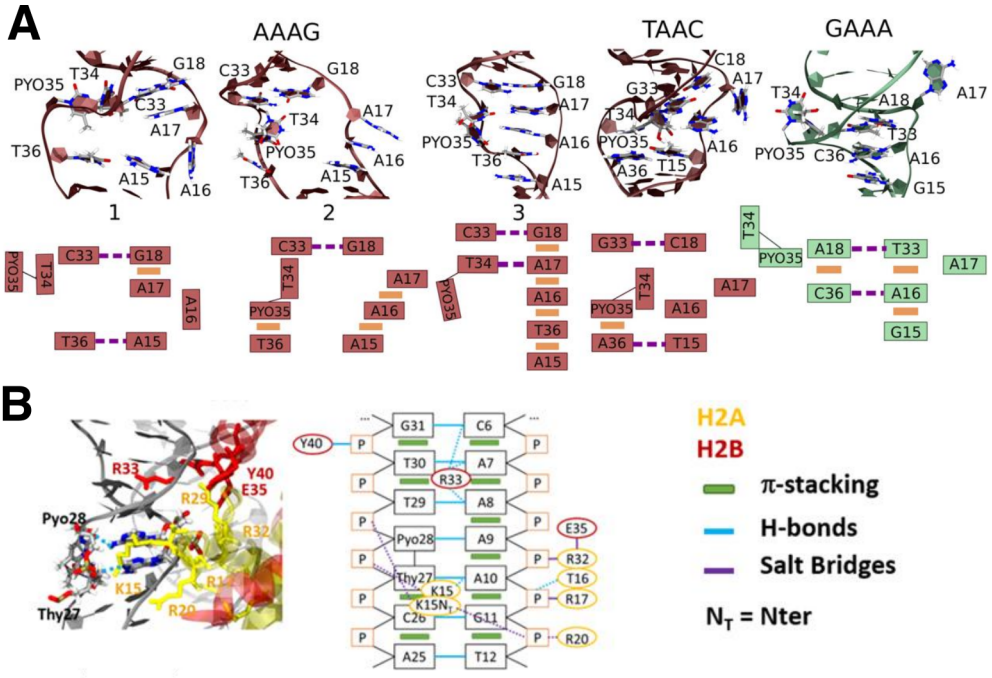

Figure 2.4: **A**. Representation of different extrahelical conformations of the 6-4 photoproduct or the complementary adenines from REST2 simulation of damaged double strands DNA (reproduced from[50]). **B**. Example of interaction network between a 6-4 photoproduct and H2A/H2B histone tails (reproduced from[36]).

solvent accessible but also if it stands close to the histone core. This first study highlights the importance of a good sampling of the histone tails as their interactions with DNA dramatically impact its conformation behaviour around the damage.

2.4 Oxidative damages

Because of its aromatic nature and the presence of ketone and amine moieties, DNA is sensitive to oxidation. This later can be direct, i.e . induced by an ionizing irradiation leading to the ejection of one electron in the solvent and followed by charge transfer and radical reactivity, or indirect, which corresponds to the reactivity with oxidizing cofactors, the most famous ones belonging to the Reactive Oxygen Species (ROS) associated to the oxidative stress in cells. The subsequent damages are numerous, affecting preferentially guanine nucleobase because of its lower ionization potential. I have focused my research on few of them: abasic sites, which can have a strong structural impact due to the lost of a nucleobase; 8-oxoGuanine, the most frequent damage resulting from guanine oxidation by the reaction with ROS or between the radical cation and water;[51] the DNA-protein cross-link involving lysine residues and oxidized guanine, a possible drawback of the DNA wrapping around the histone core.

2.4.1 Nucleosomal abasic sites conformational landscape

In collaboration with Dr. E. Bignon, I participate to the study of the structural impact of the presence of one or clustered abasic site(s) within the nucleosomal DNA. The presence of an abasic site or of its tetrahydrofuran (THF) analog can lead to the formation of an "inchworm" conformation (Figure 2.5 **A**),[27] but also to the flipping out of the abasic site or of the orphan nucleobase. Our MDs simulations, in line with the experimental structure for the THF, show that this structural and dynamical behaviour is tuned by the damage positioning around the nucleosomal and the strength of the interaction with the histone core.[35] However, we observed more fluctuations around the abasic site than for the THF, and the flipped-out adenine in the non-usual "inchworm conformation" can go back within the double strand and perturb the conformational landscape of the damaged DNA. This change is allowed by a less pronounced interaction with the histone core in the presence of the abasic site. Moreover, during our MDs simulations, truncated histone tails form direct interactions with the abasic site (Figure 2.5 **B**). Such interactions are preliminary to the formation of DNA-protein cross-link between lysine and abasic site which can quickly lead to strands scission.[52,53]

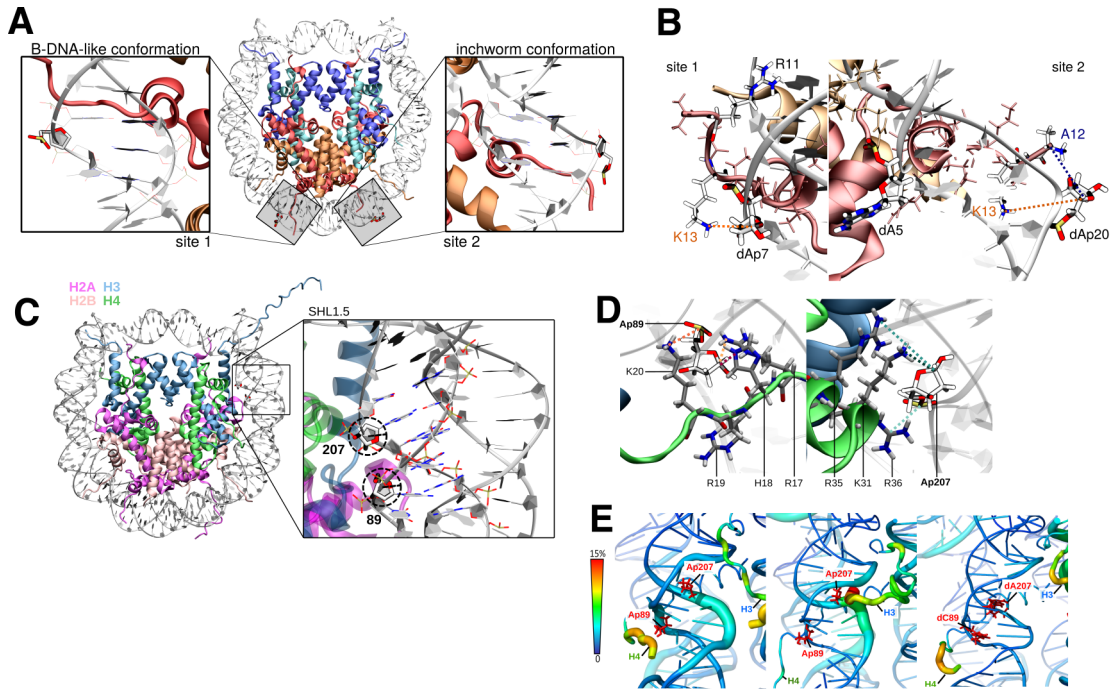

Figure 2.5: **A**. Representation of two nucleosomal abasic sites, one in the B-DNA like conformation (site 1), the other in the "inchworm" conformation site2) and **B**. Some histone tails residues interacting with the abasic sites 1 and 2 (reproduced from[35]). **C**. Representation of clustered abasic sites in nucleosomal DNA; **D**. some residues from H4 tail interacting with the abasic sites; **E**. PCA importance of the DNA and protein residue in the neighbourhood of the clustered damages (reproduced from[37]).

Besides, we applied a similar protocol to clustered abasic site lesions (Figure 2.5 **C**). These damages are highly deleterious for the genetic information by enhancing double strands breaking.

In this study,[37] two abasic sites are located at 89 and 207 positions, on complementary strands and separated by two base pairs. The dynamic behaviour has been explored along 4 replicas of one microsecond each. Comparing to free DNA, the histones induce some constraints on the damage DNA, limiting the flexibility of the nucleobases in the lesions surrounding. However, we again notice the occurrence of several interactions between the damage and lysine or arginine residues from the histone tails (Figure 2.5 **D**). These two studies benefit from the development of a unsupervised machine learning algorithm based on principal component analysis (PCA).[54] Thanks to it, we are able to highlight the importance of some residue in the overall flexibility of the nucleosomal DNA (Figure 2.5 **E**).

2.4.2 Recognition of 8-oxoGuanine damages

8-Oxo-7,8-dihydroguanine (8-oxoGuanine or 8-oxoG), is a staple of oxidative stress consequences: mutagenic, due to the formation of a Hoogsteen pair with Adenine during DNA replication,[55,56] prone to re-oxidation thanks to its lower ionization potential than guanine, oxidative stress indicators by its accumulation on RNA,[57] DNA,[58] and more specifically guanine-rich G-quadruplex DNA..[59] It is repaired by glycosylases of BER, such as OGG1,[60] despite the relatively small conformational change induced in DNA structure by this lesion.[61] Within the framework of the BERNUMOL PCSI proof of concept project in collaboration with the Dr. H. Menoni, I wish to question the recognition of 8-oxoG in the nucleosomal context by both OGG1 and CSB, a DNA remodeller protein from NER. A cooperative mechanism between the two proteins can be at play for the recognition and repair of nucleosomal DNA.

The project firstly requires the selection of the damage positions for experimental setup. To drive our choice, microseconds MD simulations have been performed by V. Bransolle during its M2 internship for 11 positions of the damages. The selected positions differ by their surrounding sequence, the position towards the histone core (inward or outward) and the suprahelical position (Figure 2.6 **A**). Then 5 replicas of one microsecond each have been performed with different initial positions of the histone tails for a control and the damaged sequences. We analysed the flexibility of the DNA strands including the damage, the solvent accessible surface area with or without taking to account the histone tails, the DNA intra and interstrand parameters and the contact between the DNA and the histone tails. Then, we started free energy profile simulations for the flipping of the 8-oxoG from two different positions: the first one close to the dyad, where the damage seems to impact the flexibility of the neighbouring DNA base pairs and far from the tail; the second one close to H2B tail. We compare with the undamaged DNA using the dihedral angle described in Figure 2.6 **B** . The preliminary profile of the flipping are given in Figure 2.6 **C**. When the damage is enclosed in a rigid part of the nucleosome, the minimal position and the free energy profile of damaged and undamaged DNA are very similar. On the contrary, close to the dyad, we observe a difference both in the position of the minimum towards the chosen reaction coordinate and on the height of the free energy barrier. These last results

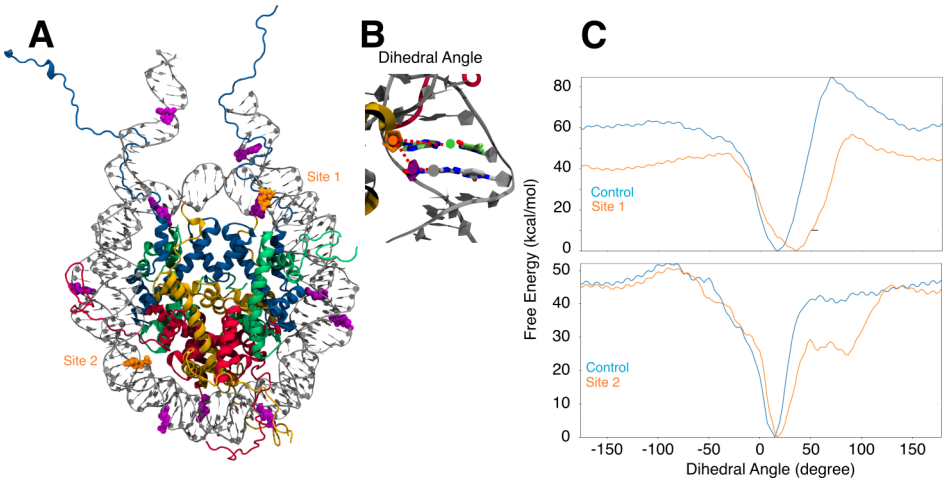

Figure 2.6: **A**. 8-oxoG selected positions on the nucleosomal DNA including a small linker part (10 base pairs on each side) in purple or orange for the sites selected for free energy calculations. **B**. The dihedral angle defining the reaction coordinate of the flipping 8-oxoG between the center of base of the nucleobase pair in 5' (green), the center of mass of the 5' sugar (orange), the center of mass of the sugar of the 8-oxoG (purple, the center of mass of the 8-oxoG (grey). **C**. Preliminary potential Mean Force profile of the rotation of the 8-oxoG in sites 1 and 2 and the comparison with the undamaged analog.

suggest a higher flexibility of the 8-oxoG lesion at this position. Even tough these results need to be confirmed by complementary simulations to improve the quality of the free energy surface, they underline that MD simulations can be a powerful tool to decipher the impact of the damage position on its recognition and repair mechanism.

2.4.3 DNA-protein cross-link formation

The work of M. Greenberg on the oxidized nucleosome has revealed the formation of DNA-protein cross-link involving lysines from histone tails. A part of M. Kermarrec's PhD was dedicated to the tracking of lysine guanine interactions by means of MDs simulation of a nucleosome containing 13 guanines tracks (**AAGGGGGCGCGGGGGAA** sequence at two positions, 35 and 51) in collaboration with Greenberg's group. This work supposes a good conformational sampling of the histone tail behaviour, while several tens of microseconds are necessary to sample relatively well the DNA-tails fuzzy interactions.[62] We chose to start our simulations from 20 histone tails structure combinations for the two DNA sequences. The contact map between lysines (or N-terminal residues) and guanines fit well the experimental attribution of the lysine residues involved in the DNA-protein cross-link (Summary in Figure 2.7 **A**). This theory-experience collaborative work supports on the one hand the consistency of experimental results and the ability of our computational protocol to describe histone tails behaviour at a molecular level .[63]

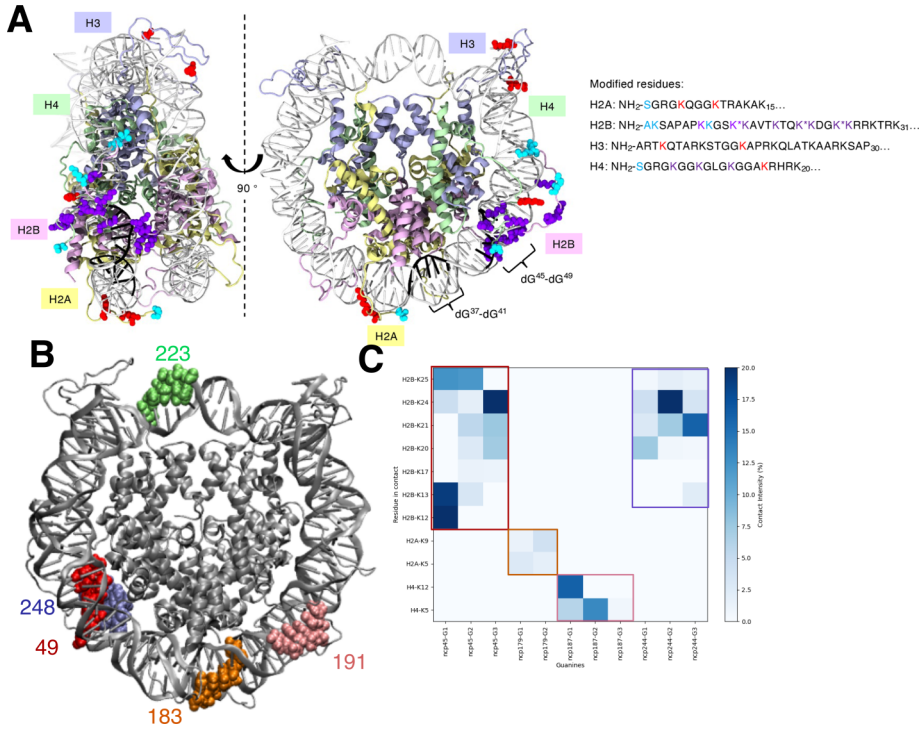

Figure 2.7: **A**. Two perspectives of the NCP with residues involved in DNA-histone cross-link: detected only by LC-MS/MS (red), detected only computationally (cyan), detected both computationally and by LC-MS/MS (dark violet). Asterisks (*) indicate that MS2 spectra do not contain enough fragment ions to distinguish modification at this lysine from that at the adjacent lysine and the corresponding histone N-terminal tail sequences (reproduced from[63]). **B**. Position of the **GGGAA** sequence -number of the last **A**- one the nucleosome. **C**. Contact map between ammonium group of lysines from H2B, H2A or H4 tails and the major groove exposed heavy atoms of guanines.

Besides, a further computational study on the guanine trimers presented in the same article has been carried out by Z. Elftmaoui during her M2 internship using a similar protocol. Five **GGGAA** sequences have been incorporated in the N-DNA sequence at different positions (see Figure 2.7 **B**). We applied the same protocol as in the previous study, with only 5 replicas per sequence. Our results are in relatively good agreement with the experimental data: no lysine-DNA contact is observed close to the dyad, fewer for the sequence at position 183 with H2A histone tail, while the three other sequences present a high number of contacts and a similar experimental formation rate of DNA-protein cross-link (Figure 2.7 **C**). The reaction could occur after the deprotonation of a radical cation guanine.[64]

2.5 Perspective

These first research topics have established the base of my computational protocol for nucleosome study in term of methods and force field selection as well on analysis approaches. Indeed,

it appears that the DNA interaction with histone, and more specifically histone tails is crucial for most of the explored systems. On the basis of literature, I also expect that the histone tails will have a great impact on the repair protein-nucleosomal DNA interactions landscape we want to explore later in the framework of the BERNUMOL project. The presence of post-tranlastional modifications such as methylation or acetylation on histone tails can also impact the pictures from these primary studies.

Within the next years, I will focus on the oxidative damages through two main projects:

- the second part of the BERNUMOL will start in 2025 with the simulations of OGG1 and CSB protein with nucleosomal DNA including damages on the selected positions by a M2 student, in strong collaboration with Dr. H. Menoni team who will carry out the experiments. We got a first view of the interactions of these proteins with free B-DNA thanks to the internship of V. Georgieva. In case of success, our mixed experiment-computation protocol will be extended to others repair proteins, and submitted to a full PCSI project.
- T. Boukéké Lesplulier is currently doing his PhD under the co-direction of Aurélien de la Lande and myself. For the moment, his research focuses on the development of a protocol to study the impact of high energy radiations on a biological system (a collagen strand). Then, his goal is to apply this protocol to a nucleosome and observe the consequences of the nucleosome core direct ionization. The conformational landscape of the resulting damages, in line with my other topics (8oxoG behaviour and radical cation transfer along nucleosomal DNA), will be explored.

Bibliography

[1] Cadet, J.; Ravanat, J.-L.; TavernaPorro, M.; Menoni, H.; Angelov, D. Oxidatively generated complex DNA damage: Tandem and clustered lesions. *Cancer Letters* **2012**, *327*, 5–15.

[2] Ravanat, J.-L.; Breton, J.; Douki, T.; Gasparutto, D.; Grand, A.; Rachidi, W.; Sauvaigo, S. Radiation-mediated formation of complex damage to DNA: a chemical aspect overview. *The British Journal of Radiology* **2014**, *87*, 20130715.

[3] Cadet, J.; Davies, K. J. A.; Medeiros, M. H.; Di Mascio, P.; Wagner, J. R. Formation and repair of oxidatively generated damage in cellular DNA. *Free Radical Biology and Medicine* **2017**, *107*, 13–34.

[4] Mao, P.; Wyrick, J. J.; Roberts, S. A.; Smerdon, M. J. UV-Induced DNA Damage and Mutagenesis in Chromatin. *Photochemistry and Photobiology* **2017**, *93*, 216–228.

[5] Yu, Y.; Cui, Y.; Niedernhofer, L. J.; Wang, Y. Occurrence, Biological Consequences, and Human Health Relevance of Oxidative Stress-Induced DNA Damage. *Chemical Research in Toxicology* **2016**, *29*, 2008–2039.

[6] Tubbs, A.; Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **2017**, *168*, 644–656.

[7] Mullenders, L. H. F. Solar UV damage to cellular DNA: from mechanisms to biological effects. *Photochemical & Photobiological Sciences* **2018**, *17*, 1842–1852.

[8] Kay, J.; Thadhani, E.; Samson, L.; Engelward, B. Inflammation-induced DNA damage, mutations and cancer. *DNA Repair* **2019**, *83*, 102673.

[9] Amente, S.; Di Palo, G.; Scala, G.; Castrignano, T.; Gorini, F.; Cocozza, S.; Moresano, A.; Pucci, P.; Ma, B.; Stepanov, I.; Lania, L.; Pelicci, P. G.; Dellino, G. I.; Majello, B. Genome-wide mapping of 8-oxo-7,8-dihydro-2'-deoxyguanosine reveals accumulation of oxidatively-generated damage at DNA replication origins within transcribed long genes of mammalian cells. *Nucleic Acids Res* **2019**, *47*, 221–236.

[10] Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature* **2001**, *411*.

[11] Hegde, M. L.; Hazra, T. K.; Mitra, S. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res* **2008**, *18*, 27–47.

[12] Jackson, S. P.; Bartek, J. The DNA-damage response in human biology and disease. *Nature* **2009**, *461*, 1071–1078.

[13] Marteijn, J. A.; Lans, H.; Vermeulen, W.; Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology* **2014**, *15*.

[14] Hu, J.; Adar, S.; Selby, C. P.; Lieb, J. D.; Sancar, A. Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes & Development* **2015**, *29*, 948–960.

[15] Menoni, H.; Hoeijmakers, J. H.; Vermeulen, W. Nucleotide excision repair-initiating proteins bind to oxidative DNA lesions in vivo. *J Cell Biol* **2012**, *199*, 1037–46.

[16] Menoni, H.; Di Mascio, P.; Cadet, J.; Dimitrov, S.; Angelov, D. Chromatin associated mechanisms in base excision repair - nucleosome remodeling and DNA transcription, two key players. *Free Radical Biology and Medicine* **2017**, *107*, 159–169.

[17] Tan, C.; Liu, Z.; Li, J.; Guo, X.; Wang, L.; Sancar, A.; Zhong, D. The molecular origin of high DNA-repair efficiency by photolyase. *Nature Communications* **2015**, *6*, ncomms8302.

[18] Holub, D.; Lamparter, T.; Elstner, M.; Gillet, N. Biological relevance of charge transfer branching pathways in photolyases. *Physical Chemistry Chemical Physics* **2019**, *21*, 17072–17081.

[19] Holub, D.; Ma, H.; Krauß, N.; Lamparter, T.; Elstner, M.; Gillet, N. Functional role of an unusual tyrosine residue in the electron transfer chain of a prokaryotic (6–4) photolyase. *Chemical Science* **2018**, *9*, 1259–1272.

[20] Holub, D.; Kubař, T.; Mast, T.; Elstner, M.; Gillet, N. What accounts for the different functions in photolyases and cryptochromes: a computational study of proton transfers to FAD. *Physical Chemistry Chemical Physics* **2019**, *21*, 11956–11966.

[21] Ma, H.; Holub, D.; Gillet, N.; Kaeser, G.; Thoulass, K.; Elstner, M.; Krauß, N.; Lamparter, T. Two aspartate residues close to the lesion binding site of Agrobacterium (6-4) photolyase are required for Mg2+ stimulation of DNA repair. *The FEBS Journal* **2019**, *286*, 1765–1779.

[22] Monari, A.; Dumont, E. Understanding DNA under oxidative stress and sensitization: the role of molecular modeling. *Frontiers in Chemistry* **2015**, *3*, 43.

[23] Gillet, N.; Bignon, E.; Dumont, E.; Monari, A. In *DNA Photodamage: From Light Absorption to Cellular Responses and Skin Cancer*; Improta, R., Douki, T., Eds.; The Royal Society of Chemistry, 2021; p 0.

[24] Gillet, N.; Dumont, E.; Bignon, E. DNA damage and repair in the nucleosome: insights from computational methods. *Biophysical Reviews* **2024**,

[25] Osakabe, A.; Tachiwana, H.; Kagawa, W.; Horikoshi, N.; Matsumoto, S.; Hasegawa, M.; Matsumoto, N.; Toga, T.; Yamamoto, J.; Hanaoka, F.; Thomä, N. H.; Sugasawa, K.; Iwai, S.; Kurumizaka, H. Structural basis of pyrimidine-pyrimidone (6–4) photoproduct recognition by UV-DDB in the nucleosome. *Scientific Reports* **2015**, *5*, 16330.

[26] Horikoshi, N.; Tachiwana, H.; Kagawa, W.; Osakabe, A.; Matsumoto, S.; Iwai, S.; Sugasawa, K.; Kurumizaka, H. Crystal structure of the nucleosome containing ultraviolet light-induced cyclobutane pyrimidine dimer. *Biochemical and Biophysical Research Communications* **2016**, *471*, 117–122.

[27] Osakabe, A.; Arimura, Y.; Matsumoto, S.; Horikoshi, N.; Sugasawa, K.; Kurumizaka, H. Polymorphism of apyrimidinic DNA structures in the nucleosome. *Scientific Reports* **2017**, *7*, 41783.

[28] Matsumoto, S.; Cavadini, S.; Bunker, R. D.; Grand, R. S.; Potenza, A.; Rabl, J.; Yamamoto, J.; Schenk, A. D.; Schübeler, D.; Iwai, S.; Sugasawa, K.; Kurumizaka, H.; Thomä, N. H. DNA damage detection in nucleosomes involves DNA register shifting. *Nature* **2019**, *571*, 79–84.

[29] Weaver, T. M.; Hoitsma, N. M.; Spencer, J. J.; Gakhar, L.; Schnicker, N. J.; Freudenthal, B. D. Structural basis for APE1 processing DNA damage in the nucleosome. *Nature Communications* **2022**, *13*, 5390.

[30] Cannan, W. J.; Tsang, B. P.; Wallace, S. S.; Pederson, D. S. Nucleosomes Suppress the Formation of Double-strand DNA Breaks during Attempted Base Excision Repair of Clustered Oxidative Damages*. *Journal of Biological Chemistry* **2014**, *289*, 19881–19893.

[31] Banerjee, S.; Chakraborty, S.; Jacinto, M. P.; Paul, M. D.; Balster, M. V.; Greenberg, M. M. Probing Enhanced Double-Strand Break Formation at Abasic Sites within Clustered Lesions in Nucleosome Core Particles. *Biochemistry* **2017**, *56*, 14–21.

[32] Wang, R.; Yang, K.; Banerjee, S.; Greenberg, M. M. Rotational Effects within Nucleosome Core Particles on Abasic Site Reactivity. *Biochemistry* **2018**, *57*, 3945–3952.

[33] 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 Research* **2016**, *44*, 8588–8599.

[34] 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. *Nucleic Acids Research* **2017**, *45*, 3654–3662.

[35] Bignon, E.; Claerbout, V.; Jiang, T.; Morell, C.; Gillet, N.; Dumont, E. Nucleosomal embedding reshapes the dynamics of abasic sites. *bioRxiv* **2020**,

[36] Matoušková, E.; Bignon, E.; Claerbout, V.; Dršata, T.; Gillet, N.; Monari, A.; Dumont, E.; Lankaš, F. Impact of the nucleosome histone core on the structure and dynamics of DNA containing pyrimidine-pyrimidone (6-4) photoproduct. *bioRxiv* **2020**,

[37] Bignon, E.; Gillet, N.; Jiang, T.; Morell, C.; Dumont, E. A Dynamic View of the Interaction of Histone Tails with Clustered Abasic Sites in a Nucleosome Core Particle. *The Journal of Physical Chemistry Letters* **2021**, *12*, 6014–6019.

[38] Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B* **2011**, *115*, 9431–9438.

[39] Kührová, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Šponer, J.; Otyepka, M.; Banáš, P. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *Journal of Chemical Theory and Computation* **2016**, *12*, 4534–4548.

[40] Mlýnský, V.; Janeček, M.; Kührová, P.; Fröhlking, T.; Otyepka, M.; Bussi, G.; Banáš, P.; Šponer, J. Toward Convergence in Folding Simulations of RNA Tetraloops: Comparison

of Enhanced Sampling Techniques and Effects of Force Field Modifications. *Journal of Chemical Theory and Computation* **2022**, *18*, 2642–2656.

[41] Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. *Journal of Chemical Theory and Computation* **2009**, *5*, 3105–3113.

[42] La Rosa, G.; Zacharias, M. Global deformation facilitates flipping of damaged 8-oxoguanine and guanine in DNA. *Nucleic Acids Research* **2016**, *44*, 9591–9599.

[43] Knips, A.; Zacharias, M. Both DNA global deformation and repair enzyme contacts mediate flipping of thymine dimer damage. *Scientific Reports* **2017**, *7*, 41324.

[44] Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.

[45] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.

[46] Zinovjev, K.; Tuñón, I. Adaptive Finite Temperature String Method in Collective Variables. *The Journal of Physical Chemistry A* **2017**, *121*, 9764–9772.

[47] Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Computational Molecular Science* **2011**, *1*, 826–843.

[48] Fu, H.; Zhang, H.; Chen, H.; Shao, X.; Chipot, C.; Cai, W. Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys. *The Journal of Physical Chemistry Letters* **2018**, *9*, 4738–4745.

[49] Paul, D.; Mu, H.; Dai, Q.; Tavakoli, A.; He, C.; Broyde, S.; Min, J.-H. *Impact of DNA sequences in the DNA duplex opening by the Rad4/XPC nucleotide excision repair complex*; preprint, 2020.

[50] Gillet, N.; Bartocci, A.; Dumont, E. Assessing the sequence dependence of pyrimidine–pyrimidone (6–4) photoproduct in a duplex double-stranded DNA: A pitfall for microsecond range simulation. *The Journal of Chemical Physics* **2021**, *154*, 135103.

[51] Cadet, J.; Douki, T.; Gasparutto, D.; Ravanat, J. L. Oxidative damage to DNA: formation, measurement and biochemical features. *Mutat Res* **2003**, *531*, 5–23.

[52] Sczepanski, J. T.; Wong, R. S.; McKnight, J. N.; Bowman, G. D.; Greenberg, M. M. Rapid DNA–protein cross-linking and strand scission by an abasic site in a nucleosome core particle. *Proceedings of the National Academy of Sciences* **2010**, *107*.

[53] Zhou, C.; Sczepanski, J. T.; Greenberg, M. M. Histone Modification via Rapid Cleavage of C4-Oxidized Abasic Sites in Nucleosome Core Particles. *Journal of the American Chemical Society* **2013**, *135*, 5274–5277.

[54] Fleetwood, O.; Kasimova, M. A.; Westerlund, A. M.; Delemotte, L. Molecular Insights from Conformational Ensembles via Machine Learning. *Biophysical Journal* **2019**, S0006349519344017.

[55] Kuraoka, I.; Suzuki, K.; Ito, S.; Hayashida, M.; Kwei, J. S. M.; Ikegami, T.; Handa, H.; Nakabeppu, Y.; Tanaka, K. RNA polymerase II bypasses 8-oxoguanine in the presence of transcription elongation factor TFIIS. *DNA Repair* **2007**, *6*, 841–851.

[56] Maynard, S.; Schurman, S. H.; Harboe, C.; de Souza-Pinto, N. C.; Bohr, V. A. Base excision repair of oxidative DNA damage and association with cancer and aging. *Carcinogenesis* **2009**, *30*, 2–10.

[57] Ishii, T.; Igawa, T.; Hayakawa, H.; Fujita, T.; Sekiguchi, M.; Nakabeppu, Y. PCBP1 and PCBP2 both bind heavily oxidized RNA but cause opposing outcomes, suppressing or increasing apoptosis under oxidative conditions. *Journal of Biological Chemistry* **2020**, *295*, 12247–12261.

[58] Gorini, F.; Scala, G.; Cooke, M. S.; Majello, B.; Amente, S. Towards a comprehensive view of 8-oxo-7,8-dihydro-2'-deoxyguanosine: Highlighting the intertwined roles of DNA damage and epigenetics in genomic instability. *DNA Repair* **2021**, *97*, 103027.

[59] Clark, D. W.; Phang, T.; Edwards, M. G.; Geraci, M. W.; Gillespie, M. N. Promoter G-quadruplex sequences are targets for base oxidation and strand cleavage during hypoxia-induced transcription. *Free Radical Biology and Medicine* **2012**, *53*, 51–59.

[60] Fromme, J. C.; Banerjee, A.; Verdine, G. L. DNA glycosylase recognition and catalysis. *Current Opinion in Structural Biology* **2004**, *14*, 43–49.

[61] Singh, S. K.; Szulik, M. W.; Ganguly, M.; Khutsishvili, I.; Stone, M. P.; Marky, L. A.; Gold, B. Characterization of DNA with an 8-oxoguanine modification. *Nucleic Acids Research* **2011**, *39*, 6789–6801.

[62] Peng, Y.; Li, S.; Onufriev, A.; Landsman, D.; Panchenko, A. R. Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails. *Nature Communications* **2021**, *12*, 5280.

[63] Wen, T.; Kermarrec, M.; Dumont, E.; Gillet, N.; Greenberg, M. M. DNA–Histone Cross-Link Formation via Hole Trapping in Nucleosome Core Particles. *Journal of the American Chemical Society* **2023**, *145*, 23702–23714.

[64] Bignon, E.; Chan, C.-H.; Morell, C.; Monari, A.; Ravanat, J.-L.; Dumont, E. Molecular Dynamics Insights into Polyamine–DNA Binding Modes: Implications for Cross-Link Selectivity. *Chemistry – A European Journal* **2017**, *23*, 12845–12852.

Chapter 3

Mapping the Guanine Oxidation hotspots

Involved People

Ranjitha Ravindranath (Master), Laleh Allahkaram (PhD), Maxime Kermarrec (Master2 and PhD), Damien Glaizal (Master), Zineb Elftmaoui (Master2)

Related Article

- R. Ravindranath, P. Mondal, N. Gillet, *Theor. Chem. Acc.*, **2021**, *140*, 89
- M. Kermarrec, E. Dumont, N. Gillet, *Biophys J.*, **2024** *Accepted*

3.1 Introduction

Before the formation of an oxidative damage, the DNA molecule can undergo ionization, i.e. lose one electron, by different ways including irradiation or redox reactions.[1–4] Beyond the natural oxidation process, this ability has been widely studied as DNA reacts as a semi-conductor and is able to transfer the charge over tenths of nucleobases pairs.[5,6] The nature of the nucleobases in the sequence tunes the charge transfer (CT) ability and the redox sensitivity[7,8] as guanine is the easiest nucleobase to oxidize[9] and relatively good charge carrier while thymine is prone to impede the transfer. These properties has been also proposed to be exploited by the cellular machinery to detect some DNA damages: the damage would stop the charge which goes back to the detector protein.[10] Therefore, numerous studies have focused on the ionization potential (IP) of nucleobases and charge transfer along free double strand DNA, both experimentally and computationally.[11–18] Among the latter, some underline the importance of the sequence on the guanine redox properties, especially of the nature of the 3' nucleobases,[13,19] compare the different abilities of purine rings to transfer a charge,[20] or highlight how a DNA damage (abasic site for instance) can impede the transfer.[21] Several models for the DNA CT mechanism have been suggested, which also partially depend on the DNA sequence (see Figure 3.1 A).[22] However, redox properties and CT are highly sensitive to the electrostatic environment around the redox cofactors, which modulates the IP, and to their relative position, which affects the elec-

tronic coupling. The diversity of the DNA folding and packing requires further studies to better understand and describe how DNA ionization occurs and leads to damages.

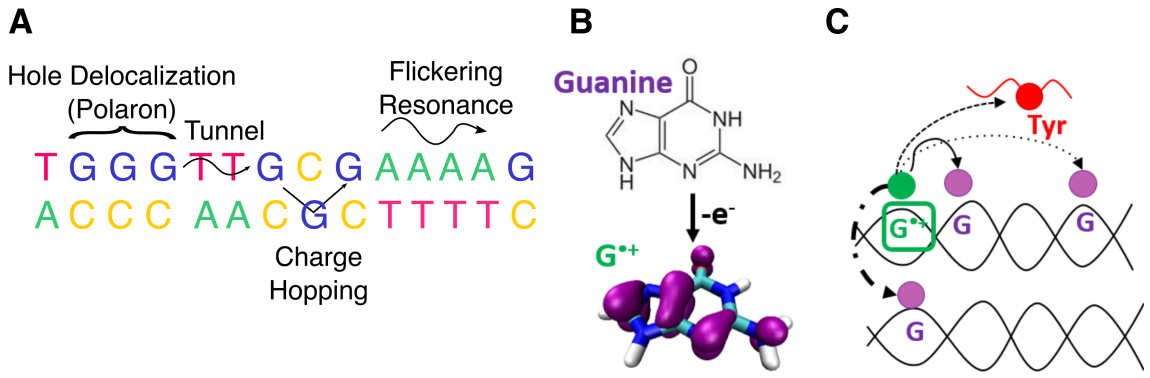

Figure 3.1: **A**.Scheme of different CT mechanisms depending on the sequence. **B**. Structure and HOMO representation of the radical cation Guanine. **C**. Scheme of the possible CT within the nucleosomal DNA.

For instance, the guanine rich G-quadruplex structure constitutes an appealing DNA fold for oxidation. It consists in a single strand DNA folded to form stacked G-quartet confining metal cations such as potassium cations. Such DNA structures can be detected all along the chromosome with the complementary cytosine I-shape folding[23] but are widely present in the telomere.[24] G-quadruplexes represent a drug target cancer,[25,26] for example to modulate the telomerase activity and thus play on the cellular aging and apoptosis.[27] Their high sensitivity to oxidative stress make them free radical scavengers: their shape drawing four columns of stacked guanines is suitable to trap the radical cation guanine (Figure 3.1 B)for a long lifetime (millisecond *versus* microsecond in double strand DNA).[28,29] Computationally speaking, they represents a very nice toy model for IP and CT methods.[30–33]

More complex, the nucleosome provides an DNA environment that can tune the IP and the CT characteristic: the mechanical constrain, the heterogeneous environment (the histone core *versus* the solvent) and the dynamical interactions with positively charged and unfolded histone tails can modulate the previously observed sequence effect.[34,35] A first computational study reports that the geometrical distortions on the G-C base pair affect the IP, and pinpoints the interaction with some histone arginines.[19] However, the dynamical dimension is missing that can lead to over-interpretation of the effect. Then, relatively few studies focus of such a complex issue whereas the experimental counterpart is still hard to set up. The experimental story started in the early 2000's with the work of the Barton's group using photoactivated rhodium DNA intercalators as oxidative precursors.[36] The observed oxidation of guanines in the nucleosomal DNA, even after an assumed long-range charge transfer (24 base pairs), tends to demonstrate that the nucleosome does not protect DNA from oxidation. Later, it has be demonstrated that the depletion of the histone tails strongly modifies the nucleosomal DNA behavior toward ionization.[35] More recent results from Greenberg's group suggest that the charge transfer efficiency varies with respect to

the hole precursor position.[37] They highlight the protective role of a tyrosine from H3, able to suppress the hole transfer (Figure 3.1 C). The role of histone tails is underlined by the measured impact of the presence of post-translational modifications (PTMs) of their lysines.[38] Considering this literature summary, it appears that a topological analysis of the oxidation hotspots, taking into account the combinatorial chemistry at play in the nucleosome is still missing. This is the goal of my ANR JCJC project, NucleoMAP, started in January 2021. We address these different points (also summarized in Figure 3.2) :

- exploration of the conformational landscape of the NCP, with a specific care to the histone tail-DNA interaction profile;
- calculation of guanines redox properties (IP, electronic coupling) in different nucleosomal context to determine which parameters modulate these chemical values using QM/MM approaches
- simulation of CT processes in the nucleosome, including possible participation of tyrosine of histones using dedicating QM/MM protocol;
- the implementation of analysis methods, based on machine learning (ML) algorithm, to deal with the large amount of created data.

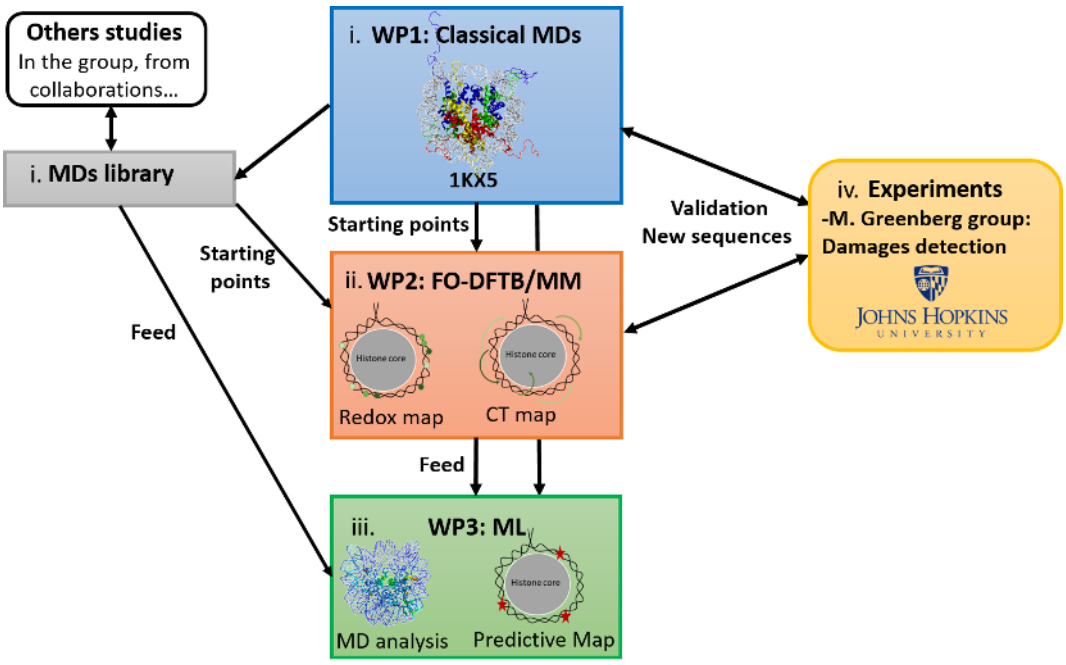

Figure 3.2: Summary of the different work packages of the NucleoMAP JCJC ANR project.

In this chapter, I will present how, in my team, we try to answer these different points, starting from the G-quadruplex toy model to validate our MM-then-QM/MM protocol.

3.2 QM/MM methods for Redox potential calculation and charge transfer simulations

3.2.1 Charge Transfer and Marcus theory

CT i are crucial in many different biological processes such as cellular respiration, photosynthesis, enzymatic reaction etc. and several methods and model have been developed to describe them using computational chemistry.[39] A CT is define as a motion of one electron (for reduction) or one hole (for oxidation) between a donor (D) and an acceptor (A). The CT rate basically depends on the redox potential of each partners in their environment, so their propensity to attract or give an electron taking into account the response of their internal geometry and of the environment to the charge modification. The corresponding energy is the reorganisation energy. The second main parameter consists in the electronic coupling H_{DA} which directly correlates with the overlap of the D-A orbitals. At equilibrium, if the linear response approximation is valid, the free energy profile of a CT can be described by the Marcus theory (see Figure 3.3 A) where the CT rate k_{CT} is given by the equation 3.1:[40,41]

$$k_{CT} = \frac{2\pi}{\hbar} \frac{|H_{DA}|^2}{\sqrt{4\pi\lambda k_b T}} \exp - \frac{(\lambda + \Delta G^\circ)^2}{4\lambda k_b T} \tag{3.1}$$

where ΔG° is the driving force of the CT or the free energy difference between the two considered electronic states D^+A and DA^+, λ the reorganization energy comprising internal and external contributions and H_{DA} is the electronic coupling between the donor and the acceptor. This equation can be applied in most of the biological CT, especially in proteins where slow long-range CT often occur through tunnelling mechanisms.[22,42] A first objective for computational chemistry is to determine accurately these different charge transfer parameters.

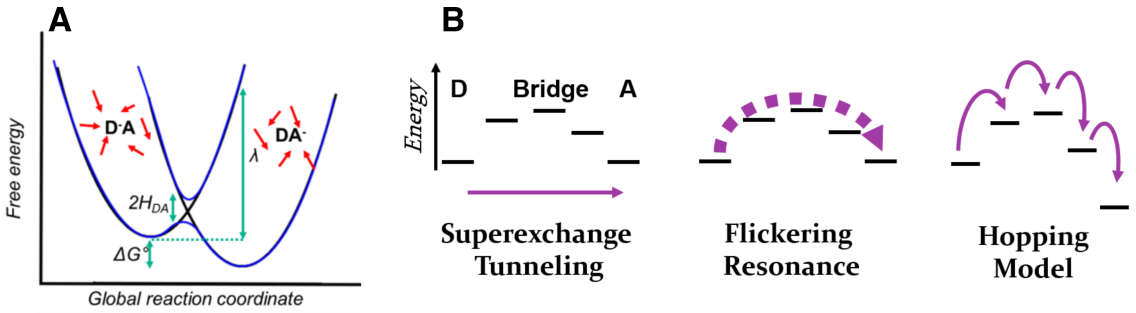

Figure 3.3: **A**. Schematic representation of the Marcus theory: diabatic states parabolas in black, adiabatic surfaces in blue and graphical representation of the driving force ΔG°, the reorganization energy λ and the electronic coupling H_{DA}.**B** Schematic representation of the long-range charge transfer mechanisms between a donor D and an acceptor A involving a bridge B.

Modelling of the CT in biomolecules has been widely explored using multiscale QM/MM

simulations, including the DA pair in the quantum part while the biological environment (biomolecules and solvent) is treated classically (see for instance[31,43–46]). This later is crucial and can tune the different parameters of the equation 3.1; understanding how the specific protein or DNA structure trigger CT remains a key question in such studies. Therefore, numerous protocols have been developed and I will focus on those that implies DFT or its parameterized derivative SCC-DFTB.[47] Indeed, QM calculations are expensive. DFT or semi-empirical approaches permit a wider sampling of the environment than post-Hartree Fock methods. However, DFT suffers from the self-interaction error, which can lead to an over-delocalization of the electronic density. Consequently, it can hinder the definition of the electronic states involved in a CT. Several strategies have been implemented to avoid this drawback, including:

- Constrained DFT (CDFT):[48,49] the charge is constrained on one redox partner during the self-consistent field procedure of the density calculation. The constraint, representing by a Lagrangian function, must fit a charge distribution given by a chosen atomic charge model. Then the two electronic states and the electronic coupling between them can be determined for a given geometry (Figure 3.4 **A**). CDFT can be implemented in a QM/MM framework to take into account a complex environment. One can notice the existence of constrained DFTB.[50]

- Fragment Orbital DFT or DFTB:[51–53] the QM part is divided into fragments which orbital are calculated individually (in the MM environment). A coarse-grained Hamiltonian is derived from the molecular orbital. In the context of CT, the HOMO (for oxidation) or the LUMO (for reduction) are considered to build this Hamiltonian: the diagonal elements are thus an estimation of IP or Electron Affinity respectively while the off-diagonal element are the electronic coupling between two fragments (Figure 3.4 **B** top).

If each charge state can be clearly defined with a suitable electronic density, then MD simulations can be run on both energy surfaces to get the bottom of the corresponding parabola. By approximating the vertical free energy gap between the two states by the potential energy gap, then one can obtain the diabatic free energy surface for each states.[46,54] In the linear response approximation, the driving force and the reorganization energy can be determined from the averaged energy gap on the conformational sampling of each states. Such kind of approach has been recently applied to the redox potential of 8-oxoG in nucleosomal DNA.[55]

However, the Marcus theory is not always suitable: it supposes the validity of the linear response approximation and a CT at the equilibrium. The conformation of the cofactors and their environment must not differ significantly between the two states. This first issue may be solved by using thermodynamic integration approaches which smooth the conformational changes. When more than 2 cofactors are involved in the CT, so we can talk about long-range CT, intermediate charge states can exist. In the case of fast CT, the relaxation of the system for

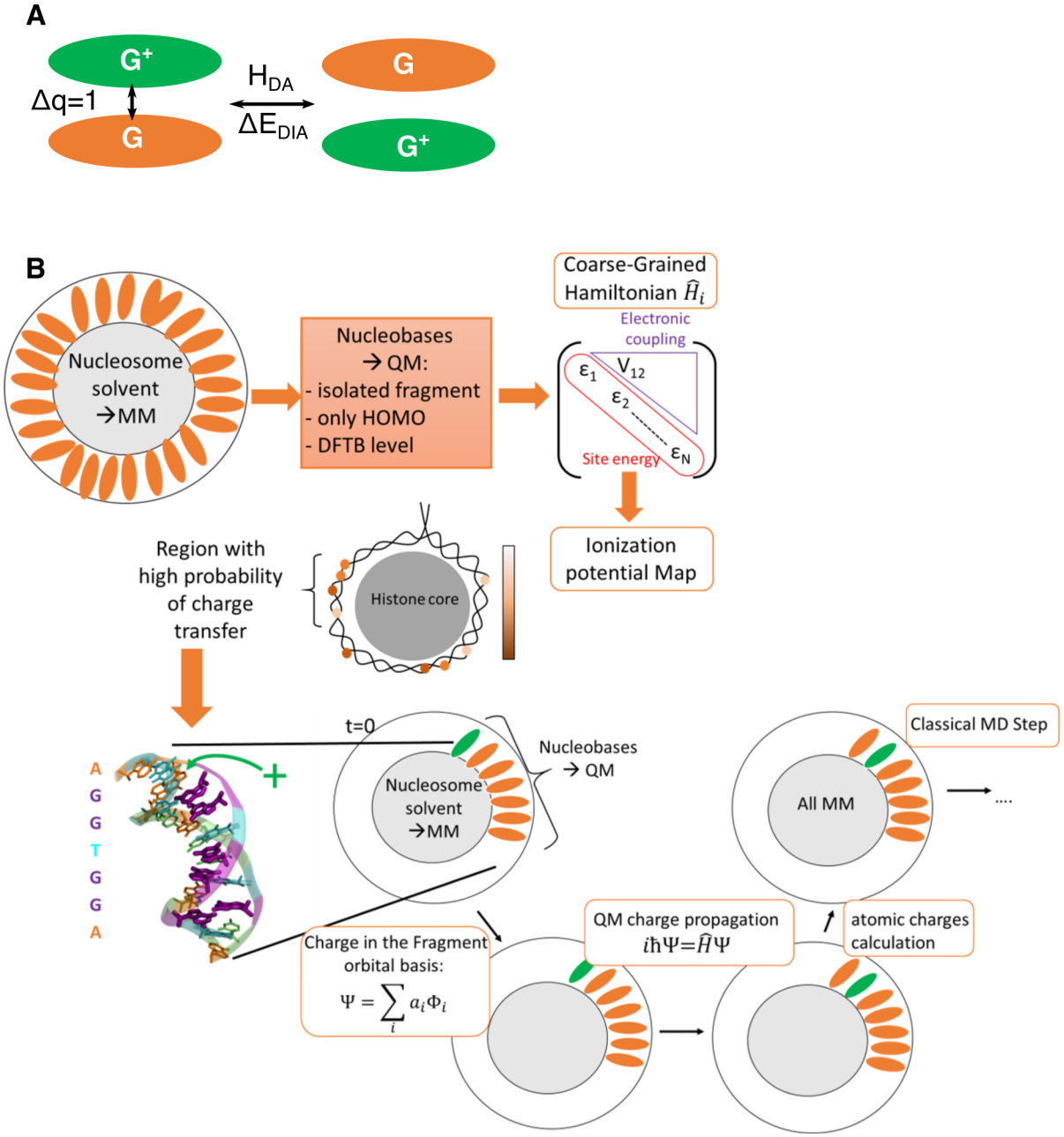

Figure 3.4: **A** Scheme of the CDFT approach used in the calculation in G-quadruplex system: two guanines are considered with a charge difference Δq of 1. The CDFT calculation gives the diabatic energy gap ΔE_{DIA} and the electronic coupling H_{DA} between the two redox states. **B** Scheme of the FO-DFTB/MM calculations, first for the determination of the coarse-grained Hamiltonian in the fragment-HOMO basis set around the nucleosome, secondly for the charge transfer simulations.

a given charge state is not complete, the equilibrium is not reached. Consequently, the Marcus theory fails in such cases, which include DNA CT.

3.2.2 Long-Range Charge Transfer

Most of the CT observed in biology occur over large distances, where several tens of nucleobases or tens of angstroms separates D and A in DNA or in protein assembly respectively. In the equilibrium regime, the driving force and the reorganization energy can be determined for D and A by similar approaches as described in the previous subsection. However, the electronic coupling becomes a more tricky point. Indeed, *in vacuo*, the electronic coupling decreases exponentially with D-A distance. Consequently, the long-range charge transfer are possible only if the charge can travel through the environment, *i.e.* a bridge, which participation allows the electronic density motion. Three main long-range CT types have been described in the literature summarized in Figure 3.3 B:

- **Hopping model**: in this model, the bridge consists in different cofactors that are able to be transiently occupied by the charge. It supposes different successive CTs, which can be ultrafast and falls out from the Marcus theory framework. The FO-DFTB/MM CT MD simulations can be well adapted to describe such transfers.[56]

- **Superexchange model**: in this model, the electron does not occupy the bridge but tunnels through it, and the gap between the D/A and the bridge redox active states is large compared to the electronic coupling between D/A and B. The thermal fluctuations bring D and A in resonance and the presence of an electronic density between them increases the coupling. The empirical pathway model from Beratan and coworkers illustrates the role of the environment in such CT: a different prefactor for the exponential decrease of the electronic coupling according to the nature of the bridge: covalent bond, hydrogen bond or vacuum.[57] We have also tested the ability of CDFT and FO-DFTB to determine at a quantum level this electronic coupling.[58]

- **Flickering resonance model**: in that case, the whole redox active system (D, A and the bridge) is brought to resonance thanks to the thermal fluctuations. The charge is then transiently delocalized through D, A and the bridge and finally stabilized on A after relaxation.

In DNA, the hole transfer mechanism depends on the DNA sequence: tunnelling or flickering resonance over several A-T,[59] while hopping can describe transfers between purine, and more specifically guanines.[9] CT at longer distance should involve a mix of superexchange and multistep hopping transfers. The hypothesis of a polaron type charge carrier was also suggested[60] with a delocalization of the charge over several base pairs. In this chapter, I will focus on the specific role of a heterogeneous environment on guanine IP and on CT involving dG-rich sequences only, that supposes fast hopping mechanism with charge delocalization on 2-3 guanines.

We thus require methods based on charge propagation model, such as Ehrenfest or surface hopping, sufficiently cheap to simulate the propagation at the required timescale (about 100 times

the CT rates at least). The charge transfer scheme based on FO-DFTB/MM approach developed in Elstner's group[56,61] fits these specifications. It has been applied both on DNA CT[30,62] and protein CT from the photolyase-cryptochrome family.[63–67] Based on the determination of the FO-DFTB/MM Hamiltonian, this approach splits each MD step into two kind of calculation:(i) the quantum propagation of the charge within the FO basis set obtained for a given geometry of the cofactors and their environment; (ii) the propagation of atomic motion at the classical level, considering the charge distribution given by the quantum step and projected on the classical atomic charges (see (Figure 3.4 **B** bottom)). Several tens of simulations can be run at the same time and then averaged to draw a kinetic description of the charge transfer and eventually provide CT rate taking into account the effect of the biomolecular environment.

3.3 Guanine ionization potential in different contexts

3.3.1 Sequence impact on Guanine ionization potential

Previous computational studies on B-DNA structures have demonstrated an interacting effect of 0.2-0.3 eV of the nature of the 3' nucleobase on the guanine IP.[13,19] following a C > T > A > G order. The nucleobase in 5' position has a clearly less important impact. To check if our method is able to reproduce such effect, we perform a serie of MM simulations of a 16 mer sequence containing a central XGY trimer for all the 16 possible X and Y combinations (Figure 3.5 **A**). The geometry of the 6 nucleobases have been extracted every ns of the 200 ns of each simulations and QM calculations have been performed at CDFT or FO-DFTB/MM level to determine the guanine ionization potential. The per-sequence results are presented in Figure 3.5 **B** top). The CDFT results correspond to the energy gap between the neutral and the radical cation guanine on a given geometry while the FO-DFTB results correspond to the HOMO energy of the central guanine. In the CDFT framework, the 6 nucleobases are treated at quantum level whereas only the central guanine figures in the quantum part in FO-DFTB/MM calculations

If our CDFT calculations are able to reproduce the 3' effect at a 0.2-0.4 eV range, the FO-DFTB results difference is less pronounced (0.2 eV). This can be due to the classic character of the 3' nucleobase in such kind of calculation. Nevertheless, we can apply a correction to our FO-DFTB results to improve the 3' effect: based on our CDFT calculations of the guanine ionization potential, we can determine the energy difference we should add to obtain, on average, a similar energy difference between the trimer using F-DFTB/MM approach. Similar corrections based on gas phase DFT calculations have been applied before to correct the difference in FO-DFTB between residues (adenine and guanine, or tyrosine and tryptophan[66]). The corrected data are given in Figure 3.5 **B** bottom.

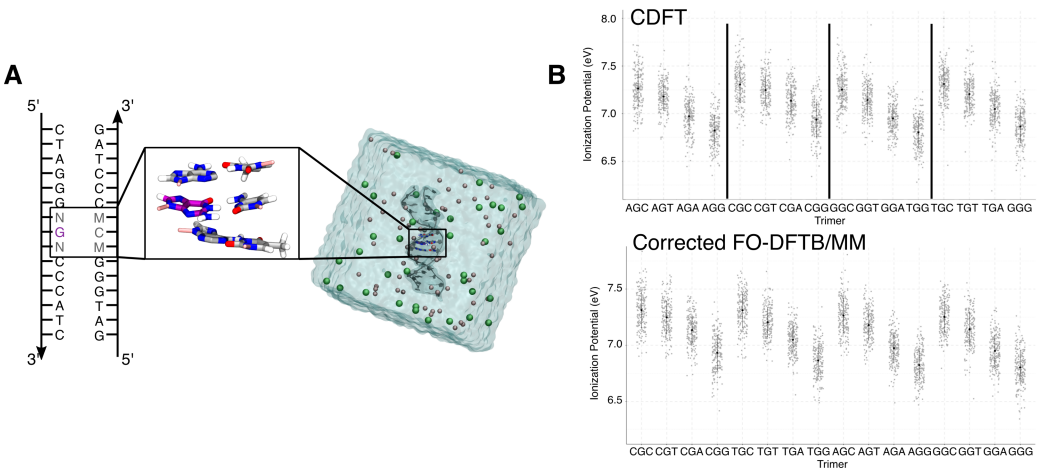

Figure 3.5: **A** Free DNA sequence use for the correction of FO-DFTB/MM results. and the three bases pairs considered in the IP calculations **B** Top: CDFT distribution of the IP for the different trimers; bottom: FO-DFTB/MM distribution of the IP for the different trimers.

3.3.2 Guanine ionization potential in G-quadruplex

In order to evaluate the ability of our approach to capture the impact of structure and environment on the guanine IP, we started the study of six G-quadruplexes structures by means of MM simulations and CDFT/MM or FO-DFTB/MM calculations. The obtained results also provide an important dataset for the development of Machine Learning algorithm for IP.

Because of the cost of the CDFT calculations, only the main representative structures obtained by clustering are taken into account. Each stacked and hydrogen-bonded guanine pair have been considered, and CDFT/MM calculations were performed constraining the charge on one guanine or the other. The vertical energy gap between these two charge states and the corresponding electronic coupling have been determined.

Our results highlight the existence of two regimes: relatively close energy gaps and coupling of tens of meV should insure a relatively fast inter-quartets while the intra-quartet transfers are depending on the hydrogen bond directionality in terms of energy gap and can be forbidden despite their highest coupling. This first study is limited to a very small number of conformations, due to the cost of the QM method.

Consequently, we used the FO-DFTB/MM approach to determine the QM properties of the guanines over 2000 structures from microsecond timescale classical simulations for each G-quadruplex systems. For stacked guanines, we obtain close IP values for the different guanines. The guanines in the central quartets tend to have a higher IP than the external ones (about 0.1 eV), due to the proximity of two cations instead of one. Besides, in some structures, the DNA backbone can influence the IP and decreases the energy by ca. 0.1 eV because of an interaction between the phosphate and the guanine nucleobases. The electronic coupling is slightly lower than in CDFT/MM calculation, which is consistent with previous comparisons between the two

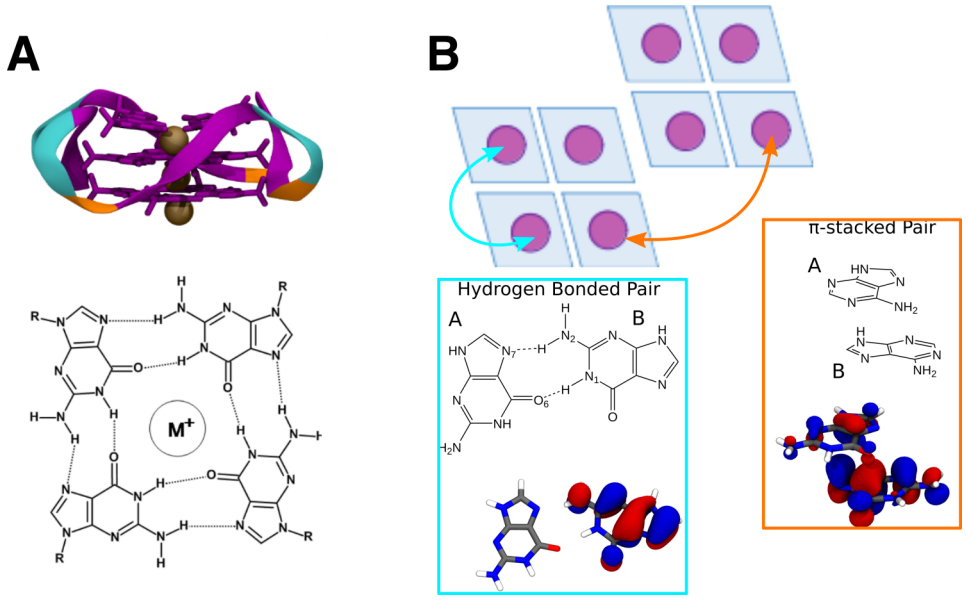

Figure 3.6: **A**. Representation of a G-quadruplex with guanine in purple (top) and of a G-quartet with guanines interacting in the Hoogsteen hydrogen-bond scheme (bottom). **B**. Two kinds of guanine pairs considered for CDFT/MM calculations: hydrogen pairs within a G-quartet (cyan); π-stacked pair between two G-quartets (orange).

methods.[58,68,69] However, the coupling between hydrogen bonded guanine is nonexistent, suggesting that this approach is unable to treat transfer within the quartet.

3.3.3 Guanine Ionization Potential in Nucleosome

Considering the previous calculations, the sequence-corrected FO-DFTB/MM approach seems to reach the good balance between computational cost and quantum accuracy in order to allow a good conformational sampling for the application to the nucleosomal DNA. Indeed, the well-known α-palindromic and the Widow 601 sequences contain 60 and 90 guanines respectively. The sampling of the nucleosome conformation landscape requires at least tens of microseconds simulations to encompass a panorama of the positions of the histone tails. Their positively charged residues can impact the IP of the guanines as already observed in G-quadruplex structures with the entrapped cations. A first step of this study consists in the exploration of the main parameters that can modulate the IP: the sequence, the solvent accessibility linked to the position toward the histone core and the proximity of protein residues.

Starting from our histone tail sampling and the 1KX5 PDB structure,[70] we drawn 20 different starting structures of a nucleosome containing the α-palindromic sequence. After the equilibration procedure, we performed 1 μs of production run for each structure for a total or 20 μs of nucleosome conformational sampling. FO-DFTB/MM calculation of each of the 60 guanines was carried out every nanosecond without any QM/MM re-optimization of the guanine geometry.

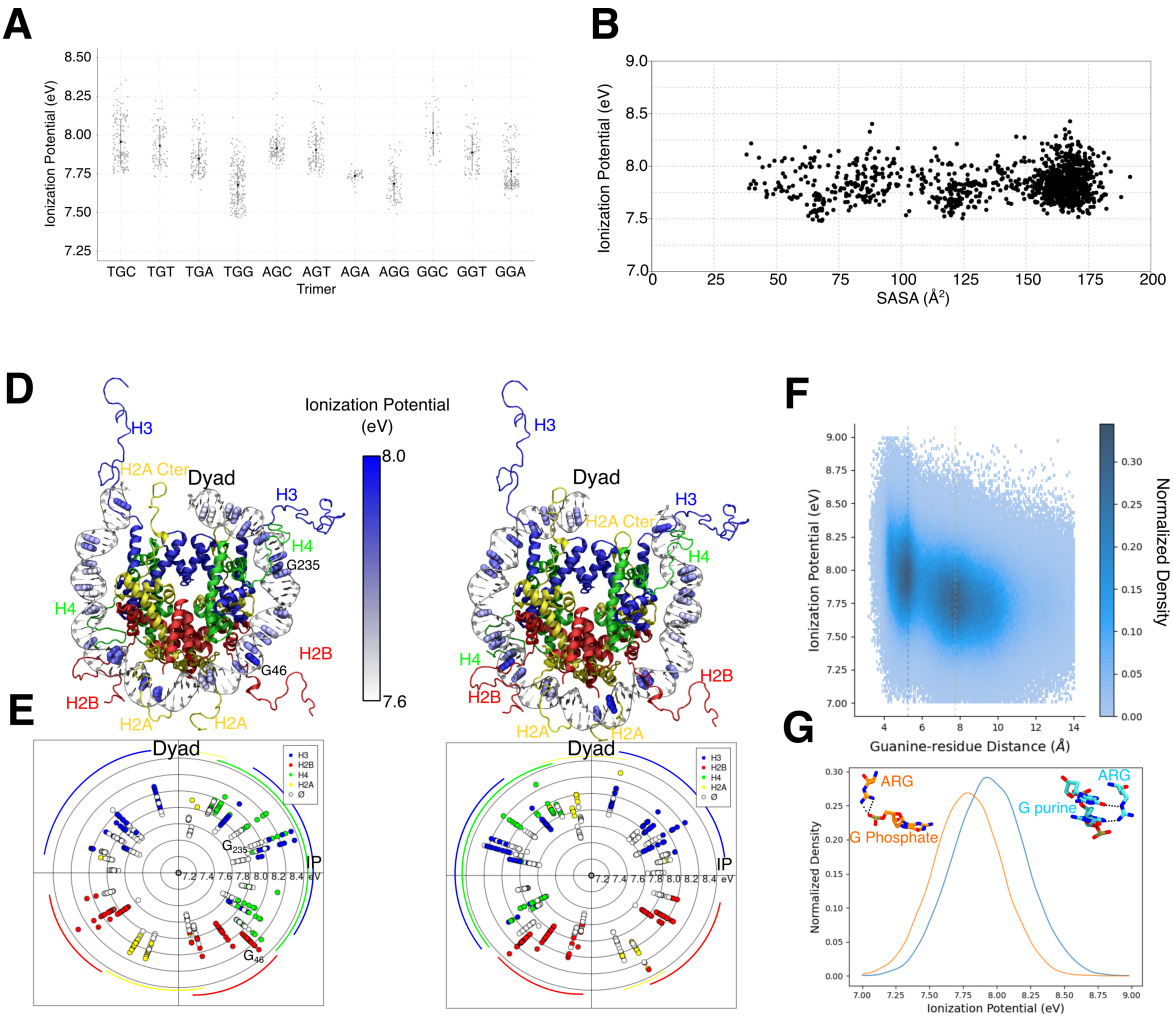

Figure 3.7: **A**. FO-DFTB/MM IP *vs* trimer sequence in the nucleosome. **B**. FO-DFTB/MM IP *vs* Solvent Accessible Surface Area without taking into account the tail. **C**. Guanine colored according to their averaged IP over all MD simulations. **D** Averaged IP of each guanine for each simulations colored according to their contacts with the histone tails. **E**. FO-DFTB/MM IP *vs* the distance to the nearest positively charged residue of histone tail (lysine, arginine, histidine or N-terminal residue). **F**. Distribution of the FO-DFTB/MM IP when protein residues are close to the guanine (cyan) or to the phosphate (orange).

We thus obtain a total of 20 000 ionization potential values for each guanine. The averaged IP ranges over 0.6 eV. Then, we analyze the correlation between different parameters and ionization potential in more detail:

- **DNA sequence** (Figure 3.7 **A**): the effect of the nucleobase in 3' position is diminished compared to free DNA, suggesting that other factors are at play. One current drawback consists in the different distribution of trimers over the nucleosome: some are absent or poorly represented.

- **Proximity of the histone core** (Figure 3.7 **B**): no correlation is observed. Nevertheless,

the solvent accessibility of the purine rings are modulated by the presence of the tails.

- **Proximity of a positively charged residue** (Figure 3.7 **C-F**): the presence of a positively charged residue in the groove (major or minor) interacting directly with the purine ring atoms of the guanine increase the ionization potential by 0.2-0.25 eV and induced a large variation in the IP behaviour. This competes with the sequence effect. The interaction with the histone tail must not be neglected in the determination of guanines redox properties.

Altogether, our results suggest that the dynamic of the nucleosome has an important role to trigger the guanine ionization potential. How does this impact the charge transfer within the nucleosomal DNA?

3.4 Charge transfer in Nucleosome

3.4.1 Charge transfer along Guanine tracks

Using the same modified nucleosome as for the determination of DNA-Protein Cross-link formation, we look at the DNA CT along the guanine tracks with the FO-DFTB/MM charge transfer scheme. The considered nucleosomes contain the **AAGGGGGCGCGGGGGAA** sequence at position 35 or 55. The IP calculations over these sequences along the 20 μs of MD simulations (2) give similar values for all the guanines within the dG tracks, ca 0.5 eV lower than the IP of adenines. Besides, the coupling values are around 50 to 100 meV for intra-strand transfer and 10 meV for inter-strand transfer, for the middle CGC sequence.

For the CT simulations, the starting structures were obtained from the 20 MDs simulations by selecting the representative structure of the main cluster obtained using a hierarchical agglomerative approach on the RMSD of the heavy atoms of the **AAGGGGGCGCGGGGGAA** nucleotides. Then, the charge was positioned on the 5' or the 3' adenine of the system and simulation were run over 50 ps using surface hopping propagator. The QM orbital basis set contains the HOMO orbital of all the 13 guanines (two of them on the complementary strand) and the 4 ending adenines (Figure 3.8 **A**).

The occupation of each guanine tracks over all CT trajectories and the distribution are given in Figure 3.8 **B**.Very fast transfers occur along the dG-tracks and do not seem to be affected by the starting position of the radical cation adenine. According to the evaluation of the spatial extension of the hole wave function, the charge is mostly delocalized between 1 and 3 residues, in agreement with the literature describing the guanine trimer as a charge trap sequence.[71,72] Sampling over different conformations thus give equivalent guanines with respect to the charge transfer properties, but a detailed description per system draw a different picture. Indeed, in some simulations, the charge stays mostly on one **GGGGG** track, with a factor 10 between the

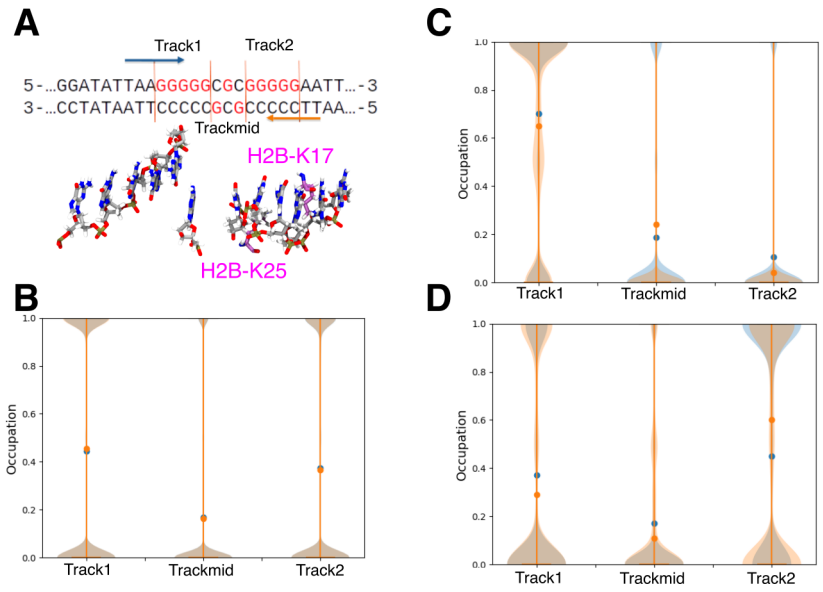

Figure 3.8: **A**. Sequence for CT FO-DFTB/MM MDs and its structure with two proximal H2B lysine residues. **B**. Charge occupation distribution (an occupation of 1 correspond to a positively charged track, 0 to a neutral track) over the 3 different tracks starting from the adenine in 5' (blue) or in 3' (orange) considering 20 different starting points. Points correspond to the averaged occupation. **C**. Charge occupation distribution over the 3 different tracks starting from the adenine in 5' (blue) or in 3' (orange) considering a starting point with two positively charged lysines close to track 2. **D** Charge occupation distribution over the 3 different tracks starting from the adenine in 5' (blue) or in 3' (orange) considering a starting point with two neutral lysines close to track 2.

occupation of the two tracks (Figure Figure 3.8 **C**). Looking at the corresponding structures, we can notice the presence of positively charged residues of the histone tails at close proximity of the neutral track. Despite the propensity of guanine-rich sequence to stabilize a radical cation, the presence of a proximal positive charge decreases the probability of radical cation guanine stabilization, in agreement with our IP study and chemical intuition.

However, it also contradicts the DNA-protein cross-link formation described in Chapter 2 which supposes the reaction between an oxidized guanine and a lysine. An alternative consists in the hypothesis of a neutral deprotonated lysine (so a local basic pH). Using the same FO-DFTB/MM approach in presence of neutral lysines gives a more similar charge distribution over the to 5-dG tracks (Figure 3.8 **D**). These preliminary results must be confirmed but they underline the crucial role of positive charge in the CT dynamics.

3.4.2 Charge transfer to Tyrosine residues

Another role played by histone with respect to the charge transfer and radical cation reactivity involves tyrosines as positive charge acceptors. This protein residue is able to undergo oxidation through a proton coupled electron transfer mechanism, and the resulting neutral radical tyrosine

can have a relative long life-time up to second time scale.[73] Experimental data from Greenberg's group has proven the protective role of H3 Tyr41 (Figure 3.9 **A**) towards guanine oxidative damages:[37,74] they observe a neat decrease of the damages occurrence for the sequence 55 in their dedicated NCP. A similar damage rate as 35 sequence can be recovered by the Y41F mutation on H3. On the contrary, the H3 F84Y mutation reduces the hole transfer efficiency along the 35 sequence thanks to its proximity to 49 guanine. This hole quenching property of the tyrosine histone could be considered as a protection of nucleosomal DNA toward the oxidation and could influence the localization of the damages.

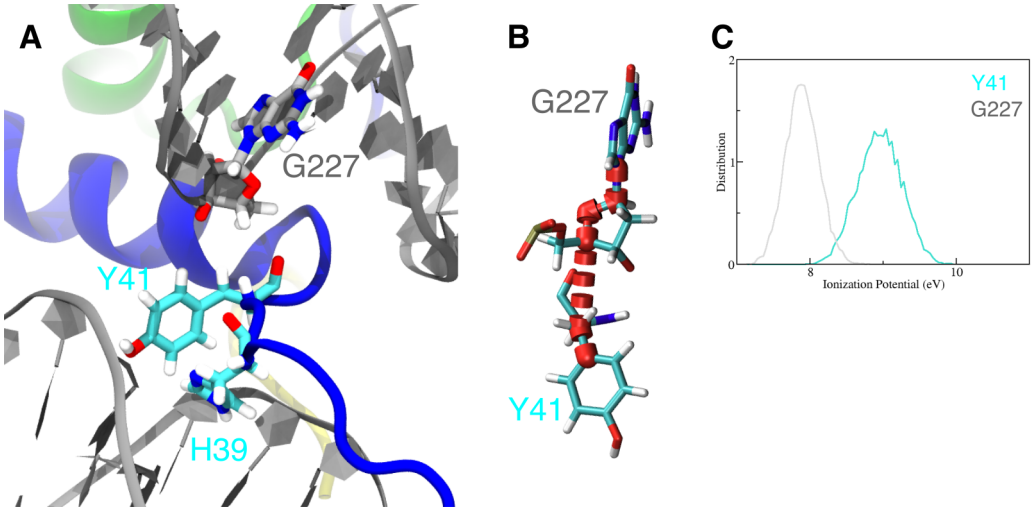

Figure 3.9: **A**. Representation of the H3-Y41, H3-H39 and dG227 for the CT from DNA to histone tail. **B**. Possible pathway using tunnelling pathway model.[75] **C**. FO-DFTB/MM IP distribution of the H3-Y41 and dG227 from 20 μs MD simulations of sequence 55.

Y41 is close to H39 which can play the role of a proton acceptor during tyrosine oxidation and allows a PCET reaction. In the experiments from Greenberg's group, the charge is created on the 55 dG tracks so Y41 can capture the charge from the guanine G227 *via* a super-exchange tunnelling mechanism involving DNA and protein backbone. Our calculations following the Beratan model for tunnelling confirm this hypothesis, with a small coupling value about 0.5 meV (Figure 3.9 **B**), and disqualify any other tunnelling pathways (for instance, including H3 Arg40 side chain, close to dG65). However, the difference in IP suggest that the CT alone will not be allowed (Figure 3.9 **C**). A possible PCET, involving the vicinal H3 His39 can be imagined. Further calculations, using CDFT or dedicated protocol for PCET reactions will be performed to improve our results on these protective CT from DNA to protein.

3.5 Machine Learning for charge transfer in biological environment

Even though our results provide insightful information of the role of histone tail in the modulation of the IP, electronic coupling and charge transfer properties on the nucleosomal DNA, we are currently limited to the analysis of averaged data or few specific situations. However, our simulations cover a wide landscape of interaction networks which diversity could be hidden by a classical analysis protocol. That is why, in collaboration with Dr Jiang, we try to develop a machine learning algorithm able to determine the IP and the electronic coupling to:

- determine these values for a large number of trajectories;
- draw a hierarchy of the parameters that modulate the IPs and the electronic coupling on a large amount of data without bias on the data selection.

A first attempt was made on the G4 results. The electronic coupling mostly depends on the relative geometry of the different redox partners. Consequently, a simple geometrical descriptor, or the coulomb matrix, can be sufficient to provide good reproducibility of the electronic coupling by a ML algorithm.[76,77] The IP is more complex as it has to take into account the fluctuation of both the cofactor and the polar environment, taking into account the bonds variation and the charge distribution. Because of the size of our system, the use of a Coulomb matrix would be too prohibitive. The SOAP descriptor appears as a good alternative: each atom is described by a Gaussian which parameters depend on the atom nature. Combined with a "long short term memory" neural network (our previous attempts used Linear Ridge Regression with already correct reproduction), this ML algorithm is able to reproduce with a good accuracy the FO-DFTB/MM results for a G-quadruplex structure (Figure 3.10 **A** and **B**).

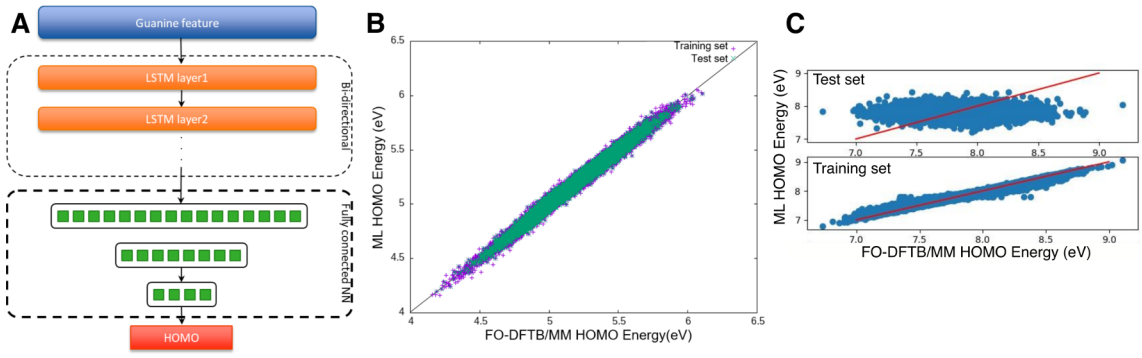

Figure 3.10: **A**. Scheme of the machine learning algorithm. **B**. Results using the machine learning algorithm for Guanine IP determination in a G-quadruplex. **C**. Results using the machine learning algorithm (with Linear Ridge Regression) for guanine IP determination in nucleosomal DNA.

The transferability of this algorithm to the nucleosome system is however not straightforward (see Figure 3.10 **C**) . The large size of the system requires a clever design of the algorithm to limit the memory demand. Moreover, the complexity of the environment makes the use of simple SOAP descriptor inefficient: a same kind of atoms (nitrogen, oxygen, hydrogen...) can be involved in neutral or charge molecule and thus must have a different description. The goal of the two months internship of D. Glaizal was to try this transfer. He tried different approaches based on the coulomb matrix (too big however for such application) or on SOAP parameters. He has shown that the algorithm is unable to discriminate the different cases. A possibility can be to use descriptors dedicated to protein description[78] and change our approach from an atomic description to a different level.

3.6 Perspective

This charge transfer project was at the heart of my ANR JCJC NucleoMAP. My current conclusion draw a picture of the nucleosomal charge transfer where the DNA sequence and the interaction with the histone tails plays a predominant role on the rate, the mechanism and the propensity of radical cation guanine to turn into oxidative damage. This study can be considered as a prequel of the damaged DNA characterisation described in 2 but the bridge between these works is not complete. As a short term perspective, I will consider to continue on:

- the improvement of the machine learning algorithm for IP and electronic coupling for its application to nucleosome. Thanks to the use of different trajectories from my own database and from literature, I aim to draw a more universal IP map for a better understanding of DNA oxidation hotspot. This project can be supported by a collaboration with Dr Jiang and short duration internships.

- the study of the PCET mechanism to Y41 and Y84 in the H3 F84Y. This study will be done thanks to the DFTB/MM based protocol for PCET currently developed by Ms Spies during her PhD, in continuation to the seminal work from my post-doctoral stay in Pr. Elstner's group.[79] Thanks to a combination of DFTB/MM framework with metadynamics (1D if only the proton transfer coordinate is taken into account; 2D when the charge difference between the cofactor is used as a second reaction coordinate to follow the charge transfer by mean of coupled-perturbed DFTB/MM (CP-DFTB/MM)). Beyond the tyrosine pair toy model, Ms Spies has applied this approach on more realistic systems consisting on β-hairpin and three α-helices structures containing a tyrosine-tyrosine, a tryptophan-tyrosine or a histidine-tyrosine pair (Figure 3.11). For her last year, we plan to apply it to Ribonucleotide Reductase and nucleosomal DNA PCET involving H3 Y41 and H39.

For a mean term project, I want to enhance the sampling of charge transfer behavior in different nucleosomal context and associate these results to the previously developed machine learning

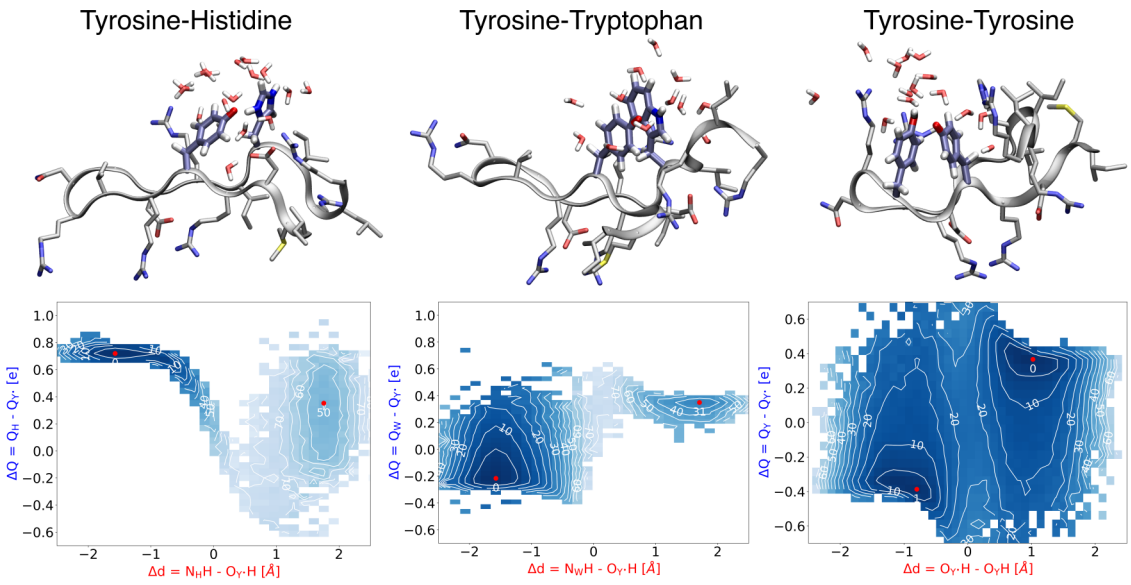

Figure 3.11: PCET free energy surface for different amino-acid pairs in a β-hairpin structure calculated at Long range Corrected-DFTB/MM level.

algorithm in order to take into account the charge transfer in the evaluation of oxidation hotspot. On the other hand, the reactivity of the radical cation guanine towards neutral lysine will be explored to determine the DNA-protein cross-link mechanisms at strategic positions based on DFTB3/MM and biased MD simulations.

Bibliography

[1] Kanvah, S.; Joseph, J.; Schuster, G. B.; Barnett, R. N.; Cleveland, C. L.; Landman, U. Oxidation of DNA: Damage to Nucleobases. *Accounts of Chemical Research* **2010**, *43*, 280–287.

[2] Ravanat, J.-L.; Breton, J.; Douki, T.; Gasparutto, D.; Grand, A.; Rachidi, W.; Sauvaigo, S. Radiation-mediated formation of complex damage to DNA: a chemical aspect overview. *The British Journal of Radiology* **2014**, *87*, 20130715.

[3] Yu, Y.; Cui, Y.; Niedernhofer, L. J.; Wang, Y. Occurrence, Biological Consequences, and Human Health Relevance of Oxidative Stress-Induced DNA Damage. *Chemical Research in Toxicology* **2016**, *29*, 2008–2039.

[4] Cadet, J.; Davies, K. J. A. Oxidative DNA damage & repair: An introduction. *Free Radical Biology and Medicine* **2017**, *107*, 2–12.

[5] Delaney, S.; Barton, J. K. Charge Transport in DNA Duplex/Quadruplex Conjugates. *Biochemistry* **2003**, *42*, 14159–14165.

[6] Arnold, A.; Grodick, M.; Barton, J. DNA Charge Transport: from Chemical Principles to the Cell. *Cell Chemical Biology* **2016**, *23*, 183–197.

[7] Núñez, M. E.; Hall, D. B.; Barton, J. K. Long-range oxidative damage to DNA: Effects of distance and sequence. *Chemistry & Biology* **1999**, *6*, 85–97.

[8] Chakraborty, R.; Ghosh, D. The effect of sequence on the ionization of guanine in DNA. *Physical Chemistry Chemical Physics* **2016**, *18*, 6526–6533.

[9] Meggers, E.; Michel-Beyerle, M. E.; Giese, B. Sequence Dependent Long Range Hole Transport in DNA. *Journal of the American Chemical Society* **1998**, *120*, 12950–12955.

[10] O'Brien, E.; Holt, M. E.; Thompson, M. K.; Salay, L. E.; Ehlinger, A. C.; Chazin, W. J.; Barton, J. K. The [4Fe4S] cluster of human DNA primase functions as a redox switch using DNA charge transport. *Science* **2017**, *355*, eaag1789.

[11] Bixon, M.; Giese, B.; Wessely, S.; Langenbacher, T.; Michel-Beyerle, M. E.; Jortner, J. Long-range charge hopping in DNA. *Proceedings of the National Academy of Sciences* **1999**, *96*, 11713–11716.

[12] Henderson, P. T.; Jones, D.; Hampikian, G.; Kan, Y.; Schuster, G. B. Long-distance charge transport in duplex DNA: The phonon-assisted polaron-like hopping mechanism. *Proceedings of the National Academy of Sciences* **1999**, *96*, 8353–8358.

[13] Voityuk, A. A.; Jortner, J.; Bixon, M.; Rösch, N. Energetics of hole transfer in DNA. *Chemical Physics Letters* **2000**, *324*, 430–434.

[14] Siriwong, K.; Voityuk, A. A. Electron transfer in DNA. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 780–794.

[15] Xiang, L.; Palma, J. L.; Bruot, C.; Mujica, V.; Ratner, M. A.; Tao, N. Intermediate tunnelling–hopping regime in DNA charge transport. *Nature Chemistry* **2015**, *7*, 221–226.

[16] Fujitsuka, M.; Majima, T. Charge transfer dynamics in DNA revealed by time-resolved spectroscopy. *Chemical Science* **2017**, *8*, 1752–1762.

[17] Peluso, A.; Caruso, T.; Landi, A.; Capobianco, A. The Dynamics of Hole Transfer in DNA. *Molecules* **2019**, *24*, 4044.

[18] Diamantis, P.; Tavernelli, I.; Rothlisberger, U. Vertical Ionization Energies and Electron Affinities of Native and Damaged DNA Bases, Nucleotides, and Pairs from Density Functional Theory Calculations: Model Assessment and Implications for DNA Damage Recognition and Repair. *Journal of Chemical Theory and Computation* **2019**, *15*, 2042–2052.

[19] Voityuk, A. A.; Davis, W. B. Hole Transfer Energetics in Structurally Distorted DNA: The Nucleosome Core Particle. *The Journal of Physical Chemistry B* **2007**, *111*, 2976–2985.

[20] Kubař, T.; Elstner, M. What Governs the Charge Transfer in DNA? The Role of DNA Conformation and Environment. *The Journal of Physical Chemistry B* **2008**, *112*, 8788–8798.

[21] Corbella, M.; Voityuk, A. A.; Curutchet, C. How abasic sites impact hole transfer dynamics in GC-rich DNA sequences. *Physical Chemistry Chemical Physics* **2018**, *20*, 23123–23131.

[22] Beratan, D. N. Why Are DNA and Protein Electron Transfer So Different? *Annual Review of Physical Chemistry* **2019**, *70*, 71–97.

[23] Panczyk, T.; Wojton, P.; Wolski, P. Mechanism of unfolding and relative stabilities of G-quadruplex and I-motif noncanonical DNA structures analyzed in biased molecular dynamics simulations. *Biophysical Chemistry* **2019**, *250*, 106173.

[24] Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 6622–6626.

[25] Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature Reviews Drug Discovery* **2011**, *10*.

[26] Neidle, S. Quadruplex nucleic acids as targets for anticancer therapeutics. *Nature Reviews Chemistry* **2017**, *1*.

[27] Epel, E. S.; Blackburn, E. H.; Lin, J.; Dhabhar, F. S.; Adler, N. E.; Morrow, J. D.; Cawthon, R. M. Accelerated telomere shortening in response to life stress. *Proceedings of the National Academy of Sciences* **2004**, *101*, 17312–17315.

[28] Banyasz, A.; Martínez-Fernández, L.; Balty, C.; Perron, M.; Douki, T.; Improta, R.; Markovitsi, D. Absorption of Low-Energy UV Radiation by Human Telomere G-Quadruplexes Generates Long-Lived Guanine Radical Cations. *Journal of the American Chemical Society* **2017**, *139*, 10561–10568.

[29] Banyasz, A.; Balanikas, E.; Martinez-Fernandez, L.; Baldacchino, G.; Douki, T.; Improta, R.; Markovitsi, D. Radicals Generated in Tetramolecular Guanine Quadruplexes by Photoionization: Spectral and Dynamical Features. *The Journal of Physical Chemistry B* **2019**, *123*, 4950–4957.

[30] Woiczikowski, P. B.; Kubař, T.; Gutiérrez, R.; Cuniberti, G.; Elstner, M. Structural stability versus conformational sampling in biomolecular systems: Why is the charge transfer efficiency in G4-DNA better than in double-stranded DNA? *The Journal of Chemical Physics* **2010**, *133*, 035103.

[31] Lech, C. J.; Phan, A. T.; Michel-Beyerle, M.-E.; Voityuk, A. A. Electron-Hole Transfer in G-Quadruplexes with Different Tetrad Stacking Geometries: A Combined QM and MD Study. *The Journal of Physical Chemistry B* **2013**, *117*, 9851–9856.

[32] Livshits, G. I.; Stern, A.; Rotem, D.; Borovok, N.; Eidelshtein, G.; Migliore, A.; Penzo, E.; Wind, S. J.; Di Felice, R.; Skourtis, S. S.; Cuevas, J. C.; Gurevich, L.; Kotlyar, A. B.; Porath, D. Long-range charge transport in single G-quadruplex DNA molecules. *Nature Nanotechnology* **2014**, *9*.

[33] Martinez-Fernandez, L.; Changenet, P.; Banyasz, A.; Gustavsson, T.; Markovitsi, D.; Improta, R. Comprehensive Study of Guanine Excited State Relaxation and Photoreactivity in G-quadruplexes. *J. Phys. Chem. Lett.* **2019**, 5.

[34] Bjorklund, C. C.; Davis, W. B. Attenuation of DNA charge transport by compaction into a nucleosome core particle. *Nucleic Acids Research* **2006**, *34*, 1836–1846.

[35] Davis, W. B.; Bjorklund, C. C.; Deline, M. Probing the Effects of DNA–Protein Interactions on DNA Hole Transport: The N-Terminal Histone Tails Modulate the Distribution of Oxidative Damage and Chemical Lesions in the Nucleosome Core Particle. *Biochemistry* **2012**, *51*, 3129–3142.

[36] Núñez, M. E.; Noyes, K. T.; Barton, J. K. Oxidative Charge Transport through DNA in Nucleosome Core Particles. *Chemistry & Biology* **2002**, *9*, 403–415.

[37] Sun, H.; Zheng, L.; Yang, K.; Greenberg, M. M. Positional Dependence of DNA Hole Transfer Efficiency in Nucleosome Core Particles. *Journal of the American Chemical Society* **2019**, *141*, 10154–10158.

[38] Yang, K.; Prasse, C.; Greenberg, M. M. Effect of Histone Lysine Methylation on DNA Lesion Reactivity in Nucleosome Core Particles. *Chemical Research in Toxicology* **2019**, acs.chemrestox.9b00049.

[39] Blumberger, J. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chemical Reviews* **2015**, *115*, 11191–11238.

[40] Marcus, R. A. Theoretical relations among rate constants, barriers, and Broensted slopes of chemical reactions. *The Journal of Physical Chemistry* **1968**, *72*, 891–899.

[41] Marcus, R. A.; Sutin, N. Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics* **1985**, *811*, 265–322.

[42] Stuchebrukhov, A. A. Long-distance electron tunneling in proteins. *Theoretical Chemistry Accounts* **2003**, *110*, 291–306.

[43] Blancafort, L.; Voityuk, A. A. MS-CASPT2 Calculation of Excess Electron Transfer in Stacked DNA Nucleobases. *The Journal of Physical Chemistry A* **2007**, *111*, 4714–4719.

[44] Antony, J.; Medvedev, D. M.; Stuchebrukhov, A. A. Theoretical Study of Electron Transfer between the Photolyase Catalytic Cofactor FADH $^{\cdot}$ and DNA Thymine Dimer. *Journal of the American Chemical Society* **2000**, *122*, 1057–1065.

[45] Blumberger, J. Free energies for biological electron transfer from QM/MM calculation: method, application and critical assessment. *Physical Chemistry Chemical Physics* **2008**, *10*, 5651.

[46] Gillet, N.; Lévy, B.; Moliner, V.; Demachy, I.; de la Lande, A. Electron and Hydrogen Atom Transfers in the Hydride Carrier Protein EmoB. *Journal of Chemical Theory and Computation* **2014**, *10*, 5036–5046.

[47] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B* **1998**, *58*, 7260.

[48] Wu, Q.; Van Voorhis, T. Constrained Density Functional Theory and Its Application in Long-Range Electron Transfer. *Journal of Chemical Theory and Computation* **2006**, *2*, 765–774.

[49] Wu, Q.; Cheng, C.-L.; Van Voorhis, T. Configuration interaction based on constrained density functional theory: A multireference method. *The Journal of Chemical Physics* **2007**, *127*, 164119.

[50] Scholz, R.; Luschtinetz, R.; Seifert, G.; Jägeler-Hoheisel, T.; Körner, C.; Leo, K.; Rapacioli, M. Quantifying charge transfer energies at donor–acceptor interfaces in small-molecule solar cells with constrained DFTB and spectroscopic methods. *Journal of Physics: Condensed Matter* **2013**, *25*, 473201.

[51] Tsuneyuki, S.; Kobori, T.; Akagi, K.; Sodeyama, K.; Terakura, K.; Fukuyama, H. Molecular orbital calculation of biomolecules with fragment molecular orbitals. *Chemical Physics Letters* **2009**, *476*, 104–108.

[52] Nishioka, H.; Ando, K. Electronic coupling calculation and pathway analysis of electron transfer reaction using ab initio fragment-based method. I. FMO–LCMO approach. *The Journal of Chemical Physics* **2011**, *134*, 204109–12.

[53] Kubař, T.; Woiczikowski, P. B.; Cuniberti, G.; Elstner, M. Efficient Calculation of Charge-Transfer Matrix Elements for Hole Transfer in DNA. *The Journal of Physical Chemistry B* **2008**, *112*, 7937–7947.

[54] Blumberger, J.; Tavernelli, I.; Klein, M. L.; Sprik, M. Diabatic free energy curves and coordination fluctuations for the aqueous Ag[sup +]Ag[sup 2+] redox couple: A biased Born-Oppenheimer molecular dynamics investigation. *The Journal of Chemical Physics* **2006**, *124*, 064507–12.

[55] Kılıç, M.; Diamantis, P.; Johnson, S. K.; Toth, O.; Rothlisberger, U. Redox-Based Defect Detection in Packed DNA: Insights from Hybrid Quantum Mechanical/Molecular Mechanics Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2023**,

[56] Kubař, T.; Elstner, M. Efficient algorithms for the simulation of non-adiabatic electron transfer in complex molecular systems: application to DNA. *Physical Chemistry Chemical Physics* **2013**, *15*, 5794–5813.

[57] Beratan, D. N.; Betts, J. N.; Onuchic, J. N. Protein electron transfer rates set by the bridging secondary and tertiary structure. *Science* **1991**, *252*, 1285–1288.

[58] Gillet, N.; Berstis, L.; Wu, X.; Gajdos, F.; Heck, A.; de la Lande, A.; Blumberger, J.; Elstner, M. Electronic Coupling Calculations for Bridge-Mediated Charge Transfer Using Constrained Density Functional Theory (CDFT) and Effective Hamiltonian Approaches at the Density Functional Theory (DFT) and Fragment-Orbital Density Functional Tight Binding (FODFTB) Level. *Journal of Chemical Theory and Computation* **2016**, *12*, 4793–4805.

[59] Zhang, Y.; Liu, C.; Balaeff, A.; Skourtis, S. S.; Beratan, D. N. Biological charge transfer via flickering resonance. *Proceedings of the National Academy of Sciences* **2014**, *111*, 10049–10054.

[60] Kendrick, T.; Giese, B. Charge transfer through DNA triggered by site selective charge injection into adenine. *Chemical Communications* **2002**, 2016–2017.

[61] Kubař, T.; Elstner, M. Coarse-Grained Time-Dependent Density Functional Simulation of Charge Transfer in Complex Systems: Application to Hole Transfer in DNA. *The Journal of Physical Chemistry B* **2010**, *114*, 11221–11240.

[62] Kubar, T.; Elstner, M. A hybrid approach to simulation of electron transfer in complex molecular systems. *Journal of The Royal Society Interface* **2013**, *10*, 20130415–20130415.

[63] Woiczikowski, P. B.; Steinbrecher, T.; Kubař, T.; Elstner, M. Nonadiabatic QM/MM Simulations of Fast Charge Transfer in Escherichia coli DNA Photolyase. *The Journal of Physical Chemistry B* **2011**, *115*, 9846–9863.

[64] Lüdemann, G.; Woiczikowski, P. B.; Kubař, T.; Elstner, M.; Steinbrecher, T. B. Charge Transfer in E. coli DNA Photolyase: Understanding Polarization and Stabilization Effects via QM/MM Simulations. *The Journal of Physical Chemistry B* **2013**, *117*, 10769–10778.

[65] Lüdemann, G.; Solov'yov, I. A.; Kubař, T.; Elstner, M. Solvent Driving Force Ensures Fast Formation of a Persistent and Well-Separated Radical Pair in Plant Cryptochrome. *Journal of the American Chemical Society* **2015**, *137*, 1147–1156.

[66] Holub, D.; Ma, H.; Krauß, N.; Lamparter, T.; Elstner, M.; Gillet, N. Functional role of an unusual tyrosine residue in the electron transfer chain of a prokaryotic (6–4) photolyase. *Chemical Science* **2018**, *9*, 1259–1272.

[67] Holub, D.; Lamparter, T.; Elstner, M.; Gillet, N. Biological relevance of charge transfer branching pathways in photolyases. *Physical Chemistry Chemical Physics* **2019**, *21*, 17072–17081.

[68] Kubas, A.; Hoffmann, F.; Heck, A.; Oberhofer, H.; Elstner, M.; Blumberger, J. Electronic couplings for molecular charge transfer: Benchmarking CDFT, FODFT, and FODFTB against high-level ab initio calculations. *The Journal of Chemical Physics* **2014**, *140*, 104105–21.

[69] Kubas, A.; Gajdos, F.; Heck, A.; Oberhofer, H.; Elstner, M.; Blumberger, J. Electronic couplings for molecular charge transfer: benchmarking CDFT, FODFT and FODFTB against high-level ab initio calculations. II. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14342–14354.

[70] Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution††We dedicate this paper to the memory of Max Perutz who was particularly inspirational and supportive to T.J.R. in the early stages of this study. *Journal of Molecular Biology* **2002**, *319*, 1097–1113.

[71] Giese, B.; Wessely, S.; Spormann, M.; Lindemann, U.; Meggers, E.; Michel-Beyerle, M. E. On the Mechanism of Long-Range Electron Transfer through DNA. *Angewandte Chemie International Edition* **1999**, *38*, 996–998.

[72] Yoshioka, Y.; Kitagawa, Y.; Takano, Y.; Yamaguchi, K.; Nakamura, T.; Saito, I. Experimental and Theoretical Studies on the Selectivity of GGG Triplets toward One-Electron Oxidation in B-Form DNA. *Journal of the American Chemical Society* **1999**, *121*, 8712–8719.

[73] Oldemeyer, S.; Franz, S.; Wenzel, S.; Essen, L.-O.; Mittag, M.; Kottke, T. Essential Role of an Unusually Long-lived Tyrosyl Radical in the Response to Red Light of the Animal-like Cryptochrome aCRY. *Journal of Biological Chemistry* **2016**, *291*, 14062–14071.

[74] Wen, T.; Kermarrec, M.; Dumont, E.; Gillet, N.; Greenberg, M. M. DNA–Histone Cross-Link Formation via Hole Trapping in Nucleosome Core Particles. *Journal of the American Chemical Society* **2023**, *145*, 23702–23714.

[75] Balabin, I. A.; Hu, X.; Beratan, D. N. Exploring biological electron transfer pathway dynamics with the Pathways Plugin for VMD. *Journal of Computational Chemistry* **2012**, *33*, 906–910.

[76] Bag, S.; Aggarwal, A.; Maiti, P. K. Machine Learning Prediction of Electronic Coupling between the Guanine Bases of DNA. *The Journal of Physical Chemistry A* **2020**, *124*, 7658–7664.

[77] Krämer, M.; Dohmen, P. M.; Xie, W.; Holub, D.; Christensen, A. S.; Elstner, M. Charge and Exciton Transfer Simulations Using Machine-Learned Hamiltonians. *Journal of Chemical Theory and Computation* **2020**, *16*, 4061–4070.

[78] Emonts, J.; Buyel, J. An overview of descriptors to capture protein properties – Tools and perspectives in the context of QSAR modeling. *Computational and Structural Biotechnology Journal* **2023**, *21*, 3234–3247.

[79] Gillet, N.; Elstner, M.; Kubař, T. Coupled-perturbed DFTB-QM/MM metadynamics: Application to proton-coupled electron transfer. *The Journal of Chemical Physics* **2018**, *149*, 072328.

Chapter 4

Understanding the nucleation mechanism assisted by lanthanides complexes

Involved People

Alessio Bartocci (Post-Doc), John Carlos (L3), Thomas Boukéké-Lesplulier (Master2), Rishab Panda (Master)

Related Article

- A. Bartocci, N. Gillet, T. Jiang, F. Szczepaniak, E. Dumont, *J. Phys. Chem. B*, **2020**, *124*, 11371

- A. Roux, Z. Alsalman, T. Jiang, J-C. Mulatier, D. Pitrat, E. Dumont, F. Riobé, N. Gillet, E. Girard, O. Maury *Chemistry. A European Journal*, **2024**, *30*, e202400900

- K. Dos Santos, A. Bartocci, N. Gillet, S. Denis-Quanquin, A. Roux, E. Lin, Z. Xu, R. Finizola, P. Chedozeau, X. Chen, C. Caradeuc, M. Baudin, G. Bertho, F. Riobé, O. Maury, E. Dumont and N. Giraud, *Phys Chem Chem Phys*, **2024**, *26*, 14573

4.1 Introduction

Nowadays, most of the current protein 3-dimensional structures available in the Protein Data Bank belong to X-Ray crystallography experiments (more than 187,000 over the 224,000 available). This large data set, used to feed the artificial intelligence algorithms such as AlphaFold2 [1] or RoseTTaFold,[2] represents about only 10% of the identified protein sequences (more than 245,000,000 unreviewed entries on UniProt). These predictive algorithms have become staple tools in structural biology but a consistent feeding must be kept. Experimental data remain crucial to cover the large part of unknown structures and the protein complexes. However, if the X-Ray techniques have evolved to be more automatized and benefit from synchroton ray-lines improvement, protein crystal, or micro-crystals,[3] are still required to provide a relevant structure. The crystal successful production remains uncertain and often requires tests of a large amount

of crystallographic conditions. The molecular mechanism of nucleation is still partially unveiled and the role of the different additive in the crystallisation medium difficult to rationalize.

In order to bring some rational design to crystallization protocol, some groups have developed molecular additives which aim to improve the crystal formation with minimal modification of the protein structure. They are called "molecular glue" as they are suspected to play the role of a cement between the proteins during the nucleation thanks to supramolecular interactions networks. Three main ways have been explored using polyoxometallates,[4] negatively charged calixarene[5] or lanthanide complex.[6] At the Laboratoire de Chimie, Oliver Maury and co-workers have designed the crystallophore (Xo4, Figure 4.1 **A**) which improves the resolution thanks to the heavy lanthanide properties for phasing and the protein crystal formation.[7] The promising properties of this complex has led to the creation of a startup, called Polyvalan (https://crystallophore.fr), and the founding of several ANR projects. A growing consortium including NMR, computational chemistry and cellular biology has emerged around Xo4 in addition to the initial synthesis-crystallography collaboration. Moreover, new crystallophores have been designed, with different properties for imaging for instance or better nucleating effect (see Figure 4.1 **B**, **C**).

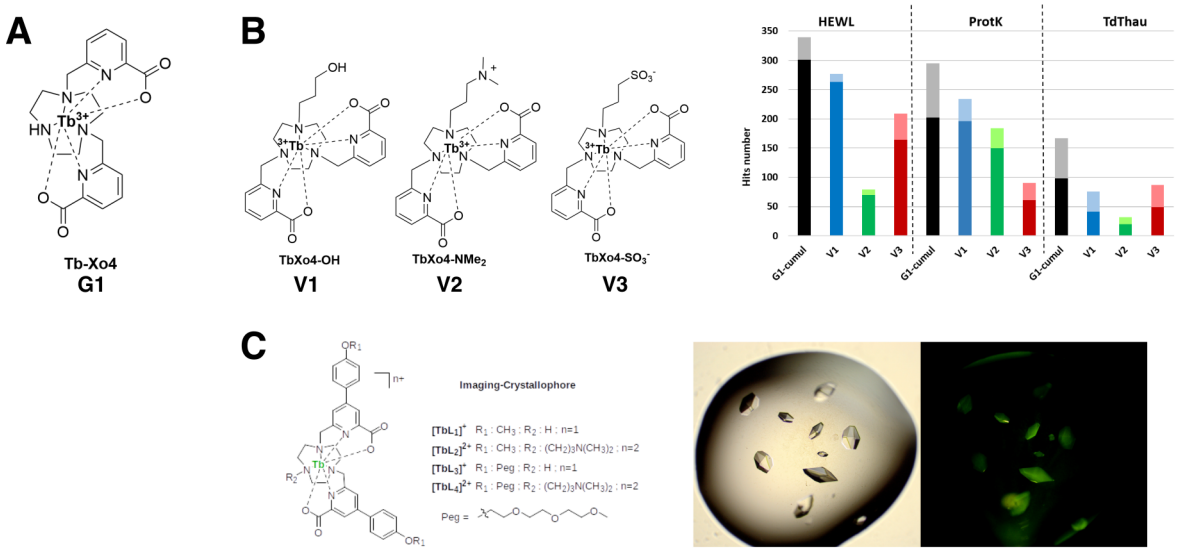

Figure 4.1: **A**. Xo4 complex. **B**. Second generation of Xo4 complexes and their crystal hits for chicken lysozyme (HEWL), Proteinase K (ProtK) and Tau Protein (TdThau).[8] **C**. imaging Xo4 and the luminescent HEWL crystal containing it.

The contribution of computational chemistry can be multiple:

- quantify the interaction energy between the Xo4 and the protein crystal packing using QM/MM calculations[7] or cheaper ProLIF[9] or MM/PBSA approaches.[10,11]
- simulate the interaction between Xo4 and the targeted protein in solution and compare it to crystallographic structure.
- decipher the contribution of the Xo4 to the nucleation mechanism in solution.

Once the computational protocol established for a given protein-Xo4 couple, we can apply it to different proteins, protein variants or other molecules from the Xo4 family to study the impact of binding site mutations or added chemical groups to the lanthanide complex. Beyond the understanding of the interaction landscape in solution for crystallography[8] or NMR,[12] our approach aims to become a predictive tool and to be extended to other lanthanide complexes such as contrast agents for IMR.

4.2 Molecular Dynamics framework for Xo4-protein interaction sampling

Protein-ligand interaction studies are mostly based on docking and/or MD simulations.[13] These simulations can start from an identified protein-ligand complex or separated entities in a solvent box. The first case is useful to quantify the interactions energies, reach the best conformation for the ligand in its pocket or to start biased MD simulations[14,15] to remove the ligand from its binding site and evaluate the corresponding binding free energy.[16] In my group, we applied such methods also to study the interaction between ligands and aptamers[17] or G-quadruplexes.[18] We also analysed the interaction of different complexes of the Xo4 family within the HEWL packing model, as described below.[8] The second kind of simulations lets the ligands explore the protein surface and unveil the driving force of the receptor-ligand interaction in solution. We applied it to the study of calixarene interaction with Cytochrome *c* protein and observed similar binding site as in the experiments and crystallographic data.[19] In addition, we are able to describe a possible entropy reduction following the calixarene ligation during our simulation by the formation and stabilization of hydrogen bond and salt bridge network.

In order to analyse the interactions between molecules from the Xo4 family and proteins, we have developed the following protocol:

- Xo4 force field parameters calculation using MCPB program in Amber.[20] During the required quantum calculations, the lanthanide is replaced by an Yttrium. This atom conserves a similar size and charge as lanthanide but avoids the spin multiplicity issues from f-orbital. This is not an problem as this electronic properties is not described by molecular mechanisms.
- MD simulation including a protein and several Xo4 (10 in general) in water solutions. We also perform simulations with more proteins and various number of crystallophore
- interaction analysis using ProLIF facilities[9] and MM-PBSA fast calculations to identify the important residues at play in the Xo4 binding.

For all the simulations, 5 replicas of at least 200 ns have been run following the conventional minimization-heating-equilibration-production procedure, at 300 K, 1 bar and using periodic conditions. The protein are described using Amberff14SB[21] force fields and the TIP3P water model[22] is used. Up to now, these simulations and their current analysis do not represent a challenge or a novelty with respect to the computational methodology.

4.3 Reveal solution binding site in AdkA protein

Adelynate kinase A from *Methanothermococcus thermolithotrophicus* (AdkA) protein is a trimeric protein with a high number of negatively charged residues.[23] Based on common electrostatic consideration, one can expect that protein carboxylate residues are well adapted to bind the positively charged lanthanide within the Xo4. In the obtained X-Ray structure, one Xo4 carboxylate is bound to a buried aspartate, D90, *via* a Mg^{2+} cation which is a quite unexpected and complex binding site. Some trace of electronic density has been found on the protein surface around the glutamate E50.

Starting MDs from the crystallographic structure put in a water box, we observe a quick release of the Xo4 from its binding site in few ns. The lanthanide complex moves towards the carboxylates of the protein surfaces such as E45, E49, E54, E136, E149 and D181 (Figure 4.2 **A** and **B**). In solution, where the Xo4 surrounding differs from the crystal packing environment, the Xo4 seems to favor a direct interaction between a protein carboxylate and its lanthanide. When the starting point of the MD contains one trimer and 10 Xo4 randomly positioned in the water box, the simulations result in fast and strong binding of the Xo4 on the protein surface at different sites: E29, E49, E54, E136, E185. Several of these sites consist in a pair of acidic residues that increase the negative electrostatic potential. R140 and Y182 can also participate to the binding stabilization by salt bridge formation or π-stack interaction respectively with the Xo4 ligand. All these sites are located on the "lateral" face of the AdkA trimer ring, which also corresponds to the protein-protein interface in the crystal packing.

Although aspartates represent one third of the acidic residues of AdkA, the Xo4 interacts almost exclusively with glutamates which possess a longer aliphatic chain and can be more flexible. We thus test if correlations exist between the rate occupancy of acidic residues and their RMSF, their solvent accessible surface area (SASA) or their electrostatic interactions with the protein (Figure 4.2 **C**). It appears that residues which have more interactions with the Xo4 are those with the less attractive interaction with the protein (so not involved in a salt bridge and surrounded by other negative charged) and with a relatively important SASA. The flexibility related to the RMSF seems less crucial.

Our computational study highlights the difference between the Xo4 behaviour in solution and its final binding in the crystallographic structure. Our results are in agreement with the

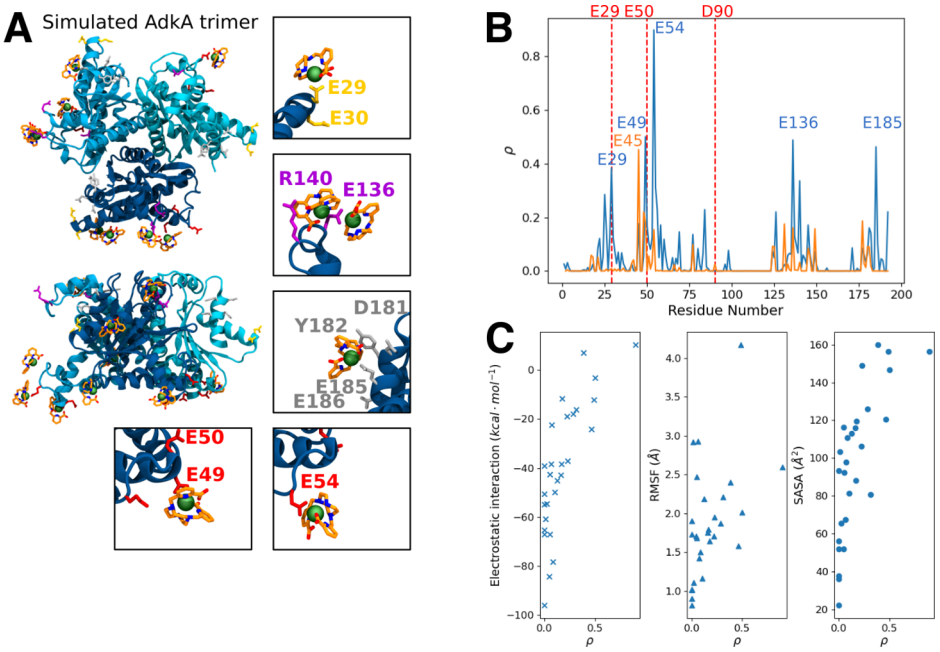

Figure 4.2: **A**. Different representative binding sites encountered during MD simulations started from solvated Xo4. **B**. Per-residue normalized contact probability ρ starting from the crystallographic (yellow, 2 Xo4 in the box) or the solvated positions (blue, 10 Xo4 in the box) of Xo4 **C**. Correlation graph between the rate occupancy ρ and the per-residue electrostatic interaction with the protein, RMSF or Solvent Accessible Surface Area (SASA).

weak signal observed close to E50 in crystallographic experiments and with the fact that the mutation of D90 or E50 does not impede the crystallization process: indeed, the numerous possible Xo4 binding sites at the protein surface and the fact that they often correspond to a pair of acidic residues maintain the presence of the Xo4 at the AdkA surface despite the mutations. The presence of the Xo4 on the protein surface decreases its global charge and probably facilitates protein-protein interactions leading to nucleation. This study is the first step to understand the nucleation mechanism assisted by Xo4. It underlines the power of MD simulations to provide interesting features for the understanding of Xo4-protein interactions in solution.

4.4 How mutation modifies Xo4 interactions in lysozyme family: an egg hunting story

One of the first protein used to test the Xo4 propensity to enhanced crystal formation is the Hen Egg White Lysozyme (HEWL). This cheap, relatively small (only 129 amino-acids) and easy to crystallize protein is a good toy model to test both experimentally and computationally the nucleating properties of the Xo4. A Xo4 binding site has been characterized in the crystallographic structure and correspond to the lysozyme substrate pocket. The complex is fixed by apolar and

π-stacking interactions with W62 and W63, and by the ionic interaction between D101 residue and the metal as demonstrated in a first study.[7] In this previous work, the binding energy has been calculated at QM/MM level on the crystallographic structure. Using the same philosophy as in the AdkA study, we want to analyze the Xo4 binding to HEWL in solution without any starting interactions. In addition, we simulate different variants of this lysozyme following the same protocol. The protein variants are chosen as they present a high global homology degree and a mutation in D101 area (see Figure 4.3 **A**). The experimental counterpart is not easy: lysozymes cannot be easily obtained by recombinant production, but they are extracted from eggs. Consequently, it depends on the availability of the chicken, quail (QEWL), turkey (TEWL) or peacock (PEWL) eggs. In our computational studies, for ostrich (OEWL) and turtle (STEWL) variants, we have mutate the residues from the available HEWL structure, assuming that the few mutations will not dramatically modify the protein folding.

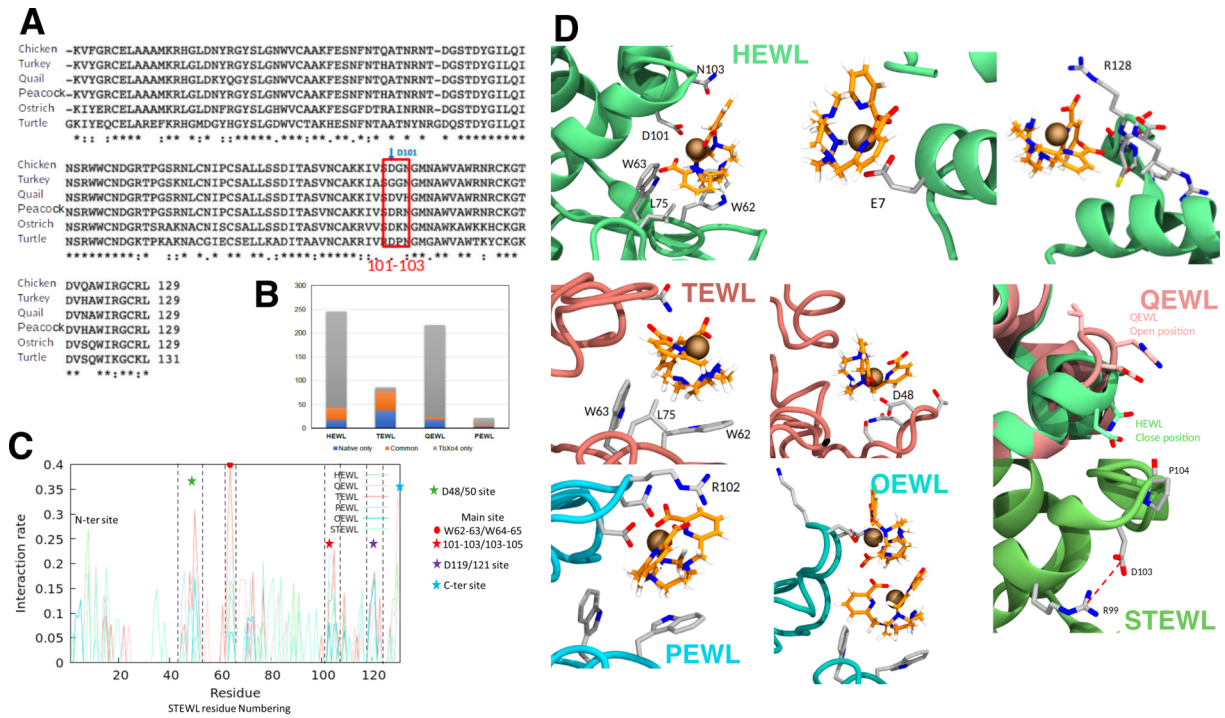

Figure 4.3:
textbfA. Sequence alignment of the different lysozyme variants from Chicken (HEWL), Turkey (TEWL), Quail (QEWL), Peacock (PEWL), Ostrich (OEWL) and Turtle (STEWL). The important mutations around the D101 site are highlighted by the red square. **B**. Experimental crystallization hits for HEWL, TEWL, QEWL and PEWL encountered in absence of Xo4 only (blue), in presence of Xo4 only (grey), or in both conditions (orange). **C**. Lysozyme residues interaction rates with Xo4 from simulation.**D**. Conformation of the bindings site for the different lysozyme variants.

Experimentally, at 10 mM, the Xo4 also enhances the number of crystallization hits for QEWL but does not improve the TEWL results. PEWL has a poor score of crystallization hits in absence or in presence of Xo4 adduct.[24] The crystallographic structure of QEWL is very similar

to the HEWL ones. Two Xo4 are found in PEWL binding site, one in similar conformation as in HEWL, the other one interacting with R102. In TEWL, no Xo4 is observed within the D101-W62/63 binding site. In parallel, the rate occupancy from MDs simulation obtained by ProLIF analysis is given in Figure 4.3 **B** for the different lysozymes. We observe that the interaction between the Xo4 and W62, W63 and L75 is present even in absence of D101 in TEWL, or in non-availability in STEWL where a salt bridge is formed between D103 and R99 (Figure 4.3 **C**). In QEWL, the mutations of 102 and 103 residues lead to a more flexible 101-103 loop. Consequently, the aspartate is not always pointing through the main binding pocket. In this second conformation, the interaction between D101 and Xo4 is not maintained. Such conformational change is also observed in OEWL simulations but leads then in the interaction with two Xo4, one within the pocket and W62-W63 and the second one with D101, N103 and the first Xo4. Finally, in PEWL, a Xo4 can be found in similar conformation as in HEWL, but with an interaction between its carboxylates and R102. No other Xo4 comes close. In addition, several other binding sites can be characterized at the N- and C-terminal ends (spatially close) and the loop comprising D48, and around D119 on the last α-helix.

Taking together these results, we can draw several conclusions:

- our simulations, totaling 1 μs, are able to recover the crystallographic sites for the different lysozyme and described different behaviour of the mutated residues. However, because of the limit of our conformational sampling, the comparison of the occupation of each site stays qualitative. Indeed, the protein-Xo4 interaction remains a transient and relatively rare event in our simulations. The convergence of the MDs simulations with respect to these events is certainly far from being reach.

- despite similarities, our MDs data reveal some different behaviour in solution and in crystal, suggesting that other sites can be occupied. An insight to the crystallographic packing shows that some residues from these secondary sites are close to the Xo4 at the protein interface (such as D119 or the terminal part).

- all lysozymes can interact with the Xo4 in solution but, experimentally, not all present a high number of crystal hits. Consequently, it seems that not all Xo4-protein complex can favor nucleation. Other molecular mechanisms must be at play to stabilize the protein-protein interactions.

4.5 Crystallophore variant

If the Xo4 increases the number of possible crystallization conditions for HEWL, it appears less efficient for some other proteins. However, its properties can be modulated by chemical modifications of the macrocycle: a pendant arm can replace the hydrogen of the free nitrogen of the

macrocycle. This arm can be synthesized with different ending moieties: polar (OH, **V1**), positively charged (MethylAmine, **V2**) or negatively charged (Sulfonate **V3**).[8] Experimentally, using the same concentration of Xo4, **V1** presents similar nucleating properties as the first generation Xo4, while **V3** and **V2** give much less monocrystal hits. In the crystallographic structure, the pendant arm induces a rotation of the macrocycle about 90° which modifies the interaction with the tryptophans 62 and 63 of the binding pocket but keeps the coordination of the lanthanide by D101. This rotation also prevents π-stacking with the C-terminal W123 of another HEWL of the packing (HEWL$_C$ in the following).

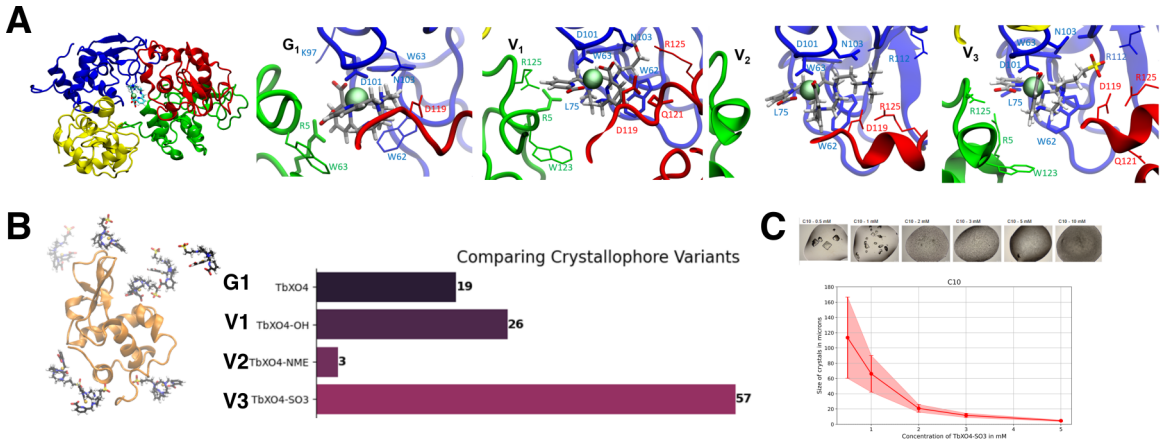

Figure 4.4: **A**. Representation of the crystal packing model with 4 HEWL and a Xo4 bound to the blue D101. Zoom on the binding site for the different variants **B**. Several **V3** complexes around HEWL and contact quantification between Xo4 variants and HEWL for the simulations with a 10:1 Xo4:protein ratio. **C**. Evolution of the crystal size with the increase of **V3** concentration and associated crystal pictures.

Starting for a crystallographic packing model with 4 HEWL and a Xo4 variant (Figure 4.4 **A**), we simulate the interaction profile of the different ligands within this tetramer in solution. Two Xo4 are necessary to maintain the tetrameric structure over 200 ns but we focus on the central one which can interact with 3 of the 4 lysosymes in the crystallographic model. During the MD simulations, the interaction profile evolves with respect to the nature of the pendant arm: **V1** has a versatile behaviour and explores several transient interactions, within the binding site or the C-term part of the second HEWL, HEWL$_B$; **V2** should prefer positively charge residues but does not keep its interaction with HEWL$_B$ and prefers staying close to N103 in the binding pocket; **V3** interacts with the positive charges around the binding site, namely R112 from HEWL$_A$ and R125 from HEWL$_B$ in a lesser extent, inducing different interactions between the macrocyle and the surrounding tryptophan. According to MM-PBSA results, these interactions profiles result in a similar binding energy for first generation Xo4 and **V1** and a ca. 20 kcal/mol less favorable interaction with **V2** and **V3**.

However, the previous picture corresponds to the interaction after nucleation but not to the Xo4 and its variant behavior in solution. Following the same protocol as for the lysozyme vari-

ants, we simulate the interaction of one HEWL with **V1**,**V2** and **V3** in solution with a 10:1 ligand:protein ratio. These simulations bring insightful information with regards to the different nucleating properties: while **V1** arbors same behavior as Xo4, **V2** does not interact with the HEWl, in agreement with the experiment. However, numerous interactions are observed between **V3** and HEWL: like AdkA, HEWL is able to bind several **V3** at the same time. Then, it appears that such a surface coverage can impede the nucleation (Figure 4.4 **B**). It leads us to the idea to decrease the **V3** concentration during crystallization experiments. A protocol was designed to quantify the size of the crystals as a function of **V3** concentration (see Figure 4.4 **C**). Even with a small amount of **V3**, for instance at *0.5 mM* which corresponds to a 20 fold decrease compare to the previous tested concentration, macro-crystal are obtained. Their size tends to decrease with respect to the **V3** concentration, in line with the hypothesis that the number of nucleation sites with a fast crystal growth increases, leading to smaller crystal for a constant amount of proteins. Consequently, **V3** reveals more efficient with HEWL than the first generation Xo4.

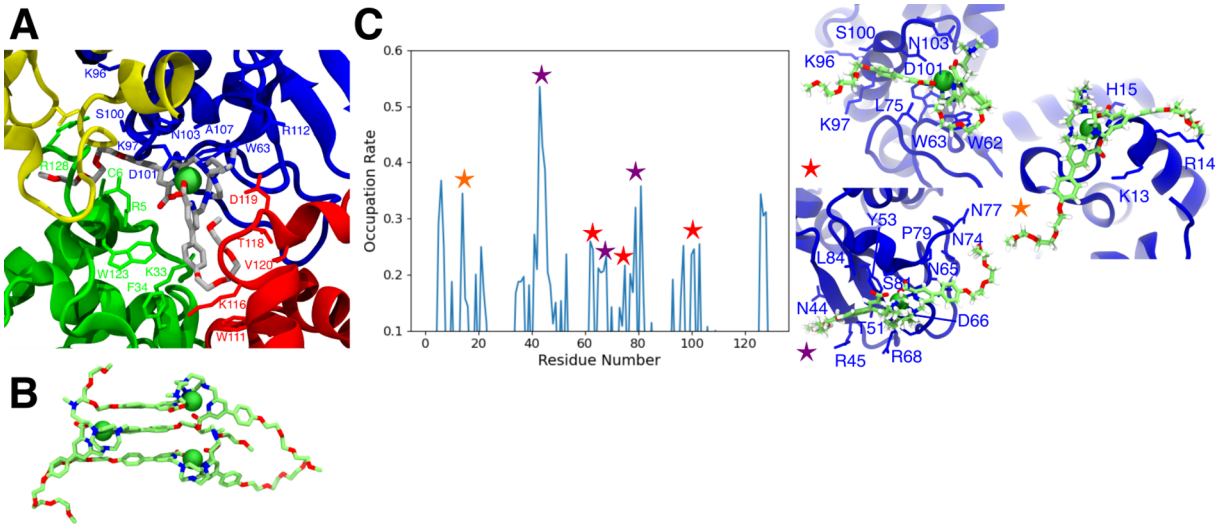

Figure 4.5: **A**. Zoom on the binding site of the imaging Xo4 in the crystal packing model. **B**. Representation of an aggregate of 3 imaging Xo4 in solution. **C**. Interaction rate of the imaging Xo4 with HEWL residues in the simulations containing 10 imaging Xo4 for one HEWL and the corresponding binding sites.

A similar protocol was applied to a fluorescent ligand, so called imaging Xo4, which facilitates the detection of the crystal by spectroscopy (see Figure 4.1 **C**). Such a property can be of particular interest in the *in vivo* protein crystallization for protein purification. The development of this approach is the heart of the GlowCryst ANR project led by Dr. Maury: the presence of a luminescent complex would allow a easy detection of the crystals within a complex medium. However, *in vitro* crystallization experiments show that crystal formation with this crystallophore variant is not as good as with the Xo4 first generation. An fruitful idea was thus to mix it these two complexes in a 10 mM/0.2 mM ratio, which provides similar results as the pure Xo4 addition to HEWL crystallization medium and conserves the spectroscopic properties of the imaging Xo4.

This suggests that the imaging Xo4 can be trapped in the solid structure during the nucleation and/or the crystal growth as its size prevents any diffusion within the crystal packing.

The force field parameterization of the imaging Xo4 requires a specific focus on the antenna dihedral angle. Indeed, the bond between the two aromatic ring has a partial aromatic character which competes with the steric hindrance between the two rings. The classical modeling is unable to describe correctly such behaviour. We thus determine a dedicated force field parameter set for this dihedral angle to reproduce the DFT scan of the rotation provided by Pr. Le Bahers. After checking its behaviour in solution, we create a crystal packing model containing the imaging Xo4 from the **V2** structure. We also present MD simulations in solution with one HEWL protein and 10 Imaging Xo4 or a mixture of Xo4 and Imaging Xo4 in a 9:1 ratio.

In our simulations of the crystal packing situation, the imaging Xo4 stays at the interface of the four proteins. While the metal is bound to D101 of the first HEWL, the pendant arm is able to interact through ionic or hydrogen bonds with the C-terminal part of the second protein (in red) (Figure 4.5 **A**). Most of the antenna chain shows only apolar interactions with the protein surface. However, the macrocycle lies outside from the hydrophobic HEWL pocket and does not interact with W62 and W63. This interaction scheme lead to a MM-PBSA binding energy 10 kcal/mol more repulsive than the HEWL tetrameric complex-Xo4 one, indicating that this complex is less bound than Xo4 in the crystal structure. Its behaviour in solution is also intriguing as 2 to 8 imaging Xo4 form aggregates thanks to *pi*-stacks between the antennae during our simulations (Figure 4.5 **B**). These aggregates occur in solution but also on the surface protein, when one imaging Xo4 is already bounded on the protein surface. Because of its arm, imaging Xo4 covers a relatively large part of the protein when interacting (Figure 4.5 **C**). Even if we observe interactions between the metal and D101, other binding sites are occupied showing a relatively poor specificity leading to the proximity of one to 4 imaging Xo4 to the protein. Because of the large size of the complex, its bulky binding landscape can hinder the protein-protein approach necessary for nucleation. On the contrary, when only one imaging is present in solution, we can still observe binding of Xo4, especially in the D101 pocket such as in previously described simulations. Only few Xo4-Imaging Xo4 interactions are transiently observed.

In conclusion, our simulations framework successfully provides information for the understanding of the different nucleation power within the Xo4 family. The structural and dynamical comparison can partially explain the crystal formation hits but also drive experiments and protocol modifications as for the variant **V3**. Then, the predicting power of our protocol must be strengthened by testing it on other proteins which present different crystallization profile in the presence of the Xo4 family. They are currently tested by Dr. Roux Gossart from Polyvalan.

4.6 Perspective

Despite the large amount of trajectories exploring the Xo4 family-protein (mostly lysozymes) interaction landscapes the brigde between solution and crystal remains to build. We have drawn a rough sketch of the Xo4 molecular glue properties beyond the crystallographic structure. During its internship, Mr Panda has performed a large panel of MD simulation with several lysosymes and different ratio of Xo4. His results suggest that the protein-Xo4 interactions induce a decrease in the protein dynamical fluctuations and flexibility, so likely an entropy decrease. This decrease may favor the protein-protein interactions. The role of the Xo4 may be to tune the protein dynamical behaviour to stabilize the protein-protein interaction. On the contrary of what the word "glue" suggests, Xo4 may not directly contribute to bridge the protein at their interfaces. Using 6 HEWL, he obtained different aggregates in presence or absence of Xo4 even if the lanthanides complexes quickly flee from protein to solution. A further analysis of these MD simulations would permit us to refine our first sketch for Xo4 role in the nucleation mechanism. Moreover, the application of our protocol to a larger panel of protein with different crystallization score in presence of Xo4, in collaboration with Dr. A. Roux Gossart from Polyvalan, will also complete our picture of the ability of Xo4 to stack on the protein surface thanks to different kind of non-covalent interaction (ionic, but also apolar or π-stacking) and to induce nucleation.

Such a knowledge will open the door to computationally driven design and selection within the Xo4 family for a better fitting on the protein-Xo4 complex. The context of the ANR Glow-Cryst, led by Dr. O Maury at the Laboratoire de Chimie, will offer the opportunity at a short term to further explore the interaction between Imaging Xo4 family and proteins in the context of *in cellulo* crystallisation process for protein purification. Besides, I wish to strengthen my collaboration with Polyvalan to provide a rationalize tool for the development of new complexes and their best usage considering a target protein. Consequently, I will continue the current projects and propose new ones to :

- complete our database on protein-Xo4 family interactions with one or several proteins and quantify the different non-covalent interactions which drive the complex formations (short term). This will help us to understand which characteristic of the protein surface (electrostatic potential, polarity, flexibility...) should be taken into account in the choice of the most adapted Xo4 for crystallization.

- develop a tool to characterize the Xo4-protein interactions beyond a distance criteria. It must be coupled to the study of the impact of the Xo4 bonding to the protein dynamical behaviour and flexibility. A final goal is to be able to predict Xo4 binding site for nucleation on a given protein surface.

Bibliography

[1] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

[2] Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876.

[3] Sauter, C.; Housset, D.; Orlans, J.; de Wijn, R.; Rollet, K.; Rose, S. L.; Basu, S.; Bénas, P.; Perez, J.; de Sanctis, D.; Maury, O.; Girard, E. Nucleating Agent Crystallophore Induces Instant Protein Crystallization. *Crystal Growth & Design* **2024**, *24*, 6682–6690.

[4] Breibeck, J.; Bijelic, A.; Rompel, A. Transition metal-substituted Keggin polyoxotungstates enabling covalent attachment to proteinase K upon co-crystallization. *Chemical Communications* **2019**, *55*, 11519–11522.

[5] Crowley, P. B. Protein–Calixarene Complexation: From Recognition to Assembly. *Accounts of Chemical Research* **2022**, *55*, 2019–2032.

[6] Engilberge, S.; Riobé, F.; Pietro, S. D.; Lassalle, L.; Coquelle, N.; Arnaud, C.-A.; Pitrat, D.; Mulatier, J.-C.; Madern, D.; Breyton, C.; Maury, O.; Girard, E. Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence. *Chemical Science* **2017**, *8*, 5909–5917.

[7] Engilberge, S.; Riobé, F.; Wagner, T.; Di Pietro, S.; Breyton, C.; Franzetti, B.; Shima, S.; Girard, E.; Dumont, E.; Maury, O. Unveiling the Binding Modes of the Crystallophore, a Terbium-based Nucleating and Phasing Molecular Agent for Protein Crystallography. *Chemistry – A European Journal* **2018**, *24*, 9739–9746.

[8] Roux, A.; Alsalman, Z.; Jiang, T.; Mulatier, J.-C.; Pitrat, D.; Dumont, E.; Riobé, F.; Gillet, N.; Girard, E.; Maury, O. Influence of Chemical Modifications of the Crystallophore on Protein Nucleating Properties and Supramolecular Interactions Network. *Chemistry – A European Journal* **2024**, *30*, e202400900.

[9] Bouysset, C.; Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints. *Journal of Cheminformatics* **2021**, *13*, 72.

[10] Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **2015**, *10*, 449–461.

[11] Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **2012**, *8*, 3314–3321.

[12] Santos, K. D. et al. One touch is all it takes: the supramolecular interaction between ubiquitin and lanthanide complexes revisited by paramagnetic NMR and molecular dynamics. *Physical Chemistry Chemical Physics* **2024**, *26*, 14573–14581.

[13] Salmaso, V.; Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Frontiers in Pharmacology* **2018**, *9*.

[14] Do, P.-C.; Lee, E. H.; Le, L. Steered Molecular Dynamics Simulation in Rational Drug Design. *Journal of Chemical Information and Modeling* **2018**, *58*, 1473–1482.

[15] Wang, J.; Arantes, P. R.; Bhattarai, A.; Hsu, R. V.; Pawnikar, S.; Huang, Y.-m. M.; Palermo, G.; Miao, Y. Gaussian accelerated molecular dynamics: Principles and applications. *WIREs Computational Molecular Science* **2021**, *11*, e1521.

[16] Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annual Review of Biophysics* **2017**, *46*, 531–558.

[17] Allahkaram, L. Modeling the dynamics and properties of key DNA lesions intermediates in realistic environments. These de doctorat, Lyon, École normale supérieure, 2023.

[18] Deiana, M. et al. A new G-quadruplex-specific photosensitizer inducing genome instability in cancer cells by triggering oxidative DNA damage and impeding replication fork progression. *Nucleic Acids Research* **2023**, *51*, 6264–6285.

[19] Bartocci, A.; Gillet, N.; Jiang, T.; Szczepaniak, F.; Dumont, E. Molecular Dynamics Approach for Capturing Calixarene–Protein Interactions: The Case of Cytochrome C. *The Journal of Physical Chemistry B* **2020**, *124*, 11371–11378.

[20] Li, P.; Merz, K. M. J. MCPB.py: A Python Based Metal Center Parameter Builder. *Journal of Chemical Information and Modeling* **2016**, *56*, 599–604.

[21] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

[22] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

[23] Vonrhein, C.; Bönisch, H.; Schäfer, G.; Schulz, G. E. The structure of a trimeric archaeal adenylate kinase1. *Journal of Molecular Biology* **1998**, *282*, 167–179.

[24] Alsalman, Z. Développement et caractérisations de complexes de lanthanide pour la biologie structurale. These de doctorat, Université Grenoble Alpes, 2023.

Chapter 5

Conclusion

Since my PhD, I have focused my research on the question of the influence of biochemical structure and more specifically non-covalent interactions on the chemical properties of biomolecule with a more specific interest on charge transfer, by means of molecular dynamics at classical and QM/MM levels. I am strongly attached to computation-experiments ping-pong for the mutual advantages that it brings in terms of results validation and methodology improvement. My concern is to unveil the molecular behaviour behind an experimental observation and propose mechanisms that can explain and predict biochemical events.

For 5 years, I have mostly dedicated my personal research on the nucleosome, which is a fascinating object with a highly combinatorial chemistry. The DNA sequence, the mechanical constrains, the heterogeneous electrostatic environment, the solvent and protein accessibility of the nucleobases and the fluctuating interactions with the histone tails can play significant role on the DNA chemical properties and sensitivity to stresses that multi-scale simulations help to decipher. I have presented in this document how I try to tune both computational and analysis protocols to manage this complexity. Strong collaborations with experimental group help me to raise relevant questions with respect to DNA damages and further understanding of cancer formation and therapy. This research benefits first from the expertise of Pr. E. Dumont, who introduced me in the DNA world, and then from two founded projects, my JCJC ANR NucleoMAP and the PCSI project BERNUMOL.

The Laboratoire de Chimie offers me an exceptional environment to interact with chemists who develop molecules for biochemical or biomedical applications. Simulations appear useful to describe the non-covalent interactions that can explain the effect of these molecules on the biological target. For instance, we are able to propose an interaction scheme between a candidate molecule for photodynamical therapy and different shape of G-quadruplexes. The computational study of crystallophore family has become a key step in the characterisation of these molecules and the design of new compounds thanks to its ability to decipher its interactions with protein in solution. It participates to understand the premises of the nucleation mechanism or the different

interaction landscapes in combination with the various experimental data (spectroscopy, X-Ray crystallography, NMR...). This work takes part of a large consortium and benefits from the complementary skills of many partners and diverse funding such ANR projects (for instance GlowCryst project).

Despite their differences in term of systems and chemical issues, these two research topics has raised the question of the analysis protocol of long simulations of highly complex systems such as nucleosome or multiple proteins-ligands structures. The flexible and transient character of the non-covalent interactions at the heart of the determination of chemical properties (such as charge transfer ability or binding energy) impedes the systematic capture of the interaction network. I consider the development of dedicated protocols, on the basis of available tools and/or machine learning algorithm trained of our large amount of molecular dynamics data, as a key of my future work on these topics.

Finally, during the last 5 years, I had the opportunity to supervise, advise and support several students from L3 to M2, PhD students and post-doctoral fellows who offer me great scientific and human exchanges. All the studies presented here belongs from their work and their involvement in the research projects which has always required to go beyond their initial formation because of the multidisciplinary aspect of the project, between chemistry, biology, physics and computation... I hope that I convinced, through this manuscript and the presentation of this team's work, my ability to lead my own research, keeping the strength of strong collaborations with developers and experimentalists and opening the way to new opportunities.

Abstract

Beyond DNA or protein sequences, non-covalent interactions drive the specificity and the diversity of macromolecules in all organisms. They are involved in the 3-dimensional structure and the dynamics of biomolecules, in the interactions between them etc. and tune their chemical properties. For few decades, simulations at molecular level have become a crucial tool to decipher the role of non-covalent interactions in the structural and chemical properties of biomolecules. They have many applications in the understanding of biological processes and the design of new compounds able to trigger biomolecules behaviour, chemistry and functions. In this manuscript, I summarize the research I have conducted since my arrival at the Laboratoire de Chimie de l'ENS de Lyon. It consists in the study of different biomolecular systems by means of molecular dynamics simulations at classical and quantum/classical level. On one hand, such approaches allow to draw a panorama of the non-covalent interactions between different partners at a molecular level. On the other hand, we are then able to quantify how these interactions and their dynamics impact the chemical properties of biomolecules. I detail three main topics: i) the structural and dynamical behaviour of DNA-damages, especially those embedded in a nucleosomal structure, which consists in a DNA fragment wrapped around a protein core; ii) the charge transfer properties of the nucleosomal DNA; iii) the interaction landscapes of a lanthanide complex and proteins in the context of the control of protein crystallization. I also present my future research projects in line with these different topics and beyond.