

# Méthodes pour l'étude de structures, dynamiques et interactions d'acides nucléiques non canoniques

**Eric Largy**

ORCID : 0000-0002-6140-9788

Laboratoire ARNA, INSERM U1212, CNRS UMR5320, Université de Bordeaux

**Mémoire en vue de l'obtention d'une  
Habilitation à diriger des recherches**

Soutenance le 19 décembre 2025

## **Jury :**

Pascale Changenet	Directrice de recherche CNRS, École Polytechnique	Rapporteur
Brahim Heddi	Chargé de recherche HDR CNRS, ENS Paris-Saclay	Rapporteur
Sylvain Ladame	Reader, Imperial College London	Rapporteur
Valérie Gabelica	Professeure, Université de Genève	
Karen Gaudin	Professeure, Université de Bordeaux	
Carmelo Di Primo	Chargé de recherche HDR INSERM, Université de Bordeaux	

Manuscrit rédigé en [FiraGO](#) et [Fira Math](#)

Fichier source préparé avec Quarto 1.7.34, Python 3.12.7, R 4.5.1.

Version pdf pour l'impression générée avec  $\text{\LaTeX}$  (LuaHBTeX 1.18.0) et compressée avec `ghostscript`  
10.06.0 pour le web.



A Mumu, que soube me fazer entender que as páginas que seguem não importam muito.

Il y a quelques fautes d'orthographe, que le  
Lecteur corrigera s'il luy plaist.

---

*Traité de mécanique des poids soustenus  
par des puissances sur des plans inclinez à  
l'horizon, 1636*

Gilles Personne de Roberval

GARDNER CHUBB  
What did we learn, Palmer?

PALMER  
I don't know, sir.

GARDNER CHUBB  
I don't fucking know either. I guess we  
learned not to do it again.

PALMER  
Yes, sir.

GARDNER CHUBB  
Although I'm fucked if I know what we did.

PALMER  
Yes sir. Hard to say.

---

Joel Coen et Ethan Coen  
*Burn After Reading*, 2008

## Remerciements

J'ai eu la chance d'évoluer dans des environnements stimulants, aussi bien académiques que privés, et de collaborer avec des chercheurs talentueux. Je tiens à remercier chaleureusement toutes celles et ceux qui m'ont accompagné et soutenu tout au long de ce parcours. Beaucoup sont mentionnés dans ce manuscrit, mais je souhaite également inclure les personnels administratifs, les collègues enseignants-chercheurs, les ingénieurs de plateformes, et plus généralement toutes les personnes dont j'ai croisé la route et qui ont, de près ou de loin, contribué à la réussite de ce travail.

J'ai eu la chance de bénéficier de plusieurs mentors d'exception au cours de ma carrière. Leurs conseils avisés et leur confiance m'ont profondément marqué. Ils m'ont souvent accordé plus de crédit que je n'en avais moi-même, et j'espère avoir été à la hauteur de leurs attentes.

J'ai moi-même eu l'opportunité de superviser des étudiants, et j'espère avoir su leur transmettre une partie de ma passion pour la recherche. Plusieurs ont poursuivi des carrières en recherche, et je suis fier d'avoir pu contribuer à leur développement professionnel, à ma petite échelle. Au-delà de leurs contributions aux projets de recherche, leur enthousiasme et leur curiosité ont été une source d'inspiration constante.

Depuis le début de ma thèse, en octobre 2008, jusqu'à aujourd'hui, j'ai beaucoup appris, et j'espère continuer à apprendre encore longtemps. Merci à toutes celles et ceux qui l'ont rendu possible et le rendent encore aujourd'hui.

Je remercie également les membres du jury pour le temps qu'ils consacreront à la lecture de ce manuscrit et pour leur présence lors de la soutenance, juste avant les vacances de Noël, ce qui mérite d'autant plus ma gratitude.

Enfin, un grand merci à ma famille pour leur soutien indéfectible et leur patience.

# Sommaire

<b>Introduction</b>	<b>1</b>
<b>Curriculum Vitae</b>	<b>9</b>
<b>1 Doctorat à l'Institut Curie/CNRS UMR176/Université Paris XI</b>	<b>27</b>
1.1 Contexte	27
1.1.1 Les G-Quadruplexes	27
1.1.2 Les ligands de G-Quadruplexes	32
1.2 Orientations des travaux de thèse	32
1.3 Axe 1. Synthèse de nouveaux ligands organiques et inorganiques de haute affinité et sélectivité	34
1.3.1 Complexes organométalliques comme ligands covalents et non-covalents des G4s	35
1.3.2 Ligands oligohétéroaryles acycliques qui discriminent fortement entre diverses topologies G4 en ciblant les sillons	38
1.3.3 Ligands de G4 possédant des propriétés physico-chimiques d'intérêt	42
1.3.3.1 Dérivés fonctionnalisés de pyridodicarboxamides	44
1.3.3.2 Ligands fluorescents distyryles	44
1.4 Axe 2. Méthodes analytiques pour l'analyse d'interactions ligands-acides nucléiques	47
1.4.1 Déplacement de sondes fluorescentes de G4 a haut débit : le test HT-G4-FID	47
1.4.1.1 Développement du HT-G4-FID	47
1.4.1.2 Découverte de ligands sélectifs de G4s par criblage	50
1.4.1.2.1 Exploration d'une ciblotheque	50
1.4.1.2.2 Exploration d'une chimiotheque	50
1.4.1.3 Caractérisation de ligands issus du design rationnel	51
1.4.1.3.1 Déterminants structurels de la sélectivité de ligands macrocycliques pour les paires de bases d'ADN mésappariées	51
1.4.1.3.2 Inhibition de l'expression génique par ciblage des G4s dans les 5'-UTR d'ARN	54
1.4.1.3.3 Instabilité génomique de minisatellites humains déclenchée par des ligands de G4	55
1.4.1.3.4 Effets de ligands de G4 halogénés sur la séquence terminale des télomères humains	56
1.4.1.3.5 Ligands fonctionnalisés pour la sélection d'aptamères anti-G4	56
1.4.2 Sélectivité de sondes fluorescentes : le test G4-POSCA	58
1.5 Conclusion	61
<b>2 Postdoctorat à l'Université de Colombie Britannique</b>	<b>63</b>
2.1 Contexte	63
2.2 Synthèse, caractérisation et optimisation d'oligonucléotides incorporants des nucléotides bifaciaux	65
2.3 Conclusion	68
<b>3 Postdoctorat à l'Institut Européen de Chimie et Biologie</b>	<b>69</b>
3.1 Chromatographie d'exclusion stérique pour l'étude du polymorphisme des acides nucléiques	69
3.2 Influence des cations sur les G4s : découverte et caractérisation de doubles commutateurs	70
3.3 Conclusion	76
<b>4 R&amp;D Scientist à Quality Assistance</b>	<b>77</b>
4.1 Contexte	77
4.2 Développement de méthodes LC/MS pour la caractérisation et quantification de protéines thérapeutiques	79
4.3 Glycosylation des protéines thérapeutiques	82

4.4	Échange hydrogène-deutérium couplé à la spectrométrie de masse pour la caractérisation de structures et interactions de protéines thérapeutiques . . . . .	84
4.5	Conclusion . . . . .	85
<b>5</b>	<b>Thèmes de recherche à l'Université de Bordeaux : résultats et perspectives</b>	<b>89</b>
5.1	Contexte . . . . .	89
5.2	HDX/MS de bio-macromolécules . . . . .	91
5.2.1	HDX/MS native d'acides nucléiques . . . . .	91
5.2.1.1	Preuve de concept . . . . .	91
5.2.1.2	Application à la biophysique des G4s . . . . .	94
5.2.1.3	Couplage à la mobilité ionique . . . . .	95
5.2.1.3.1	Cas 1 : Plusieurs conformères en solution, de même masse, avec des CCS et cinétiques d'échange distinctes . . . . .	99
5.2.1.3.2	Cas 2 : Plusieurs conformères en solution, de masses et CCS distinctes, et de même CCS et cinétiques d'échange . . . . .	99
5.2.1.3.3	Cas 3 : Même masses et cinétiques d'échanges mais CCS distinctes	102
5.2.1.3.4	Conclusion . . . . .	102
5.2.1.4	Etude de complexes G4/petites molécules . . . . .	102
5.2.1.5	Perspectives . . . . .	105
5.2.2	HDX/MS native de protéines thérapeutiques . . . . .	105
5.3	Ligands de G4s . . . . .	106
5.3.1	Ligands capables de discriminer des topologies G4 . . . . .	106
5.3.1.1	Ligands foldamères spécifiques des topologies parallèles . . . . .	108
5.3.1.2	Ligands pérylènes spécifiques d'une conformation intermédiaire de structuration de c-KIT2 . . . . .	115
5.3.2	Ligands avec des activités anticancéreuses ou antiparasitaires . . . . .	118
5.4	Développement de logiciels <i>open source</i> pour l'analyse de données . . . . .	120
5.4.1	Détermination de topologies d'acides nucléiques G4 par spectroscopie UV-vis et dichroïsme circulaire . . . . .	120
5.4.2	Analyse haut débit et haute précision d'UV- <i>melting</i> d'acides nucléiques . . . . .	123
5.4.3	Spectrométrie de masse native d'acides nucléiques . . . . .	132
5.5	Base de données d'acides nucléiques G-quadruplexes . . . . .	136
5.6	Mécanique quantique et moléculaire de biomolécules . . . . .	140
5.6.1	Contexte . . . . .	140
5.6.2	Implémentation GPU, automatisation et analyse de simulations MD . . . . .	141
5.6.2.1	Implémentation sur GPU . . . . .	141
5.6.2.2	Analyse semi-automatisée des structures et trajectoires . . . . .	144
5.6.3	Exemple d'un complexe aptamère d'ADN RKEC1-dopamine . . . . .	144
5.6.3.1	Contexte . . . . .	144
5.6.3.2	Préparation du système . . . . .	144
5.6.3.3	Analyse des résultats . . . . .	146
5.6.3.4	Considérations sur la prédiction de structures . . . . .	156
5.6.4	Application à des complexes G4/ligands. . . . .	156
5.6.5	Étude des interactions $\pi$ - $\pi$ d'oligonucléotides avec des phases stationnaires aromatiques pour leur séparation en phase inverse sans agent d'appariement. . . . .	157
	<b>Conclusions et perspectives</b>	<b>163</b>
	<b>Bibliographie</b>	<b>169</b>

## Introduction

*"In France, everything has an origin."* Chercheur canadien, 11/2025

Depuis mes premiers travaux de doctorat, jusqu'à mes travaux les plus récents, j'ai eu la chance de travailler dans des environnements très divers : recherche fondamentale en milieu académique, ou appliquée dans le milieu très régulé de l'industrie pharmaceutique, en France comme à l'étranger, dans des équipes de chimie organique, physique et analytique, de biochimie et de biophysique. À chaque étape, j'ai eu l'opportunité de m'essayer à une large gamme de techniques : synthèse organique et inorganique, développement en sciences analytiques (HPLC, MS, fluorescence), chémoinformatique, modélisation moléculaire, ou encore développement informatique. Sous cette diversité méthodologique se cache un fil conducteur fort : celui de comprendre les forces qui régissent les structures et les interactions non covalentes de macromolécules biologiques, en particulier de l'ADN, et les exploiter à des fins thérapeutiques. Certains de mes travaux les plus récents à l'Université de Bordeaux se penchent sur des questions que j'avais abordées il y a maintenant dix-sept ans, mais de nouvelles questions sont aussi, bien sûr, au centre de mon attention.

Dès mes premiers travaux de thèse, j'ai développé un amour certain pour les structures non canoniques de l'ADN (Figure 1). Celles qui ne forment pas de doubles hélices, celles qui se moquent de Watson et Crick (et Franklin), celles qui ont des bases et des sucres bizarres. Et, soyons honnêtes, tout particulièrement les G-quadruplexes (G4s). Tous les travaux que je décris ci-après ne concernent pas forcément ces bizarreries, il est même question de protéines (!), mais le lecteur attentif remarquera que ces écarts ont servi de sources d'inspiration à mes amours originelles. Reculer pour mieux sauter en somme.

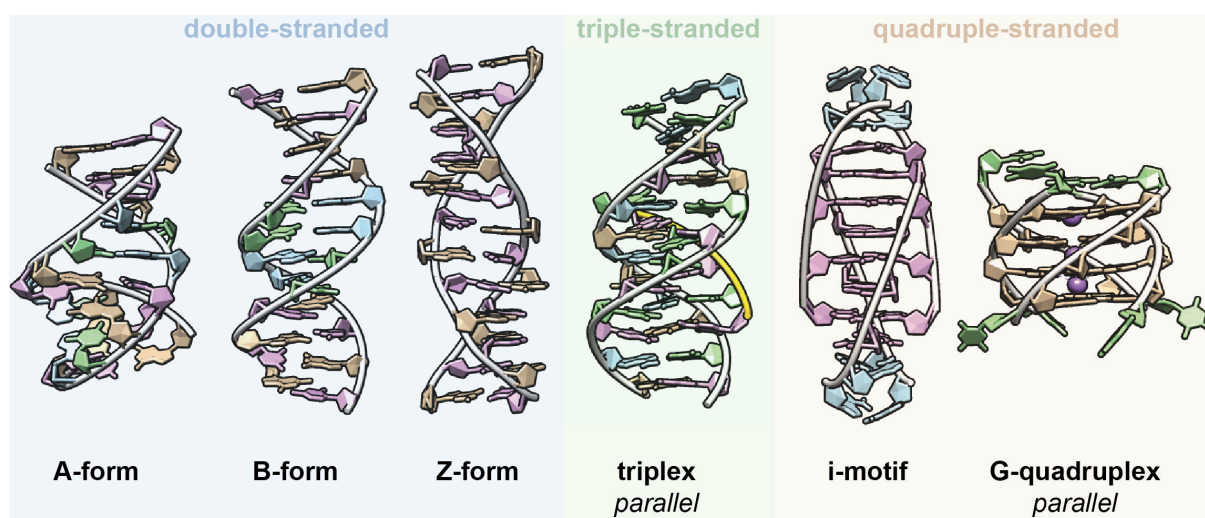


FIGURE 1. Les structures secondaires les plus connues par les amateurs de structures secondaires de l'ADN. Figure préparée pour la référence (1).

## Introduction

Parmi les autres passions se sont révélées à moi et serviront également de fil rouge dans ce manuscrit, deux sont notables :

- L'automatisation du travail. J'ai rapidement fait le constat que la fainéantise, plutôt que la nécessité, est mère de l'invention, et qu'en automatisant et/ou accélérant les processus d'analyse je pourrais me consacrer à d'autres entreprises.<sup>1</sup> Plus sérieusement, j'y ai trouvé de nombreux avantages : efficacité, reproductibilité, capacité à gérer des sets de données très larges et à générer des résultats (figures, rapports) standardisés. Mon expérience en milieu industriel régulé n'y est pas non plus étranger.
- Faire de la science ouverte et accessible : produire des figures claires et sans biais, rendre les données brutes et les méthodes/code de leur analyse publique.

Mes diverses expériences m'ont permis de rencontrer et travailler avec des personnes formidables. Si elles m'ont fréquemment laissé l'impression d'être intellectuellement limité, elles m'ont surtout beaucoup appris, tant scientifiquement qu'humainement. En écrivant ce manuscrit, j'ai eu envie de visualiser mon réseau de co-auteurs (Figure 2). Les clusters qui se forment racontent finalement assez bien ma carrière (que je survole ; ci-après). J'ai aussi découvert que je n'avais jamais rencontré certains d'entre eux et qu'en recherche comme dans la vie, les amis de mes amis sont mes amis. Quoi qu'il en soit, je les remercie tous.

Enfin, j'aimerais remercier ici tous les étudiants que j'ai eu la chance d'encadrer. Je suis particulièrement heureux que mes derniers stagiaires de master/école d'ingénieur se soient lancés dans des thèses, malgré mon insistance à leur dire que ce n'est pas un bon choix de carrière. Je ne suis sans doute pas pour grand chose dans leur réussite mais je ne les aurais au moins pas dégouttés de la recherche.

---

1. *Pour entrevoir l'essentiel, il ne faut exercer aucun métier*, nous dit Emil Cioran.



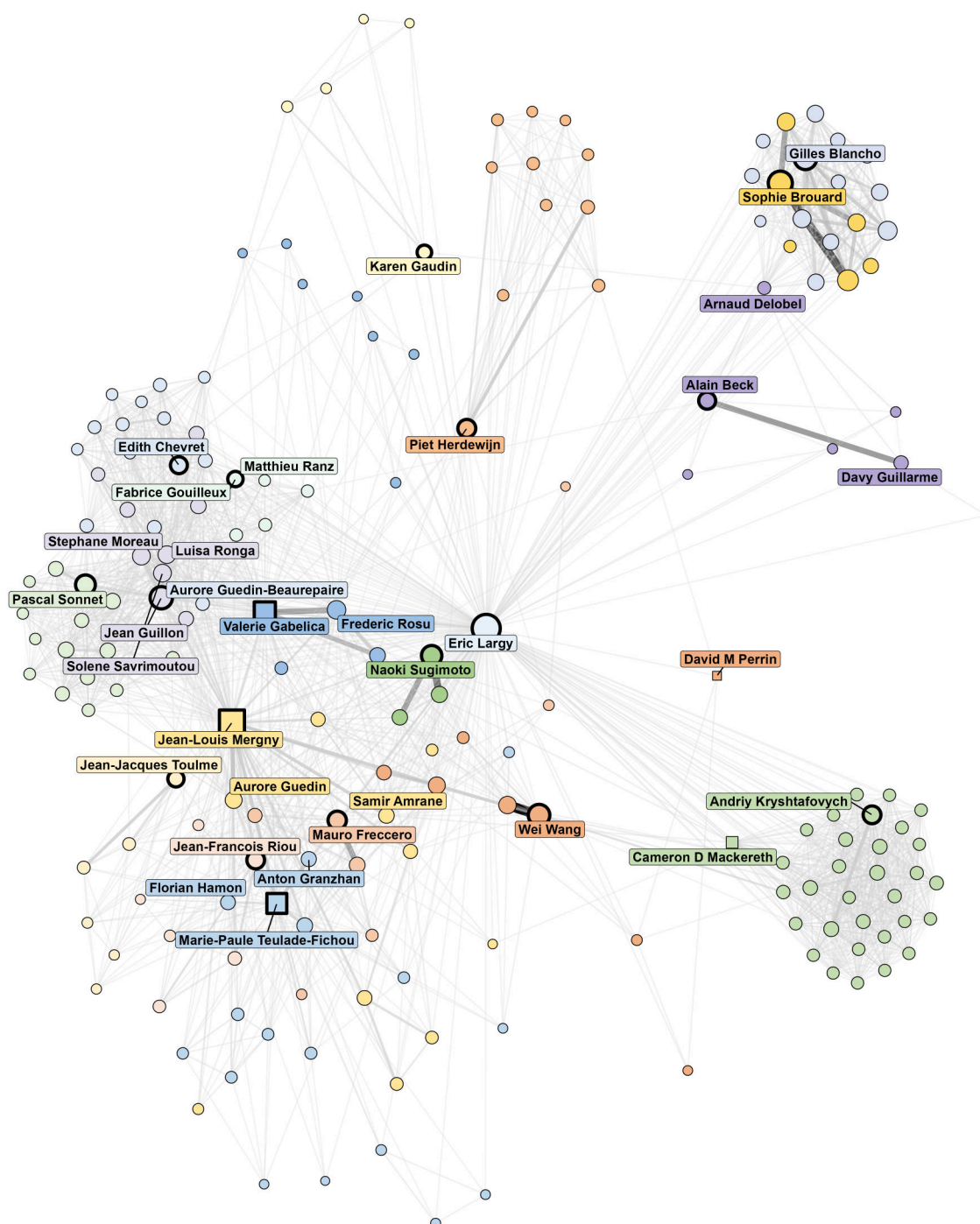


FIGURE 2. Réseau de co-auteurs de l'auteur (moi), tiré des références indexés dans Pubmed (*preprints* exclus). Pour chaque co-auteur, ses propres co-auteurs ont été listés avec la même méthodologie mais seuls ceux ayant publié avec moi sont montrés. La taille des nœuds et l'épaisseur des liens reflètent le nombre total de publications partagées avec ce réseau. Les nœuds carrés indiquent les chefs d'équipes/laboratoires qui m'ont encadrés, inspirés et supportés. Clustering effectué avec l'algorithme de Leiden (fonction d'objectif : modularité, résolution : 3, nombre d'itérations : 1000, pondération : nombre de publications partagées) (2).

## Résumé du déroulé de carrière

Mes études supérieures ont débutées par l'obtention d'un diplôme d'ingénieur chimiste à l'École Supérieure de Chimie Organique et Minérale, (à l'époque) à Cergy-Pontoise (Figure 3). J'ai complété ma formation initiale par un Master 2 recherche *Chimie fine, analytique et polymères* de l'Université de Cergy-Pontoise, effectué en parallèle de ma dernière année d'école d'ingénieur.

Mon stage de M2, qui a également fait office de stage de fin d'étude d'ingénieur, sur le développement d'une méthode de synthèse de di-organozinciques symétriques, est en apparence peu en rapport avec ma carrière académique et n'est pas détaillé ci-après. Les apparences peuvent être trompeuses, car ce sont partiellement ces travaux qui m'ont amené vers mon sujet de thèse, initialement centré sur la synthèse de complexes de métaux de transition comme ligand d'ADN G-quadruplex. L'autre étape de mes études m'ayant conduit à postuler à ce sujet de thèse est un stage estival chez Sanofi-Aventis (équipe de chimie exploratoire; site de Vitry-sur-Seine), en 2007, pendant lequel j'ai découvert le travail de chimiste à l'interface avec la biologie - aspect qui était absent de ma formation académique jusque-là. J'y ai préparé des conjugués entre le fragment FAB de l'Herceptin et des fluorophores ou molécules cages (pour la coordination d'isotopes radioactif d'indium) à destination de l'imagerie médicale de petits mammifères. Ce stage présagera d'autres de mes activités futures : caractérisation d'anticorps par chromatographie d'exclusion stérique, spectrométrie de masse, et conjugaison de fluorophores.

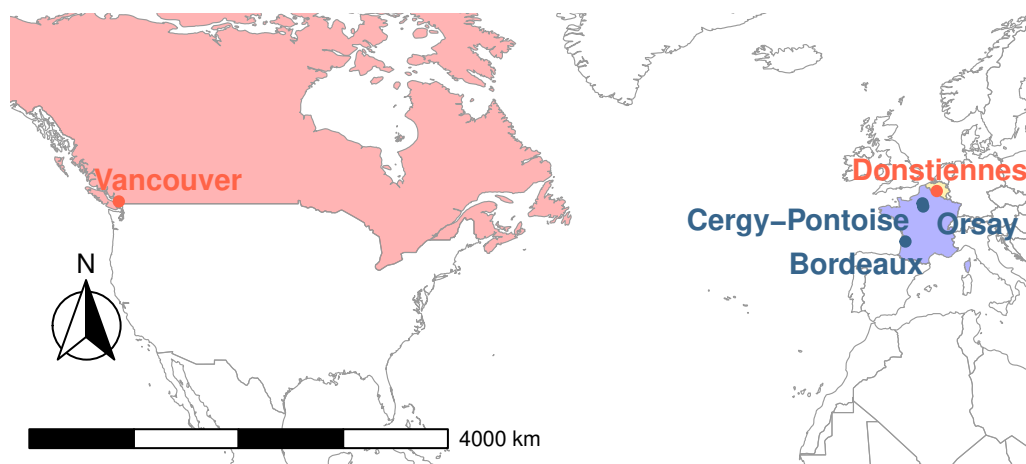


FIGURE 3. Localisation des laboratoires dans lesquels j'ai travaillé, du M2 à nos jours.

J'ai obtenu mon doctorat de chimie de l'Université Paris XI en novembre 2011 grâce à mes travaux de thèse à l'Institut Curie, sous la direction du Dr. Marie-Paule Teulade-Fichou, à l'époque directrice de l'UMR 176 du CNRS. Pour cette thèse, *Développement de méthodes analytiques pour le criblage haut débit de ligands d'ADN. Synthèse de petites molécules ligands d'ADN G quadruplex à visée anticancéreuse*, j'ai obtenu une bourse de docteur ingénieur (BDI) co-financée par le CNRS et l'Institut Curie (Chapitre 1). Je suis ensuite resté dans l'équipe jusqu'en juillet 2012 comme chercheur postdoctorant, ce qui m'a

permis de terminer plusieurs projets entrepris pendant ma thèse. Pendant ces presque quatre ans, j'ai complété ma mue de la chimie organique vers la chimie physique et analytique.

J'ai ensuite rejoint le laboratoire du Dr. David Perrin, du département de chimie de l'Université de Colombie Britannique (UBC), au Canada (Chapitre 2). Dans cette équipe, dont les intérêts sont partagés entre synthèse organique et la sélection d'aptamères d'ADN fonctionnalisés, par SELEX, j'ai continué mon apprentissage des méthodes biophysiques pour caractériser des structures d'acides nucléiques, et étendu mon répertoire méthodologique aux calculs de DFT et dynamiques moléculaires. J'ai financé ce séjour postdoctoral grâce à une bourse Banting (140 k\$).

Après une année à UBC sevré de G4s, je suis revenu en France, plus précisément à Bordeaux, dans les équipes de Jean-Louis Mergny et Valérie Gabelica à l'Institut Européen de Chimie et Biologie (IECB, ARNA, INSERM U869; Figure 4). Je me suis à nouveau intéressé aux structures non canoniques de l'ADN, à l'interface entre ces deux équipes (Chapitre 3). Dans un premier temps, j'ai entrepris avec Jean-Louis le développement d'une méthode de SEC-HPLC pour caractériser des structures d'oligonucléotides structurés; idée du regretté Jean-Louis Leroy (ICSN, Gif-sur-Yvette). Je me suis ensuite penché sur le rôle des cations dans la structuration des G4s. Ce sujet d'étude m'a permis de passer de l'équipe de Jean-Louis à celle de Valérie, et d'y découvrir la spectrométrie de masse native (après une courte mais intense initiation en thèse).

J'ai ensuite quitté le monde académique avant même d'avoir tenté ma chance aux concours. J'ai en effet accepté un poste de chercheur dans l'industrie pharmaceutique, dans laquelle j'ai mené à bien un projet financé par l'Union Européenne (Fonds européen de développement régional, « FEDER »). Cela m'a permis de mettre en œuvre un large éventail de méthodes de caractérisation et de quantification des protéines et des petites molécules thérapeutiques dans un environnement soumis à des réglementations très strictes, et avec des objectifs, des moyens et une temporalité différents de ceux de la recherche académique (Chapitre 4).

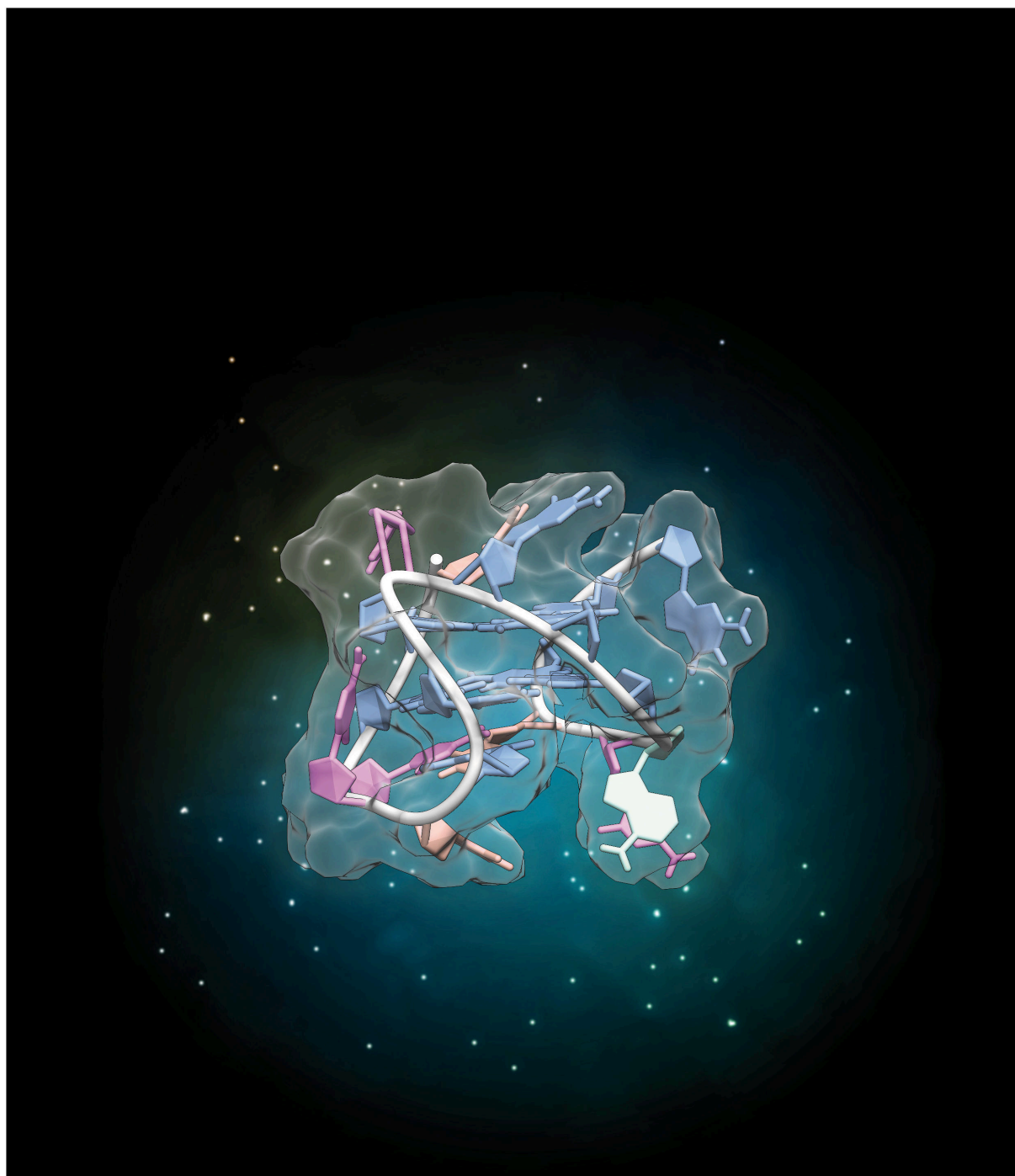
Mes travaux industriels, bien que n'impliquant pas d'acides nucléiques, ont été une source d'inspiration pour l'analyse de ces derniers, en particulier l'utilisation de l'HDX/MS pour la caractérisation de structures d'oligonucléotides et de leurs interactions avec des protéines et petites molécules. Cette expérience constitue également une base importante de mes enseignements, ainsi que de la formation continue que j'offre aux étudiants que je supervise au laboratoire (bonnes pratiques de fabrication, connaissance du monde industriel).

L'ensemble de ces travaux de recherche antérieurs à ma prise de poste à l'université de Bordeaux m'ont progressivement mené de la synthèse de petites molécules pour le ciblage de bio-macromolécules d'intérêt thérapeutique, vers la caractérisation physico-chimique et computationnelle de ces bio-macromolécules et de leurs interactions, formant un ensemble cohérent, complémentaire et en adéquation avec les projets de recherche que je développe actuellement.

Dernière étape : je suis depuis 2017 maître de conférences à l'Université de Bordeaux, rattaché à l'UFR de pharmacie pour l'enseignement et à l'unité ARNA (INSERM U1212, CNRS UMR5320) pour la recherche (Figures 5, 6). Je suis ainsi revenu à l'IECB, dans l'équipe de recherche de V. Gabelica, BALI (*Biophysics Analysis and Ligands*), mais avec de nouvelles thématiques (Chapitre 5). Suite au départ de Valérie



Institut Européen de Chimie et Biologie  
European Institute of Chemistry and Biology



Institut Européen de Chimie et Biologie  
2, rue Robert Escarpit  
33607 Pessac FRANCE

Tél. : +33(0)5 40 00 30 38  
Fax. : +33(0)5 40 00 30 04  
[www.iecb.u-bordeaux.fr](http://www.iecb.u-bordeaux.fr)

FIGURE 4. Couverture du *board report* de l'IECB (circa 2014)

pour l'Université de Genève, j'ai rejoint l'équipe PRISM de Cameron Mackereth fin 2023. J'y poursuis le travail engagé dans l'équipe de Valérie mais en explore aussi de nouvelles.

Ce mémoire est partiellement chronologique. Si les grands chapitres suivent le fil de ma carrière, j'ai regroupé les travaux par thématiques au sein de ces chapitres pour en faciliter la lecture.

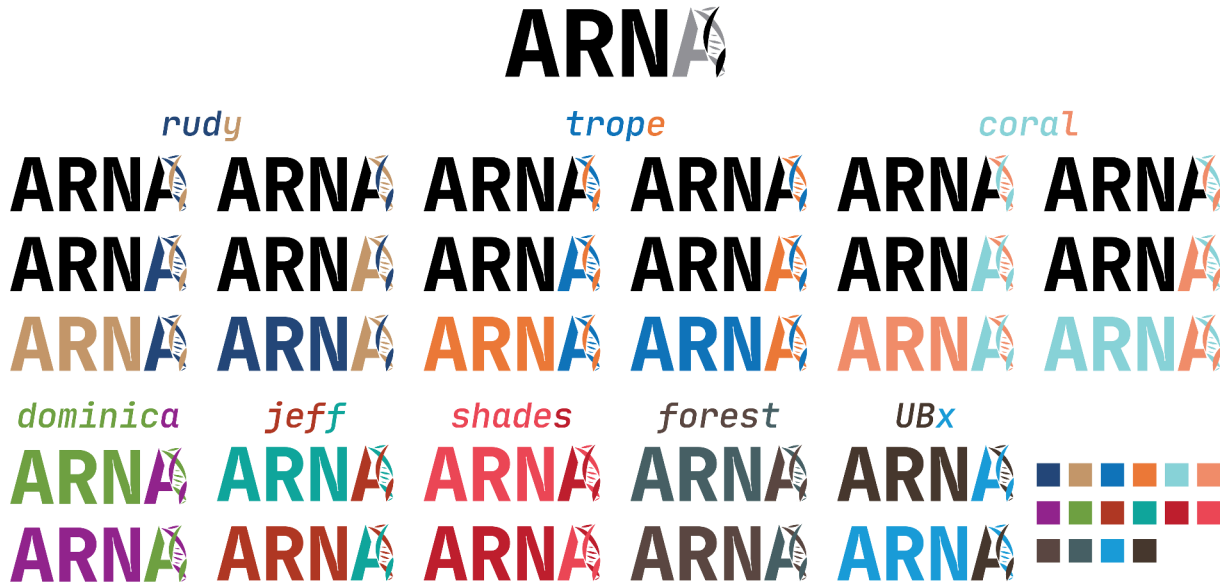


FIGURE 5. Couleurs alternatives du logo ARNA



FIGURE 6. Logo de la défunte équipe BALI



# Curriculum Vitae

## Expérience professionnelle

---

Depuis Septembre 2017	<b>Maître de conférences — Université de Bordeaux</b> <i>Laboratoire ARNA, INSERM U1212, CNRS UMR5320, équipes BALI puis Prism</i> <ul style="list-style-type: none"><li>— Développement de méthodes analytiques et outils logiciels pour la caractérisation de dynamiques structurales d'ADN et de leurs interactions par échange hydrogène deutérium couplé à la spectrométrie de masse native</li><li>— Caractérisation et prédiction de structures secondaires d'ADN par méthodes spectroscopiques et spectrométrie de masse, analysées par analyse en composantes principales et décomposition en valeurs singulières</li><li>— Construction et gestion d'une base de données de structures G-quadruplexes validées en spectrométrie de masse native</li><li>— Mise en place d'un processus semi-automatisé pour la dynamique moléculaire d'aptamères d'ADN contraints par la RMN</li></ul> <i>UFR Sciences pharmaceutiques, Collège santé, Université de Bordeaux</i> <p>Enseignant en sciences analytiques et sur les aspects réglementaires de l'analyse physico chimique dans l'industrie pharmaceutique.</p>
Août 2017 — Avril 2015	<b>R&amp;D Scientist II — Quality Assistance — Donstiennes (Belgique)</b> <ul style="list-style-type: none"><li>— Développement de méthodes LC/MS pour la caractérisation et quantification de protéines thérapeutiques, dans un environnement BPF</li><li>— Etude de structures de protéines, épitope mapping et interactions avec de petites molécules par échange hydrogène deutérium/spectrométrie de masse</li></ul>
Mars 2015 — Octobre 2013	<b>Postdoctorant — INSERM U869 — Bordeaux</b> <i>Laboratoire ARNA, équipes Mergny et Gabelica</i> <ul style="list-style-type: none"><li>— Développement de méthodes bioanalytiques pour l'étude de structures secondaires d'acides nucléiques (HPLC, IMS MS)</li><li>— Étude du rôle des cations dans le polymorphisme de l'ADN G-quadruplexe</li></ul>
Septembre 2013 — Septembre 2012	<b>Postdoctorant — University of British Columbia — Vancouver Canada)</b> <i>Perrin Lab , Chemistry department</i> <ul style="list-style-type: none"><li>— Analyses biophysiques, biochimiques et computationnelles d'oligonucléotides modifiés</li><li>— Incorporation enzymatique de nucléotides modifiés</li></ul>

Juillet 2012 –  
Janvier 2012

**Postdoctorant – CNRS UMR 176 – Institut Curie – Orsay**

*Equipe Teulade-Fichou*

- Développement de méthodes analytiques pour l'analyse d'interactions petites molécules / ADN
- Criblage haut débit de ligands d'acides nucléiques G quadruplexes
- Études biophysiques et biochimiques d'interactions petites molécules/ADN

Décembre  
2011 –

**Doctorant – CNRS UMR 176 – Institut Curie – Université Paris XI – Orsay**

Octobre 2008

*Equipe Teulade-Fichou*

- Développement de méthodes analytiques pour le criblage haut débit de ligands d'ADN.
- Synthèse de petites molécules ligands d'ADN G quadruplex à visée anticancéreuse

Juillet 2010 –  
Octobre 2008

**Vacataire – Université Paris XI – Orsay**

*IUT, Département mesures physiques*

- Cours de spectroscopie, métrologie, chimie générale, et projet professionnel
-



## Éducation

---

2011 — 2008	<b>Doctorat de chimie — Université Paris XI — Orsay</b>
	<i>Ciblage d'acides nucléiques G quadruplexes : Synthèse et développement de méthodes pour l'analyse et le criblage de ligands sélectifs multimodaux (30/11/2011)</i>
	Bourse de Docteur Ingénieur (CNRS/Institut Curie)
2008 — 2006	<b>Master de recherche en chimie fine, analytique et polymères — Université de Cergy Pontoise</b>
	<i>mention très bien</i>
2008 — 2003	<b>Diplôme d'ingénieur chimiste Ecole Supérieure de Chimie Organique et Minérale</b>

---

## Compétences

### Mesures physiques et biophysiques

Spectrométrie de masse (IMS MS, Q TOF, MALDI TOF, ion trap), échange hydrogène deutérium (HDX) couplé à la spectrométrie de masse native, U(H)PLC, 2D UPLC, spectroscopies UV-vis, fluorescence et dichroïsme circulaire, méthodes haut débit et automatisation

### Méthodes biochimiques

Purification/analyse d'oligonucléotides par PAGE, PCR, purification sur billes magnétiques, fragmentation/digestion/dérivatisation de protéines et glycanes

### Chimie médicinale

Synthèses organique et inorganique, criblage haut débit

### Chimie computationnelle et chémoinformatique

DFT, dynamique moléculaire, criblage virtuel, clustering, analyse en composantes principales, décomposition en valeurs singulières

### Langage de programmation

R (traitement de données, visualisations, bases de données, documents interactifs et web), Python, HTML, CSS, Markdown

### Langues

Français (natif), Anglais (C2), Portugais (Brésil; B1), Espagnol (B2), Néerlandais (A2)

## Financements / Récompenses

- **RIPEC C3** depuis **2023**
- **Financement ANR JCJC** (240 k€) : *DNA-HDXMS : Hydrogen deuterium exchange mass spectrometry of nucleic acids*, coordinateur, **2021 — 2025**
- **Financement projet ESR** (porté par V. Gabelica, contribution de 10%), région Nouvelle Aquitaine / Merck : *Caractériser le repliement de protéines thérapeutiques recombinantes par spectrométrie de masse couplée à la spectrométrie de mobilité ionique*, **2019 — 2021**
- **Financement MAPI (projet STEP)** pour le soutien à la transformation et à l'expérimentation pédagogiques (porté par K. Gaudin) : *Modernisation de l'UE « Sciences Analytiques » en 2ème année des études de pharmacie*, **2018 — 2020**
- **Bourse postdoctorale Banting** (140 k\$) : *Discovery and development of protein like DNAzymes that cleave target RNA with high efficiency*, **2012 — 2013**
- **Bourse de Docteur Ingénieur** (co financement CNRS / Institut Curie), **2008 — 2011**
- **Novartis funding for master degree** internships in schools of the Federation Gay Lussac, **2008**

## Directions et encadrement d'étudiants

### Postdoctorante

Dr. Nina Khristenko : *co-encadrement*, 2019 — 2021 | Cf. Section 5.2.2

- Financement projet ESR, région Nouvelle Aquitaine / Merck.
- Expertise en HDX/MS et développement de méthodes analytiques : conception et réalisation des expériences, traitement de données, interprétation des résultats et rédaction du manuscrit (3)

### Doctorant(e)s

Matthieu Ranz : *direction* (ADT 100%), thèse soutenue (U. Bordeaux, 2024) | Cf. Sections 5.2.1.2, 5.2.1.3, 5.2.1.4, 5.4.3

- Financement projet ANR
- Direction administrative et scientifique. Expertise en HDX/MS, développement de méthodes analytiques et développement logiciel, biophysique d'acides nucléiques : direction doctorale formelle, conception et réalisation des expériences, traitement de données, interprétation des résultats et rédaction de manuscrits (4–6)

⇒ *Actuellement : Postdoctorant (ARNA)*

Alexander König : *co-encadrement*, thèse soutenue (U. Bordeaux, 2023) | Cf. Section 5.3.1.1

- Financement école doctorale

- Encadrement scientifique. Expertise en spectrométrie de masse native, biophysique d'acides nucléiques et de leurs ligands : conception et réalisation des expériences, traitement de données, interprétation des résultats et rédaction de manuscrits (deux manuscrits en cours de rédaction) (1)

⇒ *Actuellement : Project scientist (dsm-firmenich)*

Silvia Ceschi : co-encadrement (6 mois, 2019) Thèse soutenue (U. Padoue, Italie) | Cf. Section 5.3.1.2

- Encadrement scientifique. Expertise en spectrométrie de masse native, biophysique d'acides nucléiques et de leurs ligands : conception et réalisation des expériences, traitement de données, interprétation des résultats et rédaction du manuscrit (7)

## Master 2 et équivalents

Clarisse Fourel : encadrement stage fin d'étude (01 au 08/2023 , ENSMAC, Pessac) | Cf. Section 5.2.1.4

- Financement sur fonds propres
- Encadrement administratif et scientifique Expertise en HDX/MS et développement de méthodes analytiques : conception et réalisation des expériences, traitement de données, interprétation des résultats.

⇒ *Actuellement : Doctorat en biologie structurale (ICSN, Université Paris Saclay)*

Cristina Dal Lago : Encadrement M2 soutenu ( 2022 , U. Padoue, Italie) | Cf. Section 5.3.2

- Financement Erasmus+
- Encadrement administratif et scientifique. Expertise en spectrométrie de masse native, biophysique d'acides nucléiques et de leurs ligands : conception et réalisation des expériences, traitement de données, interprétation des résultats et rédaction du manuscrit (8)

⇒ *Actuellement : Stagiaire en développement analytique (Aptuit, Italie)*

Wenbo Liu : co-encadrement M2 soutenu (2012-2013, UBC, Canada) | Cf. Section 2.2

- Encadrement scientifique. Expertise en biophysique des acides nucléiques, DFT et dynamique moléculaire : conception et réalisation des expériences, traitement de données, interprétation des résultats et rédaction de manuscrits (9, 10)

⇒ *Actuellement : Professeur (Université Sun Yat-sen, Chine)*

Abid Hasan : co-encadrement M2 soutenu (2012-2013, UBC, Canada) | Cf. Section 2.2

- Encadrement scientifique. Expertise en biophysique des acides nucléiques, DFT et dynamique moléculaire : conception et réalisation des expériences, traitement de données, interprétation des résultats et rédaction de manuscrits (10, 11)

⇒ *Actuellement : Software engineer (Jobber Inc., Canada)*

Emile Feugas, Mathilde Melot, Antoine Quaresima, Eve Descomps, Jade Juin, Alexis Darmaillacq : Gestion de projets (6 semaines, U. Bordeaux , 2019-2024) | Cf. Section 5.3.2

## Curriculum Vitae

- Encadrement administratif et scientifique. Expertise en spectrométrie de masse native, biophysique d'acides nucléiques et de leurs ligands : conception et réalisation des expériences, traitement de données, interprétation des résultats

⇒ Actuellement : *Consultant contrôle qualité (Nalys, Belgique)*

## Master 1 et équivalents

- Laura Fricot (2018, Université de Bordeaux) | Cf. Section 5.2.1.1 | ⇒ *Ingénieure R&D synthèse (Oleon, France)*
- Matthieu Ranz (2020, ENSMAC) | Cf. Section 5.3.2 | ⇒ *Postdoctorant (ARNA)*
- Romane Guisiano (2022, ENSIACET) | Cf. Section 5.2.1.3 | ⇒ *Doctorat en pharmacie (Institut Cochin, Université Paris Cité)*
- Avant 2017
  - Katie Mingo (MIT, Boston, USA , Cf. Section 1.4.1.3.3, (12)) | ⇒ *oto-rhino-laryngologue (Cleveland Clinic, USA)*
  - Marie Perrot (ESCOM , Cf. Section 1.4.1.2.2) | ⇒ *chargée de projet (Azur Drones, France)*
  - Reuben Ovadia U. Montpellier , Cf. Section 1.4.1.3.5) | ⇒ *European Patent Attorney (Servier, France)*
  - Julien Botton U. Paris VI , Cf. Section 1.4.1.3.4) | ⇒ *Delivery Owner (La Poste Groupe, France)*
  - Valentine Boyer (EBI , Cf. Section 1.4.1.2.2) | ⇒ *Marketing manager (Perrigo, France)*

## IUT/BTS/DEUST/Licence

- Depuis 2017 :
  - Anaïs Ferrer | ⇒ *Consultante qualification (Altogen, France)*
  - Matthieu Vida
  - Benjamin Lienard | ⇒ *Ingénieur technico-commercial (Opheleia instruments, France)*
  - Axel Florent | ⇒ *Master 2 (Bordeaux, France)*
  - Mathieu Bonnardel
  - Yann Bourdeau | ⇒ *Doctorat en chimie, (Université de Bordeaux)*
- Avant 2017 :
  - Marie Toulisse | ⇒ *Responsable cellule approvisionnement (SNCF, France)*
  - Margaux Legrand | ⇒ *Research assistant (NovImmune, France)*
  - Camille Valette | ⇒ *Technicienne de développement technologique (MANE, France)*
  - Yoann Lebreton

## Responsabilités recherche et activités annexes

- Référent Science Ouverte d'ARNA
- Référent Données de la recherche d'ARNA
- Membre de l'Atelier Bordelais des Données de recherche (ABDo)

- Membre du groupe de travail *Éthique, intégrité, genre et science ouverte* pour l'obtention du label HRS4R (Stratégie Européenne des Ressources Humaines pour les Chercheurs)
- Membre de l'équipe communication du laboratoire ARNA (activités de vulgarisation, communication du laboratoire, conception de logos et couvertures). Portfolio : <https://ericlarg4.github.io/Portfolio/>
- Consultance dans l'industrie pharmaceutique pour des problématiques d'HDX/MS de protéines thérapeutiques, et de caractérisation de structures d'oligonucléotides thérapeutiques
- Expert auprès de la Commission Européenne pour l'évaluation de financements EIC PathFinder et bourses postdoctorales Marie Skłodowska-Curie
- Reviewer pour *Advanced Science, Analytical Chemistry, Journal of Chromatography A, Journal of Medicinal Chemistry, Nucleic Acids Research, Physical Chemistry Chemical Physics, The Journal of Physical Chemistry,...*

## Responsabilités et activités d'enseignement

### UFR des sciences pharmaceutiques

- Enseignant dans les formations suivantes :
  - Docteur en pharmacie (2<sup>ème</sup> et 3<sup>ème</sup> années du *Diplôme de Formation Générale en Sciences Pharmaceutiques*) : sciences analytiques (chimie des solutions, spectroscopies, chromatographies)
  - DEUST *Production, contrôle et qualité des produits de santé Chimie analytique* (1<sup>ère</sup> et 2<sup>ème</sup> années) : chimie des solutions, spectroscopies, chromatographies, aspects réglementaires du contrôle qualité
  - Licence *TECSAN* : spectrométrie de masse et chimie des solutions
  - Master 1 IPPS : chimie des solutions
  - Masters 2 *AC2QMAPS* (*Analyse chimique, contrôle qualité du médicament et autres produits de santé*) et *ACDNP* (*analytical chemistry for drugs and natural products*) jusqu'en 2023 puis *DACQ* : Contrôle qualité du médicament, développement analytique en chromatographie, transferts de méthodes analytiques, détecteurs, qualification instrumentale, analyse de biomolécules, communication.
- Responsable des travaux pratiques de sciences analytiques (DFGSP2, ~ 160 étudiants, 63 séances / an)
- Membre du jury de délibération en deuxième année de pharmacie
- Responsable d'UEs du M2 *AC2QMAPS* puis *DACQ* : *Qualification , Recherche et développement analytique , Contrôle qualité appliqué au produit fini, Analyse de biomolécules*
- Responsable de l'UE *Research & development in analytical chemistry* du M2 *ACDNP*
- Responsable des UEs *Chimie analytique* et *Réglementaire* du DEUST *Production, contrôle et qualité des produits de santé*
- Conception et implémentation du logiciel de travaux pratiques virtuels [SANTools](#) (13)

**Centre Hospitalier Universitaire de Bordeaux**

- Enseignant pour le diplôme de *préparateur en pharmacie hospitalière* : sciences analytiques (chimie des solutions, spectroscopies, chromatographies)

**Collège des Écoles Doctorales de l'Université de Bordeaux**

Enseignant auprès des doctorants de l'Université de Bordeaux (débutera en 2026) :

- Concevoir une présentation axée recherche / innovation
- Concevoir un poster : bonnes pratiques pour communiquer visuellement des données scientifiques
- *Design a poster : best practices for communicating scientific data visually*
- Communiquer visuellement vos données scientifiques avec R : conception de figures scientifiques

## Publications

2163 citations, h-index 22.<sup>2</sup>

Les étudiants et postdoctorants supervisés apparaissent en **orange** et le candidat en **bleu**.

## Articles académiques

1. Sigl, J.; Morozov, V.; Wang, L.; Sachs, J.; Merlet, E.; **Largy, E.**; Kwon, S.; Sanchez, F.; Candela, L.; Huet, S.; Ferrand, Y.; Douat, C.; Mackereth, C.D.; Huc, Y. An artificial protein-foldamer supramolecular synthon for self-assembled hybrid architectures. *Science*, **soumis**
2. Kabbara, A.; Buré, C.; Guedin, A.; Kauffmann, B.; **Largy, E.**; Marquevielle, J.; Bonnafoos, P.; Mdarhri, Z.; Ferrand, Y.; Gabelica, V.; Rosu, F.; Andreola, M.-L.; Olivier, C.; Amrane, S. Dibenzoacridinium derivatives as a new class of G-Quadruplex ligands with anti-HIV-1 properties. *NAR Mol. Med.*, **soumis**
3. König, A.; Laffilé, V.; Thore, S.; Mackereth, C. D.; Yatsunyk, L.; Ferrand, Y.; **Largy, E.**;\* Gabelica, V.\* Helical aromatic oligoamide foldamers as selective G-quadruplex ligands. *Nucleic Acids Res.*, **2025**, *accepted*. bioRxiv : 2025.02.25.640101. <https://doi.org/10.1101/2025.02.25.640101>.
4. Tsalmpouris, A., **Largy, E.**; Guillarme, D.\* Systematic evaluation of phenyl stationary phases for oligonucleotide analysis without ion-pairing reagents. *J. Chrom. A*, **2025**, 1762, 466395. <https://doi.org/10.1093/nar/gkaf953>.
5. **Largy, E.**;\* Guedin, A., Kabbara, A.; Mergny, J.-L.\*, Amrane, S.\* Eps2Fold : A Rapid method to characterize G-quadruplex DNA structures using single absorbance spectra. *Nucleic Acids Res.*, **2025**, 53 (18), gkaf953. <https://doi.org/10.1093/nar/gkaf953>.
6. Kretsch, R. C.; Albrecht, R.; Andersen, E. S.; Chen, H.; Chiu, W.; Das, R.; Gezelle, J. G.; Hartmann, M. D.; Höbartner, C.; Hu, Y.; Jadhav, S.; Johnson, P. E.; Jones, C. P.; Koirala, D.; Kristoffersen, E. L.; **Largy, E.**; Lewicka, A.; Mackereth, C. D.; Marcia, M.; Nigro, M.; Ojha, M.; Piccirilli, J. A.; Rice, P. A.; Shin, H.; Steckelberg, A.; Su, Z.; Srivastava, Y.; Wang, L.; Wu, Y.; Xie, J.; Zwergius, N. H.; Moul, J.; Kryshtafovych, A. Functional Relevance of CASP16 Nucleic Acid Predictions as Evaluated by Structure Providers. *Proteins* **2025**, prot.70043. <https://doi.org/10.1002/prot.70043>.
7. Guillon, J.; Le Borgne, M.; Milano, V.; Guédin-Beaurepaire, A.; Moreau, S.; Pinaud, N.; Ronga, L.; Savrimoutou, S.; Albenque-Rubio, S.; Marchivie, M.; Kalout, H.; Walker, C.; Chevallier, L.; Buré, C.; **Largy, E.**; Gabelica, V.; Mergny, J.-L.; Baylot, V.; Ferrer, J.; Idrissi, Y.; Chevret, E.; Cappellen, D.; Desplat, V.; Schelz, Z.; Zupkó, I. New 2,4-Bis[(Substituted-Aminomethyl)Phenyl]Phenylquinazoline and 2,4-Bis[(Substituted-Aminomethyl)Phenyl]Phenylquinoline Derivatives : Synthesis and Biological Evaluation as Novel Anticancer Agents by Targeting G-Quadruplex. *Pharmaceuticals* **2024**, 17 (1), 30. <https://doi.org/10.3390/ph17010030>.
8. **Largy, E.**,\* **Ranz, M.**; Gabelica, V. A General Framework to Interpret Hydrogen–Deuterium Exchange Native Mass Spectrometry of G-Quadruplex DNA. *J. Am. Chem. Soc.* **2023**, jacs.3c09365. <https://doi.org/10.1021/jacs.3c09365>.

2. Valeurs calculées automatiquement à partir des données de *Google Scholar*, avec le package `scholar`(14).

9. **Largy, E.;**\* **Ranz, M.** OligoR : A Native HDX/MS Data Processing Application Dedicated to Oligonucleotides. *Anal. Chem.* **2023**, acs.analchem.3c01321. <https://doi.org/10.1021/acs.analchem.3c01321>.
10. **Khristenko, N.**; Rosu, F.; **Largy, E.**; Haustant, J.; Mesmin, C.; Gabelica, V. Native Electrospray Ionization of Multi-Domain Proteins via a Bead Ejection Mechanism. *J. Am. Chem. Soc.* **2023**, 145 (1), 498–506. <https://doi.org/10.1021/jacs.2c10762>.
11. **Largy, E.**; Alies, B.; Condesse, G.; Gaubert, A.; Livingston, T.; Gaudin, K. Teaching with Simulation Tools to Introduce the Basics of Analytical Chemistry Instrumentation. *Anal Bioanal Chem* **2022**, 414, 6709–6721. <https://doi.org/10.1007/s00216-022-04268-0>.
12. Guillon, J.; Cohen, A.; Boudot, C.; Monic, S.; Savrimoutou, S.; Moreau, S.; Albenque-Rubio, S.; Lafon-Schmaltz, C.; Dassonville-Klimpt, A.; Mergny, J.-L.; Ronga, L.; Bernabeu de Maria, M.; Lamarche, J.; **Dal Lago, C.**; **Largy, E.**; Gabelica, V.; Moukha, S.; Dozolme, P.; Agnamey, P.; Azas, N.; Mullié, C.; Courtioux, B.; Sonnet, P. Design, Synthesis, and Antiprotozoal Evaluation of New Promising 2,9-Bis[(Substituted-Aminomethyl)]-4,7-Phenyl-1,10-Phenanthroline Derivatives, a Potential Alternative Scaffold to Drug Efflux. *Pathogens* **2022**, 11 (11), 1339. <https://doi.org/10.3390/pathogens11111339>.
13. **Largy, E.**; **König, A.**; Ghosh, A.; Ghosh, D.; Benabou, S.; Rosu, F.; Gabelica, V. Mass Spectrometry of Nucleic Acid Noncovalent Complexes. *Chem. Rev.* **2021**, 122, 7720–7839. <https://doi.org/10.1021/acs.chemrev.1c00386>.
14. Ghosh, A.; **Largy, E.**;\* Gabelica, V\*. DNA G-Quadruplexes for Native Mass Spectrometry in Potassium : A Database of Validated Structures in Electrospray-Compatible Conditions. *Nucleic Acids Res* **2021**, 49 (4), 2333–2345. <https://doi.org/10.1093/nar/gkab039>.
15. Guillon, J.; Denevault-Sabourin, C.; Chevret, E.; Brachet-Botineau, M.; Milano, V.; Guédin-Beurepaire, A.; Moreau, S.; Ronga, L.; Savrimoutou, S.; Rubio, S.; Ferrer, J.; Lamarche, J.; Mergny, J.; Viaud-Massuard, M.; **Ranz, M.**; **Largy, E.**; Gabelica, V.; Rosu, F.; Gouilleux, F.; Desplat, V. Design, Synthesis, and Antiproliferative Effect of 2,9-bis[4-(Pyridinylalkylaminomethyl)Phenyl]-1,10-phenanthroline Derivatives on Human Leukemic Cells by Targeting G-quadruplex. *Archiv der Pharmazie* **2021**, 354 (8), 2000450. <https://doi.org/10.1002/ardp.202000450>.
16. Butré, C. I.; **Largy, E.**; Delobel, A. Profiling of N-Linked Oligosaccharides of a Glycoprotein by UPLC-FLR-ESI-MS After Derivatization with Fluorescent Anthranilamide. *Methods Mol Biol* **2021**, 2271, 179–188. [https://doi.org/10.1007/978-1-0716-1241-5\\_13](https://doi.org/10.1007/978-1-0716-1241-5_13).
17. Butré, C. I.; **Largy, E.**; Cantais, F.; Delobel, A. Profiling, Relative Quantification, and Identification of Sialylated N-Linked Oligosaccharides by UPLC-FLR-ESI/MS After Derivatization with Fluorescent Anthranilamide. *Methods Mol Biol* **2021**, 2271, 237–247. [https://doi.org/10.1007/978-1-0716-1241-5\\_17](https://doi.org/10.1007/978-1-0716-1241-5_17).
18. Miranda, A.; Santos, T.; **Largy, E.**; Cruz, C. Locking up the AS1411 Aptamer with a Flanking Duplex : Towards an Improved Nucleolin-Targeting. *Pharmaceuticals* **2021**, 14 (2), 121. <https://doi.org/10.3390/ph14020121>.
19. **Ceschi, S.**; **Largy, E.**; Gabelica, V.; Sissi, C. A Two-Quartet G-Quadruplex Topology of Human KIT2 Is Conformationally Selected by a Perylene Derivative. *Biochimie* **2020**, 179, 77–84. <https://doi.org/10.1016/j.biochi.2020.09.015>.
20. **Largy, E.**;\* Gabelica, V. Native Hydrogen/Deuterium Exchange Mass Spectrometry of Structured



- DNA Oligonucleotides. *Anal. Chem.* **2020**, 92 (6), 4402–4410. <https://doi.org/10.1021/acs.analchem.9b05298>.
21. Renders, M.; Dumbre, S.; Abramov, M.; Kestemont, D.; Margamuljana, L.; **Largy, E.**; Cozens, C.; Vandenameele, J.; Pinheiro, V. B.; Toye, D.; Frère, J.-M.; Herdewijn, P. Kinetic Analysis of *N*-Alkylaryl Carboxamide Hexitol Nucleotides as Substrates for Evolved Polymerases. *Nucleic Acids Research* **2019**, 47 (5), 2160–2168. <https://doi.org/10.1093/nar/gkz008>.
  22. Belarif, L.; Mary, C.; Jacquemont, L.; Mai, H. L.; Danger, R.; Hervouet, J.; Minault, D.; Thepenier, V.; Nerrière-Daguin, V.; Nguyen, E.; Pengam, S.; **Largy, E.**; Delobel, A.; Martinet, B.; Le Bas-Bernardet, S.; Brouard, S.; Soullillou, J.-P.; Degauque, N.; Blancho, G.; Vanhove, B.; Poirier, N. IL-7 Receptor Blockade Blunts Antigen-Specific Memory T Cell Responses and Chronic Inflammation in Primates. *Nat Commun* **2018**, 9 (1), 4483. <https://doi.org/10.1038/s41467-018-06804-y>.
  23. Zhou, J.; Amrane, S.; Rosu, F.; Salgado, G. F.; Bian, Y.; Tateishi-Karimata, H.; **Largy, E.**; Korkut, D. N.; Bourdoncle, A.; Miyoshi, D.; Zhang, J.; Ju, H.; Wang, W.; Sugimoto, N.; Gabelica, V.; Mergny, J.-L. Unexpected Position-Dependent Effects of Ribose g-Quartets in g-Quadruplexes. *Journal of the American Chemical Society* **2017**, 139 (23), 7768–7779. <https://doi.org/10.1021/jacs.7b00648>.
  24. **Largy, E.**; Cantais, F.; Van Vyncht, G.; Beck, A.; Delobel, A. Orthogonal Liquid Chromatography–Mass Spectrometry Methods for the Comprehensive Characterization of Therapeutic Glycoproteins, from Released Glycans to Intact Protein Level. *Journal of Chromatography A* **2017**, 1498, 128–146. <https://doi.org/10.1016/j.chroma.2017.02.072>.
  25. **Largy, E.**; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J.-L. Quadruplex Turncoats : Cation-Dependent Folding and Stability of Quadruplex-Dna Double Switches. *Journal of the American Chemical Society* **2016**, 138 (8), 2780–2792. <https://doi.org/10.1021/jacs.5b13130>.
  26. **Largy, E.**; Mergny, J.-L. Shape Matters : Size-Exclusion HPLC for the Study of Nucleic Acid Structural Polymorphism. *Nucleic acids research* **2014**, 42 (19), e149. <https://doi.org/10.1093/nar/gku751>.
  27. **Largy, E.**; **Liu, W.**; **Hasan, A.**; Perrin, D. M. A Pyrimidopyrimidine Janus-AT Nucleoside with Improved Base-Pairing Properties to Both A and T within a DNA Duplex : The Stabilizing Effect of a Second Endocyclic Ring Nitrogen. *Chemistry - A European Journal* **2014**, 20 (6), 1495–1499. <https://doi.org/10.1002/chem.201303867>.
  28. **Largy, E.**; **Liu, W.**; **Hasan, A.**; Perrin, D. M. Base-Pairing Behavior of a Carbocyclic Janus-AT Nucleoside Analogue Capable of Recognizing A and T within a DNA Duplex. *Chembiochem* **2013**, 14 (16), 2199–2208. <https://doi.org/10.1002/cbic.201300250>.
  29. Xie, X.; Choi, B.; **Largy, E.**; Guillot, R.; Granzhan, A.; Teulade-Fichou, M.-P. Asymmetric Distyrylpyridinium Dyes as Red-Emitting Fluorescent Probes for Quadruplex DNA. *Chemistry - A European Journal* **2013**, 19, 1214–1226. <https://doi.org/10.1002/chem.201203710>.
  30. **Largy, E.**; Granzhan, A.; Hamon, F.; Verga, D.; Teulade-Fichou, M.-P. Visualizing the Quadruplex : From Fluorescent Ligands to Light-up Probes. *Topics in current chemistry* **2013**, 330, 111–177. [https://doi.org/10.1007/128\\_2012\\_346](https://doi.org/10.1007/128_2012_346).
  31. Petenzi, M.; Verga, D.; **Largy, E.**; Hamon, F.; Doria, F.; Teulade-Fichou, M.-P.; Guédin, A.; Mergny, J.-L.; Mella, M.; Freccero, M. Cationic Pentaheteroaryls as Selective G-Quadruplex Ligands by Solvent-Free Microwave-Assisted Synthesis. *Chemistry - A European Journal* **2012**, 18 (45), 14487–14496. <https://doi.org/10.1002/chem.201202097>.

32. Sidibe, A.; Hamon, F.; **Largy, E.**; Gomez, D.; Teulade-Fichou, M.-P.; Trentesaux, C.; Riou, J.-F. Effects of a Halogenated G-Quadruplex Ligand from the Pyridine Dicarboxamide Series on the Terminal Sequence of XpYp Telomere in HT1080 Cells. *Biochimie* **2012**, 94 (12), 1–10. <https://doi.org/10.1016/j.biochi.2012.07.003>.
33. **Largy, E.** Hamon, F.; Teulade-Fichou, M.-P. A Streptavidin Paramagnetic-Particle Based Competition Assay for the Evaluation of the Optical Selectivity of Quadruplex Nucleic Acid Fluorescent Probes. *Methods* **2012**, 57 (1), 129–137. <https://doi.org/10.1016/j.ymeth.2012.02.008>.
34. **Largy, E.**; Saettel, N.; Hamon, F.; Dubruille, S.; Teulade-Fichou, M.-P. Screening of a Chemical Library by HT-G4-FID for Discovery of Selective G-Quadruplex Binders. *CPD* **2012**, 18 (14), 1992–2001. <https://doi.org/10.2174/138161212799958350>.
35. **Largy, E.**; Hamon, F.; Rosu, F.; Gabelica, V.; Pauw, E. D.; Guédin, A.; Mergny, J.-L.; Teulade-Fichou, M.-P. Tridentate N-Donor Palladium(II) Complexes as Efficient Coordinating Quadruplex DNA Binders. *Chemistry - A European Journal* **2011**, 17 (47), 13274–13283. <https://doi.org/10.1002/chem.201102300>.
36. Hamon, F.; **Largy, E.**; Guédin-Beaurepaire, A.; Rouchon-Dagois, M.; Sidibe, A.; Monchaud, D.; Mergny, J.-L.; Riou, J.-F.; Nguyen, C.-H.; Teulade-Fichou, M.-P. An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-Quadruplex Topologies. *Angewandte Chemie (International ed. in English)* **2011**, 50 (37), 8745–8749. <https://doi.org/10.1002/anie.201103422>.
37. Tran, P. L.; **Largy, E.**; Hamon, F.; Teulade-Fichou, M.-P.; Mergny, J.-L. Fluorescence Intercalator Displacement Assay for Screening G4 Ligands towards a Variety of G-Quadruplex Structures. *Biochimie* **2011**, 93 (8), 1288–1296. <https://doi.org/10.1016/j.biochi.2011.05.011>.
38. Renaud de la Faverie, A.; Hamon, F.; Primo, C. D.; **Largy, E.**; Dausse, E.; Delaurière, L.; Landras-Guetta, C.; Toulmé, J.; Teulade-Fichou, M.-P.; Mergny, J. Nucleic Acids Targeted to Drugs : SELEX against a Quadruplex Ligand. *Biochimie* **2011**, 93 (8), 1357–1367. <https://doi.org/10.1016/j.biochi.2011.05.022>.
39. **Largy, E.**; Hamon, F.; Teulade-Fichou, M.-P. Development of a High-Throughput G4-FID Assay for Screening and Evaluation of Small Molecules Binding Quadruplex Nucleic Acid Structures. *Analytical and Bioanalytical Chemistry* **2011**, 400 (10), 3419–3427. <https://doi.org/10.1007/s00216-011-5018-z>.
40. Halder, K.; **Largy, E.**; Benzler, M.; Teulade-Fichou, M.-P.; Hartig, J. S. Efficient Suppression of Gene Expression by Targeting 5'-UTR-Based RNA Quadruplexes with Bisquinolinium Compounds. *Chembiochem* **2011**, 12 (11), 1663–1668. <https://doi.org/10.1002/cbic.201100228>.
41. Piazza, A. A.; Boule, J.-B.; Lopes, J.; Mingo, K.; **Largy, E.**; Teulade-Fichou, M.-P.; Nicolas, A.; Boulé, J.-B. Genetic Instability Triggered by G-Quadruplex Interacting Phen-DC Compounds in *Saccharomyces Cerevisiae*. *Nucleic Acids Research* **2010**, 38 (13), 4337–4348. <https://doi.org/10.1093/nar/gkq136>.
42. Granzhan, A.; **Largy, E.**; Saettel, N.; Teulade-Fichou, M.-P. Macrocyclic DNA-Mismatch-Binding Ligands : Structural Determinants of Selectivity. *Chemistry - A European Journal* **2010**, 16 (3), 878–889. <https://doi.org/10.1002/chem.200901989>.

## Chapitres de livre

1. **Largy, E.** Application of HDX to oligonucleotides for increased characterization capabilities. In *Oligonucleotide Therapeutics Characterization by Mass Spectrometry*; Lippens, J.L., Jora, M., Eds.; Royal Society of Chemistry : Cambridge; to be published in 2026.
2. **Largy, E.**; Mergny, J.-L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. In *The Alkali Metal Ions : Their Role for Life*; Sigel, A., Sigel, H., Sigel, R. K. O., Eds.; Metal Ions in Life Sciences; Springer International Publishing : Cham, 2016; Vol. 16, pp 203–258. [https://doi.org/10.1007/978-3-319-21756-7\\_7](https://doi.org/10.1007/978-3-319-21756-7_7).
3. **Largy, E.**; Teulade-Fichou, M.-P. Screening for Quadruplex Binding Ligands : A Game of Chance ? In *Guanine Quartets*; Spindler, L., Fritzsche, W., Eds.; Royal Society of Chemistry : Cambridge, 2012; pp 248–262. <https://doi.org/10.1039/9781849736954-00248>.

## Conférences académiques

### Proceedings

1. Amrane, S.; Bedrat, A.; Renaud De La Faverie, A.; Zhou, J.; Mendoza, O.; De Rache, A.; Guedin, A.; **Largy, E.**; Amor, S.; Salgado, G.; Bourdoncle, A.; Yatsunyk, L. A.; Mergny, J.-L. Quadruplexes Are Everywhere! In *Collection Symposium Series*; Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic : Český Krumlov, 2014; pp 62–65. <https://doi.org/10.1135/css201414062>.
2. Granzhan, A.; Saettel, N.; Hamon, F.; **Largy, E.**; Guetta, C.; Teulade-Fichou, M.-P. Recognition of DNA Secondary Structures : From Structure to Fluorescent Probes. In *Collection Symposium Series*; Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic : Český Krumlov, 2011; pp 77–85. <https://doi.org/10.1135/css201112077>.

## Communications orales

1. **Largy, E.**, *Dynamics and biophysics of aptamer complexes*, Bordeaux Aptamer Mini-Symposium, 15 octobre 2025, Bordeaux
2. **Largy, E.**, *Caractérisation biophysique des oligonucléotides : stabilité, structure, dynamique et interactions*. Journée Club AFSEP Sud Ouest : Analyse des produits biopharmaceutiques : défis, innovations et perspectives. 12 juin 2025, **Invité**, Toulouse
3. **Largy, E.**, *Hydrogen-Deuterium Exchange and Native Mass Spectrometry to characterize G4 conformations and dynamics*. INTERACTION WINS 2024 Winter School, 20 février **2024**, **Invité**, en ligne
4. **Largy, E.**, **Ranz, M.**, **Guisiano, R.**, Gabelica, V., *Hydrogen-deuterium exchange native mass spectrometry of G-quadruplex DNA*. International Mass Spectrometry Conference (IMSC), 27 août **2022**, Maastricht (Pays-Bas)
5. **Largy, E.**, Gabelica, V., *HDX/Native MS of G4s*. G4thering - 8th International Meeting on Quadruplex Nucleic Acids, 27 juin **2022**, Marienbad (République Tchèque)

6. **Largy, E.**, Gabelica, V., *Biophysical characterization of G-quadruplex nucleic acids by Hydrogen-Deuterium exchange coupled to native mass spectrometry*. HDXMS2022 conference, 26 avril **2022**, Londres (Royaume-Uni)
7. **Largy, E.**, Gabelica, V., *Native HDX/MS of DNA Oligonucleotides*. 5th Online Seminar in HDX-MS : Nontraditional HDX-MS, 7 octobre **2021**, Invité, en ligne.
8. **Largy, E.**, Gabelica, V., *Native Hydrogen/Deuterium Exchange Mass Spectrometry of Structured DNA Oligonucleotides*. Journées Françaises de Spectrométrie de Masse, 15 juin **2021**, en ligne.
9. **Largy, E.**, Gabelica, V., *Native Hydrogen Deuterium Exchange Mass Spectrometry of G-quadruplexes. Nucleic acid secondary structures : G4s and beyond*, 10 décembre **2020**, en ligne.
10. **Largy, E.**, Gabelica, V., *Native Hydrogen Deuterium Exchange Mass Spectrometry of Oligonucleotides*. Orbitrap User Meeting, 3 décembre **2019**, Invité, Paris (France) (**keynote lecture**).
11. **Largy, E.**, **Fricot, L.**, **Ferrer, A.**, Gabelica, V., *Native Hydrogen Deuterium Exchange Mass Spectrometry of Oligonucleotides*. Journée Scientifique du département STS, 1<sup>er</sup> octobre **2019**, Pessac (France)
12. **Largy, E.**, **Fricot, L.**, **Ferrer, A.**, Gabelica, V., *Nucleic Acids Biophysics by In-Solution HDX/Native MS*. 67<sup>th</sup> ASMS Conference on Mass Spectrometry and Allied Topics, 2–6 juin **2019**, Atlanta, GA (USA)
13. **Largy, E.**, Haustant, J., Rosu, F., Mesmin, C., Gabelica, V., *Combining native MS, on-line HDX-MS and IMS to study biomolecule structure and dynamics*. 5th Annual European Ion Mobility Seminar, 20 mars **2019**, Pessac (France)
14. **Largy, E.**, Gabelica, V., *Continuous-flow hydrogen-deuterium exchange to study quadruplex nucleic acids structures, stability, and interactions*. BIONIC 2018, 26–28 septembre **2018**, Padova (Italy)
15. **Largy, E.**, Gabelica, V., *Hydrogen-Deuterium Exchange Mass Spectrometry (HDX/MS) to study quadruplex nucleic acid folding and interactions*. 10<sup>th</sup> Bordeaux RNA Club, 14-15 juin **2018**, Pessac (France)
16. **Largy, E.**, Gabelica, V., *Continuous-flow HDX/MS to study Nucleic acid folding and interactions*. 66<sup>th</sup> ASMS Conference on Mass Spectrometry and Allied Topics, 3–7 juin **2018**, San Diego, CA (USA)
17. **Largy, E.**, Gabelica, V., *Continuous-flow HDX/MS to study Nucleic acid folding and interactions*. COST BM1403 : Native Mass Spectrometry and Related Methods for Structural Biology, Février **2018**, Vienne (Autriche)
18. **Largy, E.**, Fast, robust and customizable analytical workflows for therapeutic proteins N- and O-glycan profiling by LC/MS<sup>E</sup>. Waters Glycans meeting, **2016**.
19. **Largy, E.**, Mergny, J. L., *Size-exclusion chromatography for the study of nucleic acid structural polymorphism*. IECB internal seminar, Mars **2014**, Pessac (France)
20. **Largy, E.**, Mergny, J. L., *Development of Size-Exclusion HPLC (SE-HPLC) for the Study of Quadruplex Nucleic Acid Polymorphism*. 4<sup>th</sup> GDR G4 day, Mars **2014**, Grenoble (France)
21. **Largy, E.**, Hamon, F., Guédin, A., Mergny, J.-L., Teulade-Fichou, M.-P. *Non-canonical binders of quadruplex DNA : groove binding and metallation*. 2<sup>nd</sup> GDR G4 day, Janvier **2012**, Pessac (France)
22. **Largy, E.**, Hamon, F., Guédin, A., Mergny, J.-L., Teulade-Fichou, M.-P. *Interaction of new metal complexes with quadruplex DNA*. 1<sup>st</sup> GDR G4 day, Mars **2011**, Paris (France)
23. **Largy, E.**, Hamon, F., Guédin, A., Mergny, J.-L., Teulade-Fichou, M.-P. *Synthesis and evaluation of new metal complexes quadruplex DNA ligand*. 14<sup>ème</sup> Journée de Chimie Organique et Chimie Organique Biologique de la Montagne-Sainte-Geneviève, Juin **2010**, Paris (France)

## Posters

1. **Largy, E.**, Mackereth, C., *GPU-accelerated Amber molecular dynamics for the study of DNA aptamer structures*. JCAD 2024, 4-6 Novembre **2024**, Bordeaux
2. **Largy, E.**, **Ranz, M.**, Gabelica, V., Mackereth, C., *Hydrogen Deuterium Exchange coupled to Native Mass Spectrometry (HDX/MS) to study telomeric G quadruplex DNA structures*. 18th NMR Retreat of Protein-RNA Interactions, 21-25 janvier **2024**, Parpan (Suisse)
3. **Largy, E.**, Gabelica, V., *OligoR : An online software suite for oligonucleotide HDX/MS and quantitative native MS data treatment and visualization*, 68<sup>th</sup> ASMS Conference on Mass Spectrometry and Allied Topics, 1–5 juin **2020**, en ligne.
4. **Khristenko, N.**, **Largy, E.**, Haustant J., Rosu F., Mesmin C., Gabelica V. *Characterizing the folding of recombinant therapeutic proteins by H/D exchange, native mass spectrometry and ion mobility spectrometry*, 68<sup>th</sup> ASMS Conference on Mass Spectrometry and Allied Topics, 1–5 juin **2020**, en ligne.
5. **Largy, E.**, Gabelica, V., *Native HDX/MS of G-quadruplex oligonucleotides*, ANNA 2019 - Advances in Noncanonical Nucleic Acids, 1–5 juin **2019**, Rogaška Slatina, Slovénie.
6. **Largy, E.**, Delobel, A., *Epitope mapping of an interleukin receptor for three therapeutic antibodies by HDX-MS*. Analytical Technologies Europe, 6–9 Mars **2018**, Barcelone, Espagne.
7. **Largy, E.**, Cantais, F., Delobel, A., *Orthogonal Liquid Chromatography-Mass Spectrometry Methods for the Comprehensive Characterization of Therapeutic Glycoproteins, from Released Glycans to Intact Protein Level*. Analytical Technologies Europe, 6–9 Mars **2018**, Barcelone, Espagne.
8. **Largy, E.**, Cantais, F., Delobel, A., *State-of-the art LC/MS Methods Applied to the Characterization of a Highly-glycosylated Fusion Protein : Etanercept*. Analytical Technologies Europe, 14–17 Mars **2017**, Bruxelles, Belgique.
9. **Largy, E.**, Catrain, A., Delobel, A., *Multi-level Mass Spectrometric Characterisation of Antibody-Drug Conjugates in a Regulated Environment*. Analytical Technologies Europe, 14–17 Mars **2017**, Bruxelles, Belgique.
10. **Largy, E.**, Catrain, A., Delobel, A., *Multi-level Mass Spectrometric Characterisation of Antibody-Drug Conjugates in a Regulated Environment*. Analytical Technologies Europe, 14–17 Mars **2017**, Bruxelles, Belgique.
11. **Largy, E.**, Catrain, A., Van Vyncht, G., Delobel, A., *Development and use of 2D-LC/MS as a versatile and powerful tool for the analysis of mAbs and ADCs in a regulated environment*. Analytical Technologies Europe, 15–18 Mars **2016**, Vienne, Autriche.
12. Hosfield, C., **Largy, E.**, Catrain, A., Cantais, F., Van Vyncht, G., Rosenblatt, M., Saveliev, S., Uhr, M., Delobel, A., *Characterizing Deamidation and Oxidation in Adalimumab with Low pH Peptide Mapping and Middle-Up Mass Spec Analysis*. 64<sup>th</sup> ASMS Sanibel, 5–9 Juin **2016**, San Antonio, TX, USA.
13. **Largy, E.**, Hamon, F., Teulade-Fichou, M.P. *A High throughput G4-FID assay for screening and evaluation of quadruplex ligands*, Fourth International Meeting on Quadruplex nucleic acids, **2013**, Singapour



14. **Largy, E.**, Hamon, F., Guedin, A., Mergny, J., Rosu, F., Gabelica, V., Bombard, S., Teulade-Fichou, M. *New metal terpyridine complexes targeting the telomeric G-quadruplex*, Third International Meeting on G-Quadruplex and G-assembly, **2011**, Sorrento, Italie.
15. **Largy, E.**, Hamon, F., Teulade-Fichou, M. *Development of a high-throughput G4-FID assay for screening and evaluation of quadruplex ligands*, Third International Meeting on G-Quadruplex and G-assembly, **2011**, Sorrento, Italie.
16. **Largy, E.**, Hamon, F., Guetta, C., Bertrand, H., Monchaud, D., Guedin, A., Mergny, J., Bombard, S., Teulade-Fichou, M. *New metal terpyridine complexes targeting the telomeric G-quadruplex*, Pacificchem 2010, International Chemical Congress of Pacific Basin Societies, **2010**, Honolulu, HI, USA.
17. **Largy, E.**, Piazza, A., Boule, J. B., Lopez, J., **Khristenko, N.**, Lacroix, L., Teulade-Fichou, M., Nicolas, A. *High affinity G-quadruplex binders Phen-DC compounds trigger genetic instability in yeast*, Pacificchem 2010, International Chemical Congress of Pacific Basin Societies, **2010**, Honolulu, HI, USA.
18. **Largy, E.**, Hamon, F., Guedin, A., Mergny, J.-L., Teulade-Fichou, M.P. *Synthèse de nouveaux ligands organométalliques de l'ADN quadruplexe*, Journée de l'Ecole Doctorale de Chimie de PARIS SUD, **2010**.
19. **Largy, E.**, Teulade-Fichou, M. *Synthèse de ligands fluorescents de l'ADN quadruplexe – Méthode d'évaluation haut débit de l'affinité d'un ligand pour diverses structures d'ADN/ARN quadruplexes*, Journée de l'Ecole Doctorale de Chimie de PARIS SUD, **2009**

## Organisation

1. 5th international Meeting on Quadruplex nucleic acids, *G4thering in Bordeaux : Chemistry, biology, and nanotechnology*, 26–28 Mai **2015**, IECB, Pessac, France
2. 1<sup>st</sup> GDR G4 day, *Structures et Rôles Biologiques des Quadruplexes d'Acides Nucléiques*, Quadruplex Research Group (GDR 3431), 30 Mars **2011**, Institut Curie, Paris, France

## Communications industrielles

### Articles

1. Cajot, C., Delobel, A., **Largy, E.**, *Epitope Mapping of an Interleukin Receptor for Three Therapeutic Antibodies by Hydrogen-Deuterium Exchange Mass Spectrometry*, LCGC, **2018**, 16, 8–14.
2. **Largy, E.**, Catrain, A., Van Vyncht G., Delobel, A. *2D-LC-MS for the Analysis of Monoclonal Antibodies and Antibody–Drug Conjugates in a Regulated Environment*, LCGC, **2016**, 14, 2935.

### Webinar

1. **Largy, E.**, *Mass spectrometry toolkit for the characterisation of protein glycosylation*, 2016.

**Notes d'application**

1. **Largy, E.**, Delobel, A., *Stress testing of Humira (adalimumab) monitored by Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS)*, Quality Assistance (Donstienne, Belgique) **2017**.
2. **Largy, E.**, Cajot C., Delobel, A., *Epitope mapping of an interleukin receptor for three therapeutic antibodies by Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS)*, Quality Assistance (Donstienne, Belgique) **2017**.
3. Dusserre-Bresson, F., **Largy, E.**, Lepers C., Mantini, C., Plennevaux, C., Vandermeers, F. Delobel, A., *Analytical approaches for the evaluation of ADCC activity of mAbs*, Quality Assistance (Donstienne, Belgique) **2017**.
4. **Largy, E.**, Cantais, F., Van Vyncht, G., Delobel, A., *Characterisation of the glycosylation of Humira<sup>®</sup> (Adalimumab) and Erbitux<sup>®</sup> (Cetuximab) by high resolution mass spectrometry*, Quality Assistance (Donstienne, Belgique) **2016**.
5. **Largy, E.**, Cantais, F., Van Vyncht, G., Delobel, A., *Characterisation of the glycosylation of Enbrel<sup>®</sup> (Etanercept) by high resolution mass spectrometry*, Quality Assistance (Donstienne, Belgique) **2016**.
6. **Largy, E.**, Catrain, A., Van Vyncht, G., Delobel, A., *Characterisation of Antibody-Drug Conjugates by High-Resolution Mass Spectrometry*, Quality Assistance (Donstienne, Belgique) **2015**.
7. **Largy, E.**, Catrain, A., Van Vyncht, G., Delobel, A., *Characterisation of Humira<sup>®</sup> (Adalimumab) by High-Resolution Mass Spectrometry*, Quality Assistance (Donstienne, Belgique) **2015**.





# Chapitre 1

## Doctorat à l'Institut Curie/CNRS UMR176/Université Paris XI

### 1.1 Contexte

J'ai effectué mon doctorat à l'Institut Curie, au sein de l'UMR176 du CNRS, rattachée à l'Université Paris XI, grâce à une bourse de docteur ingénieur du CNRS (co-financée par l'Institut Curie). Marie-Paule Teulade-Fichou (Figure 1.1), directrice de l'unité, dirige mes travaux de thèse centrés sur l'étude de l'interaction de petites molécules avec les acides nucléiques G-quadruplex (G4). Nous retrouverons cette famille de structures secondaires non-canoniques de l'ADN et l'ARN tout au long de ce mémoire, et quelques éléments de contexte sont donc nécessaires. Par soucis de concision, ces structures ne sont pas décrites ci-dessous avec le niveau de détails dont je peux habituellement faire preuve. Le lecteur curieux, mais pressé, peut se référer à un guide du G4 concis que j'ai mis [en ligne](#). Pour une description plus exhaustive des structures secondaires d'ADN et leurs interactions non-covalentes, le lecteur peut se référer à la section 2 de la référence (1).

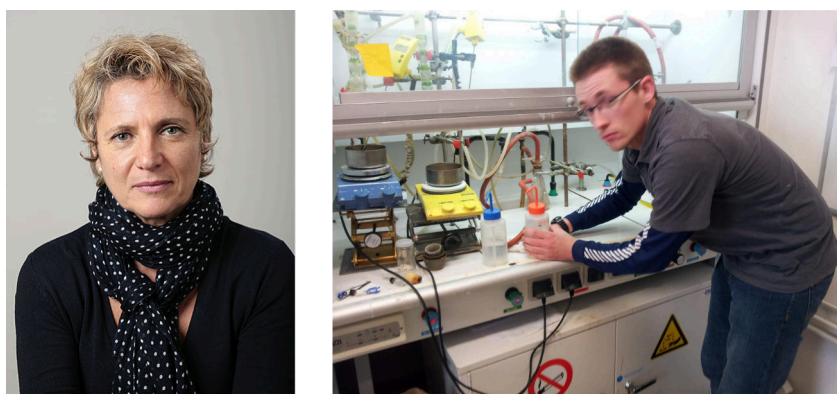


FIGURE 1.1. Marie-Paule Teulade-Fichou et Florian Hamon (dont j'ai hérité le respect du port des EPI)

#### 1.1.1 Les G-Quadruplexes

Les G-Quadruplexes (G4) sont des quadruples hélices d'acide nucléique, formées par un (intramoléculaire), deux (bimoléculaire) ou quatre (tétramoléculaire) brins d'acide nucléique riches en guanines et, de préférence, pauvres en cytosines (15, 16). Les G4 sont impliqués dans un certain nombre de processus biologiques critiques (au niveau de l'ADN et de l'ARN), ce qui en fait des cibles thérapeutiques prometteuses (par exemple, régulation des gènes, inhibition de la télomérase, lésions de l'ADN) (17–20). Ils sont parfois formés par des aptamères et peuvent être utilisés en nanotechnologie (Figure 1.2) (21, 22).



FIGURE 1.2. Couverture du numéro de *Chemical Reviews* (Volume 119, Issue 10) dédié aux nanotechnologies basées sur les acides nucléiques, dans lequel J.-L. Mergny et D. Sen s'intéressent aux G4s en particulier (22). J'ai réalisé cette couverture à leur demande.

Les G4 sont formés par un empilement d'au moins deux G-quartets, également appelés G-tétrades (Figure 1.3A). Une tétrade est elle-même composée de quatre guanines associées par huit liaisons hydrogène sur leurs faces de Watson-Crick et de Hoogsteen. Un cation métallique (typiquement  $K^+$  ou  $Na^+$ ) est coordonné par le O6 de huit guanines appartenant à deux tétrades G empilées consécutivement (Figure 1.4) (23). Par conséquent, un G-quadruplex contenant  $n$  G-quartets coordonne généralement  $n - 1$  cations. Le rayon ionique influence grandement la capacité d'un cation à promouvoir la formation des G-quadruplexes, ce que j'exposerai plus tard dans mon parcours (Section 3.2) (23–26).

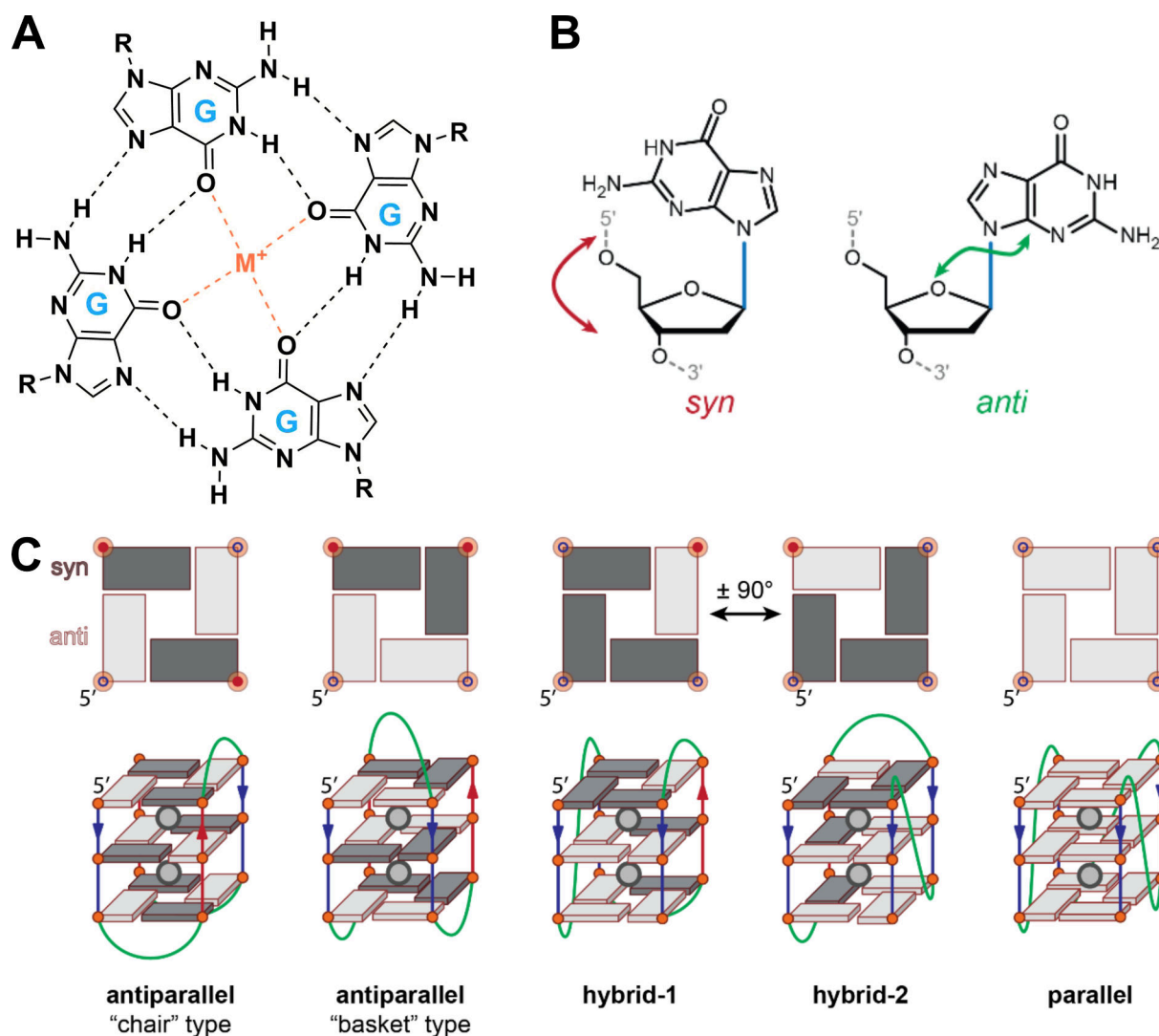


FIGURE 1.3. A. Association coplanaire de quatre guanines en G-quartet (ou G-tétrade). Un cation ( $K^+$ ,  $Na^+$ ,  $NH_4^+$ ) est coordonné par les O6 de guanines; B. Orientation *syn* ou *anti* de la liaison glycosidique; C. Différentes topologies G4 en fonction de l'orientation de la liaison glycosidique des guanines de tétrades (gris foncé ou clair) et de la géométrie des boucles (vert). Un cation (rond gris) est coordonné entre chaque couple de tétrades.

Dans un G4, les guanines impliquées dans les G-quartets constituent les *G-tracts*. Les nucléotides des termini 5' et 3', positionnés avant le premier ou après le dernier G-quartet, sont appelés *flanking sequences*. Les boucles reliant les G-tracts peuvent adopter différentes géométries, principalement la *propeller* (reliant des guanines de G-tracts adjacents mais de quartets opposés), latérale (G-tracts adjacents et même quartet), et diagonale (G-tracts diagonalement opposés et même quartet) (29).

Les G4 sont particulièrement polymorphes; certaines séquences peuvent adopter des conformations



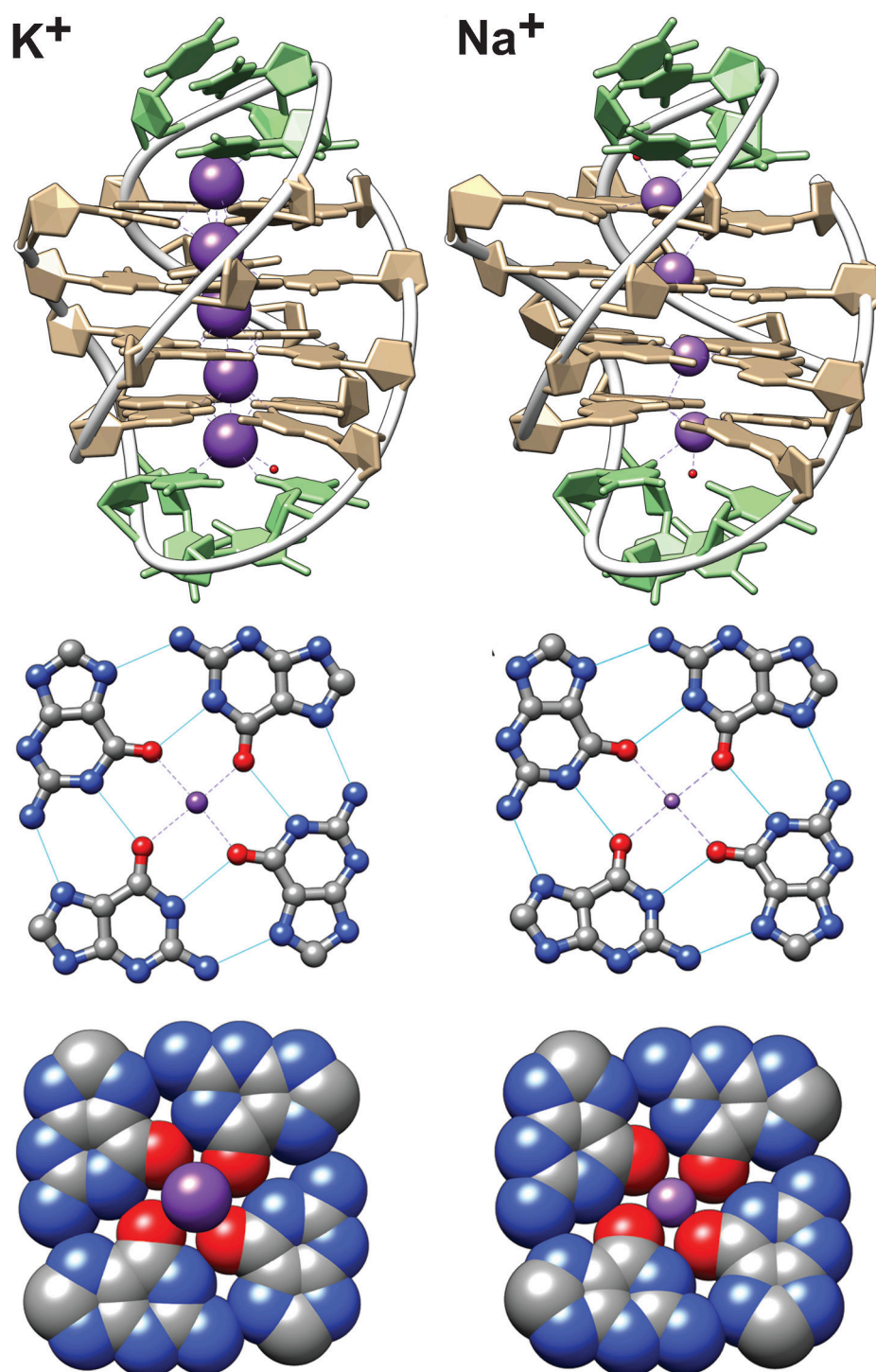


FIGURE 1.4. Haut : Structure cristalline de  $[d(T_4G_4T_4)]_2$  coordonnant  $K^+$  (PDB ID 1JPQ (27)) ou  $Na^+$  (PDB ID 1JB7 (28)). Les guanines sont représentées en brun, les thymines en vert, les cations sous forme de sphères violettes et les oxygènes des molécules d'eau complétant les sphères de coordination sous forme de sphères rouges. Milieu et bas : structure des tétrades représentées aux formats *balls and sticks* et sphères, respectivement.

distinctes, coexistant parfois en équilibre (30, 31). Les résidus des *termini* et des boucles peuvent favoriser des conformations spécifiques en formant des paires de bases avec d'autres résidus de *termini* ou des boucles et en s'empilant sur des tétrades externes (32–34). En règle générale, les séquences avec des boucles plus courtes ont tendance à adopter une géométrie *propeller* et à produire des G4s plus stables que celles avec des boucles longues (35).

L'attention particulière portée à la séquence télomérique humaine est en partie historique, avec un certain nombre d'études fondamentales parues à la fin des années 80 et au début des années 90 (36–38), en partie thérapeutique (39–41), mais découle également de son polymorphisme particulièrement impressionnant pour une simple répétition des bases GGGTTA. On pourrait écrire des pages sur ce polymorphisme, ce dont je vous ferai grâce ici puisque je l'ai déjà fait par ailleurs (23), et que la Figure 1.5 illustre plutôt bien cette diversité structurale.

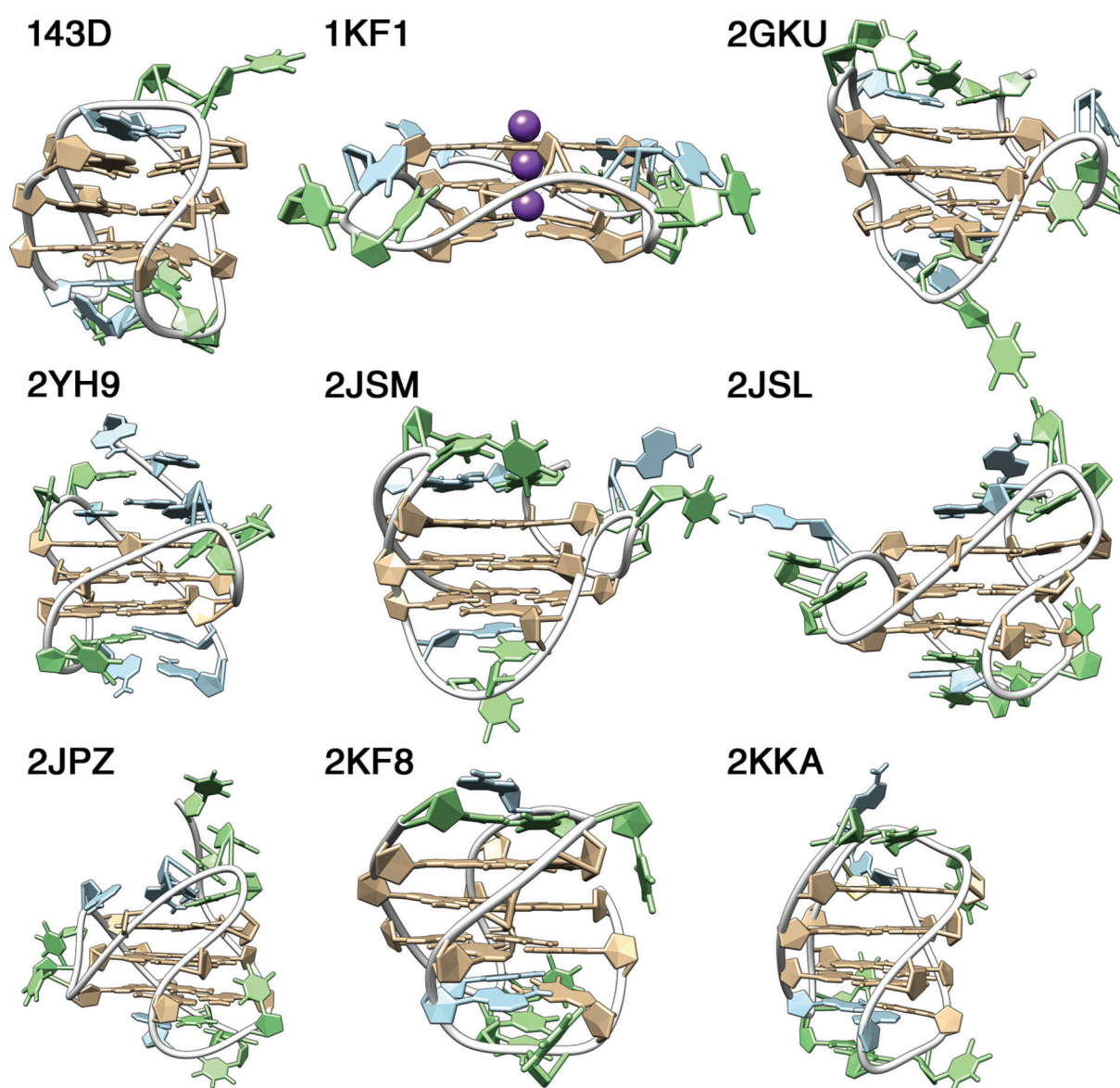


FIGURE 1.5. Exemples de structures G-quadruplex télomériques humaines déposées dans la PDB avant 2010 et résolues par RMN ou cristallographie aux rayons X : 143D (42), 1KF1 (43), 2GKU (44), 2YH9 (45), 2JSM et 2JSL (46), 2JPZ (47), 2KF8 (48) et 2KKA (49). Toutes les structures, à l'exception de 143D (Na<sup>+</sup>), ont été résolues dans des conditions K<sup>+</sup>.

Ce qu'il faut préciser en revanche, c'est que ce polymorphisme est loin de se limiter à la seule séquence télomérique humaine. Étant donnée cette polymorphie intrinsèque des G4s, ils ont rapidement été classés dans des familles partageant une même topologie. Celle-ci est généralement définie par les orientations relatives des *G-tracts* et/ou des angles de liaison glycosidique des guanines (*syn* ou *anti*; Figure 1.3B). La topologie est parallèle lorsque tous les tracts sont orientés dans la même direction (toutes les guanines empilées ont un angle de liaison glycosidique en *anti*), antiparallèle lorsque deux brins sont orientés dans la direction opposée aux deux autres (hétéro-empilement de guanines *syn* et *anti*) et hybride lorsqu'un brin est orienté dans la direction opposée aux trois autres (mélange d'homo- et d'hétéro-empilement) (Figure 1.3C) (29). Les topologies antiparallèles où les brins de direction opposée sont latéralement (type "chaise", toutes les boucles sont latérales) ou diagonalement (type "panier", une boucle diagonale et deux latérales) opposés ne sont pas équivalentes. On notera que la topologie hybride a été introduit tardivement (en 2006 (50)), et par conséquent une grande partie de la littérature antérieure ne mentionne que les topologies parallèles et antiparallèles d'une façon qui peut maintenant nous paraître erronée (51).

### 1.1.2 Les ligands de G-Quadruplexes

La découverte de molécules capables de se lier sélectivement à ces structures est une piste thérapeutique prometteuse (52–54). Les premières générations de ligands de G4s a été fortement inspirée par les intercalants d'ADN duplex. Ainsi, un des premiers ligand de G4 connu est **RHPS4**, qui est un ligand d'ADN duplex (Figure 1.6) (55). Compte tenu de la différence de taille et de forme entre les tétrades et les paires de bases, l'augmentation de l'affinité et de la spécificité peut reposer sur une simple augmentation de la surface aromatique afin d'optimiser son chevauchement avec tétrades tout en empêchant son intercalation. Le **PhenDC3**, qui présente une excellente affinité pour les G4s et une sélectivité élevée contre l'ADN duplex, est un bon exemple d'optimisation rationnelle de la taille et de la forme du cœur aromatique (56). L'ajout de chaînes latérales protonables et l'optimisation de leur nature et de leur longueur peuvent également contribuer favorablement à l'interaction, moduler la sélectivité et améliorer la solubilité dans l'eau (57–62). Un exemple précurseur est **BRACO-19**, une acridine 3,6,9-trisubstituée dont les pyrrolidines des chaînes latérales permettent des interactions favorables avec les sillons, tandis que le substituant 9-anilino se positionne dans un troisième sillon (55, 57, 63). De même, la **pyridostatine** est un dérivé de la **360A** dans lequel trois chaînes latérales ont été ajoutées et les quinolines ne sont plus N-méthylées (64).

## 1.2 Orientations des travaux de thèse

Mon sujet de thèse était initialement uniquement dédié à la synthèse de complexes métalliques comme ligands de l'ADN quadruplexe. Il devait venir continuer (une partie) du travail d'une précédente doctorante, le Dr. Hélène Bertrand, qui a produit plusieurs articles sur le sujet (65–68). Comme Hélène part au moment où j'arrive, c'est David Monchaud, chargé de recherche CNRS au laboratoire, qui me prend sous son aile et m'inculque tout ce que j'ai besoin de savoir sur l'ADN, et plus, le jour de mon arrivée. L'ampleur de la tâche et la force du café de David me donnera des palpitations. Il me forme

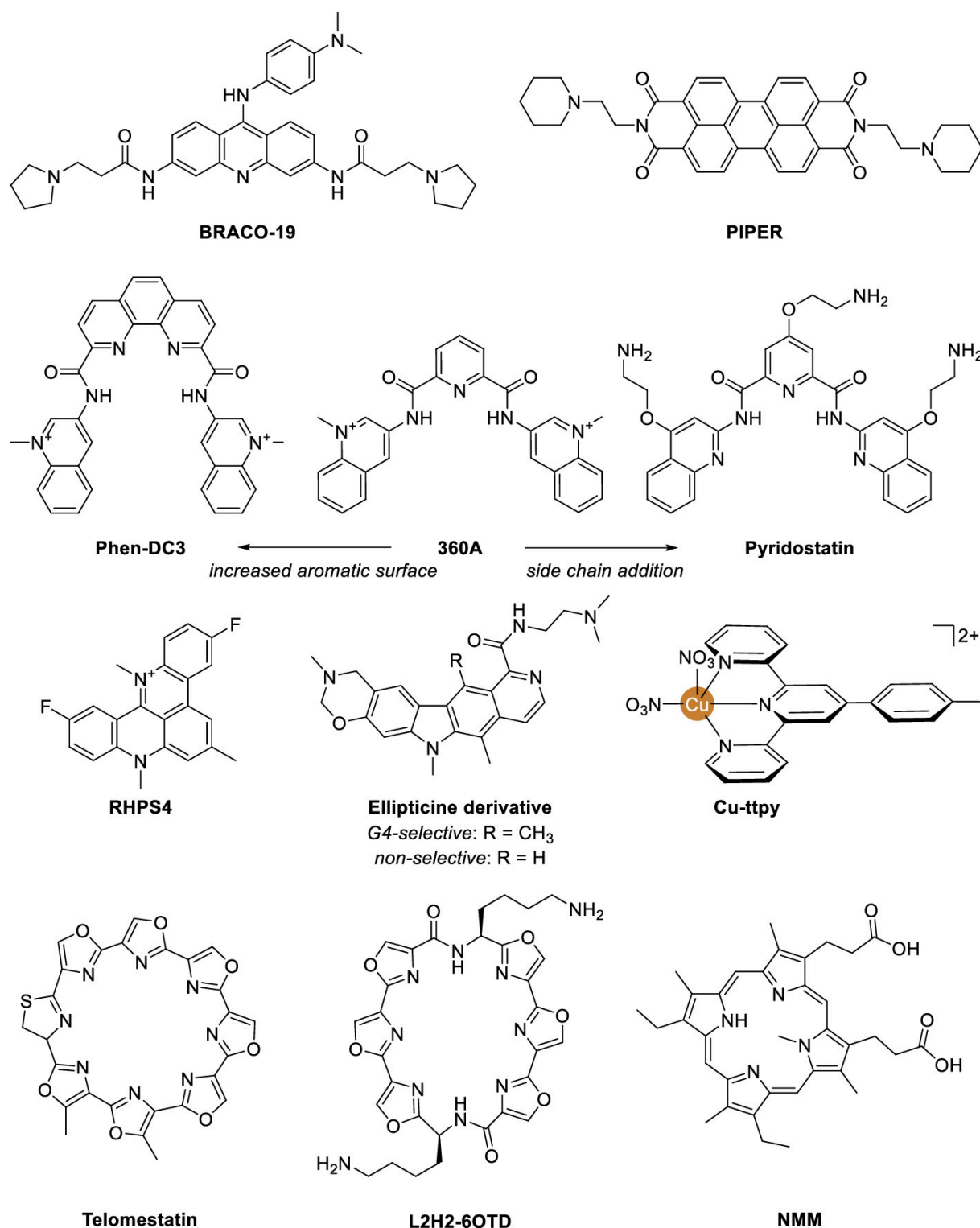


FIGURE 1.6. Ligands de G-quadruplexes . Certains étaient déjà publiés au début de ma thèse.



également sur les techniques spectroscopiques que je serai amené à utiliser (69–71), ce qui sera d'une grande importance.

Dans les faits, au début de mes travaux de thèse (en octobre 2008), de nombreuses molécules ont déjà été publiées, dont certaines - comme **360A**, **PhenDC3** ou **Cu-ttpty** - dans mon laboratoire de thèse, avec des sélectivités significatives en faveur des G4s (sans pour autant discriminer efficacement différentes topologies) (56, 66). J'entends d'ailleurs ici et là qu'on a fait le tour de la question, que les G4s sont un effet de mode et que dans 5 ans on en parlera plus, comme les triplexes.<sup>1</sup> Il s'agit d'un élément de contexte important, car il va naturellement m'imposer un défi scientifique qui n'était pas explicitement prévu : il ne s'agit pas de simplement synthétiser de petites molécules qui se lient aux G4s mais de *découvrir* des ligands à valeur ajoutée. Je présente dans ce chapitre mes tentatives de ligands covalents, fluorescents, biotinylés, et qui discriminent des topologies de G4s, mais nous verrons que j'y reviendrai bien plus tard (Section 5.3) (72). J'utilise ici le terme *découverte* car ces ligands ne sont pas tous le fruit du *design* rationnel. Pour obtenir des ligands "intéressants", je vais aussi m'intéresser au criblage de petites molécules, ce qui va m'amener à développer des techniques analytiques *ad hoc*. Cela constituera un point de bascule dans mon parcours et enterrera ma carrière de chimiste organicien - ce qui n'est pas une grande perte pour la recherche. Cela n'aurait pas été possible sans l'apport inestimable de Florian Hamon (Figure 1.1; maintenant *Head of New Drug Application group*, chez Curium Pharma), qui a rejoint l'équipe comme postdoctorant environ un an après le début de mon doctorat. C'est lui qui prendra en charge le gros du travail de synthèse, étant de son côté peu attiré par les aspects analytiques. Cette complémentarité entre nos profils et envies a sans aucun doute joué un rôle majeur pour ma thèse et la suite de ma carrière.

En résumé, deux axes de travail interconnectés se sont donc dégagés :

1. Synthèses de nouveaux ligands organiques et inorganiques d'ADN quadruplexe,
2. Développement de méthodes analytiques pour l'analyse d'interactions entre ligands et acides nucléiques,

L'application des méthodes de l'axe 2 permet de caractériser les ligands préparés à l'axe 1, mais aussi d'en découvrir de nouveaux par criblage de chimiothèques.

### 1.3 Axe 1. Synthèse de nouveaux ligands organiques et inorganiques de haute affinité et sélectivité

Comme évoqué ci-dessus, les ligands de G4s publiés au début des années 2000 sont des molécules hétérocycliques planes, inspirés des intercalants de l'ADN duplex. Leur surface aromatique est généralement étendue par rapport à ces derniers pour maximiser le  $\pi$ -stacking sur les G-quartets tout en empêchant l'intercalation dans les duplexes (1, 73). Plusieurs possèdent donc de bonnes voire excellentes sélectivités en faveur des G4s, mais pas entre plusieurs G4s (voir Section 1.4.1.2.1). Il existe également à l'époque peu de ligands possédant des fonctions supplémentaires que celle de se lier aux G4s. Une part importante de mes travaux de thèse a donc été la synthèse et caractérisation de nouveaux

1. Généralement par des personnes qui travaillent toujours dans le domaine.



ligands organiques et inorganiques de haute affinité et sélectivité, possédants un ou plusieurs modes d'interaction et/ou des fonctionnalités d'intérêt (fluorescence pour la détection de G4s, couplés à de la biotine pour des expériences de *pull down*, etc.). J'ai ainsi synthétisé, ou participé à la synthèse, de plusieurs séries de ligands que je présente brièvement ci-après.

### 1.3.1 Complexes organométalliques comme ligands covalents et non-covalents des G4s

Ce projet est fondamental dans ma carrière, et en un sens forme le trait d'union entre ma formation initiale et certains de mes projets actuels. Il s'agit de mon projet initial de thèse, celui qui était présenté sur l'annonce à laquelle j'ai répondu, et sur lequel j'ai passé trois ans - mais pas à temps plein comme on le verra ci-après.

Le projet reprend là où ce sont arrêtés (l'excellente) doctorante qui m'a précédé, le Dr. Hélène Bertrand, désormais professeure au Laboratoire des BioMolécules (ENS, Paris), et Eric Talbot, un étudiant de Master ayant comme moi étudié à l'École Supérieure de Chimie Organique et Minérale, et maintenant *Head of Synthetic Drug Design* chez *Isomorphic Labs* (Londres, RU)<sup>2</sup>. Très brièvement, plusieurs complexes métalliques, particulièrement de platine et de cuivre, avec des ligands hétérocycliques (terpyridine : **tpy**, tolyl-terpyridine : **ttpy**, dibenzoterpyridine : **BisQ**; Figure 1.7) ont été synthétisés et montre des affinités et sélectivités prometteuses pour les G4s (65, 66, 68). En particulier, il apparaît que :

1. Les complexes de platine peuvent *platiner* l'ADN, ce qui est fréquemment exploité dans le cadre de thérapies anticancéreuses (inhibition de la réplication, altération des mécanismes de réparation) (74);
2. Les ligands apicaux des complexes de cuivre(II), rendant la molécule non-plane, pourraient jouer un rôle dans cette sélectivité : ils empêcheraient l'intercalation dans les duplexes et pourraient également empêcher le *stacking* sur des G4s dont les tétrades sont coiffées de boucles, entraînant une gêne stérique (66).

Ce projet sied le jeune doctorant que je suis car je sors d'un stage de M2 recherche centré sur la synthèse de molécules organométalliques, et il se situe à l'interface avec la biologie, ce qui m'attire depuis mon stage ingénieur chez Sanofi Aventis. Il sera aussi le projet qui m'amènera à collaborer avec Valérie Gabelica et Frédéric Rosu, ce qui conditionnera beaucoup la suite des événements.

Une série de ligands hétérocycliques, azotés, tridendates commerciaux (terpyridine, tolylterpyridine) ou que j'ai synthétisés (**BisQ**, **tBisQ**, **tMebip**; Figure 1.7) a été utilisée pour obtenir quinze complexes de palladium (II), platine (II) et cuivre (II) (75). La synthèse de **BisQ** a déjà été décrite au laboratoire (65), en revanche celles de **tBisQ** et **tMebip** sont originales.

Pour **tBisQ**, l'intermédiaire clé, la **2-acétylquinoline**, a été obtenu par estérification de l'acide quinaldique, suivie d'une condensation de Claisen, puis d'une étape de saponification/décarboxylation (Figure 1.8). Le ligand a finalement été obtenu avec un rendement de 54% par une double réaction de Kröhnke, en utilisant une double condensation avec de l'ammoniaque sur la **2-acétylquinoline**,

2. À ne pas confondre avec l'artiste qui a travaillé sur les *Tortues Ninja*

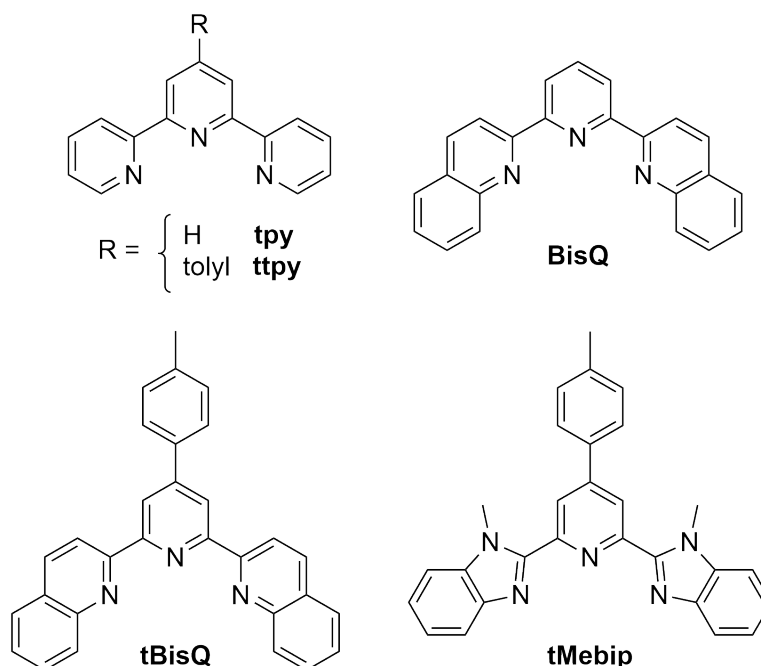


FIGURE 1.7. Ligands tridentates azotés utilisés pour la préparation de complexes de métaux de transition comme ligands de G4

l'acétate d'ammonium et le 4-méthylbenzaldéhyde dans un tube scellé à 100°C pendant 24 heures (76). Le rendement final fut de 27%.

Le **tMebip** a été préparé en trois étapes à partir de l'acide chélidamique, avec un rendement global de 45% (Figure 1.9). La formation des benzimidazoles a été obtenue par cyclisation thermique de l'acide chélidamique et de la *N*-méthyl-1,2-phénylènediamine en présence d'acide polyphosphorique, qui a été utilisé comme solvant (77). La position 4 de la pyridine a ensuite été bromée par l'oxybromure de phosphore (78), ce qui a permis d'introduire le groupe tolyle par un couplage de Suzuki (79). Le produit final a été facilement obtenu car il précipite à température ambiante dans le mélange THF/eau utilisé comme solvant. Cette synthèse a été réalisée à l'échelle du gramme et nécessite peu d'étapes de purification.

Les complexes de palladium, de platine et de cuivre de **BisQ**, **tBisQ** et **tMebip** ont été préparés en adaptant les procédures décrites dans la littérature (66, 80–82). Ces complexes ont été obtenus avec des rendements inférieurs à ceux obtenus avec **tpy** et **ttpy**, probablement en raison de leur plus grande surface aromatique, qui délocalise les doublets non liant de l'azote. Ce sont d'ailleurs mes difficultés à produire le complexe **Pt(II)-tMebip** qui m'ont amené à préparer le complexe de Pd(II), cet ion étant plus réactif. Il ne s'agit alors que du deuxième exemple publié de complexes de palladium comme ligand de G4s, le premier venant de l'équipe de Ramon Vilar (83).

Nous avons observé une augmentation de la température de dénaturation des G4s impressionnante en présence des complexes de palladium, qui suggère une interaction forte (collaboration avec Aurore Guédin et Jean-Louis Mergny, ARNA, Bordeaux) (Figure 1.10A). Les complexes de platine et de cuivre donnent des stabilisations moindres, montrant l'importance de la nature du métal. La structure du ligand organique exerce également une influence, les surfaces aromatiques plus larges favorisant le *stacking*.

### 1.3 Axe 1. Synthèse de nouveaux ligands organiques et inorganiques de haute affinité et sélectivité

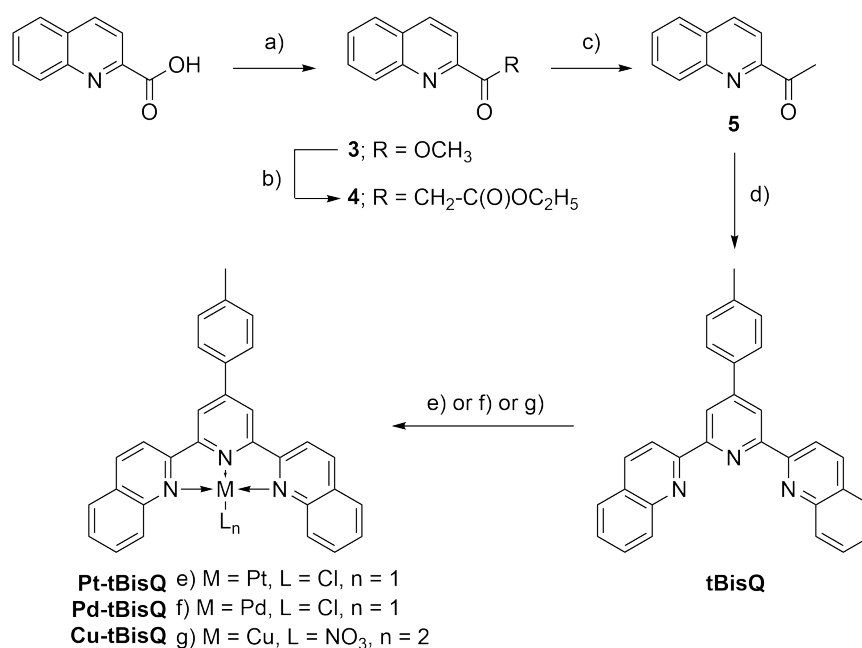


FIGURE 1.8. Synthèse de **tBisQ**. **a.** acide sulfurique, MeOH, reflux, 16 h (89%); **b.** tBuOK, acétate d'éthyle, t.a., 15 min (79%); **c.** HCl, dioxane, 100°C, 16 h (70%); **d.** 4-méthyl-benzaldéhyde, KOH, ammoniac, EtOH, 100°C, tube scellé, 24 h (54%); **e.** K<sub>2</sub>[PtCl<sub>4</sub>], nitrométhane, reflux, 24 h (94%); **f.** [Pd-(cod)Cl<sub>2</sub>] (cod = 1,5-cyclooctadiène), DMF, 50°C, 24 h (40%); **g.** Cu(NO<sub>3</sub>)<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>, acétonitrile

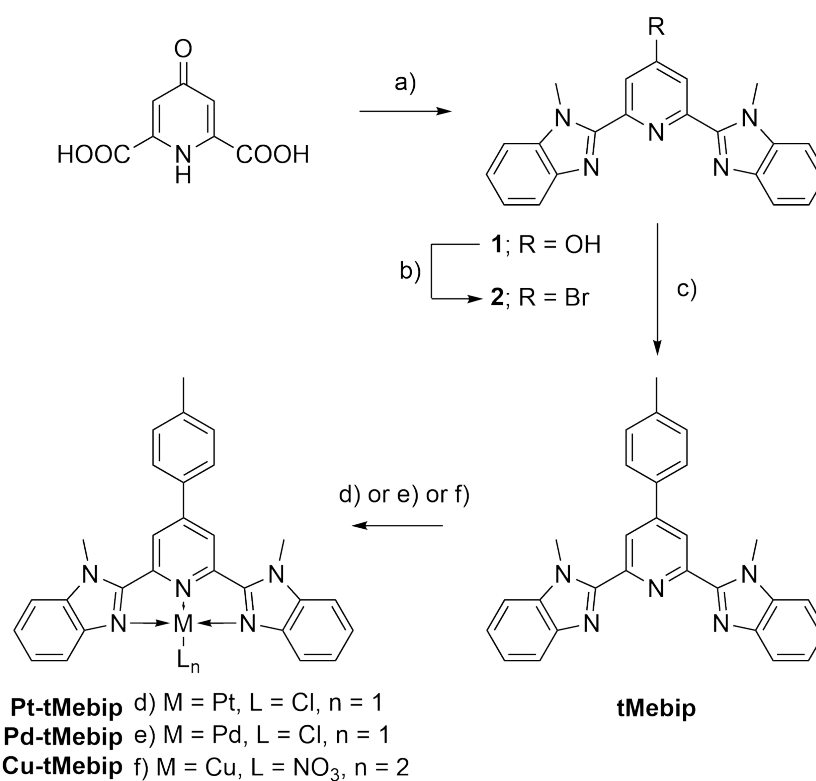


FIGURE 1.9. Synthèse de **tMebip**. **a.** *N*-méthylbenzène-1,2-diamine, acide polyphosphorique, 215°C, 17 h; **b.** POBr<sub>3</sub>, 140°C, 17 h; **c.** acide *p*-tolylboronique, [Pd(PPh<sub>3</sub>)<sub>4</sub>], K<sub>2</sub>CO<sub>3</sub>, THF/eau, reflux, 24 h (45% sur ces trois étapes); **d.** K<sub>2</sub>PtCl<sub>4</sub>, nitrométhane, reflux, 24 h (15%); **e.** [Pd(cod)Cl<sub>2</sub>], DMF, 50°C, 24 h (48%); **f.** Cu(NO<sub>3</sub>)<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>, acétonitrile, 48°C, 48 h (42%).

L'effet particulier des complexes de palladium s'explique par une coordination quantitative du cation métallique aux bases de l'ADN, que nous avons mis en évidence par spectrométrie de masse native avec Frédéric Rosu et Valérie Gabelica, alors à l'université de Liège, en Belgique (Figure 1.10B). Trois espèces ont été distinguées résultant de i) la coordination directe entre l'ADN et le complexe métallique, et du  $\pi$ -stacking non covalents entre l'ADN et les ii) formes chloré ou iii) *aqua* du complexe métallique (Figure 1.10C). Pour les complexes de platine, la forme principale liée à l'ADN est la forme *aqua* (Cl est remplacé par OH), mais quelques complexes chlorés associés à l'ADN sont encore observés. Le complexe directement coordonné à l'ADN, pour lequel un résidu de l'ADN occupe le quatrième site de coordination, est mineur pour les complexes de platine, mais très largement prédominant pour les complexes de palladium (quantitatif à température ambiante en moins d'une heure), ce qui est cohérent avec la plus grande réactivité de ce dernier.

J'ai pu confirmer cette coordination par une expérience simple de spectroscopie UV-visible, où l'on peut observer un décalage vers le rouge de la bande MLCT en présence de ligand (Figure 1.10D). La substitution d'un atome attracteur d'électrons (le chlore) par un donneur neutre (le doublet non liant d'azote d'une base de l'ADN) augmente la densité électronique du centre métallique et, par conséquent, diminue l'énergie de la MLCT (84). Je découvre à cette occasion mon appétence pour la caractérisation d'acides nucléiques par des techniques spectroscopiques "simples", ce que j'aurai l'opportunité d'exploiter à plusieurs reprises pendant ma carrière.

Finalement, les complexes **[Pd(ttpy)]** et **[Pd(tMebip)]** sont également plus efficaces que leurs homologues à base de platine pour inhiber la croissance des cellules cancéreuses *in vitro*. Plus généralement, ces résultats démontrent que l'affinité pour l'ADN G4 et le mode d'interaction des complexes métalliques peuvent être modulés en modifiant soit le métal, soit le ligand organique.

À titre personnel, la collaboration avec Valérie et Frédéric s'avérera déterminante pour la suite de ma carrière; je les rejoindrai en post-doctorat quelques années plus tard.

### 1.3.2 Ligands oligohétéroaryles acycliques qui discriminent fortement entre diverses topologies G4 en ciblant les sillons

Dans l'optique de potentielles applications thérapeutiques, les ligands de G4 doivent présenter une forte affinité pour le G4 ciblé, combinée à une faible association avec l'ADN duplex. Parmi la pléthore de ligands déjà publiés au début de ma thèse, le premier à avoir répondu à ces critères est la **télomestatine**, un macrocycle naturel à base d'oxazole (Figure 1.11) (85). La **télomestatine** est neutre, ce qui la différencie de la grande majorité des ligands de G4, qui présentent des squelettes aromatiques fusionnés avec des charges cationiques. Sa forme macrocyclique est suspectée d'être le déterminant principal de son interaction avec les G4s et de son absence d'affinité pour l'ADN duplex. Pour répondre à cette question importante dans l'établissement des règles de reconnaissance des G4s, nous avons synthétisé de nouveaux ligands neutres, mais non macrocycliques, à base de motifs oxazole et pyridine alternés (Figure 3A) (86, 87).

Les premiers exemples, **TOxaPy** et **BOxaPy**, ont été initiés lors de la thèse de Myriam Rouchon Dagois, soutenue en 2008, sous la direction de Marie-Paule Teulade-Fichou et David Grierson et supervisée par

### 1.3 Axe 1. Synthèse de nouveaux ligands organiques et inorganiques de haute affinité et sélectivité

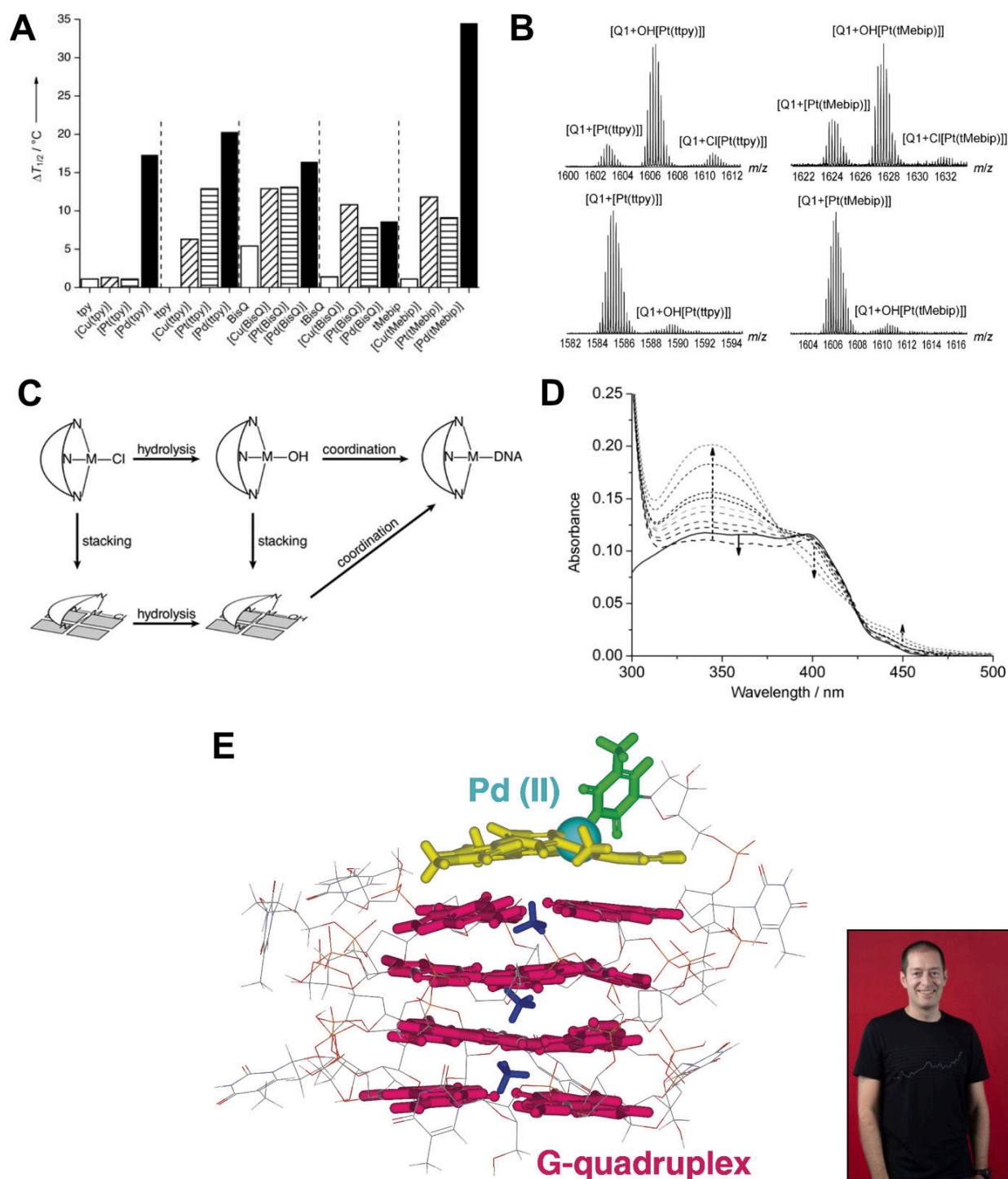


FIGURE 1.10. Les complexes de palladium stabilisent remarquablement les G4s (FRET-melting; A : barres noires), ce qui s'explique par la combinaison d'interactions non-covalentes (*stacking* sur les G-quartets), et de coordination des bases de l'ADN (C), en particulier par les complexes de Pd(II) comme mis en évidence par ESI-MS native (B) et titrations UV-vis (D). Le modèle a été optimisé par 10 ns de dynamique moléculaire dans un champ de force AMBER (parm99) après docking et coordination du ligand, lui-même optimisé par DFT (hybrid functional B3LYP, 6-31G(d) basis set), et optimisation géométrique d'une ns (E). Encart : Frédéric Rosu.

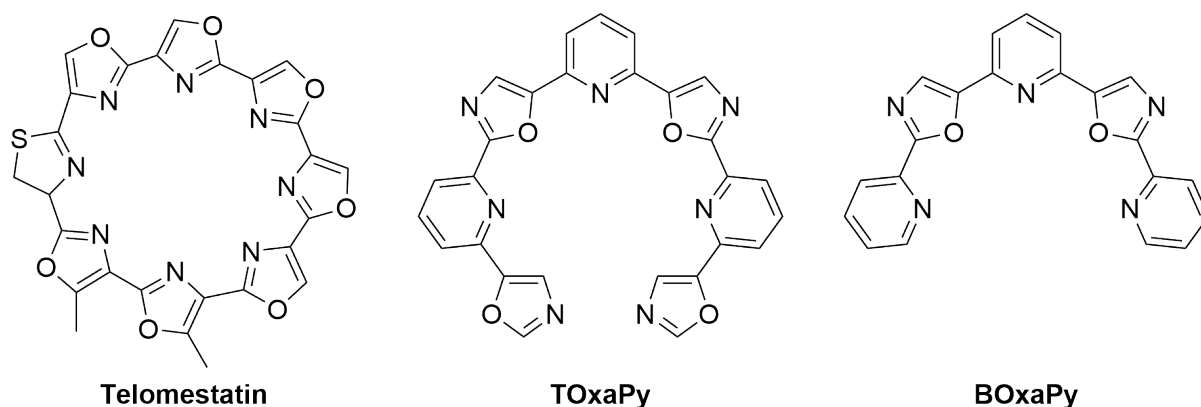


FIGURE 1.11. Les ligands **TOxaPy** et **BOxaPy** sont des analogues acycliques du ligand naturel de G4 **télomestatine**

Chi Hung Nguyen.

Leur synthèse initiale s'est appuyée sur une procédure convergente basée sur le couplage de la 2,6-bis(oxazol-5-yl)pyridine avec des dérivés de la 2-bromopyridine par la double activation des CH des oxazoles (88) (Figure 1.12). La 2,6-bis(oxazol-5-yl)pyridine a été obtenue en deux étapes par l'oxydation de la 2,6-lutidine avec du dioxyde de sélénium pour obtenir la pyridine-2,6-dicarbaldéhyde (89), qui a été soumise aux conditions de Van Leusen. Le couplage dans des conditions classiques avec deux équivalents molaires de 2-bromopyridine a conduit au penta-hétéroaryle **BOxaPy** (bisoxazoletrispyridine). La même réaction de couplage a été effectuée en utilisant le 2-bromo-5-(pyridine-2-yl)oxazole pour préparer l'homologue heptaaryle **TOxaPy** (tétraoxazoletrispyridine).

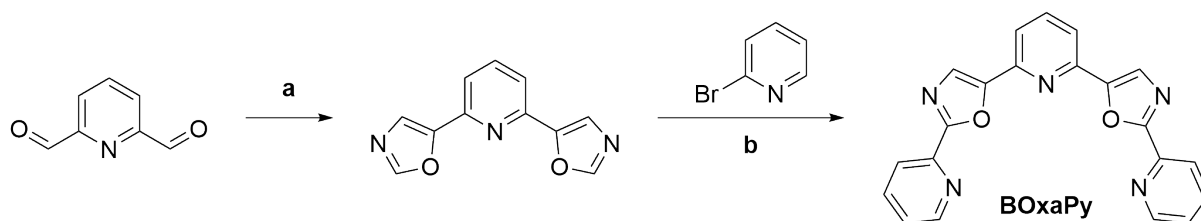


FIGURE 1.12. Première approche synthétique, produisant **BOxaPy** et un mélange de deux régioisomère de **TOxaPy**.  
**a.** TosMIC,  $K_2CO_3$ , MeOH, 50°C, 16 h (50%); **b.** CuI, Pd(OAc) $_2$ ,  $CS_2CO_3$ ,  $PCy_3 \cdot HBF_4$ , dioxane, 130°C, 24 h (45%);  
 Cy = cyclohexyl.

Nous nous satisfaisons de l'analyse en masse qui confirme l'obtention de **TOxaPy**. Cependant, un *reviewer* remarque, fort justement, qu'après la première condensation de 2-bromo-5-(pyridine-2-yl)oxazole, la seconde peut avoir lieu soit sur l'oxazole symétriquement opposé au premier site de condensation, soit sur l'oxazole qui vient d'être introduit (Figure 1.13). Cela produit un isomère géométrique de même masse, ne différant que par une jonction pyridine-oxazole. Nous avons confirmé la clairvoyance du *reviewer* par analyse LC et MS à partir de la 2,6-bis(oxazol-5-yl)pyridine deutérée sur les positions 2 des oxazoles : il y avait bien deux isomères en proportions 50/50!

Il s'agit d'un moment clé de ma thèse, où l'importance de l'utilisation de techniques analytiques multiples et orthogonales, de la révision par les pairs, et de la remise en question permanente des résultats deviennent évidentes. Je ne le sais pas encore, mais ce premier marquage isotopique au deutérium en appellera beaucoup d'autres.

Pour obtenir le **TOxaPy** sans ambiguïté, Florian Hamon et moi avons suivi une nouvelle voie dans

### 1.3 Axe 1. Synthèse de nouveaux ligands organiques et inorganiques de haute affinité et sélectivité

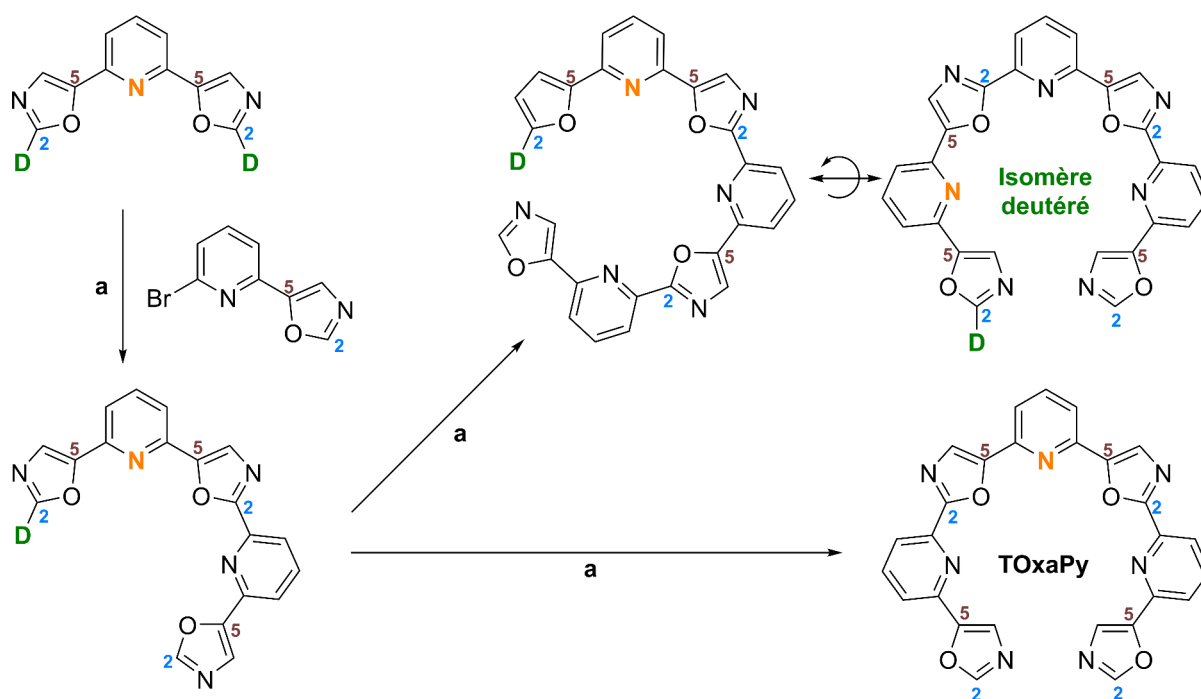


FIGURE 1.13. Synthèse initiale de TOxaPy produisant un mélange 50/50 de deux isomères variant par une unique jonction pyridine-oxazole, mise en évidence ici par deutération de la 2,6-bis(oxazol-5-yl)pyridine, dont l'azote central est coloré en orange. Les positions 2 et 5 des oxazoles sont numérotées en bleu et marron, respectivement. Conditions expérimentales : **a.** Pd(OAc)<sub>2</sub>, PCy<sub>3</sub>·HBF<sub>4</sub>, CuI, Cs<sub>2</sub>CO<sub>3</sub>, toluène, dioxane, 130°C.

laquelle des groupements oxazoles terminaux ont été synthétisés à partir d'intermédiaire bisformyle pentacyclique correspondants (Figure 1.14). Dans cette approche, la 2,6-bis(oxazol-5-yl)pyridine a été couplé à la 2-bromo-6-(1,3-dioxolan-2-yl)pyridine pour obtenir le précurseur bisacétal qui, après déprotection puis traitement au TosMIC (isocyanure de *p*-toluènesulfonylméthyl) à produit **TOxaPy** avec un rendement global de 19%.

Le dérivé heptacyclique **TOxaPy** s'est révélé totalement dépourvu d'affinité pour l'ADN duplex tout en présentant une préférence sans précédent pour certaines topologies de G4 par rapport à d'autres. En particulier, le nouveau composé reconnaît la topologie G4 (antiparallèle) de la séquence télomérique humaine formée un tampon riche en Na<sup>+</sup>, mais pas celle obtenue dans un tampon riche en K<sup>+</sup> (mélange de topologies hybrides) (Figure 1.15A). Pour confirmer les résultats du classique FRET-*melting* (Aurore Guédin et Jean-Louis Mergny, ARNA), j'ai utilisé le test G4-FID (voir Section 1.4.1) et des titrations fluorimétriques - **TOxaPy** étant fluorescent à l'état libre ( $\lambda_{exc}$  : 340 nm,  $\lambda_{em}$  : 350-550 nm) et *quenché* lorsque complexé (Figure 1.15B-C). Cette sélectivité est unique parmi les ligands publiés jusque-là (Figure 1.15D). L'interaction dépend également fortement de la taille de l'oligomère, car l'analogue pentacyclique **BOxaPy** n'interagit pas avec les G4s (Figure 1.15A).

On notera que, contrairement à la **télomestatine**, **TOxaPy** n'est pas un macrocyle. Malgré cela, il est présenté dans la publication (et ici) sous forme circulaire, sans doute pour ressembler à une **télomestatine** à laquelle il manquerait un hétérocycle. Cela induit un biais de communication visuel important, puisqu'il suggère au lecteur que ces molécules interagissent avec les G4s de façon similaire. Or, les expériences de G4-FID (Figure 1.15B), qui confirment l'absence d'interaction de **TOxaPy** avec les



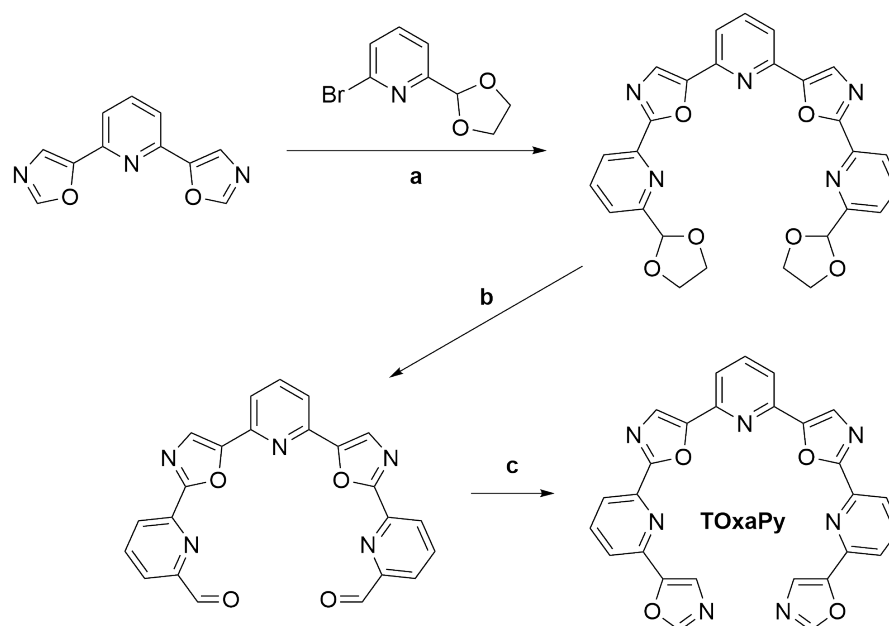


FIGURE 1.14. Synthèse régiospécifique de TOxaPy. **a.** CuI, Pd(OAc)<sub>2</sub>, Cs<sub>2</sub>CO<sub>3</sub>, PCy<sub>3</sub>·HBF<sub>4</sub>, dioxane, 130°C, 24 h (45%); **b.** HCl (2 M), 110°C, 2 h, (84%); **c.** TosMIC, K<sub>2</sub>CO<sub>3</sub>, MeOH, 50°C, 16 h (50%).

G4s télomériques en solution K<sup>+</sup>, suggèrent que la sélectivité de ce ligand s'explique par son interaction dans les sillons de G4, plutôt que par  $\pi$ -stacking sur les G-quartets. En effet, **TOxaPy** ne déplace que partiellement le thiazole orange, ce qui est cohérent avec une interaction n'ayant pas lieu sur les tétrades. L'interaction dans les sillons est *possible*, comme mis en évidence par docking par Nicolas Saettel, alors maître de conférences dans l'équipe, sans que cela ne présente une preuve formelle (Figure 1.15E).

Le principal problème de **TOxaPy** pour des études *in vitro* et *in vivo* est sa très faible solubilité dans l'eau. Une seconde série de dérivés hétérocycliques non condensés "**BOxAzaPys**", solubles dans l'eau, a ensuite été préparée en collaboration avec l'équipe du Prof. M. Freccero (Pavie, Italie) (Figure 1.16) (87). De façon très intéressante, ces composés BOxAzaPys ont une sélectivité inversée, probablement également due à une interaction dans certains sillons (Figure 1.17). Les chaînes aminées apportent d'une part la solubilité dans l'eau, et d'autre part modulent l'affinité et la sélectivité de ces composés.

Les sillons de G4s pourraient présenter des géométries plus « uniques » que ne le sont les G-quartets, et seraient donc des cibles à explorer pour obtenir des ligands sélectifs d'une topologie particulière. Ce mode d'interaction est connu pour les duplexes (et permet d'ailleurs d'améliorer la sélectivité des ligands pour des séquences particulières par rapport aux intercalants), mais n'avait pas encore été mis en évidence pour les G4s et reste à ce jour très rare (1). Notons que, si tous les G-quartets sont similaires, ils présentent un environnement qui peut varier; nous avons récemment exploré cette opportunité de discrimination (Section 5.3) (72).

### 1.3.3 Ligands de G4 possédant des propriétés physico-chimiques d'intérêt

Comme indiqué ci-avant, une partie de mes travaux de thèse était liée à la synthèse et caractérisation de ligands de G4s possédant des fonctionnalités d'intérêt. Deux stratégies ont été employées : *i.* la fonc-



### 1.3 Axe 1. Synthèse de nouveaux ligands organiques et inorganiques de haute affinité et sélectivité

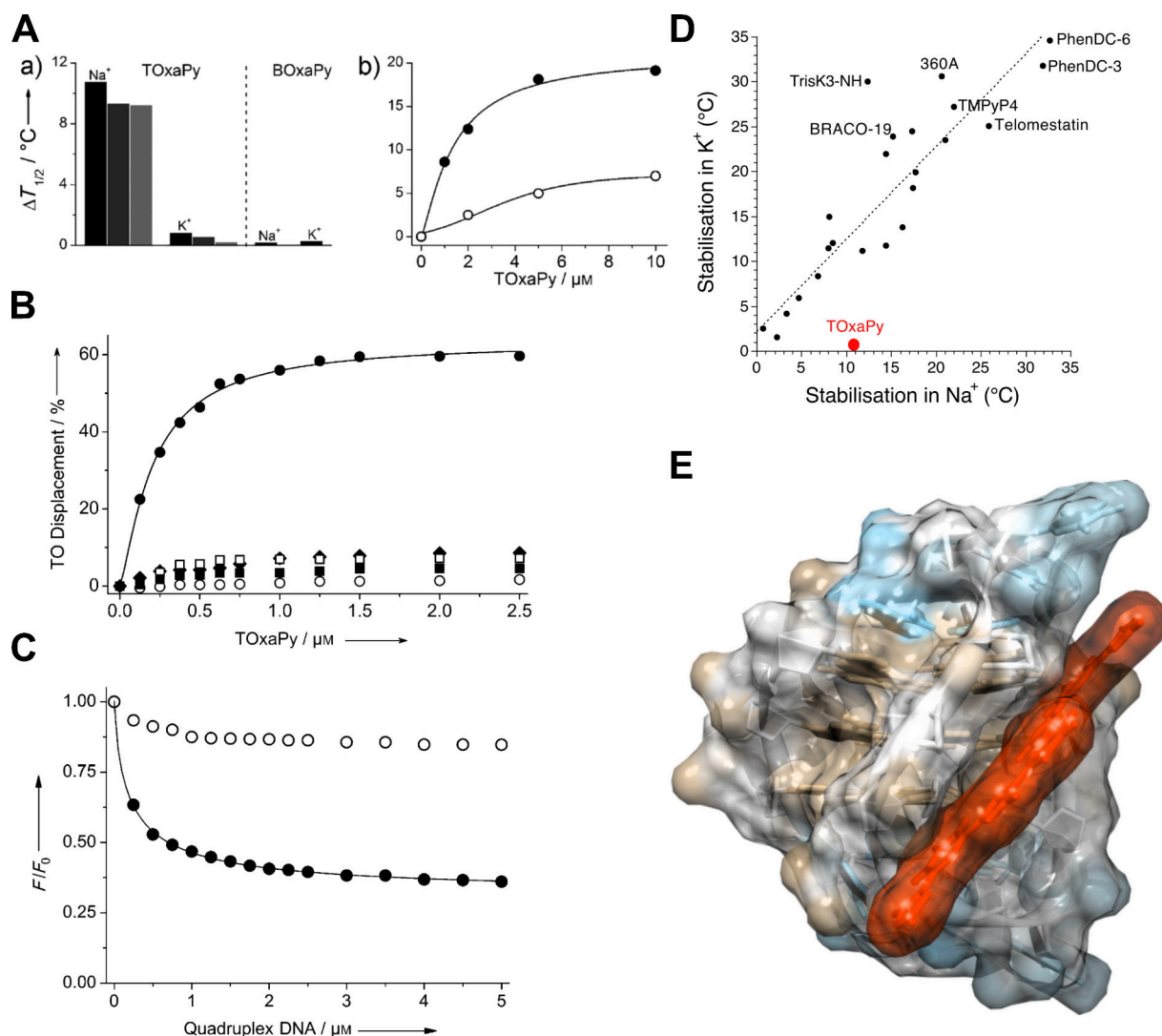


FIGURE 1.15. **TOxaPy** stabilise la topologie G4 de la séquence télomérique humaine obtenue en présence de sodium, mais pas celle observée en présence de potassium (A). Le déplacement d'un compétiteur fluorescent (test G4-FID) n'est également observé que pour cette topologie (B), tout comme le *quenching* de sa fluorescence (C). Cette sélectivité est alors exceptionnelle (D). Le déplacement n'étant que partiel, il suggère un mode d'interaction dans les sillons, illustré par des expériences de docking (E).

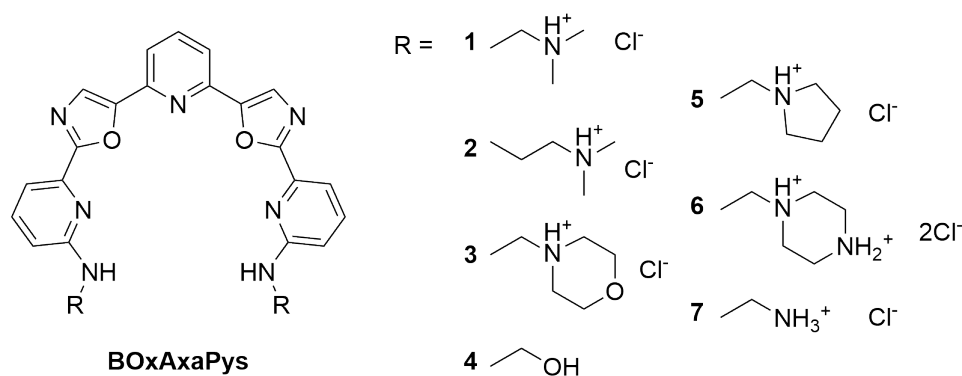


FIGURE 1.16. Structures des BOxAzaPyS

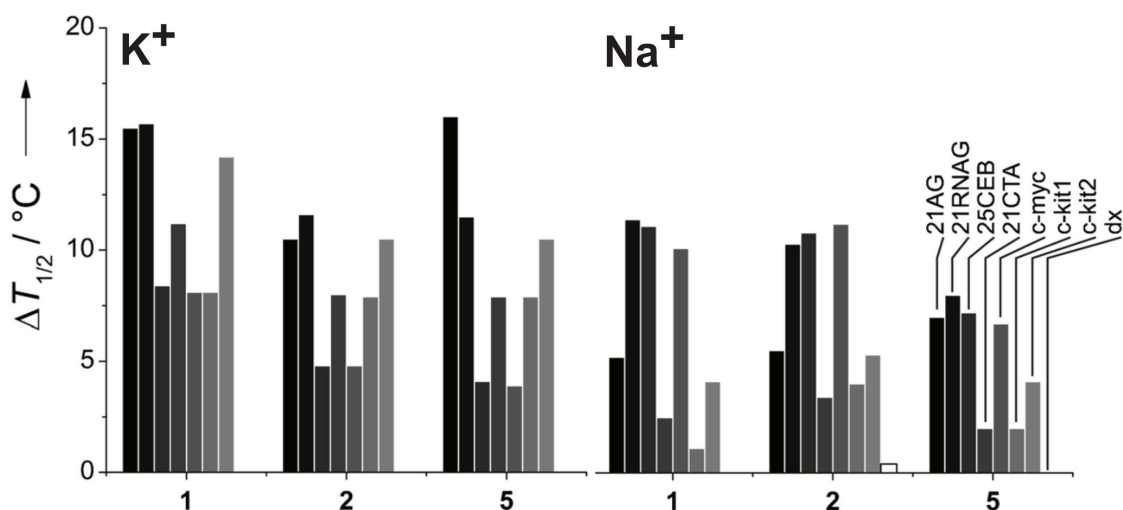


FIGURE 1.17. Les ligands **BOxAzaPys** sont des analogues de **TOxApy**, solubles dans l'eau grâce à des chaînes latérales cationiques. Leurs sélectivités sont inversées par rapport à celle de **TOxApy** : il y a plus d'interaction avec la séquence télomérique humaine en présence de potassium (A) que de sodium (B). L'affinité est également dépendante de la nature de la chaîne latérale.

tionnalisation de ligands connus (Section 1.3.3.1) et *ii.* la synthèse de ligands originaux (Section 1.3.3.2).

### 1.3.3.1 Dérivés fonctionnalisés de pyridodicarboxamides

Les ligands pyridodicarboxamides (« **PDC** »), comme **360A** (Figure 1.6), sont parmi les plus affins et sélectifs pour l'ADN G4. J'ai été impliqué dans plusieurs projets visant à leur joindre des propriétés d'intérêt. Ainsi, nous avons synthétisé les ligands suivants :

- un dérivé biotinylé qui a permis la conduite d'une expérience de SELEX pour la découverte d'aptamères G-quadruplex (90),
- un dérivé fluorescent, comme sonde d'ADN quadruplex (91),
- Des dérivés halogénés ont été synthétisés pour leur plus grande affinité et effets aux télomères (92).

Ces ligands sont décrits plus en détails en Sections 1.4.1.3.4 et 1.4.1.3.5, dans le contexte de l'utilisation de leurs propriétés.

### 1.3.3.2 Ligands fluorescents distyryles

En plus du dérivé fluorescent de **PDC**, une nouvelle famille de composés distyryles asymétriques fluorescents a été préparée par Xiao Xie, un doctorant (désormais ingénieur R&D chez *Technology Transfer Acceleration Company* / Institut Pasteur) supervisé par Anton Granzhan, et étudiée de façon à mieux comprendre comment des différences structurales subtiles influencent à la fois leurs fluorescence et sélectivité structurale (Figure 1.18) (93).

Ces dérivés interagissent sélectivement avec l'ADN G4 et ont une forte émission de fluorescence dans le rouge lorsqu'elles sont complexées (sonde fluorescentes *off/on*). Leur sélectivité optique, ils sont

### 1.3 Axe 1. Synthèse de nouveaux ligands organiques et inorganiques de haute affinité et sélectivité

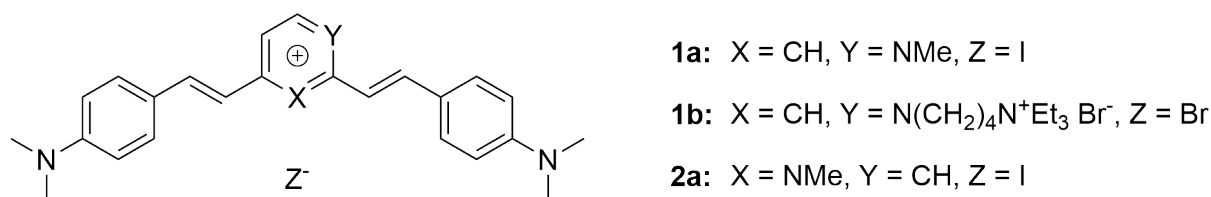


FIGURE 1.18. Dérivés distyryles fluorescents

plus de 100 fois plus fluorescent en présence de G4 qu'en absence d'ADN, ou en présence d'ADN duplex, a été mise en lumière par des titrations fluorimétriques.

Lors de précédents travaux, j'avais observé que l'ADN G4 était mal visualisé dans des gels natifs par des colorants fluorescents courants tels que le bromure d'éthidium, GelRed ou SYBR Safe, ce qui avait également déjà été noté dans des publications d'autres laboratoires (94–96). Nous avons observé que la coloration des gels PAGE avec des solutions de **1a** ou **1b** (2 mM), suivie d'une détection de la fluorescence à 670 nm, permet de détecter sélectivement les bandes correspondant au quadruplexe télomérique (22AG) et à c-myc (Figure 1.19).

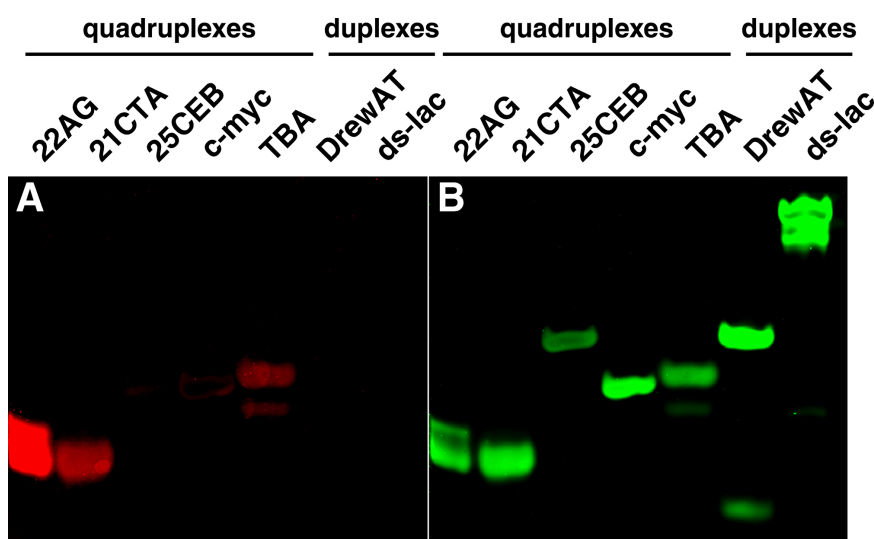


FIGURE 1.19. Une nouvelle série de ligands de G4 distyryles permet le marquage fluorescent sélectif de G4s, ici sur gel de polyacrylamide avec la molécule **1a** (A; émission dans le rouge), comparé au marquage non sélectif du thiazole orange (B; émission dans le vert).

En revanche, presque aucune fluorescence n'a été observée dans les bandes contenant des ADN duplex (ds-lac, DrewAT) et TBA. Ce quadruplexe à deux tétrades est très souvent peu lié par les ligands de G4, comme j'aurai l'occasion de le voir dans de nombreuses autres études présentées ici.

Mes contributions à la caractérisation de ces composés fluorescents, à la synthèse de complexes de platine fluorescents (jamais publié; Figure 1.21) et l'utilisation de plusieurs autres sondes lors de mes travaux de développement analytique (Section 1.4.1), m'ont amené à participer à la rédaction d'une revue exhaustive sur le sujet (11).

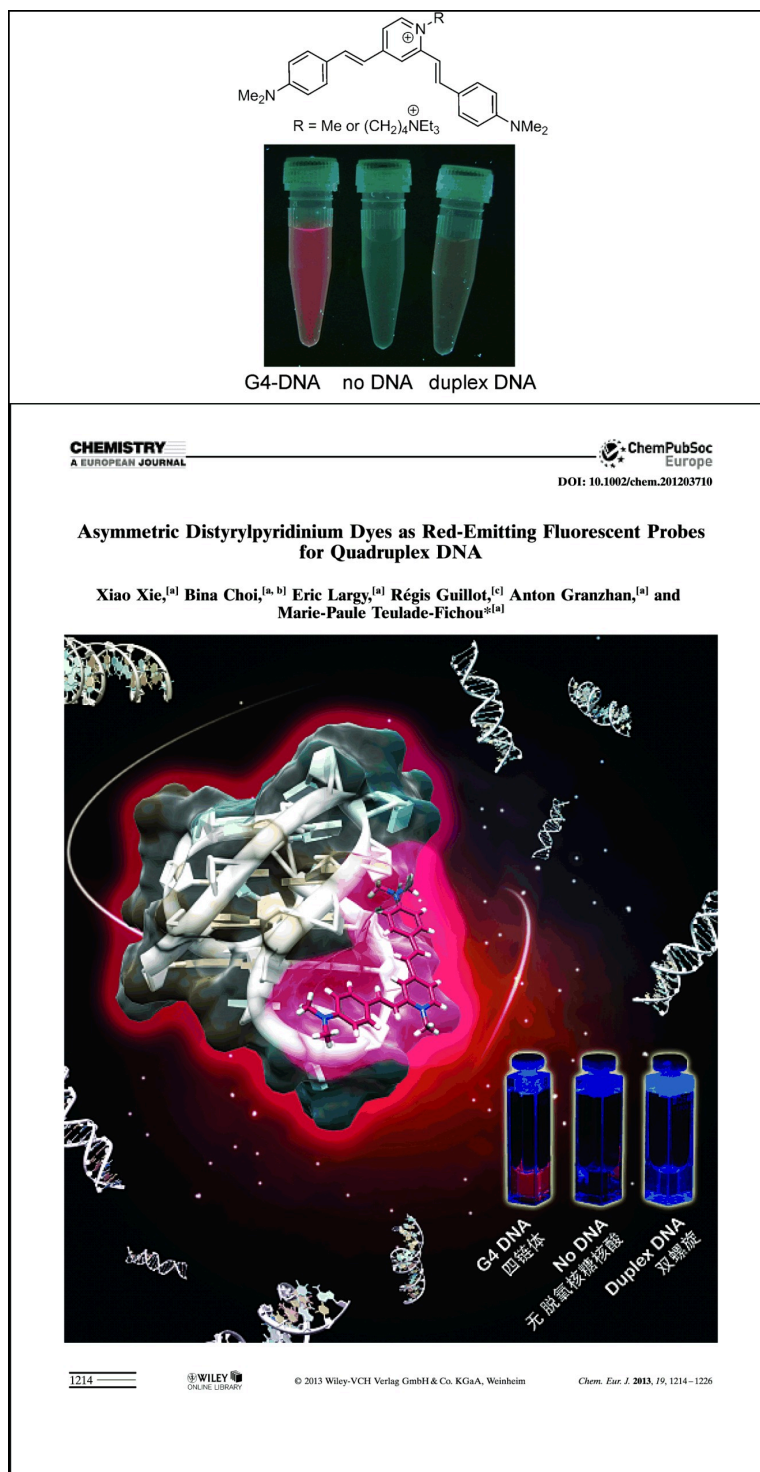


FIGURE 1.20. Frontispiece du volume 19, numéro 4 de Chemistry – A European Journal (93)

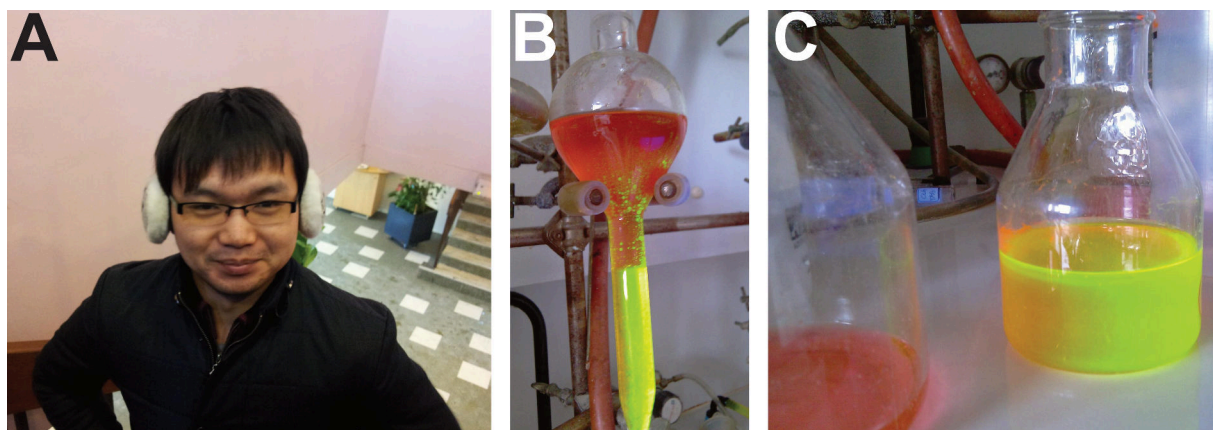


FIGURE 1.21. A. Xiao Xie, ici en conditions hivernales, a synthétisé les dérivés distyryles. B et C. Synthèse de ligands fluorescents.

## 1.4 Axe 2. Méthodes analytiques pour l'analyse d'interactions ligands-acides nucléiques

### 1.4.1 Déplacement de sondes fluorescentes de G4 a haut débit : le test HT-G4-FID

#### 1.4.1.1 Développement du HT-G4-FID

Le FID (*Fluorescence Intercalator Displacement*) est une technique qui permet de déterminer l'affinité de molécules *a priori* non fluorescentes (et ne possédant pas un autre type de propriété physico-chimiques permettant une titration) pour des oligonucléotides de diverses structures (quadruplexes, duplexes, *hairpin*, triplex), par le déplacement d'une sonde fluorescente *off/on* (Figure 1.22). Cette technique a été développée par Dale L. Boger (*Scripps Research Institute*), que j'ai eu la chance de voir présenter lors du Pacificchem 2010 (97–100).

Concrètement, l'oligonucléotide est marqué de façon non covalente par une sonde dont la fluorescence est fortement exaltée en sa présence (*on*; structure plane et isolé du solvant aqueux). L'ajout de quantités croissantes de ligand entraîne le déplacement de la sonde et donc une diminution de la fluorescence (*off*; structure en libre rotation et accessible par l'eau; Figure 1.22). Le suivi de l'intensité de fluorescence permet d'estimer l'affinité de la molécule pour l'oligonucléotide testé. Plus un ligand est affin, plus de faibles concentrations de celui-ci suffisent à déplacer la sonde.

Pour son application aux structures G-quadruplexes, on parle de test G4-FID et la sonde choisie par David Monchaud (parti depuis à l'Université de Bourgogne) lors du développement initial de la méthode est le **thiazole orange (TO)** (Figure 1.22), plutôt que le **bromure d'éthidium (EtBr)**, qui lui est préféré pour l'ADN duplex, mais fluoresce peu avec les G4s. On notera que :

- Dale Boger, qui utilisait principalement le **EtBr**, avait testé le **TO** comme sonde pour du FID sur des tiges-boucle d'ADN, en concluant qu'il constitue une alternative dans les cas où l'excitation ou l'émission des ligands étudiés chevauche celle du **EtBr** (98). C'est dans cette même optique que j'explorerais l'utilisation de sondes alternatives,
- Le FID porte mal son nom quand appliqué aux G4s car le **TO** ne s'y *intercale* pas.



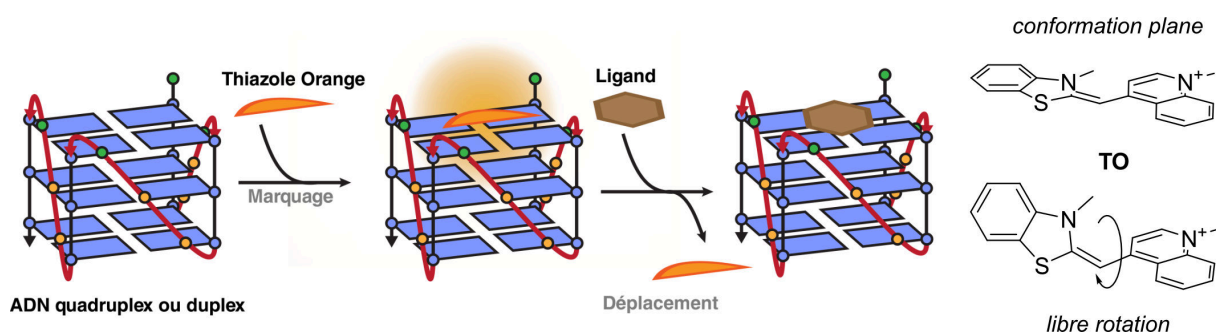


FIGURE 1.22. Principe du G4-FID. Le thiazole orange fluoresce en interagissant avec l'ADN (conformation planaire) mais pas en solution (libre rotation autour du pont méthine).

Les avantages du FID sont nombreux. C'est une méthode peu chère, rapide (notamment en microplaque; voir les sections 1.4.1.2.1, 1.4.1.2.2), exacte et précise, et dotée d'une bonne résolution. Par rapport au FRET-melting, elle a l'avantage d'être isotherme et ne nécessite qu'un fluorimètre standard. En revanche, elle ne permet généralement pas de mesurer de constantes d'affinité, même si cela est possible dans certains cas particulier que j'ai eu la chance de rencontrer (Section 1.4.1.3.2). Le FID peut aussi apporter des informations qualitatives sur le mode d'interaction. Par exemple, le déplacement partiel du **TO** par le ligand **TOxaPy** suggère que ce dernier n'interagit pas sur les G-quartets, comme le **TO**, mais plutôt dans les sillons (Figure 1.15) (86).

Après avoir utilisé la méthode telle que mise au point par David Monchaud (avec le **TO**), ou dans la version de Dale Boger (avec l'**EtBr**; Section 1.4.1.3.1), en cuvettes, j'ai souhaité développer une version à haut débit de l'essai, nommée *HT-G4-FID* (*High-Throughput Quadruplex Fluorescent Intercalator Displacement*), qui est réalisé dans des microplaques à 96 ou 384 puits (Figure 1.23) (101, 102). En effet, la réalisation du test en cuve implique une acquisition point par point des titrations, avec agitation manuelle de la cuve, et cela pour chaque couple G4/ligand. L'utilisation de plaques permet de réaliser plusieurs titrations entières en parallèle, et ouvre la porte à l'automatisation des opérations de pipetage, ce que j'ai mis en œuvre ultérieurement (Section 1.4.1.2.2).

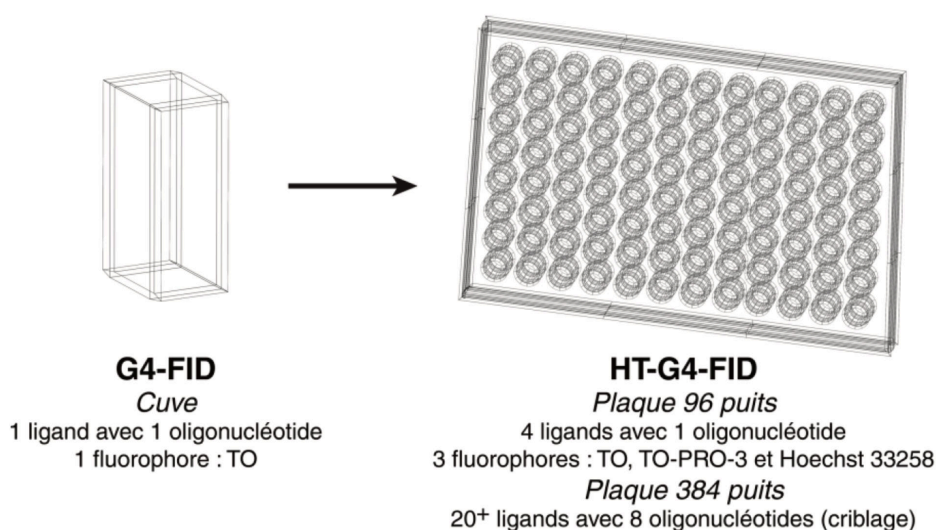


FIGURE 1.23. Transposition du G4-FID en plaque

Dans le même temps, et à l'instar de D. Boger, j'ai étendu la gamme spectrale de l'essai en utilisant

deux sondes fluorescentes supplémentaires (**Hoechst 33258** et **TO-PRO3**; Figure 1.24), dont les spectres d'absorption et d'émission sont complémentaires de ceux du **TO** (101). Cela permet de cribler des ligands susceptibles d'absorber et/ou d'être fluorescents dans la même gamme spectrale que le **TO** (comme la porphyrine **TMPyP4**), et donc de fausser les résultats. Ces améliorations ont été validées avec un panel cinq oligonucléotides d'ADN et cinq ligands d'affinité et sélectivités variées.

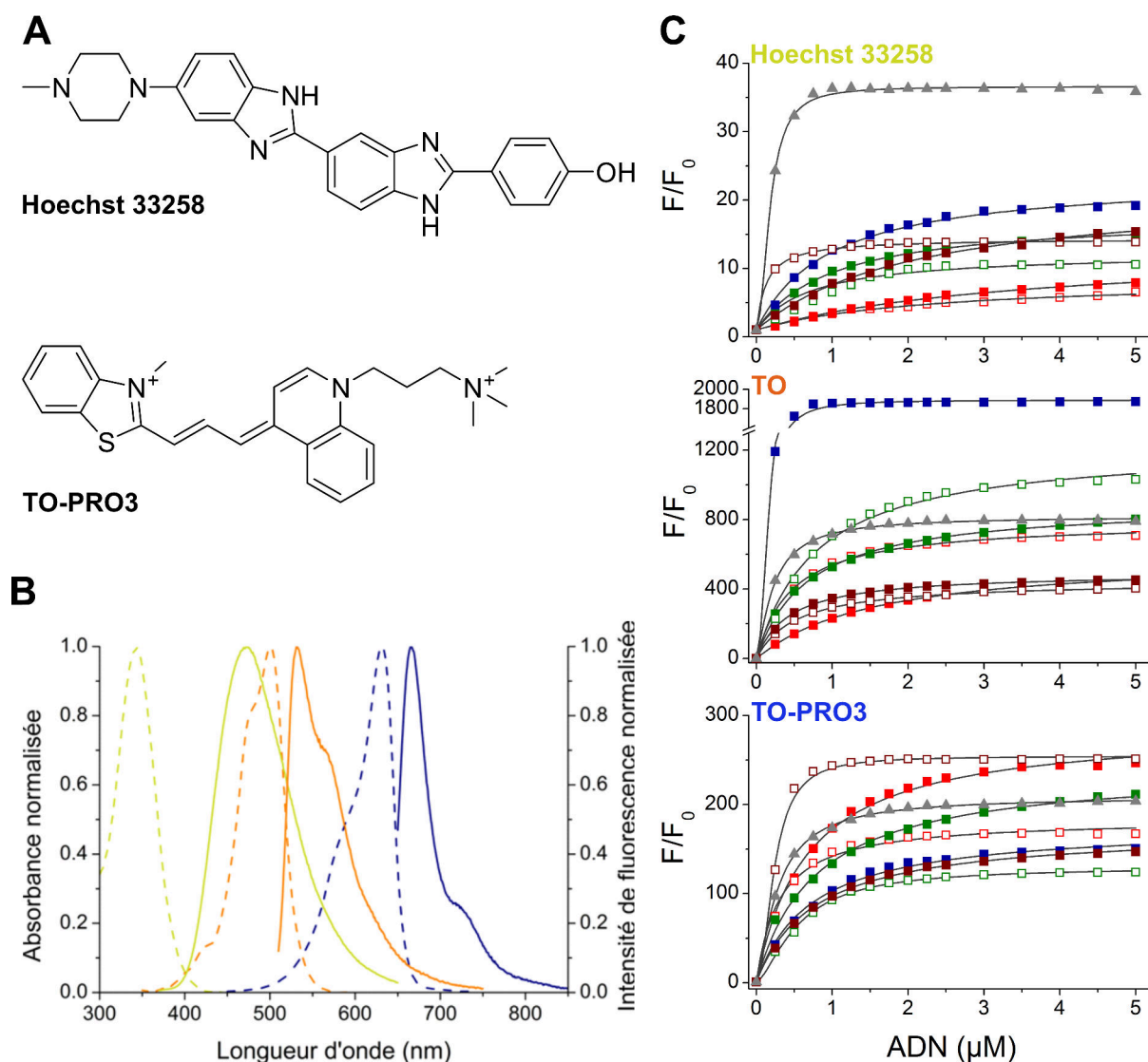


FIGURE 1.24. Extension du panel de sondes fluorescentes à **Hoechst 33258** et **TO-PRO3** (A), dont les spectres d'absorption et d'émission complètent ceux du **TO** (B). Les constantes d'affinité de ces sondes pour un panel de G4 d'intérêt ont été estimées par titration fluorimétriques (C).

Le test HT-G4-FID peut donc être appliqué à haut débit pour n'importe quel couple petite molécule / séquences d'ADN ou d'ARN de structures diverses sans qu'aucune modification soit nécessaire. Il constitue donc un outil puissant pour la découverte de nouveaux ligands capables de se lier sélectivement à l'ADN G4 vs. l'ADN duplex mais aussi entre différents G4s. Le test HT-G4-FID est fréquemment employé par la communauté scientifique, comme en témoigne le nombre élevé de citations (plus de 280 au total) (101, 102). Je présente ci-dessous mes applications les plus significatives de ce test, et par la même occasion certains projets auxquels j'ai pu contribuer pendant ma thèse.

Pour le développement de cette méthode, et certaines applications j'ai encadré deux étudiants : **Camille**

**Valette** et **Yoann Lebreton** (BTS, École nationale de chimie physique et biologie de Paris).

### 1.4.1.2 Découverte de ligands sélectifs de G4s par criblage

#### 1.4.1.2.1 Exploration d'une bibliothèque

Dans le cadre du développement du HT-G4-FID, j'ai collaboré avec l'équipe de J.-L. Mergny (Muséum National d'Histoire Naturelle, Paris) pour cribler quinze ligands de G4s connus appartenant à différentes familles contre vingt G4s et deux duplexes contrôle (102). Nous avons noté de très larges différences d'intensité de fluorescence du **TO** en fonction des oligonucléotides, ce que nous avons déjà observé précédemment mais dans une moindre mesure (Figure 1.24), et éliminé ceux pour lesquelles la fluorescence ne permettait pas de réaliser le test avec une grande sensibilité.

La plupart des composés a montré une bonne sélectivité en faveur des G4s contre les duplexes (Figure 1.25). En revanche, ils ont peu ou pas de spécificité pour une séquence G4 par rapport aux autres, ce qui limite leur potentiel thérapeutique.

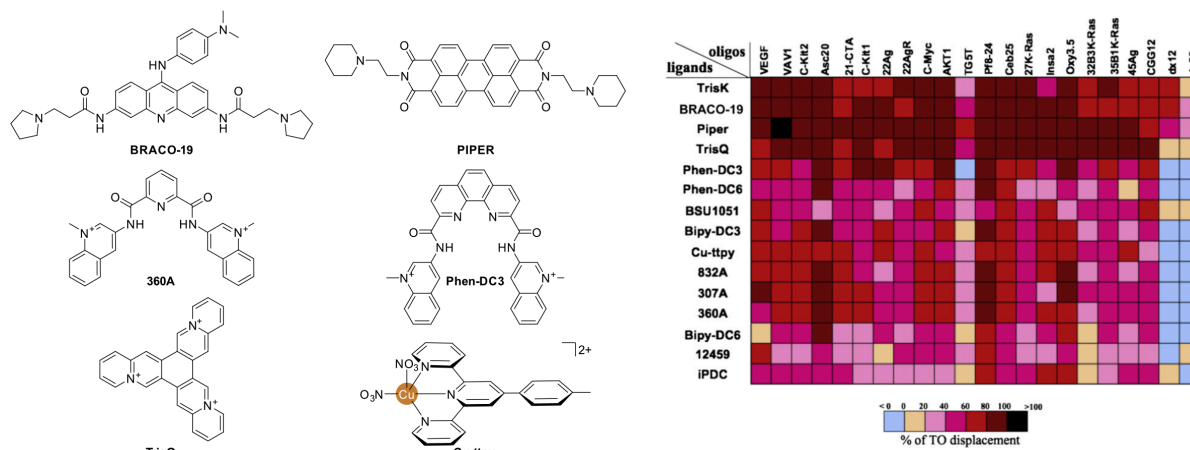


FIGURE 1.25. Exemple de ligands « classiques » de G4 (gauche) criblés par HT-G4-FID qui présentent d'excellentes sélectivités en faveur des G4s mais sans favoriser de topologies en particulier (0.5  $\mu$ M ADN, 1  $\mu$ M **TO**, 2,5  $\mu$ M ligand)

#### 1.4.1.2.2 Exploration d'une chimiothèque

J'ai présenté ci-avant la synthèse de ligands présentant une sélectivité pour certaines topologies G4s (86, 87). Si cette avancée était significative, elle n'était pas le produit d'un design rationnel. Il n'existait pas, et n'existe toujours pas, de « guide moléculaire » décrivant les règles à suivre pour cibler une structure en particulier. J'ai alors estimé que le design rationnel ne pouvait pas être la seule approche pour obtenir des ligands de G4s présentant des sélectivités significatives pour certains G4s.<sup>3</sup>

En tant qu'approche complémentaire, le criblage peut fournir de nouvelles pistes de travail pour peu que des outils analytiques adaptés soient disponibles (103), ici le test HT-G4-FID. J'ai criblé la chimiothèque de l'Institut Curie/CNRS pour identifier des candidats, avec l'assistance de deux élèves ingénieures, **Valentine Boyer** (EBI, Cergy) et **Marie Perrot** (ESCOM, Cergy) (Figure 1.26A) (104).

3. Ce qui m'arrangeait bien étant donnée ma faible appétence pour la synthèse organique.



La première étape a permis la sélection des candidats les plus prometteurs par des méthodes chimio-informatiques, en particulier des calculs de similarité moléculaire avec des ligands de référence (**PhenDC3**, **360A**, **MMQ12**; Figure 1.26B). Les molécules retenues ont été évaluées par HT-G4-FID contre divers G4s et un duplex contrôle (Figure 1.26C). La compréhension de la relation structure/activité de ces quatre ligands ellipticines (Figure 1.26D) s'est également appuyée sur des expériences de *docking* (Figure 1.26E). Nous avons ainsi mis en évidence que seul le ligand ayant un groupe méthyle en position 10 était sélectif contre l'ADN duplex, car il entraîne une perte de planéité de la molécule, ce qui en fait un mauvais intercalant (Figure 1.26D,F).

L'approche par similarité chimique n'est pas un très bon choix pour découvrir des ligands de G4 sortant du paradigme actuel (ligands hétérocycliques condensés plans), ce qui n'était cela dit pas le but premier de cette étude. Pour cela, il aurait été préférable de cribler l'entièreté de la chimiothèque, ce qui aurait malgré tout sans doute été vain étant donné que celle-ci était fortement biaisée en faveur de petites molécules hétérocycliques planes (souvent des ligands de duplexes et triplexes), à l'image des ellipticines que nous avons sélectionnées.

Cette expérience m'a amené à écrire un chapitre de livre sur le criblage de ligands de G4s (103). Si j'en conclusais qu'il ne fallait pas se reposer sur la chance seule, il semble que cela soit parfois suffisant. Je n'en dit pas plus pour ne pas divulguer la Section 5.3.1.1.

### 1.4.1.3 Caractérisation de ligands issus du design rationnel

#### 1.4.1.3.1 Déterminants structurels de la sélectivité de ligands macrocycliques pour les paires de bases d'ADN mésappariées

Les nucléotides mésappariés sont des paires de nucléotides au sein de l'ADN duplex qui ne respectent pas l'appariement classique des bases (Watson-Crick) (105). Les mésappariements peuvent se produire naturellement au cours de la réplication ou de la recombinaison homologue, ou être provoqués par des facteurs externes tels que des molécules mutagènes ou des radiations ionisantes (105, 106). L'efficacité de la réparation enzymatique des mésappariements dépend de la nature de la paire mal appariée, et des déficiences dans la réparation des mésappariements sont associées à de nombreux cancers (105). L'inhibition de la réparation est étudiée comme un moyen d'améliorer les effets des médicaments anticancéreux. Dans ce contexte, les molécules qui se lient sélectivement aux paires de bases mésappariées de l'ADN pourraient servir de base à de nouveaux agents chimiothérapeutiques.

Anton Granzhan avait préparé une série de 20 bis-intercalants macrocycliques lors de son séjour postdoctoral au laboratoire (il y est finalement revenu comme chargé de recherche CNRS, avant de rejoindre l'UMR 8601 de l'Université Paris Cité en 2025) (Figure 1.27A). J'ai été chargé après son départ d'étudier leurs interactions avec des doubles hélices d'ADN, contenant un résidu de thymine mésapparié (TX; Figure 1.27B) ou correctement apparié (TA), par FID (Figure 1.27C) (107). L'**EtBr** a été utilisé comme intercalant plutôt que le **TO** car il est plus adapté à l'ADN duplex (97).

Les dérivés 2,7-disubstitués du bisnaphtalène ont présenté la plus grande sélectivité pour les mésappariements TX par rapport à l'ADN apparié (Figure 10). Les différences spectaculaires d'affinité et sélectivité entre les diverses séries de macrocycles que nous avons observées soulignent la forte influence de la

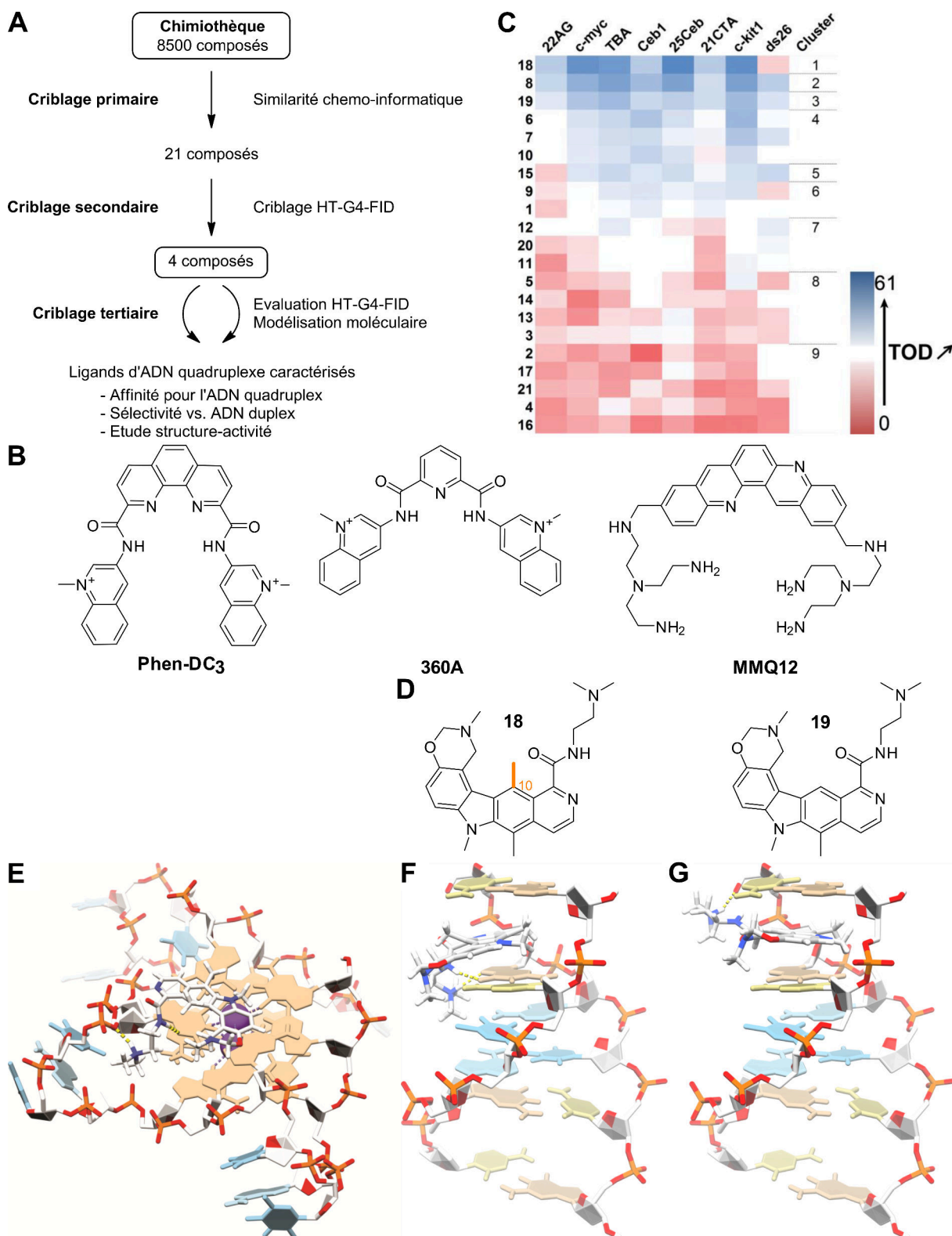


FIGURE 1.26. 8500 molécules de la chimiothèque de l'Institut Curie ont été criblées par une combinaison de méthodes chemo-informatique, HT-G4-FID, et *docking* (A). 21 ligands ont été identifiés par similarité chimique avec trois ligands de référence (B), puis testé par HT-G4-FID sur plusieurs G4 et un contrôle duplexe (C). Les ligands **18**, **8**, **19** et **6**, tous des ellipticines, ont été sélectionnés car il déplace plus le **T0** des G4s. En revanche, seul **18** est sélectif contre les duplexes et il diffère des autres ligands par l'absence d'un groupement méthyle en position 10 (D, le méthyle est l'unique différence entre **18** et **19**). Ces ligands peuvent interagir par  $\pi$ -stacking sur les quartets de guanines des G4s et liaisons hydrogène (E : exemple de **18**). Le méthyle en position 10 empêche **18** d'être plan, ce qui empêche sa bonne intercalation avec les duplex (F : gêne stérique significative), alors que **19** est un bon intercalant (G).

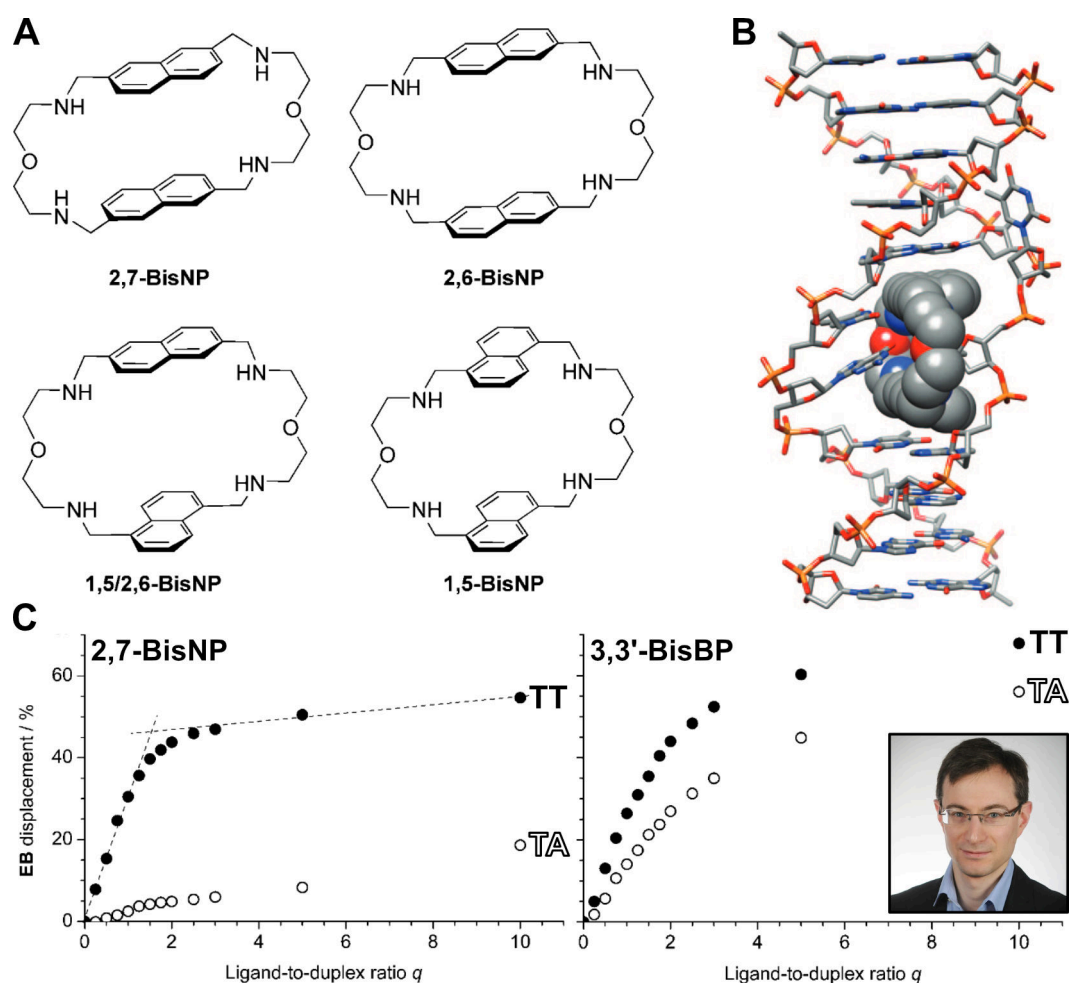


FIGURE 1.27. Le test FID dans sa version la plus classique pour étudier l'interaction entre des ligands macrocycliques et des duplex mésappariés ou non. A : Structure des ligands macrocycliques BisNP; B : Structure du **2,7-BisNP** docké dans un duplex 11-mer avec un mésappariement TT; C : Déplacement du bromure d'éthidium ( $0.4 \mu\text{M}$ ) de duplexes ( $0.1 \mu\text{M}$ ) correctement appariés (TA) ou mésappariés (TT) par **2,7-BisNP** et **3,3'-BisBP** dans un tampon cacodylate pH 6.0 ( $[\text{Na}^+] = 110 \text{ mM}$ ). Encart : Anton Granzhan.

taille et de la topologie du macrocycle sur la reconnaissance de ces mésappariements. Cette étude a permis d'établir un guide structure-propriété des ligands d'ADN contenant des mésappariements TX.

À titre personnel, il s'agit de ma toute première publication, mais j'ai surtout beaucoup appris de la rigueur scientifique d'Anton, qui reste un modèle pour moi.

#### 1.4.1.3.2 Inhibition de l'expression génique par ciblage des G4s dans les 5'-UTR d'ARN

Le Pr. Jörg Hartig (Université de Constance, Allemagne) avait montré que la formation de G4s d'ARN dans *Escherichia coli* inhibe l'initiation de la traduction car ils interfèrent avec l'accessibilité du site d'interaction des ribosomes (108). Des effets inhibiteurs similaires ont été démontrés pour les G4s d'ARN naturels basés sur le 5'-UTR (régions non traduites) dans la culture de cellules de mammifères pour l'expression de certains gènes (109–112). En outre, le niveau de suppression est proportionnel à la stabilité thermodynamique du G4 (113). Le Pr. Hartig a alors initiée une collaboration avec notre équipe pour évaluer la possibilité de réguler *in vivo* l'expression de gènes qui possèdent des G4 dans leur 5'-UTR grâce à des ligands stabilisant ces G4s (114).

J'ai utilisé le test HT-G4-FID pour évaluer l'affinité des ligands **360A**, **PhenDC3** et **PhenDC6** pour des G4s d'ARN (Figure 1.28). Lors de la caractérisation des différents complexes ARN/**TO** par titration fluorimétrique, préliminaire à l'application du test HT-G4-FID (comme dans la Figure 1.24), nous avons remarqué l'affinité très faible du **TO** pour ces G4, probablement plusieurs ordres de grandeur plus faible que celles des ligands bisquinolinium. Loin d'être un désavantage, cette large différence d'affinité m'a permis de déterminer les excellentes constantes d'affinité (1–10 nM) entre ces ligands et les G4s  $r(G_3U)_4$  et  $r(G_3U_2)_4$ . On peut en effet, dans ces conditions bien particulières, "ignorer" l'équilibre entre le G4, le **TO** et leurs complexes, et procéder comme s'il s'agissait d'une titration directe  $G4 + \text{ligand} \rightleftharpoons G4 \cdot \text{ligand}$ .

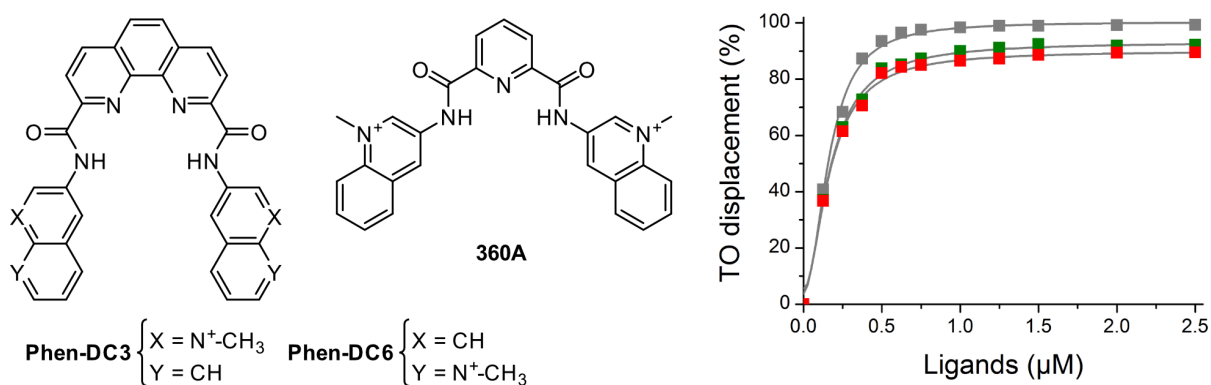


FIGURE 1.28. Courbes de déplacement du **TO** pour le G4  $r(G_3U_3)_4$  par les ligands **PhenDC3** (gris), **PhenDC6** (vert) et **360A** (rouge).

Il a ensuite été démontré que l'effet inhibiteur était plus prononcé sur les ARNm présentant une forte propension à la formation de G4, et que cet effet est amplifié quand le G4 est stabilisé par un ligand. Ces résultats suggèrent donc que ces ligands peuvent être utilisés pour réguler l'expression des gènes en ciblant les structures G4 de 5'-UTR d'ARN.

## 1.4.1.3.3 Instabilité génomique de minisatellites humains déclenchée par des ligands de G4

Les G4s sont des structures secondaires d'acides nucléiques pour lesquelles de nombreux rôles biologiques ont été proposés mais dont l'existence *in vivo* a longtemps été difficile à démontrer et le sujet de nombreux débats. Généralement, la première question à une présentation sur un sujet "G4" pouvait se résumer à : "Est-ce que ça existe vraiment vos bidules?". Il n'y a plus beaucoup de débat aujourd'hui, grâce notamment à l'utilisation d'anticorps anti-G4 (115–119), mais aussi d'études, comme celle ci-dessous, plus indirecte mais pas moins convaincante à mon humble avis.

Dans le cadre d'une collaboration avec l'équipe du Dr. Alain Nicolas (Institut Curie, Paris), j'ai ainsi utilisé le test HT-G4-FID pour démontrer la grande l'affinité des ligands **PhenDC3** et **PhenDC6** pour le G4 formé par le minisatellite humain CEB1, et leur sélectivité par rapport à l'ADN duplex (12). Cela a permis d'expliquer leur capacité à inhiber la déstructuration de ce G4 par l'hélicase Pif1 chez *Saccharomyces cerevisiae*. *In vivo*, les deux composés ont rapidement induit des réarrangements dépendants de la recombinaison de CEB1, mais n'ont pas affecté les séquences ne formant pas de G4. Pour ces expériences, j'ai encadré **Katie Mingo**, alors étudiante en médecine (MIT, Boston, USA) et désormais oto-rhino-laryngologue a la *Cleveland Clinic* (Minneapolis, MN, USA).<sup>4</sup>

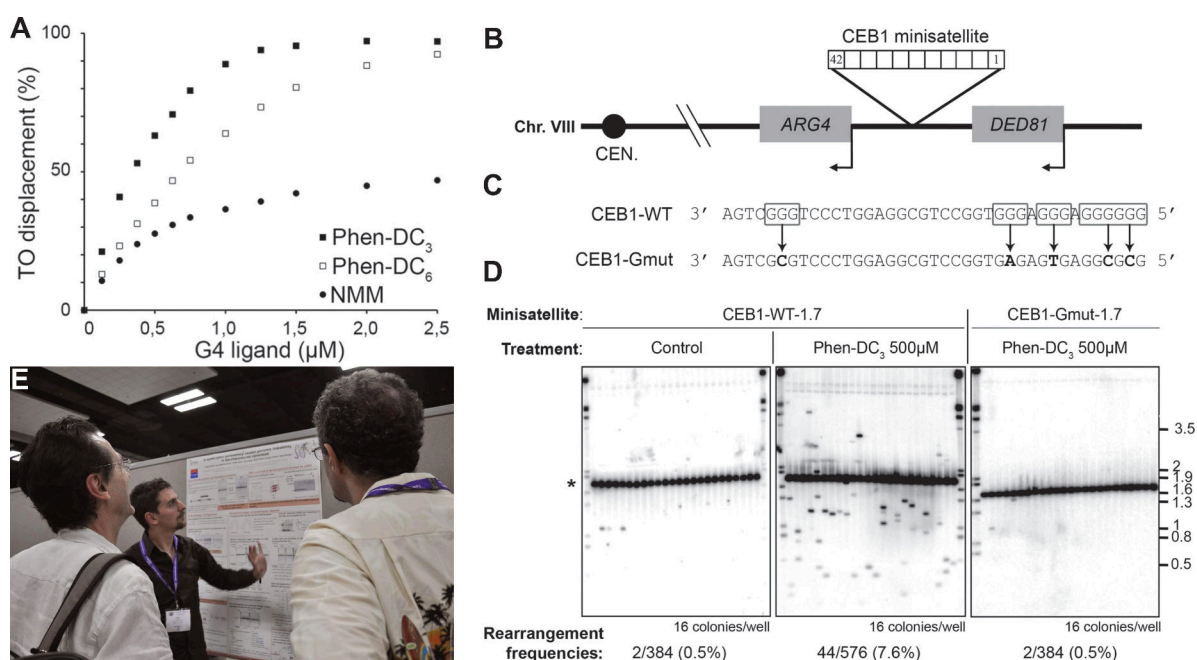


FIGURE 1.29. Les composés PhenDC déclenchent une instabilité de CEB1 dépendante des G-quadruplexes. A. Courbes de déplacement de TO montrant l'affinité de **PhenDC3** et **PhenDC6** pour le G4 CEB1. B. Représentation schématique de l'insertion du minisatellite CEB1 dans le chromosome VIII. C. Séquences des motifs CEB1-WT et CEB1-Gmut. Les mutations ponctuelles (en gras) introduites dans le motif CEB1-Gmut empêchent la formation de G4 *in vitro*. D. Analyse par *Southern blot* de cellules portant CEB1-WT-1.7 (souche AND1212-10D) ou CEB1-Gmut-1.7 (AND1227-5C) après traitement de contrôle (1% DMSO) ou 500 µM de **PhenDC3**. Le nombre de colonies analysées et les fréquences de réarrangement sont indiqués sous chaque gel. Les marqueurs de taille (kb) sont indiqués à droite. La position des allèles minisatellites parentaux (42 répétitions, 1,7 kb) est indiquée par un astérisque. E. L'auteur présentant ces résultats de génétique (!) à Jean-François Riou (gauche) et Jean-Louis Mergny (droite) lors du *Pacifichem 2010* (Photo : Florian Hamon).

4. Comme quoi les G4s mènent à tout.



#### 1.4.1.3.4 Effets de ligands de G4 halogénés sur la séquence terminale des télomères humains

Des modifications chimiques mineures de structures de ligands de G4 peuvent affecter à la fois leurs propriétés biophysiques et biologiques. En effet, des ligands de structures et d'affinités pour les G4 proches, tels que **360A**, **PhenDC3** et **PhenDC6**, ont montré des différences remarquables au niveau des changements transcriptionnels dans les gènes contenant des G4s, ce qui fut démontré dans certaines études ci-dessus et d'autres (102, 114, 120). D'autre part, l'introduction d'halogènes dans un composé biologiquement actif est souvent utilisée pour améliorer la perméabilité aux membranes (barrières cellulaire et hémato-encéphalique). En outre, les effets stériques et électroniques des atomes d'halogène améliorent parfois l'interaction avec des cibles biomoléculaires (121).

Il a été démontré que des ligands de G4 peuvent perturber l'interaction de composants télomériques essentiels, tels que POT1, et déclenchent un dysfonctionnement télomérique associé à un arrêt de croissance retardé dans les cellules tumorales (122–124). Dans ce contexte, Jean-François Riou a initié une étude sur les effets de ligands de G4s halogénés sur la séquence terminale de télomères de cellules de fibrosarcome HT1080.

J'ai assisté à la synthèse d'une série de dérivés halogénés du ligand **360A** (Figure 1.30A), pour laquelle j'ai encadré **Julien Botton** (Master, Université Pierre et Marie Curie, Paris VI) avec Florian Hamon, puis j'ai évalué leur affinité pour les G4s télomériques humains avec le test HT-G4-FID. Les dérivés bromé et iodé, mais pas le chloré, ont une affinité légèrement supérieure à celle de **360A**, tout en conservant une excellente sélectivité contre l'ADN duplex (Figure 1.30B). Malheureusement, nous n'avons pas pu rationaliser le comportement différent du dérivé chloré sur la base de l'électronégativité et de considérations stériques. Au final, les cellules de fibrosarcome HT1080 traitées jusqu'à 24 jours avec **360A-Br** ont présenté des variations mineures mais significatives de leur composition en nucléotides terminaux du brin C, avec également une déplétion partielle de POT1.

Comme pour de nombreux sujets explorés au cours de ma thèse, certains aspects de ce travail réapparaîtront bien plus tard. Récemment, j'ai ainsi participé à une étude portant sur l'influence de la substitution de ligands phényles de colonnes HPLC sur leurs interactions avec des oligonucléotides (Section 5.6.5) (125).

#### 1.4.1.3.5 Ligands fonctionnalisés pour la sélection d'aptamères anti-G4

Une collaboration avec les équipes de Jean-Louis Mergny et Jean-Jacques Toulmé (IECB, Bordeaux) avait pour but de sélectionner par SELEX des aptamères d'ADN qui se lient à un ligand de G4 spécifique. Pour cela, il nous fallait fournir des ligands fonctionnalisés avec une biotine de façon à pouvoir les immobiliser lors des expériences de SELEX. Nous avons opté pour deux ligands facilement dérivatisables, **360A** (PDC) et **Cu-ttpty**, auxquels nous avons couplé une biotine avec une chaîne suffisamment longue pour permettre leur immobilisation sans compromettre leur affinité pour les G4s.

Florian Hamon est reparti d'un intermédiaire de synthèse de **Cl-360A** (celui précédant la méthylation finale des quinolines) pour y coupler la chaîne biotinylée. Le ligand final **PDB-biotin** a ensuite été obtenu après méthylation des quinolines. De mon côté, j'ai couplé la chaîne biotinylée sur un dérivé bromé de **ttpty**, puis ai complexé du cuivre(II) pour obtenir **Cu-ttpty-biotin**.

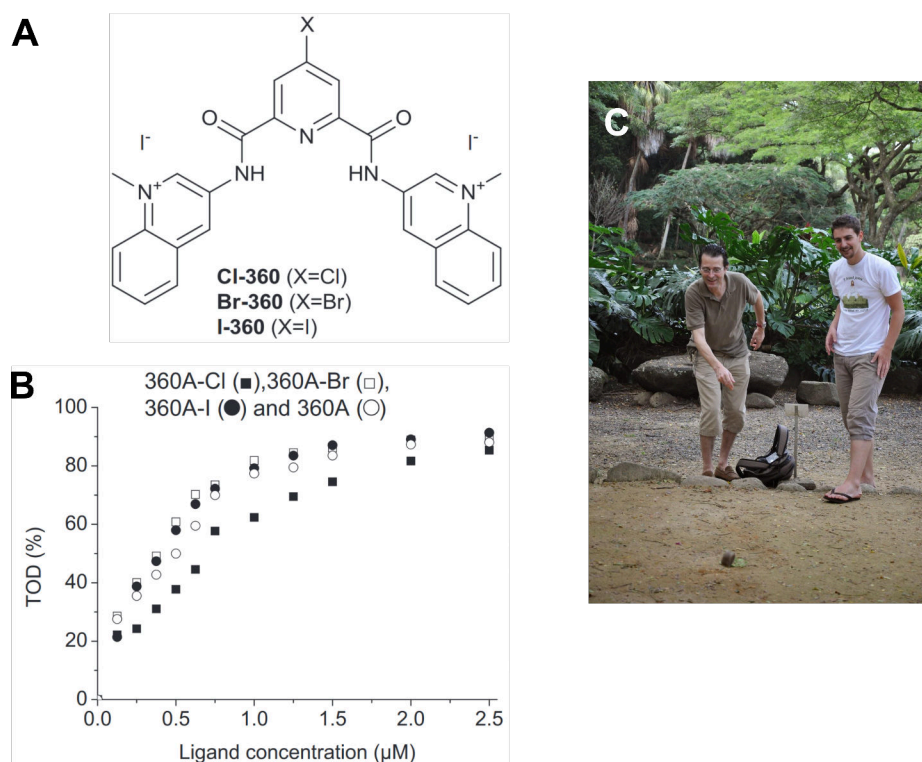


FIGURE 1.30. Les dérivés (A) bromé et iodé, mais pas chloré, de **360A** ont une affinité plus grande pour le G4 de la séquence télomérique humaine, visible par le déplacement plus efficace du **TO** (B). C. Partie d'*Ulu maika*, un ancien sport hawaïen, avec Jean-François Riou (Moanalua Gardens, Honolulu, USA; Photo : Florian Hamon)

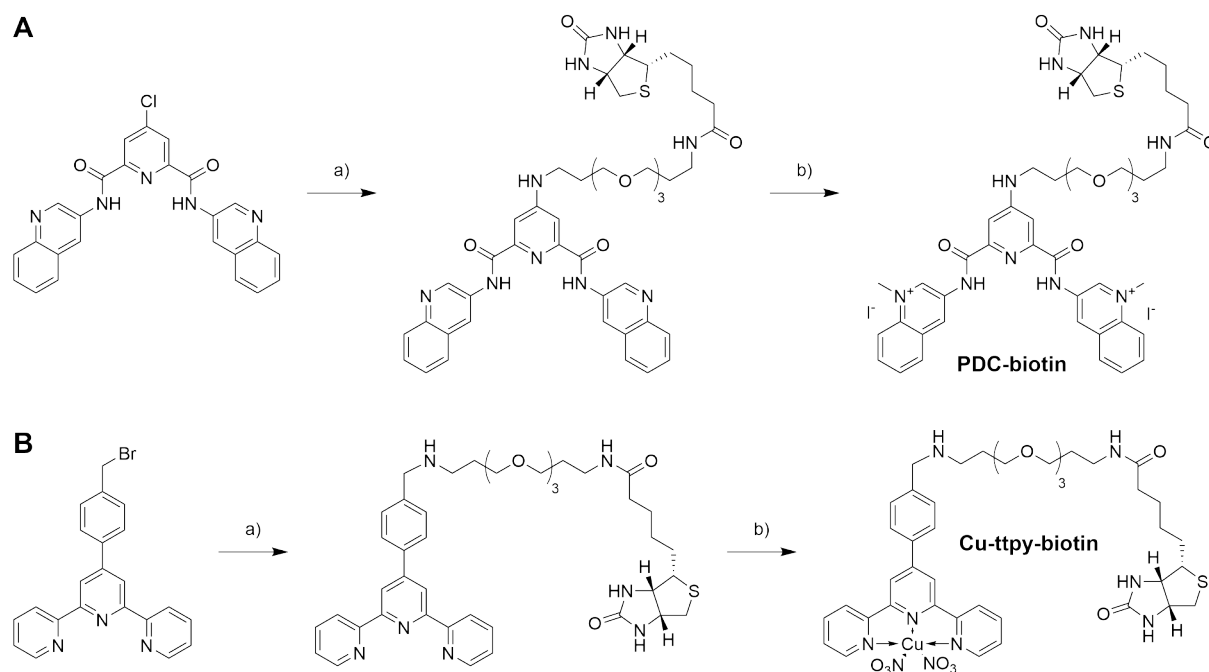


FIGURE 1.31. Synthèse de **PDC-biotin** (A) et **Cu-ttpp-biotin** (B). A. a)  $\text{NEt}_3$ , DMSO,  $90^\circ\text{C}$ , 16h, 63%; b)  $\text{CH}_3\text{I}$ , DMF,  $40^\circ\text{C}$ , 16h, 87%. B. a)  $\text{NEt}_3$ , DMF,  $60^\circ\text{C}$ , 4h, 19%; b)  $\text{CuNO}_3$ ,  $\text{CH}_3\text{CN}$ ,  $\text{CH}_2\text{Cl}_2$ ,  $4^\circ\text{C}$ , 24h, 50%.

Les expériences HT-G4-FID ont confirmé que le ligand **PDC-biotin** conserve une affinité élevée pour les G4s, comparé à **360A**, ainsi qu'une sélectivité remarquable par rapport à l'ADN duplex (Figure 1.32). À

l'inverse, **Cu-ttpty-biotin** a une affinité nettement plus faible pour les G4s et, en outre, est moins sélectif que **Cu-ttpty**, ce qui en fait un moins bon candidat pour les expériences de SELEX.

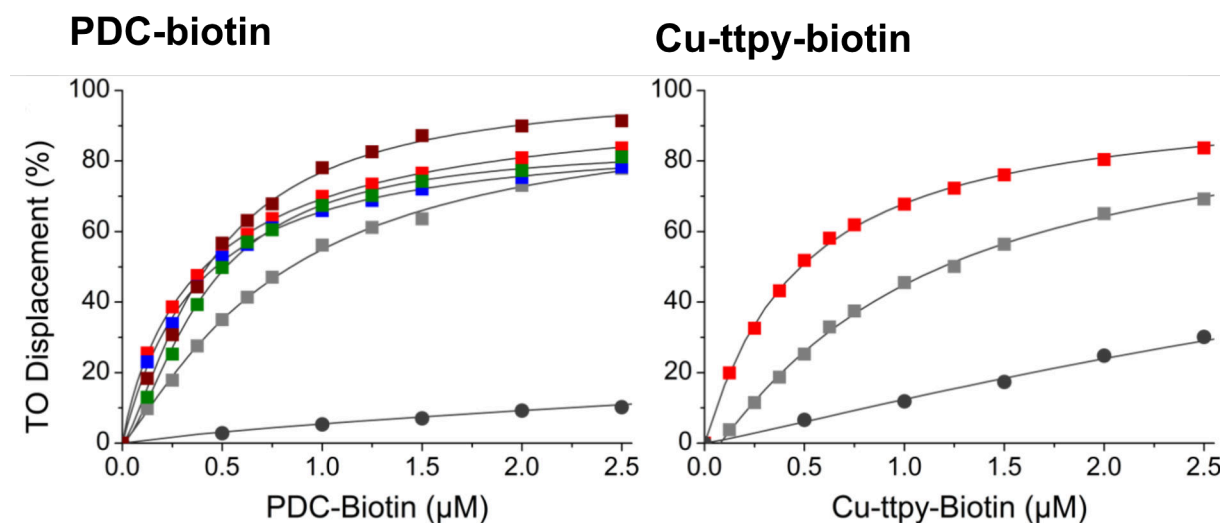


FIGURE 1.32. Courbes de déplacement du TO par PDC-biotin et Cu-ttpty-biotin avec la séquence télomérique humaine 22AG en présence de Na<sup>+</sup> (carrés gris) ou K<sup>+</sup> (carrés rouge), TBA (carrés bleu), c-kit1 (carrés vert), c-kit2 (carrés marron) et le contrôle duplex ds26 (cercles gris).

Finalement, peu des aptamères sélectionnés (5 sur 80) étaient des G4s, ce qui a montré que des sites d'interactions inattendus peuvent être observés lors d'analyse de génomes ou transcriptomes entiers. La plupart des méthodes de caractérisation de l'affinité et sélectivité de ligands de G4 incluent peu de contrôles; souvent seulement des doubles hélices. La sélectivité réelle des ligands de G4 est donc sans doute plus faible qu'on ne le pense, ce que démontre bien cette étude *in vitro*. Par extension, on peut penser que cela est pire dans les milieux *in vivo*.

#### 1.4.2 Sélectivité de sondes fluorescentes : le test G4-POSCA

Comme nous l'avons vu ci-avant, je me suis beaucoup intéressé aux ligands fluorescents de G4 (11). Cela m'a amené à concevoir une méthode originale, *G4-POSCA*, qui ne s'intéresse pas à proprement parlé à la sélectivité de petites molécules pour les G4s, mais de leur *sélectivité optique* (OS), c'est-à-dire de leur capacité à révéler la présence de G4s par fluorescence en présence d'autres cibles 'compétitrices' (91). Un des grands avantages de cette méthode est d'ailleurs de pouvoir utiliser n'importe quel compétiteur, sans modification nécessaire, ici : différents excès de duplex, d'ADN total, d'ARN total, de protéine (la HSA) et même un autre G4.

La cible G4 biotinylée est d'abord mélangée ou non à ces compétiteurs (Figure 1.33A). Après incubation en présence du ligand fluorescent (Figure 1.33B), le G4 est immobilisé sur des billes magnétiques couvertes de streptavidine et le surnageant est récupéré (Figure 1.33C). On compare alors l'intensité de fluorescence du ligand en présence de G4 seul, avec celle venant d'une interaction avec le compétiteur (mesure de la sélectivité optique : surnageant du tube bleu dans la Figure 14) et lorsque le ligand est non lié (mesure du rapport signal-sur-bruit : surnageant du tube rouge). J'ai encadré une étudiante d'IUT, **Margaux Legrand**, lors de cette étude.



## 1.4 Axe 2. Méthodes analytiques pour l'analyse d'interactions ligands-acides nucléiques

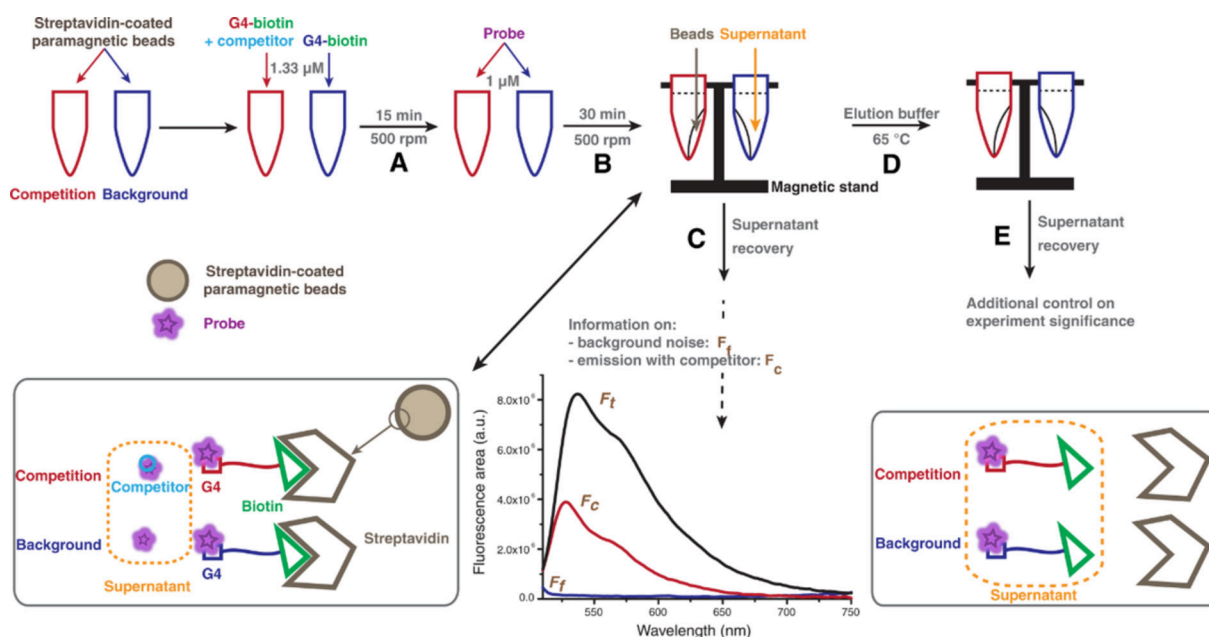


FIGURE 1.33. Principe du test G4-POSCA. A. Incubation du G4 cible (biotinylé) dans une solution contenant des billes magnétiques recouvertes de streptavidine, en présence ou absence d'un compétiteur, B. ajout du ligand fluorescent et incubation, C. récupération des surnageants ne contenant que les oligonucléotides non biotinylés puis mesure de l'intensité de fluorescence en présence ou absence de compétiteur. Les étapes D et E permettent de cliver la cible et vérifier le niveau de fluorescence (contrôle).

Pour le développement de cette méthode, j'ai utilisé plusieurs molécules fluorescentes avec des affinités, sélectivités et fluorescences diverses, que le test G4-POSCA a pu confirmer (Figure 1.34) :

- **PDC-BODIPY** et **PDC-L-TO** sont deux dérivés de **360A**, qui possède une grande affinité et une grande sélectivité pour les G4 (70, 101, 126). Nous avons vu ci-dessus que la fonctionnalisation de **360A** sur sa pyridine ne perturbe pas de manière significative son affinité pour les G4s (Section 1.4.1.3.4), ce qui est confirmé par l'excellente OS de **PDC-BODIPY**. En revanche, le **TO** peut se lier avec les G4s et les duplexes - ce qui diminue sensiblement la sélectivité de **PDC-L-TO**, mais ne fluoresce que lorsqu'il est lié ce qui lui donne un meilleur rapport signal-sur-bruit de fond (SBR).
- Le dérivé de carbazole **Cbz-2Py** est une sonde fluorescente à deux photons, qui se lie préférentiellement au petit sillon de l'ADN duplex riche en AT, mais qui peut également se lier à des structures quadruplex avec un  $K_d$  de l'ordre du micromolaire. Cette faible sélectivité est visible ici (faible OS), tout comme les excellentes propriétés de fluorescence de cette molécule (meilleur SBR).
- Le **TMPyP4**, une molécule très utilisée comme ligand de G4, qui a certes une bonne affinité pour ces derniers mais une sélectivité faible (voire inexistante) entre G4 et duplex. Cela est évident par FRET-melting, HT-G4-FID et dialyse compétitive (101, 127, 128), et se confirme ici.<sup>5</sup>
- Le **TO**, qui n'est pas sélectif mais possède un bon SBR.

Si les deux derniers ligands sont commerciaux, les trois premiers ont été préparés au laboratoire. **PDC-L-TO** a été synthétisé avant mon arrivée par Peng Yang (129), **Cbz-2Py** par Blaise Dumat (130), un collègue doctorant (maintenant CR CNRS dans l'UMR7203 de l'ENS Paris), et enfin **PDC-BODIPY** par Reuben Ovadia (Master, Université de Montpellier; désormais ingénieur brevets chez Servier) co-encadré avec Florian

5. Si le lecteur doit retirer une seule information de ce manuscrit, c'est que **TMPyP4** n'est *pas* un bon ligand de G4.

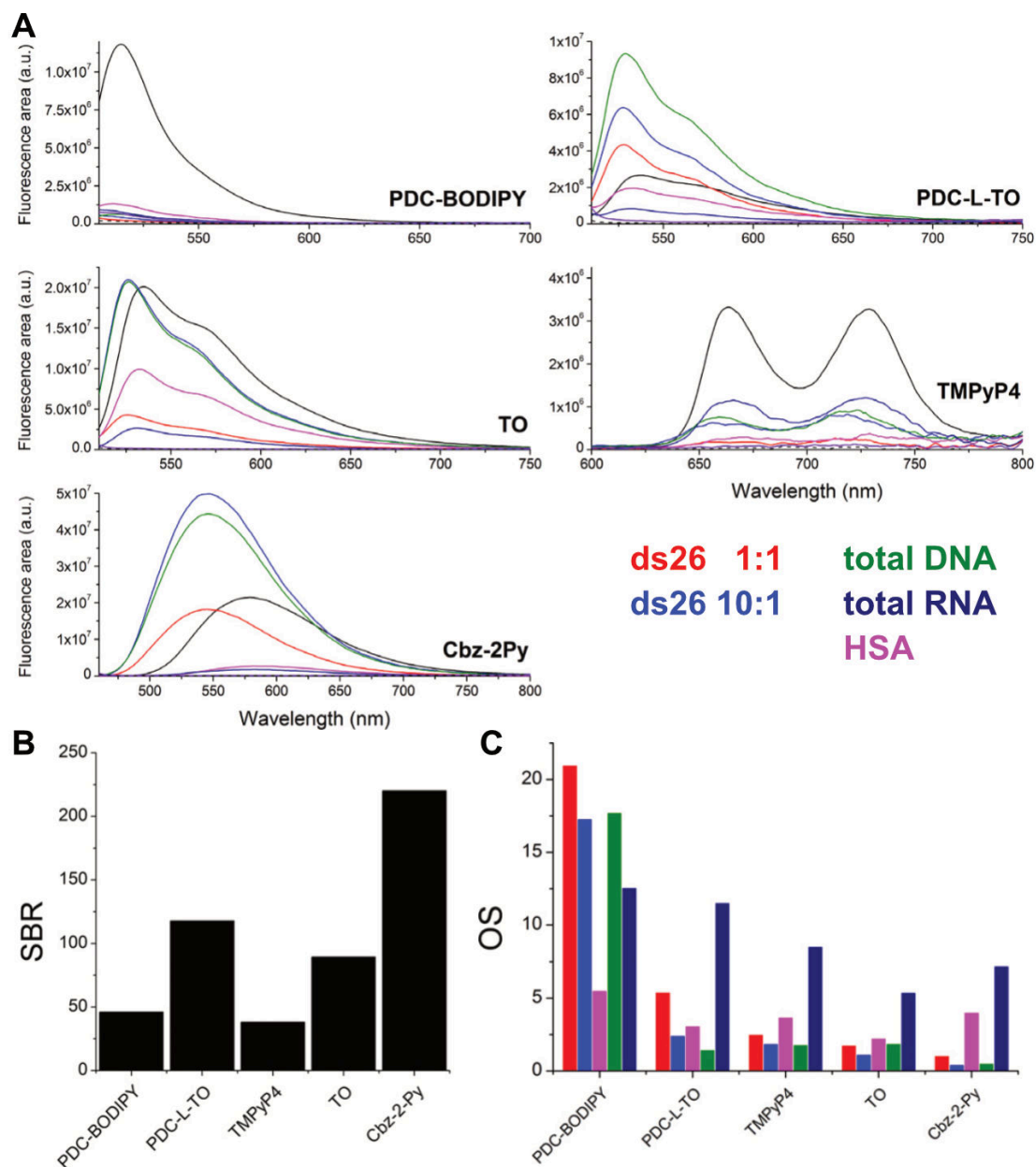


FIGURE 1.34. Résultat des tests de développement du G4-POSCA. A. Fluorescences des ligands en présence de la cible (noir, séquence télomérique humaine) ou des surnageants contenant un compétiteur (Fc; lignes colorés). B. Ratio signal/bruit de fond (SBR). C : Sélectivités optiques des différentes sondes.

Hamon (Figure 1.35). À ma connaissance, la synthèse de ce dernier n'a jamais été publiée, à l'inverse de PhenDC3-bodipy (131).

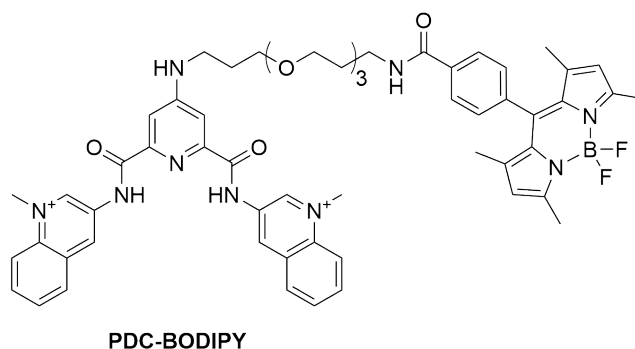


FIGURE 1.35. Structure de PDC-BODIPY

## 1.5 Conclusion

Ce travail de thèse a combiné synthèse de ligands de G4 par chimie organométallique et organique, au développement d'outils d'analyse permettant de caractériser aux mieux ces ligands et leurs complexes, mais aussi d'en découvrir de nouveaux. Nous avons démontré que nos complexes de  $\text{Pd}^{2+}$  et  $\text{Pt}^{2+}$  interagissent avec les structures G4 par des mécanismes covalents et non covalents, et que l'oligohétéroaryle acyclique **TOxaPy** présente une sélectivité topologique inédite en ciblant préférentiellement les sillons plutôt que les tétrades. Parallèlement, la mise au point du test HT-G4-FID a permis le criblage et la caractérisation rapides d'un grand nombre de ligands potentiels. Ce chapitre illustre ainsi comment la conception rationnelle de nouvelles molécules, associée à des méthodes analytiques robustes, peuvent élargir le spectre des interactions ligand-G4. Comme un clin d'oeil du destin, ce qui sera sans doute mon dernier article publié avant de soutenir cette HDR revisitera ce thème, avec un brin de chance en prime.

D'un point de vue personnel, ces années passées à l'Institut Curie ont constitué une période particulièrement enrichissante, tant sur le plan scientifique qu'humain, et ont posé les fondations de la suite de mon parcours. J'y ai trouvé ma voie, pas en chimie organique, comme je l'imaginais au départ, mais avec une motivation renouvelée pour l'exploration des interfaces entre la chimie et les systèmes biologiques.



## Chapitre 2

### Postdoctorat à l'Université de Colombie Britannique

#### 2.1 Contexte

J'ai rejoint l'équipe du Pr. David Perrin, au département de chimie de l'Université de Colombie Britannique (UBC) en septembre 2012 (Figure 2.1). Plusieurs thèmes de recherche y étaient développés, notamment la sélection d'aptamères d'ADN modifié avec des chaînes latérales d'acide aminés. L'objectif était de produire des DNAzymes pour cliver des mARN viraux, des sondes de métaux toxiques ou des aptamères se liant à la surface de cellules. D'autres projets incluaient la synthèse de molécules pour l'imagerie PET et la synthèse peptidique à des fins thérapeutiques. Plus trivialement, les étudiants et postdoctorants de l'équipe faisaient de la synthèse organique ou du SELEX. Et je n'ai donc fait ni l'un ni l'autre.

La majeure partie de l'information génétique se trouve à l'intérieur de la double hélice d'ADN, protégée par des interactions auto-complémentaires. Dans les années 80, Hélène et Dervan ont proposé l'hypothèse thérapeutique dite « antigène », qui implique le ciblage de séquences spécifiques d'ADN par des oligonucléotides complémentaires (132, 133). La reconnaissance de l'ADN pour la thérapie génique, avec des oligonucléotides présentant par nature une sélectivité de séquence, a depuis été un axe majeur de recherche, notamment pour réguler les processus biologiques comme l'activation de la transcription de l'ADN (134). Le déplacement de brin avec, par exemple, des PNA implique une séparation de la double hélice, remplaçant un set de paires de bases par un autre, ce qui n'est pas particulièrement favorable énergétiquement.

Les nucléosides Janus, proposés par Lehn, offre une approche de reconnaissance de l'ADN qui permet un gain énergétique en établissant simultanément des liaisons hydrogène avec les deux bases d'une paire, plutôt qu'en se substituant à un brin (135). La formation de nouvelles liaisons hydrogène accroît la stabilité du complexe de l'ordre de 2 kcal.mol<sup>-1</sup> par paire de bases. Les premiers nucléotides Janus étaient uniquement capables de reconnaître des paires TT ou TC (136, 137). Le véritable intérêt en termes de reconnaissance de l'ADN réside dans la synthèse de nucléosides Janus qui présentent des donneurs de protons (D) et des accepteurs (A) disposés de manière à intercepter toutes les liaisons H que l'on trouve dans une paire de bases canonique.

Lorsque je suis arrivé dans l'équipe du Prof. Perrin, deux étudiants, Wenbo Liu et Abid Hasan, avaient débuté un travail de synthèse d'analogues de nucléotides Janus, donc disposant de deux faces distinctes possédant un réseau de sites donneurs et accepteurs de liaisons hydrogènes complémentaires de différentes bases. Dans ce contexte, j'ai apporté ma contribution pour la purification et la caractérisation biophysique et computationnelle d'oligonucléotides incorporant un nucléotide Janus AT, ce qui a permis de non seulement valider mais aussi optimiser le design de ces nucléotides modifiés (9, 10).

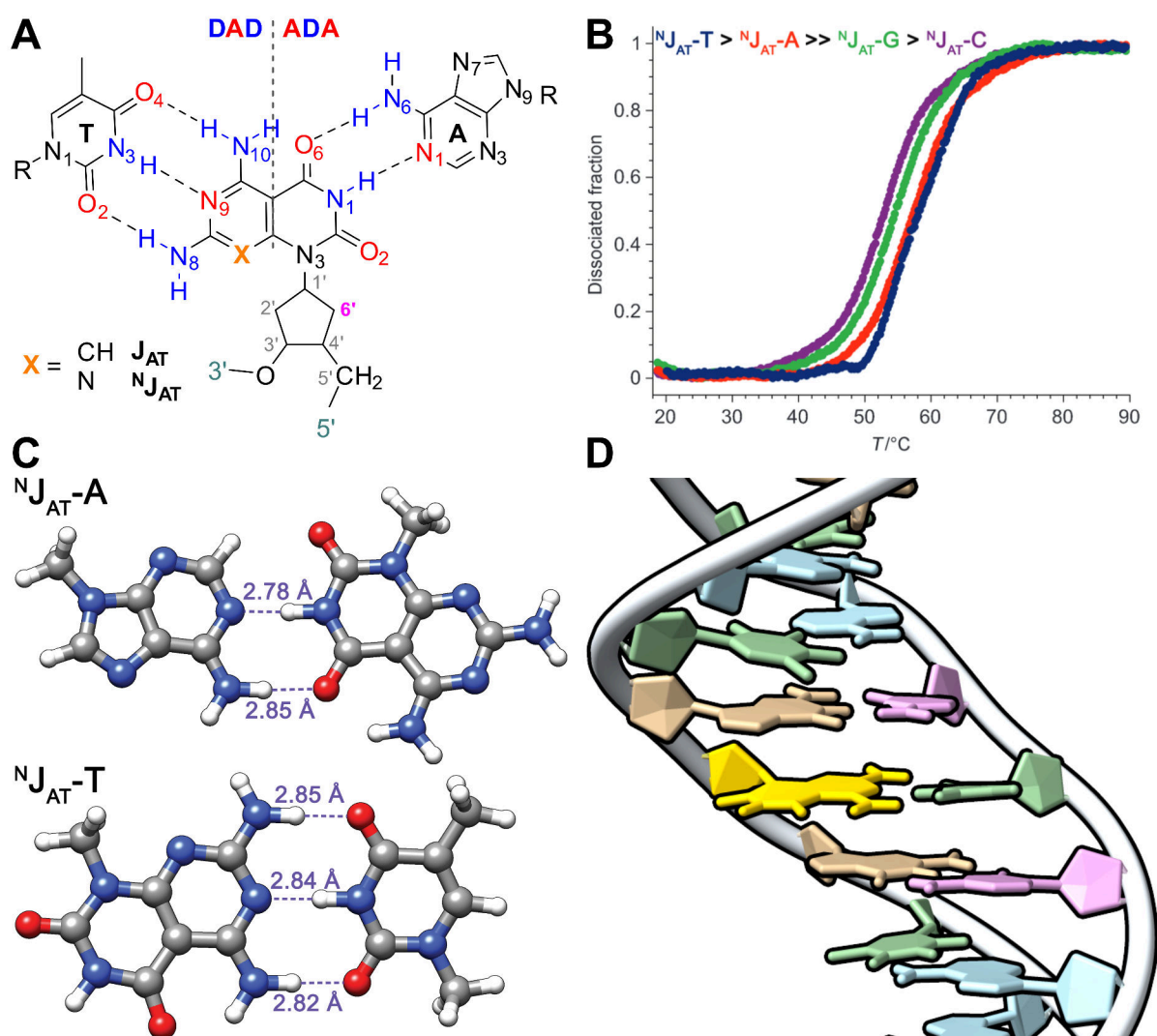


FIGURE 2.1. Bâtiment de chimie de l'Université de Colombie Britannique et, de gauche à droite, David M. Perrin, Abid Hasan et Wenbo Liu. Ce dernier ne partage pas ma passion pour le dichroïsme circulaire.



## 2.2 Synthèse, caractérisation et optimisation d'oligonucléotides incorporants des nucléotides bifaciaux

Nous avons donc développé une base Janus-AT «  $J_{AT}$  », pouvant se lier simultanément à une adénine et une thymine (Figure 2.2A,  $X = CH$ ) (9). Le *design* du nucléotide est rationnel, avec une face ADA pouvant théoriquement former trois liaisons hydrogène avec les thymines, et une face DAD pour hybridation avec les adénines. Pour éviter le clivage de la liaison *N*-glycosidique, un analogue carbocyclique du désoxyribose a été sélectionné. En outre, il a été montré que ce carbocycle améliore la stabilité des duplex et des triplex dans d'autres contextes (138). Un oligonucléotide incorporant un  $J_{AT}$  a été synthétisé à partir du phosphoramidate de  $J_{AT}$  préparé par Wenbo et Abid. J'ai ensuite purifié cet oligonucléotide sur gel et vérifié l'identité par spectrométrie de masse MALDI-TOF.





J'ai démontré avec une première étude biophysique que ce nucléotide modifié permet bien l'hybridation avec une adénine ou une thymine dans le contexte d'un double brin d'ADN. Ainsi, en présence de l'oligonucléotide complémentaire (avec une adénine face au  $J_{AT}$ ), le duplex a bien été détecté en phase gazeuse par MALDI TOF. Le dichroïsme circulaire a confirmé la formation de duplexes en solution. Les expériences d'UV-melting ont démontrés que la paire  $J_{AT}$ -T, mais pas la paire  $J_{AT}$ -A, est plus stabilisante qu'une paire AT. En outre, les paires  $J_{AT}$ -A ont générées des courbes dont les transitions particulièrement larges, avec une  $T_m$  inférieure aux paires AT mais supérieure à un mésappariement. J'ai également pris soin de montrer que  $J_{AT}$  est spécifique de A et T, en mesurant les  $T_m$  avec C et G, ces deux nucléotides entraînant une baisse de la température de dénaturation de l'ordre de 10°C.

Pour tenter d'expliquer ces résultats prometteurs mais pas optimaux, j'ai élargi mes horizons en me tournant vers la chimie computationnelle. Ainsi, par calculs DFT (BP86/TZ2P) sur des paires de nucléotides isolés, j'ai obtenu des résultats cohérents avec la formation de paires de bases, là aussi suggérant que  $J_{AT}$ -T est plus stable que  $J_{AT}$ -A. En revanche, la répulsion importante entre groupes carbonyles ne permet pas, comme attendu, l'hybridation à G ou C.

J'ai ensuite réalisé des simulations de dynamiques moléculaires de doubles hélices incorporant un  $J_{AT}$ , dans un champ de force AMBER avec solvation implicite. La double hélice est restée stable pendant la simulation de deux nanosecondes.<sup>1</sup> Lorsqu'il est associé à A,  $J_{AT}$  est cependant contraint dans une conformation qui ne maximise pas le  $\pi$ -stacking, ce que nous avons pensé pouvoir être à l'origine de la stabilité inférieure de ce duplex.

Nous avons finalement fait l'hypothèse que le problème était plus... basique. Plus spécifiquement, nous avons suggéré que le comportement anormal de la paire  $J_{AT}$ -A pouvait être dû à la protonation partielle à pH physiologique de l'azote N9, pour lequel j'avais calculé  $pK_a \approx 7.5$ . Cela conduirait à un mélange en équilibre dynamique la base neutre qui peut s'apparier et sa forme protonée pour laquelle l'appariement est peu probable.

Nous avons donc synthétisé un nucléotide de seconde génération «  $NJ_{AT}$  », dans lequel l'inclusion d'un azote endocyclique rend l'hétérocycle considérablement moins basique, minimisant ainsi le potentiel de protonation en une espèce cationique incapable de s'apparier (Figure 2.2A, Figure 2.3) (10). Les expériences de *melting* ont cette fois montré que  $NJ_{AT}$  produit des transitions nettes, et discrimine efficacement A/T par rapport à G/C (Figure 2.2B). Les expériences de dichroïsme circulaire et dynamiques moléculaires ont confirmé qu'il y a peu ou pas de modification de la structure secondaire pour les duplex formés (Figure 2.2C, D).  $NJ_{AT}$  produit donc des paires de bases stables avec A et T et est donc un candidat prometteur pour développer des stratégies de ciblage de séquences d'ADN.

Ces travaux postdoctoraux ont été possibles grâce à l'obtention d'une **Bourse Banting**, délivrée par le Conseil de recherches en sciences naturelles et en génie du Canada. Pendant ce séjour, j'ai co-encadré **Wenbo Liu** (Master) et **Abid Hasan** (Master), tous les deux participant à la synthèse de nucléotides Janus.

1. Ces simulations peuvent paraître très courtes au lecteur, en particulier au regard de mes travaux actuels, mais il s'agit d'un petit miracle de la science si l'on considère que ces calculs ont été réalisés sur mon ordinateur portable - qui s'est mû en chauffage d'appoint pendant plusieurs semaines.



FIGURE 2.3. Couverture de Chemistry - A European Journal (volume 20, issue 6) (10).

## 2.3 Conclusion

L'objectif principal de ce séjour postdoctoral était d'explorer les nucléotides Janus, capables de reconnaître simultanément deux bases complémentaires. Suite à la synthèse du dérivé  $J_{AT}$ , j'ai démontré sa capacité à former spécifiquement des paires de bases avec A et T, puis rationalisé le design d'une base optimisée  $NJ_{AT}$ , grâce à une combinaison de biophysique et chimie computationnelle.

D'un point de vue personnel, cette expérience m'a conforté dans l'orientation plus analytique de ma carrière et m'a beaucoup appris, scientifiquement et humainement. J'ai publié mes premiers travaux de chimie computationnelle (Figure 2.3), et élargi mes connaissances en biochimie des oligonucléotides et en SELEX, même si cela ne s'est pas traduit par des publications à court terme. En particulier, j'ai participé à des travaux de sélection d'aptamères possédants des bases modifiées, capables de cliver de l'ADN et de l'ARN en conditions physiologiques. Ces travaux m'ont finalement mené à l'analyse de la cinétique d'incorporation de nucléotides modifiés comme substrats de polymérases obtenues par évolution dirigée (139).

## Chapitre 3

### Postdoctorat à l'Institut Européen de Chimie et Biologie

A mon retour du Canada, j'ai travaillé comme post-doctorant INSERM à l'Institut Européen de Chimie et de Biologie dans les équipes de J.-L. Mergny et V. Gabelica (Figure 3.1). Je me suis à nouveau intéressé aux structures non canoniques de l'ADN, en développant et utilisant des techniques de chromatographie, spectrométrie de masse native et spectroscopie, à l'interface entre ces deux équipes. Je présente ci-dessous les deux principaux projets que j'ai menés.



FIGURE 3.1. L'IECB et sa fameuse vague, Jean-Louis Mergny (Crédit photo : photo Laurent Theillet, pour Sud-Ouest) et Valérie Gabelica (Crédit photo : Alexandre Darmon/Art in research pour la Fondation Bettencourt Schueller)

#### 3.1 Chromatographie d'exclusion stérique pour l'étude du polymorphisme des acides nucléiques

De nombreux laboratoires académiques s'intéressent aux structures non-canoniques des acides nucléiques et à leurs rôles dans le vivant, comme en témoignent les précédentes sections. Dans le même



temps, l'industrie pharmaceutique développe de plus en plus d'acides nucléiques thérapeutiques et a donc des besoins grandissants de caractérisation de leurs structures secondaires (140–143).

À mon arrivée dans son équipe, Jean-Louis Mergny m'a proposé de développer une méthode de chromatographie d'exclusion stérique (SEC-HPLC) pour discriminer des oligonucléotides en fonction de leur structure secondaire (ssDNA, dsDNA, i-motif, triplex, quadruplex) ainsi que de quantifier leur potentiel autoassemblage en multimères (144).

Je me suis appuyé sur une calibration externe extensive, impliquant plus de 110 oligonucléotides formant de nombreuses structures secondaires (duplexes antiparallèles, duplexes parallèles, duplexes A-tract courbés, duplexes mésappariés, triplexes, G4s, i-motifs, tiges-boucles d'ARN). En particulier, les oligonucléotides non structurés ou structurés en duplexes donnent des courbes de calibration possédant d'excellentes linéarités, ce qui permet de caractériser leur longueur et stœchiométrie avec une très grande confiance (Figure 3.2).

Comme je l'ai déjà évoqué dans le Chapitre 1, les G4s sont particulièrement polymorphes. La plus faible linéarité de la calibration n'a donc pas été une très grande surprise (Figure 3.3A). Cela dit, nous avons montré que la stœchiométrie des G4s pouvait être déterminée avec un haut niveau de confiance, tout comme leur structuration (Figure 3.3B-D).

Par ailleurs, et c'est sans doute l'aspect que je trouve le plus intéressant dans cette technique, la SEC-HPLC permet de rapidement estimer l'influence sur la structure d'acides nucléiques des conditions expérimentales (concentrations en oligonucléotide et sels, nature du tampon et des sels, température, pH), sur une grande gamme dynamique (Figure 3.3B, E, F). Cela est plus difficile, voire impossible, par d'autres méthodes de par l'incompatibilité de certaines conditions expérimentales : tampons et sels non-volatiles en spectrométrie de masse, trop faibles concentrations en oligonucléotides en RMN, coexistence de plusieurs conformères en cristallographie. La SEC-HPLC est d'ailleurs, je pense, une méthode de choix pour des études multi-techniques pour faire le lien entre des résultats obtenus dans des conditions très différentes ; par exemple des concentrations élevées en oligonucléotides et sels en RMN contre des concentrations plus faibles en spectrométrie de masse (Figure 3.3E, F).

J'ai appliqué cette méthode d'analyse dans d'autres projets, comme dans l'étude de l'influence des riboses sur la stabilité des G4s, où nous avons démontré que les résidus d'ARN ne stabilisent pas forcément ces structures, contrairement à ce qui était supposé jusque-là (151).

Une dernière chose me paraît intéressante à propos de cette étude. L'éditeur en charge de cette publication était initialement réticent à l'envoyer en relecture par des pairs et il a fallu faire preuve de conviction pour forcer le destin. Malgré cela, cette publication est fréquemment citée et, surtout, j'ai été plusieurs fois contacté par des industriels pour sa mise en place dans l'industrie pharmaceutique.

### **3.2 Influence des cations sur les G4s : découverte et caractérisation de doubles commutateurs**

Nous l'avons vu, les structures G4 présentent un polymorphisme structurel important. Une séquence donnée peut adopter plusieurs topologies en fonction de la nature et concentrations des cations, du pH,

### 3.2 Influence des cations sur les G4s : découverte et caractérisation de doubles commutateurs

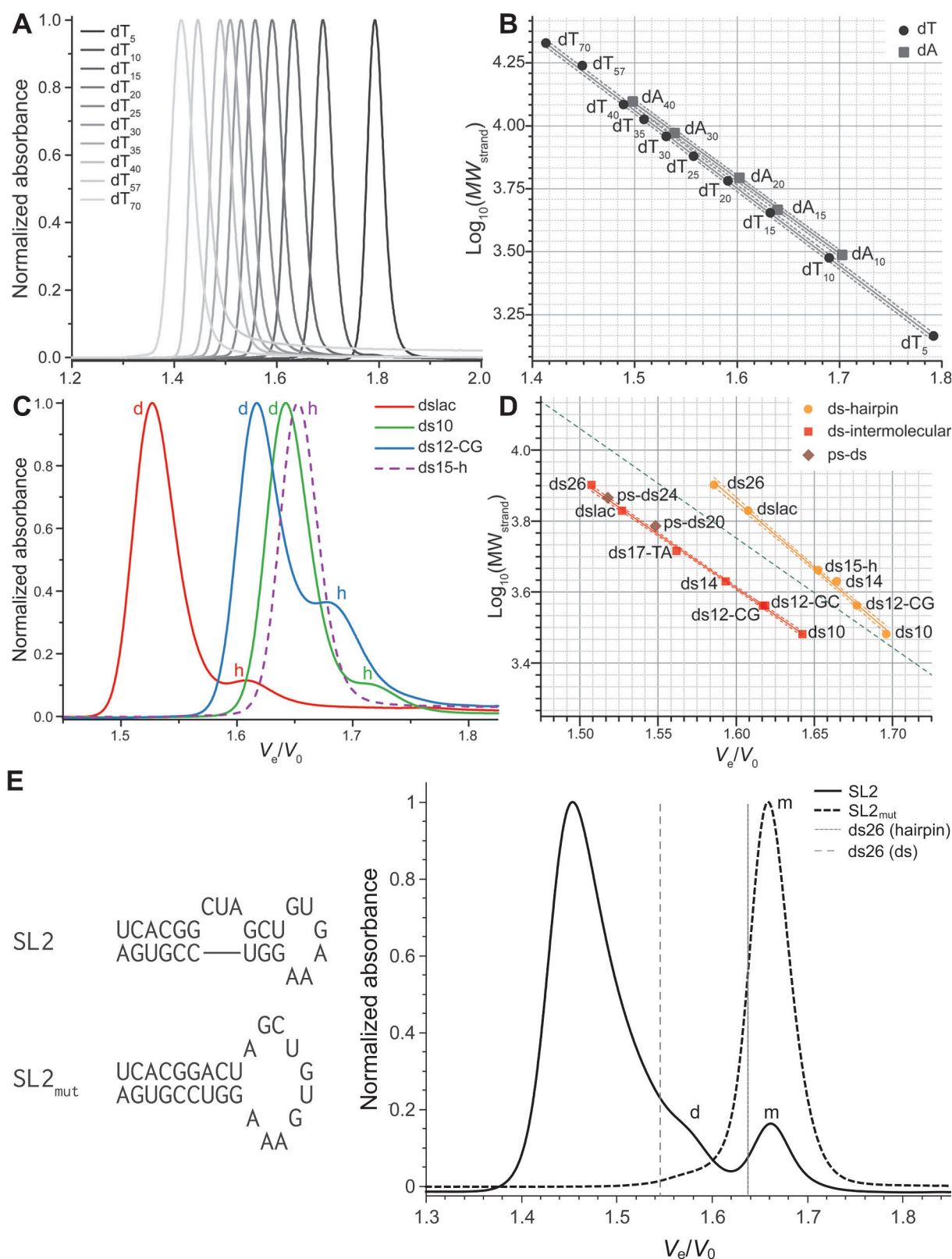


FIGURE 3.2. Chromatogrammes normalisés d'oligonucléotides polypyrimidine d(T<sub>n</sub>) non structurés (A) et auto-complémentaires (B; d : dimère, h : *hairpin*) Tracés du logarithme décimal du poids moléculaire en fonction du volume d'élution relatif pour les oligonucléotides polythymidilate (d(T<sub>n</sub>), cercles) et polyadénylate (d(A<sub>n</sub>), carrés) non structurés (C) et de duplexes intermoléculaire, intramoléculaires et parallèles (ps) (D). Application à une tige-boucle d'ARN (E) : SL2 coexiste comme un mélange de structures monomériques et dimériques, en échange rapide, tandis que SL2<sub>mut</sub>, qui contient une adénine supplémentaire, forme une tige-boucle parfaite et est utilisé comme contrôle monomérique.

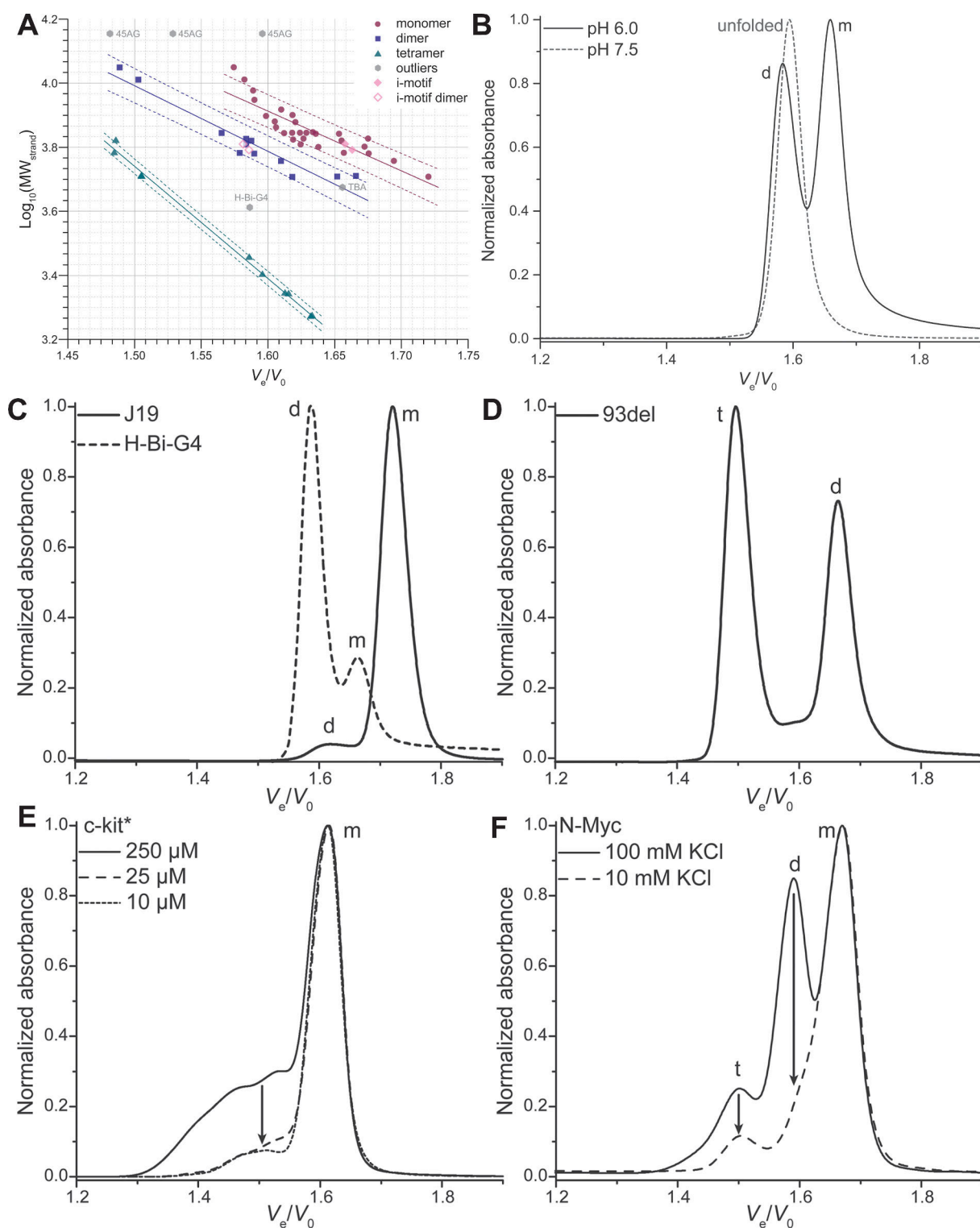


FIGURE 3.3. Tracés du logarithme décimal du poids moléculaire en fonction du volume d'élution relatif pour les oligonucléotides formant des G4s mono-, di- ou tétramériques, ou des i-motifs mono- ou dimériques (A). Exemples de chromatogrammes normalisés (B-F) : influence du pH sur la formation d'i-motif d'une séquence du promoteur d'oncogène c-MYC (B;  $A_2TC_3AC_4TC_3AC_4T_2$  (145)), mesure de la formation relative de G4 monomérique et dimérique en fonction de la séquence (C;  $GIGTG_3TG_3TG_3T$  et  $G_3ACGTAGTG_3$  (146, 147); J19 forme un G4 monomérique qui peut former de petites quantités de dimères *stackés*, tandis que H-Bi-G4 ne peut former que des G4s dimériques, le monomère ne contenant que 2 G-tracts et étant donc non structuré), formation d'un mélange de G4 di- et tétramérique *interlockés*, (D;  $G_4TG_3AG_2AG_3T$  (148); les deux espèces ne s'interconvertissent pas, ou très lentement, et la nature du tétramère n'est toujours pas connue!), influence de la concentration en oligonucléotide sur la formation de multimères par des séquences issues de promoteurs d'oncogènes c-KIT et n-MYC (E;  $G_2CGAG_2AG_4CGTG_2C_2G_2C$  (149), et F;  $TAG_3CG_3AG_3AG_3A_2$  (150)).



de sa concentration, de la température, de la présence de ligands, etc. La stabilité de ces structures est également fortement dépendante de ces facteurs. Si ce polymorphisme rend difficile l'évaluation des structures formées, il offre néanmoins la possibilité de cibler des gènes spécifiques, ou de concevoir des nanostructures.

Pour mieux comprendre le polymorphisme induit par les cations en particulier, nous avons lancé une étude visant à découvrir des séquences d'ADN dont les structures et stabilités présentent une très forte dépendance à la nature et à la concentration des cations, ce qui leur permet de fonctionner comme des commutateurs allostériques (31).

À cette occasion, j'ai encadré **Marie Toulisse** (IUT, Bordeaux; désormais Responsable cellule approvisionnement à la SNCF) avec laquelle nous avons criblé de nombreuses séquences par dichroïsme circulaire. Après avoir identifié les séquences les plus prometteuses, nous avons caractérisé thermodynamiquement et cinétiquement l'influence des cations par de nombreuses techniques spectroscopiques (UV-vis, UV-melting, CD, RMN), spectrométriques (spectrométrie de masse native, mobilité ionique) et chromatographique (méthode SEC décrite ci-avant, Section 3.1).

Le motif  $G_3T_2$  s'est montré particulièrement prometteur et nous avons caractérisé en détails les séquences 222 ( $d[(G_3T_2)_3G_3]$ ) et 222T ( $d[T(G_3T_2)_3G_3T]$ ), qui fonctionnent comme des doubles commutateurs (Figure 3.4A). Le premier switch se fait entre un G4 à trois tétrades, antiparallèle et coordonnant du sodium et un G4 à trois tétrades, parallèle, et coordonnant du potassium. L'interconversion est déclenchée par l'ajout de petites quantités de potassium à un échantillon riche en sodium. La très grande différence de stabilité entre la forme  $Na^+$  et la forme  $K^+$  explique que de petites quantités de se dernier suffisent à une interconversion quantitative et *instantanée* (plus rapide que le temps mort expérimental) à température ambiante.

Le second switch se fait entre un G4 deux tétrades, antiparallèle, et un G4 trois tétrades, parallèle, par simple augmentation de la concentration en potassium. De manière générale, ces deux oligonucléotides se structurent de façon très dépendante à la nature du cation, ce qui en fait des modèles très intéressants à explorer (Figure 3.4B). J'ai confirmé et résumé la tendance globale de stabilisation des G4s en fonction de la nature du cation :  $Sr^{2+} > Ba^{2+} > K^+ > Ca^{2+} > Na^+$ ,  $NH_4^+$ ,  $Rb^+ > Mg^{2+} > Li^+ \geq Cs^+$ .

Les thymines en 5' et 3' de 222T, mais absente de 222, rendent l'interconversion plus rapide (cinétique en CD et UV-vis; Figure 3.5A), la complexation du potassium plus coopérative (détermination de  $K_d$  par spectrométrie de masse native; Figure 3.5B, C), et inhibe la dimérisation des G4s (détermination de stoechiométrie par MS native et SEC-HPLC, comme décrit ci-avant, Section 3.1). Ces commutateurs sont donc modulables avec des modifications discrètes et minimales.

Durant le criblage, nous avons mis en évidence que de nombreuses séquences riches en guanines adoptent le même comportement, même si souvent de façon moins marquée. Ces systèmes d'ADN présentent un intérêt fonctionnel (par exemple en tant que nanomachines), mais leur polyvalence structurelle en fait également un outil méthodologique précieux pour la compréhension de la thermodynamique et de la cinétique du repliement de l'ADN G4. La modulation facile de structures G4s en fonction des conditions expérimentales en fait également des modèles très pratiques pour le développement analytique. J'aurai d'ailleurs de nombreuses fois l'occasion d'utiliser 222T dans la suite

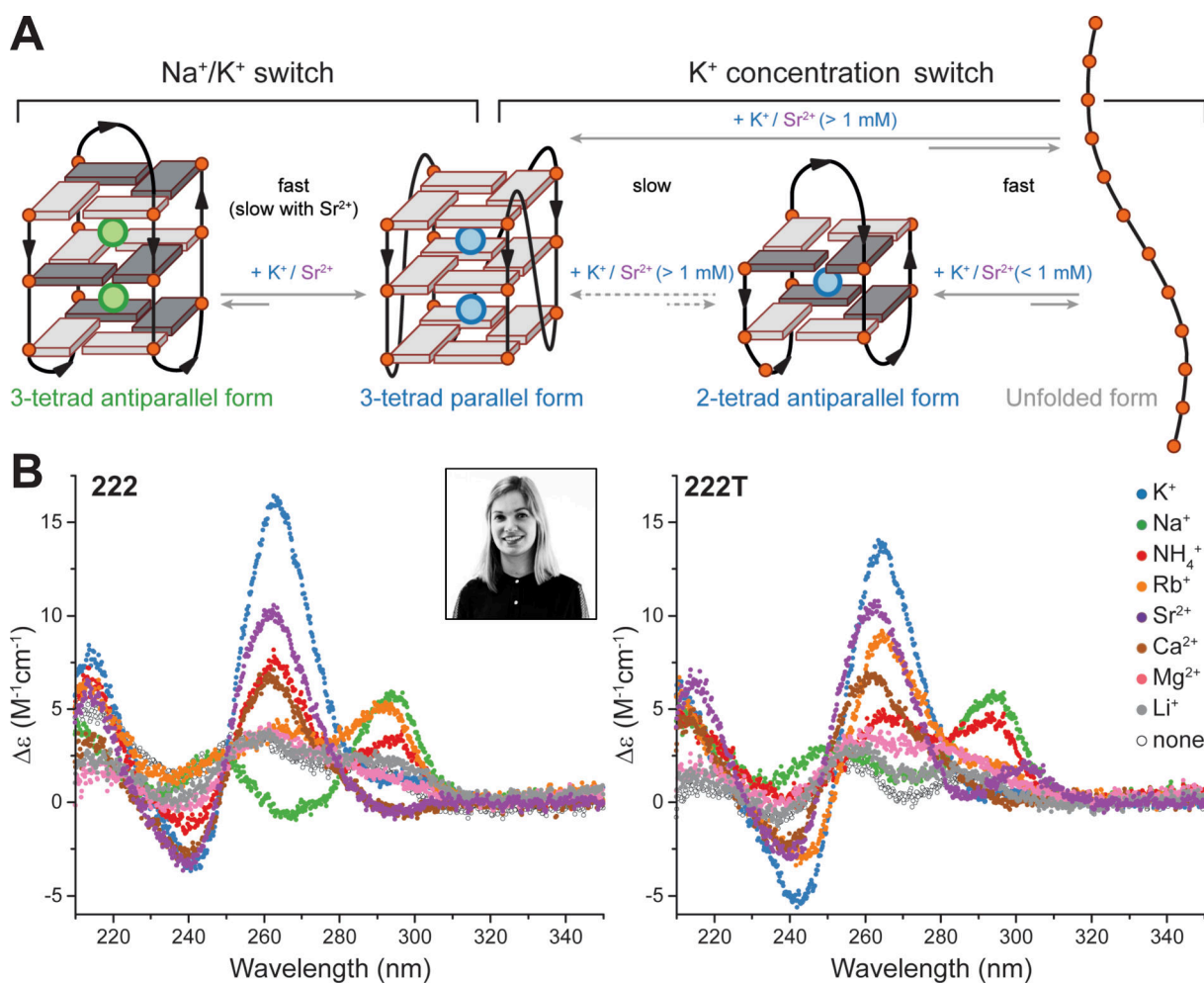


FIGURE 3.4. A. Des séquences riches en guanines peuvent commuter entre plusieurs topologies (parallèle vs. antiparallèle et 2 vs. 3 tétrades) par ajout de petites quantités de potassium à une solution riche en sodium (« Na<sup>+</sup>/K<sup>+</sup> switch ») ou contenant déjà une concentration faible ou nulle en potassium (« K<sup>+</sup> concentration switch »). B. La structure des séquences 222 et 222T est particulièrement dépendante de la nature du cation. Encart : Marie Toulisse.

### 3.2 Influence des cations sur les G4s : découverte et caractérisation de doubles commutateurs

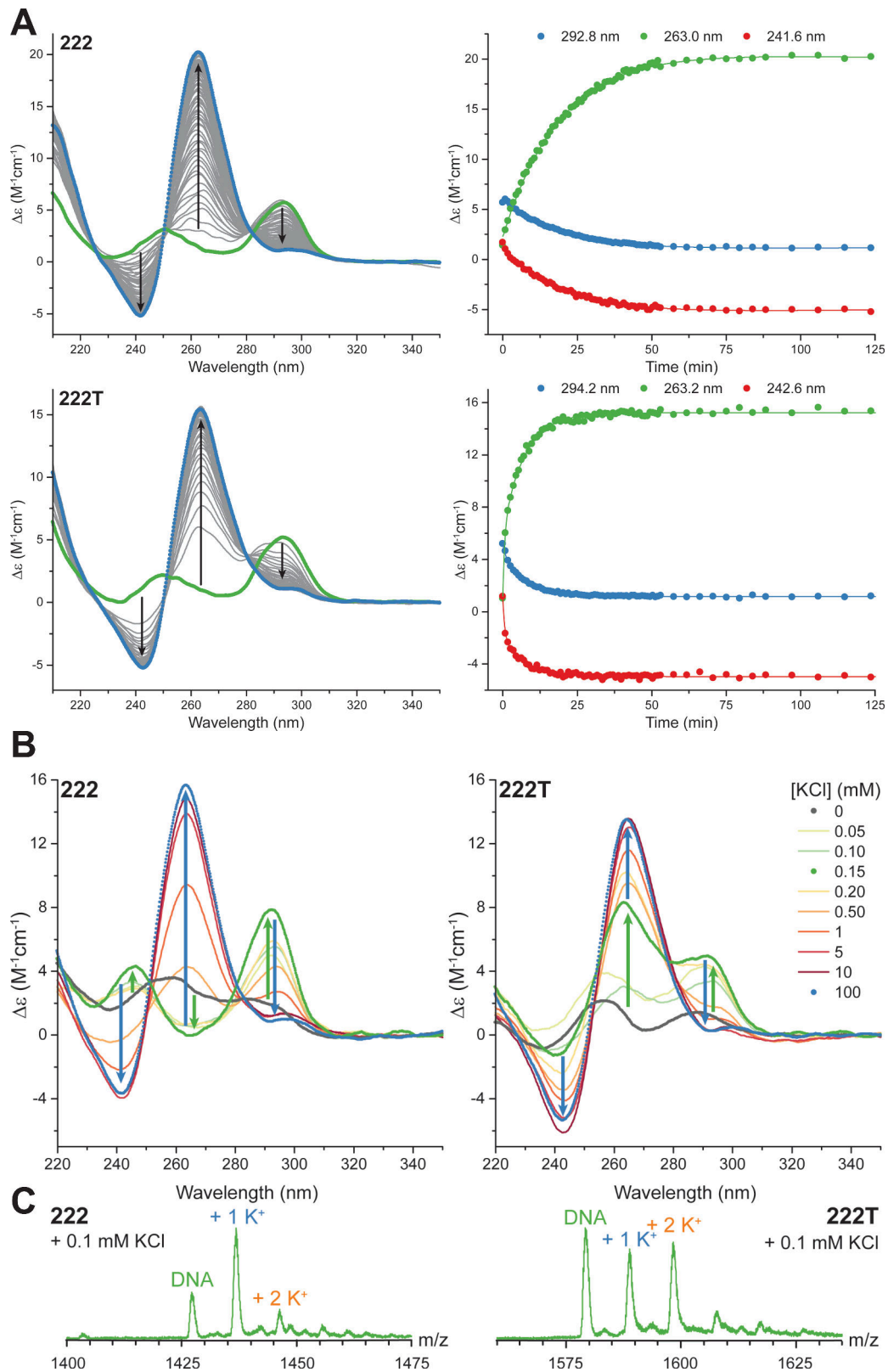


FIGURE 3.5. A. Cinétique d'interconversion  $\text{Na}^+$  vers  $\text{K}^+$  à 4°C suivie par CD, après ajout de 5 mM de KCl à une solution de 222 initialement structurée dans 100 mM de NaCl (vert; point final : bleu). B. Mise en évidence des différentes topologies formées en fonction de la concentration en KCl, par CD (B) et détermination du nombre de tétrades formées par mesure de la stœchiométrie en potassium par spectrométrie de masse native (C).

de mes travaux, notamment à mon retour à l'IECB comme maître de conférences (voir le Chapitre 5, en particulier les Sections 5.2.1 et 5.3.1.1) (6, 72, 152).

Cette publication, ainsi qu'un chapitre de livre que j'ai écrit avec Valérie et Jean-Louis sur les interactions entre G4 et métaux alcalins (23), sont fréquemment cités par la communauté G4 (> 350 citations). Il s'agit d'un de mes papiers préférés car j'y fait appel à des techniques analytiques "basiques" mais permettant, avec un traitement des données *ad hoc*, de produire des résultats très significatifs.

### 3.3 Conclusion

Au cours de ce second postdoctorat, j'ai développé une approche par chromatographie d'exclusion stérique spécifiquement calibrée pour distinguer les polymorphismes structuraux des acides nucléiques (duplexes, i-motifs, triplexes, G-quadruplexes). Cette méthode a permis de caractériser les structures et la stœchiométrie de plus d'une centaine d'oligonucléotides. Par ailleurs, l'identification, par une combinaison de spectroscopies et spectrométrie de masse native de séquences capables de commuter entre topologies de G4 selon la nature ou la concentration des cations a révélé des "interrupteurs" moléculaires très sensibles. Ces travaux illustrent l'importance d'une approche analytique multi-technique pour étudier les structures non canoniques de l'ADN, dont l'important polymorphisme a été encore un peu plus mis en lumière.

Personnellement, c'est une évidence mais il convient tout de même de l'écrire, ce contrat postdoctoral jettera les bases de mon retour en 2017, mais pas uniquement. C'est également ces travaux qui m'ont réellement lancé dans le monde de la spectrométrie de masse (native) et m'ont amené au chapitre suivant de ma carrière, dans lequel il est beaucoup question de MS (mais pas native).

## Chapitre 4

### R&D Scientist à *Quality Assistance*

#### 4.1 Contexte

Je pars au printemps 2015 vers la Belgique, où j'ai occupé un poste de chercheur dans l'industrie pharmaceutique pendant un peu plus de deux ans. J'y ai apporté un profil de chercheur nécessaire pour les projets à mener, accompagné d'une connaissance toute relative de l'industrie pharmaceutique. J'y travaille sous la direction d'Arnaud Delobel, formant une équipe de deux personnes, à laquelle viennent s'ajouter deux scientifiques "prêtés" par le département opérationnel pour six mois chacun : Anicet Catrain et Fabrice Cantais (maintenant *Analytical leader for NBE and particularly monoclonal antibody* chez Servier) (Figure 4.1). Leur aide précieuse m'aura considérablement facilité la vie, me permettant de me lancer sur mes projets tout en apprenant toutes les particularités de cette industrie très particulière.

*Quality Assistance* est une CRO, pour *contract research organisation*, c'est à dire une société à laquelle de grands groupes pharmaceutiques sous-traitent certaines de leurs analyses, le plus souvent en suivant les Bonnes Pratiques de Fabrication (BPF, ou GMP pour *Good Manufacturing Practices*). Cela implique parfois le transfert de méthodes, du client vers la CRO, ou le développement de nouvelles méthodes. Ces méthodes peuvent être développées pour les besoins spécifiques du client à partir de méthodes génériques, "prédéveloppées" à l'avance et proposées au catalogue de la CRO.

Par exemple, si la CRO souhaite vendre des analyses de *peptide mapping*, elle peut développer une méthode générique développée sur une ou plusieurs protéines représentatives de ce qui pourraient lui être soumis par un client, et démontrant sa capacité à développer et réaliser ce type d'analyse. Lorsqu'un client soumet une protéine thérapeutique à caractériser, les analystes n'ont "plus qu'à" affiner la méthode. Dans l'exemple du *peptide mapping*, cela peut être par exemple le choix d'une ou plusieurs protéases adaptées à la séquence (et donc aux sites consensus de clivage) de la protéine du client.

Le contexte réglementaire très strict de l'industrie pharmaceutique est un point de contexte très important à garder en tête dans ce chapitre. Les méthodes doivent être développées dans un environnement BPF très différent des laboratoires académiques, et doivent pouvoir être validées en suivant les guidances des autorités.

J'ai été recruté grâce à un financement FEDER (Fonds européen de développement régional) de l'Union Européenne, pour travailler sur deux grands axes :





FIGURE 4.1. A. Le site de *Quality Assistance*, à Donstienne (Belgique). B et C. Production audiovisuelle de l'auteur lors de sa présence chez *Quality Assistance*. Le lecteur curieux pourra retrouver les vidéos sur YouTube. B : L'auteur vantant les mérites du logiciel UNIFI pour le compte de la société Waters (photo : Arnaud Delobel). C : Anicet Catrain et l'auteur, faisant semblant d'utiliser la télécommande de l'automate PAL de leur plateforme HDX/MS toute neuve. D et E : Arnaud Delobel et Fabrice Catrain.

#### 4.2 Développement de méthodes LC/MS pour la caractérisation et quantification de protéines thérapeutiques

1. Le renouvellement du catalogue de méthodes génériques pour la caractérisation physico-chimique et la quantification de protéines thérapeutiques, et
2. L'implémentation d'une plateforme d'échange hydrogène-deutérium couplé à la spectrométrie de masse (HDX/MS) pour la caractérisation de structures et interactions de protéines thérapeutiques.

Dernier point divergeant significativement du monde académique : le rythme de développement. Il n'est ici pas question de passer plusieurs mois sur chaque méthode générique. J'ai dédié environ une semaine par méthode, rédaction de la SOP (*standard operating procedure*) comprise. Il ne s'agit pas ici de passer beaucoup de temps à comprendre pourquoi une méthode fonctionne, mais de la faire fonctionner de façon robuste. De même, lorsqu'une méthode ne fonctionne pas comme souhaité, une approche alternative est rapidement testée.

Pour le second point (HDX/MS), en revanche, j'ai dédié plusieurs semaines à la mise en place de la plateforme analytique puis aux développements de méthodes.

#### 4.2 Développement de méthodes LC/MS pour la caractérisation et quantification de protéines thérapeutiques

Les protéines sont de plus en plus utilisées à des fins thérapeutiques et représentent une part importante des médicaments les plus vendus (153). Elles comprennent notamment les anticorps monoclonaux et les protéines produites par la technologie de l'ADN recombinant (protéines de fusion, cytokines et facteurs de croissance). Leur caractérisation, du développement au contrôle qualité avant libération des lots, ainsi que la surveillance après la mise sur le marché, est un défi en raison de leur taille et de leur hétérogénéité inhérente au mode de production, notamment en ce qui concerne les modifications post-traductionnelles (PTM) (154–156).

L'analyse d'anticorps et autres protéines recombinantes requiert d'examiner de nombreux paramètres : identités de séquences, nature et position des PTMs, variants de charge, oligomérisation, concentration, structure et interactions. De plus, leur mode de production rend ces paramètres hétérogènes et leur formulation peut être complexe. La caractérisation des variants de ces macromolécules nécessite donc plusieurs méthodes orthogonales, chacune apportant des informations complémentaires aux autres.

J'ai ainsi développé plus de vingt-cinq méthodes pour la caractérisation de protéines et petites molécules thérapeutiques par UPLC couplée à la spectrométrie de masse à haute résolution (ESI-QTOF), aux niveaux intact, des sous-unités, peptidique et des petites molécules (impuretés, *leachables*). La caractérisation de la glycosylation (Section 4.3) et de la structure et des interactions (Section 4.4) ont fait l'objet de travaux plus poussés et sont détaillés ci-après.

Pour ces développements, j'ai principalement utilisé Humira® (adalimumab), un anticorps anti-TNF $\alpha$  de sous-classe IgG1, approuvé par la FDA et l'EMA pour différentes maladies inflammatoires, dont la polyarthrite rhumatoïde, l'arthrite juvénile idiopathique, le rhumatisme psoriasique, le psoriasis et la maladie de Crohn. L'adalimumab est un anticorps recombinant entièrement humain exprimé dans des cellules CHO. Comme pour toutes les protéines produites par les technologies de l'ADN recombinant, le produit final est un mélange de différents variants. Quand j'entame ces travaux, c'est le médicament



le plus vendu dans le monde en 2014, avec des ventes globales de plus de 13 milliards de dollars. J'ai également travaillé avec deux anticorps conjugués (ADC, pour *antibody drug conjugate*) commerciaux, le brentuximab vedotin (Adcetris®) et le trastuzumab emtansine (Kadcyla®). Ces ADCs diffèrent, entre autres choses, par la nature de la conjugaison aux cystéines ou lysines, respectivement (157).

Pour ces travaux, j'ai couplé de nombreuses phases stationnaires chromatographiques à la spectrométrie de masse haute résolution (ESI-QTOF), qui sert principalement à identifier les espèces éluées. Parmi les phases stationnaires mises en jeu, citons la phase inverse (RP; confirmation de séquence; nature, position et abondance des PTMs), l'exclusion stérique (SEC; détection et quantification de dimères, oligomères et fragments), l'échange d'ions (IEX; variants de charge), l'interaction hydrophobe (HIC; quantification des stœchiométries de conjugaison de petites molécules aux ADC), et l'interaction hydrophile (HILIC; caractérisation de la nature, position et abondance des glycans, voir Section 4.3) (158, 159).

Le couplage direct de certaines de ces méthodes chromatographiques à la spectrométrie de masse n'est pas possible du fait de l'utilisation de tampons et sels non volatils, en particulier la SEC, l'HIC et l'IEX. Une approche classique consiste à collecter des fractions, puis à les dessaler avant l'analyse en MS. Cette méthode est laborieuse car elle nécessite que l'analyste recueille les fractions manuellement ou, s'il utilise un collecteur de fractions, qu'il mesure avec précision le volume de vide entre le détecteur et le collecteur. En outre, la collecte précise des pics peut s'avérer difficile, notamment lorsqu'il s'agit de chromatogrammes mal résolus. Enfin, la génération d'un certain nombre de fractions à partir de chaque échantillon n'est pas souhaitable dans un environnement réglementé, où le suivi des échantillons est essentiel.

Pour remédier à cela, j'ai développé des méthodes d'UPLC bidimensionnelle (2D-LC) en mode *heartcutting*, mettant en jeu deux vannes, trois pompes et trois colonnes (160). Chaque *heartcut* se compose de quatre étapes (Figure 4.2A). L'échantillon est d'abord dirigé vers la première colonne (SEC, IEX, HIC), où les analytes sont séparés, puis détectés par UV ou par toute autre méthode de détection compatible avec la phase mobile (Figure 4.2A-a). Lorsque l'analyte d'intérêt sort de la colonne, une vanne deux positions à six port le dirige vers le flux de la pompe isocratique, puis vers la colonne de piégeage en phase inversée (Figure 4.2A-b). L'utilisation d'une phase mobile aqueuse acidifiée permet de concentrer les analytes sur le piège, tout en éliminant la plupart des sels et en préparant la séparation de deuxième dimension, généralement réalisée avec un système de gradient eau-acétonitrile-acide formique. La vanne rebascule à la fin du *heartcut*, dont la durée est programmée en fonction de la largeur et de la résolution du pic. Dans une troisième étape, l'analyte d'intérêt est focalisé sur la colonne de piégeage (Figure 4.2A-c). Enfin, une deuxième vanne est activée pour envoyer le contenu de la colonne de piégeage à contre-courant, vers la deuxième colonne analytique, et finalement vers le spectromètre de masse (Figure 4.2A-d).

La Figure 4.2B montre l'exemple de l'analyse de variants de charges par couplage IEX-RP/MS, où la combinaisons de ces méthodes apparaît comme particulièrement nécessaire pour identifier des variants de charge. Les variants basiques sont formés suite à la perte de lysines C-terminales, avec une différence de masse assez simple à détecter par MS seule (- 128 Da). En revanche, les variants acides se forment consécutivement à des déamidations de résidus Asn, formant des résidus Asp et isoAsp (Figure 4.2C). La différence de masse entre Asn et Asp/isoAsp est minimale (une unité de masse sur une molécule

## 4.2 Développement de méthodes LC/MS pour la caractérisation et quantification de protéines thérapeutiques

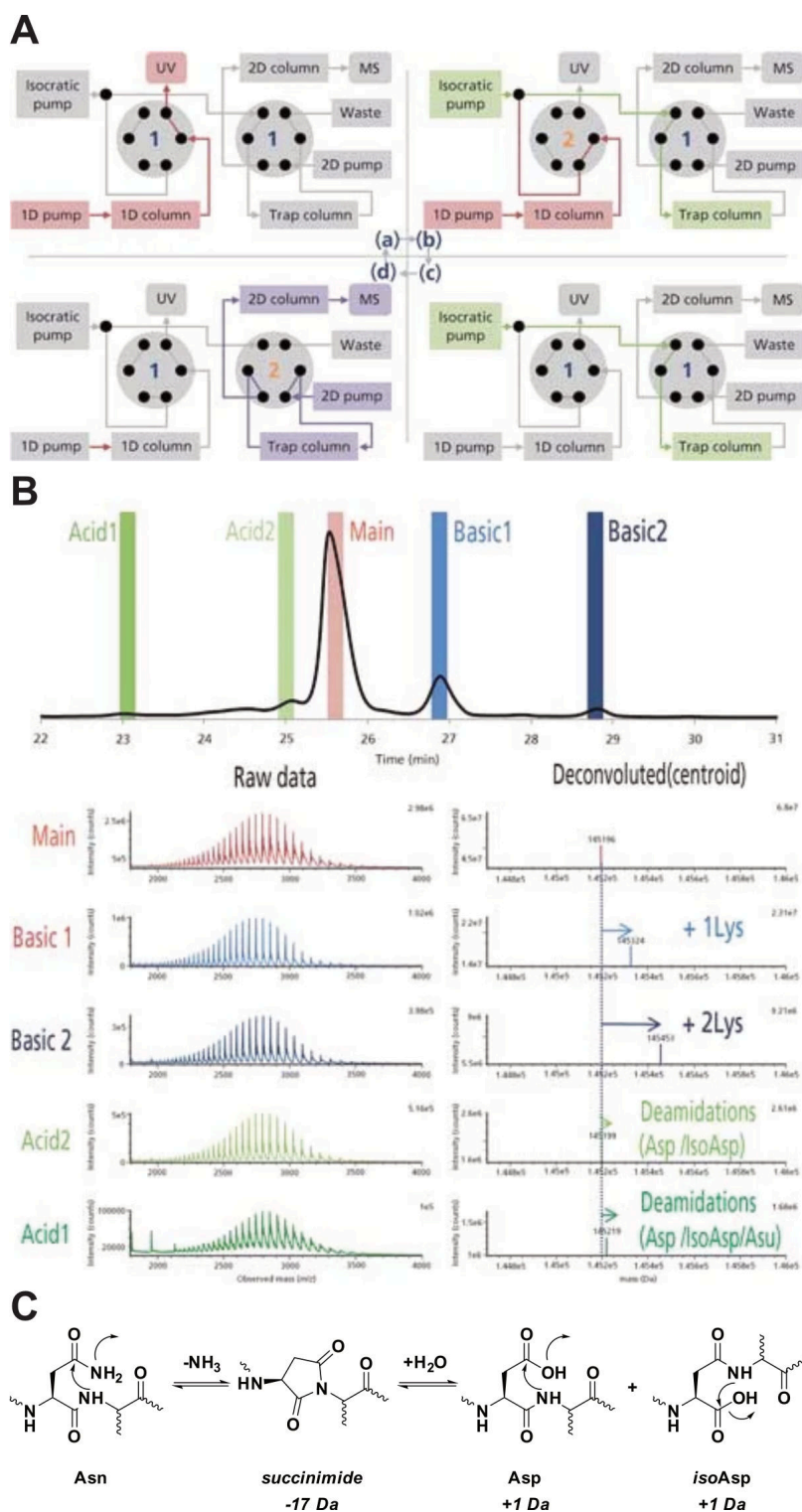


FIGURE 4.2. 2D-LC/MS de protéines thérapeutiques. A. Setup, B. Exemple d'analyse de variants de charge par IEX-RP/MS, où les fractions *heartcut*ées sur la dimension IEX sont analysées en RP/MS (gauche : données brutes, droite : données déconvoluées) C. Déamidation d'un résidu Asn en dérivé succinimide, lui-même hydrolysé pour former Asp et isoAsp, provoquant une augmentation de masse de l'ordre d'1 Da.

d'environ 150 kDa!), et la séparation LC est alors très utile. Celle-ci permet aussi de quantifier les différents variants par simple détection UV, étant entendu que les pics ont été initialement identifiés par masse.

La communication de ces résultats dans le milieu industriel m'a ensuite amené à participer à un *alpha* test de colonnes innovantes pour l'analyse d'anticorps, et au *beta* test d'un logiciel de contrôle et analyse de données LC-MS, en collaboration avec la société Waters. J'ai également collaboré avec Promega pour le développement de kits de protéases pour *peptide mapping* et pour l'évaluation de l'activité ADCC (cytotoxicité à médiation cellulaire dépendante des anticorps) (161).

Finalement, ces travaux m'ont amené à utiliser une large gamme de spectromètres de masse (ESI-QTOF, triple quadrupôle, ion-trap; MALDI-TOF), ce qui m'a permis de renforcer mon expertise sur le sujet, par exemple en mettant au point une méthode LC/MS/MS (MRM) pour la quantification de protéines thérapeutiques diluées dans des fluides biologiques.

### 4.3 Glycosylation des protéines thérapeutiques

La glycosylation est l'une des modifications post-traductionnelles les plus courantes des protéines et joue un rôle essentiel dans leur fonction, immunogénicité, clairance plasmatique et résistance aux protéases (162). Par conséquent, le profil de glycosylation des protéines thérapeutiques est considéré comme un attribut de qualité critique, tant pour les innovateurs que pour les biosimilaires, et doit faire l'objet d'une analyse approfondie (163). Les protéines thérapeutiques sont généralement produites dans différents systèmes d'expression, dont les mécanismes de glycosylation fonctionnent par étapes séquentielles et compétitives, créant ainsi des micro- (nature des glycanes pour un site donné) et macro- (nombre et emplacement des sites) hétérogénéités de glycosylation (156). En outre, la nature ramifiée des glycanes crée un autre niveau de complexité par rapport aux séquences oligonucléotidiques et peptidiques linéaires avec lesquelles j'ai l'habitude de travailler.

Afin de se conformer aux exigences réglementaires sur cette modification post-traductionnelle de plus en plus étudiée, j'ai proposé l'utilisation d'une combinaison de diverses méthodes LC ou LC/MS(/MS) opérant à différents niveaux d'analyse (glycanes libérés, peptides, intacts et sous-unités) simples à mettre en place sans utilisation de matériel et de logiciels hautement spécialisés (163–167). Ce *workflow* permet à des scientifiques non spécialisés de caractériser de manière exhaustive les *N*- et *O*-glycosylations des protéines thérapeutiques pour, *a minima*, se conformer aux exigences des autorités. Il comprend des méthodes qui sont ensuite devenues très *tendance* - en particulier l'HILIC *widepore*/MS (168–172), mais aussi d'autres peu usitées comme les colonnes à carbone graphite poreux qui présentent une excellente rétention des composés polaires, orthogonale à la phase inverse, et dont nous avons montré l'excellence pour la séparation des *O*-glycanes.

Lors de ce développement, nous avons caractérisé et publié pour la première fois la *N*- et *O*-glycosylation complète (natures, sites et occupations des sites) de l'Etanercept, une protéine de fusion parmi les médicaments les plus vendus au monde (163, 164). Cela a nécessité l'utilisation de plusieurs techniques chromatographiques, mais également de plusieurs méthodes de fragmentation en phase gazeuse (CID et ETD; Figure 4.3), et est probablement l'un des objets les plus complexes que j'ai eu à analyser.

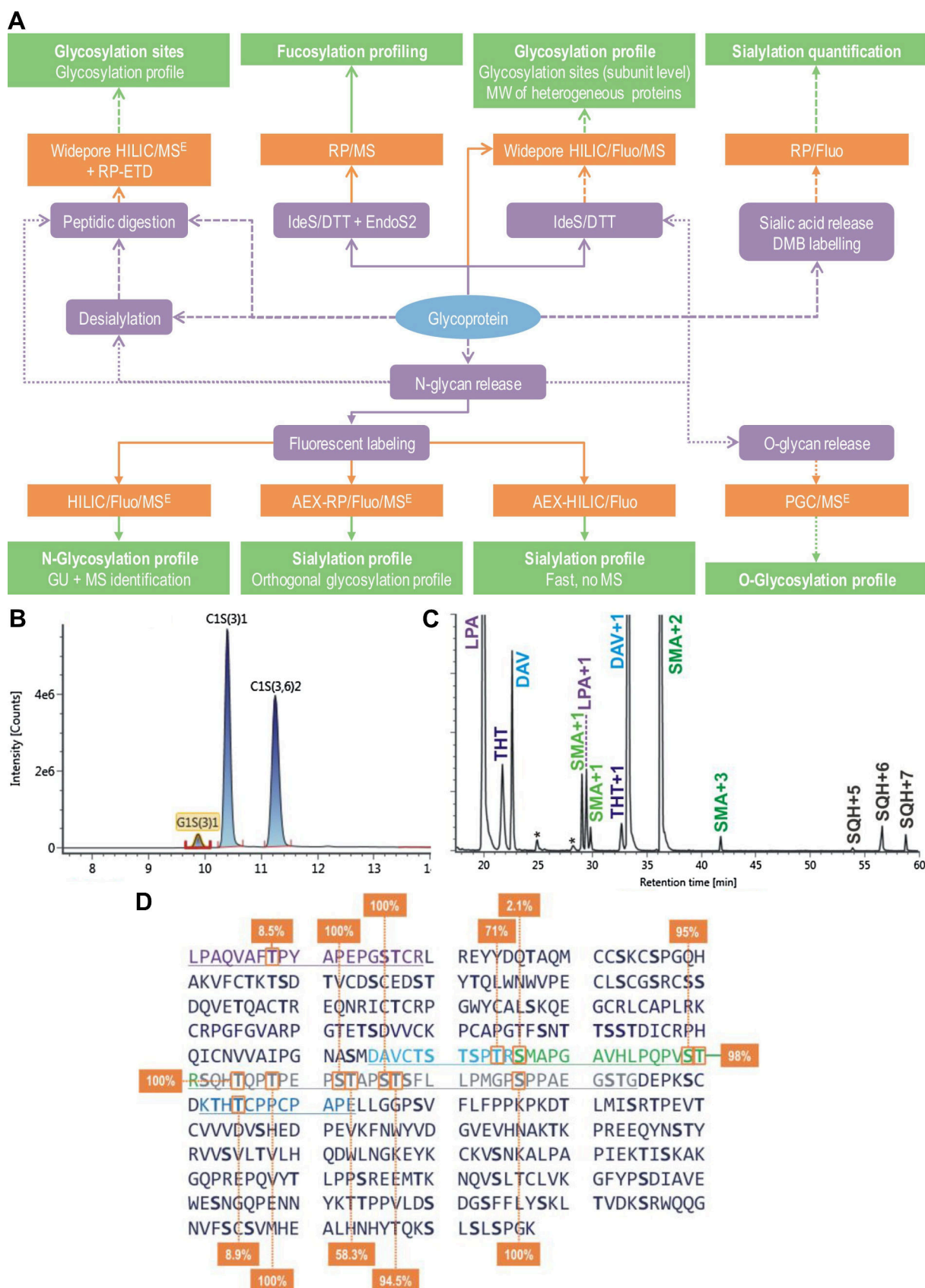


FIGURE 4.3. Vue d'ensemble simplifiée de la caractérisation des glycoprotéines thérapeutiques. Les pointes de flèches indiquent si la protéine (ouverte) ou les glycans (remplis) sont récupérés pour l'analyse (A). L'analyse concerne les *N*-glycans (traits pleins), *O*-glycans (tirets courts) ou les deux (tirets longs). Une combinaison d'analyse des glycans clivés sur colonne à carbone graphique poreux couplée à la spectrométrie de masse PGC/MS<sup>E</sup> (B) et d'analyse des peptides glycosylés sur colonne hydrophile à pore large (widepore HILIC) couplée à la spectrométrie de masse (C) a permis d'établir pour la première fois la nature de l'*O*-glycosylation, leurs sites et le degré d'occupation de ces sites pour la protéine de fusion Etanercept (D).

#### 4.4 Échange hydrogène-deutérium couplé à la spectrométrie de masse pour la caractérisation de structures et interactions de protéines thérapeutiques

Lors de ma dernière année en poste, mes activités se sont centrées sur l'analyse de structures et interactions de protéines thérapeutiques par échange hydrogène-deutérium couplé à la spectrométrie de masse (HDX/MS). L'HDX repose sur la mesure des cinétiques d'échange des protons amides peptidiques par des deutérons du solvant, qui varient très fortement en fonction de leur structuration (173). L'échange est ralenti en cas de formation de liaisons hydrogène et d'accès réduit du solvant. L'échange est quantifié par spectrométrie de masse en tirant partie de la différence de masse isotopique entre l'hydrogène et le deutérium.

Classiquement, une approche *bottom up* est utilisée (Figure 4.4). Les protéines sont incubées dans un tampon deutéré pendant un certain temps (d'une dizaine de secondes à plusieurs heures), ce qui permet la deutération des amides. La réaction d'échange est stoppée en abaissant le pH (vers 2.5) et la température. Des dénaturants et agents réducteurs peuvent être ajoutés à cette étape pour dénaturer les protéines. Les protéines sont ensuite digérées par une protéase, généralement la pepsine, à température ambiante. Les peptides sont dessalés et séparés à l'aide d'un système UHPLC réfrigéré, en phase inverse, et dirigés vers un spectromètre de masse, où ils sont ionisés par électrospray et soumis à une analyse de masse. Les niveaux de deutération sont déterminés, généralement en comparant la masse moyenne à la valeur  $m/z$  du centroïde du peptide, pondérée par l'intensité. La cinétique de deutération de chaque peptide peut ensuite être tracée.

Cette approche permet de comparer les cinétiques d'échange de peptides de plusieurs états d'une protéine, par exemple une protéine stressée (haute température, pH non physiologique, oxydant) à cette même protéine non stressée, ou encore un complexe anticorps/antigène à l'anticorps seul. En superposant les résultats de tous les peptides, il est possible d'identifier les sites perturbés (changement structural, complexation non covalente) à l'échelle de l'acide aminé. Les différences de cinétiques HDX peuvent enfin être cartographiées sur une structure de la protéine pour faciliter l'interprétation (Figure 4.4).

Le développement de méthodes HDX/MS n'est pas aisé et dépend assez fortement de l'analyte. En particulier, il est important de *bien* digérer la protéine, c'est à dire d'avoir une couverture de séquence la plus complète possible, avec une redondance élevée, de façon à ce que les séquences peptidiques partiellement superposées donnent accès à une information résolue à l'échelle de l'acide aminé. Cela est rendu difficile par le fait que la digestion doit se faire très vite (en ligne), à température ambiante et à un pH non optimal, de façon à ce que les deutérons ne soient pas rééchangés par des protons (on parle de *back exchange*). Cela devient un véritable challenge dans un environnement régulé, notamment de par l'absence de guidances claires sur la caractérisation de structures de protéines thérapeutiques.

Mon premier but a donc été la mise en place la plateforme HDX/MS au sein de la société, une première pour un sous-traitant analytique européen soumis aux bonnes pratiques de fabrication (Figure 4.1). Dans un second temps, j'ai développé des approches robustes pour la comparaison de lots, les tests de stress (175), et la détermination d'épitopes (176, 177).



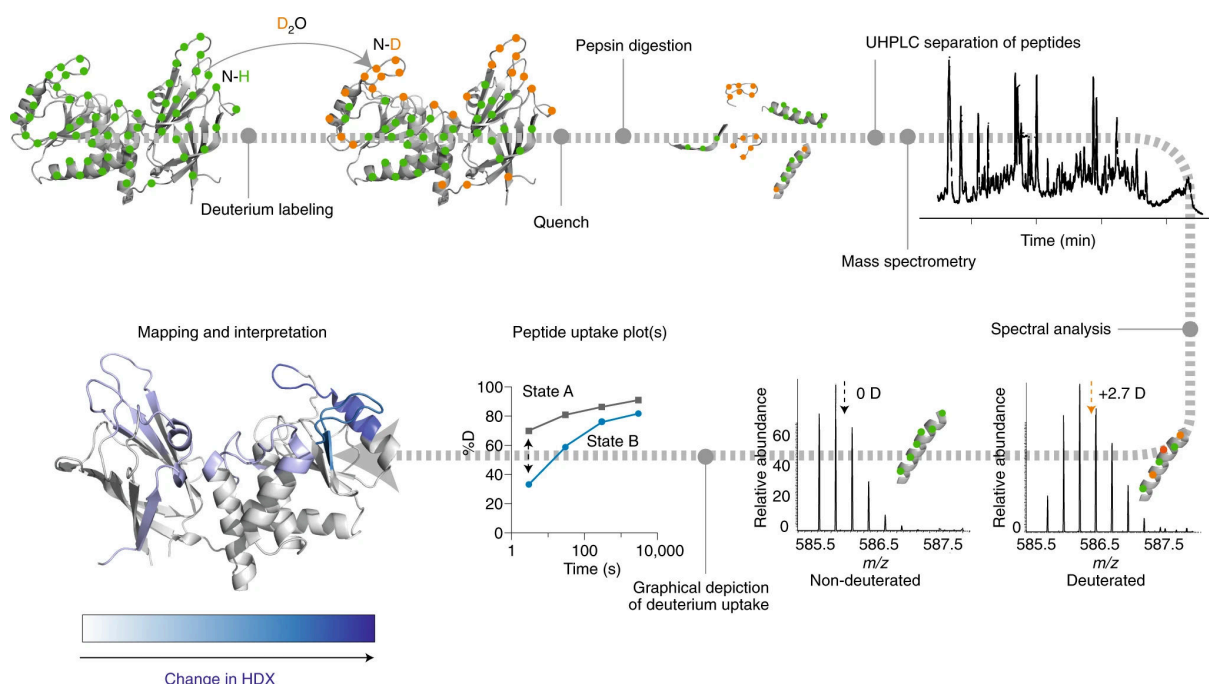


FIGURE 4.4. Expérience typique d'HDX/MS *bottom-up*. Les spectres de masse de l'exemple montrent que le peptide a un niveau de deutérium de 2,7 D (flèche orange vs. noire pour les échantillons deutéré et non-deuté, respectivement). L'état B est ici plus protégé que le A pour le peptide représenté, ce qui peut être cartographié sur la structure de la protéine. Figure reproduite sans modification, extraite de Masson et al. (174). Creative Commons Attribution 4.0 International License.

Enfin, j'ai mis en œuvre notre méthode de détermination d'épitope lors d'une collaboration avec l'entreprise *Ose Immunotherapeutics* (Nantes), qui souhaitait cibler l'expansion des cellules immunitaires « mémoires » pathogènes, comme stratégie thérapeutique prometteuse pour prévenir les attaques auto-immunes chroniques. Dans ce contexte, j'ai déterminé les épitopes de trois nouveaux anticorps monoclonaux anti IL-7R $\alpha$  par HDX/MS (Figure 4.5) (178). Grâce à cela, il a été montré que les anticorps liant le domaine d'interaction de l'IL-7 (site-1) ont des propriétés à la fois agonistes et antagonistes, tandis qu'un anticorps liant aussi le domaine d'hétérodimérisation de l'IL-7R $\alpha$  /  $\gamma$ c (site-2b) présente une activité antagoniste stricte. Une seule injection d'anticorps monoclonaux antagoniste peut induire un contrôle à long terme de l'inflammation cutanée malgré des provocations antigéniques répétées chez des singes présensibilisés.

*Ose Immunotherapeutics* avait déjà identifié les épitopes par une autre méthode avant de nous contacter, sans toutefois nous en informer. J'ai donc réalisé l'analyse de manière indépendante, à l'aveugle, ce qui a naturellement renforcé notre confiance dans les résultats lorsque *Ose* nous a confirmé avoir obtenu les mêmes régions.<sup>1</sup>

## 4.5 Conclusion

Ces années passées dans l'industrie m'ont été très profitables, puisque j'ai pu travailler sur de nombreuses techniques dans un environnement très différent des laboratoires de recherche publiques, tout en conservant une activité académique.

1. À mon plus grand soulagement...

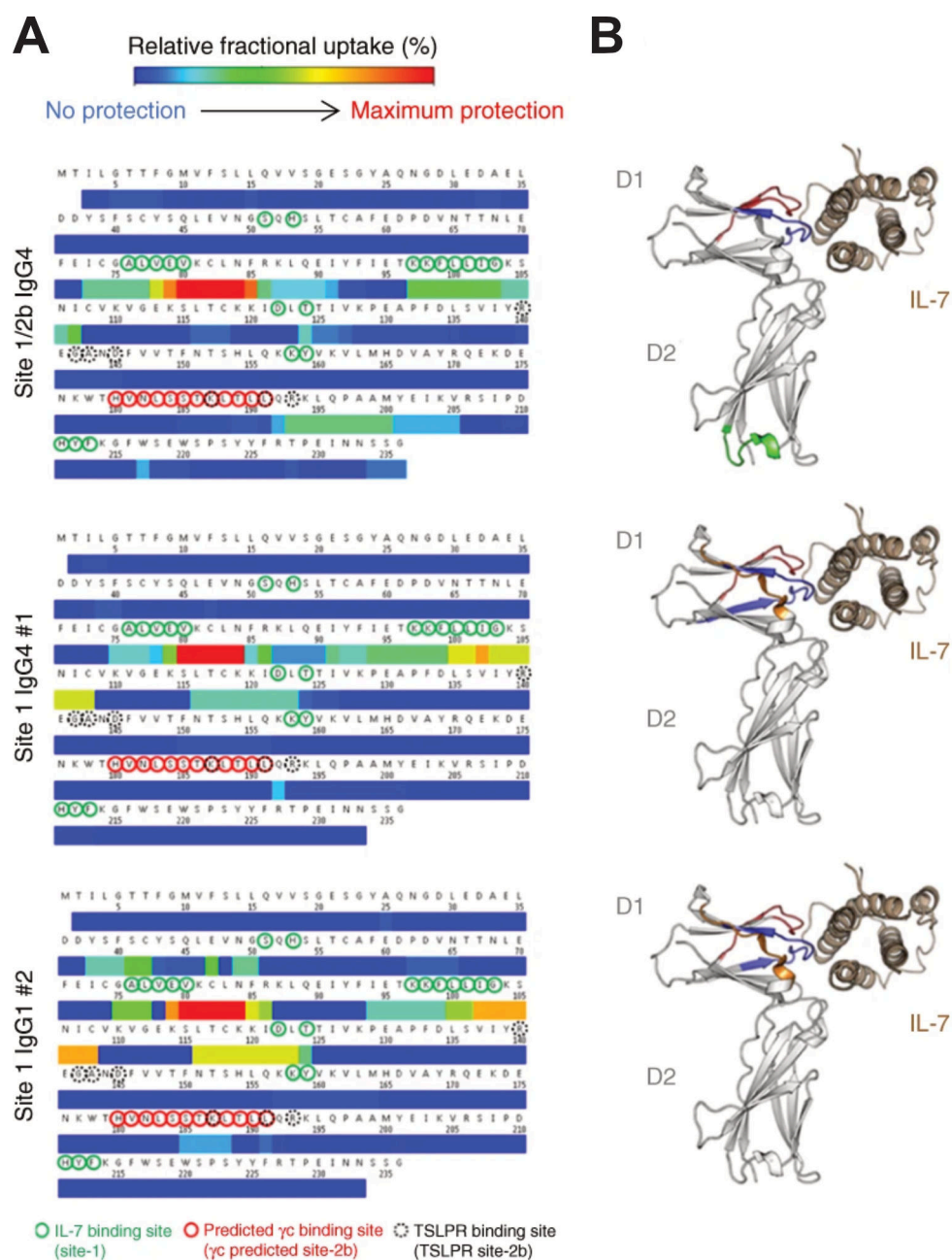


FIGURE 4.5. Détermination d'un épitope par HDX-MS. A. Heatmap des différences de deutération relative entre l'antigène IL-7Rα lié et non lié avec les anticorps IgG4 site-1/2b, IgG4 site-1 #1, et IgG1 site-1 #2. Les acides aminés essentiels à la fixation de l'IL-7 sur l'IL-7Rα (site 1) sont représentés en vert, et le site-2b prédit pour l'interaction de l'IL-7Rα avec la chaîne  $\gamma$  est représenté en rouge. B. Représentation de la structure de chaque épitope IL-7 (marron)/IL-7Rα (gris). Figure adaptée de Belarif et al. [Creative Commons Attribution 4.0 International License](#).



J'ai ainsi travaillé avec un très large éventail de méthodes de caractérisation et de quantification des protéines et des petites molécules thérapeutiques, ce qui a renforcé mes compétences techniques. Les différences de moyens et de temporalité avec le monde académique ne m'étaient pas inconnues, ayant une formation initiale plutôt industrielle, mais cette nouvelle expérience m'a conforté dans mon choix de carrière académique. J'ai apprécié l'efficacité du travail mise en place, mais j'ai regretté le relatif manque de liberté dans les orientations scientifiques et le peu de discussions techniques poussées. Tout cela est parfaitement compréhensible au regard des objectifs d'une société comme *Quality Assistance*, bien entendu.

Travailler dans un contexte réglementaire très strict, bien que chronophage et, soyons honnête, parfois pénible, a également été très formateur. Bien que ne travaillant plus dans un environnement GMP, j'essaie de conserver les bonnes habitudes que j'ai acquises à cette occasion. Et je tiens par ailleurs en très haute estime le système qualité de *Quality Assistance*. Cette expérience constitue également une base importante de mes enseignements, ainsi que de la formation continue que j'offre aux étudiants que je supervise au laboratoire (bonnes pratiques de fabrication, connaissance du monde industriel).

Enfin, bien que les travaux que je résume ci-dessus n'impliquent pas d'acides nucléiques, ils ont été une source d'inspiration pour l'analyse de ces derniers, en particulier l'utilisation de l'HDX/MS pour la caractérisation de structures d'oligonucléotides et de leurs interactions avec des protéines et des petites molécules.



## Chapitre 5

### Thèmes de recherche à l'Université de Bordeaux : résultats et perspectives

#### 5.1 Contexte

Je suis recruté par l'Université de Bordeaux en tant que maître de conférences en 2017 (Figure 5.1). Le poste est attribué au laboratoire de chimie analytique de l'UFR de pharmacie (Bordeaux, campus Carreire) pour les aspects pédagogiques, et à l'équipe BALI (*Biophysics Analysis and Ligands*) de Valérie Gabelica au sein du laboratoire ARNA (INSERM U1212, CNRS UMR 5315, Université de Bordeaux), à l'Institut Européen de Chimie et Biologie (IECB, Pessac), pour la recherche.

J'étais le seul autre chercheur permanent de l'équipe jusqu'au départ de Valérie Gabelica fin 2023 (Figure 5.2), si l'on omet la présence de Frédéric Rosu, l'indispensable ingénieur de recherche de la plateforme de spectrométrie de masse. Peu de ce qui suit aurait été possible sans lui. L'équipe était spécialiste de la spectrométrie de masse native pour étudier des complexes non-covalents, en particulier ceux formés par les acides nucléiques. Je n'introduirai pas ici cette thématique, le lecteur pouvant se référer à une récente revue exhaustive de l'équipe (1). Fin 2023, je rejoins l'équipe PRISM de Cameron Mackereth et déménage au bâtiment Bordeaux Biologie Santé (BBS, Bordeaux, campus Carreire), toujours chez ARNA.

Durant ces premières années comme maître de conférences, j'ai dirigé la thèse de **Matthieu Ranz** grâce à une ADT qui m'a été accordée en 2021 par l'Université de Bordeaux ([soutenue en 2024](#)). J'ai également coencadré la thèse d'**Alexander König**, dirigée par Valérie ([soutenue en 2023](#)).

Mes thèmes de recherche actuels sont centrés sur le développement et l'utilisation de méthodes analytiques et computationnelles pour caractériser les structures secondaires, dynamiques et interactions d'acides nucléiques, et en particulier les G4s. Bien que ces structures soient maintenant connues depuis plusieurs décennies, leur important polymorphisme les rend difficiles à étudier par des méthodes de référence (RMN, cristallographie). Ci-après, je décris brièvement les différents axes que je développe pour mieux comprendre comment ces structures, et d'autres, se forment, se comportent et interagissent. Dans ce qui suit, les Sections 5.2 à 5.5 ont été initiées dans l'équipe BALI, et les suivantes chez PRISM.



FIGURE 5.1. L'équipe BALI en septembre 2017. De gauche à droite : Stefano Piccolo, Eric Largy, Frédéric Rosu, Valérie Gabelica, Sandrine Livet, Nina Khristenko, Clémence Rabin, Anirban Ghosh et Steven Daly



FIGURE 5.2. L'équipe BALI en 2022, sans la cheffe ! : Eric Largy, Corinne Buré, Matthieu Ranz, Alexander Konig, Frédéric Rosu et Dominika Strzelecka. Crédit photo : Inserm/François Guénet.

## 5.2 HDX/MS de bio-macromolécules

### 5.2.1 HDX/MS native d'acides nucléiques

#### 5.2.1.1 Preuve de concept

Alors que l'échange hydrogène-deutérium couplé à la spectrométrie de masse (HDX/MS) est relativement répandu pour étudier la dynamique de protéines (voir Section 4.4), il n'était cependant pas exploité pour étudier les acides nucléiques. Lors de ma prise de poste à l'Université de Bordeaux, je me suis attaché à montrer qu'il s'y prête, offrant des perspectives pour explorer la structure et dynamique d'acides nucléiques, et leurs interactions avec des protéines (Figure 5.3) (152).

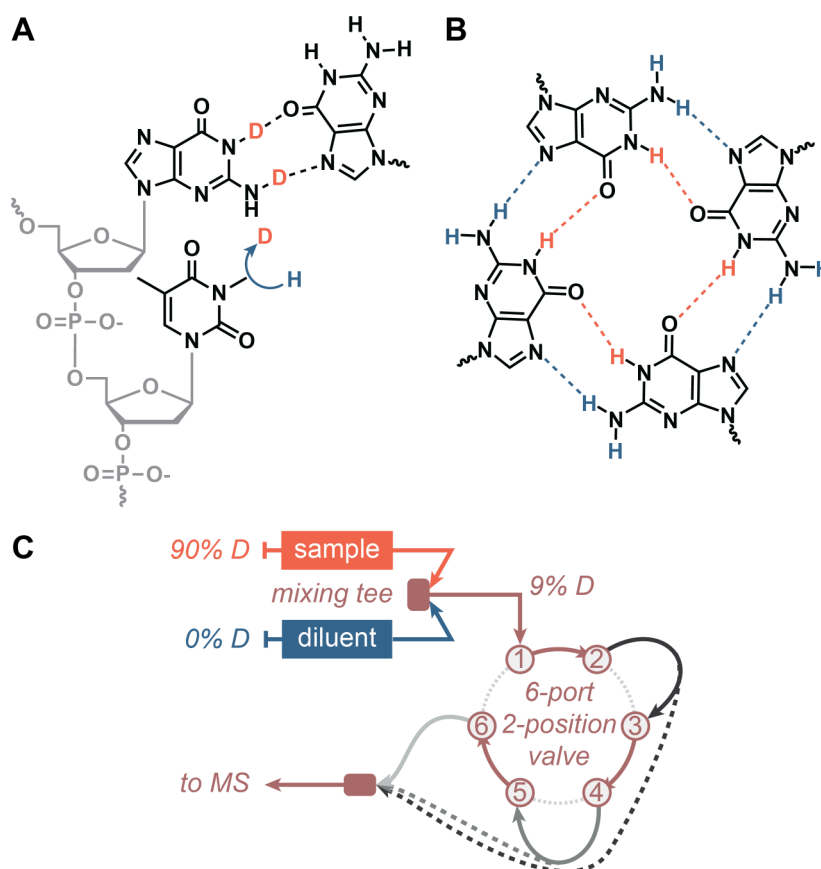


FIGURE 5.3. A. L'HDX repose sur l'échange d'hydrogène et de deutérium en fonction du "statut" de la liaison hydrogène et son accessibilité au solvant. B. Les G-quadruplexes ont été utilisés comme modèles parce que les G-quartets qu'ils contiennent sont formés par un réseau de huit liaisons hydrogène avec les fonctions imino (orange) et amino (bleu) qui peuvent être échangées sur une large plage de temps. C. Pour mesurer les cinétiques d'échange rapide (1-300 s), un montage de mélange en flux continu a été utilisé, en utilisant une vanne à 6 ports et 2 positions pour changer facilement de taille de capillaire (nuances de gris; les lignes en pointillés illustrent le couplage direct d'un capillaire à la source MS pour raccourcir le temps de mélange). Figure reproduite avec permission (152).

J'ai ainsi mis en place des méthodes de mesure d'échange allant de la seconde à plusieurs jours, couplées à la spectrométrie de masse native, et donc applicables à l'étude de complexes non-covalents d'acides nucléiques (Figure 5.3A). Deux techniques ont été développées :

1. Mesure en temps réel de l'échange (RT-HDX) : le mélange d'une solution d'ADN pré-déutéré

avec un tampon non-déutéré déclenche l'échange. Cette solution est ensuite infusée dans le spectromètre de masse, chaque scan correspondant à un temps précis de déutération.

2. Mélange en flux continu (CF-HDX) : les analytes pré-déutérés sont mélangés avec un tampon non-déutéré avant d'être introduits dans le spectromètre de masse. Le temps d'échange est contrôlé par le débit et le volume entre le point de mélange et le spectromètre (Figure 5.3C). Tous les scans MS correspondant au même temps de déutération.

Ces méthodes, très différentes des approches *bottom-up* appliquées aux protéines, nous ont permis de démontrer que :

1. Les nucléobases échangent proportionnellement au ratio H/D en solution (Figure 5.4A), ce qui permet de modéliser les distributions isotopiques; un point critique pour modéliser les données (voir Section 5.4.3).
2. Les vitesses d'échange peuvent être mesurées de façon précise et exacte, ce qui est idéal pour des applications en contrôle qualité par exemple (Figure 5.4B), et sur une large gamme de temps, adaptée à différentes structures (Figure 5.4C).
3. Les vitesses d'échange sont fortement dépendantes de l'appariement de ces bases et leurs interactions intermoléculaires avec de petites molécules, elles reflètent donc les structures des analytes (Figure 5.4D).
4. Le couplage à la spectrométrie de masse native permet de mesurer les cinétiques d'échange individuelles de complexes non covalents, pour ceux qui sont séparés en masse (Figure 5.4E). Cela permet l'utilisation de la méthode pour des études biophysiques d'acides nucléiques (Section 5.2.1.2).

Pour ces premiers travaux, j'ai encadré plusieurs étudiants : **Anaïs Ferrer** (DEUST, Bordeaux), **Yann Bourdeau**, (Licence, Bordeaux; maintenant en thèse dirigée par I. Bestel, Bordeaux), **Laura Fricot** (Master, Bordeaux) et **Émile Feugas** (Pharmacien, Master, Bordeaux).

Ce projet, financé depuis octobre 2021 par une **ANR JCJC**, vise désormais à établir l'HDX/MS comme une méthode de choix pour l'analyse de structures et dynamique d'acides nucléiques, en travaillant autour de trois axes :

1. Etablir les bases fondamentales de l'HDX/MS d'acides nucléiques, pour interpréter les vitesses d'échanges en termes d'équilibre et de dynamique, en mesurant les vitesses d'échange intrinsèques des nucléotides par RMN;
2. Développer une approche HDX/MS native *top-down*, qui couplera une séparation par MS native (par masse et forme) et une fragmentation MS/MS pour obtenir des données à l'échelle du nucléotide pour des analytes séparés par masse et structure;
3. Appliquer ces développements méthodologiques pour répondre à des questions de biophysique des acides nucléiques (thermodynamique et cinétiques de structuration), caractériser les modes interactions entre acides nucléiques et des petites molécules ou protéines d'intérêt thérapeutique, et pour améliorer le contrôle qualité d'oligonucléotides thérapeutiques (structures d'aptamères, tests de stress, comparaison de lots, etc.)

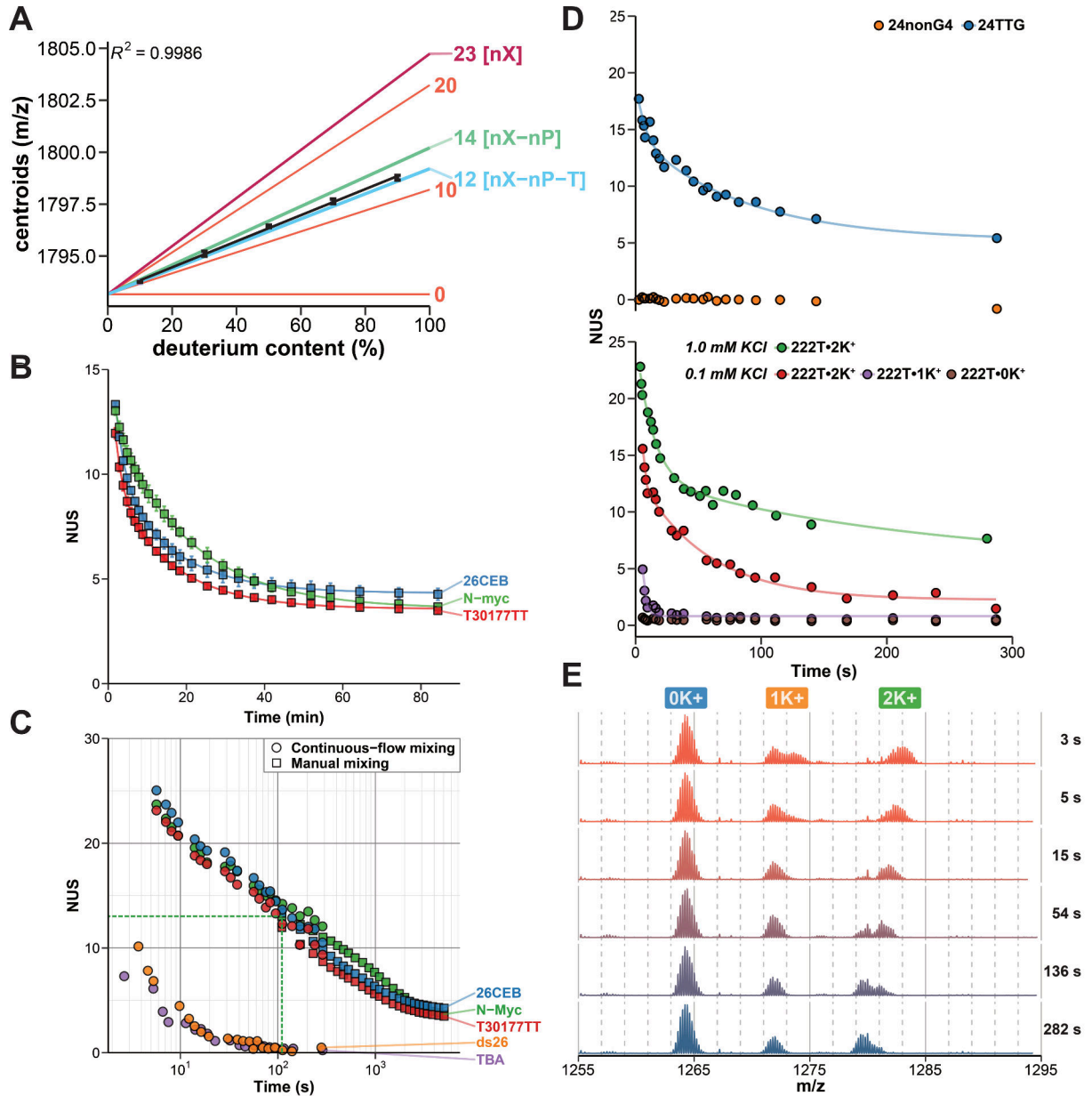


FIGURE 5.4. A. Proportionnalité entre la deutération des nucléotides et le pourcentage de deutérium en solution par mesure de ESI-MS native ( $nX$  : nombre de sites sur les bases,  $nP$  : nombre de phosphates neutralisés,  $T$  : nombre de sites des *termini* 5' et 3'). B. Mesures d'HDX/MS de trois G4s en triplicat (barres d'erreur : 1 écart-type). C. Mesure d'HDX/MS sur une large gamme de temps d'échange, en utilisant les mélanges en flux continu et manuel. D. Seuls les oligonucléotides structurés (bleu, vert, rouge, violet; G4s) sont protégés de l'échange et le degré de protection dépend de la structure. Les oligonucléotides non structurés (orange, marron) échangent dans le temps mort. E. Les cinétiques d'échanges de plusieurs complexes non covalents peuvent être mesurées simultanément grâce à la séparation en masse native. Ici l'oligonucléotide 222T est en équilibre entre les formes non structurées (0 K<sup>+</sup>), G4 à deux tétrades (1 K<sup>+</sup>), et G4 à trois tétrades (2 K<sup>+</sup>), dont les cinétiques sont visibles dans le panneau D. L'échange a lieu dans la direction D → H, donc les distributions isotopiques se déplacent vers les bas  $m/z$ , et donc de droite à gauche. Figure adaptée avec permission (152).



C'est dans ce cadre de ce projet financé par l'ANR que j'ai fait une demande d'ADT, qui m'a été accordée pour la direction de la thèse de **Matthieu Ranz**. Je présente brièvement ci-dessous nos avancées sur ce projet.

### 5.2.1.2 Application à la biophysique des G4s

Nous avons appliqué nos méthodes HDX/masse native à un large panel d'oligonucléotides se structurant en G4 de conformations et stabilités diverses (6). Cela nous a permis de confirmer les résultats de nos premiers travaux et d'établir plusieurs points importants :

1. Une large gamme de temps de deutération est nécessaire pour observer l'échange de tous les sites structurés des G4s.
2. Le couplage à la masse native permet de mesurer simultanément les cinétiques d'échange de plusieurs espèces dans un mélange conformationnel.
3. Le nombre de sites protégés reflète, entre autres, le nombre de tétrades d'un G4. Nous avons mis à disposition un set de données de référence pour déterminer le nombre de tétrades à partir d'une cinétique HDX (Figure 5.5A) (179).
4. La vitesse d'échange diminue lorsque la stabilité ( $T_m$ ,  $\Delta G^0$ ) d'un G4 augmente, mais elle dépend également de la dynamique conformationnelle et de l'éventuelle présence de motifs non canoniques (Figures 5.5B et 5.7A-C). Contrairement aux stabilités évaluées par *melting*, l'HDX/MS a l'avantage d'être une mesure isotherme, sans déplacement de l'équilibre conformationnel dû aux changements de température.

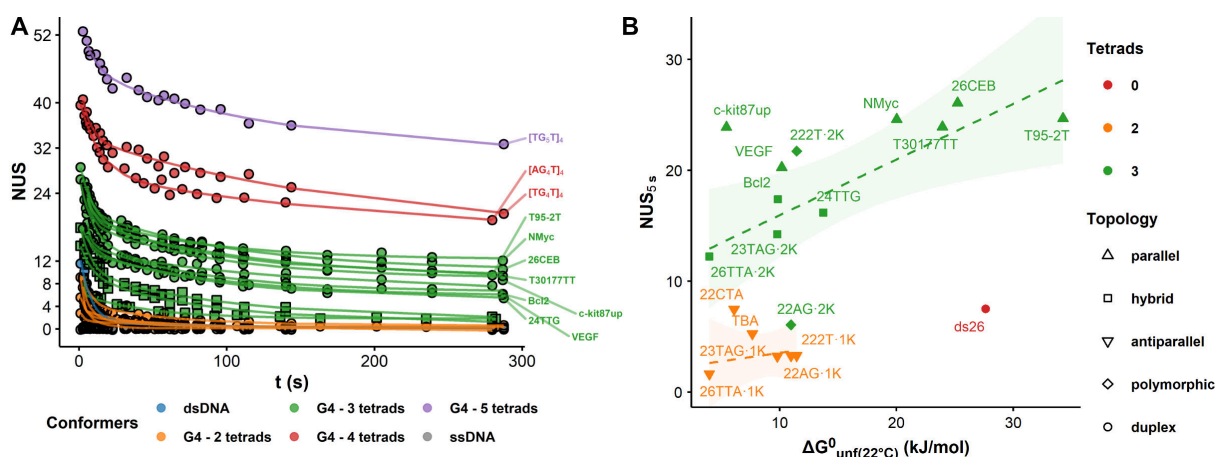


FIGURE 5.5. Les cinétiques d'échange HDX dépendent de la structure, du nombre de tétrades, et de la stabilité du G4. A. Cinétiques d'échange (NUS = nombre de sites non échangés) d'oligonucléotides G4, duplex et déstructurés (ssDNA). B. Relation entre le nombre de sites non échangés après 5 secondes et l'énergie libre de déstructuration. Figure reproduite avec permission (6).

Le mécanisme de l'échange dépend donc de la stabilité, du nombre de tétrades et de la topologie des G4s. Nous avons identifiés deux cas limites (Figure 5.6) :

1. Les G4s stables échangent rarement, et un site à la fois, *via* des fluctuations locales ('EX2'). Pour les duplexes, on parlerait de respiration, comme définit par von Hippel il y a plus de soixante

ans (180, 181). Cela se traduit par un déplacement, lent, de la distribution isotopique sur l'axe des  $m/z$ .

2. Les G4s peu stables, particulièrement ceux à deux tétrades (mais pas uniquement), peuvent aussi échanger plusieurs sites coopérativement *via* déstructuration partielle ou totale ('EX1', Figures 5.7 et 5.8B). Cela se traduit par la formation d'une seconde distribution isotopique, plus échangée que la première, et donc d'une distribution bimodale. L'intensité de la distribution originale diminue, et celle de la seconde augmente d'autant au cours de la réaction. Il est probable que la topologie du G4 influence la nature des états partiellement déstructurés. Surtout, ces états peuvent être détectés par HDX même si cette population est invisible par d'autres techniques car très peu peuplée et très transitoire. Ces cinétiques EX1 peuvent être exploitées, par déconvolution des distributions isotopiques bimodales qu'elles produisent, afin de déterminer les vitesses de déstructuration et le nombre de sites protégés des états déstructurés, ce qui permet de mieux caractériser les paysages énergétiques des G4s (Figure 5.8).

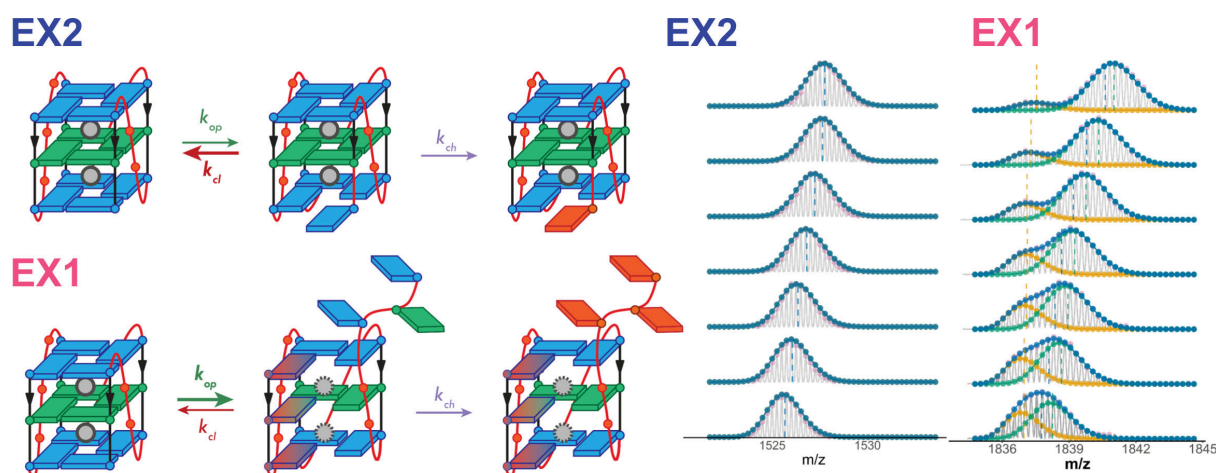


FIGURE 5.6. Schéma de principe des régimes d'échange limites EX2 et EX1 (gauche) et distributions isotopiques caractéristiques correspondantes, pendant l'échange (de haut en bas pour un échange D vers H).

On notera qu'une même espèce peut échanger par EX2 et EX1, ce qui produit des distributions bimodales avec changements concertés d'intensité (EX1) et déplacement de ces deux distributions sur l'axe des  $m/z$  (EX2). Enfin, la présence d'autres conformères de même masse mais de cinétiques d'échanges distinctes ajoute des distributions supplémentaires. Un exemple de distribution trimodale est donnée dans la Figure 5.11. Nous avons mis en place des outils logiciels pour traiter ce type de données complexes (Section 5.4.3), mais cela peut être insuffisant. Nous avons donc ajouté une dimension de séparation supplémentaire, décrite ci-dessous.

### 5.2.1.3 Couplage à la mobilité ionique

Les travaux ci-avant se heurtent à une limite : l'incapacité de séparer les conformères de même masse... par masse. Nous développons donc actuellement le couplage de l'HDX à la spectrométrie de mobilité ionique (IMS). Il s'agit d'une partie importante des travaux de thèse de de **Matthieu Ranz**. **Romane Guisiano** a également participé à ce projet comme stagiaire de M2, et est désormais en thèse sous la direction de Ralf Jockers (Université Paris Cité, Institut Cochin, UM 3, UMR 8104, UMR-S 1016).

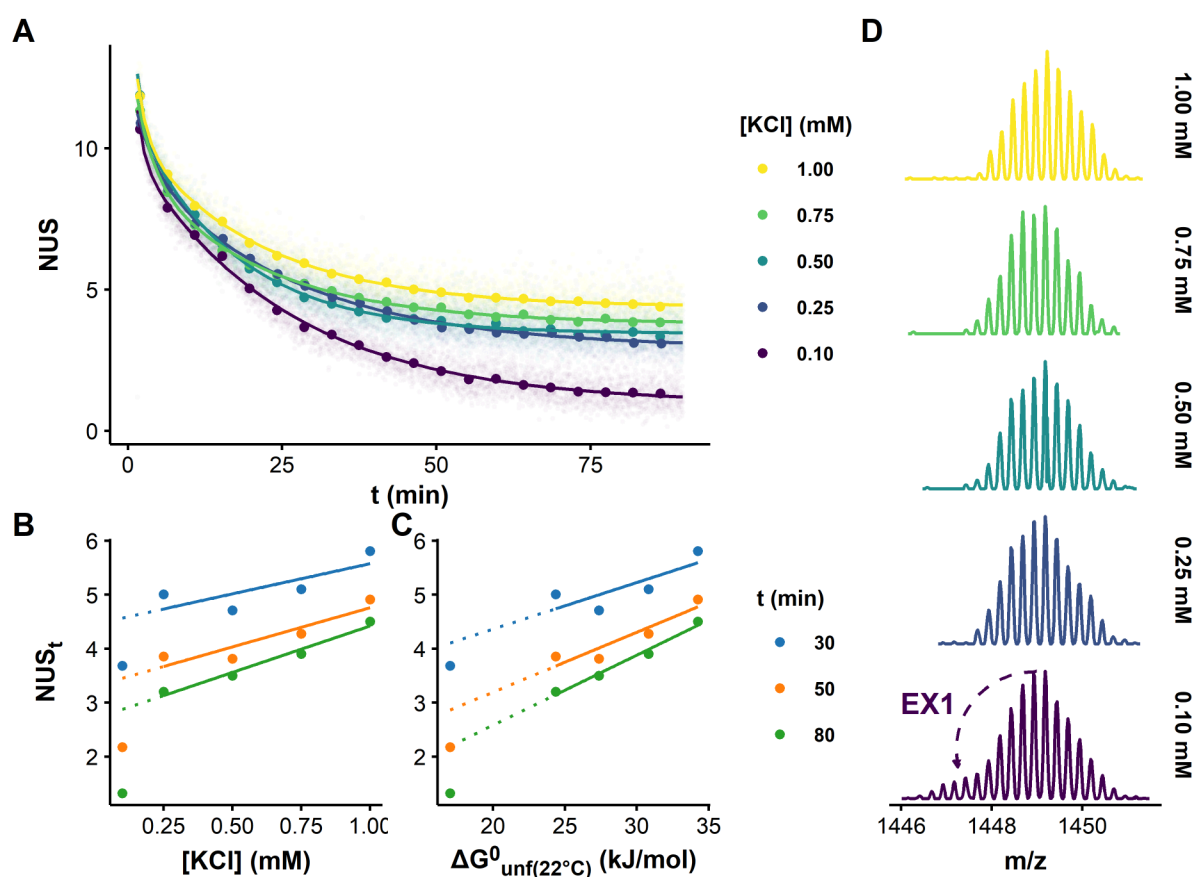


FIGURE 5.7. A. Cinétique HDX/native MS de l'aptamère anti-HIV1 T95-2T, formant un G4 parallèle en présence de différentes concentrations de KCl. B et C. Le nombre de sites protégés après 30, 50 ou 80 minutes d'échange augmente avec la concentration en KCl et avec la stabilité du G4 (qui augmente elle-même avec la concentration en KCl). Cette relation est linéaire pour un échange EX2. Lorsque l'échange se fait partiellement par EX1, la relation n'est plus linéaire. Ici, le G4 a été déstabilisé en réduisant la concentration de KCl à 0.10 mM. D. Spectres de masse de T95-2T après 6 minutes d'échange montrant la présence d'une distribution isotopique bimodale à 0.10 mM en KCl, caractéristique du mécanisme EX1. La population ayant échangé plusieurs sites coopérativement apparaît à gauche de la distribution majoritaire car elle est significativement moins deutérée (échange  $D \rightarrow H$ ). Figure reproduite avec permission (6).

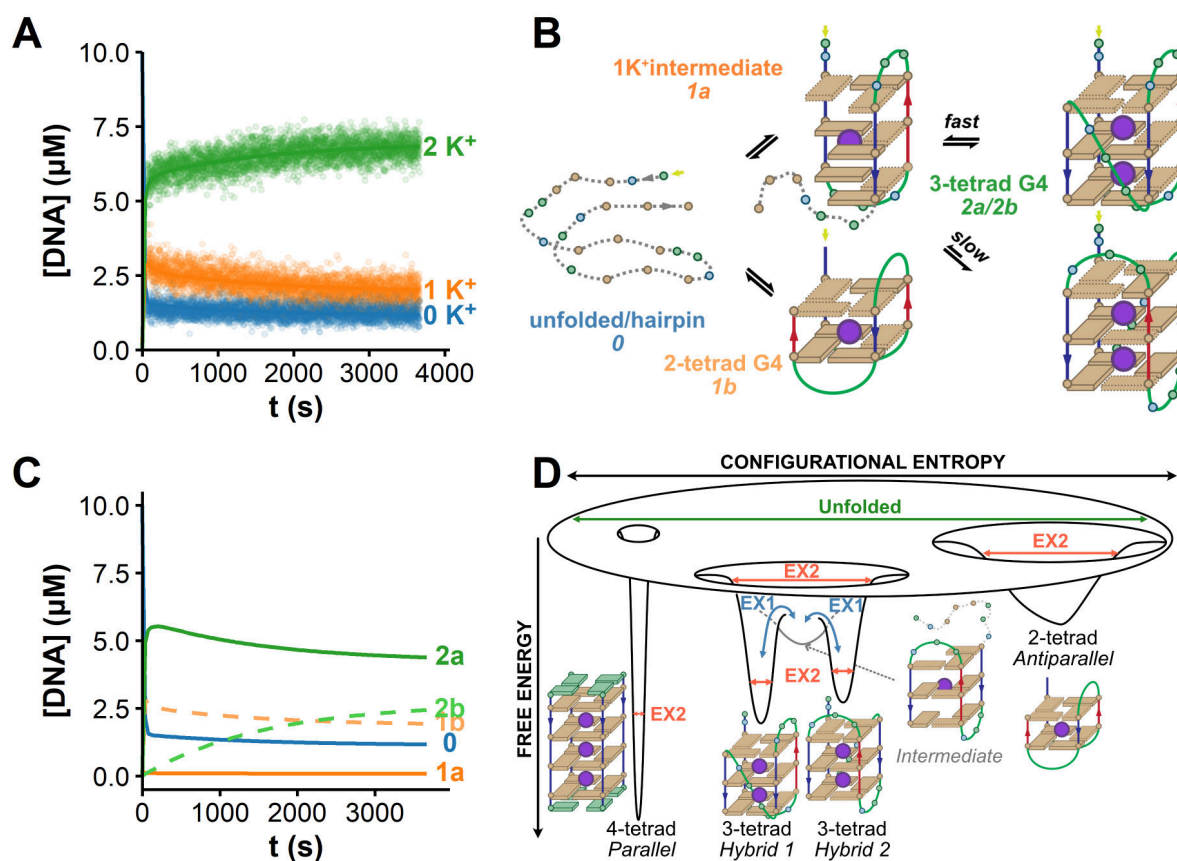


FIGURE 5.8. Structuration et déstructuration du G4 télomérique humain 23TAG. A. Cinétique de coordination du K<sup>+</sup> (1 mM) et de 23TAG (10 μM) dans 100 mM de TMAA (pH 7.0). Les lignes sont le résultat de la modélisation dynamique utilisant le mécanisme de structuration du panneau B. C. Simulation de la concentration des ensembles individuels à partir des constantes de vitesse optimisées. D. Relation entre le paysage énergétique des G4s et le mécanisme d'échange. L'augmentation du nombre de tétrades donne des entonnoirs plus profonds et plus étroits. Plus l'entonnoir est étroit, plus l'échange EX2 est lent. L'échange EX1 nécessite que l'oligonucléotide visite d'autres états qui peuvent être des intermédiaires de structuration (par exemple, G-triplex) entre différentes topologies. Figure reproduite avec permission (6).

Il existe plusieurs façon de réaliser une séparation IMS. Ici, on utilise un spectromètre de masse équipé d'un tube de dérive (Figure 5.9A). Brièvement, les ions d'oligonucléotides générés dans la source du spectromètre de masse sont introduits par paquet dans le tube de dérive contenant un gaz (généralement de l'azote, mais ici de l'hélium). Lorsque les ions se déplacent dans le tube, par application d'un champ électrique, ils entrent en collision avec le gaz, ce qui ralentit leur mouvement en fonction de leur section efficace de collision (schématiquement, les plus gros ions sont davantage ralentis). Différents conformères peuvent donc être séparés en fonction de leur mobilité ionique, avant d'atteindre le détecteur du spectromètre de masse (182). Cela vient s'ajouter à la séparation en masse, où l'on prend toujours soin de conserver les interactions non-covalentes intacts. Ainsi, dans la Figure 5.9B, les conformères sont séparés par masse sur l'axe des abscisses et par IMS sur l'axe des ordonnées. Les cinétiques d'échange de la Figure 5.9C ont été déterminées par sélection en masse de l'espèce liant 2 K<sup>+</sup> (MK<sub>2</sub>, orange), pour laquelle on a séparé deux conformères par IMS (vert et violet).

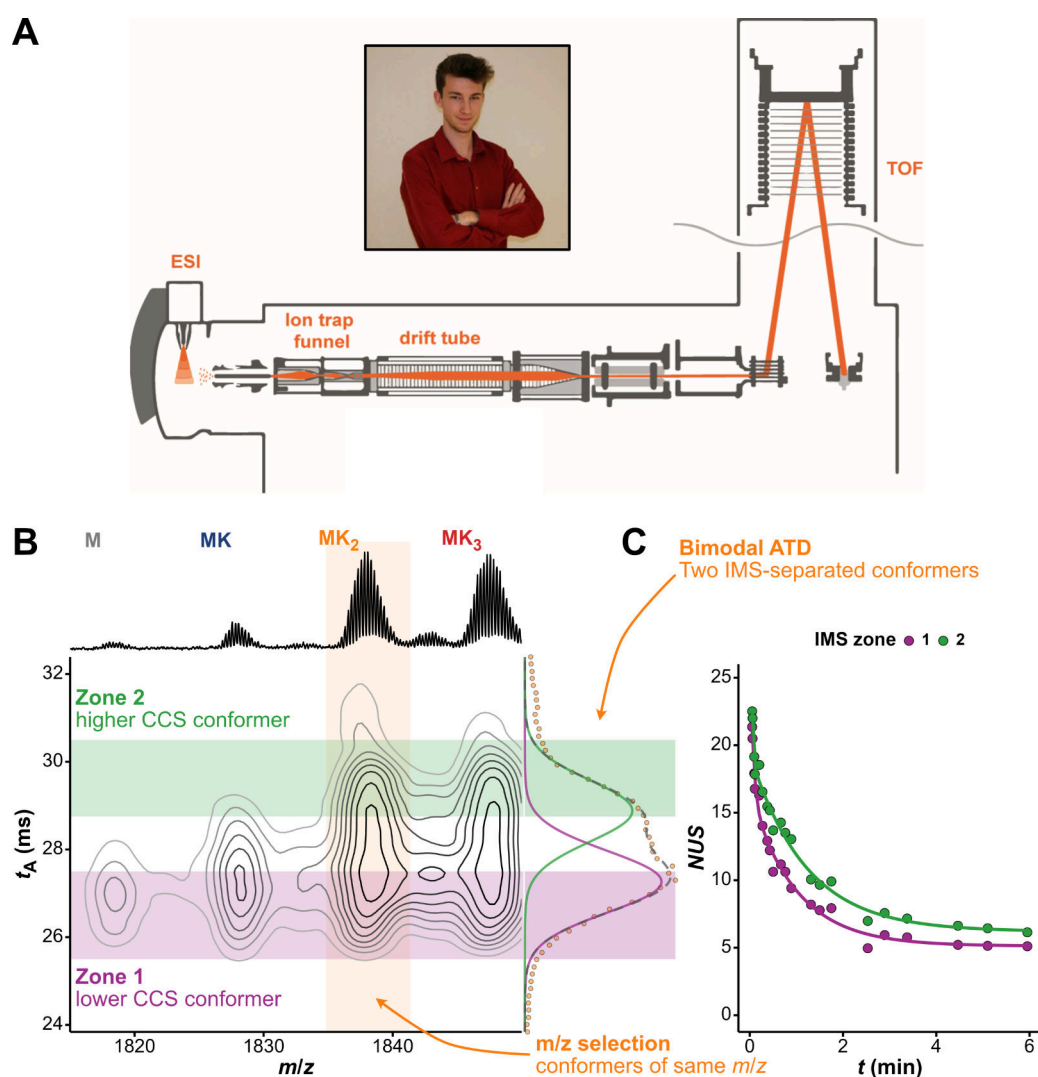


FIGURE 5.9. Principe de l'HDX/IMS-MS native. A. Schéma d'un spectromètre de masse Agilent ESI-Q-TOF équipé d'un tube de dérive pour la séparation IMS. Encart : Matthieu Ranz. B. *Contour plot* construit à partir des données HDX/IMS-MS native de 23TAG en présence d'1 mM de KCl. C. Cinétiques d'échange isotopique apparentes pour deux conformères séparés en IMS, formés par 23TAG-2K<sup>+</sup>.

Le couplage à l'HDX a nécessité d'optimiser les paramètres expérimentaux, notamment au niveau de la source et du fragmenteur du spectromètre de masse, pour obtenir un compromis satisfaisant entre

rapport signal/bruit et résolution de ces conformères par leur temps d'arrivée ( $t_A$ ).

Les résultats de l'étude d'un panel de G4s par HDX/IMS-MS native peuvent être classés en trois catégories décrites ci-dessous.

#### 5.2.1.3.1 Cas 1 : Plusieurs conformères en solution, de même masse, avec des CCS et cinétiques d'échange distinctes

Dans ce cas de figure, les conformères ne sont pas séparables par masse, mais le sont par IMS, ce qui permet d'accéder à leur cinétique d'échange propre. Nos résultats actuels montrent que nous sommes capables de résoudre des systèmes complexes, comme par exemple celui de la séquence télomérique humaine 23TAG (Figure 5.9B,C), pour laquelle nous avons discriminé les topologies hybrides 1 et 2 visibles dans la Figure 5.8B. Nous avons observé un échange plus rapide de la topologie hybride 2, qui ne s'explique pas par une différence de stabilité entre ces topologies, mais par une dynamique de déstructuration plus rapide. En effet, la déconvolution de leurs distributions isotopiques pendant l'échange, désormais résolues en IMS, permet de montrer que l'hybride 2 se déstructure plus vite que l'hybride 1 ( $0.57 \text{ min}^{-1}$  vs.  $1.76 \text{ min}^{-1}$ , respectivement, en EX1), alors que leur échange par fluctuation locale (EX2) est similaire.

Cette approche a été utilisée pour un système encore plus complexe, le variant télomérique 5YFY, qui forme au moins quatre conformères dans nos conditions, dont trois ont la même masse (Figure 5.10). À l'instar de 23TAG, ces trois conformères de même masse échangent à des vitesses distinctes à cause de leurs tendances respectives à échanger par EX1.

#### 5.2.1.3.2 Cas 2 : Plusieurs conformères en solution, de masses et CCS distinctes, et de même CCS et cinétiques d'échange

Il s'agit ici de filtrer les adduits non spécifiques, fréquents en ESI-MS native. Un seul exemple est déjà publié, où l'on s'intéresse à des complexes non-covalents de protéines (183). Dans le cas des G4s, un oligonucléotide formant des conformères à 2 tétrades ("2T") et 3 tétrades ("3T") coordonne spécifiquement 1 et 2  $K^+$ , respectivement et ont donc des masses distinctes (MK et MK2, respectivement; Figure 5.11A,B). Cependant, 2T peut coordonner un potassium supplémentaire, non spécifiquement, et ainsi avoir la même masse que 3T. L'IMS permet de les séparer en phase gazeuse, et ainsi d'accéder à la cinétique d'échange "pure" de 3T (Figure 5.11C). De même pour 2T, on peut filtrer la contribution des brins non structurés ("U"). L'exemple de la Figure 5.11 illustre cette approche pour l'oligonucléotide 222T, qui forme des conformères U, 2T et 3T. Les distributions isotopiques observées pour MK2 sont particulièrement complexes, puisque 3T échange par EX2 et EX1 et que MK2 contient également un contaminant à 2 tétrades. On obtient donc une distribution trimodale, dont deux populations isotopiques sont corrélées (EX1 de 3T), et une indépendante (contaminant non spécifique de 2T) (Figure 5.11D, zone 2). La séparation IMS permet d'accéder aux contributions de 3T uniquement, et en particulier à son échange significatif par EX1 (Figure 5.11D, zone 3).



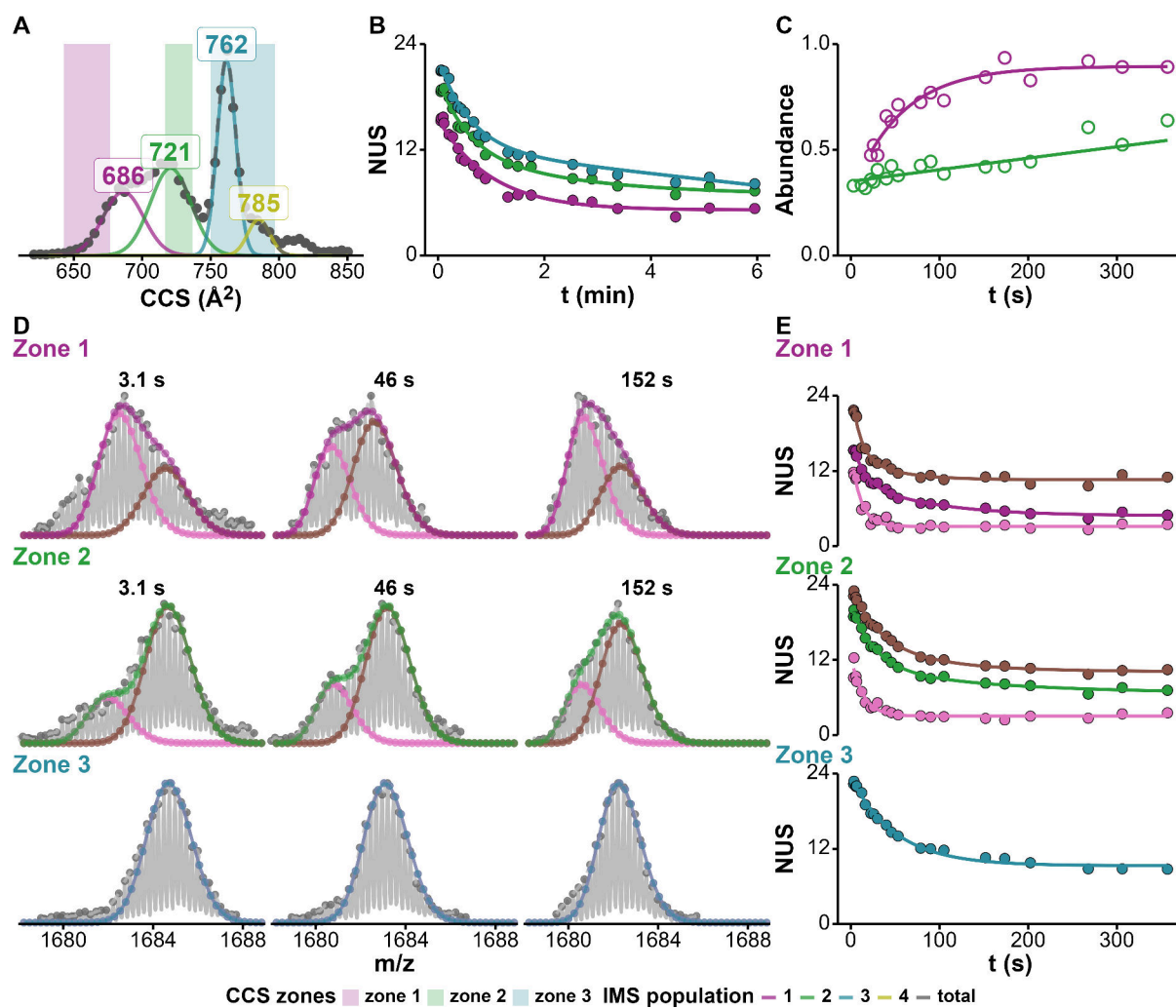


FIGURE 5.10. Détection et caractérisation de conformères mineurs de 5YEY par HDX/IMS-MS native. A. Distributions de CCS de 5YEY-2K<sup>+</sup> en présence de 1 mM de KCl. B. Cinétiques d'échange apparentes de trois zones IMS, définies dans le panneau A. Les distributions isotopiques de ces trois zones ont été déconvoluées (D), ce qui permet de quantifier leur échange par EX2 (E) et EX1 (C; zones 1 et 2). On note ici que le conformère de la zone 3 échange purement par EX2, plus lentement que ceux des zones 1 et 2, qui échangent significativement par EX1.

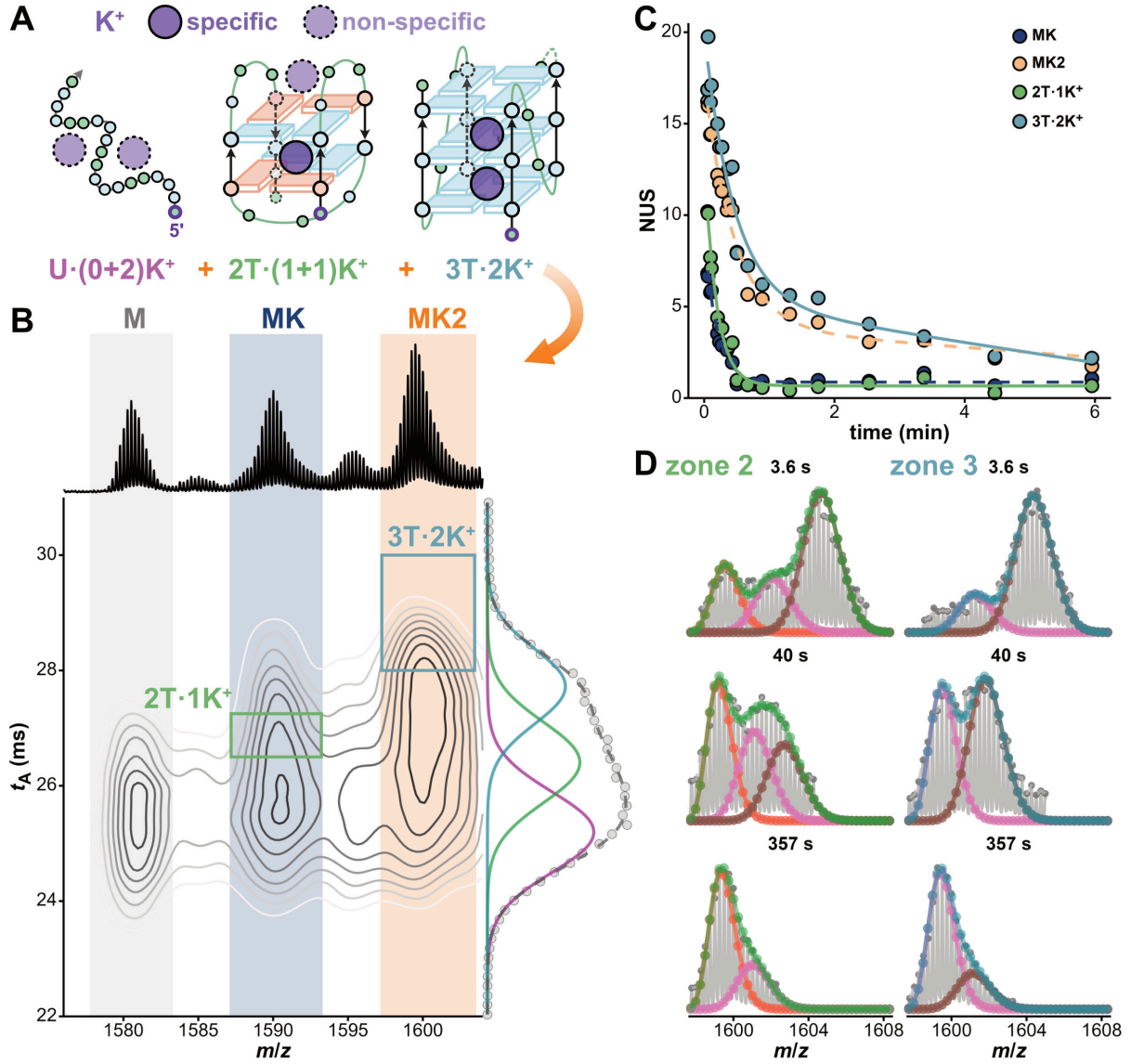


FIGURE 5.11. Filtrage des adduits non spécifiques de 222T par IMS : A. La présence d'adduits  $K^+$  non spécifiques produit des espèces de différentes conformations (U : non structuré, 2T : G4s à 2 tétrades, 3T : G4s à 3 tétrades) mais de même  $m/z$  : ici, 2T lié à un cation spécifique et un non spécifique ( $2T \cdot (1+1)K^+$ ) et U lié à deux cations non spécifiques ( $U \cdot (0+2)K^+$ ) ont le même  $m/z$  que  $3T \cdot 2K^+$  et sont donc mélangés dans MK2. Il en va de même pour MK qui contient  $2T \cdot 1K^+$  et  $U \cdot (0+1)K^+$ . Notons que les structures présentées ici pour 222T ne sont pas connues et qu'il s'agit probablement de mélanges de conformères. B. *Contour plot* HDX/IMS-MS de 222T dans 0.1 mM KCl (en haut : spectre de masse natif, à droite : distribution des temps d'arrivée). C. L'extraction de régions IMS-MS spécifiques (rectangles colorés dans le panneau B) fournit les cinétiques d'échange apparentes des espèces  $3T \cdot 2K^+$  et  $2T \cdot 1K^+$  sans les contributions non spécifiques initialement présentes dans MK2 et MK, respectivement. D. La déconvolution de la distribution isotopique multimodale pour des temps d'échange sélectionnés met en évidence la présence d'un contaminant dans la zone 2 (vert; distribution trimodale) de MK2 qui est filtré dans la zone 3 (bleu clair; distribution bimodale).

### 5.2.1.3.3 Cas 3 : Même masses et cinétiques d'échanges mais CCS distinctes

Il y a en fait ici deux possibilités :

- Les pics distincts en mobilité ionique sont la conséquence de réarrangements entre la source (où la charge de l'analyte est acquise) et le tube de mobilité (Figure 5.9A). La cinétique d'échange en solution (avant l'ESI) n'est donc pas altérée. Dans ce cas de figure, l'HDX renseigne sur les mécanismes de réarrangement en phase gazeuse et permet de mieux comprendre certains résultats d'IMS-MS.
- Il y a des conformations distinctes en solution qui ont le même comportement HDX à l'échelle de temps où l'expérience est conduite. C'est sans doute le cas de conformères ayant le même cœur G4 (même nombre de tétrades et topologie) mais des structurations et dynamiques de boucles différentes. Notre approche méthodologique permet de sonder l'échange de la seconde à plusieurs jours, ce qui permet de voir les sites des tétrades échanger, mais pas ceux des boucles. Ici c'est l'IMS qui permet de mieux interpréter les résultats HDX, et en particulier permet de détecter des conformations alternatives qui n'impactent pas les cinétiques *apparentes* d'échange.

### 5.2.1.3.4 Conclusion

Cette approche combinant HDX en solution et IMS-MS native de biomolécules intactes est peu utilisée. Nous montrons ici qu'elle fonctionne avec des G4s, et que chaque technique, HDX et IMS, permet de mieux comprendre et interpréter les résultats de l'autre. Cette approche est potentiellement utile pour d'autres macromolécules de relativement bas poids moléculaires, pour lesquelles plusieurs conformères pourraient être séparés par IMS (assemblages supramoléculaires, peptidomimétiques, et assemblages protéiques conçus *de novo* (184)). Un exemple d'utilisation d'HDX/IMS-MS native sur des protéines est donnée dans la Section 5.2.2.

### 5.2.1.4 Etude de complexes G4/petites molécules

L'étude de l'interaction entre ligands et G4s à relativement haut débit (plusieurs complexes par jour) repose généralement sur des méthodes spectroscopiques : UV-*melting*, FRET-*melting*, dichroïsme circulaire, G4-FID (voir Section 1.4) (185). Lorsque le système s'y prête, la structure des complexes formés peut être déterminée par RMN ou cristallographie. Ces techniques sont parfois limitées par le polymorphisme des analytes (ADN libre vs complexé, différentes topologies, différentes stœchiométries, etc.) et ne donne souvent accès qu'à un « instantané ». Le couplage entre HDX en solution, qui informe sur la dynamique de ces espèces en solutions, et de la spectrométrie de masse native, qui permet de séparer ces espèces par masse et structure, est prometteur pour aller au-delà de ces instantanés (6).

Il s'agit d'une autre partie importante des travaux de thèse de **Matthieu Ranz**, assisté de **Clarisse Fourel** (Étudiante ingénieure, ENSMAC, Bordeaux) au départ de ce projet. Clarisse est maintenant en thèse sous la direction d'Ewen Lescop et Laurent Catoire (ICSN, Université Paris Saclay).

Nous avons sélectionné plusieurs ligands avec divers modes d'interaction qui induisent, ou non, des altérations structurales de l'ADN G4. Cela permettra de déterminer quelles informations peuvent en

effet être collectées par HDX/MS. Nous appliquerons ensuite cette approche à des systèmes complexes dont le mode d'interaction n'est pas encore élucidé.

Nos premiers résultats montrent que les ligands interagissant par  $\pi$ -stacking sur les tétrades externes induisent un ralentissement de la cinétique d'échange dans les premières secondes/minutes par rapport au G4 libre, puis convergent vers des valeurs similaires (Figure 5.12; T95TT/**PhenDC3**). Cela s'explique par une protection accrue des tétrades externes, qui échangent normalement dans les premières secondes à minutes, mais pas des internes.

En revanche, lorsque l'échange induit un changement de conformation du G4, la cinétique d'échange est altérée sur l'ensemble des temps d'incubation, traduisant un changement de protection à l'échange de l'ensemble des sites. Dans le cas de 23TAG, l'intercalation de **PhenDC3** induit un changement conformationnel qui augmente la protection de nombreux sites (Figure 5.12; 23TAG/**PhenDC3**). L'intercalation induit également la "perte" d'un potassium, ce qui peut être directement quantifié sur les spectres de masse native acquis pendant les expériences HDX. À l'inverse, l'interaction de **PDS** avec 5YEE mène à une déprotection globale du G4, qui peut être le résultat d'un déplacement de l'équilibre vers un G4 à deux tétrades (Figure 5.12; 5YEE/**PDS**).

D'autre part, l'analyse des distributions isotopiques a permis de montrer que le mécanisme d'échange EX1 est souvent inhibé en présence de ligand (cas de **PhenDC3**/23TAG), mais pas systématiquement (cas de **PhenDC3**/TBA), ce qui suggère que certains ligands se coordonnent de façon très dynamique.

Un point faible de cette approche est la dilution nécessaire au déclenchement de l'échange, ici d'un facteur 10. Dans le cas de complexes avec des  $K_d$  submicromolaires, cela n'a virtuellement aucune incidence. En revanche, lorsque le  $K_d$  est dans l'ordre de grandeur des analytes (5–10  $\mu$ M), alors l'échange a lieu hors équilibre. Cela signifie qu'une partie du complexe *peut* se dissocier, et donc que les cinétiques mesurées pour l'ADN libre sont potentiellement "polluées" par des molécules d'ADN ayant partiellement échangées sous forme de complexe. Ce potentiel de contamination dépend de la cinétique de dissociation : si celle-ci est lente par rapport à la cinétique d'échange, alors l'incidence est faible. En revanche, dans le cas de dissociations rapides, la contamination peut être significative. Cela entraîne une différence apparente de cinétiques entre ADN libre et lié plus faible qu'elle ne devrait l'être. Une de mes perspectives à long terme est de mesurer les  $k_{off}$  et  $k_{on}$  des complexes étudiés en masse native avec une technique orthogonale comme la SPR ou le BLI.

Enfin, nous avons étendu notre gamme de ligands à un peptide, comme preuve de concept de cette approche pour étudier des complexes acides nucléiques / protéines. Le peptide RHAU18 interagit avec les tétrades terminales de l'aptamère G4 T95TT, sans changement conformationnel de ce dernier. Malgré cela, les cinétiques d'échange ne convergent pas comme dans les cas de  $\pi$ -stacking. Notre hypothèse est que le peptide protège partiellement la tétrade interne par interaction de l'extrémité C-terminale (en particulier K17 et Q18) dans le sillon du G4, mais il ne faut pas exclure d'éventuelles contributions de sites protégés sur le peptide lui-même. Pour cela, nous devons étudier l'échange du peptide, et à terme de la protéine entière, par HDX/MS *bottom-up* "classique", tel que décrit dans la Section 4.4.

L'HDX/MS semble donc avoir un excellent potentiel pour caractériser différents modes d'interaction et la dynamique des complexes formés. De plus, cette approche bénéficie des avantages de la spectrométrie de masse native, à savoir la détection et quantification des différentes stœchiométries en

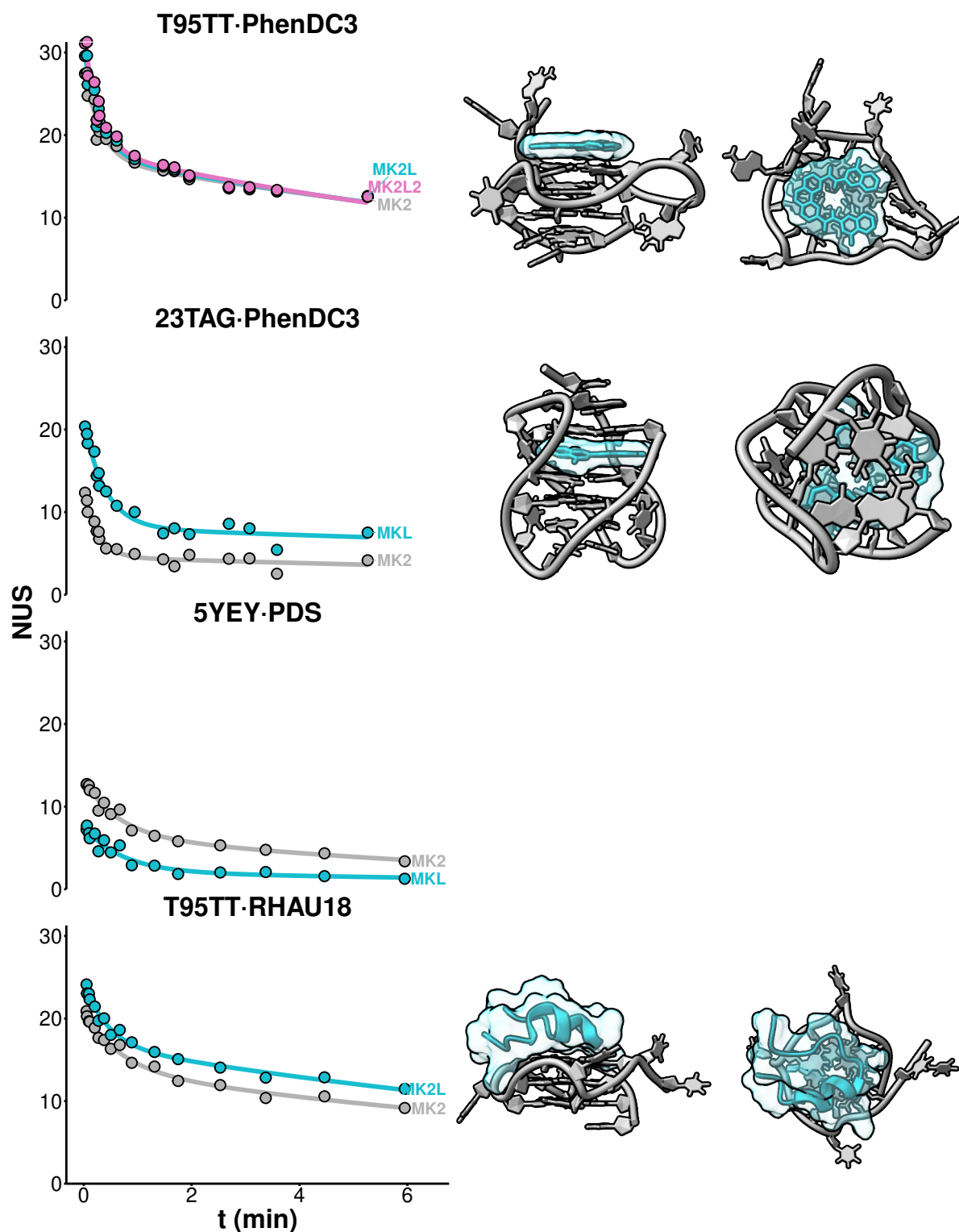


FIGURE 5.12. Comparaison des cinétiques d'échanges de complexes G4/petites molécules (**PhenDC3**, **PDS**) et G4/peptide (**RHAU18**) et des G4s libres correspondant. À droite, les structures montrent le mode d'interaction attendu pour T95TT/**PhenDC3** (PDB : 2MGN ; le G4 est c-myc dans cette structure), 23TAG/**PhenDC3** (PDB : 2MGN) et T95TT/**RHAU18** (PDB : 2N21).

oligonucléotides, cations et ligands, et la détermination des  $K_d$  pour ces espèces séparées en masse, voire en IMS. Ces travaux, feront l'objet d'une publication avec **Matthieu Ranz** comme premier auteur.

#### 5.2.1.5 Perspectives

Deux axes de travail se dégagent désormais. Le premier est méthodologique, et consiste en l'identification d'une méthode de fragmentation en phase gazeuse des analytes, sans générer de réarrangement des isotopes (*scrambling* (186–192)) qui ferait perdre l'information acquise en solution. Cette approche *top-down* permettrait l'obtention des cinétiques d'échange à l'échelle d'un petit nombre de résidus, voire même d'un résidu unique, en plus de celles mesurées à l'échelle de l'analyte intact. Ces cinétiques résolues sur la séquence permettrait de déterminer quelles bases sont impliquées dans des motifs structuraux (paires de bases, tétrades) et des interactions non covalentes avec des ligands, du cation à la protéine.

Le second axe de travail consiste à utiliser notre boîte à outils méthodologiques (résolution en temps de mélange, masse, conformation et séquence) pour étudier des complexes acides nucléiques/protéines. L'HDX/MS pourra aussi être appliqué au partenaire protéique, offrant une caractérisation complète des dynamiques d'interactions des deux partenaires. Explorer d'autres structures secondaires d'ADN et d'ARN (pas encore étudié du tout!) est également au programme, notamment en exploitant la séparation IMS.

Enfin, tangentiellement à ce projet, j'ai rejoint Carmelo Di Primo (ARNA) sur une étude fondamentale de l'utilisation du BLI pour étudier la structuration et les interactions d'oligonucléotides. À terme, j'aimerais inclure cette technique comme méthode orthogonale de mesures de cinétiques d'association et dissociation intra- et intermoléculaires.

#### 5.2.2 HDX/MS native de protéines thérapeutiques

L'échange hydrogène-deutérium ne s'est pas cantonné aux acides nucléiques dans l'équipe BALI. J'ai ainsi modestement participé à un **Projet financé par la Région Nouvelle Aquitaine**, en collaboration avec la société Merck, qui utilise l'échange hydrogène-deutérium couplé à la spectrométrie de mobilité ionique et la spectrométrie de masse native pour l'étude de protéines thérapeutiques qui présentent des conformations multiples en solution. Ce projet, innovant techniquement, a permis de déterminer les conditions optimales de production et conservation de ces protéines, identifier les conformères actifs, et participera à la protection de la propriété intellectuelle.

D'un point de vue académique, ces travaux principalement menés par une chercheuse postdoctorante, **Nina Khristenko** (Figure 5.1), nous ont permis de découvrir un nouveau mécanisme d'ionisation des protéines par électrospray en conditions natives (Figure 5.13A) (193). Durant ce projet, j'ai coencadré Nina en apportant mon expertise en HDX/MS et développement de méthodes analytiques pour la conception et la réalisation des expériences, le traitement de données, et l'interprétation des résultats.



Plus spécifiquement, nous nous sommes intéressés à la structure en phase gazeuse de protéines thérapeutiques multi-domaines, constituées de domaines globulaires reliés par des régions non structurées. Leurs distributions de charges (masse native) et de sections efficaces de collision (IMS) sont multimodales (Figure 5.13B, C), ce qui suggère l'existence de plusieurs conformations en solution suivant des mécanismes d'ionisation distincts. Nous avons confirmé la présence de plusieurs conformations en solution (et non seulement en phase gazeuse) par HDX, en reprenant l'approche en flux continu développée pour les acides nucléiques, c'est à dire en conditions natives et à l'échelle de la protéine intacte, plutôt que par approche *bottom up* (Figure 5.13D). Les états de charge les plus élevées, correspondant aux CCS de protéines les plus allongées sont théoriquement les moins protégés de l'échange, ce qui est en effet ce que nous avons observé.

Le mécanisme d'ionisation électrospray est intrigant : une fraction de la population ne suit pas le mécanisme des résidus chargés (CRM) mais ne peut pas non plus s'ioniser par pure éjection de chaîne. Nous en avons déduit qu'un mécanisme hybride est possible, dans lequel les domaines globulaires sont éjectés un à la fois d'une gouttelette mère. Les corrélations entre la charge et la surface accessible au solvant de protéines dénaturées ou intrinsèquement désordonnées sont également compatibles avec ce "mécanisme d'éjection de perles" (BEM), que nous avons proposé comme principe général de l'électrospray des biomolécules.

## 5.3 Ligands de G4s

### 5.3.1 Ligands capables de discriminer des topologies G4

Après plusieurs années sans opportunités de pouvoir jouer avec des ligands de G4, j'ai enfin eu l'opportunité de revenir à mes premières amours. Comme évoqué précédemment, une petite molécule pouvant se lier spécifiquement à un seul G4 permettrait de potentiellement réguler des processus biologiques avec peu d'effets secondaires liées aux interactions avec l'ADN duplex. Lors de ma thèse, j'ai développé une méthode permettant d'évaluer la sélectivité de ligands pour des structures G4s distinctes, mais ai dû faire le constat qu'il n'était pas encore possible de concevoir rationnellement un ligand ciblant une structure G4 particulière (Sections 1.4.1.1, 1.4.1.2.1) (101, 102). Nous avons synthétisé des molécules capables de discriminer très efficacement certaines topologies, mais sans pouvoir complètement rationaliser ces effets (Section 1.3.2) (86, 87). Je m'en suis donc remis au criblage, mais sans découvrir de design moléculaire radicalement différent (Section 1.4.1.2.2) (104).

De récentes études présentent des approches très prometteuses pour obtenir des sélectivités intra-G4 significatives (194, 195), mais beaucoup des sélectivités publiées dans la littérature se limitent à l'étude d'un nombre réduit de séquences que l'on pourrait qualifier de *cherry picking*. Un exemple récent, pris totalement au hasard, présente des ligands indo-quinolines comme sélectifs de topologies parallèles, mais seulement quatre séquences G4s (dont deux parallèles) et un contrôle duplex ont été testées (196). Dans un autre exemple récent, les auteurs présentent trois molécules qui seraient sélectives d'un seul G4 (C9orf72) avec des  $K_d$  de 2–3 mM (oui, millimolaires) (197). Troisième et dernier exemple aléatoire, où un dérivé de **TO** est présenté comme spécifique de *c-MYC* après comparaison avec trois G4s, sa faible sélectivité apparente étant cachée *Supporting Information* (198). Le **TO** est d'ailleurs à la

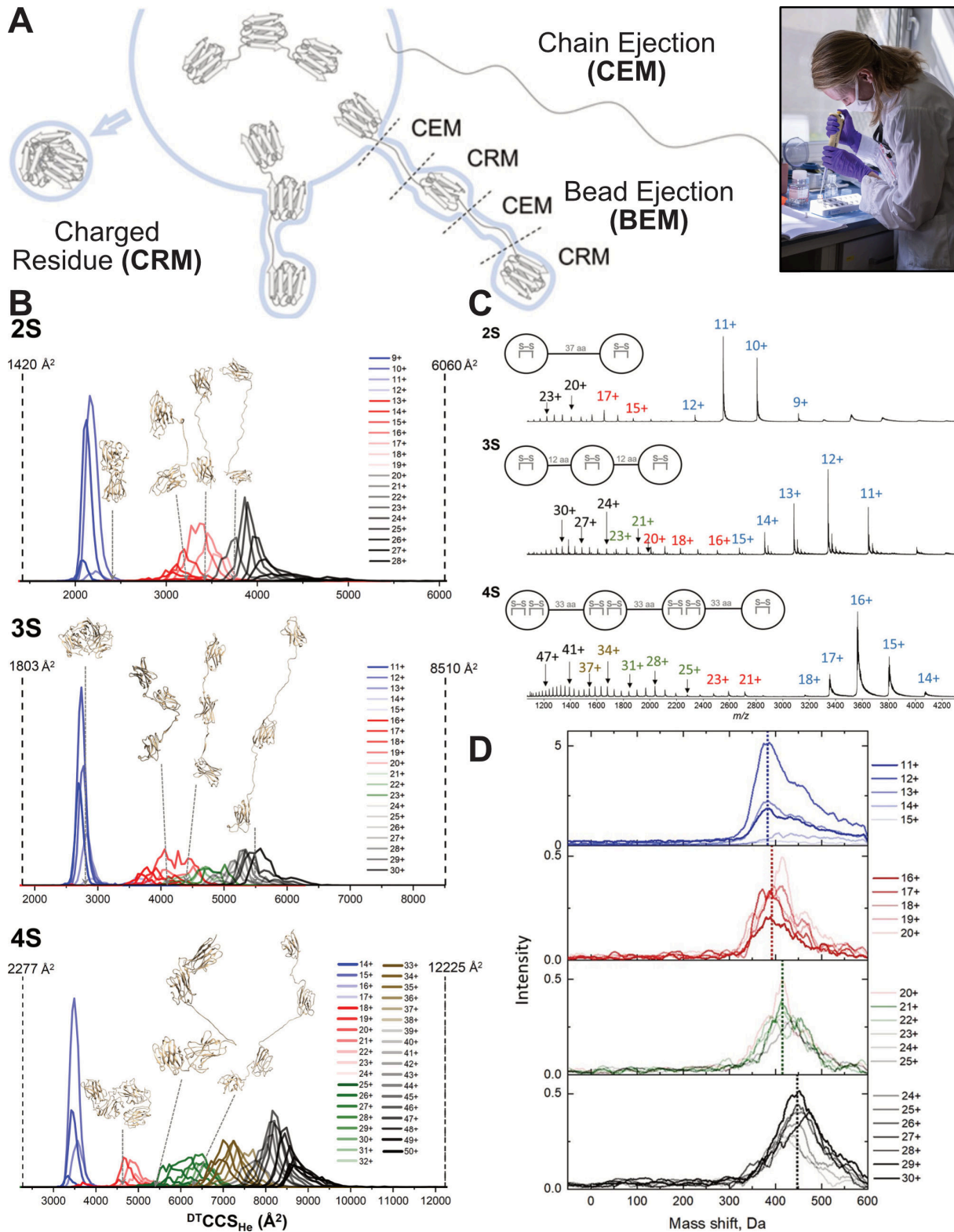


FIGURE 5.13. Ionisation en ESI natif de protéines multidomaines. A. Illustration des mécanismes coexistant proposés : le CRM produit un ensemble globulaire, chargé à un niveau proche de la limite de Rayleigh. Le CEM produit une conformation entièrement étendue. Le BEM est intermédiaire parce qu'il est un composite du CRM et du CEM : les domaines structurés peuvent être ionisés par le CRM tandis que les régions désordonnées peuvent être ionisées par le CEM. Encart : Nina Khristenko B. Distribution de CCS des protéines 2S, 3S et 4S en fonction de leurs états de charge. Les segments verticaux hachurés représentent les valeurs estimées de CCS si ces protéines étaient parfaitement globulaires (valeurs basses) ou complètement allongées (valeurs hautes). C. Spectres de masse native des protéines 2S, 3S et 4S. D. Déplacement du centroïde de masse de la protéine 3S après 53 s d'HDX en solution, en fonction des états de charge.

mode, puisque d'autres groupes ont modifié sa structure pour moduler son affinité et sa sélectivité (199, 200).

### 5.3.1.1 Ligands foldamères spécifiques des topologies parallèles

Lors de son séjour sabbatique dans l'équipe, Liliya Yatsunyk (Professeure à Swarthmore, PA, USA) a observé par ESI-MS la formation d'un complexe entre un G4 de c-myc et une molécule inconnue contaminant le spectromètre de masse. Après que Liliya a identifié le propriétaire de la molécule, l'équipe de Yann Ferrand (IECB, Bordeaux), et la nature de celle-ci, je me suis retrouvé le chanceux propriétaire du foldamère **M1**, plus tard rebaptisé **Q<sub>4</sub>** (Figure 5.14A).<sup>1</sup> Moore définit les foldamères comme "tout oligomère qui se replie dans un état conformationnel ordonné en solution, dont les structures sont stabilisées par un ensemble d'interactions non covalentes entre les unités monomères non adjacentes" (201). Il existe deux grandes classes de foldamères : les foldamères à brin unique qui se replient uniquement sur eux-mêmes (peptidomimétiques) et les foldamères à brin multiple qui peuvent également s'associer (nucléotidomimétiques). Ici, c'est de la première que j'hérite, puisque **Q<sub>4</sub>** fait partie d'une série de foldamères formés d'hétérocycles pyridine (**P**) et quinoline (**Q**) liés par des amides, ce qu'un cerveau contaminé par les ligands de G4s pourrait aussi interpréter comme des oligomères de **360A** (Figure 1.6).

De fait, ces foldamères possèdent deux caractéristiques classiques des ligands de G4 :

1. une structure basée sur des hétérocycles (Figure 5.14), bien que non condensés ici, à l'image de **TOxaPy** (Section 1.3.2)
2. Des chaînes latérales étendues et flexibles, pouvant potentiellement interagir dans les boucles et sillons des G4s. Ici chaque quinoline possède une chaîne latérale chargée positivement à pH physiologique, ce qui améliore la solubilité dans l'eau et permet de cibler les phosphates (Figure 5.14A, B).

La différence essentielle avec les ligands de G4s "classiques" réside dans le fait que ces foldamères ne sont pas plats, mais se replient en hélice grâce notamment à des des répulsions électrostatiques, des liaisons hydrogène intramoléculaires, une conjugaison et un empilement aromatique significatifs (202–204).

Avec **Benjamin Liénard** (stagiaire de licence, Bordeaux), nous avons confirmé l'interaction forte entre **Q<sub>4</sub>** et c-myc, mais également découvert que ce ligand a une préférence marquée pour certaines séquences, sans toutefois caractériser cette sélectivité complètement, ni en déterminer l'origine.

Grâce au coencadrement doctoral d'**Alexander König**, j'ai pu réexplorer ce thème de recherche pour tenter d'enfin sortir du paradigme du ligand aromatique plan condensé. Vincent Laffilé, doctorant de l'équipe de Y. Ferrand, a synthétisé une série de ces foldamères (Figure 5.14B), nous permettant de mieux caractériser les déterminants de cette sélectivité. Surtout, Alexander a souhaité tester un large panel de structures secondaires pour éviter les conclusions hâtives sur ladite sélectivité. Il a ainsi criblé par spectrométrie de masse native une large gamme de G4s et de contrôles non-G4, ce qui a

1. D'où l'intérêt de tenir à jour les *log books* des instruments.

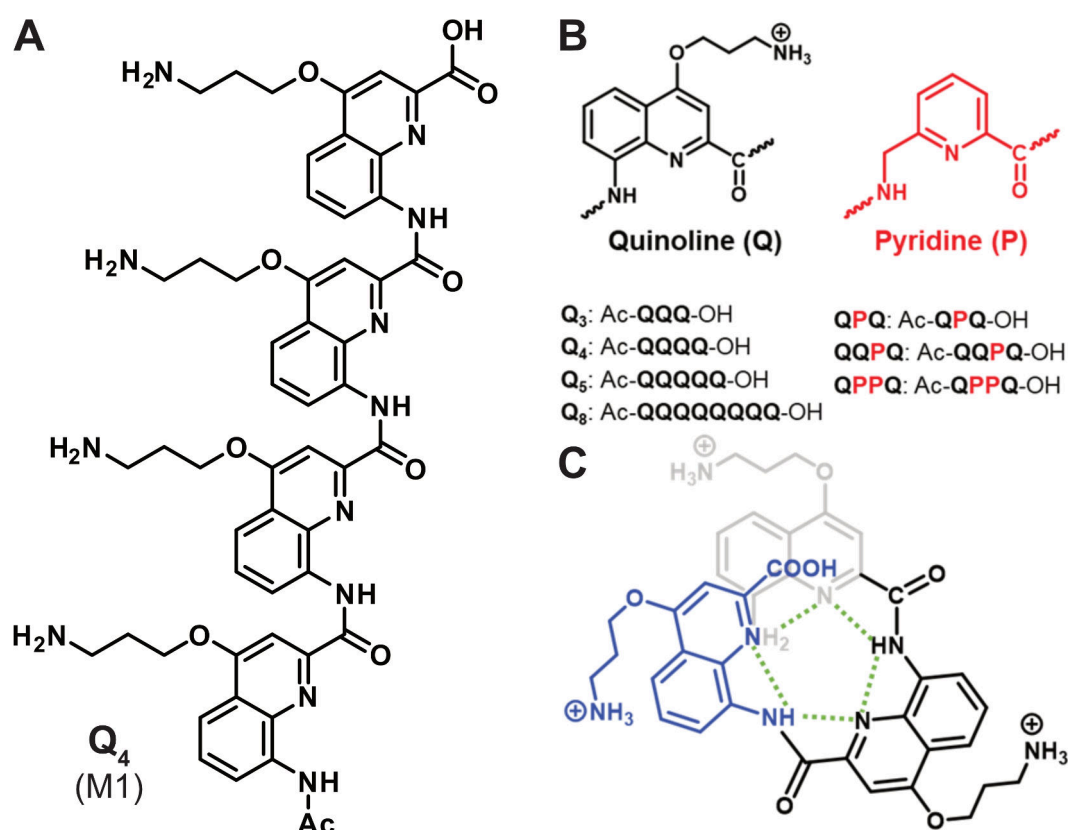


FIGURE 5.14. Série de foldamères amidoquinolines. A. Le contaminant par lequel tout a débuté, initialement nommé **M1** avant d'être rebaptisé **Q<sub>4</sub>**. B. Les différents foldamères de la série dont l'affinité et la sélectivité pour les G4s ont été évaluées. C. Vue du dessus de l'extrémité C-terminale de l'hélice formée par cette famille de foldamères.

permis de mettre en évidence la préférence marquée de **QQPQ**, **Q<sub>4</sub>**, **QPPQ** et **QPQ** pour la topologie parallèle (Figure 5.15A).

J'ai développé un script pour la quantification semi-automatisée de chacune des espèces (les différentes stœchiométries d'interaction) de ces 144 spectres de masse, et le calcul subséquent des affinités. J'ai également utilisé ce script pour quantifier les espèces des titrations entre sept foldamères et six oligonucléotides (294 spectres; Figure 5.15B, C). J'ai ensuite écrit un script générant... des scripts servant d'input pour le logiciel *DynaFit*, permettant la modélisation dynamique de ces titrations sur la base des équations chimiques de réactions (205, 206). L'exécution de ces scripts automatiquement et l'extraction des résultats ( $K_d$ , coopérativités, paramètres de *fit*) ont également été scriptées. Cette automatisation du traitement des données a permis un gain de temps notable et des gains de robustesse et répétabilité. Ces titrations ont permis la détermination de  $K_d$  de l'ordre de la centaine de nanomolaires pour les G4s parallèles, plusieurs ordres de grandeur de mieux qu'avec les autres topologies, bien que des exceptions existent (Figure 5.16).

Nous avons également observé par dichroïsme circulaire que ces ligands peuvent induire des changements de conformations de G4s non parallèles, vers la topologie parallèle, confirmant cette nette préférence structurale. Afin d'élucider l'origine de cette préférence, nous avons tenté de cristalliser certains de ces complexes avec l'aide précieuse de Stéphane Thore (ARNA, équipe PRISM). Cela nous a permis de proposer une structure pour le complexe **QQPQ**/222T (Figure 5.17), que nous avons ensuite également étudié par RMN avec Cameron Mackereth (avec un G4 légèrement différent pour obtenir des pics iminos, marqueurs de la formation des tétrades, mieux résolus; Figure 5.18).

Les deux approches montrent que l'interaction se fait par  $\pi$ -stacking sur les tétrades externes. Dans le cristal, seule une tétrade est coordonnée par le foldamère car le G4 dimérise via l'autre face. En solution, en revanche, nous avons mis en évidence que l'interaction se fait sur les deux faces, ce qui est cohérent avec les stœchiométries déterminées par spectrométrie de masse. Ce *stacking* est la norme des ligands de G4s. En revanche, cette hélice positionnée de façon coaxiale au canal ionique du G4 induit un encombrement stérique important au-dessus des tétrades. L'interaction est donc favorable lorsque ces tétrades sont relativement peu encombrées, comme dans les G4s parallèles. Dans les autres topologies, des boucles recouvrent les tétrades, rendant l'interaction stériquement peu favorable, bien que possible dans certains cas particuliers comme **QPPQ**/5YFY (Figure 5.16). Les données CD, MS et RMN supportent ici un réarrangement en un G4 hybride à trois tétrades, dont une seule présente un environnement favorable à l'interaction.

Les densités électroniques assez diffuses de l'étude cristallographique, couplées à la relative symétrie de la molécule, ne permettent pas de conclure sur le mode d'interaction. Il est théoriquement possible pour le ligand d'être positionné dans la position inversée (vis-à-vis des extrémités N et C terminales) et de pouvoir pivoter autour de l'axe de l'hélice, ce qui expliquerait les densités électroniques diffuses.

J'ai donc appliqué le *workflow* de dynamiques moléculaires que j'ai mis en place à mon arrivée dans l'équipe de Cameron, détaillé dans la Section 5.6.2, en me basant sur la structure cristalline. J'ai préalablement vérifié que la conformation hélicoïdale de **QQPQ** est la seule pertinente par calculs de mécanique quantique (Figure 5.19A), et dérivé par la même méthode des paramètres modifiés de

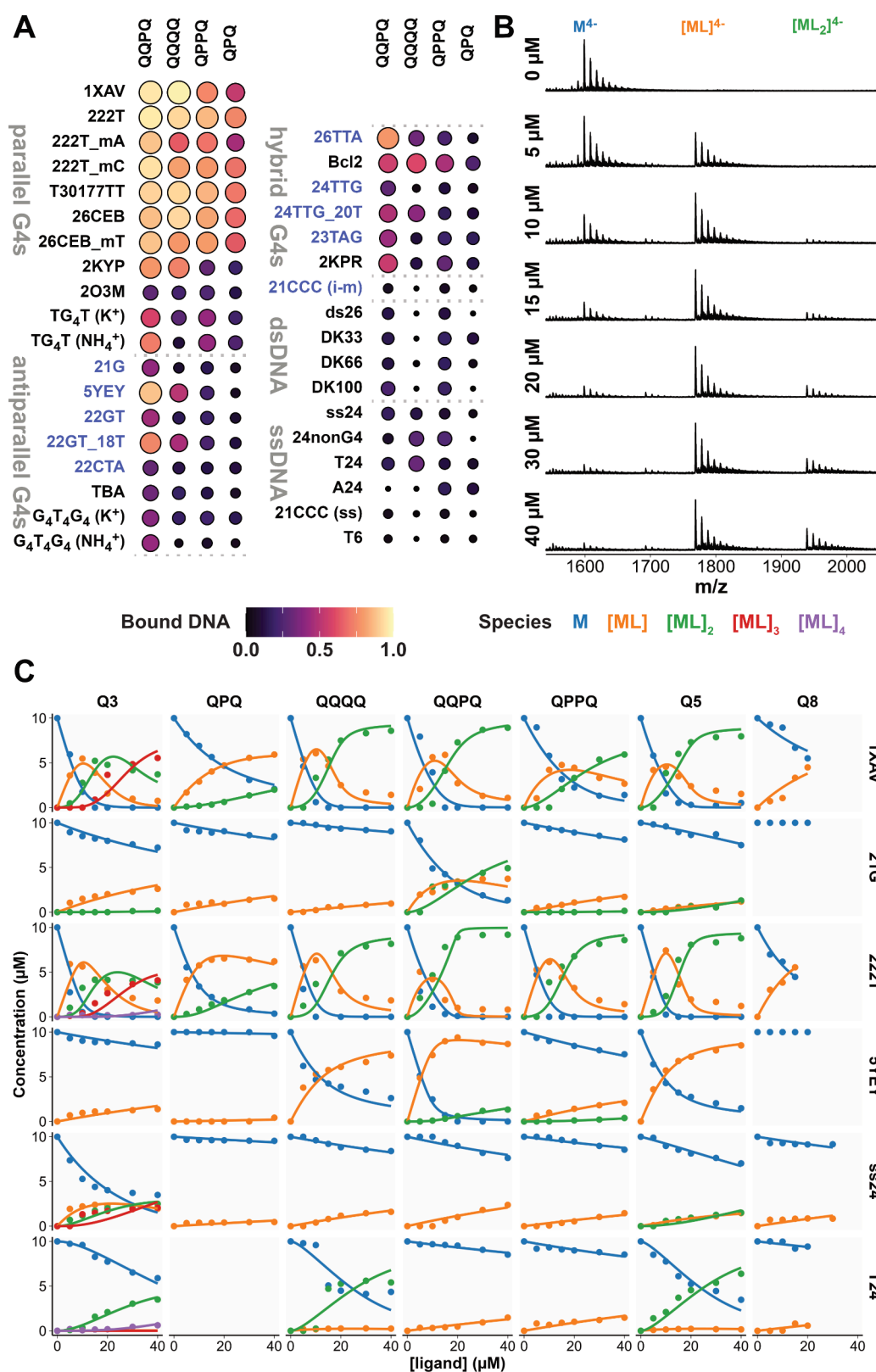


FIGURE 5.15. Utilisation de la spectrométrie de masse native pour cribler une large gamme d'oligonucléotides avec des foldamères de synthèse. A. *Heatmap* de la fraction d'ADN lié par un équivalent molaire de foldamère. 21CCC est un i-motif. B. Exemple de spectres obtenus pour la titration de 222T avec **QPQ** (M = G4 monomérique, L = ligand foldamère). C. Titrations modélisées semi-automatiquement.



Dissociation constant estimates ( $\mu\text{M}$ )									
	$K_{D1}$		$K_{D2}$		$K_{D3}$		$K_{D4}$		
	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.	
<b>2LK7</b>									
QQPQ	0.19	0.083	2.1	0.69					
<b>1XAV</b>									
Q3	0.11	0.074	1.2	0.94	4.1	2.5			
QPQ	13	0.75	110	43					
QQQQ	0.075	0.059	0.83	0.49					
QPPQ	0.41	0.29	1.9	1.2					
QPPQ	6.9	2	9.7	5.7					
Q5	0.25	0.2	0.7	0.53					
Q8	28	6.7							
<b>222T</b>									
Q3	0.1	0.1	2.8	3.1	6.1	11	64	440	
QPQ	1.8	0.25	50	13					
QQQQ	0.088	0.11	1.6	1.1					
QPPQ	0.027	0.081	0.064	0.19					
QPPQ	0.1	0.093	1.4	0.85					
Q5	0.011	0.014	0.4	0.35					
Q8	7.8	1.1							
<b>5YEY</b>									
Q3	180	28							
QPQ	1900	780							
QQQQ	6.6	1.3							
QPPQ	0.54	0.098	200	70					
QPPQ	110	23	310	760					
Q5	3.9	0.37							
<b>21G</b>									
Q3	78	19	1500	21000					
QPQ	170	26							
QQQQ	330	23							
QPPQ	8.6	2.3	14	9.2					
QPPQ	160	20	1000	4500					
Q5	210	74	47	45					
<b>SS24</b>									
Q3	12	6.2	28	87	15	63			
QPQ	580	110							
QQQQ	180	14							
QPPQ	150	21							
QPPQ	270	30							
Q5	130	42	44	41					
Q8	210	45							
<b>T24</b>									
Q3	3400	8200	0.36	0.91	320000	1.8e+08	0.0074	4.2	
QQQQ	350	3400	0.53	5.5					
QPPQ	270	24							
QPPQ	190	25							
Q5	180	490	1	3					
Q8	290	87							

FIGURE 5.16. Constantes de dissociations des réactions successives de complexation des foldamères. Les constantes submicromolaires sont colorées en bleu. Bien que les foldamères ont généralement des affinités plusieurs ordres de grandeurs meilleurs avec les G4s parallèles (1XAV, 222T) qu'avec les hybrides/antiparallèles (5YEY, 21G) ou les simples brins déstructurés (ss24, T24), des exceptions existent. En particulier, QQPQ se lie très bien à 5YEY, et Q<sub>3</sub> et Q<sub>4</sub> se lient de façon coopérative - 2 par 2! - à T24.

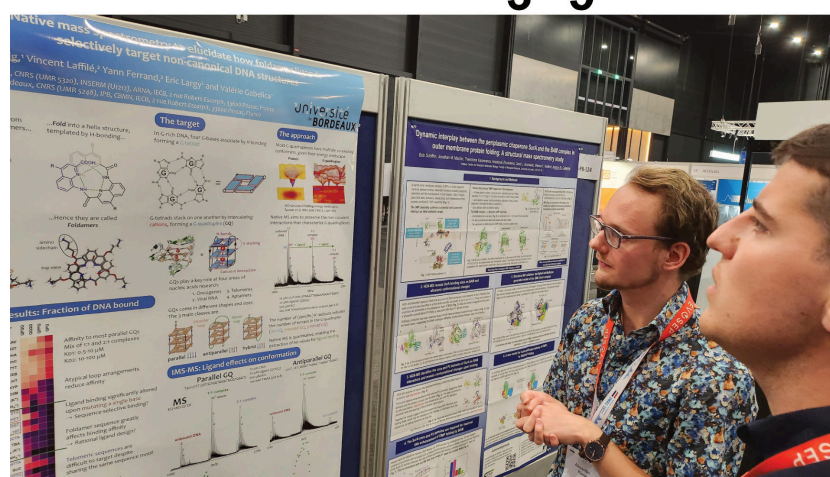
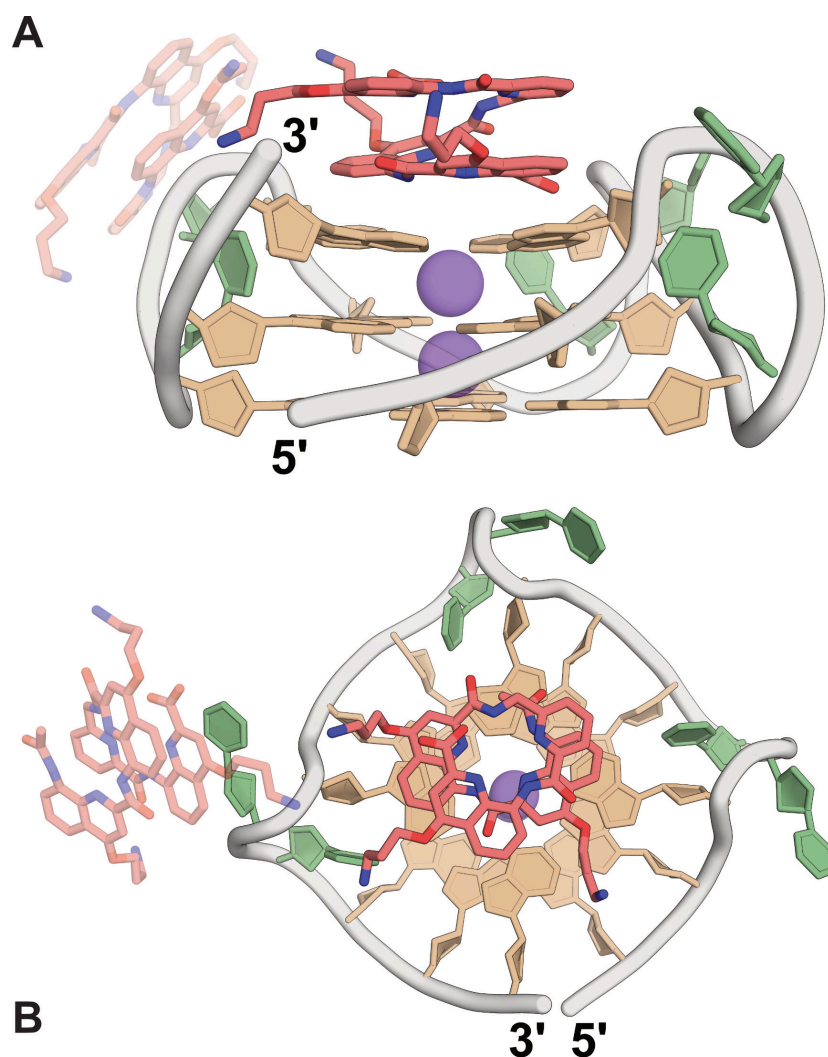


FIGURE 5.17. A. Structure cristalline du complexe **QPQQ/222T**. On notera la présence d'un second site d'interaction du ligand qui permet des contacts dans le cristal et n'existe pas en solution. B. Alexander König présente ses résultats à l'*International Mass Spectrometry Conference (IMSC)*, 27 août 2022, Maastricht (Pays-Bas).

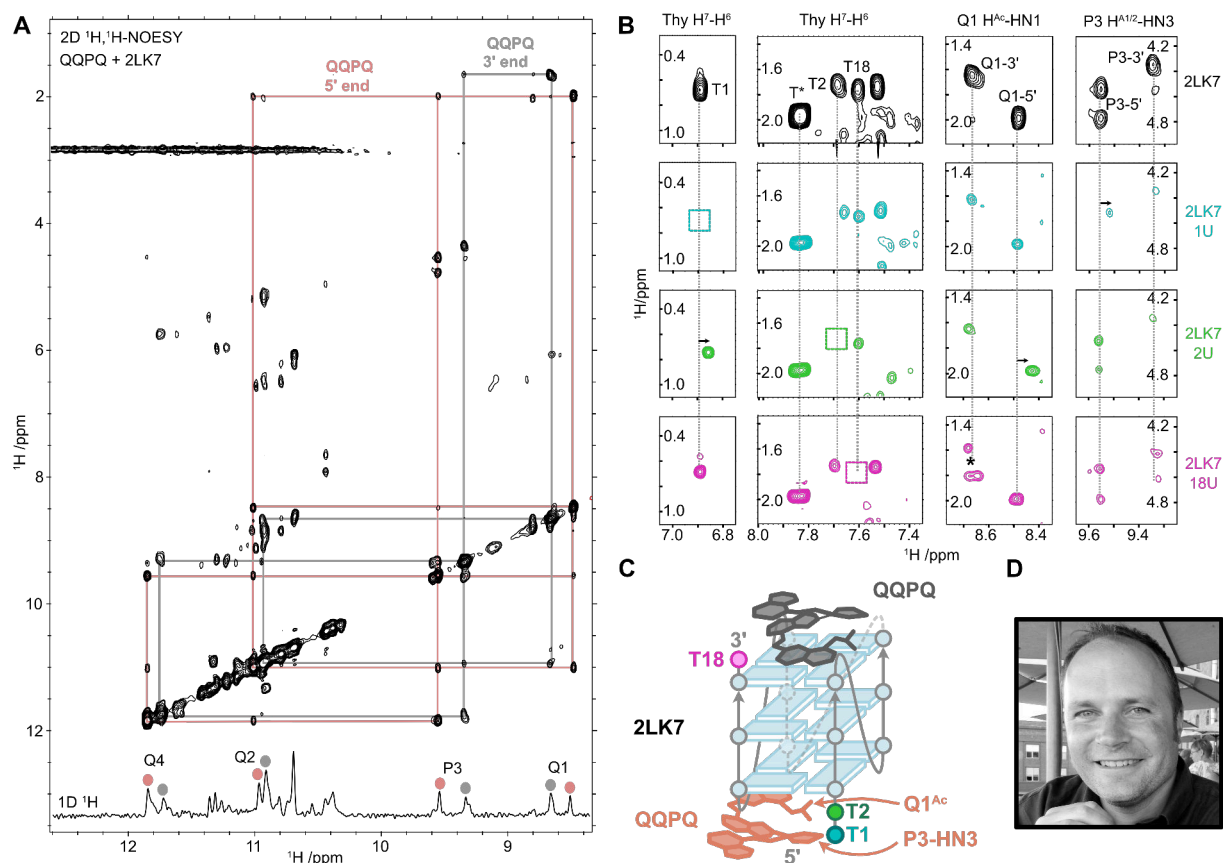


FIGURE 5.18. A. Spectre 2D  $^1\text{H}$ ,  $^1\text{H}$ -NOESY (temps de mélange de 200 ms) de 2LK7 (dTTC<sub>3</sub>TG<sub>3</sub>TG<sub>3</sub>TG<sub>3</sub>T; 100  $\mu\text{M}$ , tampon KP 10 mM, pH 7.0, dans 90/10  $\text{H}_2\text{O}/\text{D}_2\text{O}$ ) en présence de QQPQ (300  $\mu\text{M}$ ), montrant deux réseaux de corrélations NOE reliant les protons amide (NH) du foldamère. L'encart inférieur présente la région imino du spectre 1D annotée avec les signaux des deux espèces de foldamères. B) Régions sélectionnées du spectre  $^1\text{H}$ ,  $^1\text{H}$ -NOESY de QQPQ lié à 2LK7, et des mutants où un T est substitué par un U : 2LK7-1U, 2LK7-2U et 2LK7-18U, ayant permis d'identifier les sites d'interactions en 5' et 3'. Les lignes verticales indiquent les déplacements chimiques d'origine sites protons H5 de la thymine, du NH1 de Q1 et du NH3 de P3 du QQPQ. Les carrés mettent en évidence l'absence de signaux lorsque la thymidine est remplacée par la désoxyuridine. C. Schéma du complexe 2:1 QQPQ:2LK7, où le QQPQ est représenté en saumon et en gris pour les sites de liaison 5' et 3', respectivement. Les résidus T1, T2 et T18 sont colorés comme dans B. Encart : Cameron Mackereth.

champs de force pour ce ligand. Les observations les plus importantes sur un conformère parallèle de 222T ont confirmé nos hypothèses :

- Deux **QQPQ** peuvent se *stacker* en 5' et 3', sans gêne stérique, sans perte d'hélicité du foldamère, et sans préférence notable pour une interaction via les extrémités N et C-terminales.
- Tous les modes d'interactions identifiés par analyse en composantes principales (PCA) et *clustering* des trajectoires MD présentent **QQPQ** positionné de façon décentrée par rapport à la tétrade, ce qui maximise les interactions  $\pi - \pi$  (Figure 5.19B).
- Les chaînes latérales peuvent se lier au squelette phosphate de façon transitoire et sont globalement très dynamiques, à l'instar des dT en 5' et 3', ce qui justifie leur absence dans le modèle cristallographique.
- **QQPQ** peut tourner autour de son axe hélicoïdal (Figure 5.19B), parfois en complétant une rotation complète en l'espace d'une microseconde de simulation, tout en maintenant des angles interplanaires optimaux pour le  $\pi$ -stacking.
- Les amides de **QQPQ** peuvent interagir de façon discrète avec des thymine de boucles (Figure 5.19B), venant confirmer la tendance de ces foldamères à lier des oligonucléotides riches en T. Cette interaction particulière a également été observée pour 5YFY (Figure 5.19C), amenant un élément d'explication de l'affinité particulièrement bonne de **QQPQ** pour ce G4 non parallèle dont la topologie est altérée par l'interaction.

Ces travaux, récemment acceptés pour publication dans *Nucleic Acids Res.* avec **Alexander König** en premier auteur, nous permettent donc de proposer un nouveau paradigme pour la synthèse rationnelle de ligands capables de discriminer des topologies particulières de G4s : il est possible de concevoir des ligands sélectifs de G4s parallèles en associant une surface aromatique permettant le *stacking* à une structure tridimensionnelle non plane empêchant l'interaction avec les autres topologies. Cette approche avait déjà été explorée lors de ma thèse, et avant (66), avec des complexes métalliques incluant des ligands apicaux, mais la gêne stérique associée était bien plus modeste qu'avec des foldamères. Les informations structurales générées par RMN et cristallographie seront très utiles pour optimiser rationnellement ces foldamères pour le ciblage spécifique de certaines topologies G4.

Finalement, cette étude illustre le potentiel de la spectrométrie de masse native pour cribler et caractériser des ligands de G4 sans *cherry picking* et sa complémentarité avec des techniques structurales de plus hautes résolutions. Sur une note plus personnelle, je suis très friand d'études combinant plusieurs techniques orthogonales pour étudier un système d'intérêt, et je trouve très drôle qu'après des années à concevoir rationnellement ou cribler des ligands de G4s, on en trouve un *par hasard*.

### 5.3.1.2 Ligands pérylènes spécifiques d'une conformation intermédiaire de structuration de c-KIT2

Durant mes activités de caractérisation de ligands de G4 par spectrométrie de masse, j'ai également encadré **Silvia Ceschi**, une étudiante de thèse de l'université de Padoue (Figure 5.20C; maintenant postdoc au *Massachusetts General Hospital*), lors de son séjour de 6 mois au laboratoire pour ses travaux sur la séquence KIT2 du promoteur d'oncogène humain *c-KIT* (208). Plus précisément, l'étude était centrée sur la sélection d'un intermédiaire structuré en G4, à long temps de vie, pour lequel un rôle physiologique est possible compte tenu de la vitesse des processus transcriptionnels (209).

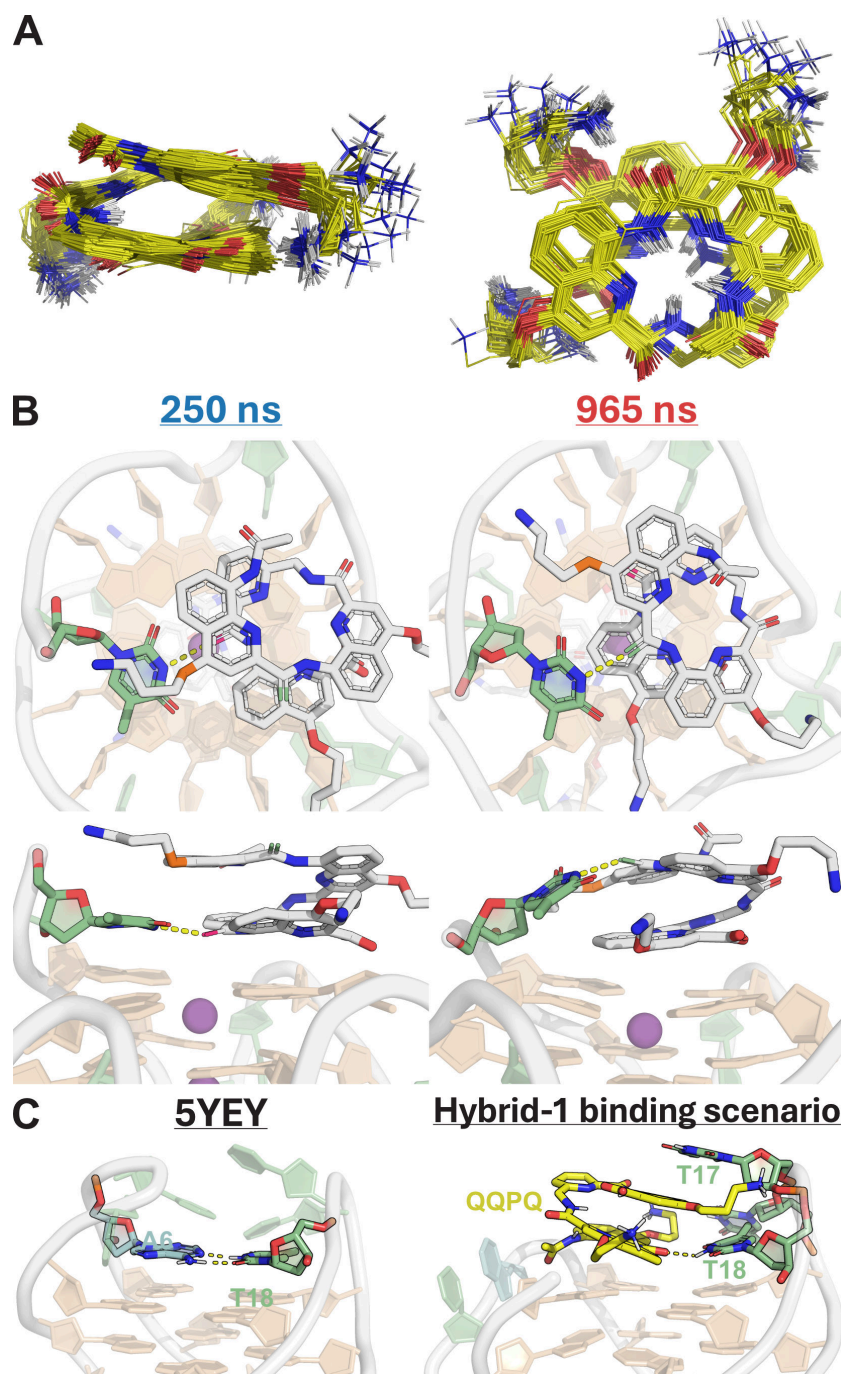


FIGURE 5.19. A. Ensemble conformationnel de **QQPQ**. Vue latérale (gauche) et vue de dessus (droite) des 133 conformères identifiés et superposés par l'algorithme GOAT (*Global Optimizer Algorithm*) : Un premier minimum local est déterminé à partir de la position initiale sur la surface d'énergie potentielle. L'énergie est ensuite augmentée dans un champs de force (GFN-FF) jusqu'à franchir une barrière, puis une nouvelle optimisation de géométrie semi-empirique (GFN2-xTB) permet d'identifier un nouveau minimum. Le processus est répété jusqu'à ce qu'aucun nouveau minimum ne soit trouvé. B. Exemple de mode d'interaction discret de **QQPQ** impliquant une thymine à l'extrémité 5' (liaison hydrogène entre O6 (rose) et le N3 de dT1 (250 ns), inhibant la rotation du foldamère. La disruption de cette interaction rétablit la liberté rotationnelle de **QQPQ** (droite, 965 ns). C. Structure publiée de 5YEQ (gauche) (207), et un complexe **QQPQ**-5YEQ de topologie *hybrid-1*, extrait aléatoirement de la simulation MD. Dans ce dernier, la paire de bases A6:T18 est rompue, laissant de l'espace au-dessus de la tétrade 5' pour que QQPQ puisse se *stacker*. T18 est disponible pour se lier à l'oxygène de l'amide de l'extrémité N-terminale de **QQPQ**, qui est également *stacké* par T17 à son extrémité C-terminale.



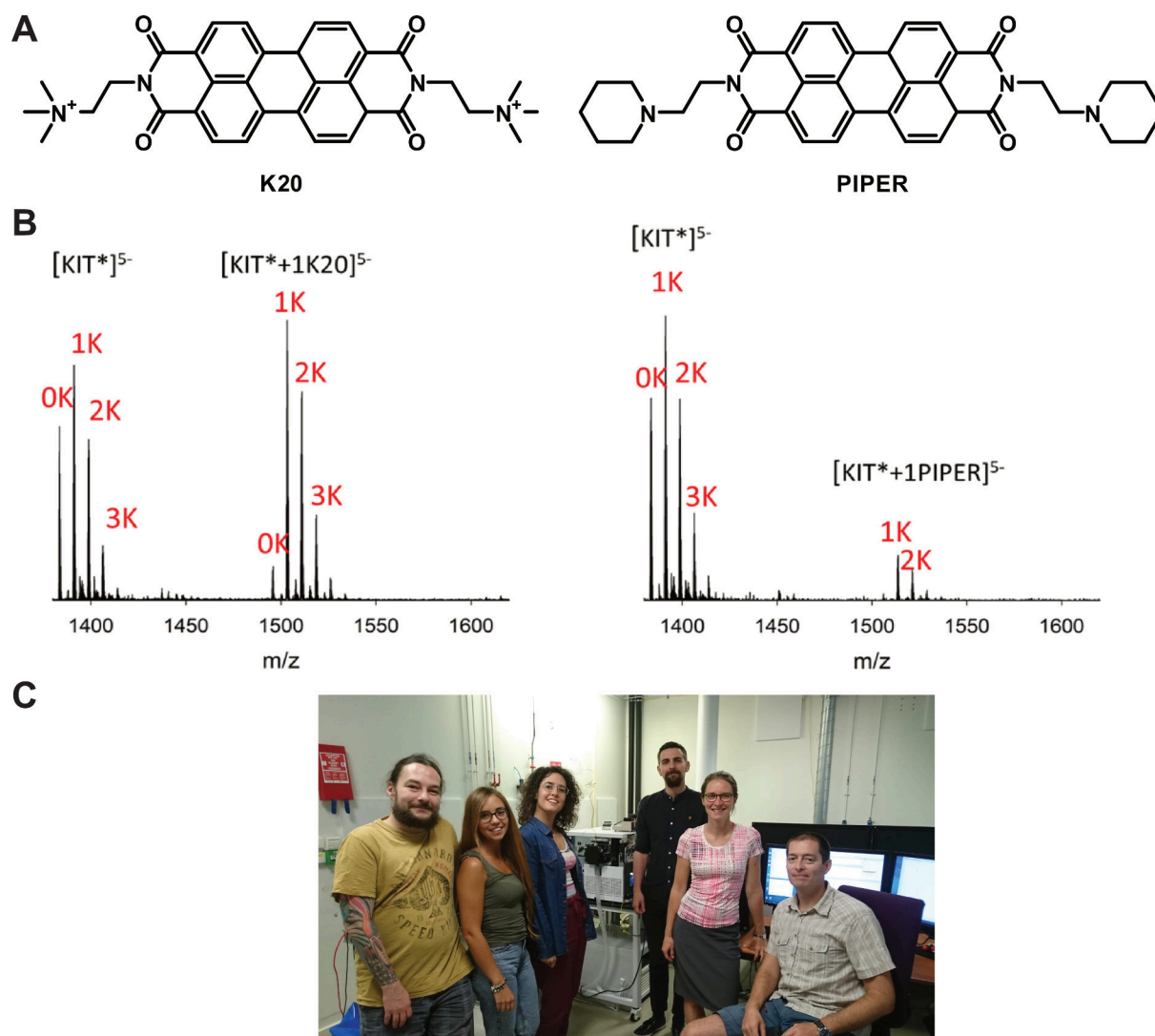


FIGURE 5.20. A. Structures de **PIPER** et **K20**. B. Spectres de masse native de KIT\* (0.01 mM) en présence d'un équivalent molaire de **K20** ou **PIPER**, dans un tampon 100 mM TMAA, 4 mM KCl. On notera ici l'utilisation de quantités de KCl peu adaptées à la spectrométrie de masse native. Adapté avec permission (7). C. L'équipe BALI en juillet 2019. De gauche à droite : Steven Daly, Silvia Ceschi, Federica D'Aria, Eric Largy, Valérie Gabelica, Frédéric Rosu.



Nous avons montré par MS native (7) :

1. que le dérivé de pérylène **K20** (Figure 5.20A) sélectionne cette topologie,
2. qu'un seul ion  $K^+$  est spécifiquement coordonné dans la structure (Figure 5.20B), qui est donc vraisemblablement composée de seulement deux quartets de guanines, et
3. que le ligand apparenté **PIPER**, dont seules les chaînes latérales diffèrent de **K20** (Figure 5.20A), ne montre pas la même sélection conformationnelle. Cette dernière est donc un processus dépendant strictement de la composition des chaînes latérales du pérylène.

Ce projet a nécessité l'utilisation de relativement hautes concentrations de potassium pour que  $KIT^+$  se structure (Figure 5.20B), ce qui a présenté un challenge technique intéressant.

### 5.3.2 Ligands avec des activités anticancéreuses ou antiparasitaires

Dans le cadre de collaboration avec le Pr. Jean Guillon (ARNA, UFR de pharmacie, Bordeaux; Figure 5.21A), j'ai participé à la caractérisation de l'interaction de petites molécules à visée anticancéreuse ou antiparasitaire avec des G4s, par spectrométrie de masse native et dichroïsme circulaire.

J'ai caractérisé avec **Matthieu Ranz** une série de dérivés 2,9-bis[4-(pyridinylalkylaminométhyl)phényl]-1,10-phénanthroline ayant des propriétés antiprolifératives contre des lignées cellulaires de leucémie myéloïde humaine (4).

Avec **Cristina Dal Lago** (Master en séjour Erasmus au laboratoire, Padoue, Italie), nous nous sommes penchés sur l'étude d'une série de dérivés 2,9-Bis[(aminométhyl)]-4,7-phényl-1,10-phénanthroline présentant une activité antiprotozoaire contre les parasites *Trypanosoma cruzi* et *Leishmania infantum* (8).

Dans les deux études, nous avons montré que certains de ces dérivés interagissent avec les G4s de façon modérée ( $K_d$  de l'ordre du micromolaire) sans établir de lien avec leurs propriétés anticancéreuse ou antiparasitaire. J'ai également encadré **Mathilde Melot** et **Antoine Quaresima** (Master, pharmaciens, Bordeaux) sur cette thématique.

Plus récemment, avec **Corinne Buré** de la plateforme de spectrométrie de masse de l'IECB, nous avons étudié des dérivés de phénylquinazolines et phénylquinolines aux propriétés anticancéreuses (210). Un dérivé m'ayant paru intéressant est **12e**, qui se lie avec des affinités de l'ordre du micromolaire aux G4s des promoteurs BCL2 et c-MYC et de la séquence télomérique humaine (24TTG) (Figure 5.21). Dans le dernier cas, des complexes avec deux **12e** se forment coopérativement ( $K_{d1} = 14$ ,  $K_{d2} = 1.9$   $\mu$ M), en déplaçant l'équilibre conformationnel de 24TTG. L'éjection d'un cation ammonium détecté en masse (Figure 5.21C) et la signature CD (Figure 5.21D) suggèrent la formation d'un G4 antiparallèle à deux tétrades, ou l'intercalation d'un des deux ligands (ce qui est beaucoup moins probable). À l'inverse, les complexes 1:1 conservent leurs deux cations ammoniums, ce qui indique que ce premier complexe conserve la topologie originale du G4, mais que celle-ci est peu favorable à l'interaction.

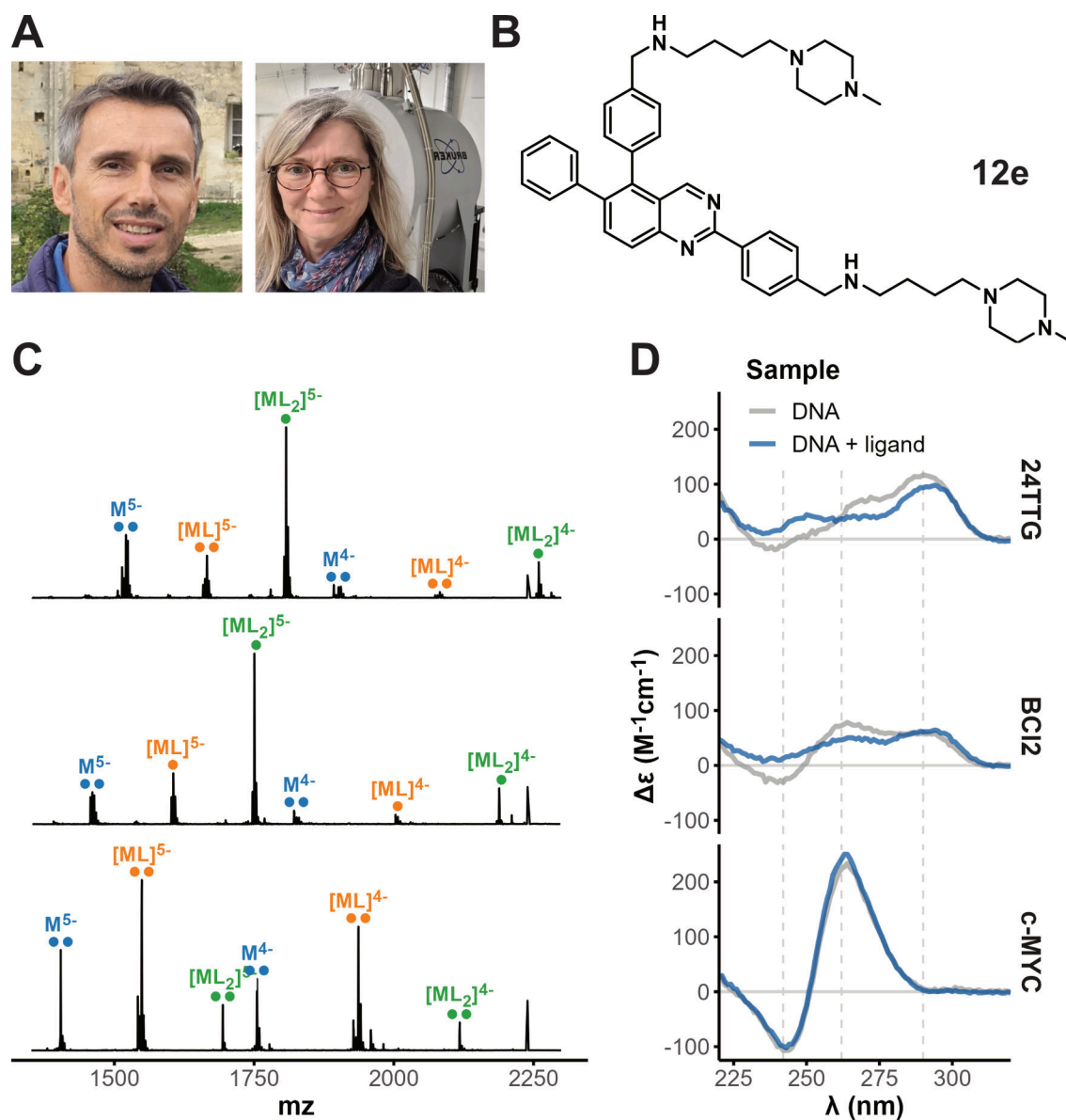


FIGURE 5.21. A. Jean Guillon et Corinne Buré. B. Structure de **12e**. C. Spectre de masse native de 24TTG, BCL2 et c-myc (de haut en bas), dans un pseudo-tampon acétate d'ammonium, en présence de deux équivalents de **12e** (M = oligonucléotide monomérique, L = **12e**). D. Spectres CD correspondants.

## 5.4 Développement de logiciels *open source* pour l'analyse de données

L'ensemble de mes contributions *open-source* sont déposées sur Github : [github.com/EricLarG4](https://github.com/EricLarG4).

### 5.4.1 Détermination de topologies d'acides nucléiques G4 par spectroscopie UV-vis et dichroïsme circulaire

Dans le cadre d'une collaboration avec J.-L. Mergny (École Polytechnique) et Samir Amrane (ARNA, INSERM U1212; Figure 5.47), qui a acquis des données spectroscopiques (UV-vis et dichroïsme circulaire) avec Amani Kabbara (doctorante, [thèse soutenue en 2022](#)), j'ai développé le logiciel [Eps2Fold](#) pour déterminer la topologie de structures d'ADN G4 à relativement haut-débit, par analyse en composantes principales de CD et de spectres UV (211). Initialement, ces derniers étaient exploités sous forme de signatures IDS (*Isothermal Difference Spectra*), qui sont obtenues par la soustraction de spectres d'oligonucléotides structurés en G4 (en présence d'un cation comme le potassium) et non-structurés correspondant (en absence du cation, donc). C'est une version isotherme d'une autre signature plus couramment utilisée pour détecter la formation d'un G4, ou le spectre "non structuré" est acquis à haute température d'où son nom : TDS, pour *Thermal Difference Spectra* (212).

L'intérêt des spectres CD pour étudier la topologie d'acides nucléiques n'est plus à démontrer (213–216). Celui de l'IDS, en revanche, est moins évident. Bien que calculé à partir de données pouvant paraître très sommaires, l'IDS contient des informations structurales plus intéressantes qu'il n'y paraît et possède l'avantage, par rapport au TDS, d'être une technique isotherme. Par exemple, lors de mon postdoctorat dans l'équipe de J.-L. Mergny, j'avais déjà observé que les signatures IDS des conformères de 22T en présence de KCl et NaCl différaient, ce qui m'avait permis d'acquérir des cinétiques d'interconversion du second vers le premier avec de simples spectres UV (31). Grâce au faible temps mort de cette approche, et au contrôle de la température qu'elle permet, j'avais accédé à l'énergie d'activation de cette interconversion en mesurant ces cinétiques UV à différentes températures.

Il n'est cependant pas toujours aisé de déterminer la signature IDS d'un échantillon. Les G4s les plus stables (en particulier les parallèles à courtes boucles) peuvent se former en présences de faibles concentrations en cations et/ou sont difficiles à dessaler, du fait de leurs excellents  $K_d$  pour ces cations. En d'autres termes, le spectre UV acquis en absence de cations ne l'est pas vraiment, et la signature IDS qui en résulte est erronée. Pour référence, les séquences T95TT et T30177TT sont entièrement structurées à 5  $\mu$ M en présence de quantités stœchiométriques de potassium (31). Pour parer à ce problème, j'ai décidé de me passer de ces spectres difficile à acquérir expérimentalement, et de plutôt les calculer théoriquement. Pour cela, j'ai développé un algorithme permettant de calculer des spectres UV théoriques d'ADN non structuré, décrit dans la Section 5.5 (Figure 5.34). J'ai ensuite pu vérifier ce problème, et l'intérêt du calcul théorique de ces spectres, sur le set de données à ma disposition (Figure 5.22).

J'ai ensuite développé le [programme Eps2Fold](#) permettant le traitement des données UV et CD (import des données, soustraction de blancs, normalisation des données, calcul des signatures, analyse PCA, *clustering*), avec lequel nous avons construit une base de données spectrales pour laquelle la structuration des oligonucléotides a été validée, notamment par RMN (Figure 5.23).

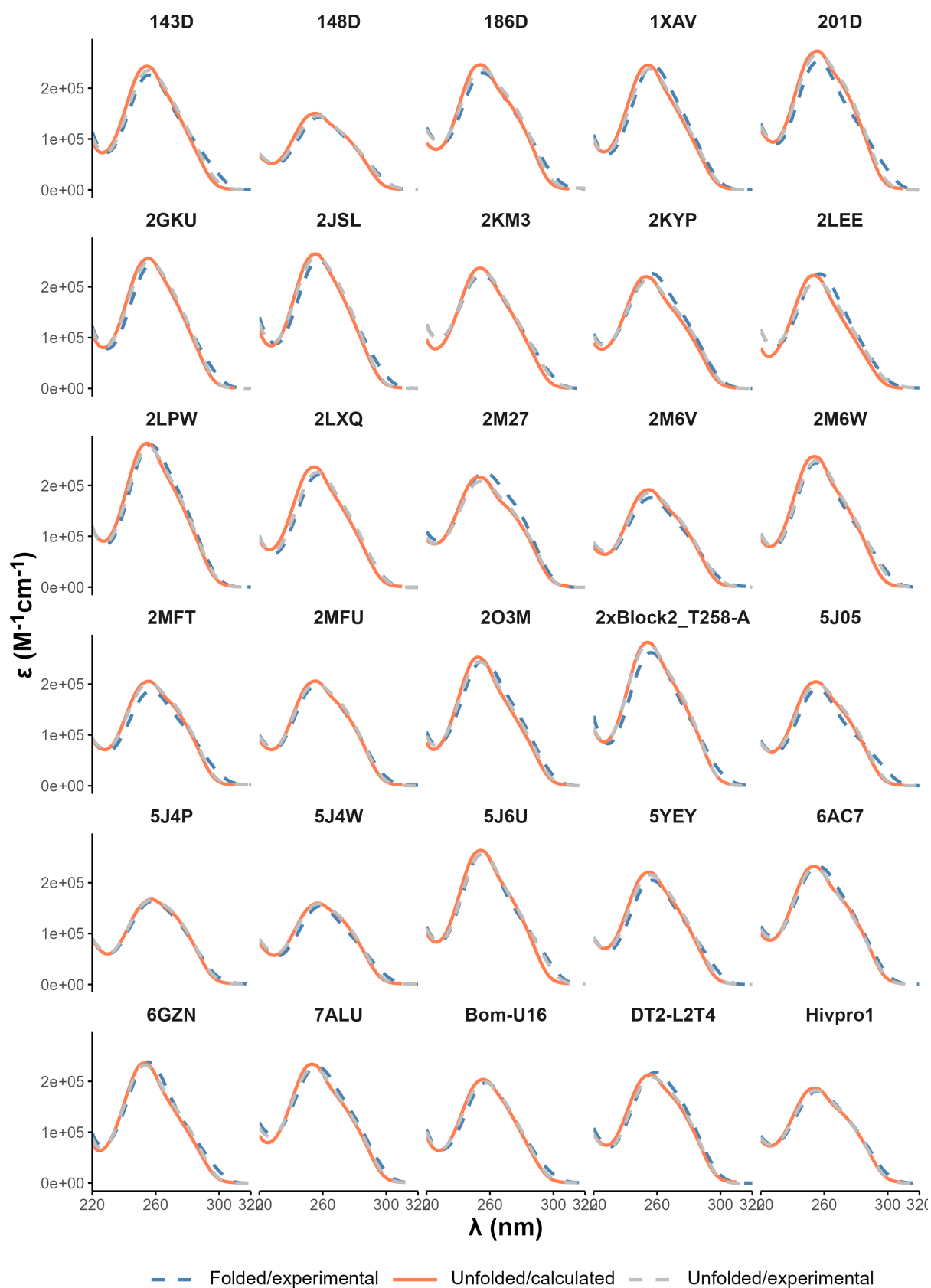


FIGURE 5.22. Comparaison des spectres UV “non structurés” calculés et expérimentaux (en absence de cation) et “structurés” (en présence de cation).

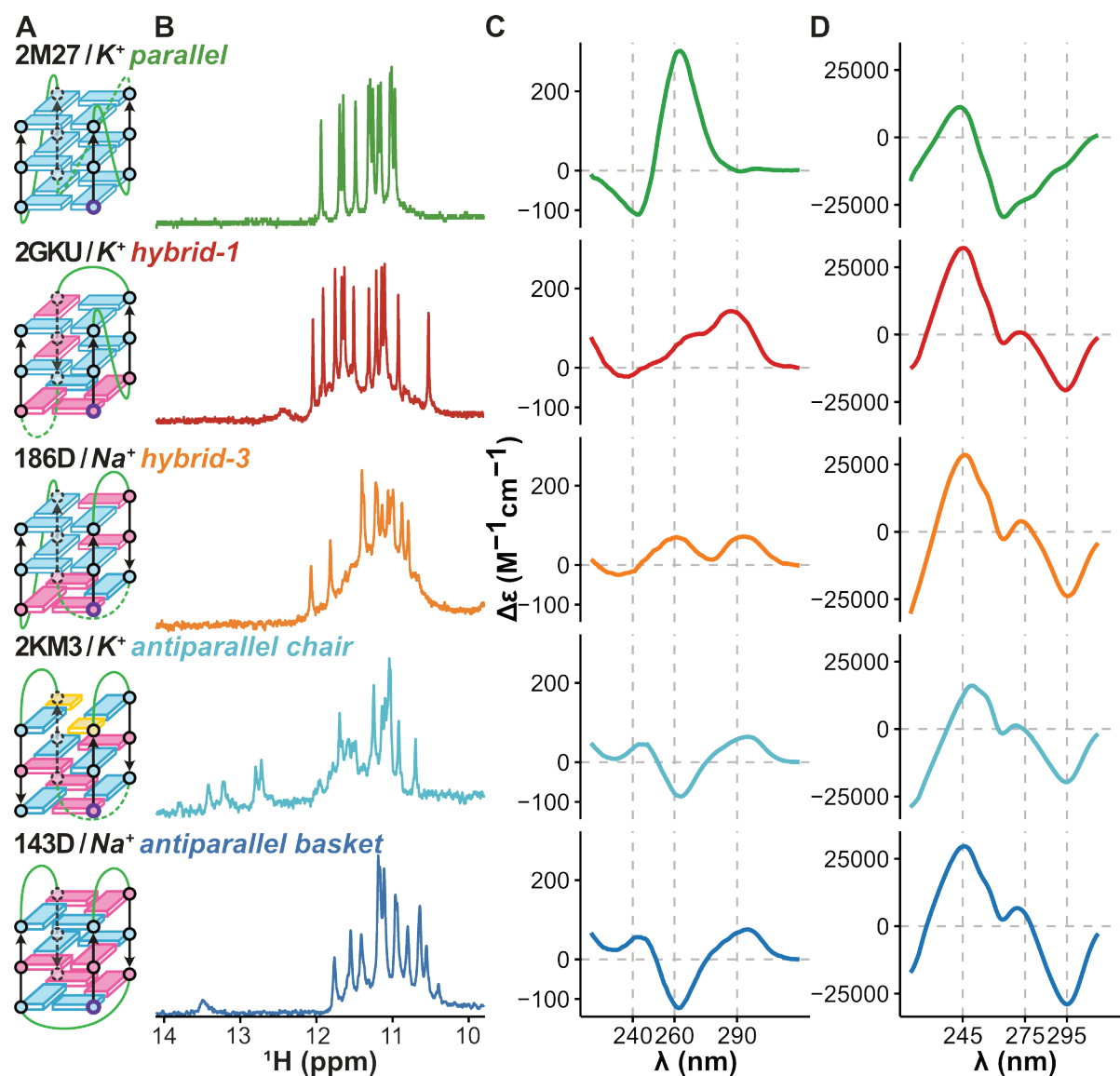


FIGURE 5.23. Exemples de données spectroscopiques pour les conformères les plus classiques. A. Schéma des structures (bleu : *anti*, rose : *syn*, jaune : cytosines). B. Région imino de spectres de RMN <sup>1</sup>H. C. Spectres CD. D. Signatures Eps2Fold. C et D sont exprimés en  $\Delta\epsilon$  (en M-1cm-1) mais sont des mesures différentes : C correspond au dichroïsme circulaire molaire  $\epsilon_L - \epsilon_R$  et D à une différence de coefficient d'extinctions molaires en fonction de la structure  $\epsilon_{destructure} - \epsilon_{structure}$ .

L'analyse en composantes principales et la détermination de *clusters* m'a amené à explorer les différents descripteurs structuraux des G4s : enchaînements d'angles glycosidiques des guanines (*syn* ou *anti*), de géométries de boucles, ou encore de sillons (Figure 5.24). En particulier, j'ai souhaité trouver un compromis entre la finesse de la description, sa capacité à décrire les différences observées expérimentalement, et sa lisibilité.

Ainsi, l'orientation *syn/anti* des liaisons glycosidiques (GBA) permet une séparation très fine des topologies mais est peu lisible (Figure 5.25D). À l'inverse, la séparation classique en topologies parallèle/antiparallèle/hybride est très peu performante, même en discriminant les antiparallèles *chair* et *basket*, mais très lisible (Figure 5.25A).

Au-delà du logiciel mise à disposition de la communauté, il est ressorti de cette étude que :

- Une combinaison donnée de géométries de *stacking* de guanines génère des signatures CD et Eps2Fold spécifiques, qui peuvent être résolues avec une granulométrie surprenante. Qualitativement, les tendances sont très similaires, et le CD intrinsèquement plus résolutif.
- Eps2Fold n'est pas un remplacement au CD, qui reste plus discriminant, mais peut venir en un complément encore plus rapide et facile à mettre en place. La technique pourrait être implémentée dans des expériences de criblage haut débit de G4s.
- L'interprétation des signatures Eps2Fold et CD ne doit pas se limiter à une inspection visuelle. Des changements structuraux discrets peuvent modifier ces signatures de façon modeste que seul un traitement des données systématique et sans biais humain peut détecter.
- Les descripteurs structuraux classiques pour décrire des résultats CD/Eps2Fold sont limités, en particulier quand ils se basent sur la polarité des brins (comme les topologies) plutôt que sur les géométries des guanines, puisque ce sont ces dernières qui expliquent le signal.
- Les descripteurs permettant des granulométries fines, comme l'enchaînement des GBA, permettent de discriminer efficacement les conformères G4s mais sont peu lisibles. Des descripteurs plus synthétiques, comme les chiralités de tétrades (Figure 5.25B) et les GBA *stacks* (Figure 5.25C, où l'on compte le nombre de *stacks* différents), présentent des compromis intéressants.

Dans la perspective de nos travaux futurs, nous visons à étendre cette approche à un éventail plus large de structures secondaires au-delà des G4s, et en explorant aussi les oligonucléotides d'ARN. En particulier, il sera essentiel d'évaluer la généralisabilité d'Eps2Fold en termes de résolution, ce qui pourrait nécessiter le développement de descripteurs structuraux adaptés à d'autres motifs de structures secondaires.

### 5.4.2 Analyse haut débit et haute précision d'UV-melting d'acides nucléiques

Les expériences de dénaturation thermique (*melting*), sont très couramment utilisées pour estimer la stabilité des structures secondaires d'acides nucléiques. Ces courbes sont le plus souvent obtenues avec une détection UV-vis, mais le CD et le FRET sont également fréquemment utilisés. Dans tous les cas, le signal est enregistré en fonction de la température et rend compte de l'état de repliement intramoléculaire et/ou d'association intermoléculaire des oligonucléotides (Figure 5.26A). Typiquement, la dénaturation de l'ADN duplex est suivie par l'hyperchromisme qui se produit à 260 nm. D'autres types



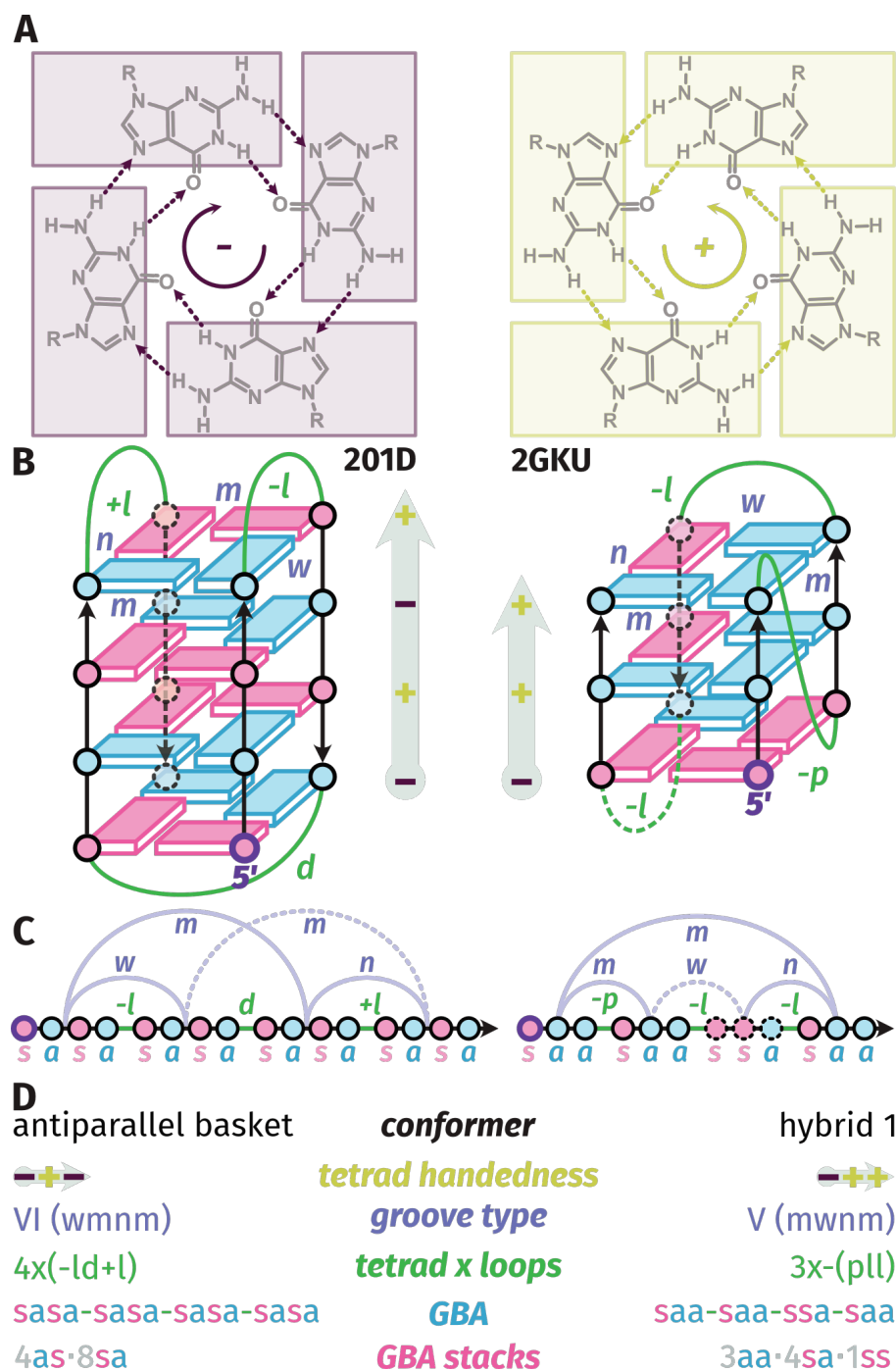


FIGURE 5.24. A. Sens de rotation (horaire (-) ou antihoraire (+)) des tétrades de guanines, définis par la direction des donneurs vers les accepteurs de liaisons H. B. Schémas des topologies antiparallèle *basket* (PDB : 143D (42)) et hybride 1 (PDB : 2GKU (44)). Les guanines bleues et roses présentent respectivement des angles de liaison glycosidique *anti* et *syn*. La chiralité des tétrades est notée + ou - selon la définition du panneau A, en les observant par le dessus et en commençant en 5' (comme définit par V. Gabelica). Les boucles présentent des géométries *propeller* (p), latérale (l) ou diagonale (d), annotées avec + ou - lorsqu'elles relient les guanines dans un sens horaire ou antihoraire, respectivement, avec l'extrémité 5' située en bas à droite de la structure (comme définit par M. Webba Da Silva (29)). Les sillons sont classés comme larges (w), moyens (m) ou étroits (n). C. Représentations linéaires des schémas du panneau B. D. Résumé des six descripteurs structuraux des topologies 143D et 2GKU du panneau B.

125

de structures peuvent nécessiter des longueurs d'onde plus spécifiques, par exemple la dénaturation des G4s est caractérisée par une hypochromie à 295 nm (212).

Dans la plupart des études, l'information quantitative extraite des courbes de dénaturation est la température de fusion ( $T_m$ ), c'est-à-dire la température de dénaturation à mi-transition (Figure 5.26B–D). Elle peut être utilisée pour évaluer la stabilité des structures secondaires formées en solution. Elle permet de classer la stabilité thermique relative de différentes séquences ou d'une seule séquence dans différentes conditions (pH, force ionique, nature et concentration des cations). Le  $T_m$  est également très utilisé comme proxy de l'affinité de complexes d'oligonucléotides avec des ligands, en comparant les valeurs en présence et absence de ligands. Il ne s'agit pas d'une méthode de mesure directe et isotherme de l'affinité, mais c'est un moyen simple et populaire de classer les ligands.

En réalité, ces courbes peuvent fournir plus que les seuls  $T_m$ . Dans de nombreux cas, il est plus pertinent de déterminer la fraction structurée/associée  $\theta$  à la température d'intérêt de l'étude (Figure 5.26E–G). Il est par exemple possible pour deux oligonucléotides de  $T_m$  similaires d'avoir un  $\theta$  différent à la température d'étude, et l'utilisation de la  $T_m$  peut donc s'avérer simpliste voire erronée. Cependant, la détermination correcte de  $T_m$  nécessite de traiter les données afin de déconvoluer les contributions des lignes de base non planes (pré et post-transition), qui résultent de la dépendance à la température des coefficients d'absorption molaires (217). Ce traitement de données souffre de sa lenteur et de biais humains réduisant la précision des résultats, notamment le choix des régions pour la modélisation des lignes de bases. Plus généralement, l'utilisation de différentes méthodes de détermination de la  $T_m$  mène à une variabilité importante des résultats ( $\Delta T_m$  de 1.4°C dans l'exemple de la Figure 5.26). Il est fréquent de voir dans la littérature des valeurs de  $T_m$  avec un nombre de chiffres significatifs qui ne sauraient être compatibles avec la précision expérimentale.

Dans le cadre de l'établissement d'une base de données de G4s (voir Section 5.5, ci-après), j'ai été amené à devoir traiter un grand nombre de *meltings* pour lesquels je souhaitais avoir une approche la plus standardisée possible. La période de confinement consécutive à la pandémie de COVID a été propice au développement d'un algorithme permettant un traitement rapide et semi-automatisé de ces données. Cette approche a été intégrée à un logiciel nous ayant permis de construire la base de données (218). Devant l'intérêt de certains collègues pour ces fonctionnalités, j'ai finalement décidé de développer un logiciel dédié, et [open-source](#), [meltR](#).

En plus des méthodes usuelles, une approche basée sur la modélisation thermodynamique non biaisée des données a été développée (Figure 5.26H), et comparée aux autres méthodes (Figure 5.27).

Le modèle repose sur l'expression des coefficients d'extinction  $\varepsilon_T$  comme la somme des contributions des formes structurées (*folded*, F) et non-structurées (*unfolded*, U), pondérées par leurs abondances relatives (Équation 5.1).

$$\varepsilon_T = \varepsilon_T^F \times \theta_T + \varepsilon_T^U \times (1 - \theta_T) \quad (5.1)$$

La fraction *foldée* est définie par  $\theta = \frac{[F]}{[F] + [U]}$ . En considérant un modèle simple à deux états  $F \rightleftharpoons U$  ayant une constante d'équilibre  $K$ ,  $\theta$  peut être exprimé par l'Équation 5.2, qui mène à l'Équation 5.3.

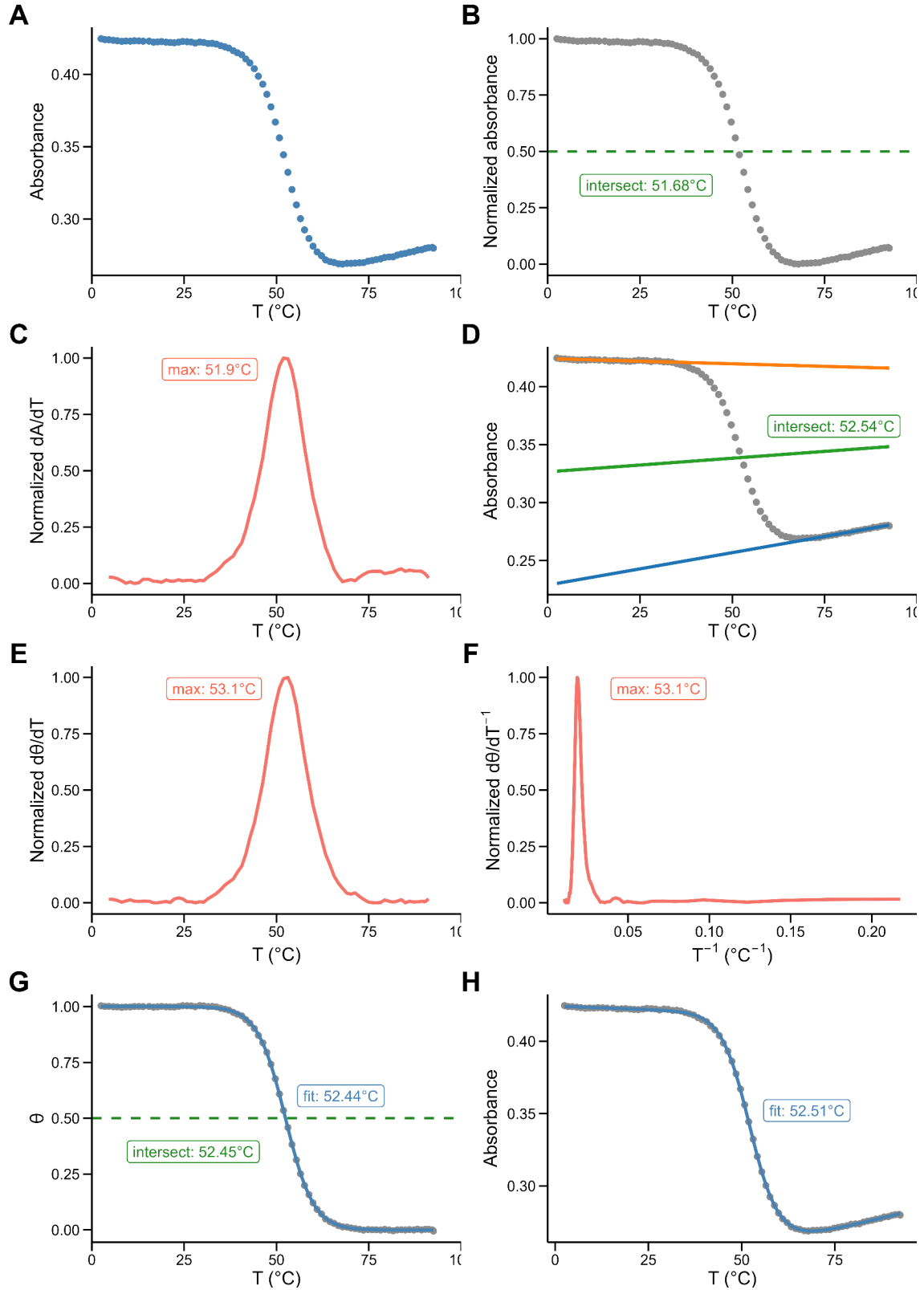


FIGURE 5.26. Méthode de traitement des données de dénaturation thermique, illustrées avec les données brutes du panneau A (22GT, courbe de refroidissement). B. Intersection de  $A_{norm} = 0.5$  (vert) avec la courbe d'absorbance normalisée à [0;1]. C. Maximum de la dérivée première  $dA/dT$ . D. Intersection de la médiane (vert) des lignes de bases avec la courbe d'absorbance brute (lignes de base à basse température en orange et à haute température en bleu obtenues par régression linéaire). E. Maximum de la dérivée première de la fraction structurée par la température  $d\theta/dT$ . F. Comme E mais par l'inverse de la température  $d\theta/dT^{-1}$ . G. Intersection à 0.5 de la fraction structurée  $\theta$ . H. Régression non-linéaire des données brutes avec un modèle thermodynamique.

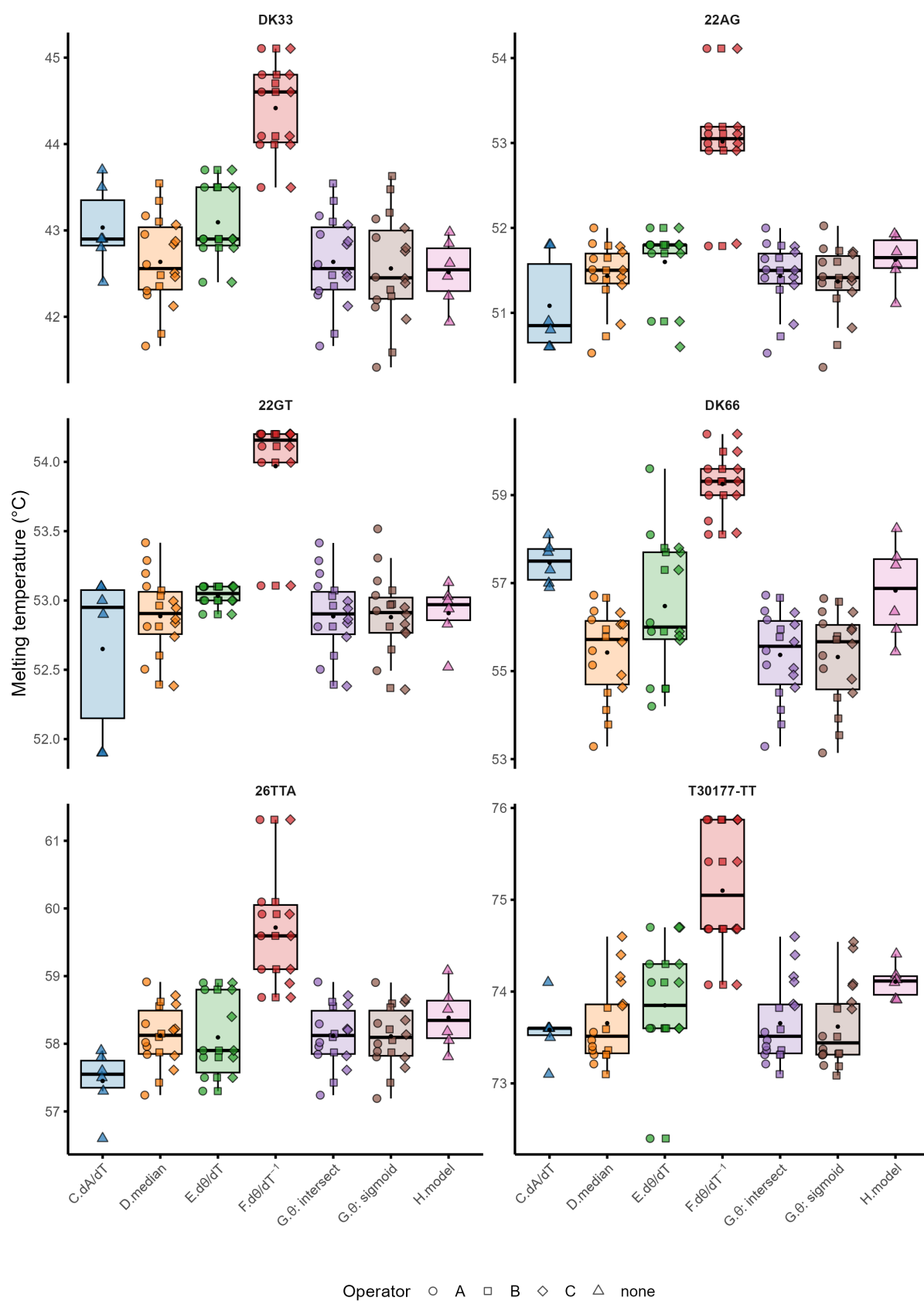


FIGURE 5.27. Comparaison de la précision des méthodes de détermination de la  $T_m$  sur 6 réplicats. Ici, les données de 6 oligonucléotides ont été traitées par 8 méthodes et 3 opérateurs différents.

$$\theta = \frac{1}{1 + K} \quad (5.2)$$

$$\varepsilon_T = \varepsilon_T^F \times \frac{1}{1 + K} + \varepsilon_T^U \times \frac{K}{1 + K} \quad (5.3)$$

$\varepsilon_T^F$  et  $\varepsilon_T^U$  peuvent être modélisés comme des fonctions linéaires de la température (Équation 5.4), où  $a$  est la pente et  $b$  l'ordonnée à l'origine de ces lignes de base :

$$\varepsilon_T = (a^F T + b^F) \times \frac{1}{1 + K} + (a^U T + b^U) \times \frac{K}{1 + K} \quad (5.4)$$

$K$  est exprimé dans l'Équation 5.5 par des quantités thermodynamiques d'intérêt.

$$-RT \ln K = \Delta G^0 = \Delta H^0 - T \Delta S^0 \quad (5.5)$$

Les éventuels changement de capacité thermique dans la gamme de températures étudiée n'ont pas été pris en compte ici, car cela a mené à des sur-paramétrisations du modèle. À la température de *melting*, l'Équation 5.6 mène aux Équations 5.7 et 5.8.

$$\Delta G_m^0 = \Delta H_m^0 - T \Delta S_m^0 = 0 \quad (5.6)$$

$$\Delta S_m^0 = \frac{\Delta H_m^0}{T_m} \quad (5.7)$$

$$\Delta G^0 = \Delta H_m^0 \left(1 - \frac{T}{T_m}\right) \quad (5.8)$$

Finalement,  $K$  peut être exprimé par  $\exp\left(-\frac{\Delta H^0(1-\frac{T}{T_m})}{RT}\right)$  ce qui mène à l'Équation 5.9 :

$$A_T = (a^F T + b^F) \times \frac{1}{1 + \exp\left(-\frac{\Delta H^0(1-\frac{T}{T_m})}{RT}\right)} + (a^U T + b^U) \times \frac{\exp\left(-\frac{\Delta H^0(1-\frac{T}{T_m})}{RT}\right)}{1 + \exp\left(-\frac{\Delta H^0(1-\frac{T}{T_m})}{RT}\right)} \quad (5.9)$$

[meltR](#) est disponible en ligne et permet de :

- Déterminer les températures de *melting* et paramètres thermodynamiques sans biais de l'opérateur,
- Déterminer les fractions structurées aux températures souhaitées,
- Analyser les courbes de chauffage et refroidissement indépendamment, afin de détecter d'éventuelles cinétiques lentes d'association/repliement,
- Automatiser toutes les étapes de l'analyse pour réduire les erreurs, accélérer et réduire la pénibilité du processus. Une dizaine d'échantillons peut être traitée en quelques secondes,



- Visualiser le résultat d'un traitement de données classique, par soustraction des lignes de base, en tenant compte de l'incertitude sur la détermination de ces lignes de bases. Plutôt que de considérer des régions discrètes et uniques pour chaque ligne de base, on entre plutôt des gammes possibles de régression. meltR permet ensuite de simuler un très grand nombre de lignes de bases dans ces régions, ce qui génère des *distributions* de  $T_m$  et  $\theta$ . Cela permet d'illustrer visuellement les biais d'opérateur, ce qui a, *a minima*, une valeur ajoutée pédagogique. La Figure 5.28 montre les différences entre 1000 couples de lignes de bases pour trois réplicats de deux rampes de la dénaturation de 22AG, ce qui entraîne des différences significatives de  $T_m$  comme illustrées Figure 5.29.

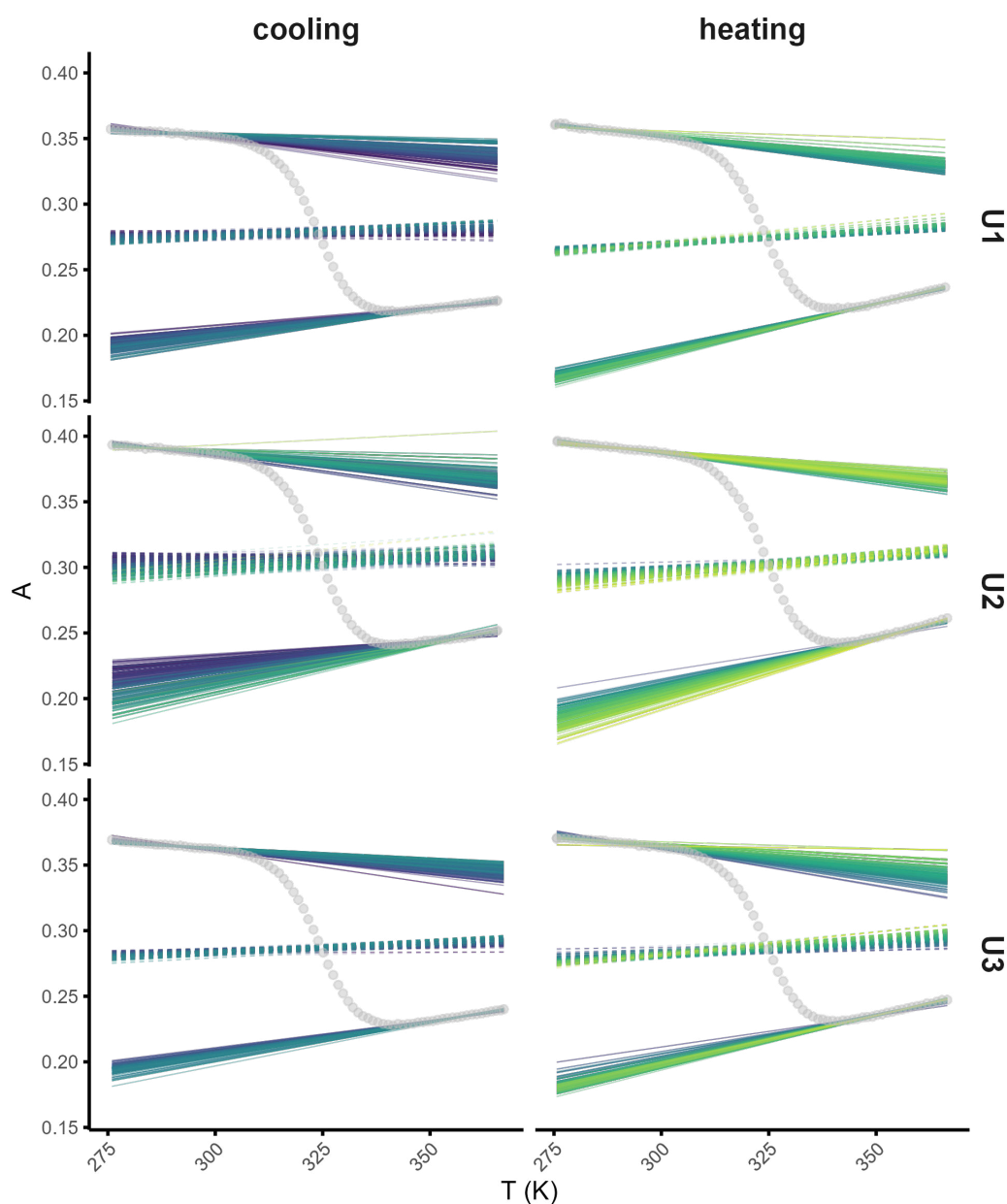


FIGURE 5.28. Traitement de données d'UV-melting par la méthode tenant compte des lignes de bases, pour trois réplicats d'une expérience à deux rampes (refroidissement puis chauffe), sur l'oligonucléotide 22AG. 1000 couples de lignes de base ont été générés à partir d'un set d'autant de régions sur lesquelles les régressions linéaires ont été effectuées. L'intersection des médianes correspondantes (en pointillés) génère 1000  $T_m$ , illustrant la potentielle variabilité inter-opérateur.

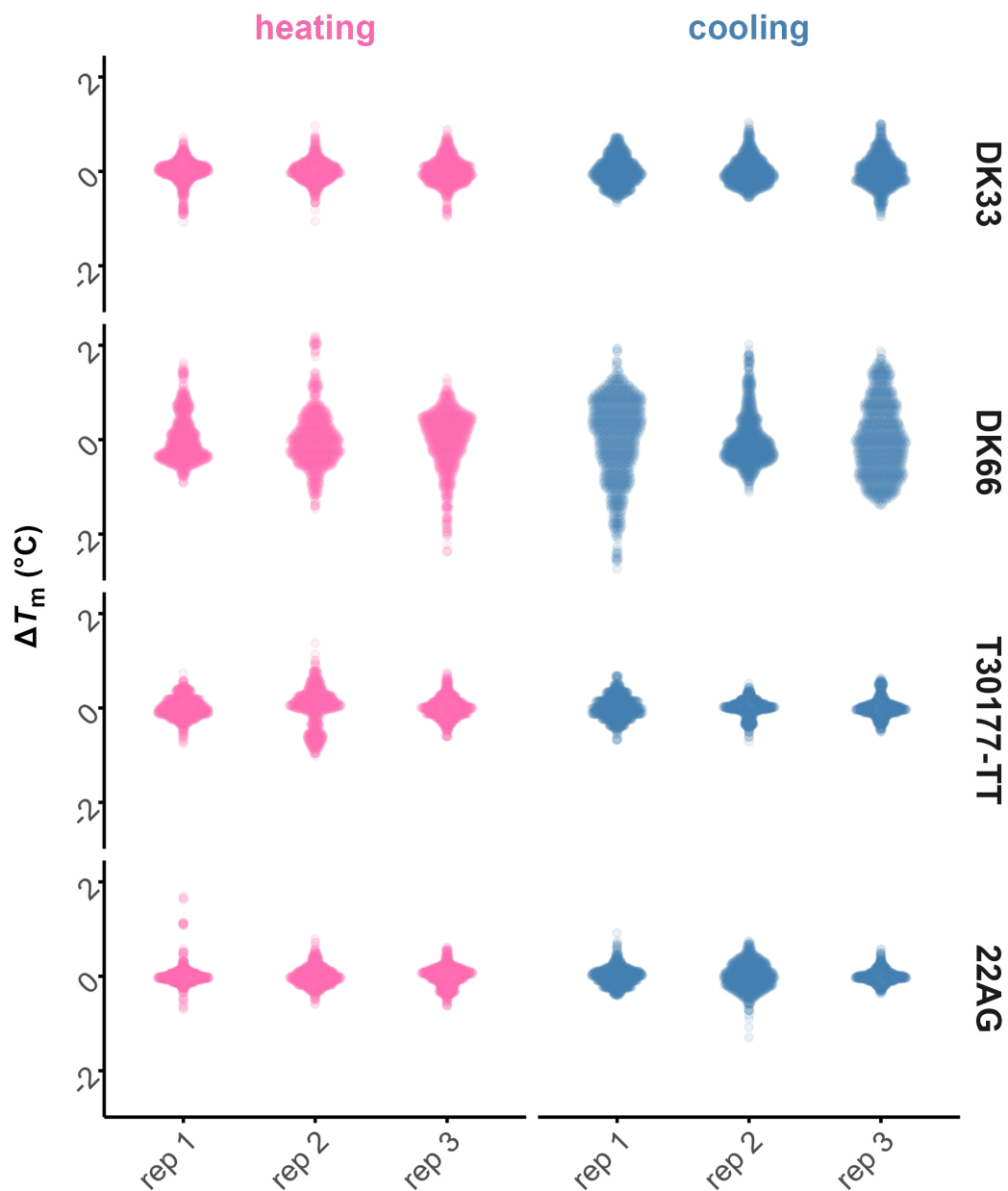


FIGURE 5.29. Distribution des différences de  $T_m$  mesurées pour quatre oligonucléotides (groupés par réplikat et rampe) relatives à leur  $T_m$  moyenne. Ces différences augmentent généralement pour les espèces particulièrement stables ou instables, pour lesquelles la transition est proche des limites haute ou basse de la gamme de température explorée. Dans ces cas de figure, les régressions linéaires doivent être faites sur de petites régions, ce qui génère de l'erreur.

J'ai encadré plusieurs étudiants pour la génération de données expérimentales et le test du logiciel afin d'en estimer les performances : **Matthieu Bonnardel** (Licence, Bordeaux) : **Axel Florent** (DEUST, Bordeaux), et **Matthieu Vida** (Licence, Bordeaux).

L'unique perspective de ce projet est de le publier. S'il ne révolutionnera pas la biophysique des oligonucléotides, il constitue un outil pratique pour la communauté.

### 5.4.3 Spectrométrie de masse native d'acides nucléiques

Comme je l'ai brièvement exposé dans le chapitre Chapitre 4, l'HDX/MS est une méthode de plus en plus utilisée pour étudier la dynamique structurale et l'interaction des protéines, généralement avec une approche *bottom-up* ([Figure 4.4]). Cette approche génère une grande quantité de données, dont le traitement est long et pénible.

Par conséquent, de nombreux logiciels ont été développés pour faciliter le traitement des données brutes HDX/MS (détection et identification des peptides, détermination de leur teneur en deutérium, modélisation et ajustement de la distribution isotopique) (219–229), le post-traitement (comparaison de la deutération entre les échantillons, évaluation de la signification statistique) et la visualisation (222, 225, 230–240).

Nous l'avons vu, j'ai lancé des travaux visant à démontrer que l'HDX/MS est également une excellente méthode pour étudier les structures secondaires d'acides nucléiques (Section 5.2.1) (6, 152). Contrairement à l'approche standard utilisée pour les protéines, cette méthode ne nécessite pas d'étapes de *quenching* ou de dénaturation, ni de séparation chromatographique. Pour rappel, deux techniques ont été développées : *i*) mesure en temps réel de l'échange (RT-HDX), où chaque scan correspond à un moment précis de deutération et *ii*) mélange en flux continu (CF-HDX; Figure 5.3C) où tous les scans MS d'une configuration donnée correspondent au même temps de deutération. Ces méthodes offrent plusieurs avantages : compatibilité avec des conditions natives, séparation de mélanges conformationnels par masse et structure, réduction du nombre d'analytes par injection tout en augmentant le nombre de points pour chaque cinétique (généralement une vingtaine contre, 4-5 en *bottom-up*). Cependant, ces approches présentent des défis spécifiques, notamment le besoin de calculer des enveloppes isotopiques théoriques pour les analytes d'ADN, dont les sites échangeables diffèrent de ceux des peptides. De plus, il est nécessaire de prendre en compte les interactions non covalentes avec des cations et des molécules organiques, qui peuvent d'ailleurs aussi contenir des sites échangeables.

Pour répondre à ces exigences, j'ai développé une application open-source nommée **OligoR** (5). Cette application, codée en R, est conçue pour traiter les données brutes de HDX/MS des oligonucléotides, allant de l'importation des données brutes de spectrométrie de masse à la génération de figures pour publication, en passant par la détermination des centroïdes et le calcul de la deutération (Figure 5.30).

OligoR permet également de déterminer le nombre de sites échangeable d'un oligonucléotide, le calcul de distributions isotopiques d'oligonucléotides et de leurs complexes directement à partir de leurs séquences et/ou formules brutes (Figure 5.31A, D), et la déconvolution des distributions bimodales (Figure 5.32), ce qui est crucial pour l'analyse des cinétiques d'échange et des constantes d'équilibre sous-jacentes que nous avons exploitées ci-avant (Section 5.2.1.2). Finalement, OligoR peut être utilisé

pour l'automatisation du traitement de données de titrations et cinétiques de complexations en spectrométrie de masse native (Figure 5.31E, F).

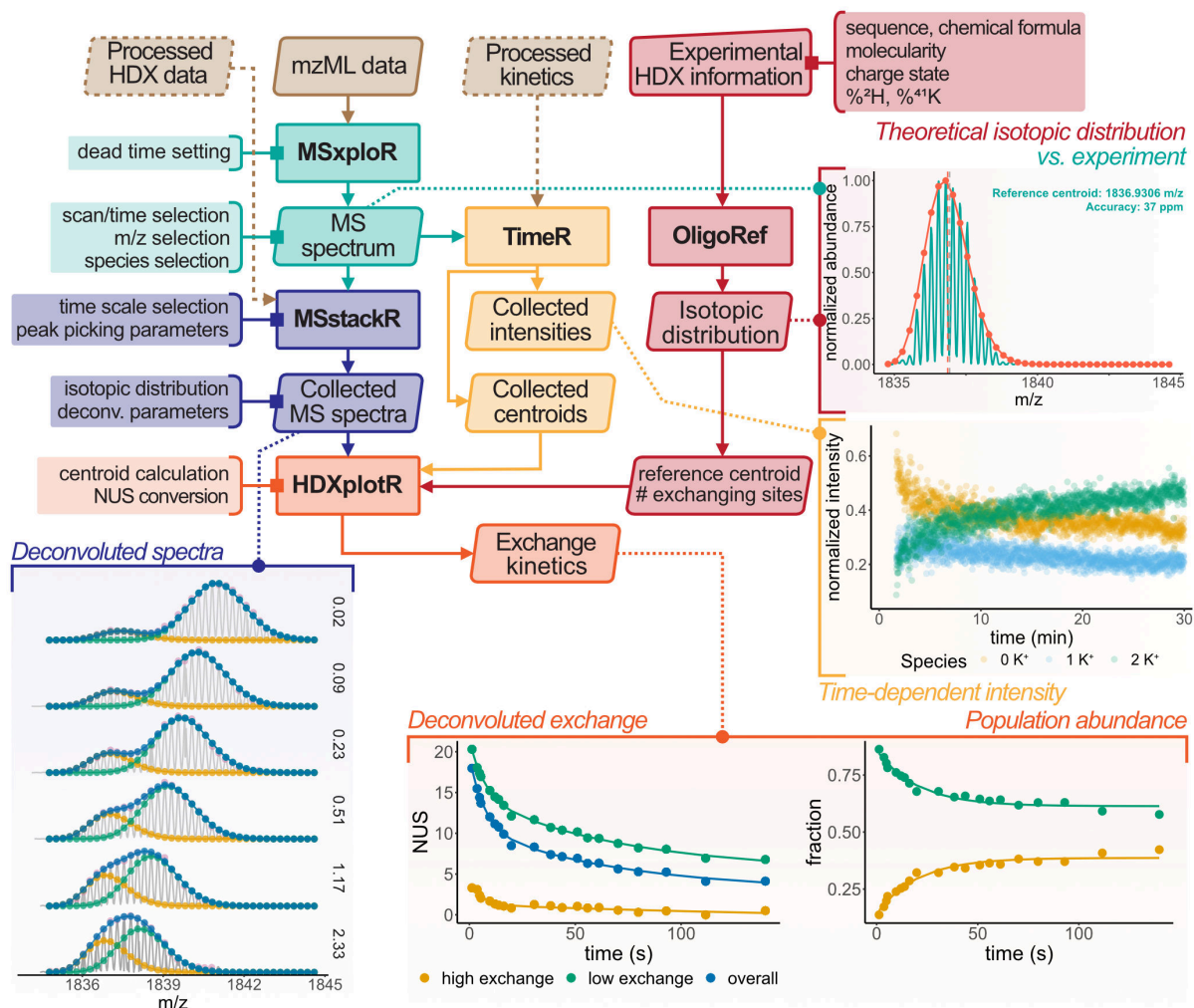


FIGURE 5.30. Flux de travail dans OligoR avec les principaux paramètres et les visualisations de résultats. Les données en entrée sont représentées en marron ; les données en pointillés sont des données prétraitées exportées sous forme de fichier .xlsx lors d'une session précédente. Les données HDX traitées sont composées de deux fichiers (spectres MS sélectionnés et déconvolués). Les figures ont été générées avec OligoR à partir des données d'HDX/MS du G4 télomérique humain 23TAG et de la cinétique d'interaction du K<sup>+</sup> avec ce même G4.

Au cœur de ce logiciel se trouvent des calculs simples, validés notamment par notre première publication sur le sujet (152). Ainsi, le nombre de sites échangeables d'un oligonucléotide est donné par l'Équation 5.10.

$$nX = 3 \times n_{dG} + 2 \times (n_{dA} + n_{dC}) + n_{dT} \quad (5.10)$$

La convolution de la distribution en deutérons, qui suit la loi binomiale (152), et de la distribution isotopique naturelle, est implémentée en utilisant une transformation de Fourier rapide (241).

Les centroïdes au temps d'incubation  $t$ , calculées comme la moyenne des  $m/z$  pondérée par leurs intensités, sont converties en nombre de sites non-échangés  $NUS$  en tenant compte du centroïde de

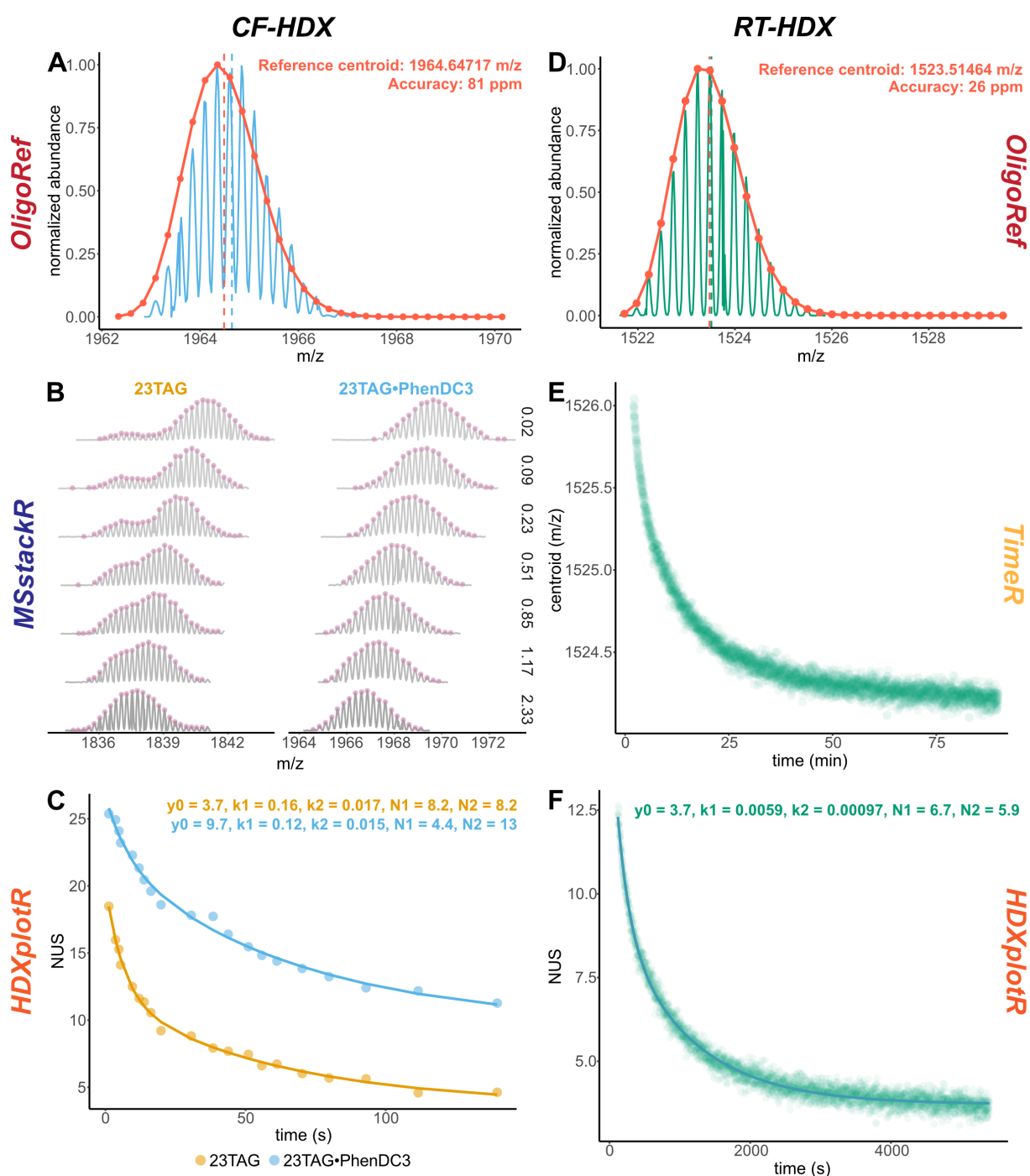


FIGURE 5.31. Exemples d'utilisation d'OligoR. À gauche : expérience CF-HDX d'un complexe 23TAG•PhenDC3 en présence de 23TAG libre, séparés en masse. A. Détermination empirique du centroïde de référence entièrement échangé (ligne bleue) et comparaison avec la théorie (ligne et points orange) pour le complexe 23TAG•PhenDC3 ( $C_{264}H_{303}O_{141}N_{100}P_{22}K_1$ ,  $z = 4^-$  avec 51 protons échangeables, 9%D). B. Spectres de masse collectés pour le 23TAG et le 23TAG•PhenDC3 pour différents points de temps de mélange. La sélection automatisée des pics est illustrée par des points roses. C. Cinétique d'échange apparente correspondante (points) et régression non linéaire avec l'Équation 5.12 (lignes). À droite : Expérience RT-HDX acquise pendant 90 minutes (temps mort : 40 s). D. Identique à A pour T30177-TT liant deux  $K^+$  ( $C_{190}H_{230}O_{119}N_{74}P_{18}K_2$ ,  $z = 4^-$  avec 43 protons échangeables, 9%D). E. Calculs des centroïdes pour chacun des 3867 scans. F. Cinétique d'échange apparente correspondante pour chaque scan (points) et régression non linéaire avec l'Équation 5.12 (lignes).

l'oligonucléotide entièrement échangé  $m/z_{\infty}$ , de l'état de charge  $z$ , des proportions de deutérium en solution avant ( $DC_0$ ) et après ( $DC_{\infty}$ ) et des masses atomiques des isotopes (Équation 5.11).

$$NUS_t = \frac{(m/z_t - m/z_{\infty}) \times |z|}{(DC_0 - DC_{\infty})(m_D - m_H)} \quad (5.11)$$

Les cinétiques d'échanges sont ensuite modélisées avec le nombre minimales de décroissances exponentielles (généralement une par ordre de grandeur de temps d'incubation), en utilisant l'Équation 5.12, où  $N_i$  et  $k_i$  sont le nombre de sites et la constante de vitesse de la décroissante  $i$ .

$$NUS_t = NUS_{\infty} + \sum_{i=1}^j (N_i \times e^{-k_i \times t}) \quad (5.12)$$

La déconvolution de distributions isotopiques est l'innovation la plus importante de ce travail. Plutôt que de se reposer sur l'utilisation de Gaussienne comme certains logiciels pour l'analyse de données HDX/MS de protéines, j'ai ici implémenté une optimisation tirant partie de la capacité d'OligoR à calculer des distributions isotopiques deutérées. L'algorithme minimise la différence d'intensité des pics  $I$  entre les distributions calculées (*model*) et expérimentales (*exp*) en modifiant la deutération de l'oligonucléotide (Équation 5.13).

$$S = \sum_i \Delta I_{i=1}^2 = \sum_i (I_{model(i)} - I_{exp(i)})^2 \quad (5.13)$$

Pour des distributions multimodales, l'optimisation est réalisée en ajustant la deutération des distributions calculées  $j$  à partir desquelles une seule intensité totale pour chaque pic est calculée (Équation 5.14).

$$I_{bimodal(i,j)} = \sum_{j=1}^n (I_{model(i,j)} \times ab_j) \quad (5.14)$$

La beauté de cette approche est que seule des distributions physiquement possibles peuvent être générées, car les calculs sont bornés par les quantités de deutérium expérimentales et dépendent directement de la séquence de l'analyte. Seuls des résultats *réalistes* peuvent donc être obtenus. En outre, comme la largeur des distributions est fixe pour une séquence et une deutération donnée, ce que reflète parfaitement OligoR, il est possible de détecter des distributions "cachées", c'est à dire deux distributions partiellement superposées qui apparaissent comme unique (voir les panneaux du bas de la Figure 5.32).

À l'inverse, dans les cas où une seule population est suffisante, les distributions bimodales vont tout de même donner de meilleurs résultats (entendre un RSS plus faible) en raison du plus grand nombre de paramètres. Cela générerait des faux positifs. Pour parer à cela, OligoR identifie automatiquement le modèle statistiquement meilleur par le calcul de la valeur  $p$  de la statistique  $F$ ,<sup>2</sup> définie dans

2. L'hypothèse nulle est que le modèle bimodal n'explique pas mieux la variance que le modèle monomodal



l'Équation 5.15, où  $df$  est le degré de liberté et  $mono$  et  $bi$  se réfèrent aux modèles mono- et bimodaux, respectivement (Figure 5.32).

$$F = \frac{\frac{RSS_{mono} - RSS_{bi}}{RSS_{bi}}}{\frac{df_{mono} - df_{bi}}{df_{bi}}} \quad (5.15)$$

Le développement du logiciel est pour le moment stoppé puisqu'il est fonctionnel et répond pleinement à nos besoins. Dans un futur proche, il sera nécessaire d'intégrer le traitement de données HDX-MS/MS pour assurer l'identification des fragments, la modélisation de leurs distributions isotopiques théoriques, la détermination de leur deutération en fonction du temps, et l'exploitation des cinétiques qui en résultent.

## 5.5 Base de données d'acides nucléiques G-quadruplexes

Les séquences d'ADN riches en guanine peuvent former des G4s (15, 243–249), et ces G4s ont une fâcheuse tendance à adopter une variété de conformations en équilibre dynamique (1, 25, 30, 144, 250). Cet équilibre est fragile. En particulier, il peut être facilement déplacé en changeant la nature et la concentration en sels (31). Le lecteur ayant fait l'effort d'atteindre cette page du manuscrit est certainement convaincu de tout cela.

Ce qui n'a pas vraiment été évoqué, et ce n'est pas un point de détail pour l'analyste, c'est que la forte dépendance de ces équilibres conformationnels aux conditions expérimentales rend difficile la comparaison de résultats obtenus avec différentes techniques analytiques. Dans le cas particulier de la spectrométrie de masse native, dont il est beaucoup question ici, l'utilisation de KCl, non volatil, est limitée à des concentrations beaucoup plus faibles qu'en conditions physiologiques. La concentration de référence pour les études de l'équipe de V. Gabelica est de 1 mM KCl, dans un tampon TMAA (100 mM) (24), même si nous avons occasionnellement infusé des solutions plus salées mais toujours de l'ordre du millimolaire (Section 5.3.1.2) (7). Ces concentrations sont deux ordres de grandeur inférieures à celles utilisées dans nombre de méthodes classiques pour étudier les structures et interactions de G4s (UV/Fluorescence/CD/FRET(-melting), RMN, SPR, ITC et bien d'autres), qui reflètent mieux les conditions physiologiques. Cela est problématique pour l'interprétation structurale des résultats de masse native, pour laquelle on est en droit de se demander si les conformères formés sont les mêmes que ceux déposés dans la PDB.

Nous avons donc souhaité établir une base de données de séquences formant des G4s, pour lesquelles il est établi que leurs structures dans le tampon "Native MS" (100 mM TMAA + 1 mM KCl) correspondent à celles de référence, c'est à dire résolues par RMN et déposée dans la PDB (26). Pour cela, Anirban Ghosh, un postdoctorant de l'équipe (désormais *Senior Staff Scientist* au CEITEC, République Tchèque), a fait l'acquisition des spectres de spectrométrie de masse native, de RMN du proton, de spectroscopie CD, et les courbes de dénaturation thermiques (par UV) en conditions *Native MS* et de référence (dépendantes des conditions expérimentales reportées dans la PDB). Le but a ensuite été de sélectionner

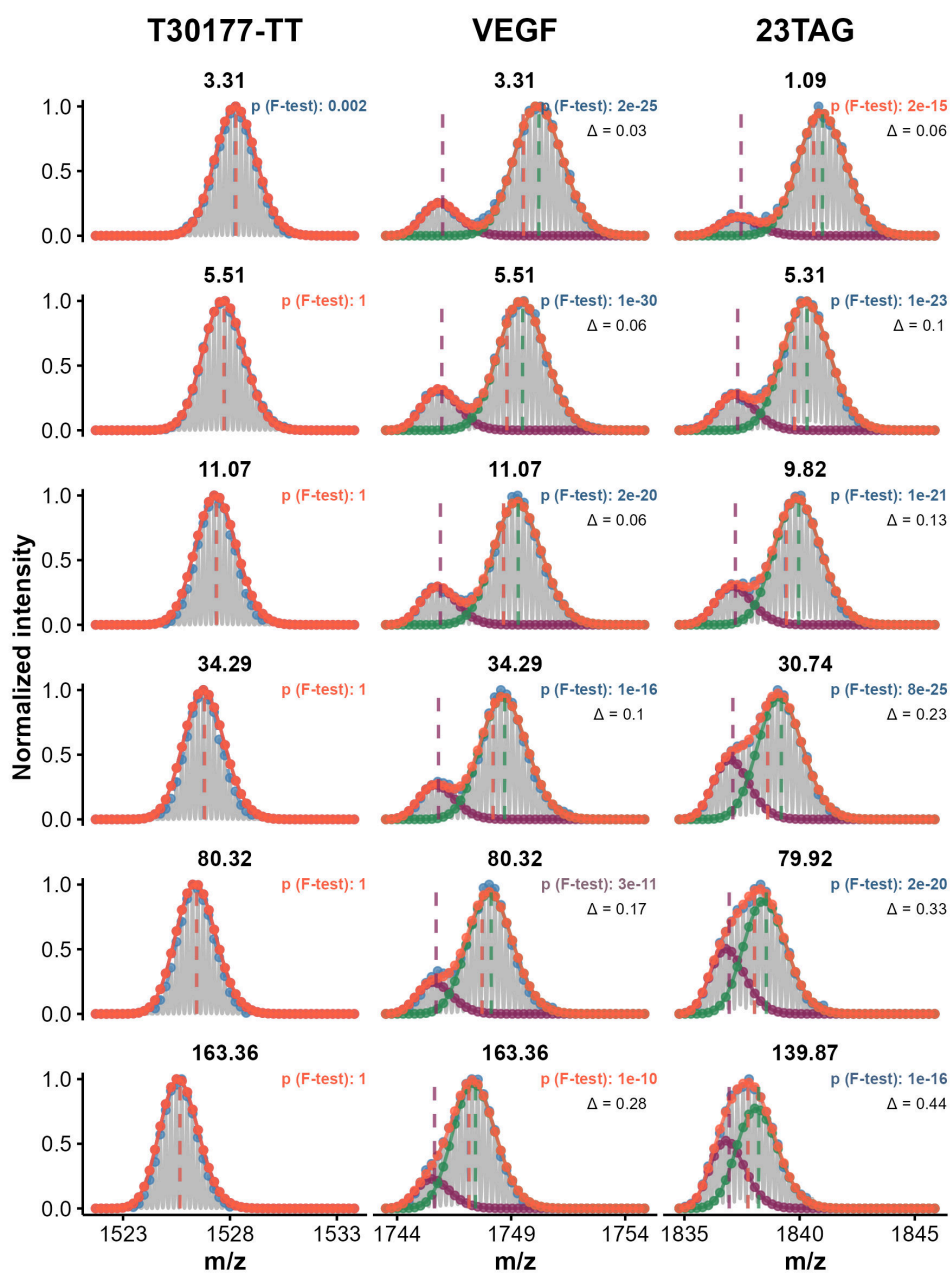


FIGURE 5.32. Exemple de modélisation de distributions isotopiques avec OligoR sur des distributions monomodales et bimodales, pour trois oligonucléotides ( $z = 4$ ,  $K^+ = 2$ ) et six temps d'incubation (en secondes, au-dessus des panneaux). Les données expérimentales sont en gris (*peak picking* en bleu), le *fit* global en orange, et les populations individuelles en vert et violet. Les lignes verticales en pointillés indiquent la position de leurs centroïdes respectifs. Dans le cas de 23TAG, l'abondance de la population isotopique fortement échangée augmente, ce qui suggère la présence d'une cinétique EX1 (5, 6, 224, 242). Au contraire, il n'y a pas de changement dans les abondances des populations isotopiques du VEGF, de sorte que la distribution bimodale résulte probablement de conformations distinctes en solution.

les séquences ayant des spectres RMN et CD similaires dans ces deux conditions (Figure 5.33), puis de publier une base de données comprenant l'ensemble de ces données expérimentales.

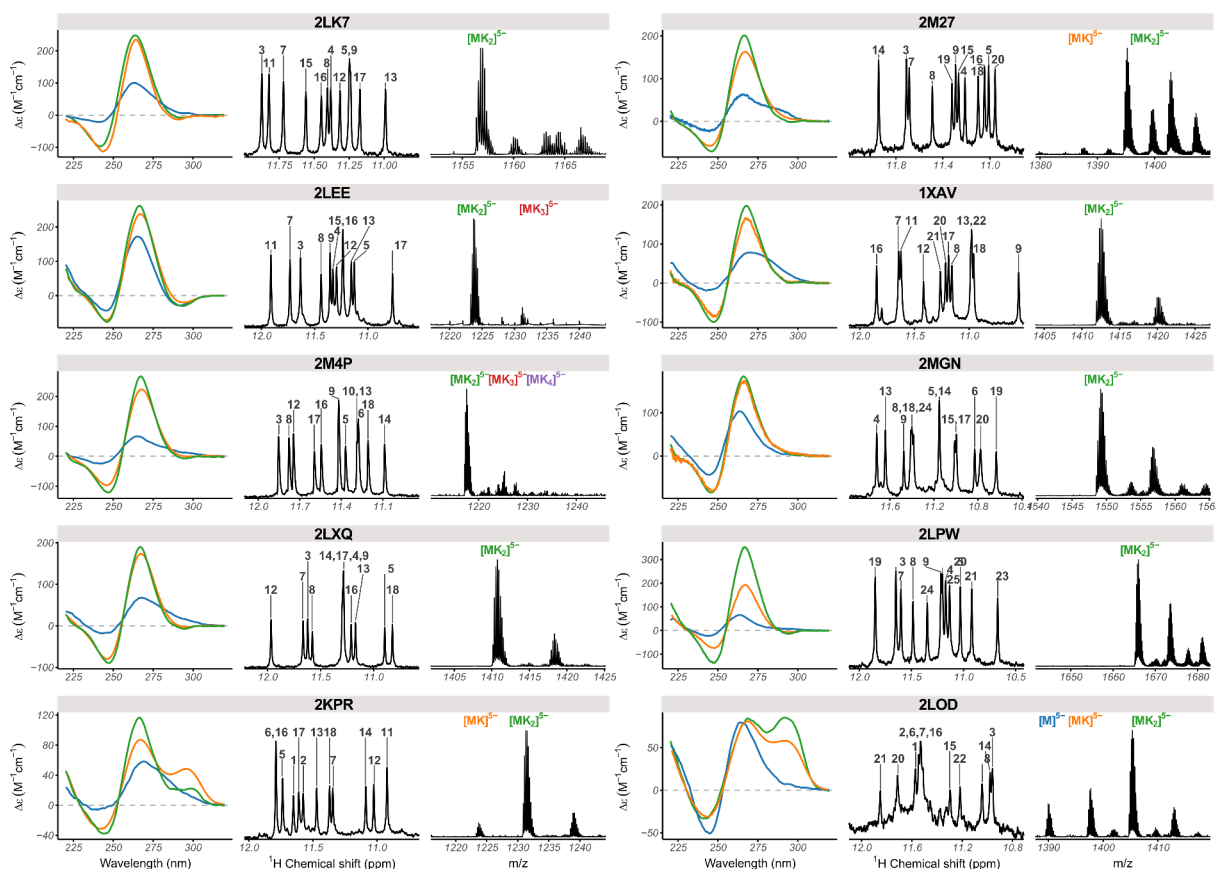


FIGURE 5.33. Exemples de spectres CD, RMN et de masse native pour dix oligonucléotides de la base de données. Les courbes CD sont colorées par concentration en KCl (vert : ~ 100 mM, rouge : 1 mM, bleu : 0).

Mes services ont été requis pour le mettre en place un traitement des données automatisé, fournissant des figures normées, et pour la mise en place de la base de données en ligne. Des activités dont je me suis principalement acquittées pendant le confinement, et qui ont également été le point de départ de meltR (Section 5.4.2).

Pour le traitement de données, j'ai développé une interface Shiny en R intégrant plusieurs fonctionnalités :

- Filtre des données par nom de la séquence, séquence, topologie, tampon, sel
- Attribution des pics des spectres de masse (automatisée) et de  $^1\text{H}$ -RMN (manuel)
- Calcul des coefficients d'extinction molaire à 260 nm  $\epsilon_{260nm}$  des oligonucléotides, par la méthode de Tataurov *et al.* (251), c'est à partir de leur séquence et de l'Équation 5.16, où  $\epsilon_i$  est le coefficient du nucléotide  $i$ ,  $\epsilon_{i,i+1}$  celui du doublet de nucléotide en position  $i; i + 1$  et  $N_b$  le nombre total de nucléotides. C'est cet algorithme que j'ai ensuite utilisé pour le projet Eps2Fold décrit en Section 5.4.1. Le logiciel associé dispose d'un onglet dédié pour le calcul des coefficients d'extinctions molaires (et de spectres d'absorbance) à partir de séquences oligonucléotidiques (Figure 5.34).

$$\epsilon_{260nm} = \sum_{i=1}^{N_b-1} \epsilon_{i,i+1} - \sum_{i=2}^{N_b-1} \epsilon_i \quad (5.16)$$

- Correction du blanc, des lignes de base, et conversion en ellipticités molaires  $\Delta\epsilon$  ( $M^{-1}cm^{-1}$ ) des données CD  $\theta$  (mdeg) avec l'Équation 5.17, où  $C$  et  $l$  sont la concentration molaire et la longueur du trajet optique, respectivement.

$$\Delta\epsilon = \frac{\theta}{32980 \times C \times l} \quad (5.17)$$

- Détermination des fractions structurées et des  $T_m$  à partir des courbes de dénaturation thermique, avec l'Équation 5.9. Il s'agit de l'embryon de ce qui deviendra meltR (Section 5.4.2).
- Réduction du bruit expérimental des spectres de masse
- Lecture/écriture de bases de données. La version finale est disponible sur Zenodo (218).
- Génération automatisée des figures et rapports [disponibles en ligne](#).

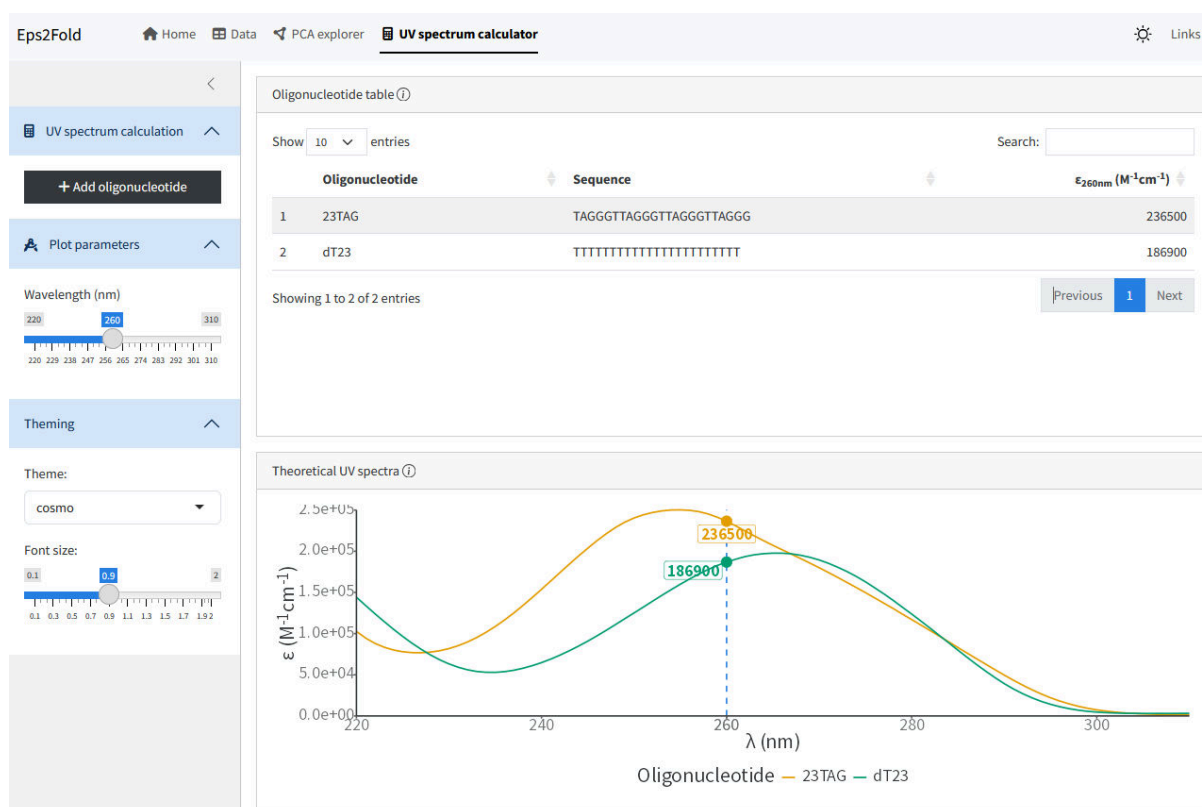


FIGURE 5.34. Copie d'écran du calculateur de spectres UV et coefficients d'extinction molaire intégré à Eps2Fold.

Au final, 28 séquences ont été retenues. La base de données [g4dbr](#) est disponible en ligne (Figure 5.35) et est utilisée, *a minima* par les membres de l'équipe. Plus de 7000 utilisateurs uniques se sont connectés du monde entier depuis sa mise en ligne.<sup>3</sup> Surtout, elle nous a permis de plus facilement et rapidement rationaliser nos travaux en termes de structure. Un exemple récent d'utilisation de cette base de données est présenté en Section 5.3.1.1.

3. Top 10 : Chine, USA, France, Japon, Italie, Inde, Singapour, Royaume-Uni, Allemagne, République Tchèque. Il est intéressant de voir que l'on retrouve des pays avec de petites populations mais actives dans la communauté G4.

PDB	Data	Description	Sequence	Length	Topology	Quartets	Diagram	Structure
2HY3		Human telomere DNA quadruplex structure in K <sup>+</sup> solution hybrid-1 form	AAAGGGTTAGGGTTAGGGTTAGGGAA	26	hybrid	3		
2JPZ		Human telomere DNA quadruplex structure in K <sup>+</sup> solution hybrid-2 form	TTAGGGTTAGGGTTAGGGTTAGGGTT	26	hybrid	3		

FIGURE 5.35. Copie d'écran de g4dbr ([https://ericlarg4.github.io/G4\\_database.html](https://ericlarg4.github.io/G4_database.html))

## 5.6 Mécanique quantique et moléculaire de biomolécules

### 5.6.1 Contexte

La thématique la plus récente que j'ai abordée, et la première initiée dans l'équipe PRISM, est l'implémentation de calculs de dynamiques moléculaires (MD) de structures d'acides nucléiques non canoniques, avec deux objectifs. Le premier, celui qui a amené l'énergie d'activation nécessaire à l'initiation de ce travail, est l'utilisation de dynamiques moléculaires contraintes (rMD), par les données RMN, comme dernière étape dans la résolution de structures d'acides nucléiques par RMN, avant dépôt sur la PDB. Le second objectif est l'utilisation de MD pour mieux comprendre, ou anticiper, des résultats expérimentaux. En particulier, je présente dans ce document des applications sur des complexes G4/petites molécules (Sections 5.3.1.1, ci-avant, et 5.6.4).

*In fine*, ces calculs de dynamique moléculaire ont nécessité la mise en place de calculs de mécanique quantique (QM), notamment pour dériver des paramètres modifiés de champs de force adaptés aux petites molécules et autres résidus non standards. D'autres utilisations de calculs QM incluent l'exploration de conformations de petites molécules (illustrer dans la Section 5.3.1.1), et l'étude d'interactions non-covalentes entre des nucléobases et des dérivés phényles servant de ligands de phase stationnaire chromatographique (Section 5.6.5).

Je me suis donc attaché à mettre en place des processus de calcul (Section 5.6.2) et de traitements de données (Section 5.6.2.2) aussi rapides et automatisés que possible, et pouvant tenir compte de contraintes RMN ou non. Le titre de ce manuscrit mentionne des structures non canoniques; on abordera en fait deux aspects distincts :

- Les structures secondaires *bizarres*, sans se limiter aux G4s. J'illustrerai cela avec l'exemple de la résolution de la structure d'un complexe aptamère d'ADN RKEC1-dopamine (Section 5.6.3)
- Les modifications de structures primaires, à l'image de ce que j'avais fait lors de mon séjour postdoctoral à l'Université de Colombie Britannique (Chapitre 2). C'est pour cette application que j'ai ajouté des calculs de QM en amont du flux MD. J'ai ainsi travaillé sur des doubles hélices hybrides ADN/ARN, dans lesquelles une des dT est modifiée en 2' par un hexose lié par une courte chaîne. Je ne présente pas ici ces résultats non publiés, mais ceux d'autres applications auxquelles j'ai étendu cette méthodologie (Sections 5.3.1.1, 5.6.5, 5.6.4).

J'ai également travaillé sur des structures de protéines, que j'élude ici.<sup>4</sup>

## 5.6.2 Implémentation GPU, automatisation et analyse de simulations MD

Les calculs de MD ont été effectués avec la suite logiciel Amber, en utilisant des champs de force... Amber. En réalité, le flux de travail que j'ai implémenté implique l'utilisation de plusieurs logiciels, certains inclus dans la suite Amber (Antechamber, Leap, pmemd, ccptraj) et d'autres non, notamment pour les calculs de QM en amont (Orca, Multifw), la visualisation et modifications des structures (PyMOL), les scripts de calculs et visualisation des résultats (R, Python), et la génération de rapports d'analyse interactifs, robustes et reproductibles (Quarto) (Figure 5.36).

### 5.6.2.1 Implémentation sur GPU

Les systèmes présentés ici sont composés d'une ou plusieurs chaînes ADN ou ARN (ou protéique) comprenant plusieurs dizaines de résidus. Ils doivent être simulés dans un solvant explicite et en présence de sels, impliquant la présence de plusieurs dizaines de milliers d'atomes (~ 40 000 pour le complexe RKEC1-dopamine). Pour des calculs de trajectoires courtes (de l'ordre de 10 ns), ce qui est suffisant pour un dépôt de structure sur la PDB, les calculs de rMD peuvent conceptuellement être effectués sur un CPU d'ordinateur de bureau. Par exemple, avec mon CPU 13th gen Intel Core i7-13700 2.10 GHz, le complexe RKEC1-dopamine est simulé avec des vitesses de l'ordre de la nanoseconde par jour de calcul. Plusieurs calculs peuvent être facilement lancés en parallèle en allouant un cœur CPU par structure, ce qui m'a permis de générer 15 structures en une dizaine de jours.

Le calcul sur un CPU au laboratoire est donc possible mais lent. Pour des trajectoires de l'ordre de la microseconde, plus conformes à l'état de l'art pour étudier des dynamiques (253, 254), l'utilisation de CPU grand public est irréaliste.<sup>5</sup> Fort heureusement, Amber bénéficie d'une implémentation CUDA qui permet des calculs bien plus rapides sur processeur graphique (GPU). J'ai donc installé cette implémentation sur les serveurs de calculs [DOREMI CALI v3](#) du [Mésocentre de Calcul Intensif Aquitain](#) de l'Université de Bordeaux. Cela me permet d'effectuer les étapes de dynamique moléculaire avec le module `pmemd.cuda` (v. 18.0.0) d'AMBER (253–255), sur un GPU NVIDIA H100 PCIe Tensor core GPU (version CUDA 12.4) qui, pour le système RKEC1-dopamine, simule environ 400 ns par jour (Figure 5.37). Pour les étapes de minimisation, peu coûteuses, je privilégie les calculs sur CPU, avec des codes en double précision. En effet, les GPUs reposent souvent sur un modèle à précision fixe située entre la simple et la double précision, et il est possible que des forces extrêmement élevées (par exemple celles résultant de *clash* atomiques, en particulier en début de calcul) provoquent des dépassements dans les calculs à virgule flottante (252). Tous les calculs de QM sont également implémentés sur CPU.

4. Papier soumis (et pas encore rejeté), à *Science*, avec Cameron Mackereth, en collaboration avec Yvan Huc, qui est le porteur de ce projet.

5. Il faudrait 2 ans et 9 mois pour une seule simulation d'une microseconde sur mon CPU de bureau.



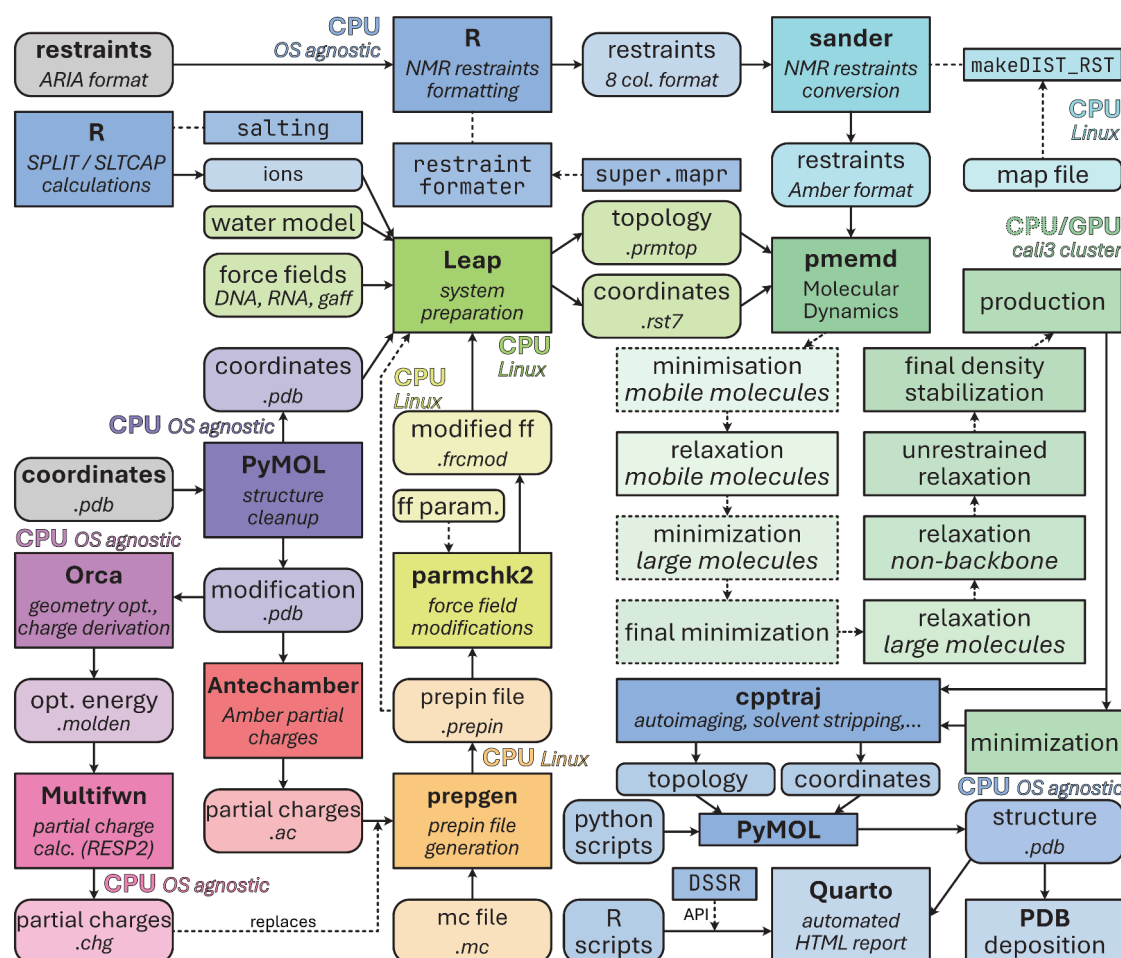


FIGURE 5.36. Flux de données pour les expériences de dynamique moléculaire avec Amber24, pouvant tenir compte de *i.* contraintes RMN et *ii.* modifications non-canoniques et ligands. Les coordonnées en entrées peuvent être le résultat de calculs consécutifs à des expériences RMN, ou tout autre source (pdb, structure “dessinées” manuellement). Après nettoyage de la structure dans PyMOL, les coordonnées peuvent être préparées dans Leap, où les champs de force ad hoc, le modèle de solvation et les ions sont ajoutés. Le nombre de ces derniers doit permettre de neutraliser les charges des acides nucléiques. Pour ajuster la force ionique, les calculs de nombre de cations et anions à ajouter par les méthodes SPLIT/SLTCAP ont été implémentés dans R. Dans le cas de modification des acides nucléiques, ou de ligands non décrits dans les champs de force d’acides nucléiques (OL21, OL3), des modifications au champs de force doivent être importées dans Leap. Pour cela, la modification est isolée dans PyMOL, coiffée par des atomes permettant de correctement simuler son environnement chimique, si nécessaire, puis sa géométrie est optimisée par mécanique quantique, ici dans le logiciel ORCA. Après calcul de l’énergie de la structure optimisée et convertie en .molden, les charges partielles des atomes sont calculées par la méthode RESP2 implémentée dans le logiciel Multifwn. Un fichier de charges partielles compatible avec Amber est généré avec Antechamber, puis les charges y sont substituées par les charges RESP2. Enfin, prepregen et parmchk2 sont utilisés pour générer les fichiers permettant à Leap de lire les modifications et de les implémenter pour les atomes de la modification, sans importer les atomes de la coiffe, comme défini dans le fichier .mc. Les fichiers de topologies et coordonnées générés dans Leap sont soumis aux différentes étapes de dynamique moléculaire avec pmemd. Une préparation robuste du système en 10 étapes (ici présentée de façon simplifiée) a été implémentée à partir du protocole de Roe et Brooks (252). Les étapes de relaxation et production MD utilise l’implémentation CUDA de pmemd que j’ai installé sur les serveurs de calculs DOREMI CALI v3 du Mésocentre de Calcul Intensif Aquitain de l’Université de Bordeaux. Les étapes de minimisation sont faites sur CPU à double précision pour éviter des *overflow* numériques si les forces mises en jeu sont élevées. Pour l’application de contraintes RMN, le fichier initial les décrivant est converti dans un format Amber via un script R, pour générer un fichier à 8 colonnes qui peut être lu et converti par le module pmemd au format Amber. Après production, une minimisation finale peut être effectuée si un ensemble doit être produit. Un pré-traitement des trajectoires est effectué en C++ avec cpptraj, puis les coordonnées finales sont converties en fichier pdb dans PyMOL. Les figures sont enfin générées avec des scripts Python et R dans un *notebook* Quarto permettant d’automatiser le traitement, la visualisation et la publication des données.

# GPU-accelerated AMBER molecular dynamics for the study of DNA aptamer structures

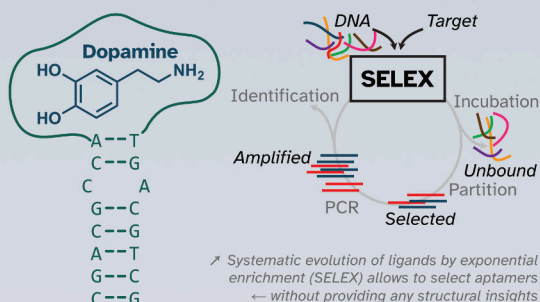
**Eric Largy, Cameron Mackereth**

ARNA - INSERM U1212 - CNRS UMR5320 - Université de Bordeaux



## Aptamers

- ◊ DNA or RNA strands folding in 3D structures allowing their binding to a molecular target with **high affinity** and **specificity**.
- ◊ Not designed *de novo* but **selected** against targets from large libraries of randomized sequences, providing **no insight into the structure** of aptamer-target complex
- ◊ Optimization/functionalization requires structural knowledge

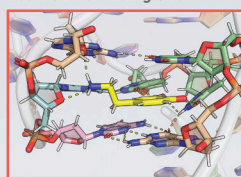


## Molecular dynamics (MD): GPU vs. CPU

- ◊ Simulation of DNA **dynamics** ± **NMR restraints** and energy **minimization**
- ◊ **AMBER force field** OL21 to model inter-atomic potential of DNA structures [3]
- ◊ **Explicit solvation with water** (OPC model with parameters for Na<sup>+</sup>, Cl<sup>-</sup> ions) [4]
- ◊ Simulations on 13th gen Intel Core i7-13700 2.10 GHz CPU: **10 ns@1 ns/day**
- ◊ Simulations on **CALI3** Nvidia H100 PCIe Tensor Core GPU: **1 μs@400 ns/day**

15 **minimized structures** by GPU-accelerated restrained MD →  
Dynamic bases adopt different configurations ↗

Zoom on the **binding site**



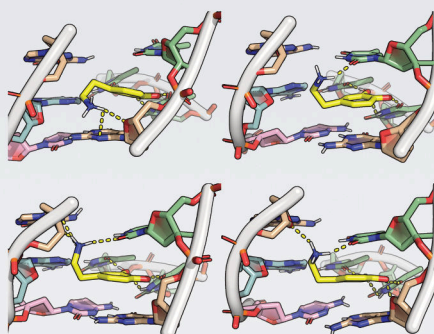
Bases are paired by **hydrogen bonds**

Formation of H-bonds during 1 μs simulations ± NMR restraints, and without dopamine →

Raw MD data analysed and visualised with a semi-automated R workflow

For each simulation frame, a vertical line is drawn if a bond exists

**Dopamine (LDP)** exhibits a stable binding configuration ↗



Different conformations of **dopamine** in absence of restraints.  
The **amino group** binds different bases during the simulation

**Sugar puckers** (an angle) can evolve if corresponding nucleotides change configuration

## Conclusions

- ◊ MD simulations of DNA aptamers implemented on the CALI3 cluster
- ◊ GPU enabled calculations yield excellent simulation speeds
- ◊ Access to microsecond scale MD results sheds light on DNA aptamer dynamics
- ◊ Data processed with R pipeline with automated HTML reporting

## References

- [1] ChemBioChem **2024**, e202409493
- [2] Science **2018**, 6412, 379
- [3] J. Chem. Theory Comput. **2023**, 19, 4299
- [4] J. Chem. Theory Comput. **2020**, 16, 4429

FIGURE 5.37. Poster présenté aux Journées Calcul et Données 2024

### 5.6.2.2 Analyse semi-automatisée des structures et trajectoires

Les trajectoires de MD génèrent des fichiers volumineux contenant une grande quantité d'informations exploitables. En aval de mon flux de production standardisé, j'ai développé une approche semi-automatisée pour le traitement et l'analyse de ces données. Compte tenu de la taille importante des fichiers, le flux d'analyse repose premièrement sur des scripts que j'ai écrits pour le programme cpptraj, écrit en C++ et beaucoup plus performant que les langages interprétés effectuant les calculs en mémoire, tels que R ou Python (Figure 5.36). Cette approche permet de générer des fichiers de trajectoire beaucoup moins lourds (par exemple sans solvant) et de calculer de nombreux paramètres diagnostiques (RMSD, respect des contraintes RMN) et analytiques (distances interatomiques, angles dièdres, etc.). Dans un second temps, je poursuis les calculs et produis les figures avec une combinaison de scripts R et Python (en particulier pour automatiser la production de figures dans PyMOL). Finalement, la production de rapports d'analyse automatisée avec Quarto est également garante de la robustesse de l'approche, et permet également l'inclusion d'éléments dynamiques et interactifs. Par exemple, pour la visualisation des trajectoires et structures, j'utilise Mol\*, un ensemble d'outils logiciels open source, accessible via le web (256). Un exemple de rapport pour le système RKEC1-dopamine décrit ci-après est [disponible en ligne](#).<sup>6</sup>

### 5.6.3 Exemple d'un complexe aptamère d'ADN RKEC1-dopamine

#### 5.6.3.1 Contexte

L'équipe de Phillip Johnson (York University, Toronto, Canada) étudie des structures d'ARN et leurs complexes avec des protéines en combinant la RMN avec des techniques de biologie moléculaire et biophysique. Récemment, Phillip s'est intéressé aux structures d'aptamères, dont on connaît souvent peu. Les publications dans lesquelles sont décrites la sélection de ces séquences ne présentent généralement que des structures dans le plan, déterminées par mFold, et pas nécessairement correctes (Figure 5.38). Hors, les aptamères sont souvent utilisés comme biocapteurs, ce qui requiert de comprendre les mécanismes d'interactions avec leur cible. Dans ce contexte, Phillip collabore avec notre équipe pour caractériser des aptamères d'intérêt, comme RKEC1 qui se lie à la dopamine (257).

#### 5.6.3.2 Préparation du système

Cameron Mackereth m'a fourni un set de 20 coordonnées du complexe RKEC1-dopamine, calculées dans ARIA par *simulated annealing* à partir de données RMN qu'il avait acquises (Figure 5.39A). De ces expériences, il a également déterminé des contraintes de distances. Le but pour moi était de mettre en place des calculs de rMD pour le dépôt de ces structures (nous en garderons 15) dans la PDB,<sup>7</sup> mais aussi pour que ce flux de travail soit réutilisable dans le futur pour d'autres systèmes. Les grandes étapes ont été les suivantes :

---

6. <https://ericlarg4.github.io/Dopamine>

7. <https://www.rcsb.org/structure/unreleased/9HIO>

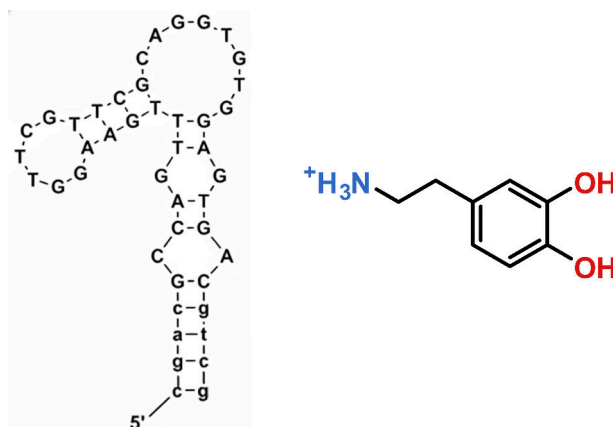


FIGURE 5.38. Structure secondaire, incorrecte, publiée de l'aptamère de la dopamine (gauche, (258)) et dopamine (droite).

1. Conversion des contraintes RMN du format ARIA vers un format à 8 colonnes, lisible par Amber, à l'aide d'un script R que j'ai développé. La fonction `makeDIST_RST` d'Amber a ensuite permis de générer le fichier de contrainte pour les simulations. De plus, un fichier de *mapping* a été préparé pour définir des noms communs pour les groupes de protons partageant une contrainte donnée (par exemple, les 3 atomes d'hydrogène du méthyl d'une thymine). Cette étape peut paraître triviale, mais sa semi-automatisation est garante de sa robustesse et un gain de temps significatif. Elle est également essentielle pour que la simulation ait lieu en tenant compte des données expérimentales. La figure Figure 5.39B illustre l'exhaustivité de ces contraintes.
2. Préparation du système à l'aide du programme `leap` de la suite Amber24 (259).
  - a. Utilisation de champ de force OL21 pour l'ADN, OL3 pour l'ARN (non utilisé pour le complexe de dopamine), complété par le champs de force général d'Amber pour les molécules organiques (GAFF; version 1.81), utile pour décrire les potentiels interatomiques des petites molécules (262). Dans la plupart des cas, je ne me suis pas limité à GAFF et ai dérivé moi-même, par QM, les paramètres de champs de force des petites molécules et résidus modifiés.
  - b. Solvation explicite dans une boîte octaédrique tronquée de molécules d'eau, en utilisant le modèle OPC avec les paramètres *ad hoc* Li/Merz des ions (ensemble 12-6) (264), avec un minimum de 14 Å entre le soluté et le bord de la boîte (Figure 5.39C). J'ai opté pour le modèle à 4 points OPC, plus coûteux que les modèles à 3 points classiquement utilisés comme TIP3P, car il reproduit de manière plus fidèle les propriétés thermodynamiques et structurales de l'eau liquide (densité, constante diélectrique), offrant ainsi une description plus réaliste des effets du solvant.
  - c. Neutralisation des charges et ajustement de la force ionique : pour le complexe de dopamine, j'ai souhaité simuler une solution contenant 140 mM de NaCl, pour mieux refléter les conditions de SELEX dans lesquelles l'aptamère a été sélectionné. Pour cela, j'ai utilisé la méthode SLTCAP (265), dans laquelle les nombres d'ions requis ( $N_{\pm}$ ; ici 35  $\text{Na}^+$  et 10  $\text{Cl}^-$ ) sont déterminés avec l'Équation 5.18, simplifiée par Machado and Pantano en l'Équation 5.19, où  $v_w$  est le volume d'eau de la boîte de simulation (environ  $3 \times 10^5 \text{ Å}^3$  ici) en unités réduites,  $c_0$  la concentration en sel,  $Q$  la charge du système (ici, 25), et  $N_0 = \frac{N_w \times c_0}{55.5}$ , avec  $N_w$  le nombre de molécules d'eau dans la boîte (ici, environ  $9 \times 10^4$ ) (266). Un point important ici est que la méthode SPLIT décrite



par Machado et Pantano ne peut pas être appliquée car notre système ne satisfait pas à la condition  $N_0 \gg Q$ ; cependant, elle donne des valeurs identiques dans la plupart des cas ou ne diffère que d'un seul ion.

$$N_{\pm} = v_w c_0 e^{\mp \text{ArcSinh}(\frac{Q}{2v_w c_0})} \quad (5.18)$$

$$N_{\pm} = N_0 \sqrt{1 + (\frac{Q}{2N_0})^2 \mp \frac{Q}{2}} \quad (5.19)$$

3. Préparation du système pour l'étape de production. Il est souvent nécessaire d'effectuer une série d'optimisations et/ou de simulations MD préparatoires afin de garantir la stabilité des simulations de production ultérieures, en particulier en présence d'un solvant explicite. Je me repose ici sur les travaux de Roe et Brooks, qui proposent un protocole simple, en dix étapes, applicable à une grande variété de systèmes (252). Le protocole consiste en une série d'optimisations d'énergie et de relaxations progressives (4000 étapes de minimisation et 40 000 étapes de dynamique moléculaire, soit 45 ps au total). Le système est divisé en molécules *mobiles*, qui diffusent rapidement dans le système (eau, ions), et *lourdes*, plus lentes à diffuser. Les premières sont relaxées avant les secondes en appliquant des contraintes positionnelles sur ces dernières. De mêmes, les bases et chaînes latérales sont relaxées avant le squelette afin de permettre la relaxation de contacts atomiques sans perturbation des éléments de structure secondaire.
4. Production de 10 ns à 1 microseconde, en fonction des besoins, de MD à pression constante, en appliquant les contraintes RMN pondérées raisonnablement.
5. Minimisation finale (pour dépôt PDB) des points finaux de rMD, avec la même pondération des contraintes qu'en étape 4.

J'ai volontairement omis tous les paramètres expérimentaux des simulations. L'approche présentée ici est générale, robuste et reproductible, ce qui me permet désormais de traiter différents systèmes de manière cohérente et unifiée.

### 5.6.3.3 Analyse des résultats

Je présente ici quelques exemples d'exploitation typique des résultats, afin d'illustrer le flux de travail mis en place. Par exemple, il est courant d'analyser l'évolution temporelle de la RMSD pour évaluer la stabilité d'une simulation de dynamique moléculaire (Figure 5.41A). Ce type de représentation montre les fluctuations structurales de chaque conformère par rapport à la conformation initiale. Toutefois, ces figures peuvent prêter à confusion : deux conformères présentant une même valeur de RMSD ne sont pas nécessairement similaires. Ils sont simplement à une distance comparable du modèle de référence, mais peuvent différer considérablement l'un de l'autre. Pour pallier à ce biais de visualisation, je privilégie les représentations sous forme de matrices de similarité (des *heatmaps* dites *pairwise RMSD*; Figure 5.41B), où chaque conformation est comparée à toutes les autres, offrant une vision plus complète de l'espace conformationnel échantillonné pendant la simulation. Ces figures permettent notamment de repérer facilement la formation de clusters structuraux au cours de la simulation, visibles par l'apparition de zones carrées le long de la diagonale.

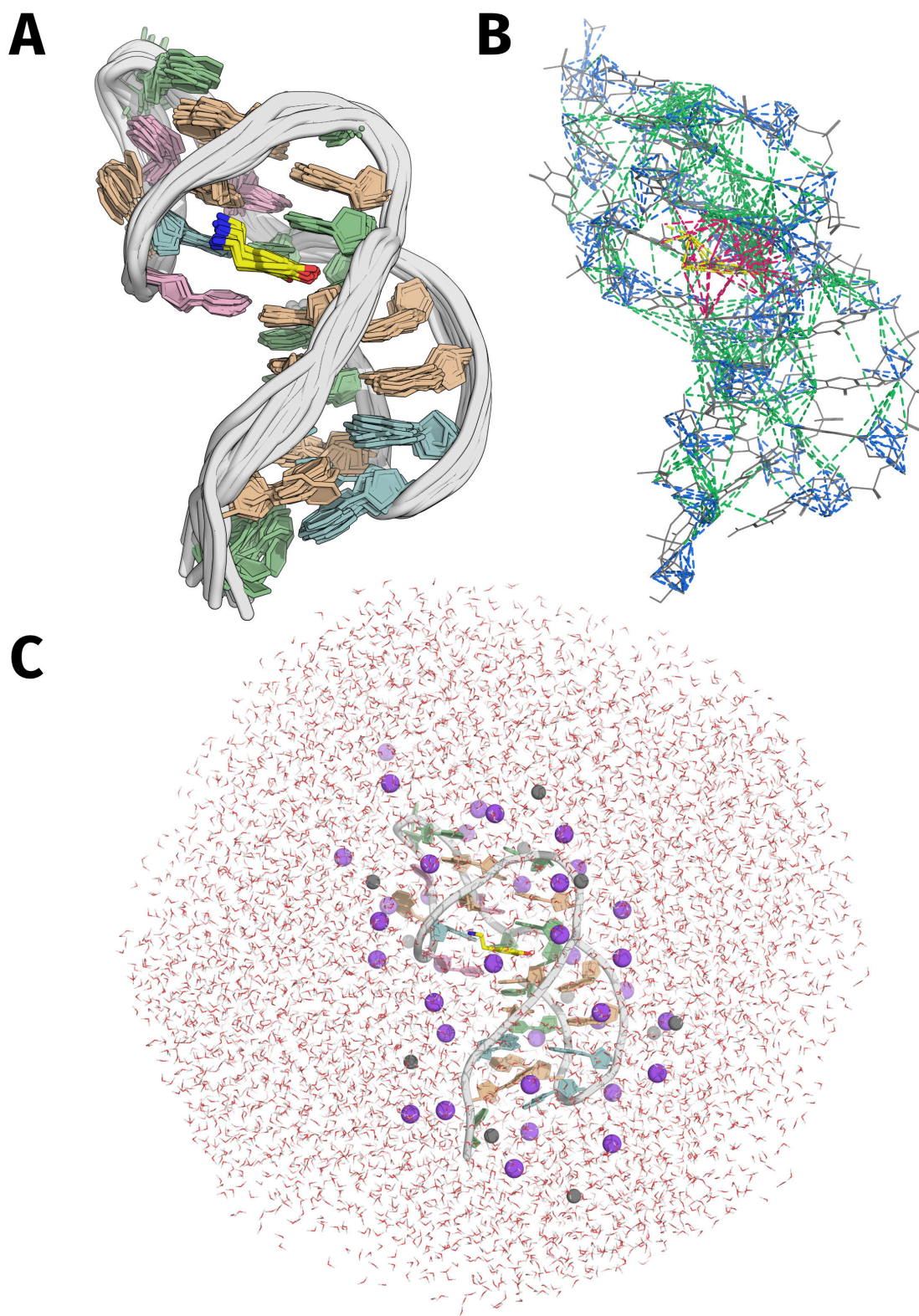


FIGURE 5.39. Préparation du système RKEC1-dopamine. A. Ensemble de 20 structures obtenues par *simulated annealing*. B. Réseau de contraintes de distances intra (bleu) et inter-résidus (vert), intra-dopamine (jaune) et entre dopamine et RKEC1 (rose) converties au format Amber puis visualisées dans PyMOL avec un script Python. C. Solvation explicite du système, dont la force ionique est ajustée avec des ions sodium (violet) et chlorure (vert).



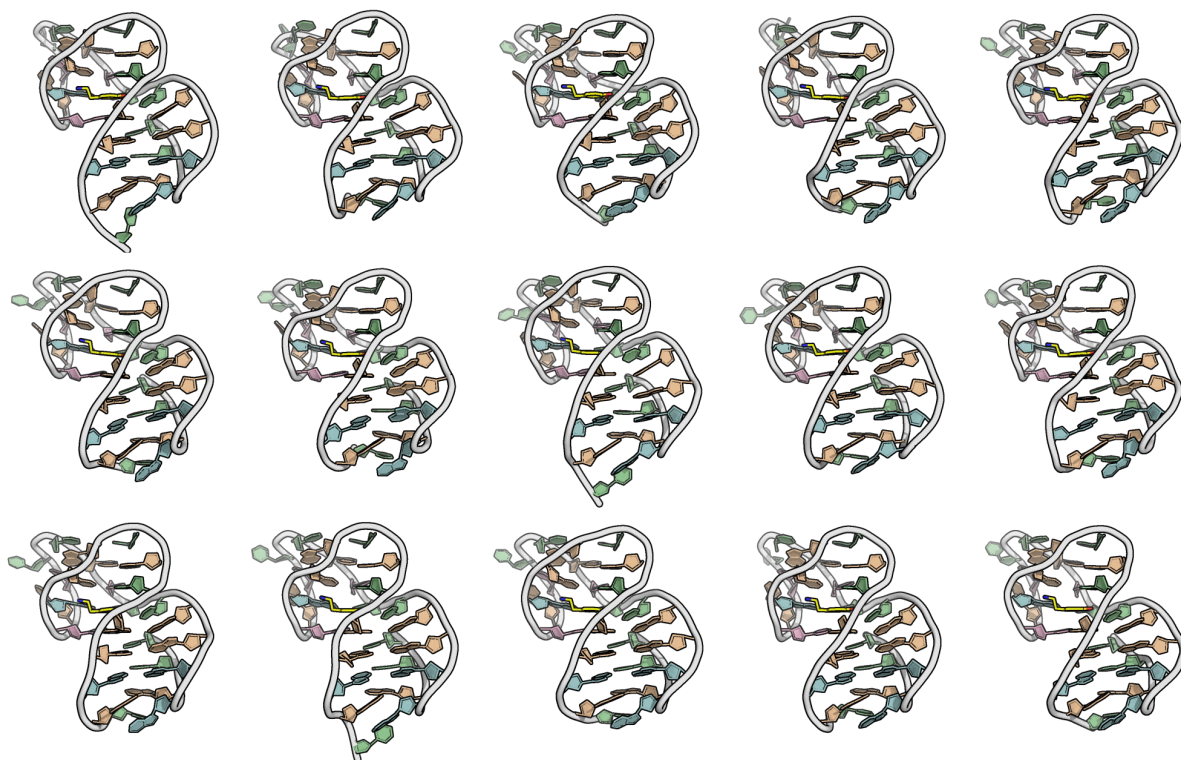


FIGURE 5.40. Ensemble de 15 conformations du complexe RKEC1-dopamine déposé sur la PDB.

Dans le contexte de l'étude de structures non-canoniques, il est fréquent de mesurer des angles dièdres, notamment celui des liaisons glycosidiques  $\chi$  (correspondant au GBA dans la Section 5.4.1; Figure 5.42) et de torsion des sucres (*pucker*  $P$ ; Figure 5.43). J'ai développé des fonctions en R qui permettent de rapidement calculer chaque angle pour chaque nucléotide pour chaque set coordonnées extrait de la trajectoire (ici, 25000; Figures 5.42, 5.43). En particulier,  $P$  et son amplitude ont été calculés avec les équations 5.20 et 5.21, respectivement, en utilisant les angles de torsion  $v_i$  (atomes définis dans les équations 5.22–5.26).

$$\tan P = \frac{(v_4 + v_1) - (v_3 + v_0)}{2v_2(\sin(\frac{\pi}{5}) + \sin(\frac{2\pi}{5}))} \quad (5.20)$$

$$\theta_M = \frac{v_2}{\cos P} \quad (5.21)$$

$$v_0 = C4' - O4' - C1' - C2' \quad (5.22)$$

$$v_1 = O4' - C1' - C2' - C3' \quad (5.23)$$

$$v_2 = C1' - C2' - C3' - C4' \quad (5.24)$$

$$v_3 = C2' - C3' - C4' - O4' \quad (5.25)$$

$$v_4 = C3' - C4' - O4' - C1' \quad (5.26)$$

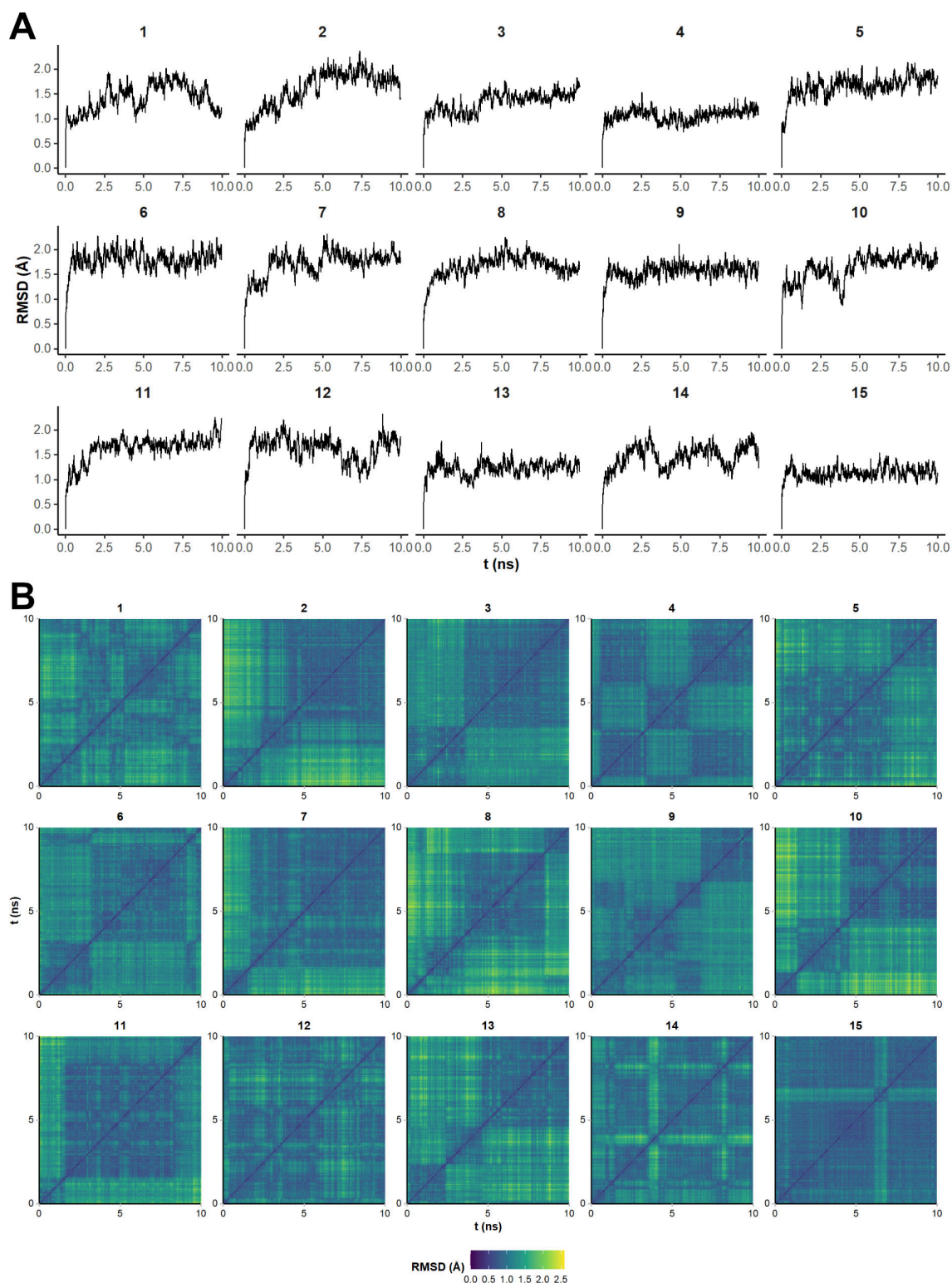


FIGURE 5.41. A. RMSD unidimensionnelle sur tous les résidus (y compris les extrémités 5' et 3'), sans hydrogène, pour l'ensemble des simulations contraintes (10 ns), la première structure étant utilisée comme référence. B. *Pairwise* RMSD (RMSD calculée pour toutes les paires de coordonnées) pour les quinze simulations. Cette représentation permet de mieux saisir l'étendue de la diversité structurale au cours de la simulation, de repérer des *clusters* (carrés le long de la diagonale), et d'éviter le biais d'interprétation.

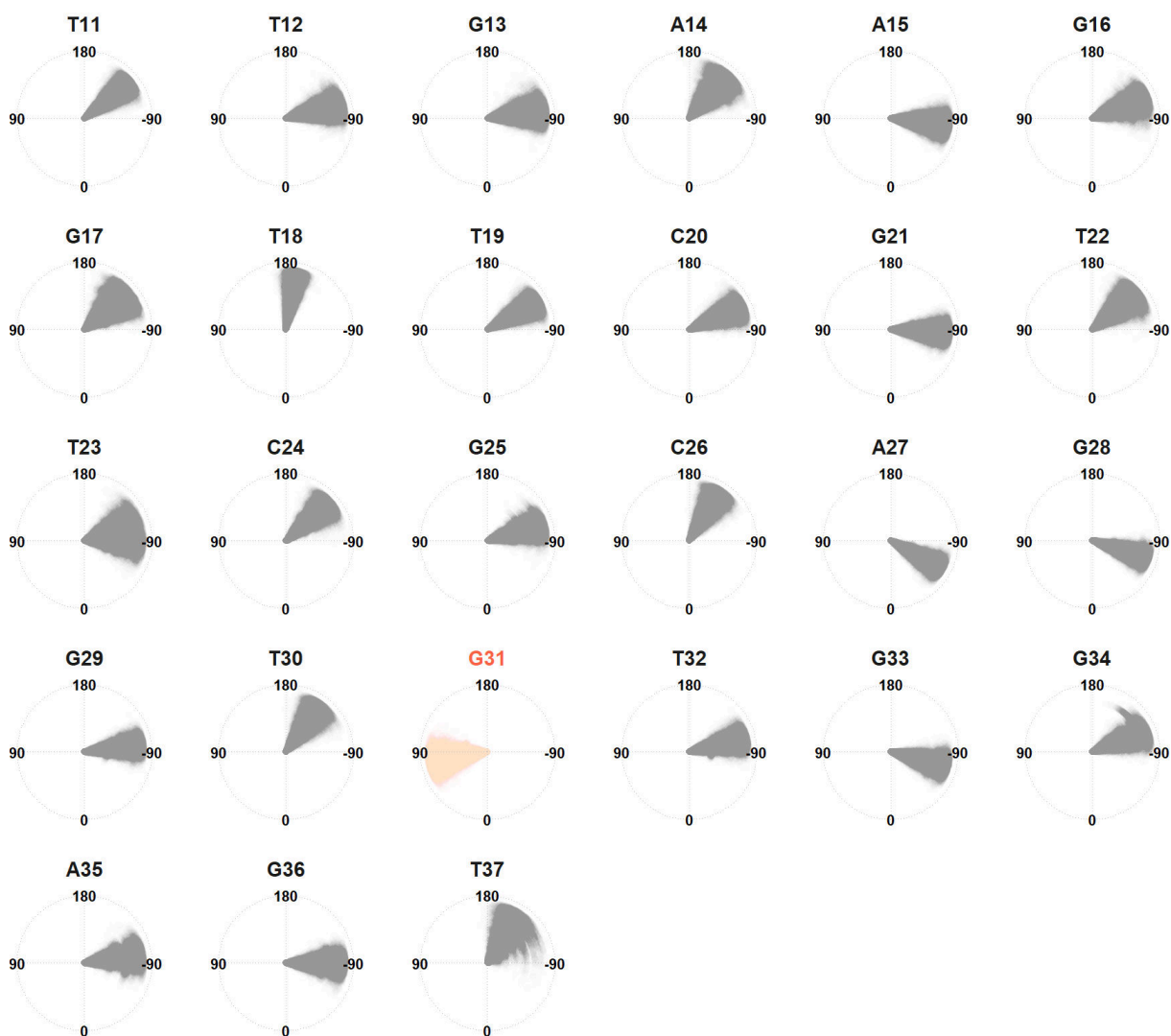


FIGURE 5.42. Angle dièdre de la liaison glycosidique  $\chi$  de chaque nucléotide au cours de la simulation d'une microseconde (de l'intérieur vers l'extérieur), exprimé entre -180 et +180 degrés. On note ici que la guanine 31 est en conformation *syn*, alors que tous les autres nucléotides sont en *anti*, qui est la conformation attendue dans l'ADN duplex. Ces angles sont globalement stables pendant la simulation, ce que nous avons également observé en absence de contraintes de distance, preuve de la robustesse de la structure calculée.

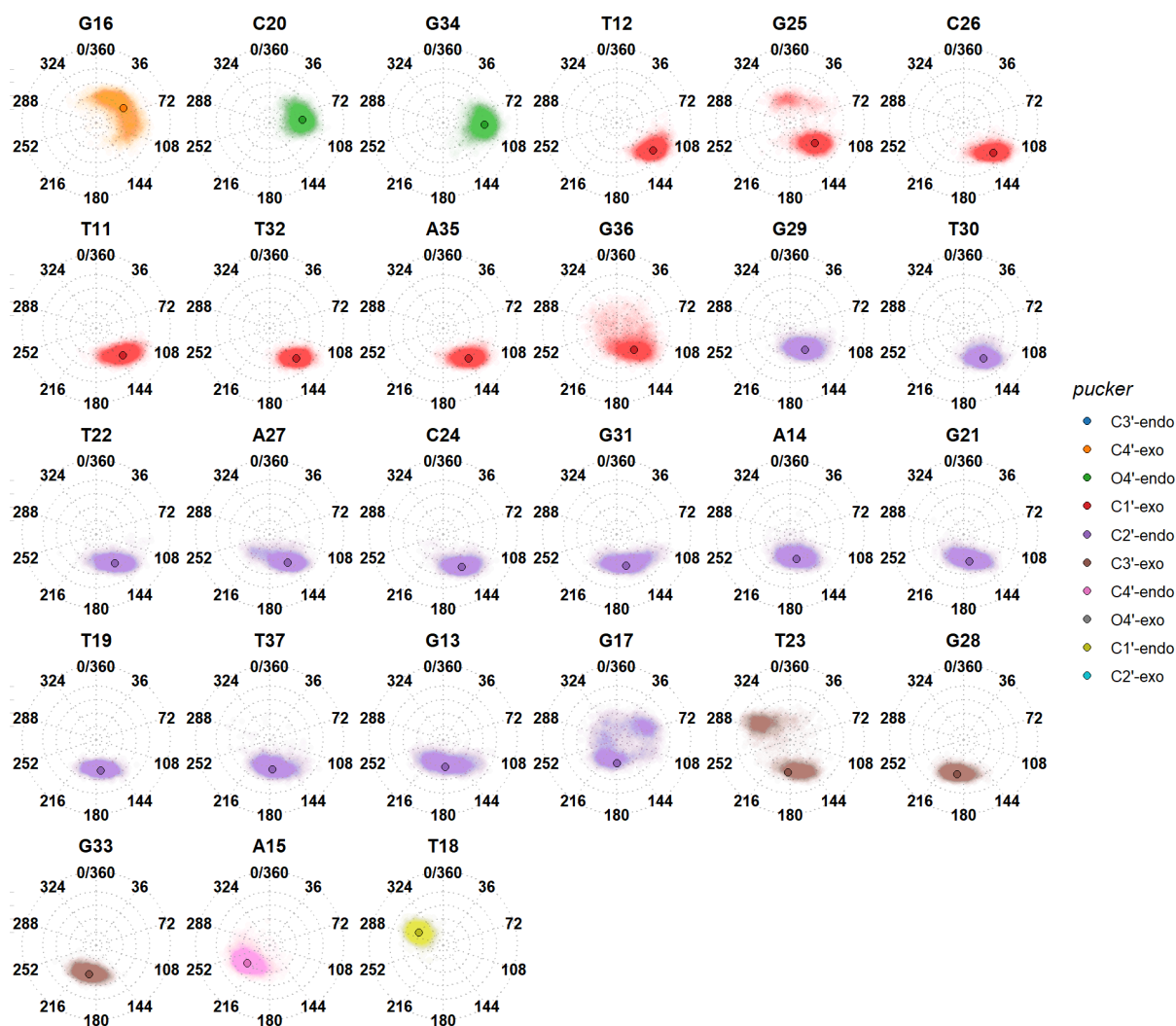


FIGURE 5.43. *Puckering P* des sucres calculé avec la méthode d'Altona & Sundaralingam (267), représenté entre 0 et 360 degrés. Chaque point représente l'une des 25000 coordonnées extraites de la simulation d'une microseconde. La position des points par rapport au centre des cercles est fonction de l'amplitude  $\theta_M$  (plus l'amplitude est grande, plus le point est vers l'extérieur). La médiane est entourée en noir. L'ADN a normalement des *puckers* C2'-endo, et l'on peut considérer ici que tous les nucléotides colorés en marron, violet et rouge ont une configuration glycosidique canonique. Un exemple de *pucker* non canonique est T18, qui forme une liaison hydrogène avec la dopamine. Sa configuration en C1'-endo est très rare, mais a été observée autour d'intercalants et pour que la structure gagne en flexibilité (268, 269).



Petite digression : la détermination d'angles dièdres, notamment  $\phi$  et  $\psi$ , est également d'intérêt pour la validation de structures protéiques. Cette validation repose sur l'utilisation de serveurs web qui comparent les valeurs soumises à un set de référence d'angles "autorisés", dont la provenance n'est pas toujours claire (Figure 5.44A). La mise en place d'analyse "maison" permet de sélectionner les données et leur analyse plus convenablement, ce que j'ai mis en place au laboratoire (Figure 5.44B).

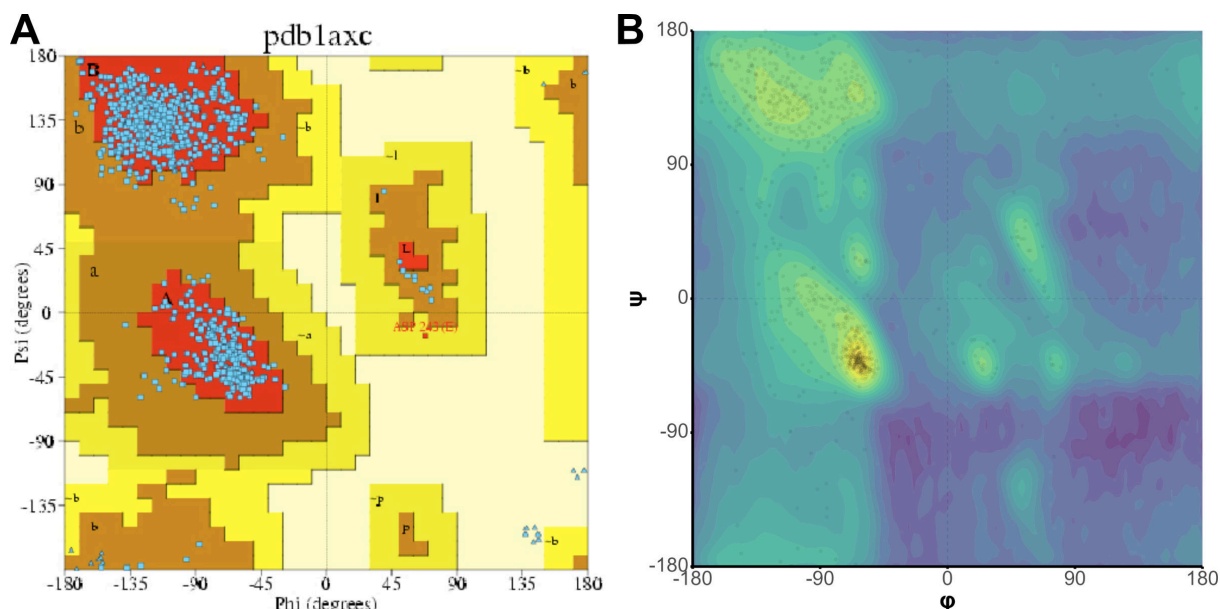


FIGURE 5.44. Diagrammes de Ramachandran produits par PROCHECK (A) et par l'auteur avec un script en R (B), sur un ensemble de 20 complexes protéiques. Chaque point correspond à une valeur expérimentale. Les zones colorées représentent les zones autorisées, qui sont dérivées de données de références. Pour B, j'ai utilisé le jeu de données *top2018* du laboratoire Richardson de l'Université Duke, déposé sur Zenodo (270). Il contient des résidus de haute qualité issus de chaînes protéiques également de haute qualité, et de faible redondance dans la PDB (271).

Un autre aspect très important de l'étude structurale est la mesure de distances interatomiques, notamment entre donneurs et accepteurs de liaisons hydrogène. Pour cela, j'ai développé un script permettant de visualiser la formation plus ou moins stable de liaisons hydrogènes sur des trajectoires entières, en fonction de critères géométriques plus ou moins strictes (angle et distance; Figure 5.45). On note ici que trois liaisons hydrogène très stables sont formées au site d'interaction avec la dopamine, et que ces interactions sont conservées en absence de contraintes.

Un corollaire à ces considérations est la caractérisation de l'appariement de bases. Ici, nul n'est besoin de réinventer la roue. En fait, la roue a déjà été inventée plusieurs fois, et j'ai retenu trois classifications pour le traitement automatisé :

- Saenger a défini une liste de 28 paires de base (plus une, plus tard), identifiées par des chiffres romains (272, 273). Une paire canonique Watson-Crick est XIX pour GC et XX pour AT.
- Leontis et Westhof (274), puis Lemieux et Major (275), on décrit les paires de base en fonction de *i*. l'établissement de liaisons H sur les faces Watson-Crick (W), Hoogsteen (H), ou sucre (S) et *ii*. l'orientation de la liaison glycosidique (*syn/trans* : c/t) des nucléotides. Une paire canonique est donc définie comme cWW.
- La classification DSSR, qui est similaire à la précédente, mais indique d'abord l'information c/t suivi par la faces interagissant définies comme Watson-Crick (W), grand sillon (*major*; M,

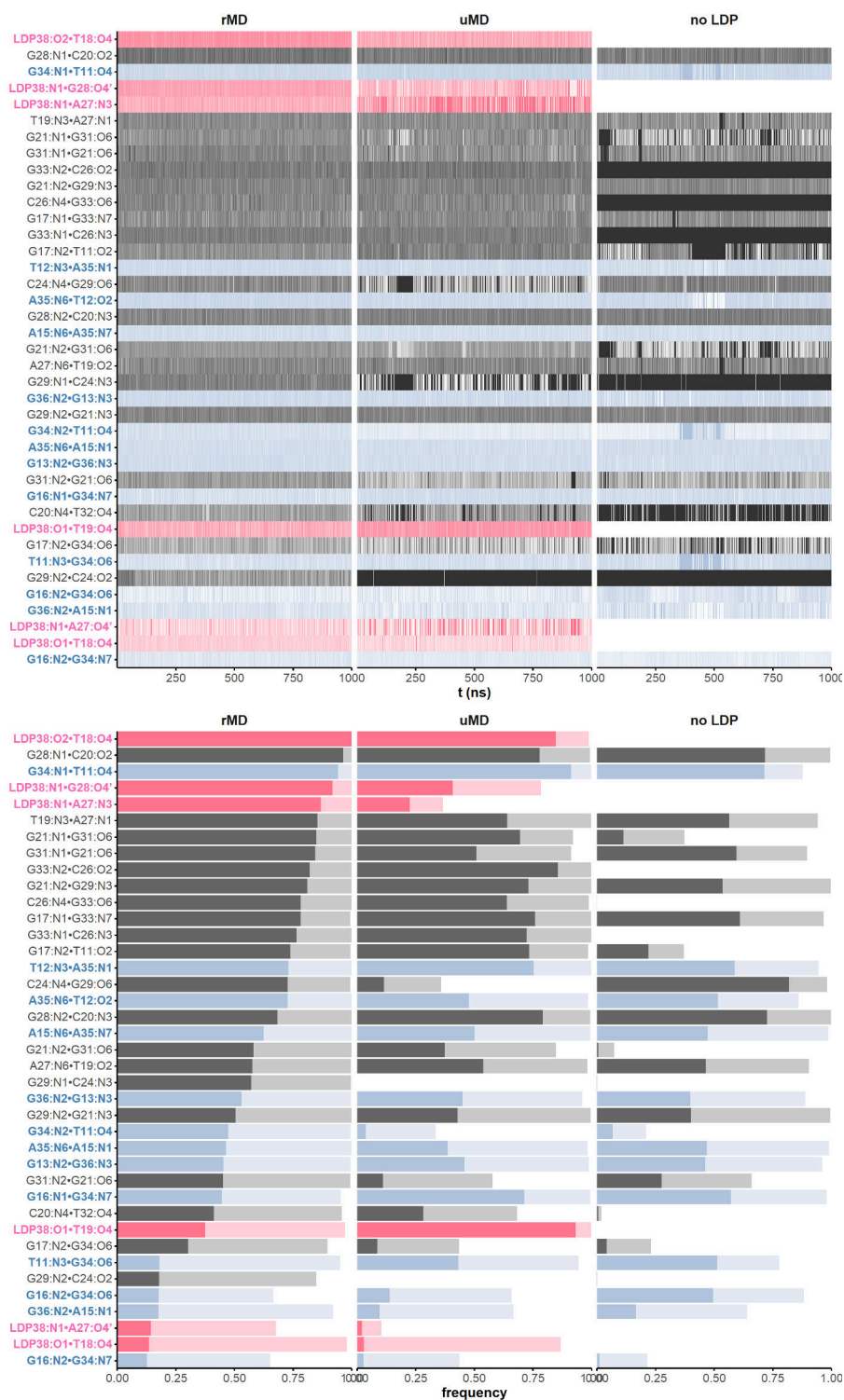


FIGURE 5.45. Formation de liaisons H au cours des simulations avec contraintes (*rMD*), sans contraintes (*uMD*) et en absence de dopamine (*no LDP*). La liste des liaisons H a été établie sur les structures minimisées des simulations *rMD*, avec une distance donneur-accepteur maximale de 3,1 Å, en conservant celles formées dans au moins 20 % des structures. Les liaisons H dans les régions 5'- et 3' (résidus 11:17 et 33:37) sont colorées en bleu, et celles impliquant la dopamine en rose. En haut : chaque ligne est un point de temps (un point toutes les 0,04 ns); plus une ligne est foncée, plus la liaison H est courte et linéaire (la transparence suit linéairement le rapport angle/longueur de la liaison). En bas : Fréquences cumulées. La fréquence de formation des liaisons est indiquée pour des seuils moins strictes (3,5 Å et 120°; couleurs claires) et plus stricts (3,0 Å et 140°; couleurs foncées). Seules les liaisons H inter-résidus dont la fréquence « stricte » est supérieure à 0,10 sont représentées.



correspondant à H ci-dessus) et petit sillon (*minor*; m, correspondant à S ci-dessus), entre lesquelles est intercalé + ou - pour indiquer si les brins sont parallèles ou antiparallèles (277). Une paire canonique est alors définie comme cW-W.

TABLE 5.1. Résultats de la caractérisation automatisée des paires de base, ici pour une seule structure.

ID	nucleotide 1	nucleotide 2	Name	Saenger	Leontis-Westhof	DSSR
1	T11	G17			cWS	cW-m
2	T11	G34	rWobble	XXVII	tWW	tW+W
3	T12	A35	rWC	XXI	tWW	tW+W
4	G13	G36		IV	tSS	tm+m
5	A15	A35		V	tWH	tW-M
6	G16	G34		VII	tWH	tW-M
7	G17	G33			tWH	tW-M
8	G17	G34	~rHoogsteen		tWH	tW-M
9	T18	G33			tWH	tW-M
10	T19	A27	rWC	XXI	tWW	tW+W
11	C20	G28	rWC	XXII	tWW	tW+W
12	C20	T32			tW.	tW-.
13	G21	G29		IV	tSS	tm+m
14	G21	G31			tWW	tW+W
15	T22	T30			tHH	tM+M
16	C24	G29	WC	XIX	cWW	cW-W
17	G25	G28	~Sheared		tSH	tm-M
18	C26	G33	WC	XIX	cWW	cW-W
19	G28	T32			cSW	cm-W

Ce que l'on peut humblement en retenir, c'est que (presque) tout a déjà été publié il y a au moins quarante ans, et qu'il suffit souvent de changer légèrement le vocabulaire pour redonner vie aux mêmes idées. Le lecteur attentif aura sans doute remarqué mon goût prononcé pour le recyclage de méthodes antérieures à ma propre naissance, comme en témoigne récemment mon intérêt pour l'HDX (mais il n'y avait pas de MS dans son implémentation initiale). Plus sérieusement, Lu *et al.* ont gentiment mis à disposition un serveur de calcul de ces classifications, ce qui m'a permis d'interroger automatiquement ledit serveur via l'API, en exploitant la librairie `httr2` (Table 5.1) (278).

Ici, le complexe RKEC1-dopamine est remarquablement non canonique. Seules deux paires de bases respectent la géométrie Watson-Crick (Figure 5.46A). En plus des paires de bases non canoniques (Figure 5.46B), plusieurs triplets de bases, pas forcément coplanaires, sont formés, rendant la structure franchement *bizarre*.

Finalement, la Figure 5.46C,D présente le site d'interaction de la dopamine, où celle-ci établit des liaisons hydrogène avec les T18 et T19 *via* les O1 et O2 de son cycle catéchol, tandis que l'ammonium interagit avec le N3 de l'adénine 27 et le sucre de G28. Cette interaction est nécessaire à la structuration

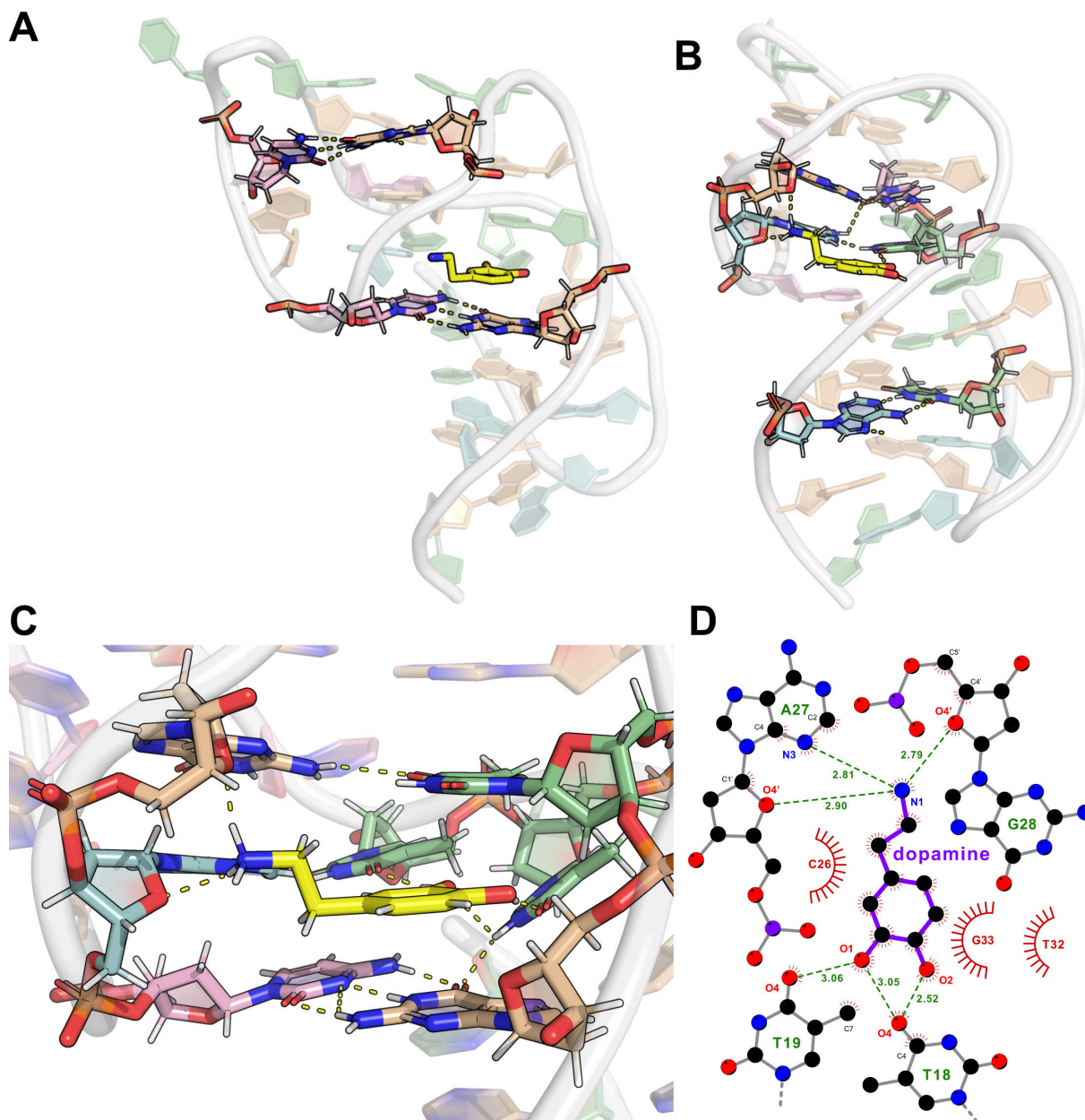


FIGURE 5.46. A. Les deux seules paires de bases canoniques “WC” B. Exemples de paires de bases non canoniques “rWC”. C. Site d’interaction de la dopamine. D. Vue projetée dans le plan avec LigPlot\* (279). Le ligand est en violet, les résidus en gris, les carbones, azotes et oxygènes en noir, bleu et rouge, respectivement, les liaisons hydrogènes en vert pointillés et les interactions hydrophobes sont symbolisées par des traits rouges.

de RKEC1, ce qui aura des implications sur la prédiction de cette structure. D'autres types d'analyse, notamment l'utilisation de la PCA, sont illustrés par ailleurs (Sections 5.3.1.1, 5.6.4).

#### 5.6.3.4 Considérations sur la prédiction de structures

Ce manuscrit présente de nombreuses méthodes permettant de caractériser des structures d'acides nucléiques, leurs dynamiques et leurs interactions. Un sujet que je n'aborde pas, du fait de mon incompétence totale sur le sujet (pour le moment), est la prédiction de structures à partir d'une séquence, comme le permet AlphaFold par exemple. Avec l'expérience, on peut prédire si une séquence peut former un G4, et même proposer une topologie en suivant des règles simples,<sup>8</sup> mais cela relève plus de la divination que de la science. Et quid de structures vraiment bizarres comme le complexe RKEC1-dopamine ?

Dans ce contexte, nous avons soumis au CASP16 (*16th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction*) la structure du complexe RKEC1-dopamine, que de nombreux groupes à travers le monde ont tenté de prédire à partir de la séquence. Des 107 modèles soumis, aucun ne s'est approché de la vérité (RMSD > 10 Å) (280). Nous pouvons avancer deux explications. La première est le manque de diversité des structures d'ADN déposées dans la PDB. Le second problème est lié au fait que RKEC1 ne se structure qu'en présence de dopamine, mais il semble que de nombreux groupes aient calculé la structure en absence du ligand, avant de tenter d'intercaler celui-ci. Ces considérations montrent bien la nécessité du travail expérimental sur les aptamères que nous avons entrepris.

#### 5.6.4 Application à des complexes G4/ligands.

Je présente brièvement ici une application récente du flux de travail MD développé dans l'équipe PRISM, une autre étant présentée dans la Section 5.3.1.1. Cette application s'inscrit dans un projet mené par Samir Amrane (ARNA; Figure 5.47), qui développe des petites molécules possédant des propriétés antivirales (HIV-1, HCV or HPV) grâce à leur affinité pour les G4s. Dans l'étude qui nous intéresse ici, une nouvelle classe de ligands dibenzoacridinium (DBA) a été synthétisée et ses interactions avec des G4s caractérisées (281). Les DBA ont une base acridine, comme **BRACO19** (Figure 1.6), mais étendue et chargée positivement par quaternarisation de l'amine aromatique avec des dérivés phényles. Des  $K_d$ s jusqu'à 0.1  $\mu$ M et une activité anti-HIV-1 avec des  $IC_{50}$  de l'ordre du micromolaire ont été mesurés pour les meilleurs complexes.

Samir a montré par RMN que le dérivé **DBA1** interagit avec la tétrade G2-G9-G12-G18 du G4 HIVpro1, en particulier avec G2, G9 et G18 (Figure 5.47A). Il lui fut alors suggéré par un pair de procéder à des expériences de *docking* pour déterminer plus précisément le mode d'interaction. S'il paraît très probable à ce stade que le ligand est *stacké* sur la tétrade, son orientation et la position de la charge positive sont inconnues.

8. En solution potassium, les G4s à deux ou quatre tétrades sont antiparallèles, mais ceux à trois tétrades ne le sont jamais, les G4s à boucles court forment des G4s parallèles, etc.

Bien que le *docking* permette d'explorer rapidement divers modes d'interaction, la flexibilité intrinsèque des G4s rend ces derniers peu compatibles avec cette méthode. De plus, la structure résolue par RMN d'HIVpro1 montre que la tétrade G2-G9-G12-G18 est peu accessible pour un ligand. Il m'a donc paru nécessaire de mettre en place une approche hybride entre le *docking*, pour générer rapidement des modes d'interactions possibles, et la dynamique moléculaire pour tenir compte de la flexibilité du G4. Concrètement, après avoir dérivé les paramètres de champs de force de **DBA1**, j'ai procédé comme suit :

1. La tétrade G2-G9-G12-G18 a été rendue accessible en basculant les résidus stériquement encombrants vers l'extérieur, par MD basée sur OpenMM (282).
2. **DBA1** a ensuite été *docké* sur la tétrade avec *LeDock* (283), sans aucune contrainte de position, ce qui a généré un ensemble diversifié de modes d'interaction, non biaisés, que j'ai regroupé en sept clusters.
3. La meilleure pose (selon le score de *docking*) de chaque cluster a été soumise à une simulation MD de 50 ns, permettant aux résidus du site d'interaction de se replier vers la tétrade, lorsque cela était possible, et de former des interactions avec **DBA1** qui n'auraient pas pu être capturées par le seul *docking*.
4. Les trajectoires obtenues ont été regroupées selon la géométrie du site d'interaction par PCA, donnant quatre modèles représentatifs reflétant les modes d'interactions les plus probables.

Deux orientations distinctes de **DBA1** ont été observées au sein des quatre clusters (Figure 5.47B). Dans tous les cas, **DBA1** établit sans surprise des interactions  $\pi - \pi$  avec la tétrade, principalement G2 et G18. G12 reste largement non lié, conformément aux observations RMN. Dans les clusters 1 et 4, la structure en croissant de **DBA1** se superpose avec la courbure de la tétrade, son groupe phényle s'insérant dans le sillon formé par les deux *termini* de HIVpro1. Ces clusters diffèrent par la position précise de **DBA1**, permettant un *stacking* supplémentaire avec la G19 en 3' (cluster 1), ou avec la C10 d'une boucle (cluster 4). Dans cette même boucle, G11, qui est *stacké* la tétrade en l'absence de ligand, se *stacke* parfois au-dessus de **DBA1**.

En revanche, les clusters 2 et 3 présentent une orientation inversée de **DBA1**, avec l'azote de l'acridinium chargé positivement situé au-dessus du canal ionique. Cette configuration favorise le *stacking* avec T1 en 5' et, dans le cluster 2, également avec C10. Dans le cluster 3, C10 se *stacke* sur G11 repositionné au-dessus de **DBA1**, formant ainsi une boucle latérale particulièrement structurée.

L'approche hybride *docking*/MD, couplée à des analyses PCA, a été particulièrement efficace pour l'étude de ce complexe ligand-G4. Bien qu'aucune contrainte n'ait été utilisée, ces simulations ont produit des résultats en parfait accord avec les expériences RMNs.

### 5.6.5 Étude des interactions $\pi - \pi$ d'oligonucléotides avec des phases stationnaires aromatiques pour leur séparation en phase inverse sans agent d'appariement.

Je clos ce chapitre avec une étude à laquelle j'ai participé pendant l'été 2025 et qui, je trouve, résume assez bien ce manuscrit. Il y a tout dedans : de la chimie analytique et physique, de la chromatographie

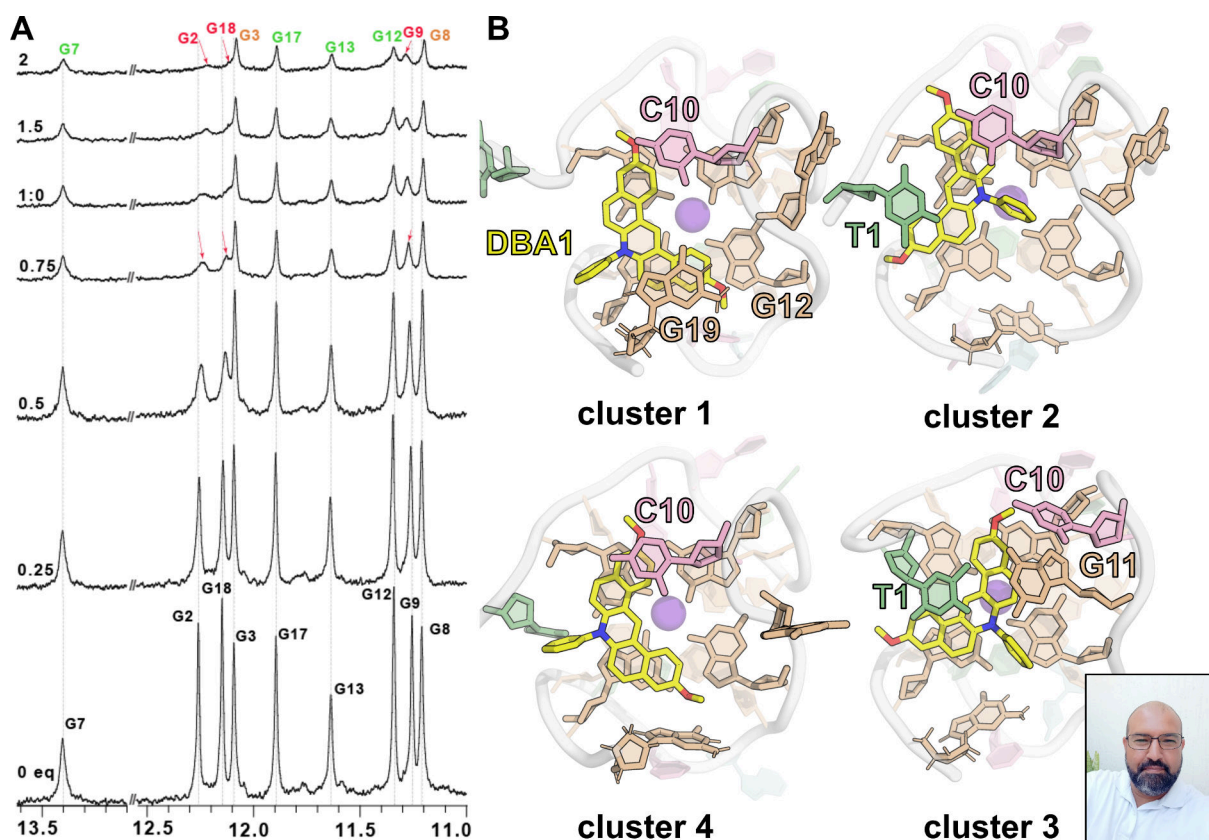


FIGURE 5.47. A. Titration <sup>1</sup>H-RMN du G4 HIVpro1 avec **DBA1** : superposition des spectres des protons imino (11.0–13.5 ppm) après addition progressive de **DBA1** (0 à 2 équivalents molaires) au G4, à une concentration de 150 μM dans un tampon 20 mM de phosphate de potassium, 70 mM de KCl (90:10 H<sub>2</sub>O/D<sub>2</sub>O). B. Vue de dessus du site de d'interaction obtenue par combinaison de *docking* et de MD non restreinte, à partir de laquelle quatre structures représentatives ont été extraites. **DBA1** est représenté en jaune, les guanines en brun, les cytosines en rose, les thymine en vert et les adénine en bleu. Encart : Samir Amrane.



et de la spectroscopie, des interactions  $\pi-\pi$  (mes préférées), des oligonucléotides et des G4s auxquelles on ne s'attendait pas. Ça manque cruellement de spectrométrie de masse, mais on notera que l'approche chromatographique développée ici y est justement particulièrement adaptée.

Tout commence (pour moi) lorsque je participe à une journée organisée par le Club Sud-Ouest de l'Association Francophone des Sciences Séparative à Toulouse, lors de laquelle je présente certaines de mes activités de caractérisation d'oligonucléotides. J'y croise Davy Guillaume, de l'Université de Genève, qui me recontacte quelques jours plus tard pour me parler de séparation d'oligonucléotides par HPLC en phase inverse. Il connaît déjà mes modestes travaux de chromatographie chez *Quality Assistance*, puisqu'il a lui-même (beaucoup plus) travaillé sur ces thématiques.<sup>9</sup>

Généralement, la rétention d'oligonucléotides, qui sont des polyanions hydrophiles, sur des phases stationnaires apolaires de phase inverse nécessite l'utilisation d'agents d'appariement (1). Ce sont typiquement des cations qui forme des paires d'ions (*ion pairs*, IP) avec les analytes, neutralisant la charge apparente et augmentant la rétention. Ces cations portent un groupe hydrophobe dont la modification permet de moduler la rétention. L'IP-RP-LC souffre cependant d'un certain manque de résolution de certaines impuretés, est plus difficile à développer que la RP-LC classique, nécessite des temps d'équilibration plus longs, et les agents d'appariement sont souvent la cause d'une suppression de l'ionisation qui détériore le signal MS. Des approches orthogonales existent, comme l'échange d'anion (AEX, non compatible avec la MS en couplage direct) ou la chromatographie hydrophile (HILIC), qui ne permet pas non plus toujours la résolution d'impuretés critiques.

Partant de ce constat, Davy a souhaité développer une méthode de RP-LC sans agent d'appariement, en utilisant des colonnes greffées avec des dérivés de phényle, notamment un biphényle et un pentafluorophényle. L'hypothèse est que ces ligands permettent la rétention d'oligonucléotides par interactions  $\pi-\pi$ , ne dépendant pas de la présence d'IP. Dans leur étude, Davy et un doctorant, Thanos Tsalmpouris, ont inclut des oligonucléotides modèles  $dX_{20}$  ( $X = A, C, G, T$ ) pour mieux comprendre l'influence de la séquence oligonucléotidique sur l'ordre d'élution des analytes ( $G < C < A < T$ ). Ils souhaitaient comprendre cet ordre, et confirmer que les interactions  $\pi-\pi$  entre colonne et analytes sont possibles, et modulables par la nature du ligand.

Pour répondre à ces questions, j'ai construit des séries de complexes entre les 4 nucléobases d'ADN et 2 ligands de colonne, un biphényle et un propyl-pentafluorophényle. Pour chaque combinaison nucléobase/ligand, j'ai préparé 4 géométries par rotation de  $90^\circ$  de l'angle de twist, tel que définit dans un duplex (284), pour un total de 32 complexes. J'ai ensuite optimisé la géométrie de ces complexes avec la fonctionnelle hybride B3LYP (jeu de bases def2-TZVP + def2/J) avec une correction non-locale de la dispersion (DFT-NL), ce qui est nécessaire pour correctement tenir compte des forces de dispersion de London dans ces systèmes dominés par des interactions attractives de van der Waals (285). Les énergies de ce type de systèmes sont généralement surestimées à cause de la superposition des jeux de bases (*Basis Set Superposition Error*, BSSE), que j'ai corrigée en implémentant une correction BB-CP (*Boys-Bernardi Counterpoise Correction*) (286). Aux aspects quantitatifs, j'ai ajouté une étude qualitative de la nature de l'interaction en calculant les potentiels électrostatiques (ESP) et les indicateurs de régions d'interaction (*Interaction Region Indicator*, IRI, Figure 5.48)

9. Notamment avec Valentina D'Atri, une ancienne postdoc de l'équipe de Valérie Gabelica, que j'avais croisée lorsque j'y étais moi-même postdoc.



Ces expériences ont permis de montrer que :

- La stabilité des complexes provient principalement d'interactions  $\pi$ - $\pi$ , dominées par les forces de dispersion. Les bases G et C bénéficient d'une stabilisation électrostatique supplémentaire grâce à une liaison polaire NH- $\pi$  entre leur amine exocyclique et l'anneau non *stacké* du biphenyle.
- Tous les complexes présentent des énergies d'interaction comprises entre -8 et -12 kcal/mol, en accord avec les énergies de *stacking* des bases nucléiques dans les hélices d'ADN (287). L'interaction de plusieurs bases le long des oligonucléotides devrait donc fournir une stabilisation suffisante pour assurer leur rétention.

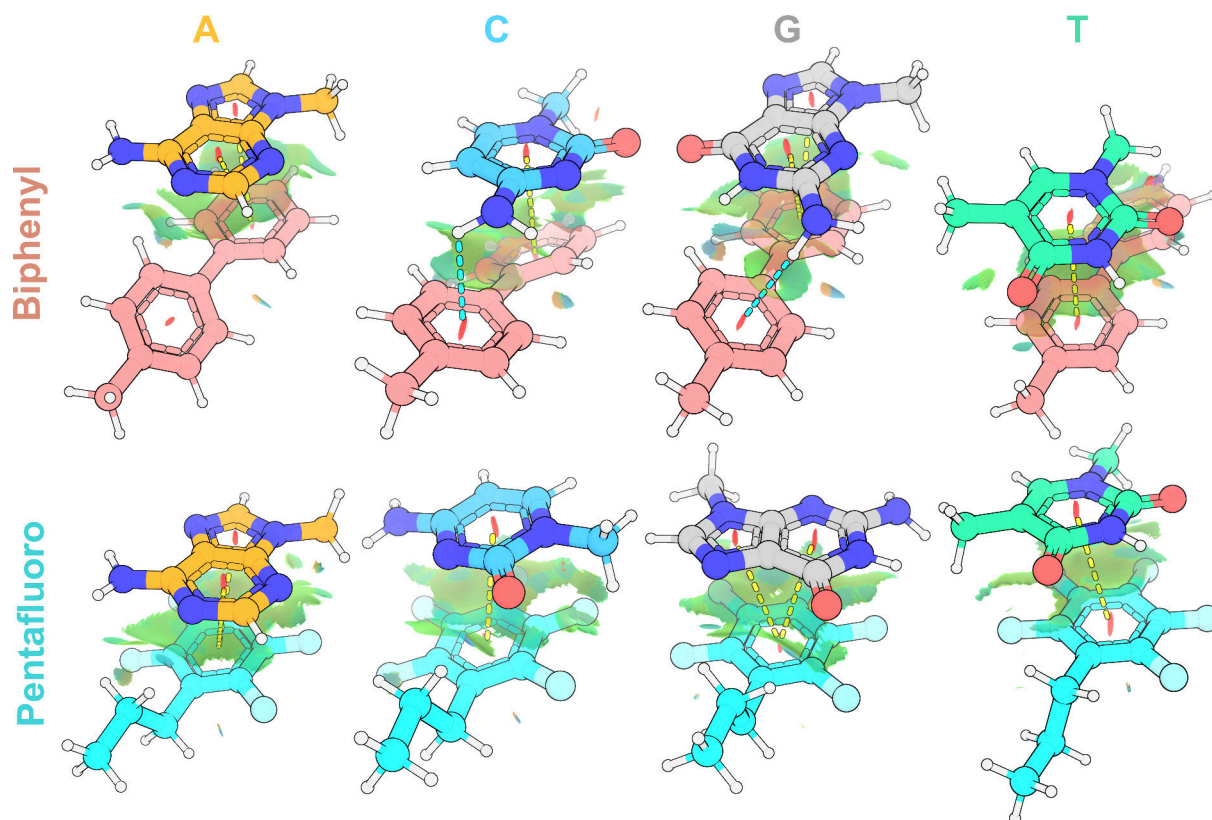
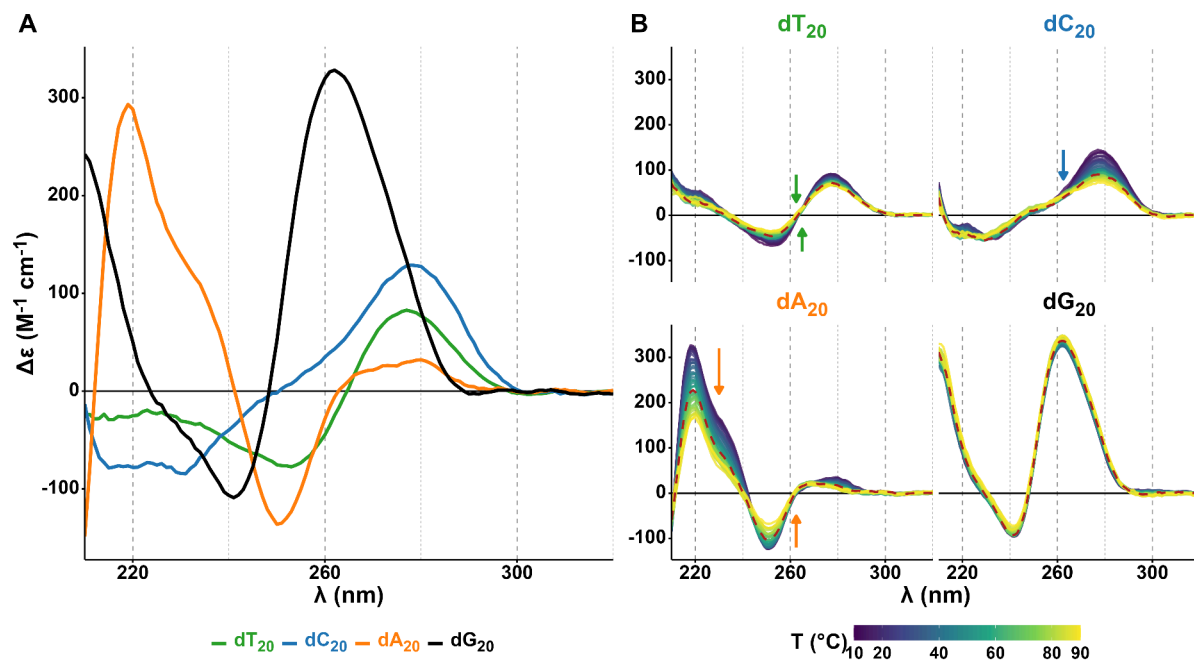


FIGURE 5.48. Géométries optimisées (niveau B3LYP + DFT-NL(VV10) / def2-TZVP) des complexes entre les bases d'ADN méthylées (de gauche à droite, A : orange, C : bleu, G : gris, T : vert) et les ligands de colonne biphenyle (en haut, rouge) et propyl-pentafluoro (en bas, bleu clair). La surface IRI (niveau = 0,75) met en évidence les interactions de van der Waals (vert) et la répulsion stérique (rouge). Les interactions  $\pi$ - $\pi$  canonique et la liaison polaire NH- $\pi$  sont également indiquées par des pointillés jaunes et cyan, respectivement.

Cela ne suffit pas, cependant, à expliquer l'ordre rétention des oligonucléotides modèles en fonction de leur structure *primaire*. L'exemple le plus criant est dG<sub>20</sub>, puisqu'il est très peu retenu alors que les guanines forment les complexes les plus stables. Bien sûr, les effets hydrophobes (C < G < A < T) contribuent très probablement (288). Mais le lecteur attentif aura compris que le problème majeur réside dans le choix des séquences, que j'ai décidé d'illustrer par CD, car il s'agit bien ici d'une question de structure *secondaire*.

Dans les conditions de chromatographie de cette étude (< 40% en méthanol ou acétonitrile, 50 mM d'acétate d'ammonium, pH 8.0), dT<sub>20</sub> et dC<sub>20</sub> sont très peu structurés, le second présentant un peu plus de *stacking* intramoléculaire transitoire (Figure 5.49). En revanche, dA<sub>20</sub>, en simple hélice, et surtout

dG<sub>20</sub>, en G4, forment des structures secondaires. Leurs nucléobases impliquées dans ces conformations ne sont alors pas disponibles pour interagir avec la phase stationnaire, ce qui diminue leur rétention d'autant que la conformation est stable. Il n'est donc pas surprenant de voir dG<sub>20</sub> très peu retenu et, à l'inverse, le très hydrophobe et linéaire dT<sub>20</sub> le plus retenu.



Je me limite ici à ma modeste contribution à cette étude par ailleurs exhaustive et très informative (125). Ce que je souhaite souligner, c'est qu'au-delà de l'étude elle-même, l'implémentation des approches de calcul que j'ai utilisées permettra l'étude, à un niveau théorique élevé, de complexes non covalents d'intérêt pour l'équipe PRISM, ou pour ARNA de manière plus générale. Par ailleurs, cette collaboration<sup>10</sup> m'a fait prendre pleinement conscience de la valeur ajoutée d'ARNA dans le paysage scientifique. Notre connaissance fine des acides nucléiques nous rend certains résultats évidents, alors qu'ils ne le sont pas pour d'autres équipes, par ailleurs excellentes dans leur domaine d'expertise.

10. et d'autres que j'ai omises ici par soucis de concision.



## Conclusions et perspectives

D'ingénieur chimiste spécialisé en chimie organométallique, j'ai lentement mais sûrement dérivé vers des approches plus physiques et analytiques, tout en conservant, je pense, un fil rouge thématique fort. J'ai trouvé durant ma thèse un sujet qui m'a passionné et qui continue de le faire, tandis que les autres thématiques que j'ai explorées m'ont permis de développer des compétences complémentaires. J'ai eu la chance de travailler dans des environnements stimulants, tant académiques que privés, et de collaborer avec des chercheurs talentueux, que je remercie vivement une nouvelle fois. Je ne reviens pas ici sur ces diverses expériences, qui sont résumées à la fin de chaque chapitre.

Dans le Chapitre 5, j'ai présenté la majorité des travaux que j'ai menés depuis mon arrivée (ou plutôt retour) à l'Université de Bordeaux, au sein des équipes BALI puis PRISM, dont certains resteront au cœur de mes recherches futures. Mes contributions récentes peuvent être regroupées en quatre axes majeurs :

1. Le développement de méthodes analytiques novatrices en spectrométrie de masse et en spectroscopies, soutenu par
2. Le développement d'outils logiciels *open-source* dédiés à l'analyse de données complexes et/ou à la gestion de bases de données,
3. L'évaluation approfondie de ligands ciblant les G4s, partiellement rendue possible grâce à
4. L'implémentation de flux de travail QM/MM pour l'étude de structures d'acides nucléiques non canoniques et de leurs complexes, récemment étendue à des systèmes protéiques.

L'extension de l'HDX/MS à l'étude des acides nucléiques a été mon projet phare depuis mon retour à Bordeaux. Ces travaux ont permis d'établir une méthode simple, précise et robuste pour l'étude des conformations, des dynamiques et des interactions des oligonucléotides d'ADN, avec des applications potentielles allant de la biophysique au contrôle qualité. Nous avons également montré que l'analyse approfondie des données brutes, grâce aux outils logiciels que nous avons développés, permet d'obtenir des informations sur la stabilité isotherme des G4s et sur leurs dynamiques de (dé)structuration. Enfin, le couplage avec la masse native permet d'étudier des systèmes complexes grâce à la séparation en masse et en mobilité ionique des divers conformères. Grâce à ces travaux, j'ai acquis une certaine visibilité au sein de la communauté HDX/MS.<sup>11</sup> Je tiens d'ailleurs à remercier Valérie Gabelica, qui m'a accordé sa confiance, m'a encouragé à développer ces approches et, très pragmatiquement, dont l'aide pour l'obtention d'un financement a été primordiale.

Plusieurs perspectives s'ouvrent désormais à moi sur ce projet. Certaines sont incluses dans le projet ANR qui s'achèvera à la fin de l'été 2026, d'autres non. À court terme, après la publication de nos travaux

---

11. On m'y appelle "*The DNA guy*", ce qui reflète, je pense, le caractère un peu atypique de ces travaux vis-à-vis de l'état de l'art.

visant à établir d'une part l'HDX/IMS-MS, et d'autre part l'HDX/MS native pour l'étude de complexes G4-ligands, deux axes de travail se dégageront :

1. Le développement de l'HDX/MS native *top-down*, consistant concrètement à ajouter une étape de fragmentation des analytes en phase gazeuse. Il s'agira d'évaluer la capacité de différentes méthodes (CID, ETD, ECD, UV-PD, IRMPD) à fragmenter des oligonucléotides sans randomiser la position des deutérons, ce qui n'est pas trivial. L'obtention de fragments dont la deutération reflète fidèlement leur état en solution permettra de caractériser les structures et interactions à l'échelle du nucléotide. En d'autres termes, il sera par exemple possible de déterminer les sites d'interactions non covalentes ou d'identifier quelles guanines sont impliquées dans des tétrades. Ces applications nécessiteront le développement d'outils logiciels assurant l'identification des fragments, la quantification de leur deutération et le tracé de leurs cinétiques. Pour cela, je pourrai m'appuyer sur OligoR, déjà opérationnel pour les expériences menées au niveau des analytes intacts.
2. L'extension de la technique à d'autres structures secondaires. Ce second axe nécessitera également des développements méthodologiques, compte tenu des temps d'échange plus courts observés dans les paires de bases. On peut faire l'hypothèse que la plupart des analytes "non-G4" nécessiteront des systèmes de mélange isotopique et de mesure plus rapides qu'actuellement. Il convient toutefois de noter que de nombreux aptamères contiennent des G4s (289), et que de tels modèles pourraient me permettre d'effectuer une transition plus en douceur depuis les systèmes à échange lent.

Il apparaît assez clairement, je pense, à la lecture de ce manuscrit que j'apprécie l'utilisation de plusieurs techniques orthogonales pour étudier des systèmes (simples). Comme j'enseigne dans un UFR de pharmacie, je me sens un plus un peu obligé de suivre les guidances de l'Agence Européenne des médicaments,<sup>12</sup> et de la FDA.<sup>13</sup> Plus sérieusement, et parallèlement aux deux axes présentés ci-dessus, je m'intéresse au développement d'une approche *bottom-up*, à l'instar de ce qui est couramment appliqué aux protéines. Dans cette approche, les analytes seraient clivés en solution avant leur analyse LC-MS. Cela est loin d'être trivial, puisqu'il faudrait définir des conditions de *quenching* (température, pH) de la réaction d'échange et mettre au point une méthode de nucléolyse compatible.

Une telle méthode présenterait un double intérêt : améliorer la résolution (axe 1) et permettre l'étude d'échanges rapides (axe 2), même si ce dernier nécessitera probablement des échanges à basse température. En définitive, notre boîte à outils isotopiques devra permettre de caractériser la dynamique de complexes acides nucléiques-protéines. Grâce à une combinaison d'approches *top-down* novatrices et *bottom-up* plus classiques, nous pourrions étudier conjointement les deux (ou plusieurs) partenaires, alors que les études HDX/MS actuelles se concentrent généralement sur la protéine seule.

Les travaux menés avec Matthieu Ranz au sein de l'équipe de Valérie Gabelica ont conduit à plusieurs publications, mais au moins deux manuscrits restent à publier. Ces dernières années, une leçon s'est

---

12. *Orthogonal analytical methods, (methods using different principles and providing different selectivities), should be developed in cases where a lack in specificity and/or selectivity leads to an inadequate control strategy for the affected impurities. Guideline on the chemistry of active substances*, 2016.

13. *Emphasis should be placed on developing orthogonal quantitative methods to definitively identify any differences in product attributes. Development of Therapeutic Protein Biosimilars : Comparative Analytical Assessment and Other Quality-Related Considerations*, 2025.

dégagée, que je perçois d'autant mieux après avoir écrit ces quelques pages : plus je travaille, plus j'ai de travail. J'ai la désagréable impression d'avoir en permanence deux publications de retard sur mon planning et de ne pas avoir le temps de... rien. *No rest for the weary*, disent les Anglo-Saxons. Évidemment, je dois jongler entre activités de recherche et enseignements, ce qui ne facilite pas les choses. Mais tous les enseignants-chercheurs sont logés à la même enseigne, et je trouve cette situation très enrichissante. Du travail, justement, il en reste beaucoup.

Par exemple, d'autres techniques analytiques devront être développées. Nous avons montré avec Eps2Fold que les données spectroscopiques pouvaient être exploitées pour déterminer la conformation des G4s, avec une granulométrie assez inattendue. Cette approche pourra être étendue à d'autres structures secondaires de l'ADN, mais aussi à l'ARN. Le dichroïsme circulaire, incontournable pour l'étude des G4s, est comparativement peu utilisé dans d'autres communautés. Mieux comprendre, *a minima* de manière empirique, les déterminants des signatures spectroscopiques UV et CD sera d'une grande utilité pour l'évaluation « macroscopique » des oligonucléotides, en particulier ceux dont la fonction est liée à la structure secondaire, comme les aptamères.

Il s'agit d'un travail assez systématique, qui produira de grandes quantités de données. Cela constituera l'occasion d'explorer d'autres méthodes d'analyses multidimensionnelles, comme l'algorithme t-SNE (*t-Distributed Stochastic Neighbor Embedding*) qui, comparé à la PCA, facilite la séparation et l'identification de sous-groupes dans des ensembles, en conservant mieux les proximités locales. L'utilisation de l'apprentissage automatique pourrait également être envisagée si la quantité de données le permet.

Une autre leçon que j'ai tirée de ces dernières années est que, même après huit années en tant que maître de conférences et à 40 ans, j'apprends toujours autant qu'en thèse. Ainsi, un autre axe de travail concerne l'utilisation avancée du BLI, pour laquelle je souhaite prêter main forte à Carmelo Di Primo. Ce projet s'inscrit, pour moi, dans la continuité de mes autres travaux, à savoir le développement de méthodes orthogonales permettant de mieux caractériser des objets polymorphes comme les oligonucléotides. Le BLI (et la SPR) sont des techniques de choix pour déterminer les constantes de vitesse d'association et de dissociation de complexes non covalents, ce qui permettra de mieux rationaliser les cinétiques d'échange isotopique de ces derniers. À titre personnel, il s'agit également d'une nouvelle opportunité d'apprentissage auprès d'un chercheur très pointu, ce qui est toujours un plaisir.

Le ligand de G4 est un sujet qui fait le lien entre ma thèse et mes activités plus récentes. Parmi celles-ci, on retiendra une étude ayant conduit à l'identification et à la caractérisation de ligands foldamères sélectifs pour les G4s de topologies parallèles. La combinaison des compétences d'ARNa en spectrométrie de masse native, biologie structurale (cristallographie, RMN) et chimie computationnelle a permis d'expliquer cette sélectivité et a transformé une curiosité de laboratoire en une étude d'intérêt pour la communauté G4.

Je n'ai pas de perspectives définies sur ce thème de recherche. Les journaux sont déjà saturés de ligands de G4s n'apportant pas grand-chose à l'état de l'art. Cela dit, je sais que je ne refuserai pas de travailler sur des ligands de haute valeur ajoutée, que j'avais définis à la fin de ma thèse comme des molécules fonctionnalisées (marquage, *pull-down*, etc.) et/ou présentant des sélectivités particulières.<sup>14</sup> D'autre

14. Qui me connaît sait que je ne dis jamais non à rien, donc bon...



part, les complexes G4-petites molécules constituent d'excellents modèles pour le développement analytique (comme celui de l'HDX/MS), car ils bénéficient d'une large gamme d'affinités, de sélectivités et de cinétiques d'association, possèdent des sites d'interactions fréquemment connus, et peuvent déplacer des équilibres conformationnels.

Plusieurs de ces activités m'ont conduit à l'implémentation de calculs QM/MM, notamment de MD sur GPU. À l'origine, cependant, cette activité a été initiée afin de m'impliquer dans des thématiques de l'équipe PRISM. J'en profite pour remercier Cameron Mackereth, qui m'a gentiment accueilli et avec lequel je partage un bureau (et des biscuits).

La thématique que je développe avec lui n'est pas totalement nouvelle pour moi, puisque j'avais déjà eu le plaisir de travailler avec des aptamères. J'ai donc mis en place un flux de travail généraliste et robuste, qui a déjà porté ses fruits pour la détermination de la structure d'un aptamère de la dopamine, ainsi que pour d'autres systèmes publiés ou non. En particulier, ces calculs se sont révélés utiles dans l'étude des modes d'interaction de ligands de G4s (un sujet récurrent).

À court terme, d'autres aptamères pourront être caractérisés via la collaboration franco-canadienne déjà bien établie. À plus long terme, j'espère que ces travaux permettront de mieux comprendre les déterminants de la structuration des aptamères et de l'ADN « non-duplex » en général, ce qui pourrait alimenter les algorithmes de prédiction. Cela pourrait être couplé à une évaluation plus systématique de leurs performances dans la prédiction de structures secondaires particulières, permettant ainsi d'affiner les besoins en développement.

Je ne peux terminer ce chapitre sur la recherche sans mentionner mon attachement à la science ouverte et au logiciel libre. Certes, cela me conduit à produire des *Supporting Information* à rallonge, mais j'ai pu constater encore récemment que certains *reviewers* commentent des figures numérotées S100+, ce qui montre l'intérêt réel pour ces données. Blague à part, il me paraît essentiel de produire une science transparente et reproductible, et c'est pourquoi j'attache une attention particulière à la mise à disposition des données pour la communauté. La mise en ligne de la base de données g4dbr en est un exemple. Bien qu'initialement construite pour l'interprétation structurale des résultats de MS native, g4dbr se révèle particulièrement pratique au quotidien pour quiconque s'intéresse aux G4s. À l'inverse, je bénéficie également des données mises à disposition par d'autres (cf. Figure 5.44B).

Pour assurer la reproductibilité, je m'attache à développer des méthodes robustes pour la production de données brutes et à automatiser leur analyse (toujours de façon transparente), afin d'éviter erreurs et biais humains. Cela m'a conduit à développer les logiciels OligoR, Eps2Fold et meltR, accessibles à la communauté, pour le traitement des données HDX, la détermination de topologies à partir de spectres UV/Vis et CD, et l'évaluation fiable des températures de fusion.

Je suis désormais impliqué au niveau institutionnel dans ces thématiques, en tant que correspondant Science ouverte, correspondant Données de la recherche, et membre du groupe de travail « Éthique, intégrité, genre et science ouverte » pour l'obtention du label HRS4R, ainsi que de l'Atelier Bordelais des Données de recherche (ABDo), récemment labellisé par le ministère. Ces engagements influencent concrètement ma façon de travailler et guideront mes projets futurs, en m'assurant de toujours partager mes données et méthodes de manière transparente et accessible à la communauté.

Enfin, et puisque c'est un sujet central de l'HDR, un mot sur les étudiants. J'ai eu la chance d'en encadrer de nombreux dès ma première année de thèse. Je dois dire que j'ai eu beaucoup de chance, car tous ont été des membres impliqués, apportant chacun leur pierre à l'édifice. De nombreux projets présentés ici n'auraient tout simplement pas pu voir le jour sans eux. J'aime transmettre mon savoir-faire, et m'être orienté vers une carrière d'enseignant-chercheur n'a pas été un accident de parcours ni un choix par défaut. Et même après m'être un peu plaint de la charge de travail plus haut, je ne compte pas les heures consacrées aux étudiants. J'ai récemment été sollicité par le collège des écoles doctorales de l'Université de Bordeaux pour assurer des formations doctorales, ce que je vais accomplir avec grand plaisir. J'espère que l'HDR me permettra de continuer à transmettre à de nombreux doctorants.



## Bibliographie

1. E. Largy, A. König, A. Ghosh, D. Ghosh, S. Benabou, F. Rosu, V. Gabelica, [Mass Spectrometry of Nucleic Acid Noncovalent Complexes](#). *Chem. Rev.* **122**, 7720-7839 (2022).
2. V. A. Traag, P. Van Dooren, Y. Nesterov, [Narrow Scope for Resolution-Limit-Free Community Detection](#). *Phys. Rev. E* **84**, 16114 (2011).
3. N. Khristenko, F. Rosu, E. Largy, J. Haustant, C. Mesmin, V. Gabelica, [Native Electrospray Ionization of Multi-Domain Proteins via a Bead Ejection Mechanism](#). *J. Am. Chem. Soc.* **145**, 498-506 (2023).
4. J. Guillon, C. Denevault-Sabourin, E. Chevret, M. Brachet-Botineau, V. Milano, A. Guédin-Beaurepaire, S. Moreau, L. Ronga, S. Savrimoutou, S. Rubio, J. Ferrer, J. Lamarche, J.-L. Mergny, M.-C. Viaud-Massuard, M. Ranz, E. Largy, V. Gabelica, F. Rosu, F. Gouilleux, V. Desplat, [Design, Synthesis, and Antiproliferative Effect of 2,9-bis\[4-\(Pyridinylalkylaminomethyl\)Phenyl\]-1,10-phenanthroline Derivatives on Human Leukemic Cells by Targeting G-quadruplex](#). *Arch. Pharm.* **354**, 2000450 (2021).
5. E. Largy, M. Ranz, [OligoR : A Native HDX/MS Data Processing Application Dedicated to Oligonucleotides](#). *Anal. Chem.* **95**, 9615-9622 (2023).
6. E. Largy, M. Ranz, V. Gabelica, [A General Framework to Interpret Hydrogen–Deuterium Exchange Native Mass Spectrometry of G-quadruplex DNA](#). *J. Am. Chem. Soc.* **145**, 26843-26857 (2023).
7. S. Ceschi, E. Largy, V. Gabelica, C. Sissi, [A Two-Quartet G-quadruplex Topology of Human KIT2 Is Conformationally Selected by a Perylene Derivative](#). *Biochimie* **179**, 77-84 (2020).
8. J. Guillon, A. Cohen, C. Boudot, S. Monic, S. Savrimoutou, S. Moreau, S. Albenque-Rubio, C. Lafon-Schmaltz, A. Dassonville-Klimpt, J.-L. Mergny, L. Ronga, M. Bernabeu De Maria, J. Lamarche, C. D. Lago, E. Largy, V. Gabelica, S. Moukha, P. Dozolme, P. Agnamey, N. Azas, C. Mullié, B. Courtioux, P. Sonnet, [Design, Synthesis, and Antiprotozoal Evaluation of New Promising 2,9-Bis\[\(Substituted-Aminomethyl\)\]-4,7-Phenyl-1,10-Phenanthroline Derivatives, a Potential Alternative Scaffold to Drug Efflux](#). *Pathogens* **11**, 1339 (2022).
9. E. Largy, W. Liu, A. Hasan, D. M. Perrin, [Base-pairing Behavior of a Carbocyclic janus-AT Nucleoside Analogue Capable of Recognizing a and T within a DNA Duplex](#). *ChemBioChem* **14**, 2199-2208 (2013).
10. E. Largy, W. Liu, A. Hasan, D. M. Perrin, [A Pyrimidopyrimidine janus-AT Nucleoside with Improved Base-pairing Properties to Both a and T within a DNA Duplex : The Stabilizing Effect of a Second Endocyclic Ring Nitrogen](#). *Chem. – Eur. J.* **20**, 1495-1499 (2014).
11. E. Largy, A. Granzhan, F. Hamon, D. Verga, M.-P. Teulade-Fichou, « [Visualizing the Quadruplex : From Fluorescent Ligands to Light-up Probes](#) » (Springer Berlin Heidelberg, 2012), p. 111-177.
12. A. Piazza, J.-B. Boulé, J. Lopes, K. Mingo, E. Largy, M.-P. Teulade-Fichou, A. Nicolas, [Genetic Instability Triggered by G-quadruplex Interacting Phen-DC Compounds in \*Saccharomyces Cerevisiae\*](#). *Nucleic Acids Res.* **38**, 4337-4348 (2010).

## Bibliographie

13. E. Largy, B. Alies, G. Condesse, A. Gaubert, T. Livingston, K. Gaudin, [Teaching with Simulation Tools to Introduce the Basics of Analytical Chemistry Instrumentation](#). *Anal. Bioanal. Chem.* **414**, 6709-6721 (2022).
14. J. Keirstead, Scholar : Analyse Citation Data from Google Scholar. (2016).
15. A. Bedrat, L. Lacroix, J.-L. Mergny, [Re-Evaluation of G-quadruplex Propensity with G4Hunter](#). *Nucleic Acids Res.* **44**, 1746-1759 (2016).
16. S. Neidle, S. Balasubramanian, Éd., « [Fundamentals of Quadruplex Structures](#) » (The Royal Society of Chemistry, 2006), p. 1-30.
17. D. Rhodes, H. J. Lipps, [G-Quadruplexes and Their Regulatory Roles in Biology](#). *Nucleic Acids Res.* **43**, 8627-8637 (2015).
18. D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, [The Regulation and Functions of DNA and RNA G-quadruplexes](#). *Nat. Rev. Mol. Cell Biol.* **21**, 459-474 (2020).
19. M. M. Fay, S. M. Lyons, P. Ivanov, [RNA G-quadruplexes in Biology : Principles and Molecular Mechanisms](#). *J. Mol. Biol.* **429**, 2127-2147 (2017).
20. P. Kharel, S. Balaratnam, N. Beals, S. Basu, [The Role of RNA G-quadruplexes in Human Diseases and Therapeutic Strategies](#). *WIREs RNA* **11**, e1568 (2020).
21. C. Roxo, W. Kotkowiak, A. Pasternak, [G-Quadruplex-Forming Aptamers—Characteristics, Applications, and Perspectives](#). *Molecules* **24**, 3781 (2019).
22. J.-L. Mergny, D. Sen, [DNA Quadruple Helices in Nanotechnology](#). *Chem. Rev.* **119**, 6290-6325 (2019).
23. E. Largy, J.-L. Mergny, V. Gabelica, « [Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability](#) » (Springer International Publishing, 2016), p. 203-258.
24. A. Marchand, V. Gabelica, [Native Electrospray Mass Spectrometry of DNA G-quadruplexes in Potassium Solution](#). *J. Am. Soc. Mass Spectrom.* **25**, 1146-1154 (2014).
25. A. Marchand, V. Gabelica, [Folding and Misfolding Pathways of G-quadruplex DNA](#). *Nucleic Acids Res.* **44**, 10999-11012 (2016).
26. A. Ghosh, E. Largy, V. Gabelica, [DNA G-quadruplexes for Native Mass Spectrometry in Potassium : A Database of Validated Structures in Electrospray-Compatible Conditions](#). *Nucleic Acids Res.* **49**, 2333-2345 (2021).
27. S. Haider, G. N. Parkinson, S. Neidle, [Crystal Structure of the Potassium Form of an Oxytricha Nova G-quadruplex](#). *J. Mol. Biol.* **320**, 189-200 (2002).
28. M. P. Horvath, S. C. Schultz, [DNA G-quartets in a 1.86 Å Resolution Structure of an Oxytricha Nova Telomeric Protein-DNA Complex](#). *J. Mol. Biol.* **310**, 367-377 (2001).
29. S. A. Dvorkin, A. I. Karsisiotis, M. Webba Da Silva, [Encoding Canonical DNA Quadruplex Structure](#). *Sci. Adv.* **4**, eaat3007 (2018).
30. H. L. Lightfoot, T. Hagen, N. J. Tatum, J. Hall, [The Diverse Structural Landscape of Quadruplexes](#). *FEBS Lett.* **593**, 2083-2102 (2019).

31. E. Largy, A. Marchand, S. Amrane, V. Gabelica, J.-L. Mergny, [Quadruplex Turncoats : Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches](#). *J. Am. Chem. Soc.* **138**, 2780-2792 (2016).
32. V. Kuryavyi, A. Majumdar, A. Shallop, N. Chernichenko, E. Skripkin, R. Jones, D. J. Patel, [A Double Chain Reversal Loop and Two Diagonal Loops Define the Architecture of Unimolecular DNA Quadruplex Containing a Pair of Stacked G\(Syn\)-G\(Syn\)-G\(Anti\)-G\(Anti\) Tetrads Flanked by a G-\(T-T\) Triad and a T-T-T triple](#)<sup>1</sup>Edited by M. F. Summers. *J. Mol. Biol.* **310**, 181-194 (2001).
33. B. Karg, S. Mohr, K. Weisz, [Duplex-guided Refolding into Novel G-quadruplex \(3+1\) Hybrid Conformations](#). *Angew. Chem. Int. Ed.* **58**, 11068-11071 (2019).
34. A. Kotar, R. Rigo, C. Sissi, J. Plavec, [Two-Quartet Kit\\* G-quadruplex Is Formed via Double-Stranded Pre-Folded Structure](#). *Nucleic Acids Res.* **47**, 2641-2653 (2019).
35. A. Guédin, J. Gros, P. Alberti, J.-L. Mergny, [How Long Is Too Long? Effects of Loop Size on G-quadruplex Stability](#). *Nucleic Acids Res.* **38**, 7858-7868 (2010).
36. E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, E. H. Blackburn, [Telomeric DNA Oligonucleotides Form Novel Intramolecular Structures Containing Guanine-guanine Base Pairs](#). *Cell* **51**, 899-908 (1987).
37. J. R. Williamson, M. K. Raghuraman, T. R. Cech, [Monovalent Cation-Induced Structure of Telomeric DNA : The G-quartet Model](#). *Cell* **59**, 871-880 (1989).
38. D. Sen, W. Gilbert, [The Structure of Telomeric DNA : DNA Quadriplex Formation](#). *Curr. Opin. Struct. Biol.* **1**, 435-438 (1991).
39. R. M. Brosh, [Put on Your Thinking Cap : G-quadruplexes, Helicases, and Telomeres](#). *Aging (Milano)* **3**, 332-335 (2011).
40. M. Dong, T. E. Mürdter, U. Klotz, [Telomeres and Telomerase as Novel Drug Targets : Reflections on the 2009 Nobel Prize in Physiology or Medicine](#). *Eur. J. Clin. Pharmacol.* **66**, 1-3 (2010).
41. A. De Cian, L. Lacroix, C. Douarre, N. Temime-Smaali, C. Trentesaux, J.-F. Riou, J.-L. Mergny, [Targeting Telomeres and Telomerase](#). *Biochimie* **90**, 131-155 (2008).
42. Y. Wang, D. J. Patel, [Solution Structure of the Human Telomeric Repeat d\[AG3\(T2AG3\)3\] G-tetraplex](#). *Structure* **1**, 263-282 (1993).
43. G. N. Parkinson, M. P. H. Lee, S. Neidle, [Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA](#). *Nature* **417**, 876-880 (2002).
44. K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, [Structure of the Human Telomere in K<sup>+</sup> Solution : An Intramolecular \(3 + 1\) G-quadruplex Scaffold](#). *J. Am. Chem. Soc.* **128**, 9963-9970 (2006).
45. J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones, D. Yang, [Structure of the Intramolecular Human Telomeric G-quadruplex in Potassium Solution : A Novel Adenine Triple Formation](#). *Nucleic Acids Res.* **35**, 2440-2450 (2007).
46. A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, [Structure of Two Intramolecular G-quadruplexes Formed by Natural Human Telomere Sequences in K<sup>+</sup> Solution](#) †. *Nucleic Acids Res.* **35**, 6517-6525 (2007).



## Bibliographie

47. J. Dai, M. Carver, C. Punchihewa, R. A. Jones, D. Yang, [Structure of the Hybrid-2 Type Intramolecular Human Telomeric G-quadruplex in K<sup>+</sup> Solution : Insights into Structure Polymorphism of the Human Telomeric Sequence](#). *Nucleic Acids Res.* **35**, 4927-4940 (2007).
48. K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, A. T. Phan, [Structure of the Human Telomere in K<sup>+</sup> Solution : A Stable Basket-Type G-quadruplex with Only Two G-tetrad Layers](#). *J. Am. Chem. Soc.* **131**, 4301-4309 (2009).
49. Z. Zhang, J. Dai, E. Veliath, R. A. Jones, D. Yang, [Structure of a Two-G-tetrad Intramolecular G-quadruplex Formed by a Variant Human Telomeric Sequence in K<sup>+</sup> Solution : Insights into the Interconversion of Human Telomeric G-quadruplex Structures](#). *Nucleic Acids Res.* **38**, 1009-1021 (2010).
50. A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, [Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution](#). *Nucleic Acids Res.* **34**, 2723-2735 (2006).
51. A. T. Phan, V. Kuryavyi, D. J. Patel, [DNA Architecture : From G to Z](#). *Curr. Opin. Struct. Biol.* **16**, 288-298 (2006).
52. N. Kosiol, S. Juranek, P. Brossart, A. Heine, K. Paeschke, [G-Quadruplexes : A Promising Target for Cancer Therapy](#). *Mol. Cancer* **20**, 40 (2021).
53. S. Neidle, M. A. Read, [G-Quadruplexes as Therapeutic Targets](#). *Biopolymers* **56**, 195-208 (2000).
54. S. Balasubramanian, S. Neidle, [G-Quadruplex Nucleic Acids as Therapeutic Targets](#). *Curr. Opin. Chem. Biol.* **13**, 345-353 (2009).
55. M. Read, R. J. Harrison, B. Romagnoli, F. A. Tanious, S. H. Gowan, A. P. Reszka, W. D. Wilson, L. R. Kelland, S. Neidle, [Structure-Based Design of Selective and Potent G Quadruplex-Mediated Telomerase Inhibitors](#). *Proc. Natl. Acad. Sci.* **98**, 4844-4849 (2001).
56. A. De Cian, E. DeLemos, J.-L. Mergny, M.-P. Teulade-Fichou, D. Monchaud, [Highly Efficient G-quadruplex Recognition by Bisquinolinium Compounds](#). *J. Am. Chem. Soc.* **129**, 1856-1857 (2007).
57. M. J. B. Moore, C. M. Schultes, J. Cuesta, F. Cuenca, M. Gunaratnam, F. A. Tanious, W. D. Wilson, S. Neidle, [Trisubstituted Acridines as G-quadruplex Telomere Targeting Agents. Effects of Extensions of the 3,6- and 9-Side Chains on Quadruplex Binding, Telomerase Activity, and Cell Proliferation](#). *J. Med. Chem.* **49**, 582-599 (2006).
58. M. Zuffo, A. Guédin, E.-D. Leriche, F. Doria, V. Pirota, V. Gabelica, J.-L. Mergny, M. Freccero, [More Is Not Always Better : Finding the Right Trade-off between Affinity and Selectivity of a G-quadruplex Ligand](#). *Nucleic Acids Res.* **46**, e115-e115 (2018).
59. A. Funke, K. Weisz, [Thermodynamic Signature of Indoloquinolines Interacting with G-quadruplexes : Impact of Ligand Side Chain](#). *Biochimie* **157**, 142-148 (2019).
60. H. Bertrand, A. Granzhan, D. Monchaud, N. Saettel, R. Guillot, S. Clifford, A. Guédin, J.-L. Mergny, M.-P. Teulade-Fichou, [Recognition of G-quadruplex DNA by Triangular Star-shaped Compounds : With or without Side Chains ?](#) *Chem. – Eur. J.* **17**, 4529-4539 (2011).
61. L. Rossetti, M. Franceschin, A. Bianco, G. Ortaggi, M. Savino, [Perylene Diimides with Different Side Chains Are Selective in Inducing Different G-quadruplex DNA Structures and in Inhibiting Telomerase](#). *Bioorg. Med. Chem. Lett.* **12**, 2527-2533 (2002).

62. E. Micheli, C. M. Lombardo, D. D'Ambrosio, M. Franceschin, S. Neidle, M. Savino, [Selective G-quadruplex Ligands : The Significant Role of Side Chain Charge Density in a Series of Perylene Derivatives](#). *Bioorg. Med. Chem. Lett.* **19**, 3903-3908 (2009).
63. N. H. Campbell, G. N. Parkinson, A. P. Reszka, S. Neidle, [Structural Basis of DNA Quadruplex Recognition by an Acridine Drug](#). *J. Am. Chem. Soc.* **130**, 6722-6724 (2008).
64. D. Koirala, S. Dhakal, B. Ashbridge, Y. Sannohe, R. Rodriguez, H. Sugiyama, S. Balasubramanian, H. Mao, [A Single-Molecule Platform for Investigation of Interactions between G-quadruplexes and Small-Molecule Ligands](#). *Nat. Chem.* **3**, 782-787 (2011).
65. H. Bertrand, S. Bombard, D. Monchaud, E. Talbot, A. Guédin, J.-L. Mergny, R. Grünert, P. J. Bednarski, M.-P. Teulade-Fichou, [Exclusive Platination of Loop Adenines in the Human Telomeric G-quadruplex](#). *Org. Biomol. Chem.* **7**, 2864 (2009).
66. H. Bertrand, D. Monchaud, A. De Cian, R. Guillot, J.-L. Mergny, M.-P. Teulade-Fichou, [The Importance of Metal Geometry in the Recognition of G-quadruplex-DNA by Metal-Terpyridine Complexes](#). *Org. Biomol. Chem.* **5**, 2555 (2007).
67. H. Bertrand, S. Bombard, D. Monchaud, M.-P. Teulade-Fichou, [A Platinum–Quinacridine Hybrid as a G-quadruplex Ligand](#). *JBIC, J. Biol. Inorg. Chem.* **12**, 1003-1014 (2007).
68. H. Bertrand, S. Bombard, D. Monchaud, M.-P. Teulade-Fichou, [New Platinum\(II\) Complexes Targeting the Loops of the Human Telomeric G-quadruplex](#). *Nucleic Acids Symp. Ser.* **52**, 163-164 (2008).
69. D. Monchaud, C. Allain, M.-P. Teulade-Fichou, [Development of a Fluorescent Intercalator Displacement Assay \(G4-FID\) for Establishing Quadruplex-DNA Affinity and Selectivity of Putative Ligands](#). *Bioorg. Med. Chem. Lett.* **16**, 4842-4845 (2006).
70. D. Monchaud, C. Allain, H. Bertrand, N. Smargiasso, F. Rosu, V. Gabelica, A. De Cian, J.-L. Mergny, M.-P. Teulade-Fichou, [Ligands Playing Musical Chairs with G-quadruplex DNA : A Rapid and Simple Displacement Assay for Identifying Selective G-quadruplex Binders](#). *Biochimie* **90**, 1207-1223 (2008).
71. D. Monchaud, M.-P. Teulade-Fichou, « [G4-FID : A Fluorescent DNA Probe Displacement Assay for Rapid Evaluation of Quadruplex Ligands](#) » (Humana Press, 2009), p. 257-271.
72. A. König, V. Laffilé, S. Thore, C. D. Mackereth, L. Yatsunyk, Y. Ferrand, E. Largy, V. Gabelica, [Helical Aromatic Oligoamide Foldamers as Selective G- Quadruplex Ligands](#) (accepted).
73. D. Monchaud, M.-P. Teulade-Fichou, [A Hitchhiker's Guide to G-quadruplex Ligands](#). *Org, Biomol, Chem*, **6**, 627-636 (2008).
74. S. Rottenberg, C. Disler, P. Perego, [The Rediscovery of Platinum-Based Cancer Therapy](#). *Nat. Rev. Cancer* **21**, 37-50 (2021).
75. E. Largy, F. Hamon, F. Rosu, V. Gabelica, E. De Pauw, A. Guédin, J.-L. Mergny, M.-P. Teulade-Fichou, [Tridentate N-donor Palladium\(II\) Complexes as Efficient Coordinating Quadruplex DNA Binders](#). *Chem. – Eur. J.* **17**, 13274-13283 (2011).
76. F. Kröhnke, [The Specific Synthesis of Pyridines and Oligopyridines<sup>1</sup>](#). *Synthesis* **1976**, 1-24 (1976).
77. P. Froidevaux, J. M. Harrowfield, A. N. Sobolev, [Calixarenes as Scaffolds : Introduction of Tridentate Rare Earth Metal Binding Units into Calix\[4\]Arene](#). *Inorg. Chem.* **39**, 4678-4687 (2000).

78. P. K. Iyer, J. B. Beck, C. Weder, S. J. Rowan, [Synthesis and Optical Properties of Metallo-Supramolecular Polymers](#). *Chem. Commun.*, 319 (2005).
79. D. Knapton, M. Burnworth, S. J. Rowan, C. Weder, [Fluorescent Organometallic Sensors for the Detection of Chemical-warfare-agent Mimics](#). *Angew. Chem.* **118**, 5957-5961 (2006).
80. S. D. Cummings, [Platinum Complexes of Terpyridine : Synthesis, Structure and Reactivity](#). *Coord. Chem. Rev.* **253**, 449-478 (2009).
81. Q.-Z. Yang, L.-Z. Wu, Z.-X. Wu, L.-P. Zhang, C.-H. Tung, [Long-Lived Emission from Platinum\(II\) Terpyridyl Acetylide Complexes](#). *Inorg. Chem.* **41**, 5653-5655 (2002).
82. X. Liu, E. J. L. McInnes, C. A. Kilner, M. Thornton-Pett, M. A. Halcrow, [Metal Complexes of 4'-\(3-Phenylpropoxy\)-2,2':6',2''-Terpyridine and 4'-\(3-Propoxy\)-2,2':6',2''-Terpyridine](#). *Polyhedron* **20**, 2889-2900 (2001).
83. J. E. Reed, A. J. P. White, S. Neidle, R. Vilar, [Effect of Metal Coordination on the Interaction of Substituted Phenanthroline and Pyridine Ligands with Quadruplex DNA](#). *Dalton Trans.*, 2558 (2009).
84. W. Zhang, C. Bensimon, R. J. Crutchley, [\(Terpyridine\)Palladium\(II\) Complexes of Phenylcyanamide Ligands](#). *Inorg. Chem.* **32**, 5808-5812 (1993).
85. M.-Y. Kim, H. Vankayalapati, K. Shin-ya, K. Wierzb, L. H. Hurley, [Telomestatin, a Potent Telomerase Inhibitor That Interacts Quite Specifically with the Human Telomeric Intramolecular G-quadruplex](#). *J. Am. Chem. Soc.* **124**, 2098-2099 (2002).
86. F. Hamon, E. Largy, A. Guédin-Beaurepaire, M. Rouchon-Dagois, A. Sidibe, D. Monchaud, J.-L. Mergny, J.-F. Riou, C.-H. Nguyen, M.-P. Teulade-Fichou, [An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-quadruplex Topologies](#). *Angew. Chem. Int. Ed.* **50**, 8745-8749 (2011).
87. M. Petenzi, D. Verga, E. Largy, F. Hamon, F. Doria, M.-P. Teulade-Fichou, A. Guédin, J.-L. Mergny, M. Mella, M. Freccero, [Cationic Pentaheteroaryls as Selective G-quadruplex Ligands by Solvent-free Microwave-assisted Synthesis](#). *Chem. – Eur. J.* **18**, 14487-14496 (2012).
88. F. Besselièvre, F. Mahuteau-Betzer, D. S. Grierson, S. Piguel, [Ligandless Microwave-Assisted Pd/Cu-catalyzed Direct Arylation of Oxazoles](#). *J. Org. Chem.* **73**, 3278-3280 (2008).
89. A. Zaid, J.-S. Sun, C.-H. Nguyen, E. Bisagni, T. Garestier, D. S. Grierson, R. Zain, [Triple-helix Directed Cleavage of Double-stranded DNA by Benzoquinoxaline-1,10-phenanthroline Conjugates](#). *ChemBioChem* **5**, 1550-1557 (2004).
90. A. Renaud De La Faverie, F. Hamon, C. Di Primo, E. Largy, E. Dausse, L. Delaurière, C. Landras-Guetta, J.-J. Toulmé, M.-P. Teulade-Fichou, J.-L. Mergny, [Nucleic Acids Targeted to Drugs : SELEX against a Quadruplex Ligand](#). *Biochimie* **93**, 1357-1367 (2011).
91. E. Largy, F. Hamon, M.-P. Teulade-Fichou, [A Streptavidin Paramagnetic-Particle Based Competition Assay for the Evaluation of the Optical Selectivity of Quadruplex Nucleic Acid Fluorescent Probes](#). *Methods* **57**, 129-137 (2012).
92. A. Sidibe, F. Hamon, E. Largy, D. Gomez, M.-P. Teulade-Fichou, C. Trentesaux, J.-F. Riou, [Effects of a Halogenated G-quadruplex Ligand from the Pyridine Dicarboxamide Series on the Terminal Sequence of XpYp Telomere in HT1080 Cells](#). *Biochimie* **94**, 2559-2568 (2012).
93. X. Xie, B. Choi, E. Largy, R. Guillot, A. Granzhan, M.-P. Teulade-Fichou, [Asymmetric Distyrylpyridinium Dyes as Red-emitting Fluorescent Probes for Quadruplex DNA](#). *Chem. – Eur. J.* **19**, 1214-1226 (2013).

94. F. Koeppel, [Ethidium Derivatives Bind to G-quartets, Inhibit Telomerase and Act as Fluorescent Probes for Quadruplexes](#). *Nucleic Acids Res.* **29**, 1087-1096 (2001).
95. Z.-y. Kan, Y. Lin, F. Wang, X.-y. Zhuang, Y. Zhao, D.-w. Pang, Y.-h. Hao, Z. Tan, [G-Quadruplex Formation in Human Telomeric \(TTAGGG\)<sub>4</sub> Sequence with Complementary Strand in Close Vicinity under Molecularly Crowded Condition](#). *Nucleic Acids Res.* **35**, 3646-3653 (2007).
96. G. W. Collie, G. N. Parkinson, S. Neidle, F. Rosu, E. De Pauw, V. Gabelica, [Electrospray Mass Spectrometry of Telomeric RNA \(TERRA\) Reveals the Formation of Stable Multimeric G-quadruplex Structures](#). *J. Am. Chem. Soc.* **132**, 9328-9334 (2010).
97. Y.-W. Ham, W. C. Tse, D. L. Boger, [High-Resolution Assessment of Protein DNA Binding Affinity and Selectivity Utilizing a Fluorescent Intercalator Displacement \(FID\) Assay](#). *Bioorg. Med. Chem. Lett.* **13**, 3805-3807 (2003).
98. D. Boger, [Thiazole Orange as the Fluorescent Intercalator in a High Resolution FID Assay for Determining DNA Binding Affinity and Sequence Selectivity of Small Molecules](#). *Bioorg. Med. Chem.* **9**, 2511-2518 (2001).
99. B. K. S. Yeung, W. C. Tse, D. L. Boger, [Determination of Binding Affinities of Triplex Forming Oligonucleotides Using a Fluorescent Intercalator Displacement \(FID\) Assay](#). *Bioorg. Med. Chem. Lett.* **13**, 3801-3804 (2003).
100. D. L. Boger, B. E. Fink, M. P. Hedrick, [Total Synthesis of Distamycin a and 2640 Analogues : A Solution-Phase Combinatorial Approach to the Discovery of New, Bioactive DNA Binding Agents and Development of a Rapid, High-Throughput Screen for Determining Relative DNA Binding Affinity or DNA Binding Sequence Selectivity](#). *J. Am. Chem. Soc.* **122**, 6382-6394 (2000).
101. E. Largy, F. Hamon, M.-P. Teulade-Fichou, [Development of a High-Throughput G4-FID Assay for Screening and Evaluation of Small Molecules Binding Quadruplex Nucleic Acid Structures](#). *Anal. Bioanal. Chem.* **400**, 3419-3427 (2011).
102. P. L. T. Tran, E. Largy, F. Hamon, M.-P. Teulade-Fichou, J.-L. Mergny, [Fluorescence Intercalator Displacement Assay for Screening G4 Ligands towards a Variety of G-quadruplex Structures](#). *Biochimie* **93**, 1288-1296 (2011).
103. E. Largy, M.-P. Teulade-Fichou, « [Screening for Quadruplex Binding Ligands : A Game of Chance ?](#) » (Royal Society of Chemistry, 2012), p. 248-262.
104. E. Largy, N. Saettel, F. Hamon, S. Dubruille, M.-P. Teulade-Fichou, [Screening of a Chemical Library by HT-G4-FID for Discovery of Selective G-quadruplex Binders](#). *Curr. Pharm. Des.* **18**, 1992-2001 (2012).
105. R. R. Iyer, A. Pluciennik, V. Burdett, P. L. Modrich, [DNA Mismatch Repair : Functions and Mechanisms](#). *Chem. Rev.* **106**, 302-323 (2006).
106. L. C. Young, J. B. Hays, V. A. Tron, S. E. Andrew, [DNA Mismatch Repair Proteins : Potential Guardians against Genomic Instability and Tumorigenesis Induced by Ultraviolet Photoproducts](#). *J. Invest. Dermatol.* **121**, 435-440 (2003).
107. A. Granzhan, E. Largy, N. Saettel, M.-P. Teulade-Fichou, [Macrocyclic DNA-mismatch-binding Ligands : Structural Determinants of Selectivity](#). *Chem. – Eur. J.* **16**, 878-889 (2010).
108. M. Wieland, J. S. Hartig, [RNA Quadruplex-Based Modulation of Gene Expression](#). *Chem. Biol.* **14**, 757-763 (2007).

## Bibliographie

109. J.-D. Beaudoin, J.-P. Perreault, [5'-UTR G-quadruplex Structures Acting as Translational Repressors](#). *Nucleic Acids Res.* **38**, 7022-7036 (2010).
110. S. Kumari, A. Bugaut, J. L. Huppert, S. Balasubramanian, [An RNA G-quadruplex in the 5' UTR of the NRAS Proto-Oncogene Modulates Translation](#). *Nat. Chem. Biol.* **3**, 218-221 (2007).
111. A. Arora, M. Dutkiewicz, V. Scaria, M. Hariharan, S. Maiti, J. Kurreck, [Inhibition of Translation in Living Eukaryotic Cells by an RNA G-quadruplex Motif](#). *RNA* **14**, 1290-1296 (2008).
112. M. J. Morris, S. Basu, [An Unusually Stable G-quadruplex within the 5'-UTR of the MT3 Matrix Metalloproteinase mRNA Represses Translation in Eukaryotic Cells](#). *Biochemistry* **48**, 5313-5319 (2009).
113. K. Halder, M. Wieland, J. S. Hartig, [Predictable Suppression of Gene Expression by 5'-UTR-based RNA Quadruplexes](#). *Nucleic Acids Res.* **37**, 6811-6817 (2009).
114. K. Halder, E. Largy, M. Benzler, M.-P. Teulade-Fichou, J. S. Hartig, [Efficient Suppression of Gene Expression by Targeting 5'-UTR-based RNA Quadruplexes with Bisquinolinium Compounds](#). *Chem-BioChem* **12**, 1663-1668 (2011).
115. G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, [Quantitative Visualization of DNA G-quadruplex Structures in Human Cells](#). *Nat. Chem.* **5**, 182-186 (2013).
116. E. Y. N. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, [G-Quadruplex Structures Are Stable and Detectable in Human Genomic DNA](#). *Nat. Commun.* **4**, 1796 (2013).
117. H. G. Kazemier, K. Paeschke, P. M. Lansdorp, [Guanine Quadruplex Monoclonal Antibody 1H6 Cross-React with Restrained Thymidine-Rich Single Stranded DNA](#). *Nucleic Acids Res.* **45**, 5913-5919 (2017).
118. M. Nadai, S. N. Richter, « [G-Quadruplex Visualization in Cells via Antibody and Fluorescence Probe](#) » (Springer New York, 2019), p. 383-395.
119. I. Maurizio, B. Tosoni, I. Gallina, E. Ruggiero, I. Zanin, S. N. Richter, « [Production of the Anti-G-quadruplex Antibody BG4 for Efficient Genome-Wide Analyses : From Plasmid Quality Control to Antibody Validation](#) » (Elsevier, 2024), p. 193-219.
120. R. Halder, J.-F. Riou, M.-P. Teulade-Fichou, T. Frickey, J. S. Hartig, [Bisquinolinium Compounds Induce Quadruplex-Specific Transcriptome Changes in HeLa S3 Cell Lines](#). *BMC Res. Notes* **5**, 138 (2012).
121. M. Hernandez, S. M. Cavalcanti, D. R. Moreira, W. De Azevedo Junior, A. C. Leite, [Halogen Atoms in the Modern Medicinal Chemistry : Hints for the Drug Design](#). *Curr. Drug Targets* **11**, 303-314 (2010).
122. S. Neidle, [Human Telomeric G-quadruplex : The Current Status of Telomeric G-quadruplexes as Therapeutic Targets in Human Cancer](#). *FEBS J.* **277**, 1118-1125 (2010).
123. M. Folini, L. Venturini, G. Cimino-Reale, N. Zaffaroni, [Telomeres as Targets for Anticancer Therapies](#). *Expert Opin. Ther. Targets* **15**, 579-593 (2011).
124. J. Gao, H. A. Pickett, [Targeting Telomeres : Advances in Telomere Maintenance Mechanism-Specific Cancer Therapies](#). *Nat. Rev. Cancer* **22**, 515-532 (2022).
125. A. Tsalmpouris, E. Largy, D. Guillarme, [Systematic Evaluation of Phenyl Stationary Phases for Oligonucleotide Analysis without Ion-Pairing Reagents](#). *J. Chromatogr. A* **1762**, 466395 (2025).



126. C. Granotier, [Preferential Binding of a G-quadruplex Ligand to Human Chromosome Ends](#). *Nucleic Acids Res.* **33**, 4182-4190 (2005).
127. J. Ren, J. B. Chaires, [Sequence and Structural Selectivity of Nucleic Acid Binding Ligands](#). *Biochemistry* **38**, 16067-16075 (1999).
128. A. De Cian, L. Guittat, K. Shin-ya, J.-F. Riou, J.-L. Mergny, [Affinity and Selectivity of G4 Ligands Measured by FRET](#). *Nucleic Acids Symp. Ser.* **49**, 235-236 (2005).
129. P. Yang, A. De Cian, M.-P. Teulade-Fichou, J.-L. Mergny, D. Monchaud, [Engineering Bisquinolinium/Thiazole Orange Conjugates for Fluorescent Sensing of G-quadruplex DNA](#). *Angew. Chem. Int. Ed.* **48**, 2188-2191 (2009).
130. B. Dumat, G. Bordeau, E. Faurel-Paul, F. Mahuteau-Betzer, N. Saettel, M. Bombled, G. Metgé, F. Charra, C. Fiorini-Debuisschert, M.-P. Teulade-Fichou, [N-Phenyl-Carbazole-Based Two-Photon Fluorescent Probes : Strong Sequence Dependence of the Duplex vs Quadruplex Selectivity](#). *Biochimie* **93**, 1209-1218 (2011).
131. B. Prasad, M. Doimo, M. Andréasson, V. L'Hôte, E. Chorell, S. Wanrooij, [A Complementary Chemical Probe Approach towards Customized Studies of G-quadruplex DNA Structures in Live Cells](#). *Chem. Sci.* **13**, 2347-2354 (2022).
132. T. Le Doan, L. Perrouault, D. Praseuth, N. Habhouh, J.-L. Decout, N. T. Thuong, J. Lhomme, C. Héène, [Sequence-Specific Recognition, Photocrosslinking and Cleavage of the DNA Double Helix by an Oligo-\( \$\alpha\$ \)-Thymidylate Covalently Linked to an Azidoproflavine Derivative](#). *Nucleic Acids Res.* **15**, 7749-7760 (1987).
133. P. B. Dervan, [Design of Sequence-Specific DNA-binding Molecules](#). *Science* **232**, 464-471 (1986).
134. T. G. Uil, [Therapeutic Modulation of Endogenous Gene Function by Agents with Designed DNA-sequence Specificities](#). *Nucleic Acids Res.* **31**, 6064-6078 (2003).
135. N. Branda, G. Kurz, J.-M. Lehn, [JANUS WEDGES : A New Approach towards Nucleobase-Pair Recognition](#). *Chem. Commun.*, 2443 (1996).
136. D. Chen, Meena, S. K. Sharma, L. W. McLaughlin, [Formation and Stability of a Janus-Wedge Type of DNA Triplex](#). *J. Am. Chem. Soc.* **126**, 70-71 (2004).
137. Y. Zeng, Y. Pratumyot, X. Piao, D. Bong, [Discrete Assembly of Synthetic Peptide–DNA Triplex Structures from Polyvalent Melamine–Thymine Bifacial Recognition](#). *J. Am. Chem. Soc.* **134**, 832-835 (2012).
138. J. Sági, A. Szemző, J. Szécsi, L. Ötvös, [Biochemical Properties of Oligo\[+\]-Carbocyclic-Thymidylates\] and Their Complexes](#). *Nucleic Acids Res.* **18**, 2133-2140 (1990).
139. M. Renders, S. Dumbre, M. Abramov, D. Kestemont, L. Margamuljana, E. Largy, C. Cozens, J. Vandenameele, V. B. Pinheiro, D. Toye, J.-M. Frère, P. Herdewijn, [Kinetic Analysis of N -Alkylaryl Carboxamide Hexitol Nucleotides as Substrates for Evolved Polymerases](#). *Nucleic Acids Res.* **47**, 2160-2168 (2019).
140. K. E. Lundin, O. Gissberg, C. I. E. Smith, [Oligonucleotide Therapies : The Past and the Present](#). *Hum. Gene Ther.* **26**, 475-485 (2015).
141. C. F. Bennett, [Therapeutic Antisense Oligonucleotides Are Coming of Age](#). *Annu. Rev. Med.* **70**, 307-321 (2019).



142. A. Goga, M. Stoffel, [Therapeutic RNA-silencing Oligonucleotides in Metabolic Diseases](#). *Nat. Rev. Drug Discovery* **21**, 417-439 (2022).
143. J. Kim, S. Woo, C. M. De Gusmao, B. Zhao, D. H. Chin, R. L. DiDonato, M. A. Nguyen, T. Nakayama, C. A. Hu, A. Soucy, A. Kuniholm, J. K. Thornton, O. Riccardi, D. A. Friedman, C. M. El Achkar, Z. Dash, L. Cornelissen, C. Donado, K. N. W. Faour, L. W. Bush, V. Suslovitch, C. Lentucci, P. J. Park, E. A. Lee, A. Patterson, A. A. Philippakis, B. Margus, C. B. Berde, T. W. Yu, [A Framework for Individualized Splice-Switching Oligonucleotide Therapy](#). *Nature* **619**, 828-836 (2023).
144. E. Largy, J.-L. Mergny, [Shape Matters : Size-Exclusion HPLC for the Study of Nucleic Acid Structural Polymorphism](#). *Nucleic Acids Res.* **42**, e149-e149 (2014).
145. J. Dai, E. Hatzakis, L. H. Hurley, D. Yang, [I-Motif Structures Formed in the Human c-MYC Promoter Are Highly Dynamic–Insights into Sequence Redundancy and I-motif Stability](#). *PLOS One* **5**, e11647 (2010).
146. N. Q. Do, K. W. Lim, M. H. Teo, B. Heddi, A. T. Phan, [Stacking of G-quadruplexes : NMR Structure of a G-rich Oligonucleotide with Potential Anti-HIV and Anticancer Activity†](#). *Nucleic Acids Res.* **39**, 9448-9457 (2011).
147. G. D. Balkwill, T. P. Garner, H. E. L. Williams, M. S. Searle, [Folding Topology of a Bimolecular DNA Quadruplex Containing a Stable Mini-Hairpin Motif within the Diagonal Loop](#). *J. Mol. Biol.* **385**, 1600-1615 (2009).
148. A. T. Phan, V. Kuryavyi, J.-B. Ma, A. Faure, M.-L. Andréola, D. J. Patel, [An Interlocked Dimeric Parallel-Stranded DNA Quadruplex : A Potent Inhibitor of HIV-1 Integrase](#). *Proc. Natl. Acad. Sci.* **102**, 634-639 (2005).
149. E.-A. Raiber, R. Kranaster, E. Lam, M. Nikan, S. Balasubramanian, [A Non-Canonical DNA Structure Is a Binding Motif for the Transcription Factor SP1 in Vitro](#). *Nucleic Acids Res.* **40**, 1499-1508 (2012).
150. M. Trajkovski, M. Webba Da Silva, J. Plavec, [Unique Structural Features of Interconverting Monomeric and Dimeric G-quadruplexes Adopted by a Sequence from the Intron of the N-myc Gene](#). *J. Am. Chem. Soc.* **134**, 4132-4141 (2012).
151. J. Zhou, S. Amrane, F. Rosu, G. F. Salgado, Y. Bian, H. Tateishi-Karimata, E. Largy, D. N. Korkut, A. Bourdoncle, D. Miyoshi, J. Zhang, H. Ju, W. Wang, N. Sugimoto, V. Gabelica, J.-L. Mergny, [Unexpected Position-Dependent Effects of Ribose G-quartets in G-quadruplexes](#). *J. Am. Chem. Soc.* **139**, 7768-7779 (2017).
152. E. Largy, V. Gabelica, [Native Hydrogen/Deuterium Exchange Mass Spectrometry of Structured DNA Oligonucleotides](#). *Anal. Chem.* **92**, 4402-4410 (2020).
153. V. Voynov, J. A. Caravella, Éd., [Therapeutic Proteins](#) (Humana Press, 2012).
154. A. Beck, E. Wagner-Rousset, D. Ayoub, A. Van Dorsselaer, S. Sanglier-Cianférani, [Characterization of Therapeutic Antibodies and Related Products](#). *Anal. Chem.* **85**, 715-736 (2013).
155. F. Debaene, A. Bœuf, E. Wagner-Rousset, O. Colas, D. Ayoub, N. Corvaia, A. Van Dorsselaer, A. Beck, S. Cianférani, [Innovative Native MS Methodologies for Antibody Drug Conjugate Characterization : High Resolution Native MS and IM-MS for Average DAR and DAR Distribution Assessment](#). *Anal. Chem.* **86**, 10674-10683 (2014).
156. H. Liu, G. Ponniah, H.-M. Zhang, C. Nowak, A. Neill, N. Gonzalez-Lopez, R. Patel, G. Cheng, A. Z. Kita, B. Andrien, [In Vitro and in Vivo Modifications of Recombinant and Human IgG Antibodies](#). *mAbs* **6**, 1145-1154 (2014).

157. N. Jain, S. W. Smith, S. Ghone, B. Tomczuk, [Current ADC Linker Chemistry](#). *Pharm. Res.* **32**, 3526-3540 (2015).
158. F. Cantais, A. Catrain, E. Largy, G. Van Vyncht, A. Delobel, « Characterisation of Humira Adalimumab by High Resolution Mass Spectrometry » (Quality Assitance, Donstiennes, Belgique, 2015).
159. A. Delobel, E. Largy, F. Cantais, A. Catrain, G. Van Vyncht, « Characterisation of Antibody-Drug Conjugates by High-Resolution Mass Spectrometry » (Quality Assitance, Donstiennes, Belgique, 2016).
160. E. Largy, A. Catrain, G. Van Vyncht, A. Delobel, 2D-LC-MS for the Analysis of Monoclonal Antibodies and Antibody-Drug Conjugates in a Regulated Environment. *LCGC : Spec. Issues* **14**, 29-35 (2016).
161. A. Delobel, F. Dusserre-Bresson, E. Largy, C. Lepers, C. Mantini, C. Plennevaux, F. Vandermeers, « Analytical Approaches for the Evaluation of ADCC Activity of mAbs » (Quality Assitance, Donstiennes, Belgique, 2017).
162. A. Beck, E. Wagner-Rousset, M.-C. Bussat, M. Lokteff, C. Klinguer-Hamour, J.-F. Haeuw, L. Goetsch, T. Wurch, A. Dorsselaer, N. Corvaia, [Trends in Glycosylation, Glycoanalysis and Glycoengineering of Therapeutic Antibodies and Fc-Fusion Proteins](#). *Curr. Pharm. Biotechnol.* **9**, 482-501 (2008).
163. E. Largy, F. Cantais, G. Van Vyncht, A. Beck, A. Delobel, [Orthogonal Liquid Chromatography-Mass Spectrometry Methods for the Comprehensive Characterization of Therapeutic Glycoproteins, from Released Glycans to Intact Protein Level](#). *J. Chromatogr. A* **1498**, 128-146 (2017).
164. A. Delobel, E. Largy, F. Cantais, G. Van Vyncht, « Characterisation of the Glycosylation of Enbrel : Etanercept by High Resolution Mass Spectrometry » (Quality Assitance, Donstiennes, Belgique, 2016).
165. E. Largy, F. Cantais, G. Van Vyncht, A. Delobel, « Characterisation of the Glycosylation of Humira (Adalimumab) and Erbitux (Cetuximab) by High Resolution Mass Spectrometry » (Quality Assitance, Donstiennes, Belgique, 2016).
166. C. I. Butré, E. Largy, A. Delobel, « [Profiling of N-Linked Oligosaccharides of a Glycoprotein by UPLC-FLR-ESI-MS after Derivatization with Fluorescent Anthranilamide](#) » (Springer US, 2021), p. 179-188.
167. C. I. Butré, E. Largy, F. Cantais, A. Delobel, « [Profiling, Relative Quantification, and Identification of Sialylated n-Linked Oligosaccharides by UPLC-FLR-ESI/MS after Derivatization with Fluorescent Anthranilamide](#) » (Springer US, 2021), p. 237-247.
168. V. D'Atri, D. Guillarme, « [Characterization of Glycosylated Proteins at Subunit Level by HILIC/MS](#) » (Springer US, 2021), p. 85-95.
169. H. Lardeux, D. Guillarme, V. D'Atri, [Comprehensive Evaluation of Zwitterionic Hydrophilic Liquid Chromatography Stationary Phases for Oligonucleotide Characterization](#). *J. Chromatogr. A* **1690**, 463785 (2023).
170. B. L. Duivelshof, A. Beck, D. Guillarme, V. D'Atri, [Bispecific Antibody Characterization by a Combination of Intact and Site-Specific/Chain-Specific LC/MS Techniques](#). *Talanta* **236**, 122836 (2022).
171. B. L. Duivelshof, S. Denorme, K. Sandra, X. Liu, A. Beck, M. A. Lauber, D. Guillarme, V. D'Atri, [Quantitative N-glycan Profiling of Therapeutic Monoclonal Antibodies Performed by Middle-up Level HILIC-HRMS Analysis](#). *Pharmaceutics* **13**, 1744 (2021).

172. B. Bobály, V. D'Atri, A. Beck, D. Guilleme, S. Fekete, [Analysis of Recombinant Monoclonal Antibodies in Hydrophilic Interaction Chromatography : A Generic Method Development Approach](#). *J. Pharm. Biomed. Anal.* **145**, 24-32 (2017).
173. J. R. Engen, T. Botzanowski, D. Peterle, F. Georgescauld, T. E. Wales, [Developments in Hydrogen/Deuterium Exchange Mass Spectrometry](#). *Anal. Chem.* **93**, 567-582 (2021).
174. G. R. Masson, J. E. Burke, N. G. Ahn, G. S. Anand, C. Borchers, S. Brier, G. M. Bou-Assaf, J. R. Engen, S. W. Englander, J. Faber, R. Garlish, P. R. Griffin, M. L. Gross, M. Guttman, Y. Hamuro, A. J. R. Heck, D. Houde, R. E. Jacob, T. J. D. Jørgensen, I. A. Kaltashov, J. P. Klinman, L. Konermann, P. Man, L. Mayne, B. D. Pascal, D. Reichmann, M. Skehel, J. Snijder, T. S. Strutzenberg, E. S. Underbakke, C. Wagner, T. E. Wales, B. T. Walters, D. D. Weis, D. J. Wilson, P. L. Wintrobe, Z. Zhang, J. Zheng, D. C. Schriemer, K. D. Rand, [Recommendations for Performing, Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry \(HDX-MS\) Experiments](#). *Nat. Methods* **16**, 595-602 (2019).
175. E. Largy, A. Delobel, « Stress Testing of Humira (Adalimumab) Monitored by Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) » (Quality Assistance, Donstiennes, Belgique, 2017).
176. E. Largy, C. Cajot, A. Delobel, « Epitope Mapping of an Interleukin Receptor for Three Therapeutic Antibodies by Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) » (Quality Assistance, Donstiennes, Belgique, 2017).
177. C. Cajot, A. Delobel, E. Largy, Epitope Mapping of an Interleukin Receptor for Three Therapeutic Antibodies by Hydrogen-Deuterium Exchange Mass Spectrometry. *LCGC : Spec. Issues* **16**, 8-14 (2018).
178. L. Belarif, C. Mary, L. Jacquemont, H. L. Mai, R. Danger, J. Hervouet, D. Minault, V. Thepenier, V. Nèrière-Daguin, E. Nguyen, S. Pengam, E. Largy, A. Delobel, B. Martinet, S. Le Bas-Bernardet, S. Brouard, J.-P. Soulillou, N. Degauque, G. Blancho, B. Vanhove, N. Poirier, [IL-7 Receptor Blockade Blunts Antigen-Specific Memory T Cell Responses and Chronic Inflammation in Primates](#). *Nat. Commun.* **9**, 4483 (2018).
179. E. Largy, M. Ranz, V. Gabelica, DNA-HDXMS\_XchangeDB : A Dataset of Hydrogen-Deuterium eXchange Native Mass Spectrometry Experiments on DNA Oligonucleotides (2023). <https://doi.org/10.5281/ZENODO.7713144>.
180. P. H. Von Hippel, N. P. Johnson, A. H. Marcus, [Fifty Years of DNA « Breathing » : Reflections on Old and New Approaches](#). *Biopolymers* **99**, 923-954 (2013).
181. A. Kabir, M. Bhattarai, K. Ø. Rasmussen, A. Shehu, A. Usheva, A. R. Bishop, B. Alexandrov, [Examining DNA Breathing with pyDNA-EPBD](#). *Bioinformatics* **39**, btad699 (2023).
182. V. Gabelica, E. Marklund, [Fundamentals of Ion Mobility Spectrometry](#). *Curr. Opin. Chem. Biol.* **42**, 51-59 (2018).
183. B. M. Hossain, L. Konermann, [Pulsed Hydrogen/Deuterium Exchange MS/MS for Studying the Relationship between Noncovalent Protein Complexes in Solution and in the Gas Phase after Electrospray Ionization](#). *Anal. Chem.* **78**, 1613-1619 (2006).
184. I. D. Lutz, S. Wang, C. Norn, A. Courbet, A. J. Borst, Y. T. Zhao, A. Dosey, L. Cao, J. Xu, E. M. Leaf, C. Treichel, P. Litvicov, Z. Li, A. D. Goodson, P. Rivera-Sánchez, A.-M. Bratovianu, M. Baek, N. P. King, H. Ruohola-Baker, D. Baker, [Top-down Design of Protein Architectures with Reinforcement Learning](#). *Science* **380**, 266-273 (2023).
185. P. Murat, Y. Singh, E. Defrancq, [Methods for Investigating G-quadruplex DNA/Ligand Interactions](#). *Chem. Soc. Rev.* **40**, 5293 (2011).

186. P. L. Ferguson, J. Pan, D. J. Wilson, B. Dempsey, G. Lajoie, B. Shilton, L. Konermann, [Hydrogen/Deuterium Scrambling during Quadrupole Time-of-Flight MS/MS Analysis of a Zinc-Binding Protein Domain](#). *Anal. Chem.* **79**, 153-160 (2007).
187. R. R. Abzalimov, D. A. Kaplan, M. L. Easterling, I. A. Kaltashov, [Protein Conformations Can Be Probed in Top-down HDX MS Experiments Utilizing Electron Transfer Dissociation of Protein Ions without Hydrogen Scrambling](#). *J. Am. Soc. Mass Spectrom.* **20**, 1514-1517 (2009).
188. K. D. Rand, M. Zehl, T. J. D. Jørgensen, [Measuring the Hydrogen/Deuterium Exchange of Proteins at High Spatial Resolution by Mass Spectrometry : Overcoming Gas-Phase Hydrogen/Deuterium Scrambling](#). *Acc. Chem. Res.* **47**, 3018-3027 (2014).
189. Q. Wang, N. B. Borotto, K. Håkansson, [Gas-Phase Hydrogen/Deuterium Scrambling in Negative-Ion Mode Tandem Mass Spectrometry](#). *J. Am. Soc. Mass Spectrom.* **30**, 855-863 (2019).
190. D. T. W. Wollenberg, S. Pengelley, J. C. Mouritsen, D. Suckau, C. I. Jørgensen, T. J. D. Jørgensen, [Avoiding H/D Scrambling with Minimal Ion Transmission Loss for HDX-MS/MS-ETD Analysis on a High-Resolution Q-TOF Mass Spectrometer](#). *Anal. Chem.* **92**, 7453-7461 (2020).
191. C. C. Mundorff, S. Hadley, L. M. Tuttle, Y. Shi, G. C. McAlister, R. Viner, R. E. Klevit, M. Guttman, [Mapping Hydrogen Migration Thresholds for Site-Specific HDX-MS](#). *Mol. Cell. Proteomics* **24**, 101075 (2025).
192. J. Anacleto, E. Kabir, M. Blanco, Y. Leblanc, C. Lento, D. J. Wilson, [Efficient, Zero Scrambling Fragmentation of Deuterium Labeled Peptides on the ZenoToF 7600 Electron Activated Dissociation Platform](#). *J. Am. Soc. Mass Spectrom.* **36**, 1175-1181 (2025).
193. N. Khristenko, F. Rosu, E. Largy, J. Haustant, C. Mesmin, V. Gabelica, [Native Electrospray Ionization of Multi-Domain Proteins via a Bead Ejection Mechanism](#). *J. Am. Chem. Soc.* **145**, 498-506 (2023).
194. A. Berner, R. N. Das, N. Bhuma, J. Golebiewska, A. Abrahamsson, M. Andréasson, N. Chaudhari, M. Doimo, P. P. Bose, K. Chand, R. Strömberg, S. Wanrooij, E. Chorell, [G4-Ligand-Conjugated Oligonucleotides Mediate Selective Binding and Stabilization of Individual G4 DNA Structures](#). *J. Am. Chem. Soc.* **146**, 6926-6935 (2024).
195. G. Qin, Z. Liu, J. Yang, X. Liao, C. Zhao, J. Ren, X. Qu, [Targeting Specific DNA G-quadruplexes with CRISPR-guided G-quadruplex-binding Proteins and Ligands](#). *Nat. Cell Biol.* **26**, 1212-1224 (2024).
196. S. Kumar, S. P. P. Pany, S. Sudhakar, S. B. Singh, C. S. Todankar, P. I. Pradeepkumar, [Targeting Parallel Topology of G-quadruplex Structures by Indole- Fused Quindoline Scaffolds](#). *Biochemistry* **61**, 2546-2559 (2022).
197. A. Cheng, C. Liu, W. Ye, D. Huang, W. She, X. Liu, C. P. Fung, N. Xu, M. C. Suen, W. Ye, H. H. Y. Sung, I. D. Williams, G. Zhu, P.-Y. Qian, [Selective C9orf72 G-quadruplex-binding Small Molecules Ameliorate Pathological Signatures of ALS/FTD Models](#). *J. Med. Chem.* **65**, 12825-12837 (2022).
198. R. Wang, M.-H. Hu, [Development of a Fluorescent Ligand That Specifically Binds to the C-MYC G-quadruplex by Migrating the Benzene Group on a Carbazole-Benzothiazolium Scaffold](#). *Bioorg. Chem.* **151**, 107690 (2024).
199. W. Long, B.-X. Zheng, Y. Li, X.-H. Huang, D.-M. Lin, C.-C. Chen, J.-Q. Hou, T.-M. Ou, W.-L. Wong, K. Zhang, Y.-J. Lu, [Rational Design of Small-Molecules to Recognize G-quadruplexes of c-MYC Promoter and Telomere and the Evaluation of Their \*in Vivo\* Antitumor Activity against Breast Cancer](#). *Nucleic Acids Res.* **50**, 1829-1848 (2022).

## Bibliographie

200. X. Lu, L. F. M. Passalacqua, M. Nodwell, K. Y. S. Kong, G. Caballero-García, E. V. Dolgosheina, A. R. Ferré-D'Amaré, R. Britton, P. J. Unrau, [Symmetry Breaking of Fluorophore Binding to a G-quadruplex Generates an RNA Aptamer with Picomolar KD](#). *Nucleic Acids Res.* **52**, 8039-8051 (2024).
201. D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, J. S. Moore, [A Field Guide to Foldamers](#). *Chem. Rev.* **101**, 3893-4012 (2001).
202. P. V. Jena, P. S. Shirude, B. Okumus, K. Laxmi-Reddy, F. Godde, I. Huc, S. Balasubramanian, T. Ha, [G-Quadruplex DNA Bound by a Synthetic Ligand Is Highly Dynamic](#). *J. Am. Chem. Soc.* **131**, 12522-12523 (2009).
203. S. Müller, K. Laxmi-Reddy, P. V. Jena, B. Baptiste, Z. Dong, F. Godde, T. Ha, R. Rodriguez, S. Balasubramanian, I. Huc, [Targeting DNA G-quadruplexes with Helical Small Molecules](#). *ChemBioChem* **15**, 2563-2570 (2014).
204. P. K. Mandal, B. Baptiste, B. Langlois d'Estaintot, B. Kauffmann, I. Huc, [Multivalent Interactions between an Aromatic Helical Foldamer and a DNA G-quadruplex in the Solid State](#). *ChemBioChem* **17**, 1911-1914 (2016).
205. P. Kuzmič, [Program DYNAFIT for the Analysis of Enzyme Kinetic Data : Application to HIV Proteinase](#). *Anal. Biochem.* **237**, 260-273 (1996).
206. P. Kuzmič, « [DynaFit—a Software Package for Enzymology](#) » (Elsevier, 2009), p. 247-280.
207. C. Liu, B. Zhou, Y. Geng, D. Yan Tam, R. Feng, H. Miao, N. Xu, X. Shi, Y. You, Y. Hong, B. Z. Tang, P. Kwan Lo, V. Kuryavyi, G. Zhu, [A Chair-Type G-quadruplex Structure Formed by a Human Telomeric Variant DNA in K<sup>+</sup> Solution](#). *Chem. Sci.* **10**, 218-226 (2019).
208. H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle, S. Balasubramanian, [A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human C-Kit Oncogene](#). *Biochemistry* **45**, 7854-7860 (2006).
209. R. Rigo, W. L. Dean, R. D. Gray, J. B. Chaires, C. Sissi, [Conformational Profiling of a G-rich Sequence within the c-KIT Promoter](#). *Nucleic Acids Res.* **45**, 13056-13067 (2017).
210. J. Guillon, M. Le Borgne, V. Milano, A. Guédin-Beaurepaire, S. Moreau, N. Pinaud, L. Ronga, S. Savrimoutou, S. Albenque-Rubio, M. Marchivie, H. Kalout, C. Walker, L. Chevallier, C. Buré, E. Largy, V. Gabelica, J.-L. Mergny, V. Baylot, J. Ferrer, Y. Idrissi, E. Chevret, D. Cappellen, V. Desplat, Z. Schelz, I. Zupkó, [New 2,4-Bis\[\(Substituted-Aminomethyl\)Phenyl\]Phenylquinazoline and 2,4-Bis\[\(Substituted-Aminomethyl\)Phenyl\]Phenylquinoline Derivatives : Synthesis and Biological Evaluation as Novel Anticancer Agents by Targeting G-quadruplex](#). *Pharmaceuticals* **17**, 30 (2023).
211. E. Largy, A. Guédin, A. Kabbara, J.-L. Mergny, S. Amrane, [Eps2Fold : A Rapid Method to Characterize G-quadruplex DNA Structures Using Single Absorbance Spectra](#). *Nucleic Acids Res.* **53**, gkaf953 (2025).
212. J.-L. Mergny, [Thermal Difference Spectra : A Specific Signature for Nucleic Acid Structures](#). *Nucleic Acids Res.* **33**, e138-e138 (2005).
213. J. Kypr, I. Kejnovska, D. Renciuik, M. Vorlickova, [Circular Dichroism and Conformational Polymorphism of DNA](#). *Nucleic Acids Res.* **37**, 1713-1725 (2009).
214. I. Kejnovská, D. Renciuik, J. Palacký, M. Vorlíčková, « [CD Study of the G-Quadruplex Conformation](#) » (Springer New York, 2019), p. 25-44.



215. R. del Villar-Guerra, J. O. Trent, J. B. Chaires, [G-quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy](#). *Angew. Chem. Int. Ed.* **57**, 7171-7175 (2018).
216. R. Del Villar-Guerra, R. D. Gray, J. B. Chaires, [Characterization of Quadruplex DNA Structure by Circular Dichroism](#). *Curr. Protoc. Nucleic Acid Chem.* **68** (2017).
217. J.-L. Mergny, L. Lacroix, [Analysis of Thermal Melting Curves](#). *Oligonucleotides* **13**, 515-537 (2003).
218. Anirban Ghosh, E. Largy, V. Gabelica, [G4dbr : A Database of Validated DNA G-Quadruplexes Structures in Native Mass Spectrometry Conditions](#) (Zenodo, 2020).
219. R. R. Abzalimov, I. A. Kaltashov, [Extraction of Local Hydrogen Exchange Data from HDX CAD MS Measurements by Deconvolution of Isotopic Distributions of Fragment Ions](#). *J. Am. Soc. Mass Spectrom.* **17**, 1543-1551 (2006).
220. D. D. Weis, J. R. Engen, I. J. Kass, [Semi-Automated Data Processing of Hydrogen Exchange Mass Spectra Using HX-express](#). *J. Am. Soc. Mass Spectrom.* **17**, 1700-1703 (2006).
221. J. K. Chik, J. L. Vande Graaf, D. C. Schriemer, [Quantitating the Statistical Distribution of Deuterium Incorporation to Extend the Utility of H/D Exchange MS Data](#). *Anal. Chem.* **78**, 207-214 (2006).
222. X. Lou, M. Kirchner, B. Y. Renard, U. Köthe, S. Boppel, C. Graf, C.-T. Lee, J. A. J. Steen, H. Steen, M. P. Mayer, F. A. Hamprecht, [Deuteration Distribution Estimation with Improved Sequence Coverage for HX/MS Experiments](#). *Bioinformatics* **26**, 1535-1541 (2010).
223. Z.-Y. Kan, L. Mayne, P. Sevugan Chetty, S. W. Englander, [ExMS : Data Analysis for HX-MS Experiments](#). *J. Am. Soc. Mass Spectrom.* **22**, s13361-11-236-3 (2011).
224. M. Guttman, D. D. Weis, J. R. Engen, K. K. Lee, [Analysis of Overlapped and Noisy Hydrogen/Deuterium Exchange Mass Spectra](#). *J. Am. Soc. Mass Spectrom.* **24**, 1906-1912 (2013).
225. R. Lindner, X. Lou, J. Reinstein, R. L. Shoeman, F. A. Hamprecht, A. Winkler, [Hexicon 2 : Automated Processing of Hydrogen-Deuterium Exchange Mass Spectrometry Data with Improved Deuteration Distribution Estimation](#). *J. Am. Soc. Mass Spectrom.* **25**, 1018-1028 (2014).
226. J. P. Salisbury, Q. Liu, J. N. Agar, [QUDeX-MS : Hydrogen/Deuterium Exchange Calculation for Mass Spectra with Resolved Isotopic Fine Structure](#). *BMC Bioinf.* **15**, 403 (2014).
227. Z. Kan, X. Ye, J. J. Skinner, L. Mayne, S. W. Englander, [ExMS2 : An Integrated Solution for Hydrogen-Deuterium Exchange Mass Spectrometry Data Analysis](#). *Anal. Chem.* **91**, 7474-7481 (2019).
228. S. Na, J.-J. Lee, J. W. J. Joo, K.-J. Lee, E. Paek, [deMix : Decoding Deuterated Distributions from Heterogeneous Protein States via HDX-MS](#). *Sci. Rep.* **9**, 3176 (2019).
229. T. Yamamoto, T. Yamagaki, H. Satake, [Development of Software for the In-depth Analysis of Protein Dynamics as Determined by MALDI Mass Spectrometry-Based Hydrogen/Deuterium Exchange](#). *Mass Spectrom.* **8**, S0082-S0082 (2020).
230. Z. Kan, X. Ye, J. J. Skinner, L. Mayne, S. W. Englander, [ExMS2 : An Integrated Solution for Hydrogen-Deuterium Exchange Mass Spectrometry Data Analysis](#). *Anal. Chem.* **91**, 7474-7481 (2019).
231. D. Bouyssié, J. Lesne, M. Locard-Paulet, R. Albigot, O. Burlet-Schiltz, J. Marcoux, [HDX-viewer : Interactive 3D Visualization of Hydrogen-Deuterium Exchange Data](#). *Bioinformatics* **35**, 5331-5333 (2019).



## Bibliographie

232. V. Hourdel, S. Volant, D. P. O'Brien, A. Chenal, J. Chamot-Rooke, M.-A. Dillies, S. Brier, [MEMHDX : An Interactive Tool to Expedite the Statistical Validation and Visualization of Large HDX-MS Datasets](#). *Bioinformatics* **32**, 3413-3419 (2016).
233. A. M. C. Lau, Z. Ahdash, C. Martens, A. Politis, [Deuteros : Software for Rapid Analysis and Visualization of Data from Differential Hydrogen Deuterium Exchange-Mass Spectrometry](#). *Bioinformatics* **35**, 3171-3173 (2019).
234. R. J. Lumpkin, E. A. Komives, [DECA, a Comprehensive, Automatic Post-Processing Program for HDX-MS Data\\*](#). *Mol. Cell. Proteomics* **18**, 2516-2523 (2019).
235. W. Puchała, M. Burdukiewicz, M. Kistowski, K. A. Dąbrowska, A. E. Badaczewska-Dawid, D. Cysewski, M. Dadlez, [HaDeX : An R Package and Web-Server for Analysis of Data from Hydrogen-Deuterium Exchange Mass Spectrometry Experiments](#). *Bioinformatics* **36**, 4516-4518 (2020).
236. A. M. Lau, J. Claesen, K. Hansen, A. Politis, [Deuteros 2.0 : Peptide-Level Significance Testing of Data from Hydrogen Deuterium Exchange Mass Spectrometry](#). *Bioinformatics* **37**, 270-272 (2021).
237. J. H. Smit, S. Krishnamurthy, B. Y. Srinivasu, R. Parakra, S. Karamanou, A. Economou, [Probing Universal Protein Dynamics Using Hydrogen-Deuterium Exchange Mass Spectrometry-Derived Residue-Level Gibbs Free Energy](#). *Anal. Chem.* **93**, 12840-12847 (2021).
238. N. Zhang, X. Yu, X. Zhang, S. D'Arcy, [HD-eXplosion : Visualization of Hydrogen-Deuterium Exchange Data as Chiclet and Volcano Plots with Statistical Filtering](#). *Bioinformatics* **37**, 1926-1927 (2021).
239. N. Seetaloo, M. Kish, J. J. Phillips, [HDFlex : Software for Flexible High Structural Resolution of Hydrogen/Deuterium-Exchange Mass Spectrometry Data](#). *Anal. Chem.* **94**, 4557-4564 (2022).
240. M. K. Janowska, K. Reiter, P. Magala, M. Guttman, R. E. Klevit, [HDXBoxeR : An R Package for Statistical Analysis and Visualization of Multiple Hydrogen-Deuterium Exchange Mass-Spectrometry Datasets of Different Protein States](#). *Bioinformatics* **40**, btae479 (2024).
241. R. G. Sadygov, [Poisson Model to Generate Isotope Distribution for Biomolecules](#). *J. Proteome Res.* **17**, 751-758 (2018).
242. D. D. Weis, T. E. Wales, J. R. Engen, M. Hotchko, L. F. Ten Eyck, [Identification and Characterization of EX1 Kinetics in H/D Exchange Mass Spectrometry by Peak Width Analysis](#). *J. Am. Soc. Mass Spectrom.* **17**, 1498-1509 (2006).
243. D. Labudová, J. Hon, M. Lexa, [Pqsfinder Web : G-quadruplex Prediction Using Optimized Pqsfinder Algorithm](#). *Bioinformatics* **36**, 2584-2586 (2020).
244. K. Elimelech-Zohar, Y. Orenstein, [An Overview on Nucleic-Acid G-quadruplex Prediction : From Rule-Based Methods to Deep Neural Networks](#). *Briefings Bioinf.* **24**, bbad252 (2023).
245. J.-M. Garant, J.-P. Perreault, M. S. Scott, [G4RNA Screener Web Server : User Focused Interface for RNA G-quadruplex Prediction](#). *Biochimie* **151**, 115-118 (2018).
246. J. Hon, T. Martínek, J. Zendulka, M. Lexa, [Pqsfinder : An Exhaustive and Imperfection-Tolerant Search Tool for Potential Quadruplex-Forming Sequences in R](#). *Bioinformatics* **33**, 3373-3379 (2017).
247. J. L. Huppert, S. Balasubramanian, [Prevalence of Quadruplexes in the Human Genome](#). *Nucleic Acids Res.* **33**, 2908-2916 (2005).
248. O. Kikin, L. D'Antonio, P. S. Bagga, [QGRS Mapper : A Web-Based Server for Predicting G-quadruplexes in Nucleotide Sequences](#). *Nucleic Acids Res.* **34**, W676-W682 (2006).

249. H. M. Wong, O. Stegle, S. Rodgers, J. L. Huppert, [A Toolbox for Predicting G-quadruplex Formation and Stability](#). *J. Nucleic Acids* **2010**, 564946 (2010).
250. S. Neidle, [Beyond the Double Helix : DNA Structural Diversity and the PDB](#). *J. Biol. Chem.* **296**, 100553 (2021).
251. A. V. Tataurov, Y. You, R. Owczarzy, [Predicting Ultraviolet Spectrum of Single Stranded and Double Stranded Deoxyribonucleic Acids](#). *Biophys. Chem.* **133**, 66-70 (2008).
252. D. R. Roe, B. R. Brooks, [A Protocol for Preparing Explicitly Solvated Systems for Stable Molecular Dynamics Simulations](#). *J. Chem. Phys.* **153**, 54123 (2020).
253. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, R. C. Walker, [Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born](#). *J. Chem. Theory Comput.* **8**, 1542-1555 (2012).
254. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, [Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald](#). *J. Chem. Theory Comput.* **9**, 3878-3888 (2013).
255. S. Le Grand, A. W. Götz, R. C. Walker, [SPFP : Speed without Compromise—a Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations](#). *Comput. Phys. Commun.* **184**, 374-380 (2013).
256. D. Sehnal, S. Bittrich, M. Deshpande, R. Svobodová, K. Berka, V. Bazgier, S. Velankar, S. K. Burley, J. Koča, A. S. Rose, [Mol\\* Viewer : Modern Web App for 3D Visualization and Analysis of Large Biomolecular Structures](#). *Nucleic Acids Res.* **49**, W431-W437 (2021).
257. Y. A. Kaiyum, E. Hoi Pui Chao, L. Dhar, A. A. Shoara, M.-D. Nguyen, C. D. Mackereth, P. Dauphin-Ducharme, P. E. Johnson, [Ligand-induced Folding in a Dopamine-binding DNA Aptamer](#). *ChemBioChem* **25**, e202400493 (2024).
258. N. Nakatsuka, K.-A. Yang, J. M. Abendroth, K. M. Cheung, X. Xu, H. Yang, C. Zhao, B. Zhu, Y. S. Rim, Y. Yang, P. S. Weiss, M. N. Stojanović, A. M. Andrews, [Aptamer-Field-Effect Transistors Overcome Debye Length Limitations for Small-Molecule Sensing](#). *Science* **362**, 319-324 (2018).
259. D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzes, T. J. Giese, A. W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. Cheatham, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan, K. M. Merz, [AmberTools](#). *J. Chem. Inf. Model.* **63**, 6183-6191 (2023).
260. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, [Development and Testing of a General Amber Force Field](#). *J. Comput. Chem.* **25**, 1157-1174 (2004).
261. M. Zgarbová, J. Šponer, P. Jurečka, [Z-DNA as a Touchstone for Additive Empirical Force Fields and a Refinement of the Alpha/Gamma DNA Torsions for AMBER](#). *J. Chem. Theory Comput.* **17**, 6292-6301 (2021).
262. O. Love, R. Galindo-Murillo, M. Zgarbová, J. Šponer, P. Jurečka, T. E. Cheatham, [Assessing the Current State of Amber Force Field Modifications for DNA-2023 Edition](#). *J. Chem. Theory Comput.* **19**, 4299-4307 (2023).
263. S. Izadi, R. Anandakrishnan, A. V. Onufriev, [Building Water Models : A Different Approach](#). *J. Phys. Chem. Lett.* **5**, 3863-3871 (2014).

## Bibliographie

264. Z. Li, L. F. Song, P. Li, K. M. Merz, [Systematic Parametrization of Divalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models](#). *J. Chem. Theory Comput.* **16**, 4429-4442 (2020).
265. J. D. Schmit, N. L. Kariyawasam, V. Needham, P. E. Smith, [SLTCAP : A Simple Method for Calculating the Number of Ions Needed for MD Simulation](#). *J. Chem. Theory Comput.* **14**, 1823-1827 (2018).
266. M. R. Machado, S. Pantano, [Split the Charge Difference in Two! A Rule of Thumb for Adding Proper Amounts of Ions in MD Simulations](#). *J. Chem. Theory Comput.* **16**, 1367-1372 (2020).
267. C. Altona, M. Sundaralingam, [Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. New Description Using the Concept of Pseudorotation](#). *J. Am. Chem. Soc.* **94**, 8205-8212 (1972).
268. A. H. J. Wang, G. Ughetto, G. J. Quigley, A. Rich, [Interactions between an Anthracycline Antibiotic and DNA : Molecular Structure of Daunomycin Complexed to d\(CpGpTpApCpG\) at 1.2-Å Resolution](#). *Biochemistry* **26**, 1152-1163 (1987).
269. X. Shen, B. Gu, S. A. Che, F. S. Zhang, [Solvent Effects on the Conformation of DNA Dodecamer Segment : A Simulation Study](#). *J. Chem. Phys.* **135**, 34509 (2011).
270. C. Williams, D. Richardson, J. Richardson, High Quality Protein Residues : Top2018 Mainchain-Filtered Residues (2021). <https://doi.org/10.5281/ZENODO.5777651>.
271. C. J. Williams, D. C. Richardson, J. S. Richardson, [The Importance of Residue-level Filtering and the Top2018 Best-parts Dataset of High-quality Protein Residues](#). *Protein Sci.* **31**, 290-300 (2022).
272. W. Saenger, [Principles of Nucleic Acid Structure](#) (Springer New York, 1984).
273. R. F. Gesteland, T. Cech, J. F. Atkins, *The RNA World : The Nature of Modern RNA Suggests a Prebiotic RNA World* (Cold Spring Harbor Laboratory Press, 2006) *Cold Spring Harbor Monograph Series*.
274. N. B. Leontis, E. Westhof, [Geometric Nomenclature and Classification of RNA Base Pairs](#). *RNA* **7**, 499-512 (2001).
275. S. Lemieux, [RNA Canonical and Non-Canonical Base Pairing Types : A Recognition Method and Complete Repertoire](#). *Nucleic Acids Res.* **30**, 4250-4263 (2002).
276. X.-J. Lu, H. J. Bussemaker, W. K. Olson, [DSSR : An Integrated Software Tool for Dissecting the Spatial Structure of RNA](#). *Nucleic Acids Res.*, gkv716 (2015).
277. X.-J. Lu, [DSSR-enabled Innovative Schematics of 3D Nucleic Acid Structures with PyMOL](#). *Nucleic Acids Res.*, gkaa426 (2020).
278. H. Wickham, [Http2 : Perform HTTP Requests and Process the Responses](#). (2023).
279. R. A. Laskowski, M. B. Swindells, [LigPlot+ : Multiple Ligand-Protein Interaction Diagrams for Drug Discovery](#). *J. Chem. Inf. Model.* **51**, 2778-2786 (2011).
280. R. C. Kretsch, R. Albrecht, E. S. Andersen, H.-A. Chen, W. Chiu, R. Das, J. G. Gezelle, M. D. Hartmann, C. Höbartner, Y. Hu, S. Jadhav, P. E. Johnson, C. P. Jones, D. Koirala, E. L. Kristoffersen, E. Largy, A. Lewicka, C. D. Mackereth, M. Marcia, M. Nigro, M. Ojha, J. A. Piccirilli, P. A. Rice, H. Shin, A.-L. Steckelberg, Z. Su, Y. Srivastava, L. Wang, Y. Wu, J. Xie, N. H. Zwergius, J. Moulton, A. Kryshchuk, Functional Relevance of CASP16 Nucleic Acid Predictions as Evaluated by Structure Providers. *Proteins* (2025).

281. A. Kabbara, C. Buré, A. Guedin, B. Kauffmann, E. Largy, J. Marquevielle, P. Bonnafeous, Z. Mdarhri, Y. Ferrand, V. Gabelica, F. Rosu, M.-L. Andreola, C. Olivier, S. Amrane, Dibenzoacridinium Derivatives as a New Class of G-quadruplex Ligands with Anti-HIV-1 Properties. *NAR Mol. Med.* (submitted).
282. B. Helmich-Paris, B. De Souza, F. Neese, R. Izsák, [An Improved Chain of Spheres for Exchange Algorithm](#). *J. Chem. Phys.* **155**, 104109 (2021).
283. Z. Wang, H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian, T. Hou, [Comprehensive Evaluation of Ten Docking Programs on a Diverse Set of Protein–Ligand Complexes : The Prediction Accuracy of Sampling Power and Scoring Power](#). *Phys. Chem. Chem. Phys.* **18**, 12964-12975 (2016).
284. J. Jerbi, R. F. Fink, J. Peña-García, D. Arias-Olivares, J. Contreras-García, J. P. Cerón-Carrasco, [Quantum Effects Explain the Twist Angle in the Helical Structure of DNA](#). *ChemPhysChem* **25**, e202400372 (2024).
285. S. Grimme, J. Antony, T. Schwabe, C. Mück-Lichtenfeld, [Density Functional Theory with Dispersion Corrections for Supramolecular Structures, Aggregates, and Complexes of \(Bio\)Organic Molecules](#). *Org, Biomol, Chem*, **5**, 741-758 (2007).
286. B. Brauer, M. K. Kesharwani, J. M. L. Martin, [Some Observations on Counterpoise Corrections for Explicitly Correlated Calculations on Noncovalent Interactions](#). *J. Chem. Theory Comput.* **10**, 3791-3799 (2014).
287. P. Mignon, [Influence of the  \$\pi\$ -Interaction on the Hydrogen Bonding Capacity of Stacked DNA/RNA Bases](#). *Nucleic Acids Res.* **33**, 1779-1789 (2005).
288. T. Fornstedt, M. Enmark, [Separation of Therapeutic Oligonucleotides Using Ion-Pair Reversed-Phase Chromatography Based on Fundamental Separation Science](#). *J. Chromatogr. Open* **3**, 100079 (2023).
289. A. Cucchiarini, M. Dobrovolná, V. Brázda, J.-L. Mergny, [Analysis of Quadruplex Propensity of Aptamer Sequences](#). *Nucleic Acids Res.* **53**, gkaf424 (2025).

