



HAL
open science

Expériences in silico sur biomolécules : Une approche multi-modèles pour décrypter les mécanismes physico-chimiques à l'échelle atomique

Marie Brut

► **To cite this version:**

Marie Brut. Expériences in silico sur biomolécules : Une approche multi-modèles pour décrypter les mécanismes physico-chimiques à l'échelle atomique. Micro et nanotechnologies/Microélectronique. UPS Toulouse - Université Toulouse 3 Paul Sabatier, 2021. tel-04735195

HAL Id: tel-04735195

<https://laas.hal.science/tel-04735195v1>

Submitted on 14 Oct 2024

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

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

Mémoire présenté en vue de l'obtention de
l'Habilitation à Diriger des Recherches

Université de Toulouse

Ecole Doctorale Sciences de la Matière
Spécialité Physique

Expériences in silico sur biomolécules :
Une approche multi-modèles pour décrypter
les mécanismes physico-chimiques
à l'échelle atomique

Présentée par : **Marie Brut**
Maître de Conférences à l'Université Toulouse 3

Soutenue publiquement le 11 mai 2021 devant un jury composé de :

Fabienne ALARY	Université Toulouse 3	Présidente
Eric HENON	Université de Reims	Rapporteur
Pierre JOSEPH	LAAS-CNRS Toulouse	Examinateur
Damien LAAGE	UMR PASTEUR, ENS Paris	Rapporteur
Georges LANDA	LAAS-CNRS Toulouse	Garant
Chantal PREVOST	LBT - IPBC Paris	Rapporteure

- Activités de recherche menées au LAAS-CNRS -
Equipe *Modélisation Multi-niveaux des Matériaux*

Préambule

Ce manuscrit synthétise l'essentiel des travaux de recherche que j'ai menés au LAAS-CNRS depuis mon recrutement en tant que Maître de Conférences à l'Université Paul Sabatier en 2011. Ma principale motivation est d'avancer dans la **compréhension des mécanismes physico-chimiques qui gouvernent l'activité des biomolécules grâce à des approches théoriques, principalement à l'échelle atomique.**

Mon travail s'articule autour de deux activités complémentaires : i) d'une part le développement de **nouvelles méthodologies, principalement pour l'étude de la relation structure-fonction de molécules biologiques** ; ii) d'autre part, l'application de ces méthodes à des systèmes biologiques et bio-hybrides. Ces deux activités sont intrinsèquement liées car la spécificité de chaque système d'étude nous conduit à choisir et adapter les outils, voire à en développer *ex nihilo*. Par ailleurs, mettre en œuvre des approches originales nous permet de poser un regard nouveau sur les systèmes étudiés et de proposer des expériences *in silico* sur-mesure. Ces deux activités se nourrissent ainsi mutuellement. Enfin, la diversité des cas d'études nous est apportée par les collaborations interdisciplinaires développées avec physiciens, biophysiciens, chimistes et biologistes, sans qui ces travaux n'existeraient pas. **Le lien à l'expérience est donc à la fois moteur et complémentaire de nos calculs.**

Les molécules biologiques sont caractérisées par la fonction qu'elles assurent au cœur des cellules. Cette fonction est conditionnée par leur capacité à adopter des conformations spécifiques et à interagir avec d'autres partenaires. Cette capacité est elle-même conditionnée par le caractère dynamique de leur organisation structurale.

Décrypter ces processus est nécessaire pour accéder à la compréhension fondamentale des mécanismes cellulaires, mais aussi de leur dysfonctionnement, et pour proposer de nouvelles stratégies thérapeutiques. Les biomolécules sont également exploitées de part leurs propriétés uniques (auto-assemblage, reconnaissance spécifique, catalyse, actionnement) pour créer de nouvelles fonctions ou encore comme matériau. La maîtrise de ces mécanismes nécessite de caractériser les interactions inter et intramoléculaires de façon précise, à l'échelle atomique. Cette tâche est ardue, et cela pour au moins trois raisons. D'une part, les molécules biologiques sont des macromolécules comportant des milliers, voire des dizaines de milliers d'atomes ; les problèmes associés comportent donc un **nombre considérable de variables et de degrés de liberté**. Par ailleurs, les processus biologiques sont des **processus dynamiques** couvrant une très large gamme d'échelles de temps et d'espace, lesquelles doivent être caractérisées dans leur ensemble pour accéder à la compréhension complète des mécanismes. Enfin, les liaisons chimiques qui gouvernent les processus biologiques sont des **liaisons de faible énergie**, difficiles à caractériser, mais dont l'effet cumulatif dicte la structure et l'auto-assemblage des macromolécules tout en autorisant la souplesse et la dynamique conformationnelles nécessaires pour assurer leur capacité de reconnaissance spécifique et leur fonction biologique.

Le premier chapitre de ce manuscrit porte sur la complexité intrinsèque des biomolécules et sur les défis scientifiques et techniques qu'elle pose en matière de modélisation. J'y discute le fondement de la simulation des biomolécules qui repose sur **l'exploration de leur structure afin de comprendre leur fonctionnement.** Nous cherchons à identifier par le calcul les états actifs et inactifs dans leur paysage énergétique, c'est-à-dire les minima, les états métastables et les barrières à franchir pour passer d'un état à un autre. Ce passage se fait via des fluctuations dont la dynamique est essentielle pour la compréhension des mécanismes fonctionnels. Ce paysage peut être **modifié par des perturbations** telles que des mutations, des interactions, etc, permettant de peupler des régions autrement inaccessibles et dont la prise en compte est essentielle pour comprendre le fonctionnement intrinsèque des systèmes considérés. Les simulations jouent donc un rôle crucial pour développer cette vision, complétant les expériences qui offrent une vision souvent statique.

Différentes techniques expérimentales existent, chacune présentant ses avantages et ses limites. La cristallographie aux rayons X, centrale en biologie structurale, fournit une structure à haute résolution, mais statique et dans les conditions d'une molécule cristallisée à basse température. La résonance magnétique nucléaire permet quant à elle d'accéder à des informations sur la dynamique, mais sur des structures de tailles limitées. La cryo-microscopie électronique a connu des progrès très rapides ces dernières années et permet d'accéder à la structure de grands édifices avec une résolution qui se rapproche de l'échelle atomique, mais à des températures non physiologiques. Par ailleurs, l'ensemble de ces techniques est insuffisant pour montrer la dynamique du système. L'analyse des structures expérimentales, comme point de départ, doit être complétée par des simulations qui décryptent le comportement conformationnel. La **dynamique moléculaire**, basée sur la résolution des équations de mouvement de Newton, reste l'outil de prédilection pour échantillonner l'espace des conformations. Je discuterai l'utilisation que nous en faisons pour accéder aux changements conformationnels ou aux grandeurs thermodynamiques des systèmes étudiés, mais aussi pour accéder aux réactions chimiques au sein des enzymes en couplant localement la dynamique classique à un traitement quantique.

Je présenterai également l'approche par **Modes Statiques** que nous développons pour explorer les déformations induites des biomolécules en réponses à des contraintes qui leur sont appliquées. Cette méthode que nous avons initialement développée pour prendre en compte la flexibilité des molécules en interaction (ajustement induit) est aujourd'hui utilisée de multiples façons pour mener des expériences *in silico* de manipulation des molécules. A la manière des expériences sur molécules uniques menées par exemple à l'aide de pinces optiques, nous appliquons des contraintes locales sur un atome ou une collection d'atomes, et mesurons la réponse mécanique du système. Je montrerai ainsi **comment nous envisageons les questions qui nous sont posées sous un angle mécanistique et créons des expériences sur mesure pour y répondre avec une précision atomique.** Cette méthode s'apparente à une approche moléculaire de la mécanobiologie : nous cherchons à mieux appréhender la relation entre une contrainte locale, mécanique ou chimique, et les effets qu'elle induit sur l'activité biologique. Cela peut être par exemple l'effet de l'amarrage d'un ligand sur un récepteur induisant un changement conformationnel activant la fonction, la conversion de l'énergie chimique d'une réaction en énergie mécanique générant un mouvement fonctionnel, ou encore l'effet d'une modification physico-chimique sur la structure et l'activité.

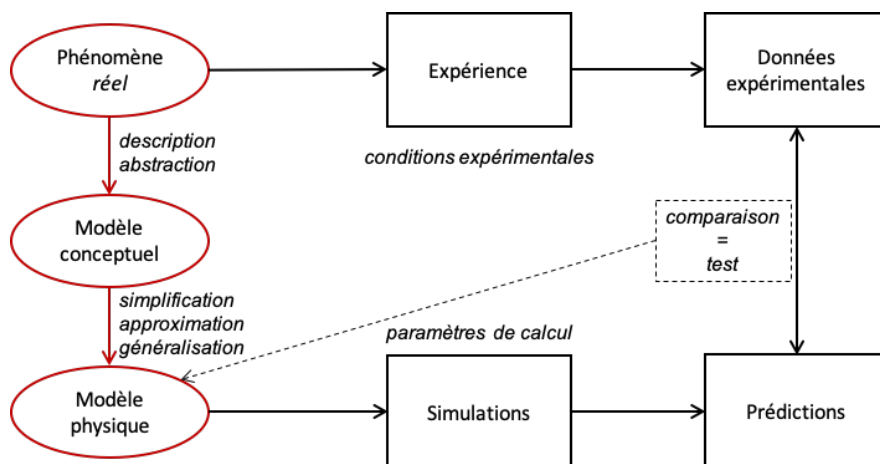
Dans un second chapitre, je présenterai mes activités d'un point de vue applicatif selon deux axes distincts : d'une part l'étude des protéines comme cibles thérapeutiques, et d'autre part l'étude de molécules biologiques pour leur intégration dans des dispositifs technologiques. Je détaillerai certains de nos cas d'étude et y associerai le panel de méthodologies utilisées pour traiter ces problématiques dont la ligne directrice vise des applications santé. **L'étude des protéines comme cibles thérapeutiques** constitue notre principale activité. Notre objectif est d'une part d'améliorer la connaissance fondamentale de protéines d'intérêt, de comprendre pour quelles raisons et par quels mécanismes elles peuvent devenir des cibles thérapeutiques, mais aussi l'origine des résistances qu'elles peuvent présenter, occasionnées par l'apparition de mutations spécifiques. Nous nous intéressons également aux **mutations que nous intégrons comme un outil de caractérisation** des protéines, en évaluant leurs effets sur les propriétés des systèmes traités, en terme de structure et d'activité. Enfin, les mutations caractéristiques d'une pathologie sont d'intérêt pour prendre en compte l'hétérogénéité moléculaire, en oncologie tout particulièrement, et ainsi proposer des stratégies thérapeutiques spécifiques à certaines variations génétiques. C'est le principe des thérapies ciblées.

S'affranchir de ces difficultés est donc nécessaire pour accéder aux mécanismes qui régissent les propriétés structurales et fonctionnelles des biomolécules. Les progrès expérimentaux nous fournissent de plus en plus de structures de biomolécules, seules et en complexes, mais aussi d'informations sur leurs propriétés dynamiques, leurs énergies d'interactions, leur activité biochimique, etc. Cependant, il demeure encore impossible d'accéder à toutes les échelles du vivant, et les couvrir au mieux nécessite la combinaison de nombreuses méthodes dans une approche dite intégrative. Les méthodes théoriques y trouvent aujourd'hui particulièrement leur place, portées par de constants développements en matière de modèles, mais aussi de ressources informatiques qui, conjointement, permettent d'accéder à des niveaux de description inatteignables par les méthodes expérimentales. Il devient ainsi possible de simuler de très grands systèmes sur des échelles de temps avoisinant la milliseconde, mais aussi des réactions chimiques dans des conditions proches de l'environnement *in vivo*.

Sommes-nous encore loin du microscope computationnel ?

A cette question, nous pouvons répondre que la modélisation moléculaire permet aujourd'hui la compréhension des phénomènes physiques et chimiques à l'échelle atomique, mais aussi à plus grande échelle, grâce à des modèles efficaces mais sacrifiant précision et résolution. Ces techniques *in silico* sont déjà largement appliquées pour aider à interpréter et comprendre les résultats expérimentaux, tester de nouvelles théories, mais aussi prédire le comportement physique ou chimique de la matière et faire le lien entre les caractéristiques du système à l'échelle atomique et ses propriétés macroscopiques (structurales, dynamiques, thermodynamiques). Mais jusqu'où sommes-nous capables de les utiliser ? **Ces expériences numériques sont-elles assez matures pour nous permettre d'explorer de nouvelles voies et de guider, voire remplacer l'expérience ?**

La réponse est oui, si nous arrivons à proposer des modèles suffisamment simples pour que les simulations soient techniquement réalisables, mais suffisamment réalistes pour reproduire au mieux les propriétés physico-chimiques des phénomènes étudiés, avec le niveau de précision souhaité. Déjà, la position centrale du calcul est démontrée par son impact croissant dans des domaines aussi variés que la conception de médicaments, de biomatériaux, la biocatalyse, les bionanotechnologies, etc. Les simulations jouent un rôle clé dans notre compréhension fondamentale des bases moléculaires de maladies, dans la façon dont les biomolécules biologiques sont au cœur des processus cellulaires, du fonctionnement de machines biologiques telles que le ribosome, les pompes à ions, les moteurs protéiques, les récepteurs membranaires, etc. Leur intégration sur différentes échelles spatio-temporelles constitue aujourd'hui un vrai défi qui doit être relevé pour prétendre relier mécanismes moléculaires et cellulaires. Pour y parvenir, le lien à l'expérience demeure un prérequis fondamental. Tous les calculs que nous menons se placent dans cette optique et visent à compléter les apports des techniques expérimentales.



La simulation des événements physico-chimiques passe par la formulation et la validation mathématique de modèles qui se doivent d'être réalistes, mais aussi réalisables d'un point de vue calculatoire. La comparaison aux données expérimentales permet de les valider et de les corriger jusqu'à établir leur caractère prédictif (adapté de Gunsteren *et al.* [225].)

Au travers de trois principaux cas traités, j'illustrerai les fondements de **notre approche générale pour l'étude des cibles thérapeutiques, une approche mécanistique pouvant ensuite être enrichie de calculs réactionnels**. Nous partons du principe que les changements conformationnels sont intrinsèques à la nature et à la structure des systèmes, et sont conduits par des stimuli externes (interactions, charges, réactions chimiques, etc). Je montrerai ainsi comment nous relierions biomécanique et activité biologique au travers de deux principaux exemples. Le premier est celui de la dihydrofolate réductase, une cible pour divers traitements sur laquelle nous n'avons travaillé qu'à l'aide des Modes Statiques et avec laquelle nous avons commencé à nous intéresser aux effets des mutations. Le second, celui de l'oncoprotéine Ras, est celui qui sera le plus largement développé et sur lequel nous avons utilisé une approche multi-méthodes particulièrement complète. Nous visons ici à comprendre les mécanismes par lesquels des mutations peuvent induire une diminution de l'activité enzymatique, laquelle peut être restaurée par l'ajout de nouvelles mutations que nous avons prédites et qui peuvent déboucher sur de nouvelles stratégies thérapeutiques. Le dernier cas porte sur la protéine cGAS, un déclencheur de la réponse immunitaire qui pourrait constituer une nouvelle cible anti-cancer. Ce dernier système est abordé uniquement par dynamique moléculaire, classique et QM/MM, afin de comprendre ses mécanismes d'activation par l'ADN, ainsi que les réactions qu'il catalyse. **Nos travaux sur ces systèmes sont décrits de manière indépendante et par ordre chronologique, ordre qui montre l'enrichissement continu des méthodes que nous utilisons.**

Je passerai ensuite à notre second axe de travail, portant sur l'étude des biomolécules comme objets des nanotechnologies. Je présenterai différents projets pour la plupart directement suscités par nos collaborations internes au LAAS, en particulier autour de l'étude de brins d'ADN fonctionnels pour leur intégration dans des dispositifs de détection. **Les questions adressées par l'intégration de biomolécules comme agents structurants ou fonctionnels sont nombreuses et nouvelles.** Je montrerai comment nous utilisons les Modes Statiques pour évaluer les propriétés de ces objets. Celles-ci doivent en effet être contrôlées et maintenues dans le cadre de l'intégration, ce qui implique notre capacité à répondre à des questions fondamentales, dont certaines sont adressées dans nos travaux, à savoir : comment fonctionnaliser les biomolécules sans les dénaturer ? comment diriger leur sélectivité ? comment diminuer leur complexité tout en conservant leur spécificité et leur fonction ? peut-on caractériser et optimiser leur capacité d'interaction avec leur environnement ? Ces nombreux niveaux de compréhension ne peuvent être adressés que par une batterie d'outils de modélisation. En plus des Modes Statiques, certains sont déjà à notre disposition, comme la dynamique moléculaire pour étudier la stabilité des structures en fonction de paramètres expérimentaux (température, concentration saline), ou encore pour évaluer l'énergie d'interaction d'un capteur avec sa cible. De nouveaux outils restent cependant à développer, et tous doivent se compléter pour **guider les choix clés autour des constructions moléculaires, interpréter et prédire les performances de dispositifs qui rassemblent les mondes biologiques et non-biologiques.** Je décrirai deux outils que nous avons développés pour aider à avancer dans cette problématique, notamment un algorithme de prédiction de séquence pour proposer des acides nucléiques aux propriétés satisfaisant un cahier des charges établi par les besoins de l'expérience, ou encore une adaptation des Modes Statiques pour la prédiction de mouvements de grande amplitude. Ce deuxième outil a initialement pour vocation de comprendre la mécanique de moteurs moléculaires, mais je souhaite l'adapter à d'autres systèmes de grande taille, ce que j'aborderai dans mes perspectives de travail.

Le dernier chapitre sera consacré à ces perspectives, lesquelles auront déjà été abordées dans le manuscrit au fur et à mesure de la description des projets passés et en cours. Elles se veulent plus méthodologiques qu'applicatives, et surtout généralisables, dans l'idée de rendre l'approche par Modes Statiques compatible avec tout type de niveau de représentation et de modèle énergétique. **Nous souhaitons ainsi être capables de balayer toutes les échelles afin de nous adapter et de pouvoir choisir le niveau de résolution à atteindre selon la question posée. Cet effort ouvrira le champ des applications possibles et constitue un passage nécessaire pour maintenir un dialogue fructueux avec l'expérience.**

Table des matières

Préambule	i
1 Introduction	1
1 Parcours - Vers la physique pour la biologie	1
2 Motivation - Comprendre la relation structure-fonction pour mieux la maîtriser .	2
2.1 De la séquence à la structure	2
2.2 De la structure à la fonction	3
2.3 De la séquence à la fonction	5
2.4 De l'importance des liaisons faibles en biologie	5
2.5 De l'importance de l'eau en biologie	6
3 Vers le "microscope computationnel" : où en sommes-nous ?	7
3.1 Challenges et outils de la biologie structurale	7
3.2 Bref état de l'art méthodologique	9
3.3 Bilan : quelles questions veut-on et peut-on adresser ?	16
4 Les Modes Statiques : un outil versatile pour la manipulation des biomolécules .	17
4.1 Concept et méthode	17
4.2 Algorithme	18
4.3 Un outil validé et évolutif	18
4.4 Pour une approche d'ingénierie in silico	20
2 Activités de recherche	21
1 Objectifs scientifiques	21
1.1 Les biomolécules comme cibles thérapeutiques.	21
1.2 Les biomolécules pour l'intégration de fonctions actives.	23
2 De la compréhension des mécanismes fondamentaux aux cibles thérapeutiques .	24
2.1 Exploration de la relation structure - activité par une approche mécanistique	25
2.2 Modulation de l'activité enzymatique par le criblage de mutations : apport du couplage des approches mécanistiques et physico-chimiques	43
2.3 Modulation de l'activité enzymatique par réarrangement structural induit : apport de la dynamique moléculaire classique et hybride	53
3 Les biomolécules comme objets des nanotechnologies	59
3.1 Réponse mécanique, stabilité et reconnaissance des acides nucléiques : des outils pour comprendre et optimiser	60
3.2 Réponse mécanique et mouvements de grande amplitude des assemblages moléculaires . .	73
Conclusions et Perspectives	81
Curriculum Vitæ	87
1 Informations personnelles	87
2 Domaine de recherche	87
2.1 Situation professionnelle	87
2.2 Formation	88
3 Encadrements	88
3.1 Thèses de doctorat	88
3.2 Participation à des travaux de thèses avec production	90
3.3 Post-doctorat	90
3.4 Stages niveaux Licence et Master	90

4	Diffusion des travaux	90
4.1	Revue Internationale à Comité de Lecture	90
4.2	Conférences Internationales à Comité de Lecture	92
4.3	Valorisation de logiciel	92
4.4	Communications scientifiques	92
5	Activités pédagogiques.	95
5.1	Enseignement	95
5.2	Informations significatives sur les conditions d'exercice	95
6	Responsabilités scientifiques et administratives	95
6.1	Vie de l'Université.	95
6.2	Vie du Laboratoire	95
6.3	Comités d'expertise	95
6.4	Activité éditoriale	95
6.5	Participation à des jurys de thèses.	95
6.6	Organisation de congrès	96
6.7	Projets de recherche sur financements	96
	Bibliographie	97

1 Parcours - Vers la physique pour la biologie

J'ai suivi un parcours en physique fondamentale à l'université Paul Sabatier jusqu'à la Maîtrise. J'ai ensuite suivi les enseignements du Master "3N" (Nanophysique, Nanocomposants, Nanomesures) par intérêt pour les bionanotechnologies. Au fil des années et des enseignements, j'ai découvert l'infiniment petit, avec l'envie de sonder le cœur de la matière qui nous constitue et nous entoure. J'étais particulièrement fascinée par **les molécules du vivant**, immenses édifices structurés, capables de se mouvoir, de se reconnaître et d'interagir. Une gigantesque machinerie à comprendre, **mais par quel moyen ?** Je n'étais pas familière des simulations atomiques et n'affectionnais pas les ordinateurs.

"Les ordinateurs, plus on s'en sert moins, moins ça a de chance de mal marcher." [189]

C'est durant mon stage de Master en 2005 que j'ai compris qu'**un ordinateur pouvait être un outil pour répondre à ces questions**. Je suis arrivée au LAAS attirée par ses activités en nanotechnologies et y ai fait la rencontre du calcul et des biomolécules. Ce stage donnait alors du sens à ma trajectoire universitaire, m'initiant au **développement de modèles physiques dédiés aux molécules biologiques**. C'est sur cette problématique que j'ai soutenu ma thèse en 2009, thèse durant laquelle j'ai commencé à développer la méthode des Modes Statiques pour prédire les déformations induites des biomolécules. Je me suis ainsi familiarisée avec ces systèmes complexes, les développements algorithmiques, mais aussi les techniques et outils usuels de simulations atomistiques, tels que la dynamique moléculaire ou les calculs *ab initio*.

J'ai alors souhaité poursuivre dans cette voie, mais davantage en immersion en biologie, tout en gardant mon identité de physicienne. En 2009, j'eus l'immense chance d'être intégrée à l'équipe du Pr Michael Levitt à Stanford, dans le laboratoire *Computational Structural Biology* du centre Bio-X. Ce centre pluridisciplinaire a pour vocation de répondre à des questions biologiques grâce à l'interaction d'équipes de recherches en médecine, biologie, ingénierie, physique et sciences computationnelles. Ce séjour post-doctoral m'a permis de découvrir de nouveaux niveaux et de nouvelles techniques de modélisation. Il m'a également permis de croiser ces disciplines au quotidien, de travailler au sein d'un groupe de calcul, tout en interagissant étroitement avec les biologistes qui nourrissaient nos travaux de leurs questions. Ce dialogue interdisciplinaire demeure aujourd'hui la pierre angulaire de chacun de mes projets, dès lors que l'on s'affranchit de la première difficulté, qui constitue aussi la première richesse de ces collaborations, à savoir **traduire une question biologique en modèle physico-chimique**.

Cette pluridisciplinarité s'est d'autant plus renforcée depuis mon recrutement en 2011 au LAAS-CNRS. L'ingénierie pour le Vivant et l'Environnement y constitue un axe stratégique et la simulation des biomolécules y renforce deux champs d'activités : i) les recherches *"pour la biologie"*, soutenues par des développements devant permettre de **décrypter les mécanismes du vivant** pour des applications santé ; ii) mais aussi les recherches *"par la biologie"*. Ce second volet s'appuie sur les propriétés des biomolécules, afin de les modifier, les manipuler, les contrôler, **les coupler aux matériaux traditionnels des technologies** et donc de les détourner pour des applications principalement orientées vers la santé et l'environnement. Il faut alors **intégrer un nouveau niveau de complexité** au-delà de celui qui est inhérent à ces systèmes déjà complexes, et prendre en compte des interactions avec un environnement non biologique.

Dans la suite de ce manuscrit, les stratégies de calcul et scientifiques mises en œuvre pour l'étude des biomolécules sont exposées selon ces deux voies, tant pour la compréhension que pour l'optimisation, voire la prédiction de nouvelles propriétés.

2 Motivation - Comprendre la relation structure-fonction pour mieux la maîtriser

“Mieux vaut regarder là où on ne va pas, parce que, là où on va, on saura ce qu’il y a quand on y sera ; et de toute façon, ce ne sera jamais que de l’eau.” [189]

Comprendre la relation structure-fonction des biomolécules signifie **accéder aux bases structurales et biochimiques qui leur confèrent leur activité biologique**. Des approches pluridisciplinaires aux interfaces physique, chimie et informatique sont alors requises pour permettre de décrypter les mécanismes fondamentaux qui régissent la vie cellulaire. Ceci nécessite de caractériser les propriétés physico-chimiques et structurales des biomolécules ainsi que leurs réseaux d’interactions. Les champs d’applications sont vastes et visent en particulier à identifier de nouvelles cibles et stratégies thérapeutiques, mais aussi à aller vers l’ingénierie des biomolécules et vers leur intégration dans des systèmes complexes, biologiques ou non.

La relation structure-fonction des biomolécules, protéines et acides nucléiques en particulier, peut être adressée en se demandant comment la nature a modelé une structure pour remplir une fonction. Ce problème fondamental sous-entend les notions de conservation, de variation, et donc d’évolution, basées sur des informations séquentielles et structurales [48].

Comprendre ce paradigme demeure un des enjeux fondamentaux de la biologie structurale et nécessite plusieurs niveaux d’études que je décris de façon non exhaustive afin de positionner mon travail : i) les relations séquence-structure et ii) structure-fonction sont des préalables à la prédiction directe de iii) la relation séquence-fonction ; iv) elles nécessitent un complexe réseau d’interactions intra et intermoléculaires et v) dépendent des conditions du milieu.

2.1 De la séquence à la structure

La fonction biologique des biomolécules repose sur leur capacité à interagir avec des partenaires de façon spécifique. Cette capacité implique une affinité forte et donc une relative adaptabilité conformationnelle. Les longues chaînes de polymères qui les constituent se replient au travers d’interactions intra et inter-moléculaires dans une **configuration quasi-unique dont la stabilité et la dynamique résultent d’un compromis entre rigidité et flexibilité structurale**. Une caractérisation fine des interactions est donc nécessaire pour comprendre les mécanismes de repliement et accéder à la prédiction de structures à partir de la séquence. Il s’agit là d’un premier grand défi [89,192].

Lorsqu’en 1953, R. Franklin et al. découvrent la structure régulière de l’ADN en double hélice, on s’attend implicitement à ce que la structure des protéines admette également des symétries. La première résolution par diffraction de rayons X d’une protéine, obtenue en 1958 par J. Kendrew et al., met à mal cette hypothèse en révélant une structure compacte et apparemment désordonnée. Les protéines ne sont (généralement) pas symétriques, mais pourtant hautement ordonnées (je n’aborde pas ici le cas des protéines intrinsèquement désordonnées [28]). En 1960, C.B. Anfinsen obtient le prix Nobel pour ses travaux sur "la relation entre la séquence des acides aminés et la conformation biologiquement active" [15]. En dénaturant une protéine et en observant son repliement spontané dans la même configuration tridimensionnelle unique, il propose l’"hypothèse thermodynamique", postulant que la structure 3D d’une protéine dans un environnement physiologique est celle pour laquelle l’énergie libre de Gibbs est minimale, c’est-à-dire celle pour laquelle le maximum d’interactions est formé.

La structure peut donc être directement prédite à partir de la séquence. Cependant, le nombre de configurations à explorer est quasiment infini, alors qu’une protéine peut se replier spontanément en quelques microsecondes. Il s’agit du fameux paradoxe de Levinthal auquel différents modèles théoriques tentent de répondre : dans les conditions *in vivo*, tout comme dans les simulations atomistiques, il est **impossible d’explorer l’espace conformationnel de façon exhaustive dans une démarche aléatoire** [67]. Le repliement doit être guidé par des considérations cinétiques, thermodynamiques, ou encore des motifs, ou "points d’ancrage" conservés par l’évolution et guidant le repliement (Fig.1.1).

Quelle que soit l'approche adoptée, nous savons aujourd'hui que l'espace des phases du repliement d'une protéine comporte de nombreux minima locaux souvent proches en énergie. Au final, le gain d'énergie d'une protéine dans sa forme native n'est que de ~ 10 kcal/mol par rapport à sa forme dénaturée. **Suivre le processus de repliement** défie la résolution temporelle des méthodes expérimentales [75]. Les simulations de dynamique moléculaire constituent aujourd'hui une alternative largement utilisée, permettant de prendre en compte le rôle de l'environnement (solvant, pH, température) et de balayer les échelles de temps. Le coût des calculs demeure cependant une limitation non négligeable, avec comme conséquence, la limitation en taille des systèmes accessibles [77, 82, 151, 168, 192, 201].

Une autre voie peut-être adoptée si l'on veut obtenir la structure native sans les chemins qui y mènent. La modélisation par **homologie** se base sur l'existence de classifications structurales de protéines (ou ARN) qui partagent le même repliement malgré des homologies de séquences parfois faibles, mais révélant une conservation sélective des résidus fonctionnellement importants. Afin de résoudre ce problème, c'est-à-dire d'identifier et d'utiliser les motifs de séquence pertinents et de les associer à des structures récurrentes, il existe une grande variété d'algorithmes, banques de données, serveurs web, etc, [170], souvent discutés dans la littérature [102].

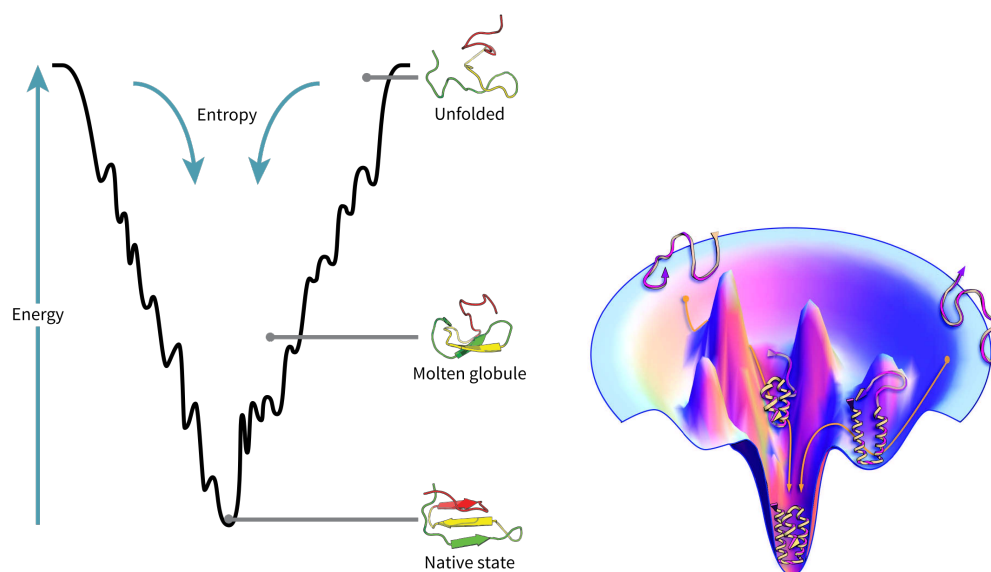


FIGURE 1.1 – Modèle du repliement en entonnoir [243]. Le paysage énergétique a la forme d'un entonnoir présentant des minima locaux capables de piéger des intermédiaires dont le nombre diminue du haut vers le bas, ainsi que l'énergie associée. L'entropie représentée par la largeur de l'entonnoir décroît elle aussi alors que le nombre de contacts natifs augmente. (Sources : wikipedia à gauche et Dill et al. [71] à droite).

2.2 De la structure à la fonction

Les protéines sont de véritables **machines moléculaires assurant une multitude de fonctions** : elles catalysent des réactions chimiques, ont un rôle structural, participent à la transmission des signaux cellulaires, au stockage, au transport, à la réponse immunitaire, à l'expression et régulation de l'information génétique, etc. C'est aussi le cas des acides nucléiques, des ARN en particulier, qui au-delà de leur rôle dans l'expression de l'information génétique, peuvent catalyser des réactions chimiques [76, 155, 201, 241].

Un prérequis inhérent à la fonction des biomolécules repose sur leur capacité à inter-agir. La vision selon laquelle la séquence détermine une structure (quasi)unique a permis de conforter un temps le fameux modèle "clé-serrure" [22] dans lequel une protéine se trouve dans une conformation préexistante parfaitement complémentaire à celle de sa cible. Si ce modèle rigide peut parfois fonctionner, la réalité est plus complexe : les biomolécules sont des objets dynamiques en perpétuel mouvement dans un paysage énergétique multidimensionnel extrêmement complexe. **Une structure est un ensemble de conformations** existant avec une certaine probabilité (thermodynamique) et séparées par des barrières d'énergie (cinétique).

La **dynamique constitue donc elle aussi un prérequis à la fonction** [144, 152] et demeure centrale du point de vue expérimental comme de celui du calcul, pour lequel la dynamique moléculaire reste l'outil privilégié [4, 73, 114, 206]. La figure 1.2 présente les échelles de temps correspondant aux mouvements par flexibilité croissante ainsi que les méthodes de résolution associées ; deux populations d'états sont aussi présentées ainsi que la modification de leur équilibre en réponse à une perturbation du système (amarrage d'un ligand, mutation, etc). Notons qu'il s'agit là du **principal mécanisme de régulation de la fonction**, on parle alors de régulation allostérique [242].

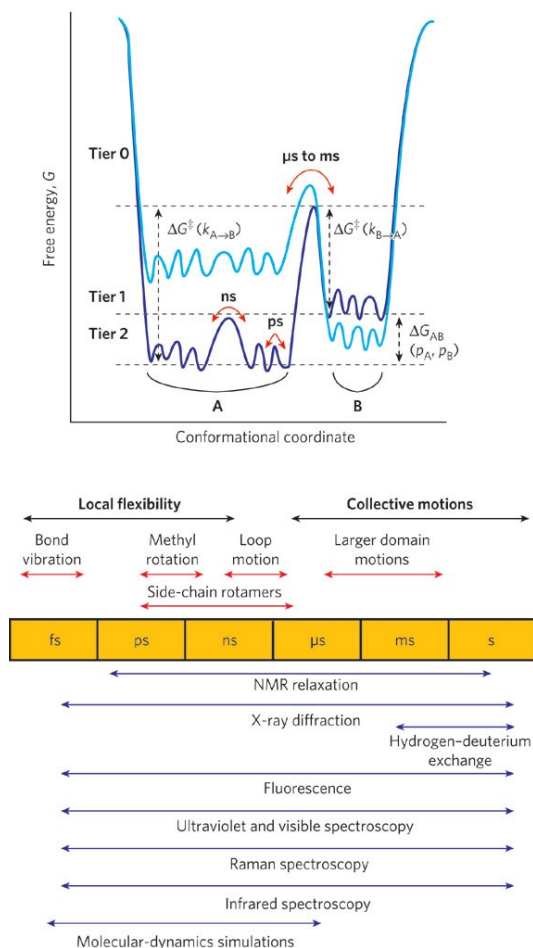


FIGURE 1.2 – En haut, paysage énergétique d'une protéine présentant deux populations A et B, ainsi que leur réponse à une perturbation (en bleu clair et bleu foncé). En bas, échelle de temps des mouvements dynamiques par amplitude croissante et méthodes associées à leur résolution [90].

Lorsqu'elles interagissent avec un partenaire, les biomolécules peuvent en effet admettre un réarrangement "induit". Ce réarrangement peut être local (les valeurs des RMSD pour les structures multiples d'une même protéine sont inférieures à 1,2Å dans la plupart des cas [124]), ou de grande amplitude (mouvements collectifs d'atomes), constituant un défi coûteux en terme de ressources de calcul [168]. Dans les deux cas, la flexibilité du système va lui permettre d'assurer sa fonction en lui permettant de répondre aux modifications de l'équilibre entre les forces qui maintiennent son repliement et les nouvelles forces résultant de l'interaction. **La flexibilité est donc intrinsèque à la biomolécule et lui permet d'assurer sa fonction** en i) s'adaptant pour autoriser des interactions spécifiques ; ii) préconfigurant son site actif pour la catalyse de réactions [184]. Différents modèles sont proposés selon la flexibilité, c'est-à-dire le niveau de réarrangement permis de la protéine (Fig.1.3).

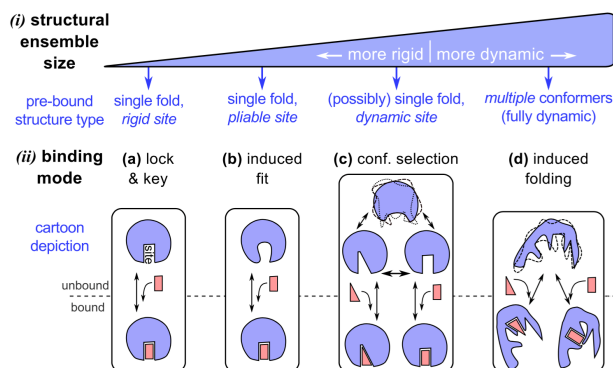


FIGURE 1.3 – De gauche à droite sont représentés des modèles d'interactions selon une flexibilité croissante : a) le modèle "clé-serrure" rigide; b) le modèle d'ajustement induit qui autorise un réarrangement du site actif; c) la sélection d'une conformation pré-existante lors de l'interaction; d) le repliement induit par l'interaction (source [144]).

Enfin, **cette mobilité est couplée aux réactions chimiques dont les molécules sont le siège**. Ce couplage entre interactions, mouvement et chimie leur confère les extraordinaires propriétés qui sous-tendent les processus biologiques au cœur des cellules, faisant d'elles des catalyseurs, des récepteurs de signaux, des commutateurs, des moteurs, ou encore des pompes à ions, cas sur lesquels je reviendrai plus longuement dans ce manuscrit.

Une compréhension fine des mécanismes qui régissent ces réactions est donc également cruciale pour accéder à la fonction des biomolécules, mais aussi à leur régulation [16].

2.3 De la séquence à la fonction

Comme nous l'avons vu en section 2.1, certaines séquences ou motifs ont été conservés au fil du temps pour leur rôle crucial dans la structuration des biomolécules, mais aussi pour leur fonction. Identifier ces motifs et comprendre leur relation à la fonction doit donc permettre de prédire cette fonction sans connaissance intermédiaire de la structure ou de la dynamique, mais aussi de synthétiser de nouvelles molécules. C'est la base du design *de novo* de protéines [64, 95], mais aussi des technologies ADN sur lesquelles je reviendrai [197]. De très nombreux algorithmes sont développés dans ce domaine relevant de la bioinformatique, la principale difficulté reposant sur la manipulation et la combinatoire d'immenses banques de données [210].

L'étude des effets des mutations peut également être un outil pour relier séquence et fonction [78]. Une mutation ponctuelle peut en effet affecter le repliement, la dynamique, mais aussi la structure, la stabilité et donc la réactivité d'un site actif. Une modification de séquence peut ainsi drastiquement modifier la fonction d'une biomolécule, être associée à une pathologie, et donc être d'intérêt thérapeutique [63]. C'est sur cette problématique que portent la plupart des activités pour la santé que je décrirai par la suite.

2.4 De l'importance des liaisons faibles en biologie

Les biomolécules sont capables d'auto-organisation et présentent une architecture dynamique répondant aux perturbations extérieures. **Repliement, dynamique, interactions, réactions chimiques, ces mécanismes intriqués reposent sur de complexes réseaux d'interactions** établis au sein des biomolécules, mais aussi entre les biomolécules et leur environnement (ligands, solvant, ions). Ces interactions réversibles sont des liaisons chimiques **non covalentes et de faible énergie**, de type électrostatiques, van der Waals et hydrogène [27] (voir figure 1.4). Elles varient en terme de spécificité et d'intensité, bien que celle-ci soit dans tous les cas bien inférieure à celle d'une liaison covalente. Bien qu'individuellement faibles, elles ont un rôle crucial dans les processus biologiques de part leur nombre [58, 186].

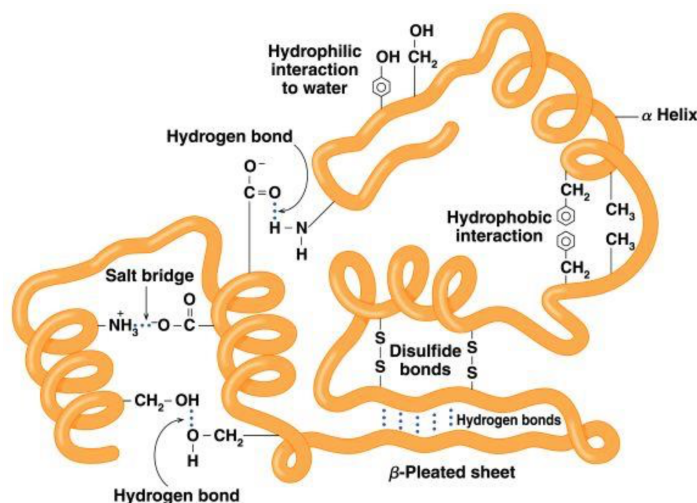


FIGURE 1.4 – Interactions faibles impliquées dans le repliement d'une protéine [219]. On retrouve les mêmes forces dans le cas des acides nucléiques (à l'exception du pont disulfure).

Les mêmes forces interviennent dans les chaînes d'acides aminés et d'acides nucléiques. Les différences de structures et de propriétés proviennent de la polarité des chaînes, des contraintes stériques internes, de la composition interne des molécules. Dans le cas des protéines en effet, les repliements sont particulièrement complexes en raison de la diversité des chaînes latérales des acides aminés (chaînes chargées, neutres, polaires, apolaires, aromatiques). Les acides nucléiques peuvent présenter une structure plus régulière. Leurs bases s'hybrident par liaisons hydrogène et s'empilent sous l'effet des forces de van der Waals, cette attraction s'opposant à la répulsion électrostatique entre groupements phosphate.

- *L'interaction électrostatique* intervient entre deux atomes ou groupements chargés qui se repoussent ou s'attirent si leurs charges sont de même signe ou de signe opposé, respectivement. L'énergie E de l'interaction est décrite par la loi de Coulomb $E \sim kq_1q_2/r$, où q_1 et q_2 sont les charges en jeu, r est la distance entre atomes et k une constante dépendant du milieu.

- *L'interaction de van der Waals* repose, selon la nature des groupements chimiques en jeu, sur plusieurs contributions que l'on peut attribuer à la variation de la distribution des charges électroniques autour des atomes en fonction du temps. Elle intervient entre deux dipôles électriques, permanents ou induits, et portés par deux molécules qui permettent de distinguer les forces de Keesom (entre molécules polaires), de Debye (entre molécules polaires et apolaires) et de London (entre molécules apolaires). L'énergie associée variant en $1/r^6$, ces forces sont donc attractives à courte portée. Cette attraction augmente au fur et à mesure qu'atomes et molécules se rapprochent jusqu'à être séparés par une distance correspondant au rayon de van der Waals, distance en dessous de laquelle les forces deviennent fortement répulsives en raison du recouvrement des nuages électroniques.

- *La liaison hydrogène* est en grande partie d'origine électrostatique, impliquant les charges partielles d'atomes voisins qui s'attirent. Elle implique un groupe "donneur d'hydrogène" (accepteur d'électron : groupements OH ou NH₂ par exemple) et un atome "accepteur d'hydrogène" (donneur d'électron : O, N, F). L'atome auquel l'hydrogène est lié par covalence capte plus fortement les électrons que l'atome d'hydrogène qui présente alors une charge partielle positive. Celui-ci peut interagir avec un atome ayant à l'inverse une charge partielle négative par interaction électrostatique. Ces interactions peuvent être inter ou intra-moléculaires et sont particulièrement importantes dans les processus de repliement et de reconnaissance, tout comme pour les propriétés et effets du solvant en milieu biologique.

2.5 De l'importance de l'eau en biologie

Les interactions faibles sont fortement affectées par la présence de l'eau et des ions en milieu physiologique. Les molécules d'eau, polaires, interagissent au travers de liaisons hydrogène qui assurent leur cohésion. En milieu liquide, ces interactions se font et se défont en permanence, les molécules d'eau sont libres d'interagir entre elles ou avec d'autres molécules polaires. Elles entrent en compétition avec les interactions pouvant s'établir au sein des biomolécules et agissent comme un écran affaiblissant les liaisons électrostatiques et hydrogène entre groupements polaires. Les molécules non polaires ont alors tendance à se regrouper de manière à présenter la plus faible surface de contact avec l'eau, créant les *interactions hydrophobes*, lesquelles résultent du positionnement de groupements hydrophobes qui s'attirent alors mutuellement sous l'effet des forces de London. **L'effet hydrophobe est un des premiers facteurs structurants** puisqu'il modèle les biomolécules en groupant les groupements hydrophobes en leur cœur et les résidus hydrophiles sur leur surface. Il conduit le repliement des chaînes d'acides aminés et d'acides nucléiques, mais aussi l'auto-assemblage des membranes lipidiques [18].

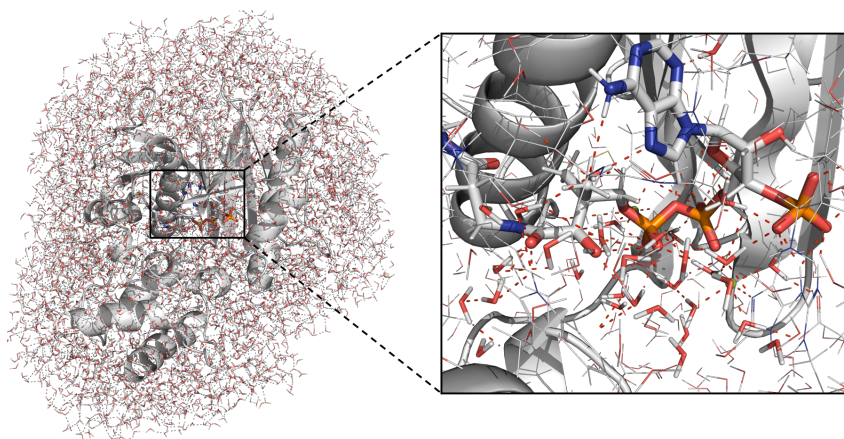


FIGURE 1.5 – A gauche, un complexe de protéines entouré d'une couche de solvant de 10 Å d'épaisseur. A droite, le réseau de molécules d'eau structuré à l'interface protéine-ligand dans le site actif.

Par ailleurs, la couche d'hydratation présente en surface des biomolécules, des protéines en particulier, est indispensable à leur activité et à leur dynamique, c'est-à-dire à leur fonction biologique [19]. Cette couche, parfois appelée "eau biologique" dans la littérature, bien que ce terme prête à confusion, présente des propriétés sensiblement différentes de celles de l'eau "libre". Elle est plus dense et plus dynamique, pouvant agir comme un "lubrifiant" des mouvements biologiques et des processus d'interactions. D'une part, elle agirait comme un **réservoir d'entropie**, des "coups" thermiques enrichissant le paysage énergétique et donnant accès à de nouvelles conformations de faible énergie [7]. D'autre part, l'eau pourrait aussi intervenir dans la **spécificité de la reconnaissance des interactions** (voir Fig.1.5). Elle empêcherait des repliements internes et maintiendrait accessibles des sites de liaisons, agissant comme une extension de la surface de la biomolécule, ainsi facilement adaptable. Elle participerait enfin au gain enthalpique associé à la formation d'interactions non-covalentes et à la perte d'entropie associée aux effets de désolvatation et à la réorganisation du solvant dans les sites de liaison. A l'heure actuelle, certains aspects de ces propriétés dynamiques sont encore méconnus et sujets à controverse, et donc largement discutés dans la littérature [24, 93, 109, 128, 248].

L'eau a enfin un **rôle fondamental à jouer dans les réactions chimiques**, que ce soit au niveau de l'activité enzymatique, dans le transport d'électrons, la coordination des ions, etc. Elle est donc au tout premier plan de l'ensemble des fonctions biologiques.

3 Vers le "microscope computationnel" : où en sommes-nous ?

3.1 Challenges et outils de la biologie structurale

Comprendre la façon dont les biomolécules se structurent et interagissent pour accomplir leur fonction nécessite de **combiner techniques expérimentales et modèles physico-chimiques** [33, 86]. On assiste ainsi à l'essor de ce qui constitue un domaine à part entière de la biologie : la biologie structurale intégrative vise à intégrer des données à différentes échelles de résolutions spatiales et temporelles, afin de décrypter les mécanismes fondamentaux des processus biologiques [121] (Fig. 1.6).

Du point de vue expérimental, le nombre de structures disponibles témoigne des progrès réalisés. La principale technique de résolution reste à ce jour la cristallographie aux rayons X [108, 214]. Sur presque 170000 structures disponibles en 2020 dans la PDB¹, 90% sont résolues par cette technique qui donne, avec une résolution atomique, l'image statique d'une molécule ou d'un complexe dans un état cristallin dont l'obtention est un défi en soi, parfois limitant. Par ailleurs, certaines molécules ou portions de molécules trop flexibles ne peuvent être résolues par cette technique. Parmi les alternatives, on peut citer la spectroscopie RMN (résonance magnétique nucléaire) [110] qui peut donner accès à plusieurs conformations d'une même structure en solution. Elle est cependant limitée par la taille des systèmes (< 20kD) et par une résolution spatiale généralement moindre. Depuis quelques années, la cryo-microscopie électronique vient rivaliser avec ces techniques [200], donnant accès à la structure d'assemblages macromoléculaires de très grande taille en solution, avec une résolution jusque-là plus faible, mais en constant progrès et ayant récemment atteint le record de 1,25Å [50]. Bien d'autres méthodes existent encore, chacune présentant ses limites techniques et de résolution [208, 209].

Au final, les structures disponibles résultent de la combinaison de données issues de différentes méthodes expérimentales, surtout pour les grands assemblages, mais aussi de modèles théoriques et/ou empiriques pour atteindre une résolution atomique. Les progrès passent donc par l'amélioration des technologies mais aussi des modèles physiques et du traitement en masse des données, lequel implique la puissance des ressources informatiques et l'efficacité des logiciels [11, 125, 135].

1. *Protein Data Bank*, rcsb.org

Cependant, malgré les progrès en terme de résolution et de variété d'échantillons accessibles, ces techniques ne permettent pas, pour l'instant, d'accéder à la dynamique des systèmes, ou seulement par extrapolation, ni au suivi des réactions chimiques qui gouvernent leur activité. Les informations fournies concernent les positions statiques d'atomes alors que ceux-ci bougent en permanence avec des degrés de liberté déterminés par les liaisons dans lesquelles ils sont engagés.

D'intenses efforts sont menés pour développer des outils permettant de caractériser les interactions faibles, d'extraire des informations sur le repliement et la dynamique des biomolécules, etc (comme les lasers à électrons libres et à rayons X [85, 203]), ou de déterminer la structure des biomolécules directement dans la cellule (comme la RMN en cellule [162, 202]).

Le suivi des mouvements biologiques, couvrant des amplitudes spatiales et temporelles très étendues, la résolution précise des positions des atomes impliqués dans les mécanismes des réactions chimiques, la distribution de l'eau en conditions physiologiques, les réseaux d'interactions inter et intra-moléculaires, l'identification des sites de liaison, l'effet de l'environnement, des mutations, etc, nécessitent **une résolution que les techniques expérimentales ne peuvent pas, ou que partiellement atteindre**. Aucune ne peut donner accès au chemin suivi lors du repliement ou des réactions chimiques par exemple. **La modélisation est donc plus que jamais nécessaire pour combiner, interpréter les données et remplir les espaces laissés vides par l'expérience.**

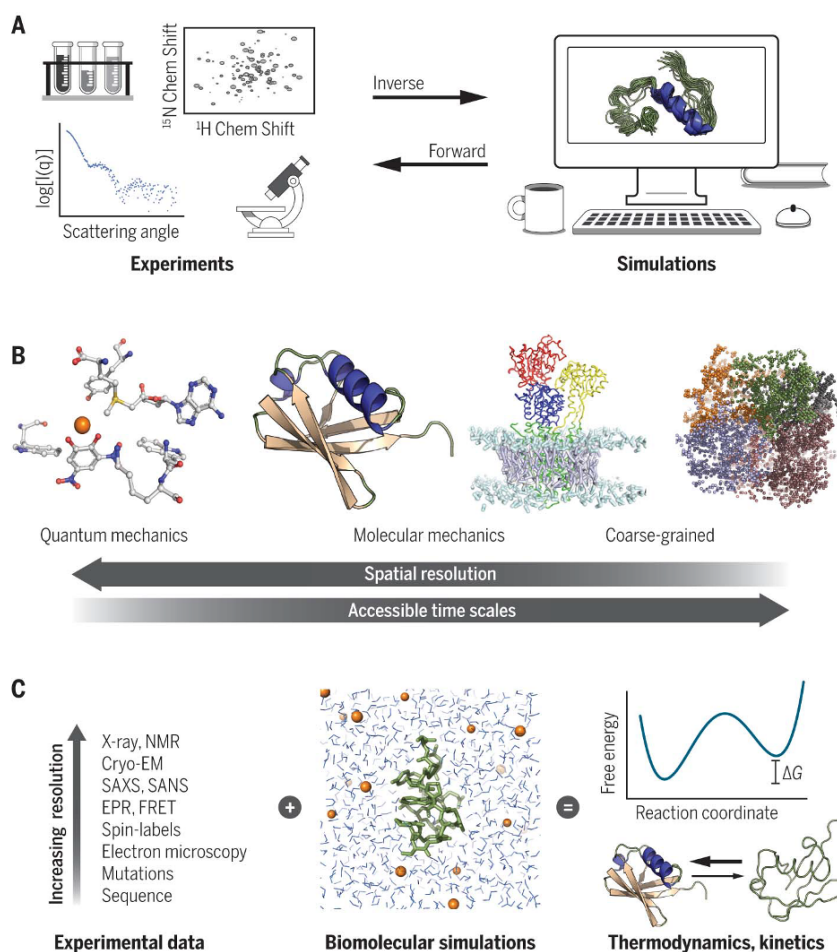


FIGURE 1.6 – Complémentarité simulations/expériences en biologie structurale [33]. **A**. Les données expérimentales sont utilisées pour nourrir les modèles théoriques qui en retour, peuvent prédire des ensembles de conformations. **B**. Simulations à différentes résolutions : de la structure électronique aux modèles gros-grains, les temps de simulations accessibles augmentent en diminuant la complexité des systèmes. **C**. Combinaison de données expérimentales et de modèles théoriques pour accéder aux propriétés thermodynamiques et cinétiques.

Pour décrire les mouvements des biomolécules en conditions physiologiques, la **dynamique moléculaire** reste de loin la méthode de prédilection. Comme les autres méthodes que je listerai rapidement, elle s'appuie sur le calcul de l'énergie potentielle associée à une conformation donnée, puis sur l'exploration du paysage énergétique pour déterminer la configuration la plus stable, *i.e.* de plus basse énergie [4].

Le coût de cette exploration a longtemps été extrêmement limitant, compte tenu de la taille des systèmes et des échelles de temps à couvrir, nécessitant de choisir entre temps de calcul et précision du modèle physique. D'incroyables progrès techniques (super calculateurs dédiés, processeurs graphiques, architectures parallèles) et méthodologiques (algorithmes, parallélisation), ont permis d'adresser ces limitations et d'atteindre des temps de simulation de l'ordre de la micro à la milliseconde [227].

Du point de vue méthodologique, un second obstacle porte sur la précision des modèles énergétiques sur lesquels reposent les simulations, notamment des **champs de forces capables de retranscrire les observables** [73, 79]. Là encore, d'importantes avancées ont permis de proposer de nouveaux modèles [159], les champs de force polarisables par exemple [103, 112]. Cependant, leur utilisation courante demeure difficile, conditionnée par l'accès aux avancées techniques que je viens d'évoquer. Ces progrès permettent néanmoins de simuler aujourd'hui des événements biologiques tels que les changements conformationnels, les repliements, les interactions protéines-ligands, les phénomènes de transport, les réactions chimiques.

La route demeure cependant longue pour accéder à l'ensemble de ces mécanismes "en routine" et le coût des calculs limite encore la taille des systèmes accessibles ainsi que la simulation d'événements allant au delà de la milliseconde. Davantage de progrès doivent être faits pour **rendre les simulations plus rapides et moins coûteuses tout en étant plus prédictives**. Au delà des progrès techniques, cela peut passer par le développement de **méthodologies multi-échelles** et l'**amélioration des techniques d'échantillonnage** de structures. Parmi les défis scientifiques qui en dépendent, on peut citer la prédiction des interactions protéines-protéines, où plus largement des grands complexes macromoléculaires, la prédiction des repliements (des acides nucléiques notamment), le criblage/design de médicaments, le design de biomolécules fonctionnelles, etc [73, 164].

3.2 Bref état de l'art méthodologique

Mon intention n'est pas de présenter ici un état de l'art exhaustif des méthodes de simulation existantes, mais de nous positionner par rapport à celles que nous utilisons et développons. L'ensemble des méthodes est très largement décrit dans la littérature. Je ne décris pas par exemple les techniques de détermination et de raffinement d'ensembles de structures [32, 127], les méthodes de docking [173, 188, 215], de design *de novo* de protéines fonctionnelles [17, 120, 125], de design et criblage de médicaments [68, 136], ou encore l'apport de l'intelligence artificielle [154, 163, 177, 220, 231, 236].

Nous nous intéressons ici aux approches théoriques pertinentes pour l'étude de la relation structure-fonction, laquelle nécessite plusieurs niveaux de simulation que nous étudions :

- l'identification des interactions "structurantes"
- l'évolution de la structure dans le temps (évolution "intrinsèque", dynamique)
- la réponse de la structure à des perturbations internes (mutations) ou externes (interactions)
- l'activité de la biomolécule (siège de réactions chimiques)

Toutes les échelles temporelles et spatiales doivent donc être couvertes : de la femto/pico seconde pour traiter les liaisons chimiques, à la seconde pour les mouvements de grande amplitude et les processus d'interactions. Parallèlement, les modèles doivent décrire la structure électronique et couvrir des déplacements d'atomes sur des distances allant de l'ordre du dixième à la dizaine d'angströms. Les modèles de structure et de calcul de l'énergie associée doivent donc être adaptés au système et au mécanisme étudié. **La résolution des problèmes en biologie passe donc généralement par l'utilisation d'approches multi-méthodes.**

Nous disposons aujourd’hui de différents outils dédiés aux biomolécules et permettant de traiter ces différentes échelles, au prix de concessions sur le coût ou la précision. Des choix doivent en effet être faits à plusieurs niveaux décrits par la suite (voir Fig. 1.7) : le modèle de représentation du système, d’interaction des particules, la méthode d’exploration de la surface d’énergie potentielle, et enfin les paramètres retranscrivant les conditions choisies pour la simulation.

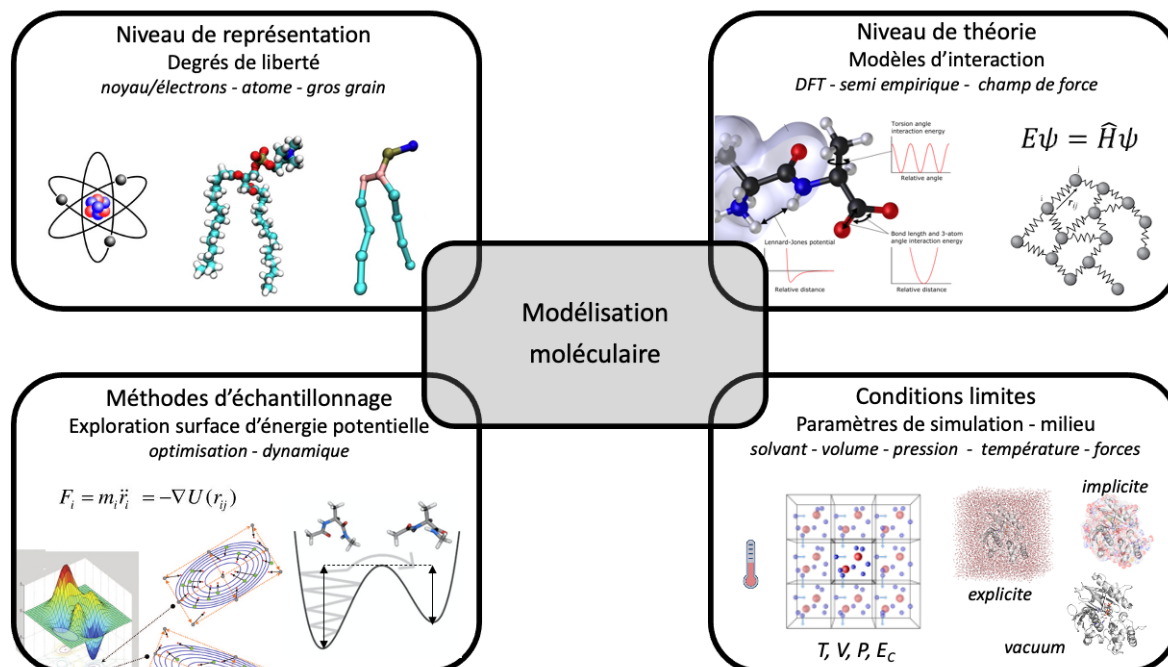


FIGURE 1.7 – Quatre choix se présentent selon le système à modéliser (taille, échelle de temps à atteindre, problème à traiter) et le mécanisme étudié (réaction chimique, changement conformationnel, etc) : i) le modèle de représentation du système, ii) le modèle d’interaction des particules, iii) la méthode d’exploration de la surface d’énergie potentielle, iv) les conditions de simulation.

Modélisation et niveaux de théorie : de la structure électronique aux modèles gros-grains

1) Pour étudier des mécanismes nécessitant de se placer au niveau de la liaison chimique, et pour des systèmes de quelques dizaines à quelques centaines d’atomes, des calculs basés sur la **structure électronique** peuvent être mis en œuvre. Les approches quantiques sont très coûteuses, et donc incompatibles avec des systèmes comprenant plusieurs milliers d’atomes. Elles peuvent néanmoins être utilisées localement, avec d’autres méthodes permettant d’accéder à l’ensemble du système. L’équation de Schrödinger ne pouvant être résolue analytiquement de façon exacte, différents modèles et approximations peuvent être utilisés, évaluant plus ou moins finement les interactions entre électrons, et permettant de traiter des systèmes de plus grande taille par ordre d’approximation croissant [34].

- Les méthodes *ab initio* basées sur la détermination de la fonction d’onde (méthodes de type Hartree-Fock (HF), basées sur un modèle de particules indépendantes, ou post-HF, prenant en compte la corrélation électronique ajoutée comme un terme perturbatif) sont les plus précises et les plus prédictives, mais sont inutilisables sur nos systèmes, sauf pour la paramétrisation de ligands par exemple.

- Les méthodes *DFT* (*Density Functional Theory*) ne traitent pas des fonctions d’ondes mais des densités électroniques. La valeur exacte de la densité conduit à l’énergie totale du système dans l’état fondamental, énergie décrite comme la somme des énergies cinétiques, d’attraction électrons-noyaux, de répulsion coulombienne entre électrons et d’échange-corrélation. Le calcul du terme d’échange-corrélation, le plus complexe, nécessite lui-même des approximations telles que la LDA (*local density approximation*), basée sur un modèle de gaz d’électrons uniforme, la GGA (*generalized gradient approximation*), utilisant le gradient de la densité pour rendre compte des variations de distributions

dans le nuage électronique, ou encore les fonctionnelles hybrides, dans lesquelles est intégrée une partie de l'échange exact, calculée en théorie HF. On citera enfin les pseudopotentiels qui ne traitent explicitement que les électrons de valence, considérés en interaction avec le potentiel créé par le noyau et les électrons de cœur. Les calculs DFT sont plus accessibles mais encore coûteux et ne peuvent être utilisés que sur des systèmes de taille restreinte (de l'ordre de la centaine d'atomes).

- Enfin, les *méthodes semi-empiriques* (SE) simplifient les calculs en utilisant des données issues de calculs HF ou DFT et ajustées sur des résultats expérimentaux [59]. Les modèles SE permettent d'accélérer les calculs d'un facteur de plusieurs ordres de grandeurs par rapport aux méthodes *ab initio* et sont largement utilisés pour les systèmes biologiques. Ils peinent cependant parfois à décrire suffisamment précisément les interactions non covalentes, restreints par les limitations des méthodes dont ils sont issus.

2) La **mécanique moléculaire classique** (MM) permet quant à elle de simuler des molécules de type protéines ou ADN/ARN. Etant donnée la température à laquelle se produisent les mécanismes d'intérêt (ambiante ou physiologique), et puisque les énergies impliquées sont de l'ordre de 1-10 $k_B T$ (k_B étant la constante de Boltzmann), les processus sont largement déterminés par les lois de la mécanique statistique classique. Les atomes sont alors représentés par des particules sphériques dont les interactions sont décrites par une fonction d'énergie dont le formalisme et la précision sont critiques pour le calcul. Dans la plupart des **modèles tout-atomes**, l'énergie du système est donnée par une somme de termes d'énergies potentielles associés à un "**champ de force**" dont la forme est généralement :

$$U_{total} = \sum_{liaisons} K_r (r - r_0)^2 \quad (1.1)$$

$$+ \sum_{angles} K_\theta (\theta - \theta_0)^2 \quad (1.2)$$

$$+ \sum_{torsions} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{impropres} K_{imp} (\phi - \phi_0)^2 \quad (1.3)$$

$$+ \sum_{\substack{paires \\ non\ liées\ i,j}} \left(\epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \quad (1.4)$$

Quatre termes décrivent les interactions covalentes, *i.e.* les énergies d'élongation de liaison (1.1), de déformation angulaire (1.2) et de rotation (torsions propres et impropres (1.3)). Les interactions non covalentes sont traitées comme la somme des interactions de van der Waals et électrostatiques (1.4).² **La forme de ce potentiel peut paraître simple, mais la complexité du traitement repose sur le nombre et la variété des liaisons à traiter.** Cette variété implique aussi la paramétrisation d'un très grand nombre de termes constituant le champ de force. Ces paramètres sont dérivés de données expérimentales ou de calculs quantiques menés sur petites molécules et affinés pour les protéines et les acides nucléiques [4]. De plus en plus, ce domaine profite également des progrès expérimentaux qui permettent de développer de nouveaux champs de forces à partir de données obtenues sur les macromolécules [73]. Au final, ce traitement additif des interactions entre toutes les paires d'atomes doit **représenter au mieux le paysage énergétique du système considéré.**

Les champs de forces de première génération réussissent à bien retranscrire les propriétés structurales et dynamiques des biomolécules. Ils affichent cependant certaines faiblesses, notamment dues au découplage total entre les termes du potentiel, ou encore pour reproduire les effets de la redistribution de charges induite par l'environnement et non traitée par le potentiel de Coulomb qui considère des charges fixes attribuées à chaque atome.

2. K_r , K_θ , K_χ et K_{imp} sont les constantes de forces des potentiels respectivement décrits, r , θ , χ et ϕ la longueur, l'angle, l'angle dièdre et l'angle dièdre impropre de la liaison considérée, et r_0 , θ_0 , et ϕ_0 leurs valeurs à l'équilibre. Dans (1.3), n est la périodicité de la torsion et δ l'angle de phase. Dans (1.4), ϵ_{ij} correspond à la profondeur du puits du potentiel de Lennard-Jones, ϵ_0 à la permittivité du milieu, r_{ij} est la distance séparant les deux atomes en interaction q_i , et q_j leurs charges respectives et $R_{min_{ij}}$ la valeur de r_{ij} pour laquelle le potentiel s'annule.

Une nouvelle génération de champs de force plus sophistiqués cherche à compenser ces faiblesses, par exemple en ajoutant des termes croisés élongation/déformation angulaire, ou en mimant certains comportements anharmoniques. On note surtout l'apparition des champs de force polarisables, prometteurs, mais coûteux [112]. Ces derniers sont basés sur différents modèles tenant compte de la polarisabilité des molécules (dipôles induits, multipôles distribués, densité électronique) et modulent les interactions électrostatiques, surestimées dans les champs de force non polarisables pour compenser l'absence de polarisation. Ces nouveaux champs donnent de très bon résultats pour l'étude de certains mécanismes, par exemple pour les interactions protéines-ligands [103].

3) D'autres modèles que les modèles tout-atomes existent pour traiter les systèmes de très grande taille et/ou atteindre des temps de simulation plus longs, difficilement atteignables en dynamique moléculaire. Dans ces **modèles dits "gros grains"** (*coarse grained*, CG), les particules ne sont plus des atomes mais des ensembles d'atomes [118, 153, 185], ce qui réduit considérablement la complexité du système, et "lisse" le paysage énergétique, plus ou moins selon le niveau de résolution du modèle. Des exemples de représentations gros-grains usuelles sont donnés sur la figure 1.8.

L'utilisation de modèles gros-grains est souvent accompagnée de modèles énergétiques simplifiés. Je ne citerai ici que le plus "simple", mais robuste : *Gaussian Network Model* (GNM) qui donne une forme quadratique à l'énergie intramoléculaire. Il peut être paramétré pour être utilisé à différentes échelles et rendre compte de divers mouvements fonctionnels [247].

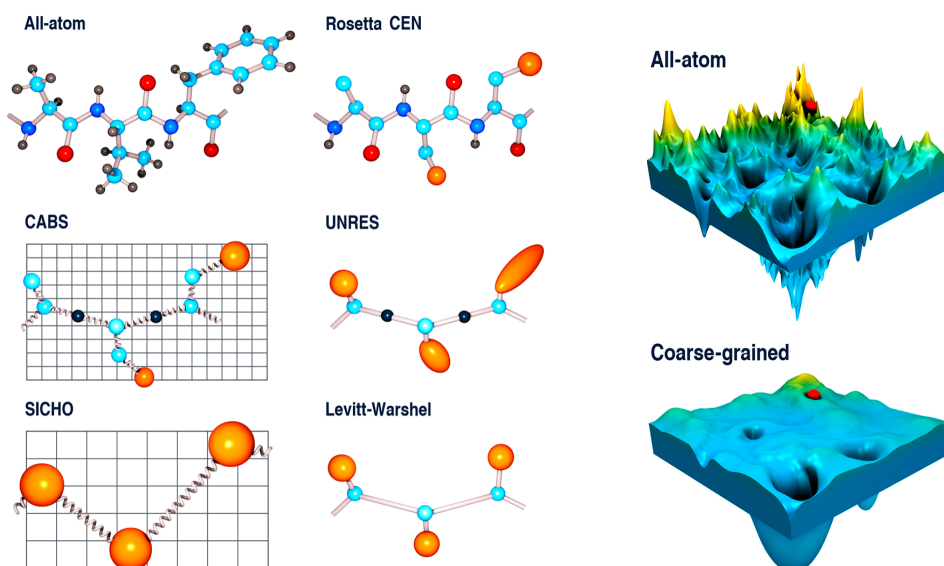


FIGURE 1.8 – A gauche, représentations tout-atomes et gros-grains d'un tripeptide (modèles Rosetta, CABS, UNRES, SICHO, et Levitt-Warshel). A droite, illustration du paysage énergétique simplifié par le modèle gros-grain (tout-atomes en haut, Levitt-Warshel en bas). (Source : [118])

4) Il est enfin possible de coupler les modèles lorsque tous les atomes du système ne nécessitent pas le même niveau de traitement. C'est le cas lorsque l'on veut étudier les réactions enzymatiques [198], ou encore pour évaluer les interactions protéine/ligand [107]. Les **approches hybrides QM/MM** (*quantum mechanics/molecular mechanics*) [199, 238] permettent d'utiliser différents niveaux de description et de traiter localement la chimie dans une région quantique, tout en prenant en compte le reste du système par mécanique moléculaire classique. Cette méthode est aujourd'hui largement employée. Elle doit être néanmoins paramétrée avec précaution, les principales difficultés reposant sur les liaisons entre régions QM et MM, ainsi que sur le choix de la région QM [191]. Ces problèmes sont minimisés par le choix d'une région QM très étendue, mais qui augmente drastiquement le coût des calculs. Les méthodes DFT ou semi-empiriques sont traditionnellement employées pour décrire l'énergie associée [111].

Choisir le niveau de représentation du système, c'est-à-dire les degrés de liberté permis, constitue la première étape de la modélisation. La seconde repose sur le choix du modèle décrivant l'énergie du système en fonction de ses degrés de liberté. Nous venons de voir que l'on peut aujourd'hui choisir parmi une large gamme de champs de forces, connaissant les spécificités et les limites de chacun, afin d'aller au plus près de la réalité du mécanisme étudié. La troisième étape consiste enfin à parcourir la surface d'énergie potentielle pour accéder à la structure et la fonction associée du système. De nombreuses approches existent pour réaliser cette exploration, généralement par l'apport de forces extérieures. Quelques unes de ces techniques sont décrites dans le paragraphe suivant.

Différentes méthodes d'exploration de la surface d'énergie potentielle

La rupture d'une liaison chimique, la rotation d'une chaîne latérale, l'effet d'une mutation sur la structure, l'amarrage d'un ligand, le repliement d'une protéine, etc. Simuler tous ces mécanismes nécessite de calculer l'énergie d'un état, voire le chemin entre deux états. Compte tenu du nombre de degrés de liberté à traiter dans ces systèmes, la surface est très accidentée, avec des bassins de toutes tailles et profondeurs. La recherche d'un minimum global correspondant à une structure unique est donc quasiment impossible d'accès. Mais surtout, elle n'est pas pertinente car la réalité du système passe par la description d'un ensemble statistique de configurations r , d'énergie potentielle $V(r)$, et associées à une probabilité $P(r) \sim e^{(-V(r)/k_B T)}$ ³. Les états d'équilibre représentatifs sont donc caractérisés par les faibles valeurs de $V(r)$.

Le concept de minimisation d'énergie est ainsi fondamental. Considérant que r contient N variables indépendantes ($r = (r_1, r_2, \dots, r_n)$ étant les coordonnées des atomes/particules), on cherche à déterminer la valeur de chacune de ces variables permettant de se placer à un minimum global de la fonction $V(r)$. Une exploration exhaustive de l'espace des conformations, dans le cas des macromolécules, paraît inenvisageable. De nombreux algorithmes existent en revanche par une approche heuristique [130].

La recherche d'un minimum peut tout d'abord s'opérer de manière statique, par des algorithmes de minimisation d'énergie. Ceux-ci peuvent être utilisés pour raffiner une structure, par exemple pour relaxer un système soumis à une perturbation. Des distorsions locales, des problèmes de clash stérique peuvent ainsi être résolus en laissant le système "glisser" au fond du puits de potentiel. Cependant, ces méthodes sont inappropriées pour passer des barrières énergétiques, c'est-à-dire aller explorer d'autres puits alentour, et donc échantillonner efficacement l'espace des conformations. Parmi les méthodes usuelles, on peut distinguer celles qui se basent sur les dérivées premières de l'énergie (de type *gradient conjugué* ou *steepest descent*), rapides, mais qui peuvent peiner à converger, et les dérivées secondes (*Newton Raphson*), plus lourdes, mais capables de converger sur une valeur exacte du minimum. Dans les deux cas, le minimum trouvé ne peut être que local et d'autres approches doivent être envisagées pour trouver, si ce n'est un minimum global, au moins plusieurs autres minima locaux.

La dynamique moléculaire en particulier se positionne comme référence pour explorer efficacement un très grand nombre de conformations. Elle se base sur l'intégration des équations du mouvement dans le temps. Les forces F_i sont d'abord calculées sur chaque atome i , à intervalles de temps t réguliers, selon : $F_i = -\nabla V(r_i)$, puis utilisées pour attribuer les nouvelles vitesses v_i et positions r_i puisque : $dv_i/dt = F_i/m_i$ et $dr_i/dt = v_i$. On peut alors réitérer le processus jusqu'à atteindre le temps de simulation souhaité. L'algorithme de résolution utilisé est généralement l'algorithme de Verlet, lequel utilise les positions et les accélérations à l'instant t , ainsi que les positions du pas précédent $t - \delta t$ pour déterminer les nouvelles positions à $t + \delta t$. Le pas de temps δt est choisi de l'ordre de la femtoseconde, garantissant la conservation de l'énergie du système [4]. Notons enfin que le calcul des forces peut être fait à partir de n'importe quel potentiel vu précédemment, quantique ou classique. On distinguera ainsi les dynamiques *ab initio* et newtonienne. Les méthodes hybrides QM/MM peuvent également être utilisées, ainsi que les modèles gros grains [221] (voir Fig. 1.9).

3. k_B étant la constante de Boltzmann et T la température

De nombreuses approches sont développées afin de limiter les temps de calcul, au risque de perdre de la précision dans l'exploration. Pour caractériser des mouvements à grande échelle, une voie est celle de la **simplification des modèles** (gros grains [178]) ou la **limitation des degrés de liberté** en s'appuyant sur les mouvements collectifs d'atomes. Les modèles de réseaux élastiques peuvent alors être employés et couplés à un calcul de modes normaux. On assume alors que les modes de plus basses fréquences, associés aux mouvements collectifs d'atomes ou de particules, peuvent être utilisés pour représenter certains mouvements fonctionnels et "diriger" l'exploration [21,123]. Ce type d'approche connaît des succès sur les grands assemblages moléculaires en particulier mais présente plusieurs limites, notamment celle la perte de résolution et de réalisme physique, mais aussi surtout, l'absence de lien *a priori* entre modes normaux et mouvements fonctionnels [25]. Ce n'est qu'après la simulation que l'on peut confirmer ou non l'implication des modes normaux dans la fonction biologique.

D'autres approches consistent à **biaiser les calculs en "lissant" la surface ou en favorisant la rencontre d'événements rares**. On peut citer en particulier la technique *umbrella sampling* [116] qui introduit une contrainte harmonique pour forcer l'exploration d'une région le long d'une coordonnée de réaction, la dynamique moléculaire dirigée (*steered molecular dynamics*) [105] qui introduit une force basée sur la distance entre coordonnées initiales et coordonnées cibles tout en générant des états intermédiaires. On peut également citer la *métadynamique* [20,129], qui ajoute un terme énergétique défavorable aux régions déjà explorées afin de forcer le système à s'éloigner de son état initial et à passer les barrières énergétiques, ou encore la dynamique moléculaire accélérée [88] qui fait appel à un potentiel qui diminue la hauteur des puits et accélère l'échantillonnage.

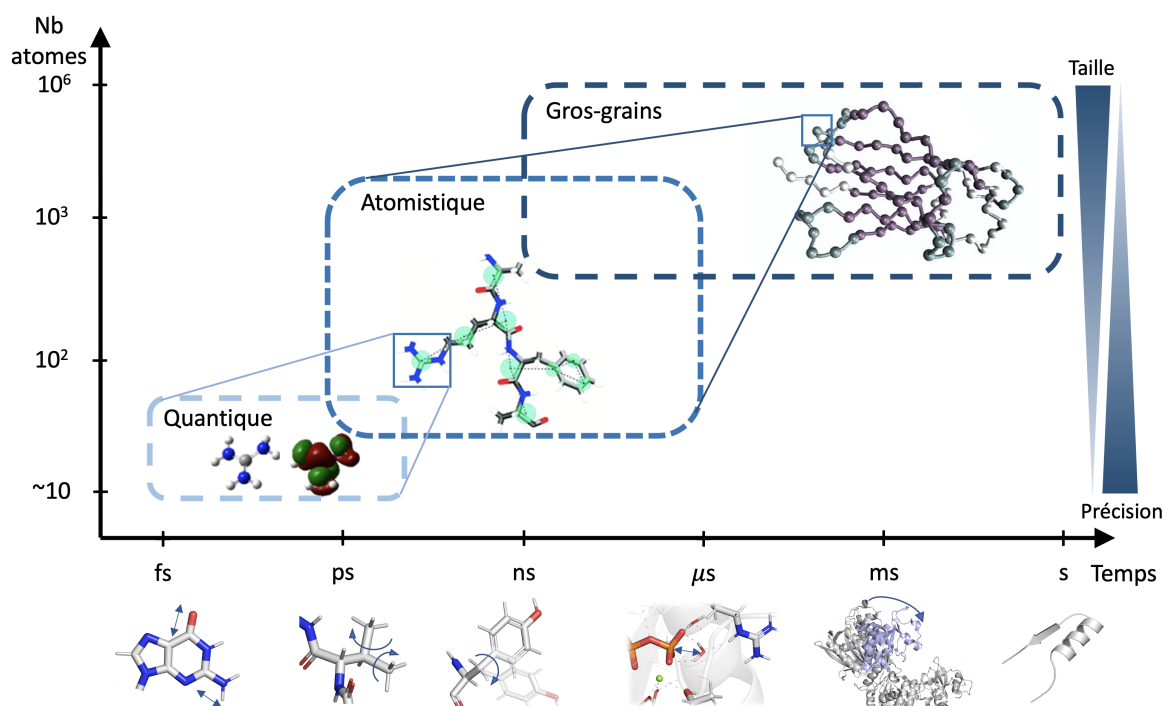


FIGURE 1.9 – Représentation des différentes échelles de taille et de temps accessibles par dynamiques moléculaires *ab initio*, classique tout-atomes et gros-grains. Des exemples d'événements sont donnés en dessous, de la gauche vers la droite, par temps croissant : *fs*) vibration de liaison, *ps*) rotation d'un groupe méthyle, *ns*) rotation d'une chaîne latérale, *μ s*) réaction chimique, *ms*) mouvements collectifs d'atomes, *s*) repliement.

Enfin, des alternatives à la dynamique moléculaire existent pour **échantillonner l'espace des conformations**, comme la *dynamique de Langevin* [104], basée sur l'équation de Langevin plutôt que la loi de Newton. Des termes sont inclus pour ajouter un coefficient de friction ζ et simuler les impulsions aléatoires générées par le solvant. La *dynamique brownienne* utilise une approche semblable avec un coefficient ζ suffisamment grand pour permettre de simuler une marche aléatoire [97]. Les simulations de type *Monte Carlo* permettent d'accéder aux propriétés structurales et thermodynamiques du système en générant un grand nombre de configurations associées à une probabilité d'existence.

Elles ne sont cependant pas compétitives pour explorer l'espace des configurations par rapport à la dynamique moléculaire sur des systèmes biologiques et ne donnent pas d'information sur leur évolution temporelle. Des techniques ont pour but d'optimiser cet aspect, en particulier les *chaînes de Markov* qui permettent de générer une configuration avec la connaissance de la configuration précédente, mais pas des configurations antérieures [99, 157]. Ces méthodes et bien d'autres encore sont largement décrites dans la littérature [4, 13, 98, 169].

Conditions de simulations : quels paramètres pour intégrer les facteurs d'environnement ?

Un dernier choix stratégique devant être opéré est celui des **conditions de simulations qui, dans l'idéal, doivent être aussi proches que possible des conditions expérimentales**. En dynamique moléculaire classique, il est ainsi possible d'intégrer des conditions physiques avec contrôle de pression P et de température T comme grandeurs indépendantes. Le nombre de particules N étant conservé, il est possible de travailler dans différents ensembles : micro-canonique NVE (volume V et énergie E constants), canonique NVT (V , T constants), et isotherme-isobare NPT (P , T constantes). Différents thermostats permettent de contrôler la température en ajustant la vitesse des atomes/particules pour faire tendre T vers une valeur cible (thermostats de Berendsen, Nosé-Hoover, Andersen ou Langevin [101]). La pression peut quant à elle être contrôlée en ajustant le volume de la boîte de simulation ou les coordonnées atomiques (barostats de Berendsen et Andersen [14, 26]).

Par ailleurs, la prise en compte de **la contribution du solvant est cruciale pour retranscrire les propriétés des systèmes biologiques**. Les simulations sont donc menées en solution, dans une boîte de solvant pouvant être de différentes formes (cubique, octaédrique, etc) et dans laquelle on s'affranchit des effets de bord en ayant recours aux **conditions aux limites périodiques**. La boîte de simulation est alors dupliquée dans toutes les directions et entourée par ses images. Ainsi, un atome qui sort par un côté de la boîte y rentre par le côté opposé avec la même vitesse. Le nombre d'atomes reste constant dans la boîte et le système électriquement neutre. Dans ce **traitement explicite du solvant**, les molécules d'eau sont ajoutées physiquement au système, avec un nombre de degrés de liberté qui dépend du modèle. Inclure explicitement le solvant est donc impossible dans les approches *ab initio* ou DFT car il augmenterait le nombre d'atomes et le temps de calcul de façon rédhibitoire. En mécanique classique en revanche, soluté et solvant sont représentés de façon cohérente, et le mouvement des molécules d'eau peut alors être calculé pour étudier leur trajectoire, les interactions établies avec le soluté, et leurs conséquences sur l'évolution de sa structure. Plusieurs modèles explicites existent, selon le nombre de sites utilisés pour définir les atomes (de 3 à 6) [226]. Le modèle TIP3P, le plus utilisé, considère une molécule d'eau à 3 sites d'interactions (un par atome). Les charges et rayons associés aux atomes, les longueurs de liaisons ainsi que la valeur de l'angle de valence sont définis et fixes. Les interactions faibles intermoléculaires sont évaluées au travers d'un potentiel électrostatique et de Lennard-Jones.

Le traitement explicite du solvant rend compte de conditions de simulation réalistes, mais il nécessite d'ajouter un très grand nombre de molécules d'eau au système. L'évaluation des interactions solvant/solvant augmente alors drastiquement la durée du calcul. Afin d'alléger les coûts de calculs, mais aussi d'accéder plus efficacement aux grandeurs thermodynamiques du système (comme par exemple l'énergie de solvatation qui nécessiteraient un traitement statistique sur un nombre de configurations trop important pour donner un résultat pertinent), le solvant peut être traité de façon implicite. On considère alors que la contribution de l'eau sur le système est principalement de nature électrostatique et qu'elle peut être moyennée dans le temps et l'espace. Il existe plusieurs **modèles de solvant implicite**, le plus connu étant le modèle de Born généralisé (*Generalized Born Model*) [166]. Basé sur la théorie de Poisson-Boltzmann, il traite les molécules d'eau comme un continuum de forte constante diélectrique qui entoure le soluté, représenté comme une cavité de faible constante diélectrique, et agit comme une perturbation du système. Ce traitement implicite présente le grand avantage d'être peu coûteux en terme de temps et de ressources de calcul par rapport au traitement explicite des molécules d'eau. Il peut également être plus efficace pour échantillonner l'espace des conformations, couplé à une faible viscosité du milieu [5]. Il calcule enfin des valeurs moyennes, par principe, et permet donc une évaluation directe des interactions solvant/soluté pour les estimations d'énergie libre [182].

Nous disposons donc aujourd’hui d’un large panel de modèles et méthodologies permettant de simuler les mécanismes inhérents à la fonction des biomolécules. **Le choix de la stratégie à adopter dépend de la propriété étudiée, du niveau de précision attendu et de la taille du système à traiter, au regard de la puissance des ressources disponibles.**

Les progrès futurs passeront par la constante amélioration des modèles énergétiques et des ressources informatiques afin d’accéder de façon prédictive, et aussi exacte que possible, à l’ensemble des mécanismes sur l’ensemble des échelles. La représentation tout-atomes demeure néanmoins incontournable pour comprendre les mécanismes d’interactions, le fonctionnement et le dysfonctionnement des biomolécules, et donc aller vers des stratégies de design.

3.3 Bilan : quelles questions veut-on et peut-on adresser ?

“Tout ce qui va arriver peut et doit être prévu.” [189]

Les techniques de simulation sont aujourd’hui assez matures pour offrir une vision unique et détaillée au niveau atomique des propriétés moléculaires, et pour permettre une comparaison quantitative et sur des échelles de temps pertinentes avec l’expérience. Elle le sont suffisamment pour conduire à une **interaction efficace et productive entre théorie et expérience, offrant de nouvelles possibilités de validation et de développement de modèles.**

Le calcul peut ainsi se mettre au service de problématiques aussi variées que l’identification de nouvelles cibles, la proposition de nouvelles stratégies pharmacologiques, la vectorisation de médicaments, l’étude des mécanismes de résistance, etc, pour ce qui concerne les aspects purement biologiques à visée médicale. Les avancées en matière de connaissance fondamentale ont par ailleurs ouvert de nouveaux champs d’applications possibles, dans lesquelles les propriétés fonctionnelles, de reconnaissance spécifique, ou encore d’auto-assemblage des biomolécules sont utilisées et détournées dans des procédés et dispositifs technologiques. Dans les deux cas, **on ne peut s’intéresser au mécanisme d’action de la molécule d’intérêt sans considérer son intégration dans un système souvent complexe, ni la façon dont cet environnement va agir sur la fonction.**

Pour traiter ces questions et **asseoir le caractère prédictif des calculs**, nous devons porter nos efforts sur des développements méthodologiques permettant : de traiter les effets d’environnement à la fois de manière réaliste et efficace, d’accéder à toutes les échelles d’espace et de temps, de simuler les réactions chimiques en incluant leur caractère dynamique, d’améliorer les algorithmes de reconnaissance moléculaire, ou *docking*, afin d’effectuer du criblage virtuel haut débit, quelles que soient la taille et la flexibilité des partenaires en jeu. **Les méthodes multi-échelles et multi-modèles sont certainement la clé pour relier les modifications chimiques et les contraintes physiques locales aux observables macroscopiques.**

Concernant les "perturbations" intramoléculaires, il s’agit principalement de mutations qui peuvent être générées *in silico* et dont les effets sur la structure et la dynamique peuvent être évalués. Concernant les contraintes externes, nous les considérons comme étant de deux types. D’une part, les paramètres de l’environnement (solvant, pH, salinité, pression, température) peuvent être inclus comme nous l’avons vu dans le paragraphe précédent. D’autre part, un second type de contraintes peut être considéré sous forme de forces extérieures appliquées au système, soit pour le soumettre à une interaction (forces intermoléculaires), soit pour simuler une contrainte "mécanique". Ce dernier type de contrainte ouvre la voie à la **manipulation de molécules *in silico* dans le but d’évaluer leur capacité à répondre à une contrainte extérieure.**

D'un point de vue expérimental, ce type de manipulation existe via des techniques biophysiques telles que les pinces optiques, magnétiques, la microscopie à force atomique, etc [145,160], par exemple pour mesurer l'élasticité de brins d'ADN, ou encore la force développée par une protéine motrice. D'un point de vue calculatoire et conceptuel, cette approche peut paraître artificielle. Pourtant, les cellules exercent, détectent, transmettent des forces en permanence, forces qui traduisent une interaction mécanique et continue avec l'environnement, et qui déterminent la façon dont les cellules prolifèrent, se différencient, ou encore interviennent dans la tumorigenèse. La mécanobiologie étudie ainsi les effets des contraintes mécaniques sur les cellules [234]. En leur sein, la communication entre objets moléculaires, leur rôle dans la signalisation cellulaire, les pathologies et les traitements associés sont le plus souvent étudiés par des approches de biochimie (affinités de liaison, cinétiques de réactions). Le rôle des contraintes mécaniques commence cependant à être également considéré, ouvrant la voie à la mécanobiologie moléculaire et à une nouvelle vision de la régulation de la fonction biologique [126,250].

Notre connaissance des phénomènes de mécanotransduction à l'échelle moléculaire, induits par des stimuli chimiques ou physiques, est encore immature et nécessite des efforts pour comprendre par exemple comment des protéines peuvent être activées ou désactivées par ces mécanismes, et ainsi devenir de nouvelles cibles médicamenteuses ou de nouvelles sources d'inspiration pour les bionanotechnologies.

C'est sur cette problématique que portent nos travaux et les méthodologies que nous développons au laboratoire. **Nous combinons les méthodes usuelles de simulation à des approches mécanistiques nous permettant de manipuler les biomolécules en les soumettant à différentes contraintes pouvant être rencontrées au cœur des cellules, lors de leur intégration dans un dispositif technologique, ou bien conçues et appliquées afin d'acquérir une meilleure compréhension de la façon dont ces machines moléculaires répondent aux contraintes.** Nous pouvons ainsi accéder sous un angle original, mais aussi très peu coûteux, à leur caractéristiques intrinsèques, comprendre leurs mécanismes fondamentaux, simuler les modifications pouvant conduire à un dysfonctionnement, et proposer de nouvelles stratégies pour contrôler leur réponse. Les fondements de cette méthodologie, l'approche par "*Modes Statiques*" a été développée durant ma thèse et est introduite dans la section suivante.

4 Les Modes Statiques : un outil versatile pour la manipulation des biomolécules

4.1 Concept et méthode

La méthode des Modes Statiques (MS) a été conçue pour **calculer à faible coût les déformations induites par une contrainte appliquée à une biomolécule**, initialement pour prendre en compte la flexibilité totale de deux partenaires en interaction dans un algorithme de docking et ainsi traduire les effets d'ajustement induit.⁴ Elle peut être appliquée à tout type de système, mais elle est, par essence, particulièrement **adaptée aux spécificités des biomolécules** (flexibilité conformationnelle, interactions, faible coût de calcul permettant de s'affranchir de la taille des systèmes macromoléculaires). Elle permet de caractériser de façon systématique la réponse d'un système d'atomes (ou de particules) soumis à des contraintes ou modifications physico-chimiques, internes ou externes, traduites comme des forces appliquées aux atomes.

4. Les méthodes de docking, ou amarrage moléculaire, visent à prédire l'interaction de petites molécules dans le site de liaison d'une protéine cible, ou encore les modes d'interaction de macromolécules (complexes protéines-protéines en particulier). Plusieurs méthodes existent, mais considèrent généralement que le récepteur, voire les deux partenaires, sont rigides ou semi-rigides, afin de limiter le nombre de degrés de liberté et de faciliter la recherche de poses. Les prédictions sont basées sur la complémentarité des surfaces et/ou le calcul de l'énergie du complexe. La flexibilité totale ne peut-être prise en compte en raison de la complexité et de la lourdeur des calculs associés. La dynamique moléculaire peut-être utilisée sur un complexe pour en affiner la structure, mais en aucun cas comme un outil de traitement systématique. [173,215]

L'idée est de pouvoir explorer un nombre de conformations inaccessible avec des méthodes telles que la dynamique moléculaire, trop coûteuse en temps. Dans une optique de simplification, les déformations permanentes d'une molécule soumise à une excitation extérieure (telle qu'une interaction) sont calculées, sans état intermédiaire, ni notion de temps. Les Modes Statiques diffèrent des Modes Normaux qui représentent des vibrations spontanées en proposant au contraire des déformations induites. Cette approche permet ainsi de sortir du cadre de la dynamique intramoléculaire pour considérer des **interactions intermoléculaires**.

Le calcul est basé sur la matrice Hessienne H dont les éléments sont relatifs aux constantes de forces décrivant les interactions entre atomes (ou particules). Un intérêt notable de cette approche est qu'elle peut être appliquée à des systèmes tout-atomes comme à des systèmes gros-grains, tout en utilisant des modèles quantiques ou classiques, selon la taille des systèmes et la précision recherchée. Elle permet par ailleurs de conserver les informations physico-chimiques pertinentes en manipulant des modèles énergétiques adaptés et les paramètres du milieu, tout en permettant de traiter les biomolécules comme des objets d'ingénierie manipulables et modifiables à façon.

Une fois les Modes Statiques calculés, il est possible de **les combiner pour créer une excitation spécifique**, à l'aide d'outils que nous développons au fil des ans et des projets, et que je présenterai dans la suite de ce manuscrit. Nous cherchons ainsi à proposer un véritable outil d'ingénierie moléculaire *in silico*, permettant d'**évaluer les propriétés biomécaniques et intrinsèques d'une molécule**, mais aussi de caractériser, d'optimiser et de proposer des constructions moléculaires.

4.2 Algorithme

Le calcul est basé sur la matrice Hessienne H d'une structure à l'équilibre, obtenue après minimisation énergétique par la méthode et le modèle énergétique de son choix. Les éléments de H sont utilisés pour simuler l'excitation systématique des atomes du système et calculer les déformations résultantes, ou *Modes Statiques*, chacun correspondant à une contrainte spécifique sur un site atomique donné. Afin d'éliminer les 3 translations et les 3 rotations qui laissent la molécule invariante, 3 atomes sont tout d'abord fixés en translation, en azimut et en longitude. On obtient une matrice H de taille $(3N - 6) \times (3N - 6)$, N étant le nombre d'atomes.

La version actuelle du code se place dans l'approximation harmonique. Dans ce cas, l'énergie totale E peut s'écrire : $E = E_0 + H_{i,j}x_i x_j$ où les termes $H_{i,j}$ sont les éléments de la matrice Hessienne, x_i les coordonnées cartésiennes des déplacements atomiques pour une configuration à l'équilibre d'énergie E_0 . Le vecteur force, gradient de l'énergie, peut alors s'écrire $-F_i = \sum_j H'_{i,j}x_j$ (1). La méthode retenue pour résoudre ce système consiste à considérer un déplacement donné x_l , résultant de l'application de la force associée F_l , (x_l est associé à la contrainte de Lagrange, F_l agit comme une force de réaction). Dans un premier temps, la l ème ligne de H , associée à la contrainte F_l est éliminée du système (1), ainsi que la l ème colonne, contenant les termes en x_l . En nommant cette colonne B , et A le reste de la matrice $(3N - 7) \times (3N - 7)$, il reste à résoudre le système $A.X + B = 0$, où le vecteur X , appelé un "Mode Statique", contient les déplacements atomiques imposés par la contrainte. Par élimination successive des lignes de la matrice H , l'ensemble des MS peut être déterminé. Ces modes, caractéristiques du système, sont ensuite stockés pour un usage ultérieur, pouvant être combinés à volonté pour calculer les champs de déformations résultants de contraintes complexes .

4.3 Un outil validé et évolutif

L'algorithme de calcul des Modes Statiques est implémenté dans le **code Flexible** [38], déposé à l'APP [62]. Pour chaque système traité, une **banque de données des Modes Statiques** est créée pour pouvoir ensuite être **réutilisée dans diverses procédures de post-traitement**. Ces procédures, basées sur des combinaisons linéaires de modes, en raison du cadre de l'approximation harmonique dans lequel nous travaillons, nous permettent alors de **manipuler les modes** en créant l'excitation voulue à l'aide d'outils que nous développons, et d'obtenir une réponse immédiate du système (voir Fig. 1.10). Nous revendiquons ici une approche simple mais efficace lorsqu'utilisée dans son domaine de validité.

Durant ma thèse, j'ai validé le caractère prédictif de cette approche sur des systèmes variés parmi lesquels je citerai : la déformation de polymères thermosensibles (PNIPAM) [38], la réponse conformationnelle de brins d'ADN contenant des acides nucléiques contraints [40], l'effet de mutations sur la stabilité du site actif d'une enzyme (DHFR) [39], le repliement d'un peptide (β -amyloïde) induit par l'approche d'un ion [183], ou encore la simulation du changement conformationnel d'une protéine (protéase du HIV-1) induit par l'amarrage d'un ligand [43]. Cette liste n'est pas exhaustive et ne sera pas développée ici car je ne détaille dans ce manuscrit que mon travail post-thèse.

Cependant, un point important à noter ici est que la méthode des MS, initialement développée pour être incluse dans une procédure de docking, ne l'a pas encore été véritablement. Au travers de ces cas d'études, nous avons tout d'abord cherché à montrer que les MS permettaient d'**accéder à des déformations pertinentes, induites par des contraintes spécifiques**. Si le dernier exemple donné ci-dessus constitue une preuve de concept de la légitimité d'utiliser les MS dans une approche de docking flexible, le ligand n'était pas physiquement présent dans les simulations, son effet était traduit au travers des forces qu'il aurait exercé sur la protéine en étant positionné dans son site actif. Si des travaux ont été entrepris pour implémenter un algorithme de docking incluant les MS, ce dernier n'est pas finalisé et demeure un projet d'envergure pour le futur sur lequel je reviendrai à la fin du manuscrit. Le cas du repliement du peptide β -amyloïde au voisinage d'un ion Zn^{2+} incluait en revanche l'ion et ainsi qu'un calcul des interactions électrostatiques et de van der Waals entre les deux partenaires. Il s'agissait donc d'un exemple très simple, mais validé, de changement conformationnel induit par l'approche d'un ligand.

Même sans aller jusqu'au docking, ces exemples ont prouvé **l'intérêt et l'efficacité de l'utilisation de l'approche MS comme outil pour manipuler les molécules *in silico* et pour sonder leur capacité intrinsèque à répondre à une contrainte**. Il s'agit de la voie que je continue de développer et que je présenterai rapidement dans le paragraphe suivant.

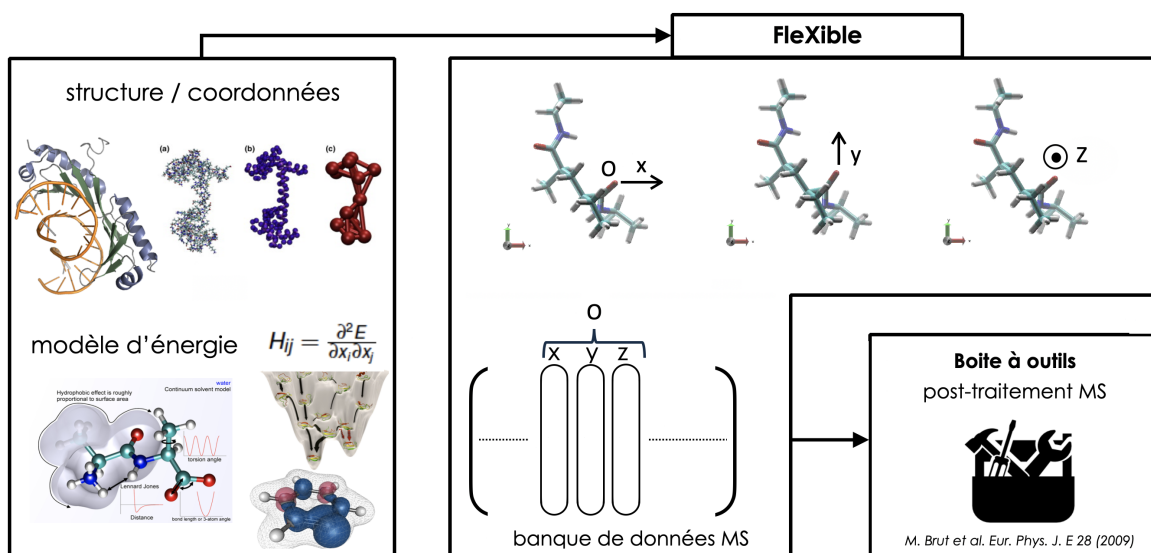


FIGURE 1.10 – La méthode de MS est implémentée dans le code FleXible qui nécessite en entrée les coordonnées et la matrice hessienne H du système considéré. FleXible calcule l'ensemble des MS à partir des éléments de H , en simulant le déplacement de chaque atome du système (ici un atome d'oxygène), dans chaque direction x , y , z de l'espace. Chaque déformation résultante, correspondant à une contrainte spécifique sur un site atomique donné, est appelée un "mode statique". Les modes statiques sont ensuite stockés dans une banque de données pour un usage ultérieur à l'aide d'outils que nous développons.

4.4 Pour une approche d'ingénierie *in silico*

Je montrerai dans ce manuscrit, au travers d'exemple variés, comment les Modes Statiques peuvent être utilisés pour explorer les propriétés intrinsèques de (bio)molécules. Ceci est rendu possible grâce à une "batterie" d'outils que nous développons au fur et à mesure des projets et des besoins, afin de **permettre la mise en place d'une véritable "ingénierie moléculaire *in silico*", comparable à une approche numérique de la mécanobiologie moléculaire.**

Les Modes Statiques peuvent ainsi être utilisés à différentes échelles, seuls ou combinés à d'autres méthodologies, pour caractériser les propriétés mécaniques des biomolécules et les relier à leur activité. Notre objectif est de développer des chemins de calculs prédictifs, pour comprendre et maîtriser la relation structure-fonction, en relation avec des applications dans les domaines de la santé et des biotechnologies, et accompagnant les procédés technologiques du laboratoire ainsi que les questions biologiques amenées par nos collaborations extérieures.

La figure 1.11 illustre quelques uns des outils les plus utilisés dans cet objectif et sur lesquels je reviendrai dans la suite de ce manuscrit.

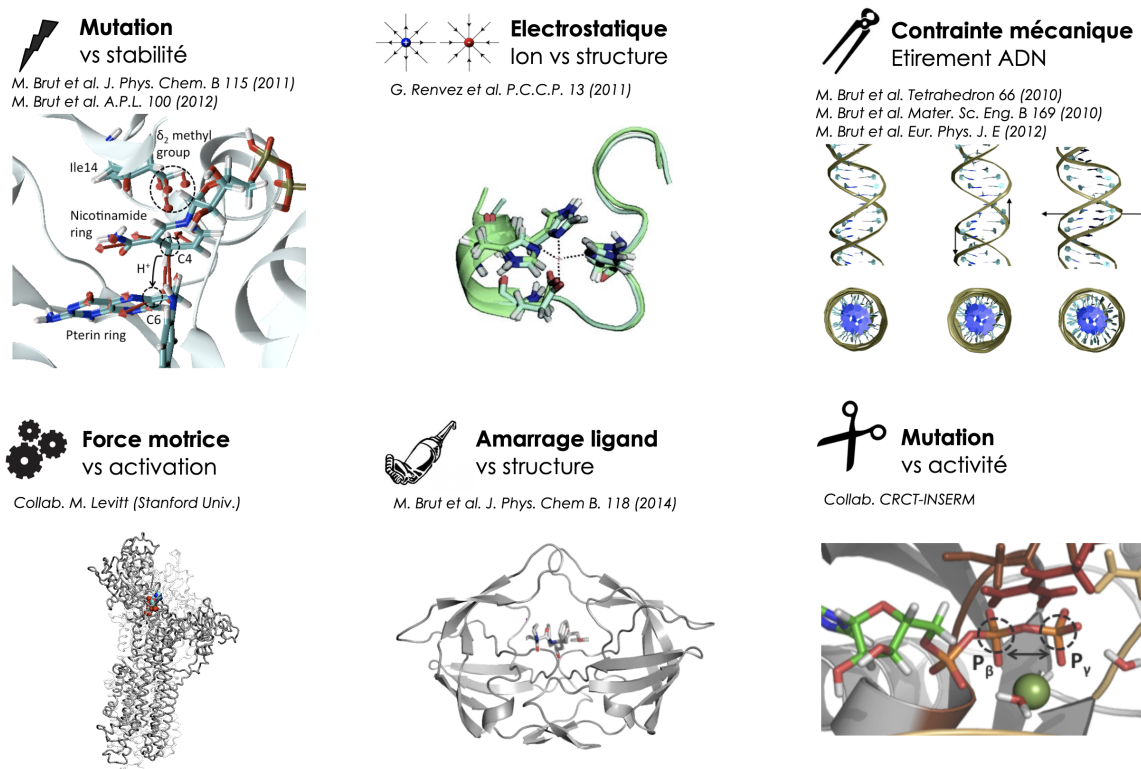


FIGURE 1.11 – Quelques exemples de la "boîte à outils" de traitement des Modes Statiques. Les MS peuvent être exploités comme approche *in silico* pour la mécanobiologie moléculaire et évaluer la réponse d'un système soumis à des contraintes internes (mutations) ou externes (interactions, contraintes mécaniques, effets de charges, etc).

1 Objectifs scientifiques

Mes activités de recherche adressent deux axes distincts : d'une part l'étude des protéines comme cibles thérapeutiques, et d'autre part, l'étude de molécules biologiques (principalement ADN/ARN, mais aussi protéines et membranes lipidiques de façon préliminaire) pour leur intégration dans des dispositifs technologiques.

Je présente dans ce chapitre le contexte et les objectifs de nos travaux ainsi que les méthodologies que nous utilisons et développons pour traiter ces problématiques. Je détaille ensuite certains de nos cas d'étude dont la ligne directrice vise des applications santé qui seront présentées.

1.1 Les biomolécules comme cibles thérapeutiques

Contexte

L'étude des protéines comme cibles thérapeutiques constitue ma principale activité. A l'heure actuelle, la recherche biomédicale se concentre majoritairement sur une petite proportion du génome humain qui a par ailleurs déjà été étudiée de manière intensive. L'objectif final est de concevoir une molécule médicament visant à modifier l'activité biologique d'une cible. Sa réussite repose sur le développement de plusieurs étapes :

- identifier une cible potentielle, ce qui implique d'assigner à un gène/une protéine une fonction donnée,
- décrypter à l'échelle atomique les mécanismes qui régissent sa relation structure-fonction et les mécanismes induisant un dysfonctionnement de cette relation,
- identifier une stratégie et un site "d'attaque", c'est-à-dire une cavité où lier une molécule médicament pouvant modifier la fonction,
- identifier une molécule médicament pertinente et efficace, répondant à des critères stricts [230].

Parmi ces étapes, nous ne travaillons pas sur la première, qui ne relève pas de nos domaines de connaissances et compétences. L'identification de nouvelles cibles réclame en effet de décrypter les informations portées et exprimées par le génome ou le protéome, mais aussi de les relier à des pathologies. L'identification de la cible (détermination d'une cible pertinente pour une pathologie donnée) et sa validation (vérification des effets de sa perturbation sur l'activité cellulaire et les marqueurs de la maladie) constituent un domaine de recherche à part entière [81, 167, 187, 233]. Des projets nationaux et internationaux mobilisent notamment la communauté autour de la collecte et du traitement systématique de données génomiques, protéomiques, chimiques et liées aux maladies pour identifier des cibles clés. Nous ne travaillons pas non plus sur la dernière étape, qui relève de la pharmacologie et consiste notamment à faire du criblage haut débit pour identifier des molécules capables de se lier à la cible et d'en modifier l'activité biologique [68].

Nous travaillons en revanche principalement sur la deuxième étape, qui peut nourrir la troisième, et qui consiste à **caractériser la cible en terme de structure, de flexibilité, de réactivité et de réseau d'interactions. L'objectif est d'une part d'améliorer la connaissance fondamentale de protéines d'intérêt, de comprendre pour quelles raisons et par quels mécanismes elles peuvent devenir des cibles thérapeutiques, mais aussi l'origine des résistances qu'elles peuvent présenter, occasionnées par l'apparition de mutations spécifiques.**

Nous nous intéressons en particulier au **rôle des mutations** que nous étudions selon trois directions. Tout d'abord, inclure une mutation "artificielle" et évaluer les effets qu'elle induit sur la structure et l'activité de la protéine peut nous renseigner sur l'importance et le **rôle du site muté**, en vue de comprendre le fonctionnement de la protéine (il s'agit du principe de la mutagenèse dirigée). Par ailleurs, inclure une mutation apparaissant sur une cible et modifiant/inhibant l'effet d'une molécule médicament doit permettre de comprendre les **mécanismes de résistance** et d'aider à proposer de nouvelles stratégies efficaces, même en présence de telles mutations. Enfin, inclure une mutation caractéristique d'une pathologie, évaluer le dysfonctionnement de la protéine mutée en comparaison au fonctionnement de la forme native (*i.e. biologiquement active*), est crucial pour relier la réponse médicamenteuse à la variation génétique. C'est le principe des **thérapies ciblées**, dont l'essor est accéléré par l'explosion de l'acquisition de données génomiques.

Parmi les cas traités, je détaillerai en particulier le cas de l'oncoprotéine Ras sur lequel nous travaillons depuis près de 10 ans. Cette protéine représente un enjeu particulier pour la communauté de part son rôle clé dans le développement de cancers, mais surtout parce que malgré d'intenses efforts menés depuis plusieurs décennies, aucune stratégie thérapeutique ne s'est réellement avérée efficace contre cette cible. De façon générale, malgré de grandes avancées expérimentales et théoriques, **la découverte de nouvelles stratégies thérapeutiques a ralenti et nécessite le développement de nouvelles approches pour faire face à des cibles considérées comme non "druggable"**. C'est le cas de Ras, et nous avons développé, au travers de cet exemple, un chemin de calcul prédictif qui peut être appliqué à d'autres cibles.

Les cibles thérapeutiques peuvent être classées en plusieurs catégories : la plupart sont des enzymes, des récepteurs, et des canaux ioniques ; en moindre proportion, on cible également des acides nucléiques, ou encore des cibles d'anticorps [193]. Nous travaillons principalement sur des enzymes, bien que nous ayons aussi travaillé sur un cas de récepteur (récepteur opiacé μ OR) et de canal ionique (Ca^{2+} ATPase). Ces cas, entre autres, seront développés dans la suite du manuscrit.

Méthodologie

Nous utilisons de façon systématique des simulations de type dynamique moléculaire et minimisation d'énergie pour rechercher des états d'équilibres en solution à partir de structures cristallines. L'intérêt est d'obtenir des conformations représentatives des états natifs, mais aussi des structures de départ comme prérequis pour le calcul de Modes Statiques, le cas échéant. Les calculs sont pour la plupart menés à l'aide du logiciel AMBER [51]. Différentes méthodes d'échantillonnage décrites en section 3.2 sont utilisées, telle que l'*umbrella sampling* et la dynamique moléculaire accélérée, par exemple.

Nous utilisons également des calculs DFT, en particulier pour la paramétrisation des ligands, mais surtout des calculs hybrides QM/MM, pour accéder à la simulation de processus chimiques au sein des protéines, ou encore pour accéder à une meilleure représentation de certaines caractéristiques du système, tels que la distribution de charges ou de molécules d'eau au sein des sites actifs.

Parmi les méthodologies utilisées de façon récurrente, nous employons la méthode des Modes Statiques comme ligne conductrice sur presque tous les systèmes étudiés, et ce pour adresser deux objectifs : i) soit pour caractériser le système et cartographier des sites d'intérêt déterminés pour leur rôle dans une réponse spécifique induite (*i.e.* afin d'identifier de nouveaux sites et mécanismes permettant de piloter des fonctions moléculaires) ; ii) soit pour soumettre le système à des forces physiques telles qu'elles pourraient être rencontrées en conditions expérimentales afin d'évaluer sa réponse).

Le détail des méthodologies appliquées sera donné pour chaque système décrit dans ce manuscrit.

1.2 Les biomolécules pour l'intégration de fonctions actives

Contexte

Au croisement des nanosciences et de la biologie, **les bionanotechnologies utilisent des molécules biologiques comme matériau et agent structurant. L'objectif est de les intégrer dans des nano-dispositifs** pour couvrir un large champ d'applications : technologies sur puces, nanobiocapteurs, nanodiagnostic, vectorisation intelligente de médicaments, constructions nanostructurées et fonctionnelles, etc [94, 156, 180]. Cette intégration passe par le besoin de **manipuler, contrôler, voire modifier ces systèmes complexes fournis par la nature**. Les protéines, d'une part, sont de véritables machines moléculaires, offrant des **fonctions spécifiques, sophistiquées et ajustables**. Quant aux acides nucléiques, leurs **propriétés d'auto-assemblage et d'hybridation** laissent imaginer à l'infini de nouvelles solutions, pour la nanostructuration et la détection par exemple. Enfin, les bicouches lipidiques qui peuvent s'auto-assembler en forme de vésicules en solution, peuvent servir de modèles pour l'étude des propriétés élastiques/mécaniques de cellules, de diffusion d'ions et de molécules au travers des membranes cellulaires, pour l'étude des interactions cellules/nanoparticules ou encore à des fins de vectorisation de médicaments.

Un exemple majeur est celui des nanotechnologies ADN qui ont rapidement évolué et trouvent aujourd'hui des applications en analyse, en thérapie et en nanofabrication [70, 246]. Initiées par Seeman en 1982 [196], elles utilisent les propriétés de reconnaissance programmables de l'ADN, et ont notamment donné naissance à l'origami ADN pour la construction de nano-objets [197]. Elles ont également permis, parmi d'autres exemples, de créer de nouvelles générations de capteurs à l'aide de courts brins d'ADN/ARN programmés ou sélectionnés pour se lier de façon spécifique à une cible [181].

Cependant, malgré un nombre de publications croissant, ces technologies sont encore sous-représentées dans le paysage industriel. Ce constat peut notamment être expliqué par le fait que le comportement des biomolécules est fortement dépendant de leur environnement. Or une grande partie du travail dans ce domaine est expérimental, les chercheurs ne connaissent souvent les performances d'un composant qu'après l'avoir créé et testé.

Pour développer et exploiter pleinement ces nouvelles technologies jusqu'à leur déploiement industriel, nous devons maîtriser leurs propriétés et comprendre leur origine aux échelles atomique et moléculaire. Cela nécessite un effort théorique important, plutôt que de systématiser l'approche expérimentale classique essai/erreur. La modélisation est alors incontournable pour décrypter et prédire les propriétés des biomolécules dans un contexte et un environnement nouveaux, afin d'assister et de guider l'expérience. L'enjeu repose sur leur intégration et le contrôle de leur relation structure/fonction pour des applications bio-puces et la nanostructuration de matériaux.

Méthodologie

Au-delà de la complexité intrinsèque du traitement des biomolécules, une nouvelle difficulté repose sur la nécessaire mise en place d'une modélisation permettant de prendre en compte leur intégration dans des milieux non biologiques, ou en interaction (parfois covalente) avec des objets non biologiques, voire inorganiques. Dans ce cadre, nous avons en particulier travaillé sur des applications aux technologies ADN (brins d'ADN/ARN utilisés comme capteurs) et sur l'activation de mouvements fonctionnels de grande amplitude avec le cas des moteurs moléculaires.

Ici encore, nous nous servons de la dynamique moléculaire pour accéder à diverses propriétés structurales et thermodynamiques. Dans les cas que je présenterai sur l'utilisation d'acides nucléiques comme capteurs, nous avons notamment évalué des énergies d'interactions capteur/cible pour évaluer l'affinité et la spécificité des interactions.

Nous utilisons aussi les Modes Statiques pour appliquer des forces locales et évaluer par exemple la stabilité des repliements. Nous avons par ailleurs développé des algorithmes de prédiction de séquence d'acides nucléiques par rapport à des cahiers des charges fixés par des chercheurs du LAAS qui ont motivés ces travaux, et "calibré" les séquences obtenues en fonction de températures de fusion obtenues à l'aide de modèles empiriques intégrés à nos calculs.

Enfin, l'approche par Modes Statiques a été adaptée à des modèles de types gros-grains pour l'étude de l'activation de mouvements de grande amplitude au sein de moteurs moléculaires. Ce type de calcul est de nouveau adapté pour évaluer la réponse mécanique de membranes biologiques modèles.

Ces outils, utilisés, et pour la plupart développés au sein de notre équipe, seront présentés dans la partie 3.1. Implémentés pour un besoin donné, ils peuvent bien sûr être adaptés à d'autres systèmes, d'autres questions, et viennent s'ajouter à notre "boîte à outils" *in silico*. Ils ont pour but d'aider les technologues à maîtriser l'intégration de biomolécules dans de nouvelles générations de dispositifs. A terme, ils devraient être intégrés dans une stratégie plus large de calcul multi-échelles pour se rapprocher au mieux des conditions et des problématiques expérimentales.

2 De la compréhension des mécanismes fondamentaux aux cibles thérapeutiques

Le premier cas de protéine que j'ai traité est celui de la protéase du VIH-1, durant ma thèse. L'intérêt pharmaceutique de cette enzyme en a fait une cible potentielle dès le début des thérapies antivirales dans les années 1990, au point qu'elle a constitué le premier exemple de réussite de conception de médicament par une approche rationnelle [240]. Ses données structurales et fonctionnelles sont largement décrites dans la littérature et ont permis de confronter les résultats obtenus avec les Modes Statiques afin de valider la pertinence et le caractère prédictif de notre approche sur un premier exemple de macromolécule. Grâce à un calcul principal, nous avons retrouvé un nombre important de propriétés identifiées au cours des deux décennies précédant ces travaux, à la fois par l'expérience et le calcul. Nos résultats ont ainsi permis de "cartographier" les résidus impliqués dans les mouvements fonctionnels et la stabilité du site actif, d'identifier des sites allostériques, des corrélations/communications entre domaines, ainsi que des mouvements de boucles induits par la simulation de l'arrimage d'un ligand [43]. Ce dernier point en particulier a permis de confirmer la compatibilité de la méthode avec une approche de docking flexible, ce pour quoi elle a été initialement conçue.

Ainsi, au travers de ce premier cas, j'ai jeté les bases de ce qui constitue aujourd'hui encore notre approche générale pour l'étude des cibles thérapeutiques : en nous appuyant sur un pré-calcul de Modes Statiques stockés dans une base de données, nous identifions des sites stratégiques jouant un rôle dans la stabilité et la fonction, et anticipons des réponses d'intérêt en appliquant des stimuli externes, simples ou multi-sites. Selon les problématiques abordées, ces calculs sont complétés par des calculs de dynamique moléculaire pour explorer différentes conformations et évaluer des grandeurs thermodynamiques, ou encore des calculs hybrides QM/MM pour étudier en détail le site actif et les réactions chimiques dont il est le siège.

Je présente dans la suite trois systèmes sur lesquels nous avons travaillé ces dernières années. Le premier est celui de la dihydrofolate réductase, une cible pour divers traitements (chimiothérapies, malaria, etc) sur laquelle nous n'avons travaillé qu'à l'aide des Modes Statiques et avons commencé à nous intéresser aux effets de mutations. Le second, celui de l'oncoprotéine Ras, est celui qui sera le plus largement développé car il s'agit d'un projet de long terme mené en parallèle avec des collègues biologistes, et sur lequel nous avons utilisé une approche multi-méthodes particulièrement complète. Nous visons ici à comprendre les mécanismes par lesquels des mutations peuvent induire une diminution de l'activité enzymatique, laquelle peut être restaurée par l'ajout de nouvelles mutations que nous avons prédites et qui peuvent déboucher sur de nouvelles stratégies thérapeutiques. Le dernier cas porte sur la protéine cGAS, un déclencheur de la réponse immunitaire qui pourrait constituer une nouvelle cible anti-cancer. Ce dernier système est abordé uniquement par dynamique moléculaire, classique et QM/MM, afin de comprendre ses mécanismes d'activation par l'ADN, ainsi que les réactions qu'il catalyse. **Nos travaux sur ces trois systèmes sont décrits de manière indépendante dans les paragraphes suivants et montrent comment nous intégrons chronologiquement de nouvelles méthodologies.**

2.1 Exploration de la relation structure - activité par une approche mécanistique

Je décris ici les cas traités sous un angle purement structural et mécanistique. Afin d'étudier la relation entre mouvements fonctionnels et activité biologique, nous avons utilisé des simulations de dynamique moléculaire classique, mais surtout l'approche des Modes Statiques afin de caractériser les sites actifs via leur réponse à des contraintes imposées au système. Les cas décrits dans cette partie ont permis d'asseoir le caractère prédictif des Modes Statiques et la pertinence de les utiliser selon différentes approches, adaptables aux systèmes et aux questions qui nous sont posées.

A. Relation entre flexibilité et stabilité du site actif d'une enzyme : cas de la dihydrofolate réductase de *E. coli*

Cadre de travail : Ce travail mené hors cadre contractuel a été initié au travers de discussions avec F. Nepveu (UMR 152 IRD, Toulouse) et a fait l'objet d'une publication [39]. Il a été suivi par la thèse de M. du Manoir, co-encadrée avec A. Berry (CPTP-U1043) dans le cadre d'un financement du labex ParaFrap (étude des mécanismes de résistance de *Pf*DHPS aux traitements anti-paludiques).

Résumé

Les Modes Statiques sont utilisés pour évaluer les propriétés biomécaniques de l'enzyme DHFR, en particulier pour caractériser les mouvements de boucles et les résidus importants pour son activité. **Nous étudions ici la relation entre la structure/flexibilité du site actif et la stabilité du complexe folate/cofacteur, nécessaire à l'activité catalytique de l'enzyme.** Nous montrons que les mouvements du backbone sont particulièrement importants pour assurer la communication entre domaines et soulignons notamment le rôle de l'hélice αa dans le mouvement de la boucle Met20. Nous étudions pour cela la flexibilité du site actif dans différents états liés et montrons que la boucle Met20 est rigidifiée en état fermé par de nouvelles liaisons faibles intra et intermoléculaires qui stabilisent le complexe et favorisent le transfert d'hydrure. Enfin, si diverses simulations, dont celles incluant les mutations I14V et I14A, confirment que I14 est un résidu clé pour l'activité catalytique, nous isolons et caractérisons à l'échelle atomique quel groupement spécifique permet d'assister le positionnement du ligand afin de diriger le cycle nicotinamide vers le cycle ptérine.

Contexte

La dihydrofolate réductase (DHFR) se retrouve dans la plupart des organismes avec une forte homologie structurale. En présence du cofacteur NADPH¹, elle catalyse la réduction du 7,8-dihydrofolate (DHF) en 5,6,7,8-tétrahydrofolate (THF). Son rôle est essentiel pour la synthèse des purines et de la thymidine, et par conséquent, pour la division cellulaire. Son inhibition entraîne la perturbation de la synthèse d'ADN conduisant à la mort cellulaire. Pour cette raison, cette enzyme a été largement étudiée à des fins thérapeutiques et constitue une cible pour les chimiothérapies, ou encore les traitements contre le paludisme ou la tuberculose [179]. Elle est largement étudiée comme **système modèle** pour explorer la relation entre changement conformationnel et activité catalytique [237].

La **structure** de la DHFR consiste en deux sous-domaines entre lesquels se trouve le site actif. Le principal contient trois boucles flexibles jouant un rôle crucial dans la liaison du ligand, la catalyse et la libération du produit. Il s'agit des boucles Met20 (résidus 9-24), F-G (116-132) et G-H (142-150), présentées sur la figure 2.1.

1. NADPH est la forme réduite du NADP (nicotinamide adénine dinucléotide phosphate).

Le **cycle catalytique** de cette enzyme comprend les étapes suivantes : liaison du cofacteur NADPH, liaison du dihydrofolate, transfert d'hydrure du NADPH au dihydrofolate et protonation, et libération du tétrahydrofolate et du cofacteur.

Au cours de ce cycle, deux conformations de la partie centrale de la boucle Met20 (résidus 16-19) ont été observées en solution, en fonction de l'état lié au ligand (voir Fig. 2.1 b).

Lorsque seul le folate (FOL) est lié, la boucle Met20 se trouve dans une conformation dite *occluse*, stabilisée par la formation de liaisons hydrogène entre le résidu N23 de Met20 et le résidu S148 de la boucle G-H.

Quand le cofacteur est également lié, l'enzyme se trouve dans une conformation dite *fermée* : les liaisons hydrogène entre N23 et S148 se rompent tandis que de nouvelles liaisons se forment entre G15/G17 de la boucle Met20 et N122 de la boucle F-G. La boucle Met20 recouvre alors le cofacteur, tandis que le cycle nicotinamide stabilise le complexe et le protège du solvant.

C'est ce mouvement de la boucle Met20, entre conformation fermée et occluse, que nous souhaitons étudier.

Pour cela, nous évaluons et caractérisons la propension de la boucle Met20 à subir des changements conformationnels en présence et en absence du cycle nicotinamide.

Méthodologie

Deux conformations de la DHFR d'*E. Coli* ont été utilisées dans ce travail, en référence aux précédentes études théoriques qui ont porté sur les changements de flexibilité de l'enzyme lors de la liaison au cofacteur [171,195]. Parmi les étapes du cycle de l'enzyme, nous avons choisi de traiter des changements conformationnels qui se produisent pendant la transition du complexe binaire E: FOL au complexe ternaire E: FOL: NADPH. Nous avons pour cela utilisé les structures cristallines 1RX7 et 1RX2 extraites de la PDB, lesquelles correspondent respectivement à la conformation occluse (DHFR liée au substrat), et à la conformation fermée (DHFR liée au substrat et au cofacteur). Chaque structure a d'abord été soumise à un calcul de minimisation d'énergie. Nous avons pour cela utilisé le logiciel AMBER [52] avec le champ de force ff99SB [92] pour générer les conformations relaxées et calculer la matrice Hessienne en solvant implicite.

Les **Modes Statiques** ont ensuite été calculés à l'aide du logiciel FleXible et utilisés pour identifier les atomes impliqués dans le mouvement de la boucle Met20 et la stabilité du complexe. Pour cela, nous avons utilisé une procédure d'**optimisation sous contrainte** que je présente rapidement ici et sur laquelle je reviendrai plusieurs fois dans la suite de ce manuscrit. Cette procédure consiste à optimiser les contraintes à appliquer sur chaque atome et permettant de maximiser une réponse spécifique du système. Ici, la réponse est géométrique et évaluée au travers de distances interatomiques d'intérêt au sein du complexe protéine/ligands. Pour cela, tout en conservant l'amplitude de la force appliquée sur un atome N_0 , on optimise sa direction pour obtenir la variation maximale de distance entre deux atomes N_1 et N_2 . Deux matrices (3×3), m_{N_1} et m_{N_2} , sont tout d'abord extraites des Modes Statiques, contenant les déplacements de N_1 et N_2 , résultant de l'application de trois forces dans les directions x, y, z sur l'atome N_0 . La variation de la distance $N_1 N_2$ est alors associée à la matrice $m = m_{N_1} - m_{N_2}$, et sa valeur maximale est donnée par la plus grande valeur propre de la matrice $m^t m$, déterminée par la méthode de puissance itérée.

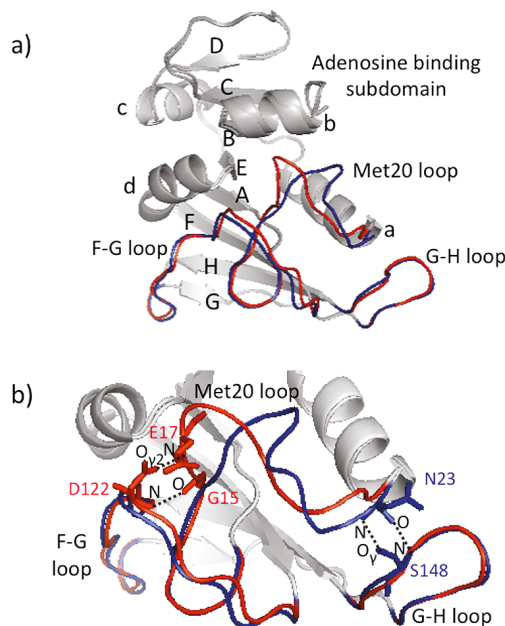


FIGURE 2.1 – Superposition de deux conformations de la DHFR avec la boucle Met20 à l'état fermé (en rouge, PDB 1RX2) et à l'état occlus (en bleu, PDB 1RX7). a) est une vue d'ensemble de l'enzyme. Les hélices α sont notées de a à d et les feuilletts β de A à H. b) est une vue rapprochée de la boucle Met20 et de ses interactions avec les boucles F-G et G-H.

Résultats

Mouvements de boucle. Des études antérieures, notamment basées sur les valeurs du facteur β , ont suggéré que la boucle Met20 est plus flexible lorsque seul le folate est lié. Pour vérifier que les Modes Statiques nous permettent d'accéder à ce type d'information, nous avons successivement optimisé les contraintes à appliquer sur tous les atomes de la molécule dans les conformations fermées et occluses pour maximiser les variations de distances entre les groupements représentés sur la figure 2.1 b) : CO(G15) - NH(D122), NH(E17) - CO γ 2(D122), CO(N23) - NH(S148), et NH(N23) - CO γ (S148).

Distance E17-D122. Nous partons de la structure occluse et appliquons la procédure d'optimisation sous contrainte pour déterminer quelle force induit la variation maximale de la distance E17-D122 (calculée entre les atomes C α), ce qui constitue un premier critère d'évaluation de la fermeture de la boucle. Le résultat (Fig. 2.2) montre qu'appliquer une contrainte sur les atomes C α des résidus E17 et D122 induit les plus grandes variations, ce qui est attendu, mais aussi que d'autres atomes C α des boucles Met20 et F-G, appartenant respectivement aux résidus R12-N18 et D116-P126 sont également impliqués dans ce mécanisme. En outre, les atomes C α des résidus W30-F31 (hélice α a), H45-T46 et S49-I50 (hélice α b), et I94 (feuillet β E), sont également impliqués dans le mouvement de boucle Met20. Ces résultats sont particulièrement intéressants car **tous ces résidus ont été identifiés pour leur rôle dans la transition occluse/fermée et sont par ailleurs fortement conservés** [9, 54].

Si nous considérons maintenant la contribution moyenne des atomes par résidu (Fig. 2.2 b), il apparaît que seuls G15-M20 (boucle Met20) et E120-F125 (boucle F-G) contribuent vraiment au mouvement de boucle. La déformation se propage facilement via les atomes du backbone, tandis que seules les chaînes latérales des résidus les plus proches permettent la transmission. Ce résultat n'est pas surprenant puisque, dans la conformation occluse, de nombreuses liaisons faibles entre chaînes latérales sont perturbées et ne peuvent propager la déformation. Ces représentations nous permettent de localiser les résidus induisant la réponse demandée, mais le contenu des résultats du calcul d'optimisation nous permet d'atteindre un niveau de détails supplémentaire. Sur la figure 2.2 c), sont représentées les forces optimisées, calculées sur les résidus E17 et D122. Leur direction est une **indication précise de la façon dont les atomes doivent être contraints** pour obtenir le mouvement de boucle souhaité, tandis que leur norme est proportionnelle à l'amplitude de la déformation induite. En d'autres termes, plus la flèche est longue, plus il est facile de minimiser la distance E17-D122. Cette représentation permet notamment une lecture aisée de la contribution de chaque atome. Par exemple, il est remarquable d'observer que la manière optimale de déplacer les groupements NH(E17) et CO γ 2(D122) n'est pas de contraindre ces atomes eux-mêmes, mais les atomes C γ et C β de E17 et D122, respectivement.

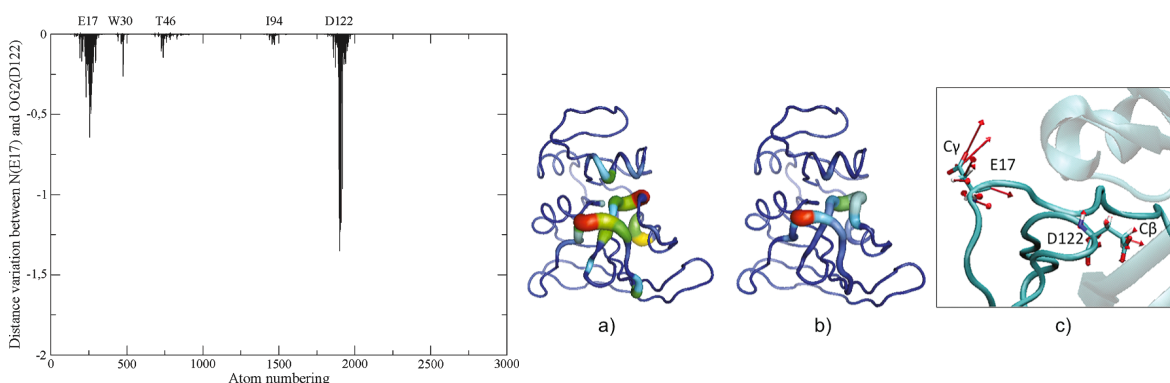


FIGURE 2.2 – A gauche : variations de distance entre les groupements NH(E17) et C γ O2(D122) induits par une contrainte optimisée sur chaque atome du système DHFR/folate en configuration occluse (en Å). A droite : localisation des résidus correspondants (du bleu au rouge par amplitude croissante). a) Les couleurs sont attribuées aux résidus selon les contributions des atomes C α . b) Les couleurs sont attribuées par rapport à la contribution moyenne des atomes par résidu. c) Représentation des directions des forces appliquées sur E17 et D122, leurs normes sont proportionnelles à la variation de distance induite.

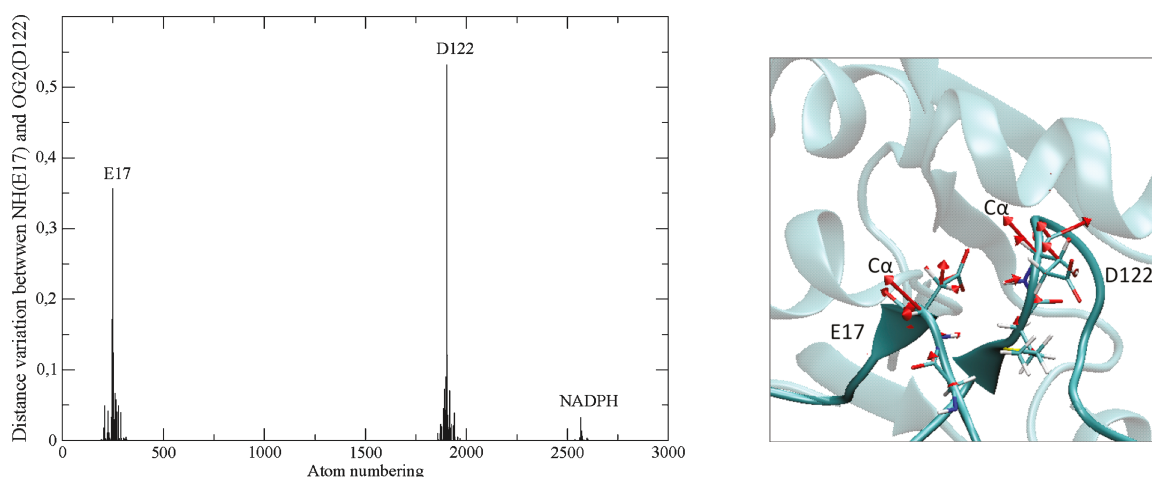


FIGURE 2.3 – À gauche : variations de distance entre les groupements NH(E17) et $C\gamma O_2(D122)$ induits par une contrainte optimisée sur chaque atome du système DHFR/folate/NADPH en configuration fermée (en Å). À droite : représentation des forces appliquées sur E17 et D122 pour maximiser l'élongation de la distance NH(E17) - $C\gamma O_2(D122)$.

La même exploration a été menée sur la conformation fermée, lorsque le cofacteur NADPH est également lié. On cherche donc maintenant à éloigner les groupements NH(E17) et $C\gamma O_2(D122)$. On observe sur la figure 2.3 que les pics correspondant à E17 et D122 sont toujours visibles, mais avec une amplitude plus faible. Tous les autres ont disparu par rapport aux résultats obtenus sur configuration occluse. Cela signifie que la molécule est beaucoup plus rigide en présence du cofacteur. L'ensemble des liaisons faibles formées entre des résidus de la boucle F-G et les ligands dans la poche de liaison stabilise la conformation de la boucle Met20, ce qui rend difficile d'étirer de la liaison E17 - D122. Cette conclusion corrobore les résultats d'un travail antérieur mené par spectroscopie de force sur molécule unique [9], montrant que **la liaison au ligand augmente la stabilité** de la DHFR, grâce au réseau d'interactions protéine-ligand. On relève également un pic correspondant au cofacteur NADPH, et plus précisément aux atomes du sucre lié au cycle nicotinamide, lequel est très proche des résidus E17 et D122, ce qui confirme et détaille l'implication de la liaison du cofacteur dans la stabilisation de boucle Met20 à l'état fermé. Enfin, comme le montre la partie droite de la figure 2.3, contrairement à ce que nous avons constaté pour l'état occlus, la manière optimale de déplacer NH(E17) et $CO\gamma_2(D122)$ est maintenant de contraindre les atomes $C\alpha$, les chaînes latérales étant stabilisées par de nouvelles interactions, et il n'est pas favorable de contraindre les atomes qu'elles impliquent. Cela confirme encore la **rigification de la DHFR à l'état fermé**.

Distance G15-D122. Les mêmes calculs ont été menés sur les variations de distance $CO(G15) - NH(D122)$ (voir Fig. 2.4). Les détails ne seront pas donnés ici mais sont décrits dans l'article [39] et permettent d'aboutir à des conclusions similaires à celles obtenues sur la distance E17-D122. Partant de la conformation occluse, nous obtenons de nouveau deux pics de plus grande amplitude correspondant aux atomes des résidus G15 et D122. D'autres variations significatives sont induites par des contraintes appliquées sur les résidus suivants V13-M16 et M20 (Met20), S49 (hélice αb), D116-A117 et G121-T123 (boucle F-G). Tous ces résidus étaient déjà impliqués dans les variations de distance E17-D122, ce qui est attendu puisque E17-D122 et G15-D122 sont interconnectés et localisés à proximité les uns des autres. Enfin, les liaisons hydrogène existant entre N18 et S49 nous permettent de comprendre comment S49 est impliqué ici. Les forces qui entraînent S49 (hélice βb) sont mises en évidence sur la partie droite de la figure 2.4.

Nous avons enfin effectué les mêmes calculs pour DHFR en état fermé. Les résultats sont présentés au centre de la figure 2.4 et vont dans le sens de la rigidification du système. Seuls les pics correspondants aux résidus G15 et D122 sont conservés. La contribution du NADPH apparaît également, indiquant son rôle dans la stabilisation globale du complexe. Nous concluons de nouveau qu'**il est difficile de conduire un mouvement de boucle via les stimuli externes en conformation fermée**.

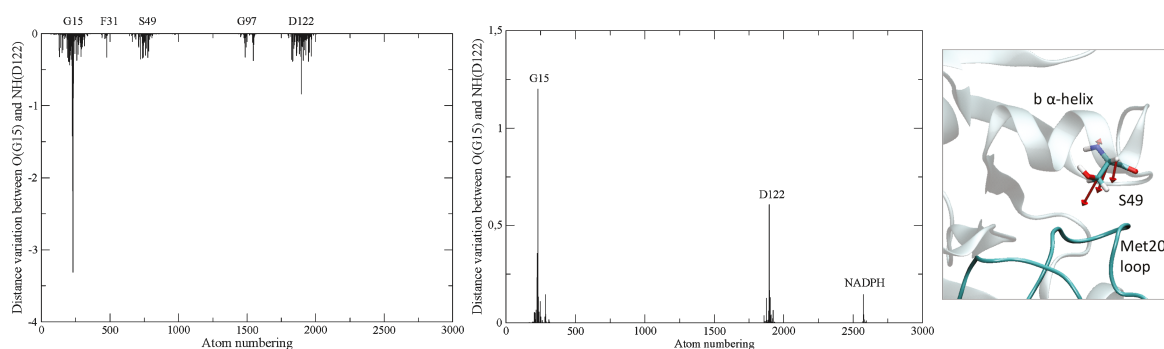


FIGURE 2.4 – A gauche : variations de distance entre les groupements CO (G15) - NH (D122) induits par une contrainte optimisée sur chaque atome du système DHFR/folate en configuration occluse (en Å). Au centre : résultats du même calcul sur le système DHFR/folate/NADPH en configuration fermée. A droite : représentation des forces appliquées sur S49 pour optimiser le passage de la forme occluse vers la forme fermée.

Distance N23-S148. De nouveau, nous partons de l'état occlus (DHFR/folate) et optimisons les forces permettant de minimiser la distance séparant les résidus N23 et S148. Les résultats ne sont pas présentés en raison de leurs similitudes avec les caractéristiques trouvées ci-dessus : les résidus qui induisent les plus grandes déformations sont N23 et S148 eux-mêmes, ainsi que les résidus voisins des boucles Met20 et G-H (A19-L24 et N147-S150, respectivement). Ici encore, certains résidus de l'hélice αa (A29-F31) et de la boucle F-G (H114-I115) sont également impliqués. Ce résultat est attendu puisque la boucle G-H ne subit pas de changement significatif. En mesurant la distance N23-148, nous évaluons comment la boucle Met20 se rapproche ou s'éloigne de la boucle G-H. **Il est donc pertinent d'identifier certains résidus de la boucle F-G qui poussent la boucle Met20 vers la boucle G-H, ainsi qu'une partie de l'hélice αa , dont le mouvement est corrélé à celui de la boucle Met20.**

Stabilité du site actif. En utilisant la même approche, nous nous sommes également concentrés sur le positionnement des ligands au sein du site actif, et avons exploré la légitimité d'une approche mécanique pour étudier la propension de l'enzyme à réaliser l'activité catalytique.

Distance C_4 (NADPH)- C_6 (DHF). Une des étapes du cycle enzymatique réalisé par la DHFR passe par le transfert d'hydrure d'un atome donneur appartenant au cycle nicotinamide (C_4) vers un atome accepteur du cycle ptérine du folate (C_6). L'environnement des deux ligands (*i.e.* DHFR) catalyse ce mécanisme de sorte que la perturbation de la position des ligands entraîne une perte d'activité enzymatique [213]. Nous avons utilisé le même protocole d'optimisation que précédemment pour évaluer le rôle des résidus du site actif sur le positionnement des ligands. La partie gauche de la figure 2.5 montre les variations de distance entre les atomes C_4 et C_6 imposées par des contraintes appliquées sur l'ensemble des atomes du système. Une représentation logarithmique est utilisée pour faciliter la lisibilité. Les valeurs sont en effet très faibles, conformément aux conclusions précédentes sur l'importante stabilité du site actif en conformation fermée. Les plus grands pics correspondent au cofacteur NADPH et au substrat DHF, principalement aux atomes des cycles nicotinamide et ptérine, comme le montre la figure 2.5 (partie droite). Cependant, un pic apparaît de façon remarquable sur la gauche, correspondant aux atomes du résidu I14, et plus exactement au groupe méthyle $\delta 2$. Le rôle de ce résidu a déjà été mis en évidence dans des études antérieures [9, 195, 239], ou depuis, dans de nouvelles études [149], mais pas avec le niveau de détails obtenus ici, et par ailleurs, à l'aide de dynamiques moléculaires classiques ou QM/MM beaucoup plus coûteuses. **Les Modes Statiques sont donc également capables d'extraire efficacement ce type d'information.**

En l'occurrence, les atomes de la chaîne latérale de I14 établissent un réseau de liaisons hydrogène en se plaquant contre le cycle nicotinamide du cofacteur. Cette connexion rigide dirige cette région du NADPH vers le cycle ptérine du folate, comme illustré sur la figure 2.5. On peut nettement observer comment les forces motrices suivent la direction de l'alignement I14($\delta 2$)- C_4 (NADPH)- C_6 (DHF), perpendiculairement au plan des cycles nicotinamide et ptérine, et de quelle façon elles peuvent ainsi jouer sur la distance donneur-accepteur.

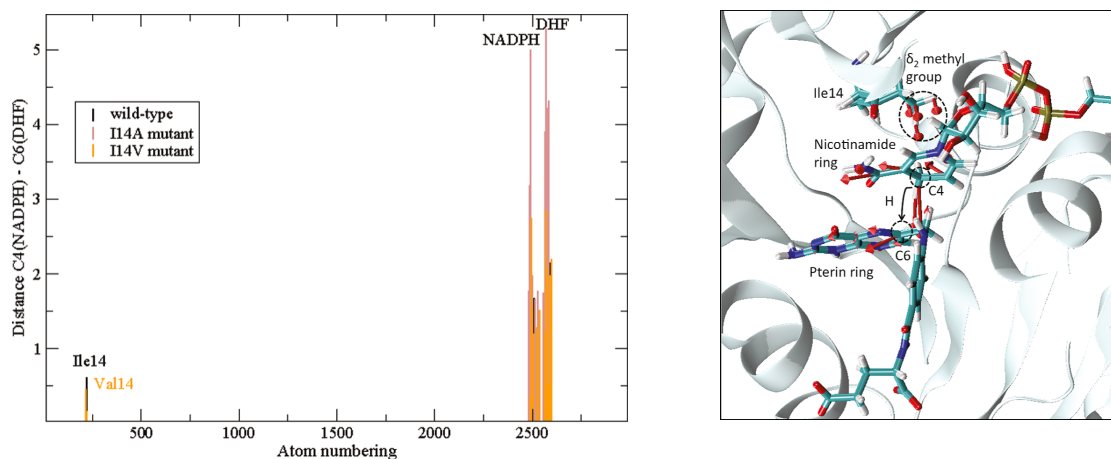


FIGURE 2.5 – A gauche : variations de distance entre les atomes C4(NADPH) et C6(DHF) induites par une contrainte optimisée sur chaque atome du système DHFR/DHF/NADPH en configuration fermée (en représentation logarithmique pour une meilleure lisibilité). Les calculs ont été menés sur la forme native (noir), ainsi qu'en incluant les mutations I14A (en rose) et I14V (en jaune). A droite : représentation des forces appliquées sur le résidu I14, le cofacteur et le substrat pour minimiser la distance C4(NADPH)-C6(DHF).

Pour éclairer le rôle joué par le groupe méthyle de I14, ce résidu a été successivement remplacé par une valine (I14V) et une alanine (I14A) (à partir de la structure 1RX2 modifiée *in silico*). La mutation I14A possède une chaîne latérale plus courte, consistant en un simple groupe méthyle. En conséquence, I14A ne devrait plus pouvoir former de liaisons avec le cycle du nicotinamide puisque le groupement méthyle est bien présent, mais trop éloigné. Après réitération de la procédure d'optimisation sur la structure équilibrée, nous pouvons comparer le résultat obtenu sur DHFR I14A par rapport aux résultats obtenus sur la forme native. Les deux pics correspondant aux cycles nicotinamide et ptérine sont toujours présents, mais avec des amplitudes beaucoup plus importantes, atteignant 2Å. De plus, le pic au niveau du résidu 14 a disparu. La chaîne latérale de A14 est trop courte pour plaquer le cycle nicotinamide contre la ptérine. Par ailleurs, le site actif devient nettement plus flexible, les ligands sont davantage libres de bouger, le réseau d'interactions créées entre le résidu 14 et son environnement étant perdu. Cela suggère une perte probable d'activité concernant le transfert d'hydrure, perte qui a depuis été confirmée par des calculs de chemins réactionnels [72]. Ce résultat corrobore le rôle important joué par le résidu I14 dans la stabilisation du complexe et est cohérent avec d'autres études, expérimentales et théoriques [72, 149, 212, 213].

Afin d'aller plus loin dans notre proposition, le même travail a été mené en incluant le mutant I14V. La chaîne latérale de la valine est en effet similaire à celle de l'isoleucine à laquelle on retire le groupe méthyle δ_2 . Ce dernier calcul doit donc permettre de vérifier notre conclusion sur le rôle spécifique de ce groupement. On retrouve de nouveau les deux pics, correspondant respectivement aux atomes de NADPH et de DHF, avec des amplitudes autour de 0,2Å, soit dix fois plus faibles qu'en présence de I14A, mais trois fois plus grandes que celles de la forme non mutée (type sauvage). La longueur de la chaîne latérale est donc indispensable pour la stabilisation du complexe. Par ailleurs, un troisième pic est visible, correspondant au groupe CO (backbone) de I14V, lequel est orienté vers le cofacteur NADPH. Ce groupe établit une liaison hydrogène avec le groupe amide du NADPH et contribue à moduler la position du cycle nicotinamide. Pourtant, les amplitudes indiquent que l'effet est inférieur à celui de I14.

La comparaison de l'enzyme de type sauvage et des mutants I14A et I14V nous permet de caractériser le rôle de la chaîne latérale Ile14 dans le déplacement du cycle nicotinamide. Sa longueur est essentielle pour créer une contrainte spatiale qui optimise la configuration du site actif. De plus, il joue un rôle clé dans stabilisation complexe, nécessaire pour favoriser le transfert d'hydrure du donneur à l'accepteur. Il est enfin à noter que ce résidu est hautement conservé dans DHFR.

Conclusions et perspectives

La méthode des Modes Statiques est appliquée ici pour traiter une question mécanique générale au fonctionnement des enzymes, à savoir le rôle des mouvements de boucles dans l'activité catalytique. En nous intéressant au système modèle du transfert d'hydrure dans la DHFR d'*E. coli*, nous nous sommes attachés à montrer comment les Modes Statiques, par une approche mécanique et grâce au choix de critères pertinents (ici le suivi des distances interatomiques caractéristiques de l'activité biologique), peuvent nous permettre d'accéder à l'identification de résidus et de mécanismes clés comme indicateurs de l'activité chimique. Nos résultats ont ainsi permis de souligner le rôle de résidus hautement conservés chez DHFR, mais aussi des modes de communication entre domaines intervenant dans la régulation de l'activité, tout en accédant à un niveau de détails à l'échelle atomique.

Nous nous sommes particulièrement intéressés à la flexibilité et à la stabilité du site actif, ainsi qu'aux conséquences de l'introduction de mutations spécifiques sur des sites d'intérêt avec deux objectifs : i) confirmer le rôle prépondérant de résidus spécifiques en modifiant leur nature, ii) modifier l'activité enzymatique en choisissant la mutation de façon pertinente.

La méthode des Modes Statiques peut donc se positionner comme une alternative à la dynamique moléculaire, mais aussi à des expériences de mutagenèse *in vitro*, toutes beaucoup plus coûteuses. Ce travail nous a donc engagé à poursuivre les applications à visées thérapeutiques, en particulier en utilisant les informations tirées des expériences de mutations *in silico* pour explorer les mécanismes de résistance thérapeutique, mais aussi comme soutien à des stratégies de conception *de novo*.

C'est en suivant cette trajectoire que nous nous sommes engagés dans l'étude de l'oncoprotéine Ras sur laquelle je reviens de façon détaillée dans la partie suivante.

C'est aussi à la suite de ce travail qu'a débuté une collaboration avec le Pr A. Berry, dans l'idée d'appliquer le même chemin de calcul à une seconde enzyme qui intervient en amont dans la voie des folates et constitue une cible, conjointement avec DHFR, dans des traitements anti-paludiques visant un des parasites vecteurs de la maladie, *Plasmodium falciparum*.

B. Une ouverture vers la compréhension des mécanismes de résistance à la sulfadoxine-pyriméthamine chez *Plasmodium falciparum*

Contexte

La voie des folates chez *P. falciparum* a été intensivement étudiée [100] et a permis de proposer des traitements visant l'arrêt de la synthèse des purines de l'ADN dans les parasites afin de limiter la charge infectieuse, voire de détruire des germes. Les folates sont synthétisés *de novo* chez la plupart des micro-organismes, mais certains, comme *P. falciparum*, peuvent également nécessiter un apport de formes intermédiaires, d'où la stratégie de cibler la voie des folates à deux niveaux : *PfDHPS* et *PfDHFR*. J'ai décrit le fonctionnement de DHFR dans la partie précédente. Quant à DHPS (dihydroptéroate synthase), elle catalyse la condensation de 6-hydroxyméthyl-7,8-dihydroptérine-pyrophosphate (DHPP) avec l'acide para-aminobenzoïque (pABA) pour former une intermédiaire des folates, le 7,8-dihydroptéroate. Ces deux enzymes sont respectivement ciblées par la pyriméthamine et la sulfadoxine (SP), lesquelles agissent comme inhibiteurs compétitifs. **Des mutations de *PfDHFR* et *PfDHPS*, apparues sous pression médicamenteuse, sont cependant à l'origine de souches résistantes : en modifiant la structure des enzymes, elles diminuent leur affinité à la SP.** Se pose alors la question des mécanismes par lesquels elles sont liées à la résistance : le sont-elles toutes ? Quel niveau de résistance est associé à chaque mutation ou combinaison de mutations ?

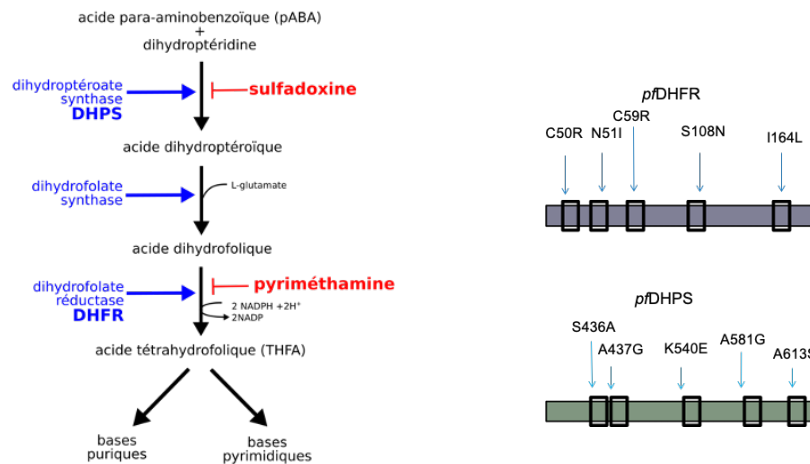


FIGURE 2.6 – À gauche est présentée la voie des folates et l'ordre d'intervention des enzymes DHPS et DHFR, ciblées respectivement par leurs inhibiteurs : la sulfadoxine et la pyriméthamine. À droite, les principales mutations responsables de résistance à la SP chez *P. falciparum* (Source thèse M. du Manoir).

Nous travaillons avec l'équipe *Parasites eucaryotes intracellulaires: immunité et chimiorésistance* du CPTP-U1043, spécialiste de cette thématique, dans le but d'évaluer les risques et les conséquences de l'émergence d'isolats de *P. falciparum* hautement résistants à la SP en Afrique et d'identifier de nouvelles stratégies thérapeutiques sur *PfDHPS* moins sensibles aux mutations. Cette collaboration bipartite et interdisciplinaire (biologie / physico-chimie) a débuté dans le cadre de la thèse de M. du Manoir, qui avait pour objectif de relier les caractéristiques fonctionnelles moléculaires de l'enzyme *PfDHPS* aux données épidémiologiques et phénotypiques recueillies par l'équipe du CPTP. Ce travail de thèse s'est attaché à faire un état des lieux récent de la prévalence des mutations des gènes *pfdhfr* et *pfdhps* à travers 7 sites d'Afrique Centrale. L'étude a porté en particulier sur un octuple mutant combinant trois mutations sur le gène *pfdhfr* et cinq sur le gène *pfdhps* (CirnI + vagKgs) qui a été découvert à Yaoundé en 2015 ; nos collègues biologistes l'ont retrouvé au Nigeria et au Cameroun, et pour la première fois à haut niveau de prévalence à Maroua (nord du Cameroun). Sur plusieurs sites d'Afrique Centrale, ils ont également décrit l'augmentation de la prévalence de la mutation K540E, typique des parasites résistants de l'Afrique de l'Est et sur laquelle se base l'implémentation de la SP par l'OMS. Ces observations laissent penser que des parasites avec un très haut niveau de résistance à la SP émergeraient en Afrique Centrale. **En l'absence d'alternative à la SP, des efforts de recherche sont actuellement nécessaires, à la fois pour mieux connaître l'épidémiologie des résistances, en particulier en Afrique Centrale, mais aussi pour mieux comprendre et prédire l'évolution dans le temps de ces résistances.**

Dans ce cadre, nous avons choisi de focaliser notre étude sur *PfDHPS* car de nombreuses études théoriques et expérimentales ont été menées ces dernières années sur *PfDHFR*. Ces études ont notamment mené à l'identification d'un potentiel inhibiteur, actuellement en phase d'essais cliniques [1]. La structure et les mécanismes de fonctionnement et de résistance de *PfDHPS* à la SP ne sont en revanche que très peu étudiés, et donc méconnus. Le champ d'investigation relatif à cette enzyme demeure ainsi vaste et ouvert, et c'est dans ce contexte que nous avons proposé de répondre aux objectifs suivants : i) modéliser la structure et le fonctionnement de l'enzyme *PfDHPS* à l'échelle atomique pour les différents allèles d'intérêt et valider ces calculs avec les données phénotypiques *in vitro* ; ii) modéliser le *fitness* de l'enzyme *PfDHPS* pour prédire l'évolution des résistances et l'impact sur la viabilité du parasite ; iii) proposer de nouvelles stratégies thérapeutiques possibles pour cibler *PfDHPS*. Ce programme ambitieux n'a été réalisé à ce jour que de façon préliminaire et nous sommes actuellement en recherche de financement pour avancer ce projet.

Méthodologie et résultats préliminaires

Nous cherchons ici à accéder à différents niveaux de compréhension et de prédiction : depuis la caractérisation structurale de l'enzyme *PfDHPS*, peu connue à ce jour, jusqu'à la compréhension des effets des mutations sur l'activité de l'enzyme vs la résistance à la SP (*i.e. fitness*) dans le but de proposer de nouvelles stratégies efficaces sur les cibles portant des mutations significatives. Nous souhaitons également explorer une potentielle corrélation entre mutations et niveaux de résistance.

Modéliser la structure de l'enzyme *PfDHPS* à l'échelle atomique. Le principal verrou expliquant le peu de travaux de simulation publiés sur *PfDHPS* repose sur l'absence de structure cristallisée pour cette enzyme. Nous avons mené des calculs exploratoires (dynamique moléculaire) sur le plus récent modèle d'homologie dont nous disposons [165], construit à partir de six protéines homologues. L'indice de confiance de ce modèle a été estimé à plus de 90% par l'équipe qui l'a réalisé, bien qu'un insert spécifique à *P. falciparum* présente un indice de confiance faible. Ce modèle, après équilibrage par dynamique moléculaire (AMBER 16, solvant explicite, champ de force ff99SB) a été comparé à une structure de *P. vivax* récemment publiée [245]. Celle-ci présentant la plus grande homologie de séquence avec *PfDHPS*, en particulier au niveau du site actif, elle nous a permis de vérifier notre modèle (présenté sur la figure 2.7).

A partir de là, une seconde étape a consisté à évaluer les changements conformationnels associés aux mutations. Pour cela, nous avons généré, en plus de la forme sauvage, des systèmes comportant chaque mutation seule pour évaluer son rôle de façon indépendante, mais aussi des combinaisons de mutations conformément aux haplotypes identifiés sur le terrain, en particulier ISgKAA, IagKAA, ISgKgs, ISgegA et vagKgs².

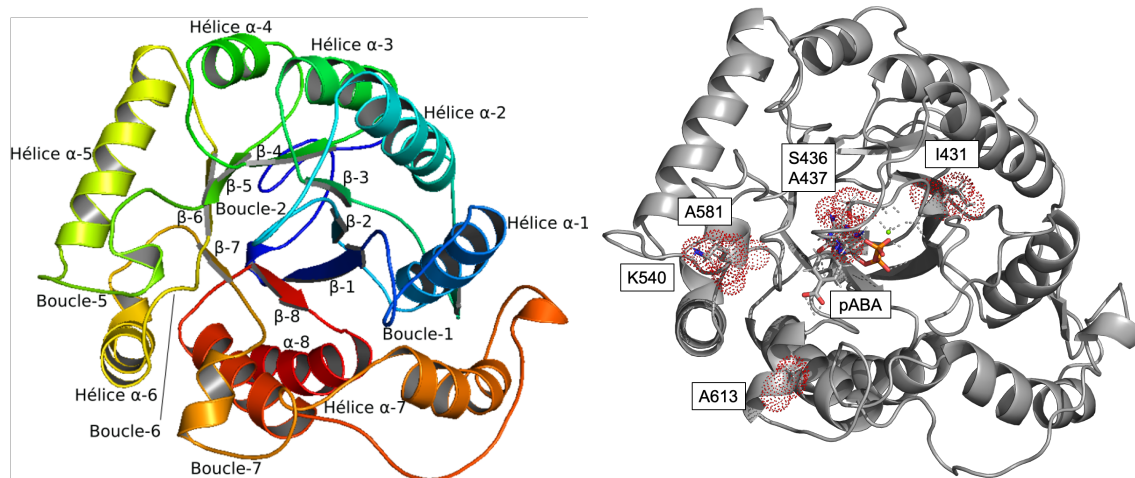


FIGURE 2.7 – A gauche : annotation des motifs de l'enzyme *PfDHPS* sur le modèle d'homologie utilisé [165]. La plupart des mutations liées à la résistance à la sulfadoxine sont situées sur des boucles : les mutations S436A et A437G se trouvent sur la boucle 2, K540E sur la boucle 5, A581G sur la boucle 6 et A613S sur la boucle 7. La mutation I431V, récemment identifiée, est la seule à être localisée sur un feuillet β (β -2). La boucle 2 est rarement résolue, suggérant une importante mobilité. Une hypothèse est que les boucles seraient stabilisées en présence de ligands, leur mobilité est donc d'intérêt pour une étude structurale et fonctionnelle. A droite : Localisation des six résidus dont les mutations sont associées à des résistances à la sulfadoxine.

Dans un premier temps, nous avons travaillé sur les structures de *PfDHPS* sans ligand afin d'évaluer l'impact des mutations sur leur flexibilité intrinsèque. Nos principales observations sont résumées dans le tableau suivant (Fig. 2.8).

2. Le acides aminés mutés sont notés en minuscule, la séquence correspond aux sites 431, 436, 437, 540, 581, 613.

	431		436		437		540		581		613	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
WT	1	S480	1	D482	1	V440	1	D539 T537	1	D539 T542 Y544	1	R610 N617
vSAKAA	1	S480 D482	1	D482 L531	1	-	Inst	-	1	D539 T542 Y544	1	R610 N617
IaAKAA	1	S480 D482	1	-	1	0	Inst	-	1	D539 Y544	1	-
ISgKAA	1	S480 D482	1	R608	1	S435	Inst	-	1	T542	1	R610
ISAcAA	1	G433 S480	2	S401 D404 R610 D430	1	S436 V440	1	R610	1	N534	1	N617
ISAKgA	1	S480	1	S436 D482	1	S436 I441	Inst	-	1	D539 T542	1	R610 N620
ISAKAs	1	S480	1	-	1	S435 V440	Inst	-	1	T542	1	R610 M616 N617 N624
vagKAA	1	V397 S480	1	-	1	-	Inst	-	1	-	1	R610 M616
IagKAA	2	G433 S480 G433 D482	1	-	1	V440	Inst	-	1	-	1	R610
ISgKgs	1	G433 D482 S480	1	E434 K531	1	V440	Inst	-	1	T542	2	K540 H614
vSgKgs	2	S480	2	-	1	V440	1	D539	2	T542	1	R610 M616
IagKgs	2	S480 D482 G433 S480	1	-	1	V440	Inst	-	1	D539	1	R610 N617 V622
vagKgs	2	S480 D482	1	E434	1	V440	Inst	-	1	D539 T542 Y582	2	M616 I623 N624
vagegs	1	S480	1	E434 K531	1	V440	Inst	-	1	T542	1	R610 N617
ISgegA	1	G433 S480	2	D482	1	V440	1	R610	1	R610	1	-

FIGURE 2.8 – Tableau résumant les effets des mutations sur les structures des haplotypes étudiés, obtenus par analyse des trajectoires de dynamique moléculaire. Pour chacun des haplotypes (correspondants aux lignes du tableau), les conformations des résidus d'intérêt (présentés indépendamment dans chaque colonne) sont indiquées de façon quantitative (colonne (a)) et qualitative (colonne (b)). La colonne (a) indique de nombre d'états observés dans les états stabilisés (généralement 1 ou 2) ou les états instables (Inst). La colonne (b) indique les résidus avec lesquels des liaisons hydrogène sont établies.

Mutation I431V. L'impact de cette mutation récemment découverte est encore inconnu. Dans la forme sauvage, nous observons que le résidu I431, situé sur le feuillet β -2, forme une liaison hydrogène avec le résidu S480 du feuillet β -3. Dans les systèmes vSAKAA, IaAKAA, ISgKAA, IagKAA et vagKgs, le résidu 431 forme une deuxième liaison hydrogène avec le feuillet β -3, au niveau du résidu 482. Or ces systèmes portant tous une mutation dans cette région (431, 436, ou 437), nous pouvons émettre l'hypothèse que leur présence pourrait être liée au rapprochement des feuillet β -2 et 3.

Mutation S436A. Le résidu 436 se situe sur la boucle 2, proche du ligand pABA ou de son analogue. Sur ce site, on ne relève de modification significative par rapport à la forme native que sur les systèmes portant une glycine en position 437 (mutation A437G). Cette seconde mutation libère de l'espace et facilite la rotation de la chaîne latérale du résidu 436 qui forme alors des liaisons avec les boucles 5 et 7 en l'absence de ligand (en particulier avec le résidu V440). On ne peut conclure sans avoir mis les ligands en présence afin de vérifier si ces nouvelles liaisons sont maintenues (et les résidus impliqués ne sont par conséquent plus disponibles pour les ligands) ou si la liaison aux ligands est plus favorable.

Mutation A437G. On substitue ici un atome d'hydrogène à un groupement CH_3 . Nous ne nous attendons pas à observer de modification de la structure ou du RMS, ce que nous obtenons en effet. Cette mutation n'affecte pas les propriétés intrinsèque de PfdHPS, mais probablement l'interaction à la sulfadoxine, ce qui sera vérifié après introduction des ligands. Notons que la prévalence de cette mutation est en constante augmentation, elle ne doit donc pas affecter l'activité de PfdHPS.

Mutation K540E. Cette mutation est synonyme de très forte résistance. On remplace ici une chaîne latérale longue et chargée positivement par une chaîne plus courte et chargée négativement. Le résidu 540 se trouve sur la boucle 6, sur l'extérieur de la protéine. Son environnement ne permet pas de le stabiliser et nous avons observé des oscillations constantes, autorisant des interactions ponctuelles avec des voisins immédiats, en particulier avec D539 dans la forme sauvage et le système vSgKgs. Dans l'ensemble des systèmes, le résidu est donc instable, excepté dans les cas où il est muté, cas dans lequel il peut se stabiliser au travers d'interactions avec le résidu R610 appartenant à la boucle 7. Si cette liaison est suffisamment stable, elle peut affecter, voire empêcher la liaison de la sulfadoxine.

Mutation A581G. La substitution d'une alanine par une glycine n'affecte pas la structure et la dynamique de la boucle 6 de façon significative. Le résidu 581 continue, dans la plupart des cas, à établir des ponts hydrogènes avec les résidus voisins de la boucle 5 (D539, T542, Y544 en particulier). Cependant, selon les liaisons établies, la distance inter-boucles peut sensiblement varier et induire une réorientation du résidu 540.

Mutation A613S. La substitution d'une alanine en serine qui possède une chaîne latérale légèrement plus longue et hydrophile (ajout d'un groupe OH) ne modifie pas la structure et peu le réseau d'interactions locales. On observe différentes orientations du groupement OH qui lui permettent d'interagir avec ses différents voisins au cours du temps. Néanmoins, ces voisins, R610 en particulier, sont très proches du site de liaison de la sulfadoxine et pourraient affecter son greffage.

Dans un second temps, nous avons ajouté les ligands (DHPP, pABA/sulfadoxine) par alignement avec la structure liée de *Yersinia pestis* (PDB 3TYZ)) et reconduit les dynamiques moléculaires afin d'évaluer les réseaux et énergies d'interactions de la protéine avec son ligand naturel (pABA) ou son inhibiteur (sulfadoxine). Les interactions entre pABA/sulfadoxine et *Pf*DHPS sont établies avec les mêmes résidus, à savoir S436, G579 et R610. Il est donc normal de s'attendre à ce que les mutations de ces sites ou de leurs proches voisins affectent la liaison de pABA et de la sulfadoxine à leur cible.

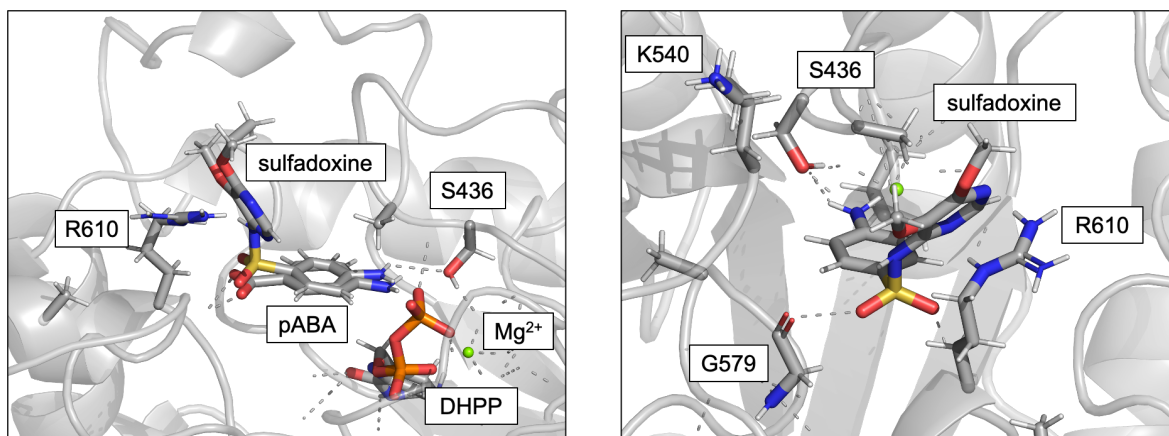


FIGURE 2.9 – À gauche sont représentées les positions de pABA et de la sulfadoxine superposées dans le site actif de *Pf*DHPS. Les deux ligands interagissent avec leur cible par des liaisons hydrogène établies avec les résidus S436 et R610. À droite, le complexe est réorienté de 90° pour mettre en évidence la position de la chaîne latérale de R610, plaquée contre la sulfadoxine, et faisant face à K540 sur la boucle voisine.

Les résultats que nous avons obtenus sont encore préliminaires et nécessitent des simulations plus longues afin d'obtenir des valeurs d'énergies libres pertinentes. Ces premiers résultats révèlent néanmoins que la liaison au substrat naturel (pABA) et à la sulfadoxine semble possible mais est affectée à différents degrés selon les mutations, c'est-à-dire que les ligands sont déplacés au sein du site actif. Or, un déplacement même infime peut drastiquement modifier l'affinité et l'activité enzymatique. Nous devons donc obtenir une configuration du site actif affinée autant que possible pour comparer les formes natives et mutées.

Des mécanismes particuliers sont à creuser à la lumière des résultats obtenus sur *Pf*DHPS non liée. En effet, les mutations ne doivent pas affecter la liaison à pABA, mais seulement à la sulfadoxine, sans quoi le parasite ne survivrait pas. C'est donc certainement sur la partie extérieure de la sulfadoxine qu'il faut concentrer les efforts. En l'occurrence, nous avons vu que le mutant K540E peut se lier à R610 en l'absence de ligand. Or il apparaît nettement sur la partie droite de la figure 2.9 que la sulfadoxine est positionnée entre ces deux résidus. Il est donc nécessaire de mettre en compétition les liaisons K540E-R610 et sulfadoxine-R610 et de déterminer les énergies associées afin de confirmer ou d'infirmer que la liaison K540E-R610 l'emporte. **Les mécanismes relatifs à chaque mutation et chaque haplotype peuvent être très différents, tout en menant tous à une perte d'affinité de la sulfadoxine pour sa cible, d'où la nécessité de tous les explorer en détail.**

Perspectives

Ce travail doit être poursuivi pour comprendre les mécanismes de résistance et ainsi établir des critères pertinents pour la proposition de nouveaux traitements. Nous devons en effet comprendre comment les mutations empêchent l'action de la sulfadoxine sans affecter l'activité de l'enzyme. Cette évaluation du *fitness* de l'enzyme doit être réalisée en parallèle par approches biologiques (collaboration expérimentale) et calculatoires. Concernant la modélisation, la méthode des Modes Statiques sera appliquée pour étudier les propriétés biomécaniques de la cible et identifier des résidus d'intérêt par rapport aux critères établis à l'aide des simulations de dynamique moléculaire décrites ci-dessus (par exemple des résidus impliqués dans l'amarrage des ligands ou l'activité enzymatique). Ces résidus seront comparés aux sites de mutations connus. **Leurs effets seront hiérarchisés et mis en perspective par rapport aux niveaux de résistance observés.**

Si une telle corrélation est possible, elle validera notre approche et offrira trois possibilités : i) prédire le niveau de résistance d'un nouvel haplotype identifié par l'équipe expérimentale et de résistance inconnue ; ii) proposer de nouveaux mutants à tester *in vitro* afin d'évaluer le *fitness* de l'enzyme modifiée ; iii) identifier des sites et stratégies pertinents pour cibler *Pf*DHPS. Dans ces calculs, le lien au *fitness* se fera en comparant les réponses des enzymes selon leur liaison au ligand naturel ou à la sulfadoxine. Des calculs complémentaires de type QM/MM pourront être requis pour **accéder à la réactivité chimique de l'enzyme et la caractériser qualitativement et quantitativement.**

De manière itérative, les sites identifiés pourront être soumis à des substitutions et les nouveaux mutants testés en suivant le même protocole afin de **vérifier si la modification « artificielle » que nous générons est favorable ou défavorable** au *fitness* de l'enzyme. A ce stade, nous obtiendrons deux types d'informations pouvant servir d'entrée à la recherche de nouveaux inhibiteurs : une liste de sites « actifs » et la nature chimique de la substitution à générer pour obtenir une réponse donnée de l'enzyme. On s'attend bien sûr à ce que certains de ces sites soient déjà connus pour leur rôle dans le fonctionnement de l'enzyme ou l'émergence de résistances. Ceci permettrait par ailleurs de corroborer le caractère prédictif de notre approche. On s'attend également à voir ressortir des sites encore non identifiés, dont la substitution modifiera l'activité enzymatique et/ou l'amarrage des ligands dans un sens favorable ou défavorable.

Deux stratégies pourront alors être envisagées au regard de ces informations. (i) Un site stratégique dont la substitution n'affecte ni la liaison à pABA ni l'activité enzymatique pourra être anticipé comme potentielle source de résistance, et donc comme pouvant émerger. Ce site pourra alors être exclu de la recherche des sites de liaison de la drogue. (ii) Suivant le même protocole calculatoire, nous pourrions rechercher les sites jouant un rôle dans le maintien de l'activité enzymatique. Ces sites pourront être mutés *in silico* pour identifier des substitutions bloquant cette activité. Ces nouvelles mutations, défavorables au parasite, ont a priori peu de chance d'émerger par pression médicamenteuse, car ce serait au détriment de la fonction de l'enzyme. Elles présentent donc un double intérêt : i) nous pourrions explorer la possibilité de poches de liaison autour de ces sites conservés ; ii) la nature et le site de la substitution bloquant l'activité enzymatique seront les indicateurs d'une nouvelle stratégie thérapeutique à adopter mimant leur effet. Ces sites pourront par ailleurs être mutés *in silico*, et l'activité des nouveaux mutants testées *in vitro* pour validation.

Ce programme est ambitieux, mais nous travaillons depuis plusieurs années maintenant sur une cible sur laquelle nous avons prouvé que cette stratégie fonctionne. Il s'agit du cas de l'oncoprotéine Ras que je développe dans la partie suivante (2.2). Auparavant, je termine cette première partie sur un dernier cas d'exploration de la réponse biomécanique d'un récepteur opioïde qui me permet de présenter une application des Modes Statiques aux mécanismes d'activation/inhibition.

C. Modélisation de la réponse biomécanique du récepteur opioïde μ

Cadre de travail : Ce travail a été mené dans le cadre d'une collaboration avec l'équipe de M. Dauchez (UMR 7369, CNRS, Université de Reims Champagne-Ardenne), et N. Marie (ERL 3649, CNRS, Université Paris Descartes). Il a été soutenu par un financement CNRS (Défi Mécanobiologie) et a fait l'objet du stage de L. Casse.

Résumé

Le récepteur opioïde μ est un récepteur couplé aux protéines G. Il est la cible de la morphine et est à l'origine des effets analgésiques et addictifs qui lui sont associés. Aujourd'hui encore, la relation structure/fonction/dynamique de ces complexes demeure inconnue de manière précise. Cette question fondamentale ouvre donc de nombreuses perspectives thérapeutiques. Nous avons utilisé les Modes Statiques pour explorer l'effet d'interactions de petites molécules au travers de l'étude de la plasticité spontanée interne ou sous stress mécanique externe du récepteur dans son environnement membranaire. Nous avons en particulier identifié des résidus stratégiques dans la transmission des mouvements et établi des corrélations entre grands domaines structuraux, ce qui apporte des éléments de compréhension à la mécanique d'activation/inhibition du récepteur.

Contexte

C'est dans le cadre d'un consortium interdisciplinaire rassemblant biologistes, pharmacologues, physiiciens et biophysiciens, que nous nous sommes intéressés au récepteur opioïde μ (μ OR), qui est la cible de la morphine et est à l'origine de ses effets analgésiques et addictifs. Malgré de récentes avancées sur sa structure tridimensionnelle, sa mécanique d'activation/inhibition en fonction de contraintes exercées par un ligand reste totalement incomprise. En effet, les différentes étapes et modes d'interactions sont totalement spécifiques du récepteur mais dépendent aussi et surtout de son comportement dynamique et de ses réponses en fonction de contraintes externes qui doivent donc être prises en compte. Il existe de nombreux agonistes de μ OR ayant des propriétés pharmacologiques différentes. Pourtant, nous ne disposons aujourd'hui que de peu de données sur la mécanique d'activation du μ OR en fonction de la nature des ligands. La résolution de structures par cristallographie de complexes permet de comprendre l'interaction fine dans l'état final mais ne donne pas accès aux processus d'interactions eux-mêmes, ni aux conséquences mécanistiques associées.

Dans ce projet, nous avons proposé de combiner notre expertise des Modes Statiques à celle de l'équipe rémoise de M. Dauchez qui développe des techniques de docking inverse et d'amarrage moléculaire multiple. L'objectif est ici d'**explorer l'effet d'interaction de ligands sous stress mécanique** et d'accéder aux mécanismes impliqués dans le processus d'activation et de fonctionnement du récepteur μ OR. Une troisième équipe nous a apporté l'expertise biologique des récepteurs opioïdes et la possibilité de validation des prédictions par apport de données biologiques issues de mutagenèse dirigée et par évaluation de l'activité du récepteur.

Les données expérimentales fournies par nos collaborateurs biologistes ont permis dans une première étape de construire la démarche *in silico* devant permettre de déterminer comment des ligands de μ OR interagissent avec ce récepteur et modifient sa mécanique d'activation.

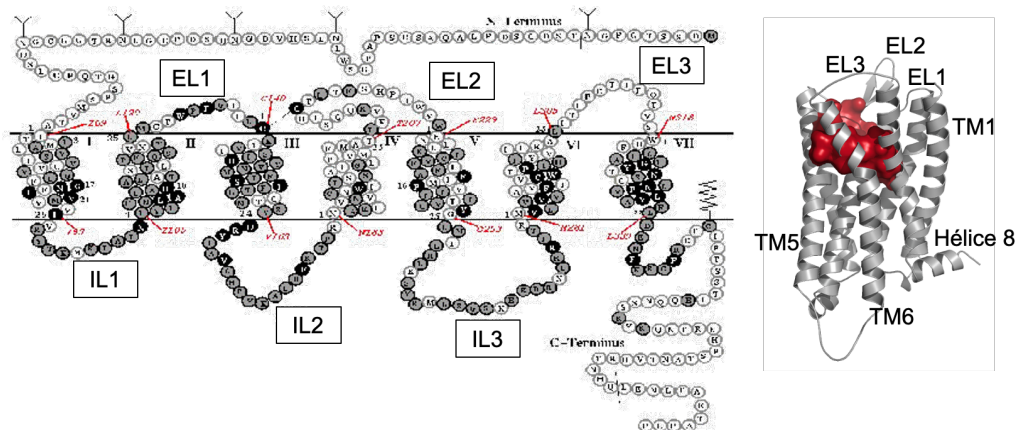


FIGURE 2.10 – A gauche : représentation de la structure secondaire de μ OR constitué de sept hélices transmembranaires (notées TM), de trois boucles intracellulaires (ICL) et de trois boucles extracellulaires (ECL). A droite : la structure tridimensionnelle est représentée en gris et le site de liaison de la morphine en rouge.

Méthodologie

Nous avons dans un premier temps cherché à caractériser l'interaction cible (récepteur μ OR) / ligands (morphine et traitements de substitution : buprénorphine et méthadone). Nous avons pour cela travaillé en étroite collaboration avec nos collègues rémois afin de mutualiser nos résultats et de nourrir tour à tour les calculs de l'autre équipe. L'équipe de l'UMR 7369 a ainsi construit les ligands et déterminé leurs paramètres compatibles avec les champs de forces GROMACS et AMBER, affiné les structures cristallographiques du récepteur en forme active (PDB 5C1M [96]) et inactive (4DKL [143]) à l'aide du logiciel GROMACS, amarré les trois ligands sur le récepteur dans les différentes conformations précédemment obtenues. Ce sont sur ces structures que nous avons travaillé et que nous avons i) évalué la stabilité des complexes obtenus et calculé leurs énergies d'interaction ; ii) caractérisé le site actif du récepteur en terme de volume et de résidus impliqués ; iii) calculé les Modes Normaux et les Modes Statiques des complexes (mouvements naturels et induits) ; iv) utilisé les Modes Statiques pour évaluer les sites d'intérêts et les corrélations internes (*i.e.* mécanismes d'allostérie). Sur ce dernier point en particulier, nous avons utilisé un outil d'analyse permettant de déterminer quels sont les modes "délocalisés", c'est-à-dire quels atomes sous contrainte engendrent le déplacement d'un grand nombre d'atomes. A partir des Modes Statiques, les déplacements $x_{i\alpha}$ des atomes i dans la direction $\alpha(x, y, z)$ sont utilisés pour calculer les moments μ_n et en déduire la localisation l associée :

$$\mu_n = \sum_i \left[\sum_{\alpha} (x_{i\alpha})^2 \right]^n \quad \text{et} \quad l = \frac{\mu_1^2}{\mu_2} \quad (2.1)$$

Si les déplacements $x_{i\alpha}$ sont nuls sauf pour un atome $i = 0$, on peut en déduire que :

$$\mu_n = \left[\sum_{\alpha} (x_{0\alpha})^2 \right]^n \quad \text{et donc que} \quad l = 1 \quad (2.2)$$

Si les déplacements des atomes sont tous équivalents, alors $\sum_{\alpha} (x_{i\alpha})^2$ est indépendant de i et :

$$\mu_n = N \left[\sum_{\alpha} (x_{0\alpha})^2 \right]^n \quad \text{et par conséquent} \quad l = N \quad (2.3)$$

Ainsi, pour la localisation d'un mode, on trouve une valeur de l comprise entre 1 et N , c'est-à-dire un indicateur de la propagation de la déformation. Cette approche a déjà été utilisée pour la localisation de Modes Normaux dans des systèmes cristallins [23] et je montre ici qu'elle est aussi adaptée pour extraire des informations pertinentes des Modes Statiques sur la **propagation des déformations au sein des systèmes de grande taille**.

Résultats

Construction et évaluation de la stabilité des complexes récepteur/ligands. Pour comprendre l'impact des différents ligands agonistes ou antagonistes, il est nécessaire d'obtenir des structures des complexes, ce qui a été réalisé à partir de la structure du récepteur inactif. L'amarrage moléculaire a été prédit pour trois ligands induisant des effets analgésiques et/ou addictifs différents : la morphine, la buprénorphine (principe actif du subutex) et la méthadone. Cette première étape, menée par nos partenaires de Reims, a permis de vérifier que les sites de fixation des ligands se superposaient et correspondaient au site identifié pour un autre antagoniste du récepteur se trouvant dans la structure cristallographique (PDB 4DKL). Des simulations de dynamique moléculaire ont alors été conduites afin de pouvoir évaluer la stabilité des complexes et de la quantifier en calculant l'énergie d'interaction pour chacun de ces ligands. Pour les trois ligands, l'énergie d'interaction est du même ordre de grandeur. Il est à noter que l'énergie d'interaction de la buprénorphine ($-11,28\text{kcal.mol}^{-1}$) est plus basse que celle observée pour la morphine ($-8,25\text{kcal.mol}^{-1}$) traduisant le pouvoir d'inhibiteur irréversible de la buprénorphine qui empêche ainsi la morphine de venir se fixer sur le récepteur. Ces simulations ont aussi permis une étude fine de la contribution énergétique de chacun des résidus impliqués dans l'interaction, et ainsi d'identifier les résidus favorables et défavorables à l'interaction avec le récepteur. En comparant les résultats pour chacun des ligands, il a été possible d'identifier deux résidus (D147 et K233) qui sont défavorables à l'interaction de la morphine et favorables à l'interaction de la buprénorphine ou de la méthadone. Une piste est donc de substituer ces résidus polaires par des acides aminés aromatiques renforçant ainsi le pouvoir de stacking entre les ligands et le récepteur μ .

Caractérisation des sites de liaison. Nous avons dans un premier temps travaillé sur les structures actives et inactives du récepteur sans ligand (extraites des PDB 5C1M et 4DKL et relaxées avec AMBER en solvant implicite). Nous avons utilisé le serveur PockDrug³ qui identifie et caractérise les potentielles poches de liaison en terme de *druggabilité*. Elles sont pour cela classées selon le volume de la cavité, mais aussi selon la probabilité qu'un ligand vienne s'y fixer (prise en compte des possibles interactions électrostatiques, de l'indice d'hydrophobicité, de la proportion de résidus polaires et aromatiques, etc). Dans notre cas, dans les deux structures, la première poche identifiée (en jaune et notée P_0 sur la figure) correspond bien à la poche dans laquelle viennent se fixer les ligands.

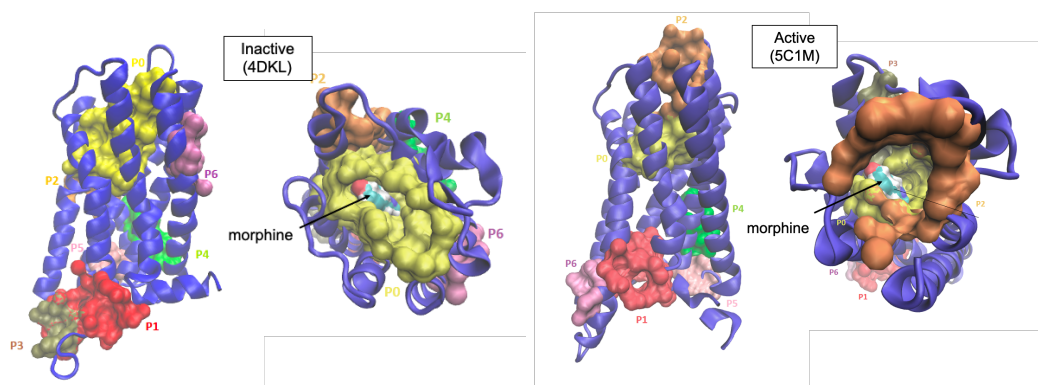


FIGURE 2.11 – Localisation des poches de liaisons prédites par PockDrug sur les conformations inactive et active du récepteur μ OR. On montre dans les deux cas la vue selon le plan membranaire et la vue extracellulaire.

A partir des complexes générés précédemment, nous avons évalué le volume occupé par chacun des ligands au sein de μ OR. Ce volume a été estimé à 289\AA^3 pour la morphine, 377\AA^3 pour la méthadone et 505\AA^3 pour la buprénorphine. Par ailleurs, le volume de la cavité de fixation dans le récepteur μ étant de $1908,8\text{\AA}^3$ (pour 4DKL), il peut être envisagé que plusieurs ligands pénètrent dans la cavité. Au regard de la différence de taille des ligands, principalement entre la morphine et la buprénorphine, il est possible que deux fois plus de morphine pénètre dans la cavité que de buprénorphine. Il est donc pertinent d'envisager ensuite une étude de l'effet-dose en ajoutant successivement plusieurs ligands dans la procédure de docking (amarrage moléculaire), afin d'évaluer combien de ligands peuvent être fixés, et quel est l'impact de ce paramètre sur la structure, la dynamique et l'activité du récepteur.

3. <http://pockdrug.rpbs.univ-paris-diderot.fr>

L'étude du site actif nous a également permis d'identifier les résidus impliqués dans l'interaction, ce qui est essentiel pour les soumettre ensuite à des forces par notre approche de Modes Statiques. Nous avons identifié 11 résidus pour la morphine, 15 pour la méthadone, et 17 pour la buprénorphine. Il est à noter que ces résidus sont conservés dans les trois cas : 9 des 11 résidus impliqués dans l'interaction avec la morphine se retrouvent dans l'interaction avec les substituants, 14 des 15 résidus impliqués dans l'interaction avec la méthadone sont communs au cas de la buprénorphine (voir tableau ci-dessous).

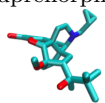
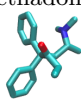

Ligand	Volume (\AA^3)	Résidus de μ OR en interaction
buprénorphine 	505	Q124 V143 I144 D147 Y148 M151 L219 E229 L232 K233 V236 I296 V300 K303 W318 I322 Y326
méthadone 	377	Q124 V143 I144 D147 Y148 M151 L232 K233 V236 W293 I296 V300 W318 I322 Y326
morphine 	289	D147 Y148 N150 M151 K233 V236 W293 I296 H297 V300 I322 Y326

TABLE 2.1 – Présentation des ligands étudiés et de leurs sites de liaison au récepteur μ OR. Le volume occupé a été calculé à l'aide du serveur 3v (Voss Volume Voxelator [229]) et d'une grille de haute résolution (0.5\AA).

Réarrangement conformationnel du récepteur. La comparaison des structures active et inactive révèle en particulier un plan de cassure au milieu du récepteur, en dessous de la zone de fixation du ligand, au milieu de la membrane (en rouge sur la figure 2.12). Ces points de cassures sont connus pour jouer un rôle important dans les changements structuraux lors de la transduction de signal dans ce type de récepteur. A partir des structures cristallographiques, il a été possible de les localiser, d'évaluer l'amplitude de leur courbure et d'identifier les résidus impliqués. La comparaison entre les structures active et inactive a montré que l'hélice transmembranaire la plus impactée dans le changement conformationnel est l'hélice TM2, l'amplitude de l'angle passant de $11,3^\circ$ à $36,4^\circ$. On peut donc supposer que la transduction du signal au travers du récepteur μ est réalisée par l'intermédiaire de l'hélice TM2. Or cette hélice, se trouve à proximité de l'hélice 8 impliquée dans l'interaction avec la protéine G et permettant l'activation de la cascade de signalisation.

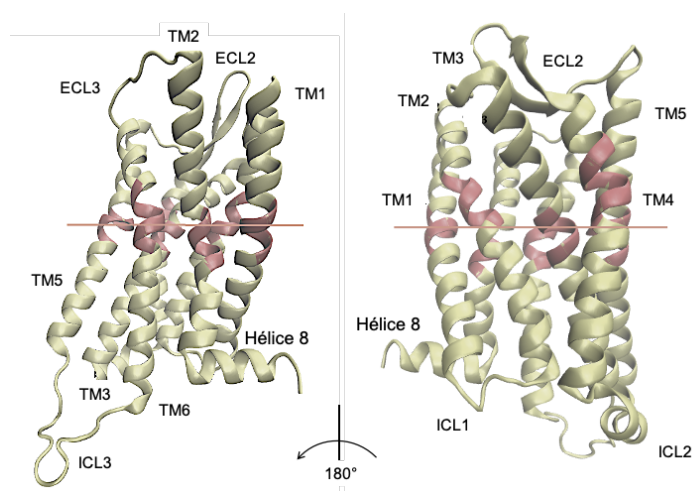


FIGURE 2.12 – Représentation en rouge des résidus impliqués dans le plan de cassure de μ OR.

Nous nous sommes donc interrogés sur la nature du mouvement induisant la courbure des hélices α à ce niveau : se fait-elle spontanément ? Est-elle due à la nature des résidus ? Induite par l'amarrage du ligand ? Par les contraintes exercées par la membrane ? Ces deux dernières possibilités doivent être vérifiées par l'application de forces locales grâce aux Modes Statiques, respectivement sur le site de fixation du ligand et sur la membrane. En parallèle, pour approfondir la compréhension du mécanisme du changement conformationnel, une étude par modes normaux a été conduite. Elle a permis d'identifier deux modes distincts pouvant expliquer la transition entre l'état actif et l'état inactif. La somme de ces modes permet d'observer le déplacement des hélices TM1, TM6 et H8 ainsi que la partie extracellulaire de l'hélice TM2. L'ensemble de ces données montre que l'hélice TM2 joue un rôle important dans la transition conformationnelle entre les deux états du récepteur μ .

Impact des ligands sur le changement conformationnel. Les Modes Normaux permettent de prédire les mouvements naturels mais pas le mécanisme qui les induit, ce que nous proposons de faire à l'aide des Modes Statiques. Les calculs, à ce stade préliminaires, ont été menés sur le récepteur seul et en complexe pour une caractérisation « à l'aveugle » du récepteur, c'est-à-dire que des contraintes ont été appliquées indifféremment sur tous les atomes.

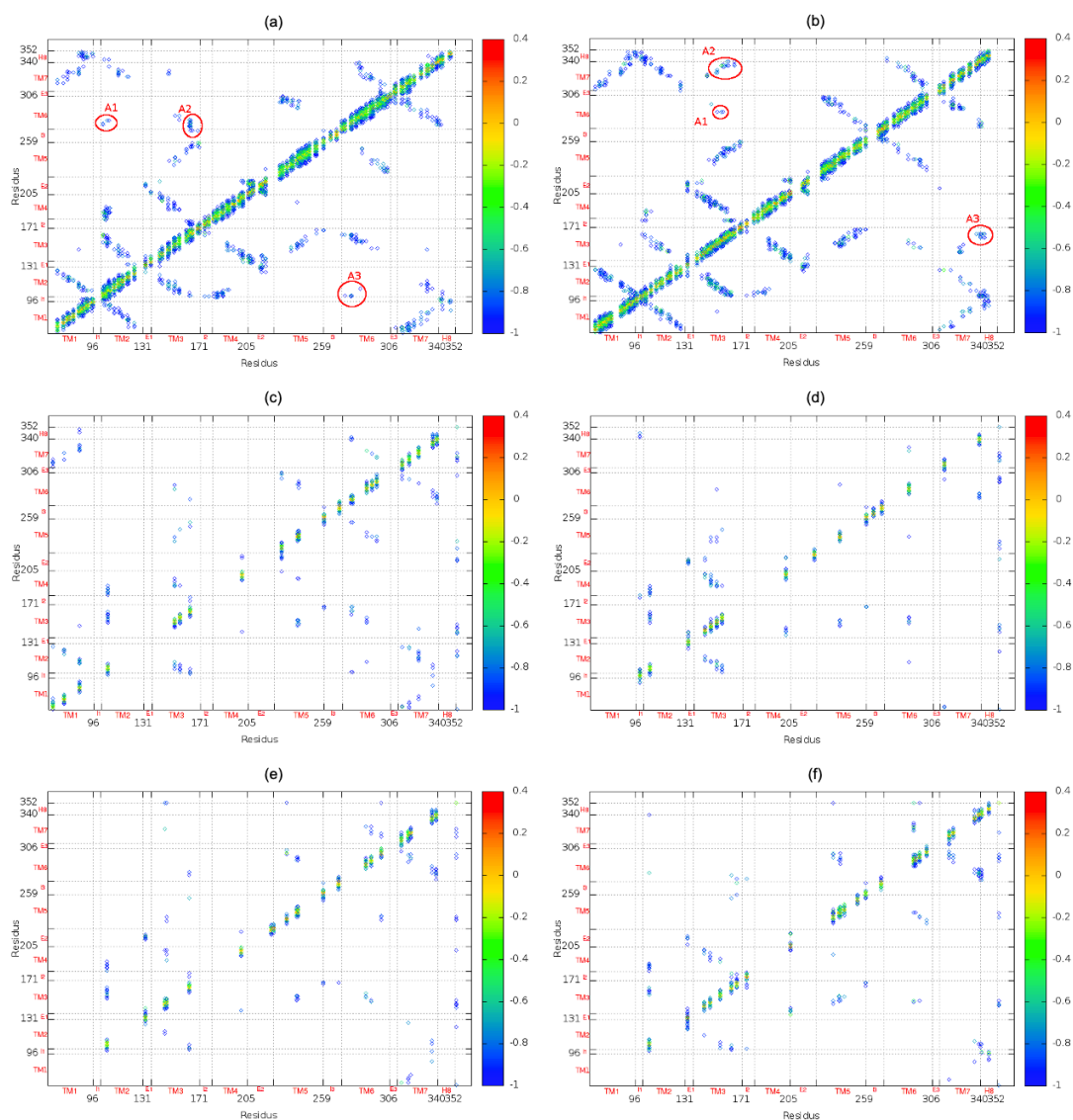


FIGURE 2.13 – Localisation des Modes Statiques moyennée par résidu sur les structures sans ligand (inactive 4DKL (a) et active 5C1M (b)) et avec ligand (buprénorphine 1 (c) et 2* (d), méthadone (e) et morphine (f)).
* Le cas buprénorphine 2 correspond à une deuxième pose issue du docking avec un score moindre et s'est avérée instable lors des dynamiques moléculaires. Elle a été retirée des calculs par la suite.

Pour chacune des contraintes, nous avons calculé la réponse induite sur l’ensemble des déplacements atomiques afin de localiser celles induisant un déplacement significatif au sein du récepteur et d’identifier des résidus d’intérêt. Ces résidus seront la cible de l’application de forces choisies en conséquence, dans une prochaine étape. La figure 2.13 montre les résultats des calculs de *localisation* des déformations extraites des Modes Statiques. Nous y représentons, pour chaque système étudié, les résidus déplacés (en ordonnée) pour chaque résidu contraint (en abscisse). Pour une meilleure lisibilité, nous ne faisons apparaître que les résidus donnant une valeur de localisation $l > 20$ et induisant des déplacements plus grands que 0.1Å . Les calculs sont faits pour chaque atome mais les résultats présentés ici comptabilisent les moyennes par résidu pour simplifier la lecture. Toutes les informations détaillées à l’échelle atomique sont en revanche stockées et disponibles.

Notons tout d’abord que les contraintes appliquées génèrent davantage de déplacements (en terme de propagation et d’intensité) dans les structures sans ligands (a et b) que dans les complexes. Les graphes a et b, correspondant aux formes inactive et active respectivement, sont très semblables, à certaines différences près (entourées en rouge). En particulier, en conformation inactive, le déplacement de certains résidus de TM2 engendrent des déplacements plus importants de résidus situés au milieu de TM6, notamment M281 (A1). De même, le déplacement de R165, situé en-dessous du site actif sur TM3, entraîne celui de plusieurs résidus situés dans la partie inférieure de TM6 dont T279 (A2). Enfin, des résidus de TM6 entraînent les déplacements de A102 et T103, localisés sur TM2, déplacements qui n’apparaissent plus sur la structure active (A3). En conformation active, les déplacements de résidus situés au milieu de TM3 induisent en particulier le déplacement de F289, appartenant à un trio d’acides aminés conservés (I155, P244, F289) se trouvant juste en-dessous de la poche de liaison. Des études antérieures ont déjà émis l’hypothèse que F289 serait lié à la transition inactive/active [96]. Les mêmes résidus de TM3 induisent également le déplacement d’autres résidus situés dans la partie basse de TM7 (A2). Enfin, la zone A3 indique que des résidus de l’hélice 8 interviennent dans le déplacement de résidus situés dans la partie inférieure de TM3 et notamment de R165.

Position	Buprénorphine	Méthadone	Morphine
TM2	S119 T120 Q124	Q124	-
TM3	I142 V143 I144 D147 Y148 M151	I144 I146 D147 Y148	I146 D147 Y148 M151
ECL2	D216 C217 T218	-	-
TM4	-	V202	-
TM5	V236	K233 I234 V236	V236 A240
TM6	K303	I296 V300	W293 I296 H297 V300
TM7	W318 H319 I322 Y326	W318 I322 Y326	I322

TABLE 2.2 – Principaux résidus déplacés par les contraintes imposées au niveau des différents ligands. Les résidus en gras sont ceux pour lesquels nous avons identifié au moins deux occurrences.

Enfin, les principaux résidus déplacés (table 2.2) ont pour la plupart été identifiés comme étant en interaction directe avec le ligand ou dans son voisinage proche. Il est donc pertinent de les retrouver, mais remarquable de noter qu’ils induisent eux-mêmes des mouvements délocalisés. C’est le cas par exemple de Trp293, déjà identifié pour son rôle dans la stabilisation à l’état actif ou inactif du récepteur. Nous pouvons ainsi observer des **corrélations claires entre les mouvements des résidus des différentes hélices**. Un mouvement spécifique de l’hélice TM2 entraîne en effet l’hélice TM6. Un couplage semblable apparaît entre TM3 et TM6, entre TM3 et l’hélice 8, etc. Parmi les résidus identifiés, nous retrouvons les résidus D147 et K233 déjà identifiés par dynamique moléculaire.

Perspectives

Ce projet est actuellement suspendu, faute de financement. Les premiers résultats présentés doivent servir de base aux futurs calculs afin d’actionner les résidus identifiés et d’évaluer leur impact grâce aux Modes Statiques. Les résultats issus de ces calculs serviront d’entrée pour prédire les amarrages multiples dans différentes conformations. Nous pourrions alors tester différents ligands à un coût raisonnable et nous intéresser à l’effet dose, c’est-à-dire étudier la modulation de l’activité du récepteur en fonction du type et du nombre de ligands. L’objectif final reste de proposer des sites à affinités différentes. Le partenaire biologiste validera ces résidus par mutagenèse dirigée en évaluant les propriétés de ces récepteurs mutants (affinité ligand/récepteur et activité sur les voies de signalisation).

2.2 Modulation de l'activité enzymatique par le criblage de mutations : apport du couplage des approches mécanistiques et physico-chimiques

Cibler l'hétérogénéité moléculaire en biologie : cas des oncoprotéines Ras

Cadre de travail : Ce travail a été initié en 2011 en collaboration avec G. Favre (CRCT-INSERM) et soutenu par deux financements de la Fondation RITC et de la Région Occitanie (Prématuration). Il a fait l'objet d'un stage (I. Soussi), d'un post-doc (V. Ramadoss) et de la thèse de R. Tichauer qui a conduit à la publication de deux articles [217,218]. Nous travaillons également avec J.D. Pedelacq et G. Czaplicki, tous deux à l'IPBS, respectivement pour la résolution de structures et le criblage haut débit de molécules médicaments.

Résumé

Notre objectif est de **restaurer la capacité d'hydrolyse de mutants de l'oncoprotéine Ras** pour les rendre inactifs et ainsi inhiber la prolifération tumorale. Pour répondre à ce défi nous avons développé une stratégie originale qui couple l'approche des Modes Statiques pour proposer de nouveaux mutants capables de rétablir l'activité de Ras, à des approches expérimentales moléculaires et cellulaires développées au CRCT (équipe de G. Favre) afin de valider ces prédictions. Nous avons par ailleurs utilisé des dynamiques QM/MM pour caractériser, à l'échelle atomique, les effets des mutations sur les propriétés du site actif (structure, stabilité, distribution du solvant, densité électronique), et pour les relier au dysfonctionnement de l'activité biologique au sein des protéines Ras mutées.

Contexte

Les protéines Ras agissent comme des interrupteurs moléculaires dans des réseaux de signalisation complexes qui contrôlent notamment la prolifération et la différenciation cellulaire. Ces protéines cyclent entre un état actif lorsqu'elles sont liées au GTP (*guanosine triphosphate*) et un état inactif lorsqu'elles sont liées au GDP (*guanosine diphosphate*). Le passage de l'état inactif à l'état actif se fait sous le contrôle de protéines GEFs (*Guanine nucleotide exchange factors*) qui permettent l'échange du GDP par le GTP, le retour à l'état inactif dépend de la protéine GAP (*GTPase activating proteins*) qui catalyse l'hydrolyse du GTP en GDP en enrichissant le site actif de Ras d'une arginine (R179) [35,147].

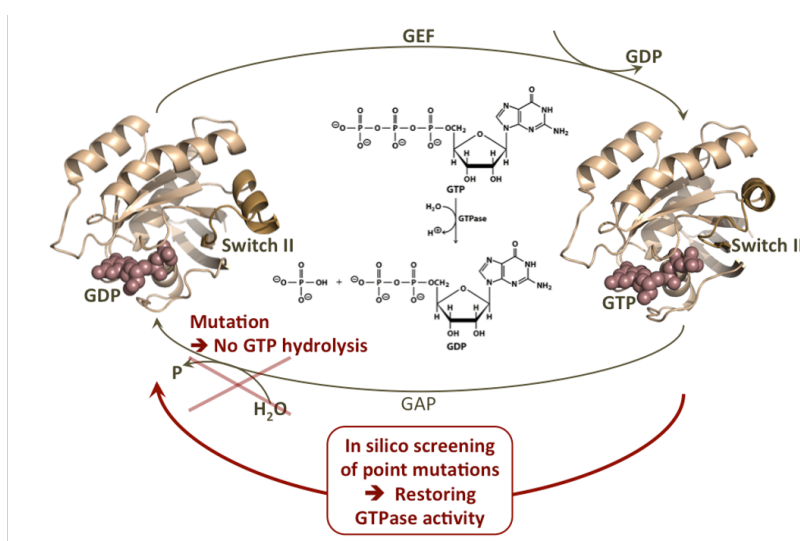


FIGURE 2.14 – Ras cycle entre deux états : actif (lié au GTP) et inactif (lié au GDP). Des mutations de Ras bloquent l'hydrolyse, maintenant ainsi Ras dans un état actif. Nous cherchons à proposer de nouvelles mutations visant à restaurer l'hydrolyse au sein de formes mutées de Ras.

Associées à plus d'un tiers de toutes les tumeurs humaines, les mutations somatiques de Ras favorisent l'oncogenèse de certains cancers et sont des marqueurs de mauvais pronostic [134,176]. Il existe trois isoformes de Ras (KRas, NRas, HRas) que l'on retrouve toutes mutées, la position et le type de substitution variant selon les différents cancers auxquels elles sont associées. Cependant, selon la base de données COSMIC⁴, les mutations ponctuelles se trouvent principalement sur les sites 12, 13 et 61 des trois isoformes. Les mutations de KRas sont fréquentes dans les cancers du pancréas, colorectal, du poumon et du col de l'utérus. Les mutations NRas et HRas prédominent respectivement dans le mélanome et le cancer de la vessie. Pour cette raison, la voie de signalisation Ras a attiré une attention considérable en tant que cible pour les thérapies anti-cancer [8,66]. Dans cette étude, nous avons concentré nos efforts sur NRas dont 60% des mutations sont des substitutions de Q61 (45% Q61R, 36% Q61K, 10% Q61L, 7% Q61H, 1% Q61P, <1% Q61E). Ces mutations sont considérées comme responsables du dysfonctionnement de Ras à différents niveaux: elles réduiraient sa capacité intrinsèque à hydrolyser le GTP [148] et la rendrait insensible à GAP [3].

Depuis 40 ans, les études théoriques et expérimentales se multiplient pour décrypter les mécanismes sous-jacents à l'activité de Ras. Ils ne sont pourtant que partiellement déterminés et deux voies sont proposées, suggérant un mécanisme dissociatif ou associatif [84,117,190]. Les substitutions de G61 conduisent toutes à une baisse drastique du taux d'hydrolyse, mais par des mécanismes différents selon leur nature. Pour comprendre ces mécanismes, nous devons d'abord comprendre le rôle que tient Q61. Des études évoquent i) son implication directe dans le transfert de proton, ii) son effet stabilisant sur l'état de transition dans un mécanisme impliquant le GTP lui-même comme base de la réaction, iii) son rôle dans le positionnement des molécules d'eau comme prérequis pour l'hydrolyse, iv) son effet indirect, par contraintes stériques, sur la stabilisation d'une configuration pré-organisée du site actif. Enfin, parmi les mécanismes proposés, le nombre de molécules d'eau présentes dans le site actif et impliquées dans l'hydrolyse reste sujet à controverse [49].

Parmi les nombreuses approches envisagées pour cibler directement ou indirectement Ras [8,69,138,174], la première était basée sur le développement de petits antagonistes du GTP, mais l'échec a été dû à la forte affinité de liaison du GTP à Ras avec laquelle il était difficile de rivaliser. Une deuxième stratégie infructueuse a consisté à développer de petites molécules capables de mimer GAP et de restaurer une activité GTPase normale. Après l'échec des thérapies ciblant directement Ras, de nouveaux efforts se sont tournés vers l'inhibition de l'association Ras/membrane. Plusieurs stratégies ont alors été développées pour inhiber la farnésylation de Ras, notamment à travers la conception de peptidomimétiques CAAX qui entrent en compétition pour la liaison à la FTase, mais aucune d'entre elles n'a connu de succès à ce jour. De nouvelles approches ont également visé à cibler les effecteurs de Ras en aval de la voie. Cependant, il a été démontré que certains de ces inhibiteurs interféraient avec d'autres enzymes. Plus récemment, une nouvelle poche de liaison a été identifiée chez Ras, dans laquelle de petites molécules pourraient inhiber l'échange de nucléotides assisté par SOS et empêcher la réactivation de Ras [45]. Cette étude a renouvelé l'espoir de réussir à cibler Ras, et c'est finalement par une approche très différente qu'une équipe vient de proposer un inhibiteur non-réversible, se liant de façon covalente au mutant KRas G12C [161].

Cependant, plusieurs mutants ont été décrits pour chaque gène Ras, différant par le site et le type de mutation, mais aussi par les résultats biologiques et cliniques. Jusqu'à présent, cette **hétérogénéité moléculaire** n'a pas été suffisamment prise en compte dans les stratégies thérapeutiques. Pourtant, alors que ces mutants ont tous pour effet de rendre Ras hyperactif, ils impliquent des mécanismes différents et peuvent ne pas réagir de manière identique aux thérapies développées. Il est donc nécessaire de déchiffrer, à l'échelle atomique, les conséquences des mutations somatiques sur l'activité de Ras, afin de proposer de nouvelles voies pertinentes pour contrôler l'activation constitutive des mutants.

Notre objectif est de restaurer la capacité d'hydrolyse du mutant pour le rendre inactif et ainsi inhiber la prolifération tumorale. Cette idée n'est certes pas nouvelle, cependant les stratégies appliquées à ce jour ont toutes échoué. Pour répondre à ce défi nous avons développé une **stratégie originale qui couple l'approche des Modes Statiques avec les approches expérimentales moléculaires et cellulaires** développées par l'équipe de Pr. G. Favre au CRCT (Fig. 2.14).

4. <https://cancer.sanger.ac.uk/cosmic>

Méthodologie

Ce travail repose sur trois étapes : i) la caractérisation, basée sur l'exploitation des Modes Statiques et des dynamiques moléculaires, des mécanismes par lesquels des mutations spécifiques génèrent un dysfonctionnement de Ras, mais aussi des résidus impliqués ; ii) les expériences de mutations *in silico*, pour valider le rôle des résidus précédemment identifiés, par les Modes Statiques, et par des dynamiques QM/MM ; iii) la validation des prédictions de modification fonctionnelle au travers d'expériences cellulaires, mais aussi par des caractérisations structurales.

Nous avons travaillé sur les trois isoformes HRas, NRas et KRas afin d'obtenir toutes les structures des formes natives et mutées, ainsi que les banques de données de Modes Statiques pour chaque système. La structure initiale de la forme native de NRas-GTP a été construite à partir des PDB 3CON (NRas-GDP) et 1QRA (HRas-GTP). La structure résultante a tout d'abord été relaxée en solvant implicite, puis soumise à des calculs de dynamique moléculaire jusqu'à l'équilibre à l'aide du logiciel AMBER (champ de force ff99SB). Les paramètres utilisés pour le GTP sont ceux de Bryce⁵. GAP a été ajoutée par superposition avec le PDB 1WQ1, seuls les résidus 780–795 de GAP, qui se lient directement à NRas, ont été conservés pour alléger les calculs. Tous les détails sont donnés dans les références [217, 218]. Les six mutations (Q61R, Q61K, Q61L, Q61H, Q61P et Q61E) ont été ajoutées sur le système natif, lequel a de nouveau été équilibré par la même procédure.

Les **Modes Statiques** ont été calculés sur toutes les structures (native et mutées) et ont permis une première caractérisation biomécanique de NRas, en particulier en utilisant notre algorithme d'optimisation sous contrainte (décrit dans la section 2.1). Cette procédure nous a notamment permis d'identifier les résidus dont les mouvements sont corrélés à ceux du GTP et de localiser un site "d'attaque" que nous proposons pour moduler l'activité de NRas.

A partir des mêmes structures, nous avons mené des dynamiques moléculaires en solvant explicite, en **dynamique classique et hybride QM/MM**, toujours avec le logiciel AMBER. Concernant les calculs QM/MM, la région quantique contient les résidus 12–13, 32, 35 et 60–61 de Ras, le résidu 789 de GAP, le GTP, le cation Mg^{2+} , les deux molécules d'eau qui lui sont coordonnées, ainsi que les molécules d'eau se trouvant à moins de 5 Å de l'atome P_γ du GTP (voir Fig. 2.15). Cette région a été traitée avec l'hamiltonien PM3 [211].

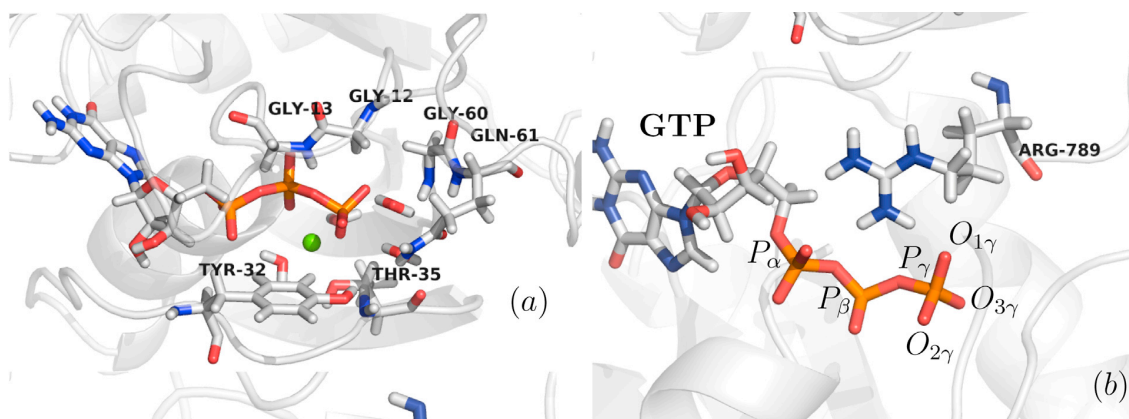


FIGURE 2.15 – Représentation du site actif de NRas et des résidus traités dans la région quantique.

Les trajectoires de dynamique moléculaire ont été analysées afin d'évaluer la stabilité du site actif et les réarrangements conformationnels dus aux mutations par calcul des RMSD des résidus d'intérêt (12, 13, 32, 35, 60, 61, 789). Nous avons également déterminé la fonction de densité de probabilité pour comparer la distribution de molécules d'eau au sein des différents sites actifs, ainsi que la contribution de chaque atome à la densité électronique du GTP par des calculs DFT (analyse de population de Löwdin et évaluation de l'orbitale HOMO avec le logiciel Orca interfacé avec AMBER [158], approximation GGA, fonctionnelle PBE et base Def2-SVP).

5. <http://research.bmh.manchester.ac.uk/bryce/amber/>

Résultats

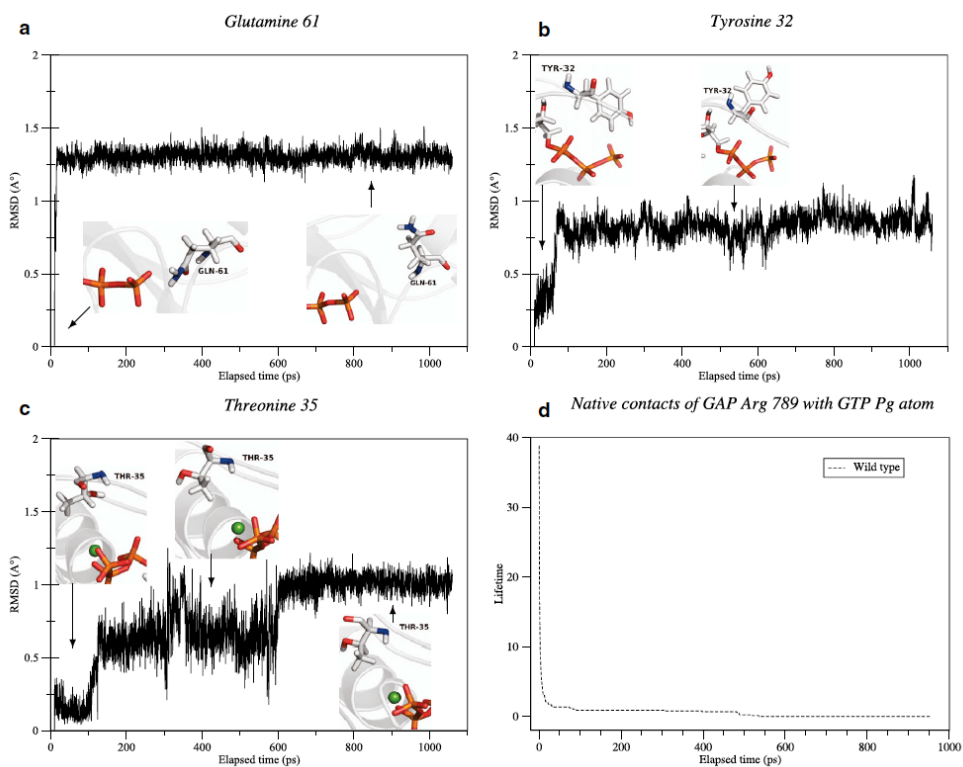
Caractérisation biomécanique du site actif. A l'aide des Modes Statiques, nous avons évalué de façon systématique la contribution de chaque atome sur les variations de distances caractéristiques de l'hydrolyse telles que les distances P_{β} - $O_{\beta\gamma}$, P_{β} - P_{γ} et P_{γ} - O_{H_2O} . Les résultats ne sont pas présentés ici pour raison de confidentialité, mais nous pouvons résumer les principales conclusions. Les calculs indiquent que **les mutants présentent une perte globale de stabilité du site actif, avec un gain de flexibilité du GTP et une interaction plus faible du GTP avec la molécule d'eau catalytique.** Ces conclusions sont particulièrement marquées pour le mutant Q61R, le plus agressif. Nous avons par ailleurs identifié un résidu clé, nommé X, qui n'est pas directement impliqué dans l'hydrolyse du GTP mais dont **certaines substitutions permettent de restaurer l'activité GTPase** perdue par les mutations sur site 61 (notamment en facilitant l'orientation correcte de la molécule d'eau catalytique et par effet de charges).

Nous avons construit, *in silico*, une famille de **doubles mutants Q61R-X** et montré que certains de ces systèmes présentent des **propriétés améliorées par rapport à celles de NRas Q61R**. Nos collaborateurs du CRCT ont également généré, par biologie moléculaire, le mutant Q61R constitutivement actif et les doubles mutants Q61R-X. Ces mutants ont ensuite été exprimés en cellule [46]. Les résultats ont montré un niveau d'activation plus élevé du mutant Q61R par rapport à la forme native. En revanche, un double mutant Q61R-X présente une diminution de 80% de son niveau d'activation par rapport au mutant Q61R. De plus, il n'est plus capable ni d'induire la transformation tumorale ni d'activer les voies de signalisation oncogénique dans des cellules NIH3T3, comme le fait le mutant oncogénique Q61R. Au total, ces résultats suggèrent fortement que la mutation du résidu X sur NRas Q61R empêche sa fonction transformante.

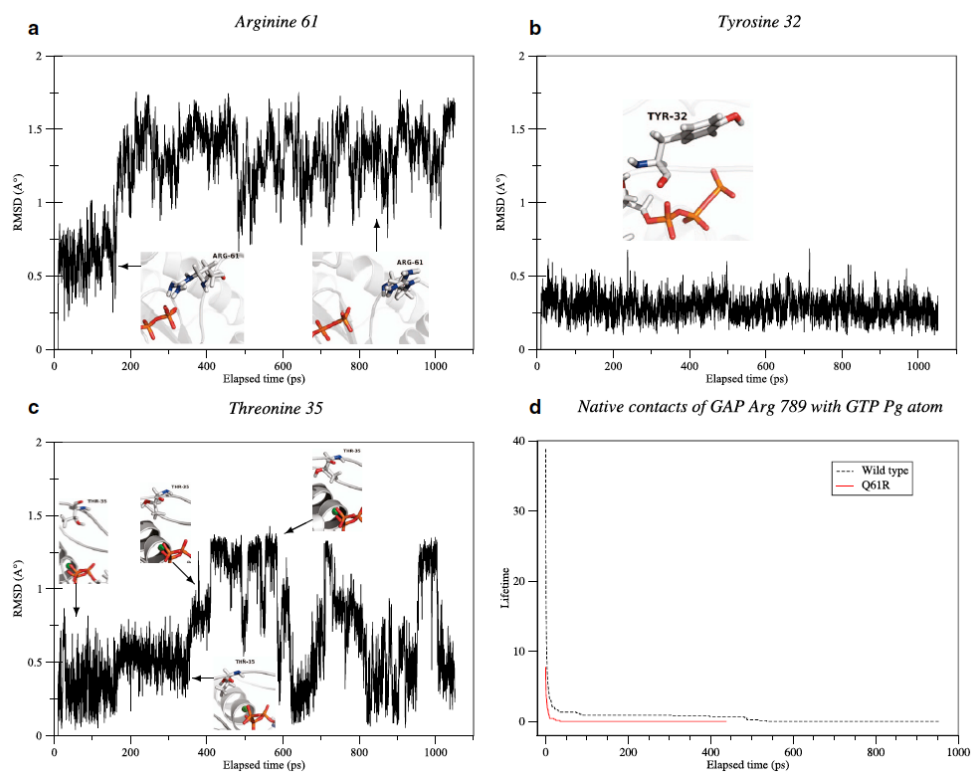
Au-delà de la **validation des prédictions** obtenues *in silico*, l'identification du site X doit permettre de **proposer une nouvelle stratégie pour cibler NRas**. C'est là qu'interviennent nos collaborateurs de l'IPBS, lesquels travaillent sur la résolution de la structure des différents mutants et procèdent au criblage haut-débit de ligands devant permettre de mimer l'effet de la mutation X. Dès lors, notre travail consiste à caractériser précisément les effets des mutations sur l'activité de NRas pour les guider dans leur tâche, ce que nous faisons au travers des études décrites ci-après et basées sur des calculs de **dynamique moléculaire QM/MM**.

Stabilité structurale du site actif. Les simulations ont montré que toutes les substitutions de Q61 induisent un réarrangement structural du site actif. La stabilité et le réseau d'interactions des résidus Y32, T35 et G60 sont affectés en fonction de la substitution. Ceci est particulièrement vrai pour le mutant Q61R, dans lequel la flexibilité de Y32 diminue, T35 devient instable et G60 interagit avec le résidu GAP T791 au lieu de former des liaisons hydrogène avec des molécules d'eau du site actif, comme c'est le cas dans la forme native. Enfin, les contacts natifs entre le résidu R789 de GAP et l'atome P_{γ} du GTP sont perdus dans tous les mutants, et des contacts non-natifs se forment avec le résidu Q61 dans la forme sauvage seulement (voir Fig. 2.16). Les résidus 35 et 60 étant conservés dans les GTPases et impliqués dans la stabilisation de la configuration catalytique du site actif, un tel changement dans la nature et la stabilité des interactions dans lesquelles ils sont impliqués doit contribuer à la baisse du taux d'hydrolyse observé pour les mutants.

Occupation du solvant au sein du site actif. Nous nous sommes intéressés dans un second temps à la distribution du solvant pour apporter des éléments nouveaux aux questions posées sur le nombre de molécules d'eau impliquées dans la réaction d'hydrolyse. Nous avons pour cela calculé les fonctions de distribution radiale à l'aide d'un code développé par R. Tichauer, lequel détermine la densité de solvant en fonction de la distance par rapport à un atome de référence. Les pics successifs sont interprétés comme décrivant les différentes couches d'hydratation. Un second code a également été implémenté pour représenter cette probabilité de présence de l'eau, mais projetée sur un plan (les détails sont donnés dans la thèse de R. Tichauer et dans l'article [218]). Cette représentation présente l'avantage de localiser les sites d'hydratation précisément dans le site actif, et plus seulement dans un rayon donné. Les résultats sont présentés sur la figure 2.17. J'en donne ici les principales conclusions.



(a) NRas forme native



(b) NRas mutant Q61R

FIGURE 2.16 – RMSD des résidus Q61 (a), Y32 (b) et T35 (c), et temps de vie des contacts natifs entre l'atome P γ (GTP) et le résidu R789 (GAP) (d) pendant les dynamiques QM/MM. Les données sont représentées pour la forme non mutée de NRas (en haut, a) et pour son mutant Q61R (en bas, b).

Pour le type sauvage, les molécules d'eau restent proches de l'atome $P\gamma$, formant un arc de densité autour du groupe phosphate, étendu entre les résidus 12 à 35 de NRas. La densité d'eau la plus élevée est localisée entre les résidus 12, 59 et 60. Les molécules d'eau sont maintenues dans cette région par des liaisons hydrogène formées avec les atomes d'azote de la chaîne peptidique au niveau des résidus G12 et G60. Un deuxième pic de densité est observé à proximité du résidu T35. La présence de deux pics, dans deux régions distinctes, indique la présence d'une molécule d'eau supplémentaire dans le site actif, en plus de la molécule d'eau lytique, pour une catalyse efficace de la réaction d'hydrolyse.

Il est remarquable de noter que **dans tous les mutants, ce positionnement précis des molécules d'eau est perdu**. Pour Q61E, Q61P, Q61L et Q61R, la délocalisation et l'excès de solvant peuvent empêcher GAP R789 d'interagir avec l'atome $P\gamma$ du GTP ainsi qu'avec le résidu 61 en raison de l'encombrement stérique. Par ailleurs, ces éléments écrantent très probablement l'effet électrostatique de GAP R789 sur NRas. Dans les mutants Q61H et Q61K, un seul pic apparaît au voisinage de T35, ce qui corrobore le besoin d'une seconde molécule d'eau pour catalyser la réaction. Comme dans la forme native, on note le rôle crucial du résidu G60 qui maintient la position de l'eau face au groupe phosphate γ du GTP. Le résidu 61 a alors un rôle essentiel mais indirect en stabilisant G60, et donc l'eau, dans une configuration pré-organisée pour la catalyse. Les substitutions de Q61, selon leur nature et le réarrangement structural induit, altèrent ce mécanisme et le pouvoir catalytique de NRas.

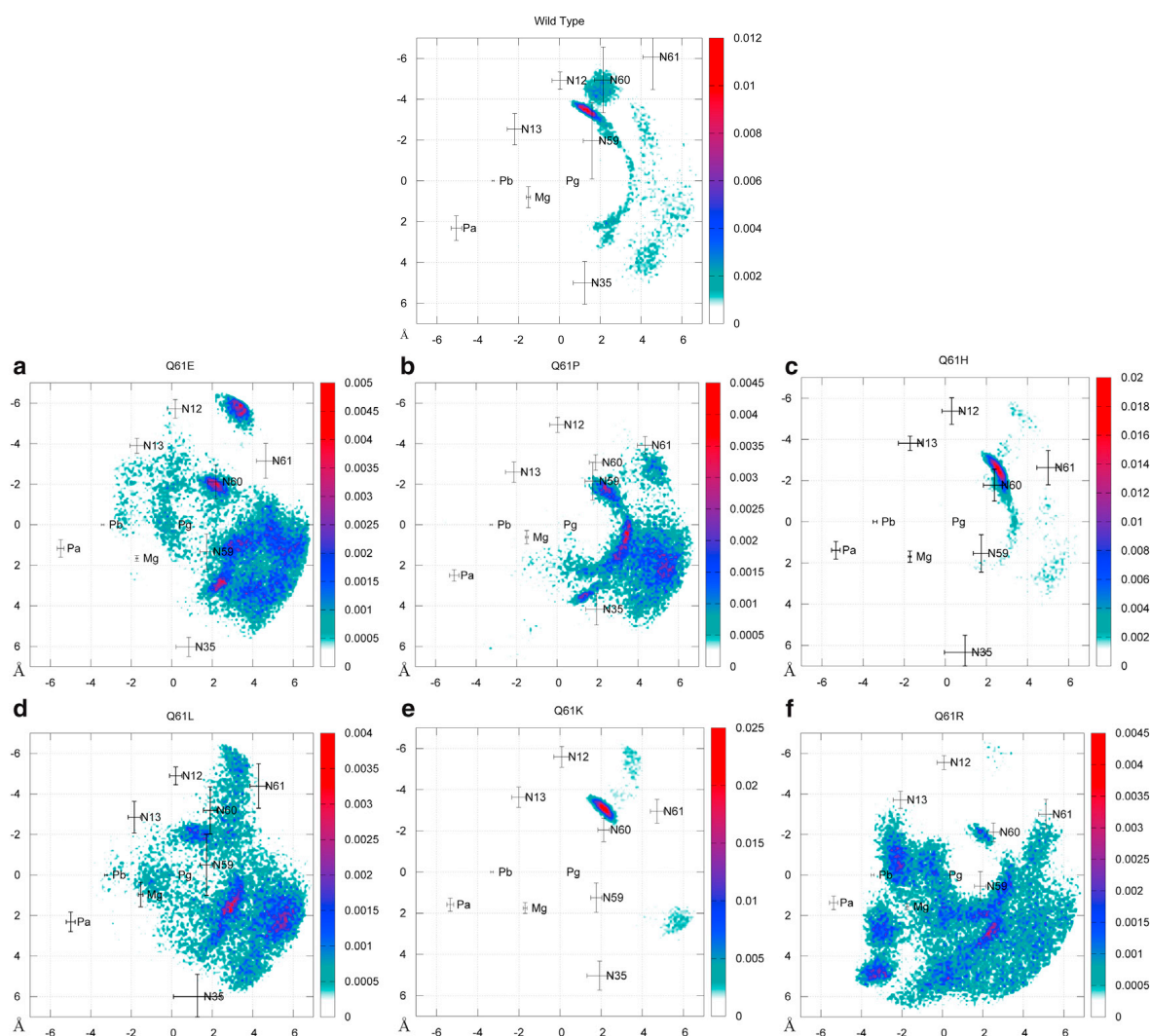


FIGURE 2.17 – Distribution radiale (projection) des molécules d'eau à l'intérieur du site actif de NRas (forme native et différents mutants). Les coordonnées moyennes des atomes d'azote du backbone des résidus d'intérêt (12, 13, 35, 59, 60, 61) ainsi que celles des atomes de phosphore du GTP et de l'ion magnésium sont aussi représentées (les barres d'erreurs correspondent aux écarts-types).

Densité électronique. Il a été déterminé expérimentalement qu'un effet catalyseur majeur consiste en un transfert de charge du groupe phosphate γ vers le groupe β [53, 190, 232]. Cet effet est induit par l'environnement fourni par NRas, renforcé par le résidu GAP Arg789. Nous l'avons évalué sur les structures les plus représentatives de NRas (native et mutées), issues du clustering effectué sur les trajectoires QM/MM. Les calculs DFT menés sur ces structures (analyse de Löwdin pour évaluer la densité d'électrons portée par chaque atome, et de l'orbitale KS HOMO) montrent que, **dans la forme NRas native, le groupe phosphate β du GTP détient plus de charge négative que le groupe γ .** Cet accord avec l'expérience suggère que la conformation du site actif que nous avons retenue, y compris la position des molécules d'eau, est bien représentative.

La même analyse réalisée pour les mutants montre que le transfert de charge est entravé par toutes les substitutions de Gln 61 considérées. En effet, la distribution de charge sur le GTP est telle que le groupement phosphate γ détient plus de charge négative que le groupement β . Par conséquent, la baisse du taux d'hydrolyse dans ces systèmes peut également être attribuée à la **perte de la stabilisation du nucléotide dans un état de type GDP normalement induite par l'environnement protéique.** Par ailleurs, on note que le cation Mg^{2+} détient une plus grande charge réduite négative dans la forme native que dans tous les mutants, ce qui va dans le sens d'études qui ont suggéré que le mécanisme réactionnel est partiellement réalisé par cet ion qui stocke temporairement les électrons du groupe phosphate γ [190].

Enfin, la représentation des orbitales KS HOMO (*Kohn-Sham Highest Occupied Molecular Orbital*) permet d'identifier l'emplacement des **électrons les plus énergétiques, et donc les plus susceptibles de réagir.** Pour la forme native de NRas, ces électrons sont distribués sur les groupements phosphate α , β , γ du GTP de telle sorte que la distribution de charge suit également la tendance déjà indiquée par l'analyse de population de Löwdin : $\alpha_{PO_4^-} > \beta_{PO_4^-} > \gamma_{PO_4^{2-}}$. Or, dans tous les mutants, sauf Q61R, ces contributions sont fortement déplacées, dans certains cas sur les molécules d'eau (Q61E, Q61L), ou encore sur les atomes d'oxygène γ . La seule exception est celle de Q61R, qui conserve la tendance $\alpha_{PO_4^-} > \beta_{PO_4^-} > \gamma_{PO_4^{2-}}$, bien que le groupe phosphate β contienne moins de charge que dans la forme native, et que le groupe phosphate γ en contienne plus au contraire. Les résultats sont illustrés sur la figure 2.18.

Influence du modèle énergétique : comparaison des dynamiques classiques et QM/MM.

Nous avons été contraints de mener des simulations de dynamiques très courtes dans l'approche QM/MM présentées jusque-là. La production a été exécutée pendant 1 ns par portions de 50 ps en raison du manque de procédure automatisée dans AMBER pour ajuster la région quantique au fur et à mesure des entrées/sorties des molécules d'eau dans le site actif. En effet, les molécules d'eau initialement présentes dans le site actif sont libres de se déplacer à l'extérieur et peuvent être remplacées par des molécules d'eau provenant du solvant. Nous avons dû interrompre régulièrement les calculs et ajuster le traitement des molécules d'eau de manière adaptative, ce qui, à notre connaissance, a été fait pour la première fois dans ce système.

Il est donc légitime de s'interroger sur le spectre de mécanismes observés et sur leur représentativité par rapport à ce que nous aurions obtenu sur des temps de calculs plus longs. Notre réponse repose sur deux arguments : selon le suivi des paramètres globaux des simulations (évolution des énergies, RMSD, etc), nous avons vérifié que l'équilibre est rapidement atteint. Par ailleurs, la prise en compte du couplage électrostatique entre le complexe et le solvant nous permet d'accéder à des propriétés que nous n'aurions pas acquises par un traitement classique, même sur des temps beaucoup plus longs. C'est ce que nous avons montré en refaisant l'ensemble des calculs en dynamique classique sur une durée 100 fois plus longue. Nous présentons rapidement le comparatif dans la suite de cette section. Ce travail est présenté en détails dans l'article [217] et concerne en particulier les formes NRas native et mutée Q61R.

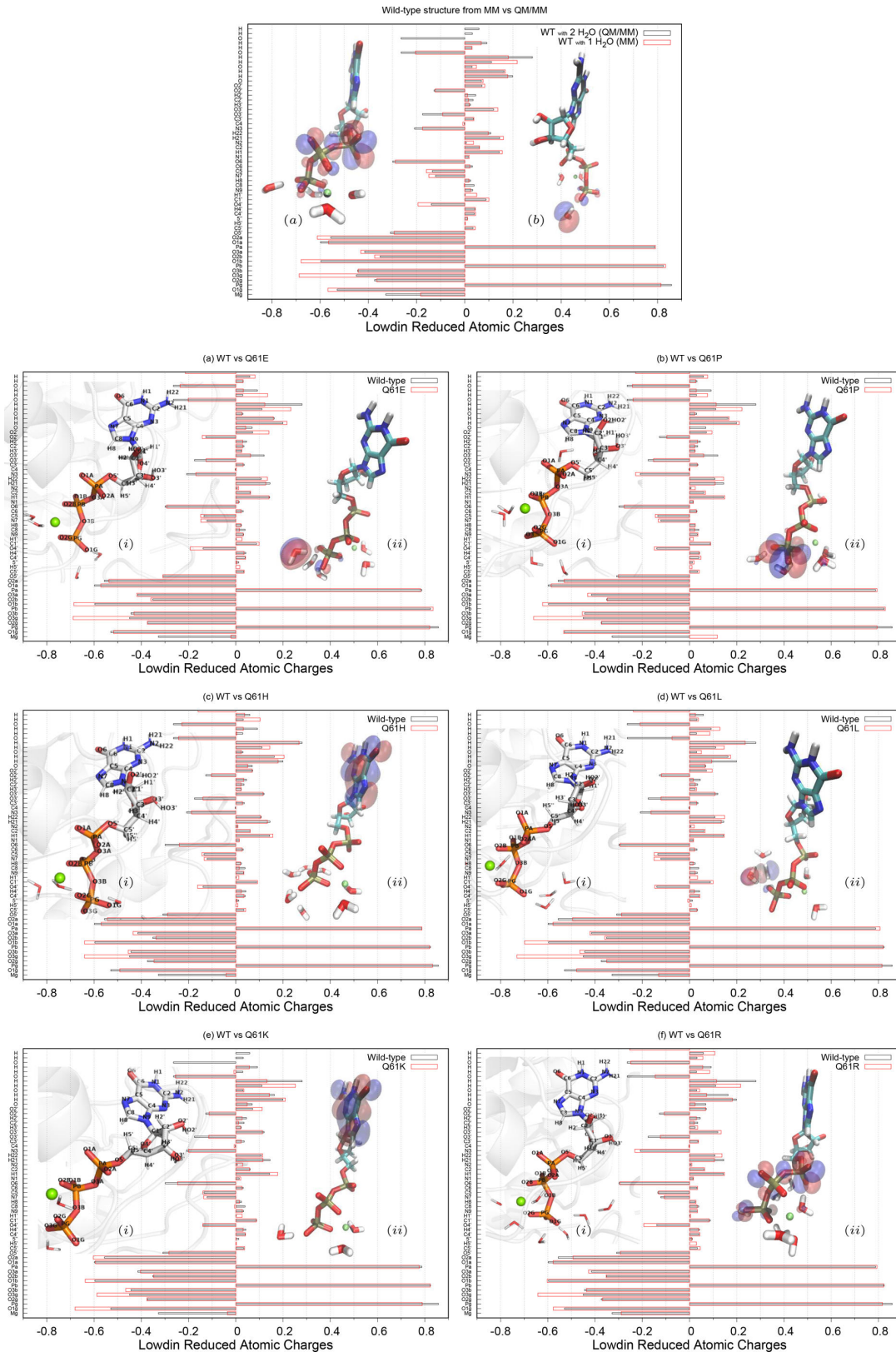


FIGURE 2.18 – Analyse de population de Löwdin du GTP, de l'ion Mg^{2+} et des molécules d'eau présentes dans le site actif de NRas. La première figure compare les résultats obtenus pour deux conformations issues des calculs QM/MM (en noir) et classiques (en rouge). Les six autres graphes ((a) à (f)) comparent, pour les structures QM/MM, les caractéristiques de NRas native (en noir) et mutée (en rouge). Sont aussi représentées les orbitales KS HOMO associées. Les lobes sont représentés en bleu si positifs et en rouge si négatifs.

Stabilité structurale du site actif. Globalement, il apparaît que, même sur un temps plus long, la dynamique classique ne permet pas d’observer certains changements de structure induits par la mutation Q61R et observés par approche QM/MM. En particulier, les modifications dans le réseau d’interactions du résidu Gly60, la destabilisation du résidu Thr35, et la perte de contacts natifs entre le résidu GAP Arg789 et l’atome P γ du GTP. Ces différences peuvent cependant être attendues, au moins en raison de la charge positive supplémentaire introduite par la mutation.

Occupation du solvant dans le site actif. Concernant la distribution radiale des molécules d’eau autour des atomes P α et P γ du GTP, des conclusions similaires peuvent être formulées dans un premier temps d’après les trajectoires de dynamique classique et QM/MM. En effet, pour NRas native et le mutant Q61R, un pic de densité d’eau se situe près de l’atome P γ . Cependant, dans un rayon de 5Å autour de cet atome, il y a moins de molécules d’eau dans la forme native que dans la forme mutée Q61R. Une différence plus marquée se profile en analysant la présence d’eau autour de l’atome P α , puisque le pic qui apparaissait en QM/MM n’est plus présent ici. L’eau est comme déportée, ce que l’on peut expliquer par une configuration différente du GTP et qui **met l’accent sur une limite de la mécanique classique**. Celle-ci fait en effet appel à des champs de forces déterminés sur molécules isolées. Les paramètres du GTP, les distances inter-atomiques en l’occurrence, ont été déterminées sur le GTP non environné et ne peuvent retranscrire ici l’effet produit par le milieu (protéine, solvant, charges). Au contraire, le traitement quantique des liaisons leur permet de s’allonger selon la distribution des charges, et de pré-configurer le GTP dans un état favorable à la catalyse. Enfin, malgré la présence de deux molécules d’eau comme vu en QM/MM pour la forme native, il s’avère qu’elles ne sont plus positionnées de la même façon et l’on peut voir sur la figure 2.19 que les résidus 12 et 60 qui étaient impliqués dans le positionnement très localisé de l’eau n’y participent plus. La même représentation associée au mutant Q61R ne montre par ailleurs pas de réarrangement du site actif en dynamique classique. En outre, la délocalisation de l’eau observée d’après les calculs QM/MM n’est plus visible ici.

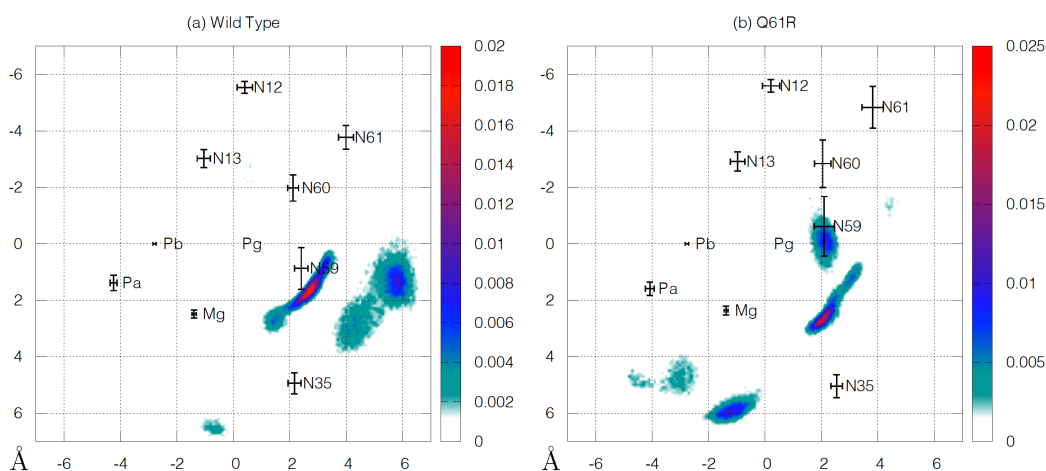


FIGURE 2.19 – Distribution des molécules d’eau dans le site actif de NRas (forme native à gauche (a) et mutée Q61R à droite (b)) d’après les trajectoires de dynamique moléculaire classique, à comparer avec les résultats QM/MM de la figure 2.17.

Densité électronique. L’analyse des différences de distribution d’eau observées entre NRas native et mutée d’après les calculs QM/MM avait suggéré l’impact direct du solvant sur la densité électronique du ligand. Nous avons également cherché à évaluer cet effet sur la structure de NRas issue de dynamique classique. Le résultat pour la forme native est donné sur la figure 2.18. Il apparaît que la distribution de charges associée à un état de type GDP n’est plus visible ici. La dynamique classique peine donc à fournir une configuration du site actif en accord avec l’expérience, et favorable à l’hydrolyse.

Ingénierie des doubles mutants. Dans cette dernière partie, nous décrivons les résultats obtenus en suivant le même protocole sur trois doubles mutants : Q61R- $X_{(a)}$, Q61R- $X_{(b)}$ et Q61R- $X_{(c)}$. Concernant la distribution de l'eau (Fig. 2.20), il est remarquable d'observer que pour les deux premiers, le principal pic de densité présent dans la forme native est retrouvé au niveau des résidus G12 et G60. Perdus dans le mutant Q61R, ces pics sont restaurés par ajout des mutations $X_{(a)}$ et $X_{(b)}$. Le second comporte même l'arche de densité, presque autant que dans la forme native. Le troisième en revanche résout moins bien l'effet de délocalisation de l'eau bien qu'elle soit tout de même plus faible que dans NRas Q61R et que le solvant soit de nouveau présent au voisinage de G12 et G60. Enfin, concernant la distribution des charges, seul le mutant Q61R- $X_{(c)}$ permet de retrouver une préconfiguration de type GDP comme nous l'avions précédemment trouvée pour la forme native (Fig. 2.21).

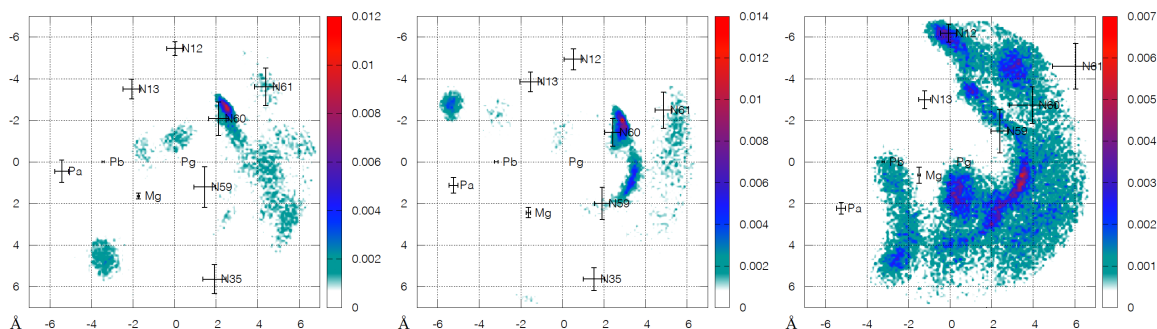


FIGURE 2.20 – Distribution des molécules d'eau dans le site actif de trois doubles mutants de NRas traités en dynamique QM/MM, respectivement Q61R- $X_{(a)}$, Q61R- $X_{(b)}$ et Q61R- $X_{(c)}$ de gauche à droite.

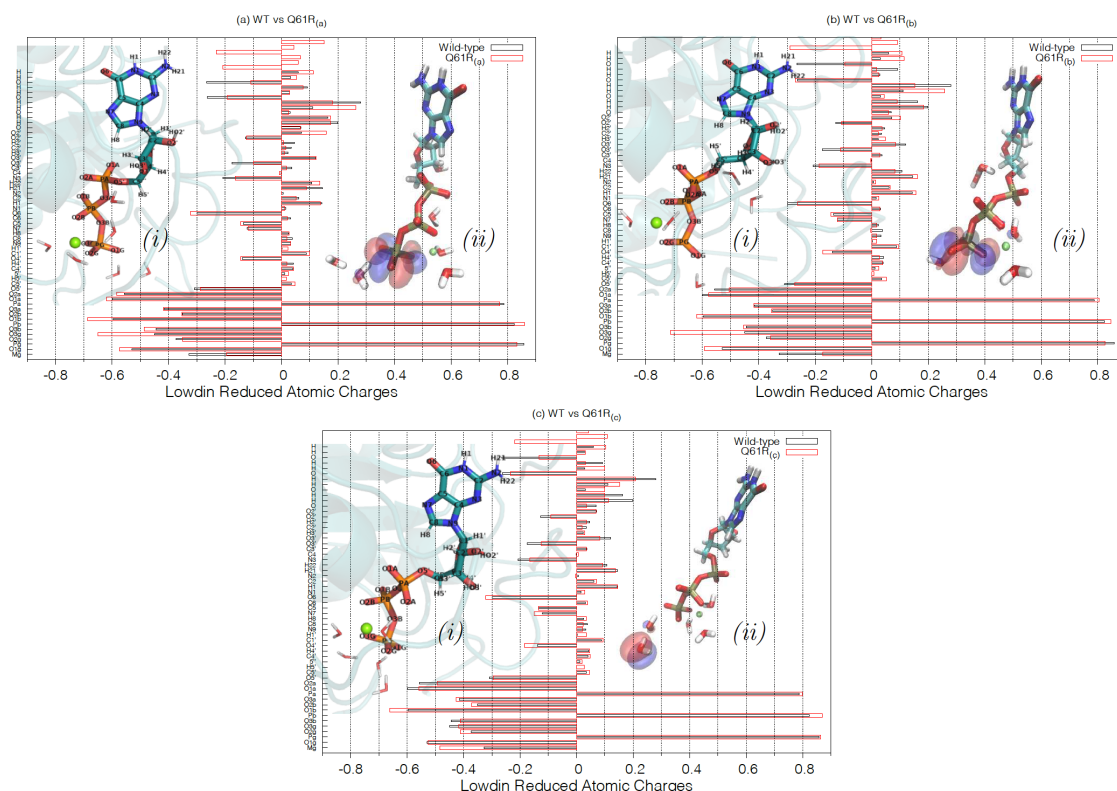


FIGURE 2.21 – Analyse de population de Löwdin du GTP, de l'ion Mg^{2+} et des molécules d'eau présentes dans le site actif des doubles mutants Q61R- $X_{(a)}$, Q61R- $X_{(b)}$ et Q61R- $X_{(c)}$ de NRas. Nous comparons ici les résultats obtenus, en rouge, à ceux précédemment obtenus pour la forme native, en noir. Sont aussi représentées les orbitales KS HOMO associées.

Perspectives

Ce travail nous a permis d'**améliorer la compréhension fondamentale des propriétés structurales et fonctionnelles de NRas** et de nous interroger sur la dépendance des résultats au modèle énergétique employé. Il a aussi permis de **proposer et d'éprouver le protocole construit avec nos partenaires biologistes**, à savoir : i) caractériser la relation structure/fonction en terme de propriétés biomécaniques auxquelles nous accédons avec les Modes Statiques et dont nous nous servons pour prédire des mutations pertinentes pour moduler l'activité ; ii) caractériser finement les éléments essentiels à la réaction grâce à une approche de physico-chimie, exigeante, mais riche de possibilités pour accéder aux effets de chaque substitution ; iii) valider les prédictions par expériences *in cellulo*.

La comparaison entre les formes native et mutées a révélé les spécificités de chaque mutant et a montré la nécessité de les prendre en compte dans les thérapies ciblées qui devront être proposées. Nos prédictions ont montré que **jouer sur la nature des substitutions nous permet ainsi de restaurer une à une les caractéristiques de la forme fonctionnelle de NRas. Ce sont autant d'éléments qui nous permettent d'améliorer la compréhension fondamentale de cette cible et de proposer de nouvelles stratégies médicamenteuses prenant en compte l'hétérogénéité moléculaire.** Ces mutations ont été testées *in cellulo* par nos collègues du CRCT et ont permis de valider le caractère prédictif de notre approche, montrant que l'hydrolyse du GTP dans NRas Q61R peut être restaurée par certaines substitutions sur le site X. Ces résultats ouvrent la voie à de nouvelles stratégies thérapeutiques pour cibler NRas Q61R, ce qui est actuellement tenté dans le cadre d'un projet Région Prématuration. Sur la base des résultats obtenus avec les doubles mutants, nous développons une stratégie de sélection de molécules inhibitrices en deux étapes qui seront assurées avec l'IPBS. La première repose sur la résolution des structures du mutant Q61R et des doubles mutants associés pour identifier des régions du mutant Q61R qui seraient modifiées par la double mutation. Ce travail donnera le rationnel pour le criblage de petites molécules. La seconde étape, déjà en cours, repose sur la sélection par criblage *in silico* de molécules qui interagissent avec les domaines protéiques identifiés et qui miment la modification de structure induite par la mutation X. Le CRCT procédera enfin à la validation pharmacologique des molécules identifiées sur les modèles de NIH3T3 qui expriment les différents mutants et sur les modèles de lignées de mélanome humain.

D'un point de vue calcul, un long travail, déjà amorcé, reste à terminer sur la simulation de la réaction d'hydrolyse au sein de NRas et de ses mutants. Par ailleurs, la même stratégie pourra être appliquée aux mutants de KRas, qui concernent principalement le site 12 et sont associés aux cancers des poumons, du pancréas et du côlon. Ce travail a également été amorcé dans l'équipe.

2.3 Modulation de l'activité enzymatique par réarrangement structural induit : apport de la dynamique moléculaire classique et hybride

Synthèse de 2',3'-cGAMP par la protéine cGAS activée par l'ADN

Cadre de travail : Ce travail a été initié par des discussions avec A. Bancaud (LAAS-CNRS), Y.-L. Lin et P. Pasero (IGH, UMR 9002 CNRS - Université de Montpellier). Il a fait l'objet du stage de J. Soler, co-encadré avec P.A Fernandes (Université de Porto), dans le cadre du Master européen TCCM (Theoretical Chemistry and Computational Modelling). Il a conduit à la rédaction de deux articles.

Résumé

L'activation de cGAS par l'ADN cytosolique induit la synthèse de 2',3'-cGAMP, un messager qui déclenche une réponse immunitaire au sein des cellules eucaryotes. Nous avons réalisé des simulations de dynamique moléculaire classique et QM/MM, respectivement pour i) valider que cGAS est activée par l'ADN en clarifiant le rôle de la boucle d'activation et ii) éclairer en détail son mécanisme réactionnel.

Ainsi, pour produire 2',3'-cGAMP, deux voies catalytiques et deux mécanismes potentiels sont proposés et analysés. Nos résultats suggèrent que la formation du produit final se fait après génération de l'intermédiaire GMP-3'-ATP. Concernant la catalyse, un anion hydroxyde est nécessaire pour générer un oxyanion réactif par un mécanisme de réaction dont le taux est en accord avec les données expérimentales. Ce travail apporte de nouveaux éléments dans la compréhension des mécanismes fondamentaux qui gouvernent l'activité chimique de cGAS, en particulier, les états de transition sont ici caractérisés et peuvent fournir de nouvelles informations pour cibler cGAS.

Contexte

Pour activer sa fonction, la protéine cGAS (synthase de GMP-AMP cyclique) se lie à l'ADN cytosolique et utilise l'adénosine triphosphate (ATP) et le guanosine triphosphate (GTP) pour produire un messenger cellulaire appelé 2',3'-cGAMP ([G(2',5')pA(3',5')p]) [80]. cGAMP se lie à l'enzyme STING (*stimulator of interferon genes*) et active ainsi une cascade enzymatique qui conduit *in fine* à déclencher une réponse immunitaire. Récemment, le rôle de la voie cGAS-STING dans la réponse anti-tumorale a été mis en évidence, faisant d'elle une cible d'intérêt pour l'immunothérapie [133]. Par ailleurs, des mutations de cGAS réduisent son efficacité catalytique et il a été montré que certaines lignées de cellules cancéreuses perdent l'expression ou la fonctionnalité de cGAS ou STING, ce qui conduit à l'inactivation de cette voie et permet la croissance et la réplication des cellules tumorales [249].

La forme humaine de cGAS contient 522 résidus. Le fragment N-terminal (res. 1-160) n'intervient ni dans la liaison à l'ADN, ni dans la catalyse, et est par ailleurs mal résolu. Il n'intervient donc pas dans nos calculs qui ne prennent en compte que le domaine catalytique (res. 161-522). Cette région contient deux lobes : le lobe N et le lobe C entre lesquels se situe un motif de liaison au zinc (incluant H390, C396, C397 et C404 qui appartiennent à la sphère de coordination de Zn^{2+}) (voir Fig.2.22). Une longueur d'ADN double brin (ADNdb) d'environ 15 paires de bases est suffisante pour activer cGAS *in vitro*, mais un ADN plus long est nécessaire pour une activation complète *in vivo* [56]. La liaison n'est pas spécifique de la séquence de l'ADN car cGAS se lie au squelette sucre-phosphate. Elle induit une rigidification du lobe N ainsi qu'un changement de conformation de la boucle d'activation (res. 210 à 220). Dans sa conformation inactive, cette boucle se déplace vers le site actif tandis que dans sa conformation active, elle interagit avec l'ADN et facilite la catalyse [244]. Les résidus catalytiques se composent de la triade E225, D227 et D319. Certaines structures cristallographiques suggèrent également la présence de deux cations assignés à Mg^{2+} et que nous avons inclus dans les calculs.

À notre connaissance, une seule voie catalytique a été proposée (Fig.2.22 [80]). Nous l'avons nommée "voie GTP-ATP" car le GTP est placé au-dessus de l'ATP. Néanmoins, nous proposons également une autre voie possible dans laquelle les positions de l'ATP et du GTP sont inversées. Nous la nommons "voie ATP-GTP" dans la suite de ce travail. Les deux voies sont explorées dans notre étude.

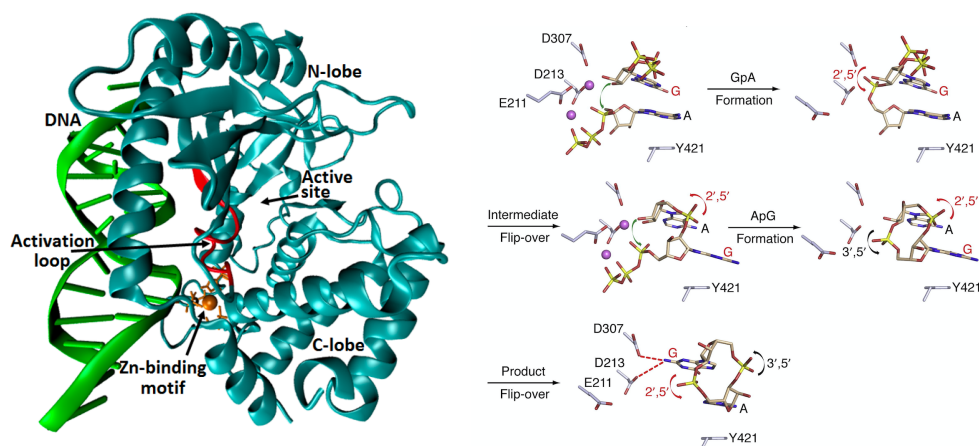


FIGURE 2.22 – A gauche: présentation de la structure de cGAS situant les deux lobes, le site de liaison à l'ADN, la boucle d'activation et le site actif. A droite : voie catalytique GTP-ATP par Gao et al. [80].

Méthodologie

Un premier effort a été mené pour générer les structures de départ. Nous avons travaillé sur deux groupes : l'apoenzyme (avec Zn^{2+} et avec/sans ADN), et les complexes de Michaelis (avec Zn^{2+} , ADN, deux ions Mg^{2+} , et les réactifs ou intermédiaires des deux voies catalytiques étudiées). Pour obtenir les géométries apo, nous avons utilisé la structure PDB 4O68 [244] comme modèle pour la structure cGAS-Zn, et 6CTA [249] pour la structure cGAS-Zn-ADN. Nous avons également utilisé le PDB 5VDP [87] pour compléter les boucles manquantes. Enfin, la structure cGAS-Zn-noDNA est obtenue après élimination de l'ADN de la structure cGAS-Zn-DNA. Pour obtenir les complexes comprenant les réactifs pour la voie GTP-ATP, le PDB 6CTA a été utilisé comme modèle (en conservant l'ADN, l'ATP et les ions métalliques). Le GTP manquant a été ajouté à partir de 4KB6 [61]). La structure des réactifs pour la voie ATP-GTP a été obtenue en échangeant les positions des réactifs dans le complexe construit pour la voie GTP-ATP. Pour l'intermédiaire de la voie GTP-ATP, le GTP et l'ATP ont été remplacés par l'intermédiaire issu du PDB 4K9A (cGAS avec 5'-pG(2',5')pA) [80]. Le pyrophosphate manquant de la guanosine est construit en utilisant les groupes phosphate β et γ de l'ATP du PDB 6CTA. Enfin, l'intermédiaire pour la voie ATP-GTP a été entièrement reconstruit en orientant l'adénosine, la guanosine et le pont phosphate 3',5' pour correspondre à celle du produit 2',3'-cGAMP de 5VDP.

Nous obtenons quatre complexes de Michaelis (avec deux réactifs et deux intermédiaires) et trois structures apo. Sur chacune, des minimisations ont été effectuées pour vérifier la stabilité et la coordination des ligands. Le logiciel Amber16 a été utilisé avec les champs de force FF14SB pour l'enzyme, BSC1 pour l'ADN, GAFF pour l'ion Zn^{2+} . Les paramètres des réactifs (ATP, GTP), des intermédiaires (AMP-2'-GTP et GMP-3'-ATP) et des ions Mg^{2+} ont été obtenus à partir de la base de données de Bryce⁶. Leurs charges ont été calculées par calculs DFT (GAUSSIAN09, B3LYP/6-31G(d)). Tous les calculs ont été menés en solvant explicite avec ajout de contre-ions pour neutraliser la charge totale. La procédure classique de minimisation, chauffage à 300K avec relâchement progressif des contraintes, équilibrage dans l'ensemble NPT a été réalisée, suivie d'une production de 500 ns dans l'ensemble canonique. Nous avons également fait appel à la dynamique moléculaire accélérée pour étudier les changements conformationnels dans la boucle d'activation [150].

Les structures initiales pour les calculs QM/MM sont extraites des trajectoires de dynamiques moléculaires classiques. La région quantique comprend les chaînes latérales de 3 résidus catalytiques (E225, D227 et D319), la chaîne triphosphate, le ribose du GTP, le ribose de l'ATP, les deux ions Mg^{2+} et les molécules d'eau qui se lient directement aux ions du site actif et au groupement phosphate α du GTP. Les atomes de la région QM sont traités à un niveau de théorie DFT B3LYP/6-31G(d). Pour améliorer la description de la densité électronique sur les points stationnaires apparaissant lors de la réaction, des *single points* ont été effectués avec MPWB1K/ 6-311++G(2d,2p) avec la correction de dispersion de Grimme (D3).

Résultats

Activation de cGAS par l'ADN. Des simulations de dynamique moléculaire classique ont été réalisées sur cGAS avec et sans ADN lié. Bien que les deux structures ne subissent pas de changement notable au cours des calculs lorsque l'on considère le RMSD global, on relève des différences locales, en particulier sur la position de la boucle d'activation. Pour quantifier l'impact de ce mouvement, le volume du site actif pendant les DM a été calculé sur dix instantanés également espacés sur les trajectoires. Les résultats sont visibles sur la figure 2.23. Les volumes moyens sont de $1351 \pm 185 \text{ \AA}^3$ et de $1876 \pm 231 \text{ \AA}^3$ sans et avec ADN respectivement, ce qui signifie que les deux simulations explorent bien deux états conformationnels. Les différences entre ces deux états apparaissent sur les structures (a) et (b) de la figure 2.23. En particulier, l'hélice 1 (res. 161-200) se déplie pour améliorer la liaison à l'ADN, augmentant ainsi le volume du site actif et ouvrant l'entrée de la cavité. En parallèle, la boucle d'activation est exclue du site actif, permettant aux substrats d'atteindre la triade catalytique.

6. <http://research.bmh.manchester.ac.uk/bryce/amber/>

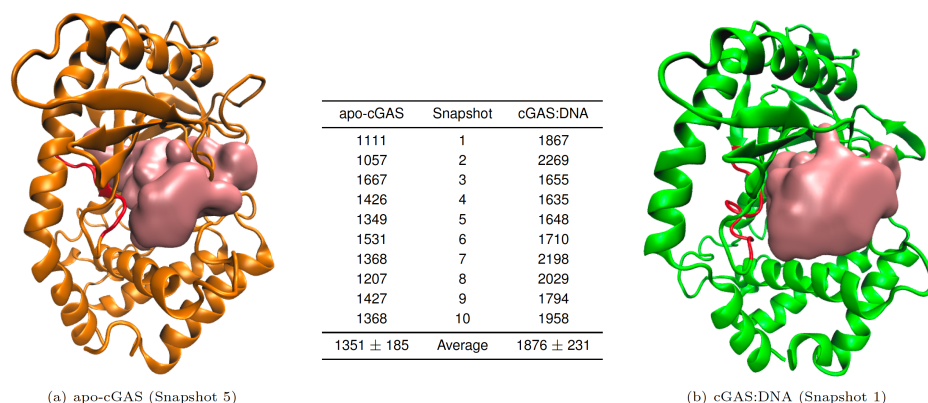


FIGURE 2.23 – À gauche et à droite sont présentées les structures de cGAS sans ADN (a) et avec ADN lié (b) (l'ADN n'est pas représenté par souci de clarté). La poche du site actif apparaît en rose et la boucle d'activation en rouge. Au milieu : les valeurs du volume de la cavité lors des dynamiques moléculaires (\AA^3).

Changement conformationnel de la boucle d'activation et implication dans la catalyse.

Les calculs précédents explorent les états inactifs et actifs de l'enzyme, sans renseigner sur la transition permettant de passer de l'un à l'autre. Pour simuler le changement de conformation, nous avons utilisé la dynamique moléculaire accélérée (aMD). Si nous partons de la structure active et que nous retirons l'ADN, la simulation aMD montre que la boucle repasse en conformation inactive après environ 500 ns, ce qui, compte tenu de l'accélération, correspond à un événement se déroulant sur une échelle de temps de l'ordre de la μs - ms . Ceci suggère que cGAS est un capteur robuste d'ADN car il n'est actif que lorsque l'ADN est présent et reste inactif lorsque l'enzyme n'y est pas liée.

Sur la figure ci-contre, je présente la superposition des formes cGAS-apo (en orange) et cGAS-ADN (en vert), pour la voie catalytique GTP-ATP. Les ions Mg^{2+} et les ligands GTP et ATP sont représentés dans la structure cGAS-ADN.

Dans la forme apo, le résidu V218 occupe les positions d'un ion Mg^{2+} et de la chaîne triphosphate de l'ATP, leur interdisant l'accès au site actif. Il occupe donc une position stratégique alors qu'il est rejeté loin de la cavité dans la conformation active, autorisant ainsi la catalyse.

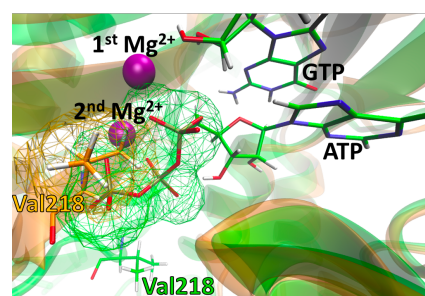


FIGURE 2.24 – Superposition des formes cGAS-apo en orange et cGAS-ADN en vert.

Mécanisme réactionnel. Dans la voie GTP-ATP, la liaison 2',5' doit être formée pour produire un intermédiaire qui pourra lui-même produire 2',3'-cGAMP. Nous nous attendons donc à ce que le groupe 2'-hydroxyle du GTP soit proche du cation Mg^{2+} qui catalyse la réaction. Dans la voie ATP-GTP, c'est la liaison 3',5' qui doit être formée pour conduire à la formation de GMP-3'-ATP. Plusieurs calculs nous ont permis de conclure que la voie ATP-GTP est favorisée. Notamment, les simulations menées sur les structures avec réactants montrent que dans le cas GTP-ATP, les deux groupes hydroxyles du ribose du GTP se retrouvent loin du premier cation Mg^{2+} , une configuration peu favorable à la réaction. Dans le système ATP-GTP, le groupe 2'-hydroxyle de l'ATP, initialement coordonné à Mg^{2+} , est réorienté par la chaîne latérale de Ser378, et est remplacé par le groupe 3'-OH dans la sphère de coordination de Mg^{2+} . Ce changement conformationnel de l'ATP améliore par ailleurs le π -stacking entre l'adénosine et la guanosine des substrats, et engage le groupe 2'-hydroxyle dans de nouvelles liaisons hydrogène qui stabilisent cette configuration. Dans la voie GTP-ATP, cette interaction Ser378 - guanosine n'est pas possible, ce qui constitue **un des arguments en faveur de la voie ATP-GTP**, laquelle mène à une conformation du site actif plus favorable à la réaction.

Des dynamiques ont également été menées avec les produits intermédiaires pour évaluer leur stabilité dans les deux voies. Les deux intermédiaires construits (2',3'-cGAMP et GMP-3'-ATP) restent correctement positionnés et coordonnés au magnésium. En revanche, l'intermédiaire AMP-2'-GTP ne peut pas être formé à partir de la configuration GTP-ATP, comme nous venons de le voir, alors que GMP-3'-ATP peut en revanche être produit par la voie ATP-GTP.

A partir de là, seule la voie ATP-GTP a été explorée en détails (ONIOM [60]) et a conduit à proposer deux mécanismes (présentés sur la figure 2.25).

(a) Le premier est un mécanisme concerté : l'oxygène O3' de l'ATP attaque l'atome P_{α} du GTP et l'atome d'hydrogène du groupe 3'-hydroxyle est transféré à D227, seule base à proximité qui peut l'accepter. Le pont phosphate entre les groupements α et β est brisé et le pyrophosphate PP_i est formé. Ici, la molécule d'eau coordonnée au premier Mg^{2+} ne participe pas directement à la réaction.

(b) Dans le second mécanisme, la molécule d'eau est remplacée par un anion hydroxyde. La première étape de la réaction consiste en la déprotonation du groupe 3'-hydroxyle. Transformé en oxyanion, il peut alors attaquer l'atome P_{α} et conduit à la formation du PP_i . Les détails des mécanismes sont donnés sur les figures 2.26 et 2.27. Après examen des énergies des états de transition et des énergies finales, il apparaît que ce second mécanisme est nettement plus favorable.

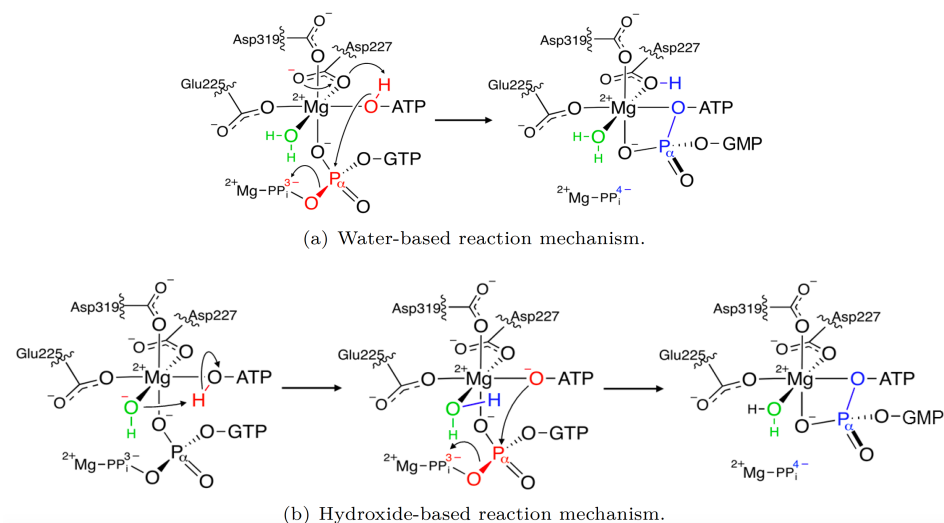
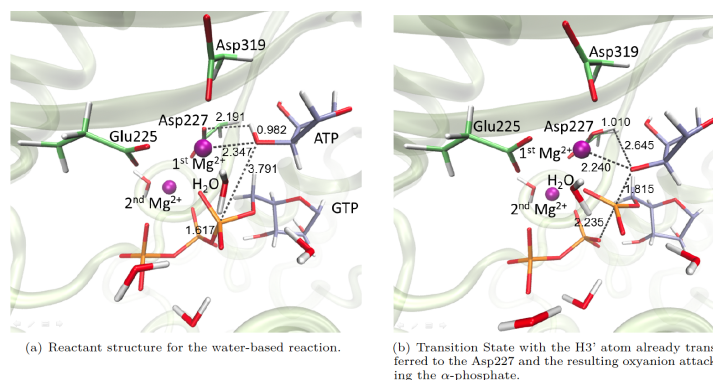


FIGURE 2.25 – Mécanismes proposés pour la voie ATP-GTP. (a) Une molécule d'eau est coordonnée à Mg^{2+} et n'intervient pas dans le mécanisme qui implique directement les substrats ATP et GTP. (b) La molécule d'eau est remplacée par un ion hydroxyde qui participe à la réaction. Les liaisons brisées et les mouvements de charges sont représentés en rouge tandis que les liaisons formées sont représentées en bleu.



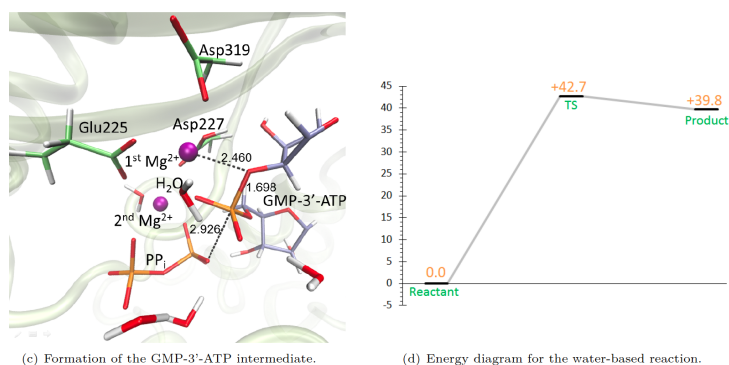


FIGURE 2.26 – Représentation des états stationnaires dans le mécanisme réactionnel (a) et diagramme énergétique associé (en kcal/mol). La triade catalytique est représentée en vert, les substrats en bleu, les cations Mg²⁺ en violet. La région QM apparaît en représentation "bâtonnets".

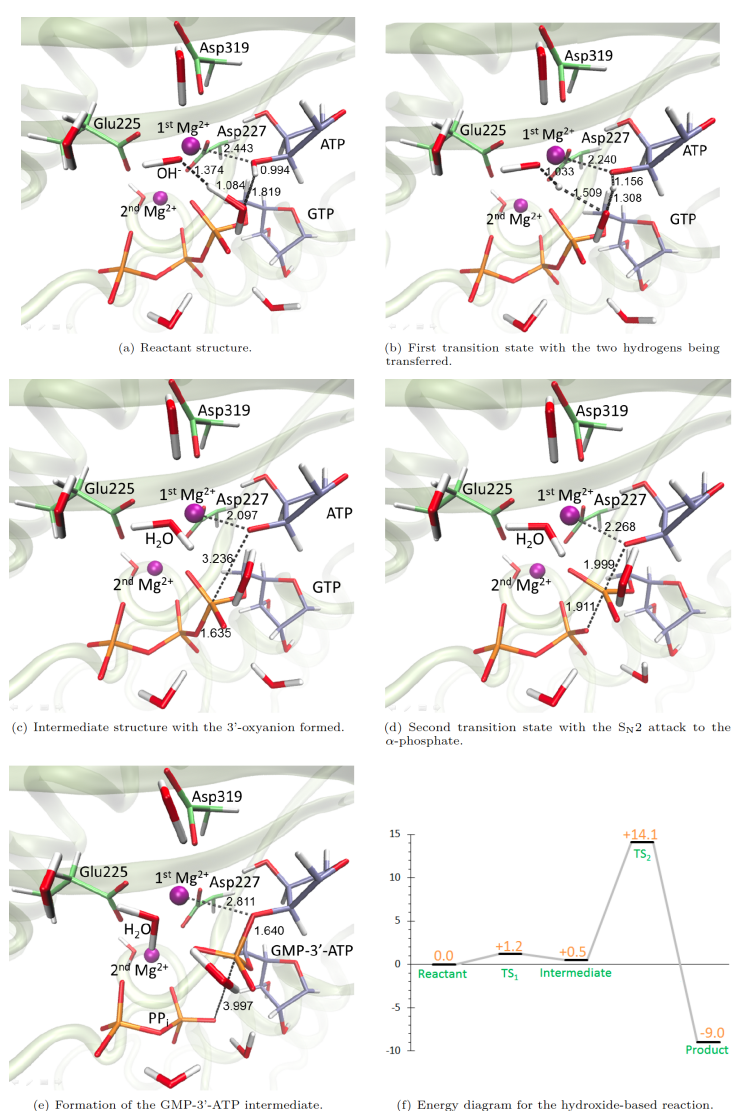


FIGURE 2.27 – Représentation des états stationnaires dans le mécanisme réactionnel (b) et diagramme énergétique associé (en kcal/mol). La triade catalytique est représentée en vert, les substrats en bleu, les cations Mg²⁺ en violet. La région QM apparaît en représentation "bâtonnets".

Perspectives

A court terme, afin de compléter cette étude, il conviendrait d'utiliser l'intermédiaire GMP-3'ATP pour étudier la formation de la liaison 2',5' qui pourrait conduire à produire 2',3'-cGAMP. Les simulations devraient également être reconduites à partir d'autres points extraits des trajectoires pour affirmer que les résultats ne dépendent pas de la géométrie et ainsi consolider notre proposition.

Au-delà, je souhaiterais appliquer à cGAS la stratégie de calcul que nous avons appliquée à Ras, c'est-à-dire envisager les mécanismes d'activation/inactivation, mais aussi les caractéristiques de l'activité enzymatique d'un point de vue mécanistique. Les Modes Statiques nous permettront d'identifier des résidus clés, dont certains ne sont peut-être pas ressortis des dynamiques, et de vérifier ces propositions par des mutations appliquées par le calcul, voire l'expérience si nous arrivons à consolider une collaboration expérimentale.

Par ailleurs, de récentes études ont permis de mettre en évidence des mutations spécifiques de cGAS menant à un dysfonctionnement de son activité à différents niveaux. Les expériences *in silico* que nous avons développées sur Ras seraient tout à fait pertinentes ici : i) d'une part du point de vue de leur impact sur la mécanique de cGAS que nous aborderons avec les Modes Statiques : ii) mais aussi du point de vue réactionnel : puisque les mécanismes enzymatiques sont bien caractérisés à présent, nous pourrions évaluer les effets des mutations sur ce plan également.

3 Les biomolécules comme objets des nanotechnologies

Au cours des dernières décennies, la capacité des chercheurs à manipuler la matière à l'échelle des atomes et des molécules, combinée aux avancées dans divers domaines (microfluidique, chimie de surface, micro / nanotechnologie, détection électronique / optique), a permis de réaliser des percées majeures dans les nanotechnologies. Renforcées par des progrès spectaculaires dans l'intégration de systèmes nanométriques et les découvertes en biologie moléculaire, ces percées se traduisent désormais par l'intégration de composants biologiques dans des nanosystèmes "bio-hybrides". S'ouvrent alors des perspectives inédites pour de nouveaux matériaux avancés et des machines synthétiques bio-inspirées, offrant des fonctionnalités uniques et flexibles, non nécessairement dédiées aux questions biologiques.

La rencontre des mondes biologiques et non biologiques soulève des questions fondamentales totalement nouvelles, appelant à des stratégies de simulation innovantes. Il existe aujourd'hui un réel besoin de modélisation, tant pour la recherche fondamentale que pour accélérer l'innovation. Nous avons vu jusque-là comment les Modes Statiques permettent d'imaginer une ingénierie *in silico* pour sonder les propriétés de systèmes complexes que sont les biomolécules. De ce point de vue, ils constituent donc un outil "sur-mesure" pour aborder la problématique de leur intégration dans des dispositifs dans lesquels elles seront détournées comme matériau structurant ou fonctionnel. C'est sur cette problématique spécifiquement que portait mon projet de recherche lorsque j'ai été recrutée au LAAS-CNRS en 2011. Les technologies ADN y étaient alors initiées au travers de l'utilisation d'aptamères, de courts brins d'acides nucléiques considérés comme d'excellents candidats pour une nouvelle génération de capteurs.

L'objectif de cette dernière section est de montrer comment nous avons utilisé les outils "traditionnels" de simulation, comme la dynamique moléculaire, mais aussi les Modes Statiques, pour traiter de nouvelles questions dans le contexte de l'intégration de biomolécules. Je décrirai en particulier les travaux que nous avons menés sur l'étude de brins d'ADN pour la détection, notamment au travers d'outils que nous avons développés pour optimiser des séquences aux propriétés choisies. J'ouvrirai également la discussion sur l'adaptation des Modes Statiques pour la description de mouvements de grande amplitude, approche que nous avons validée sur le cas de moteurs moléculaires et que nous voudrions maintenant utiliser sur les déformations mécaniques de membranes biologiques modèles.

3.1 Réponse mécanique, stabilité et reconnaissance des acides nucléiques : des outils pour comprendre et optimiser

Je décris dans cette partie les différents cas relatifs à l'utilisation de brins d'ADN que nous avons traités. Les premiers ont été abordés d'un point de vue purement mécanistique et demeurent préliminaires, avec l'objectif de comprendre la propagation de déformations imposées à leur structure. Les suivants se positionnent dans l'optique d'évaluer les effets de séquence sur la structure, la stabilité et les capacités de reconnaissance d'acides nucléiques utilisés pour la détection. Je présenterai les principaux résultats obtenus grâce à différents outils, la dynamique moléculaire pour l'étude d'interaction ADN/cible, les Modes Statiques pour évaluer la stabilité des Structures, et enfin des outils d'"ingénierie de séquence" que nous avons développés pour guider le choix des nos collègues expérimentateurs dans une procédure d'optimisation.

A. Expériences sur molécule unique : réponse de l'ADN au stress mécanique

Cadre de travail : Ce travail mené hors cadre contractuel a été initié au travers de discussions avec V. Viasnoff (UMI 3639 BMC, Singapour) et a fait l'objet d'une publication [42].

Résumé

De nombreux travaux expérimentaux cherchent à élucider les mécanismes de translocation de l'ADN (ou ARN) au travers d'un nanopore afin de pouvoir étudier les mécanismes de repliement, ou encore de l'isoler un simple-brin en vue de son séquençage. Les Modes Statiques ont été utilisés afin de recréer des forces extérieures appliquées à l'ADN et de comprendre comment des contraintes locales permettent de l'étirer et le dégrafer. Ils montrent que les forces les plus favorables sont des forces appliquées dans le plan des bases, et non des forces de cisaillement. Ils montrent également que les déformations générées se propagent le long du backbone et sont absorbées par les sucres qui accommodent la contrainte, mais qu'elles n'atteignent pas les bases pour les intensités appliquées, ce qui peut être approfondi pour caractériser des propriétés plus macroscopiques de l'ADN.

Contexte

Ce travail fait suite à des travaux très préliminaires que j'avais amorcés après ma thèse pour évaluer les déformations d'un brin d'ADN greffé sur un substrat de silicium Si(100) via une chaîne alkyle et une terminaison thiol [41]. Nous avons pour cela forcé son élongation en contraignant la position de ses groupements phosphates. Les calculs avaient alors montré que la déformation se propage le long du squelette sucre-phosphate, mais n'atteint pas les bases, étant absorbée par les sucres flexibles dont on peut observer les changements de phase. Nous avons également montré que le linker ne participaient pas à l'accommodation de l'ADN sur la surface, point particulièrement important pour comprendre les effets et donc choisir le mode de fonctionnalisation et d'immobilisation du brin.

L'application de forces locales sur l'ADN a été reproduite ici pour adresser, à l'échelle atomique, les mécanismes d'étirement. Ce phénomène est plus largement étudié à l'échelle du polymère, ce qui présente un fort intérêt, notamment pour l'étude des mécanismes de compaction/décompaction. La résolution à l'échelle atomique est cependant essentielle pour étudier les interactions ADN/protéines ou pour **explorer localement le désappariement de paires de bases** par exemple.

La translocation de biopolymères est un processus biologique essentiel et omniprésent. L'utilisation de nanopores offre une alternative efficace et peu coûteuse aux approches *in vitro* pour effectuer la translocation. Au-delà des informations qu'elle peut apporter sur les propriétés structurales et dynamiques, elle permet aussi la manipulation de molécules uniques et ouvre de nouvelles perspectives en matière de séquençage [194].

Alors que le mécanisme de translocation de l'ADN simple brin (ADNsb) est plutôt bien compris [139], le manque d'informations structurales et dynamiques rend difficile l'élucidation du processus par lequel un double brin (ADNdb) se sépare et passe par le nanopore. Des techniques de manipulation mécanique sur molécule unique ont par exemple été utilisées pour pousser l'ADN à travers le nanopore [65], ou pour forcer sa décompaction [228]. Plusieurs dispositifs existent à cette fin, par exemple les techniques de spectroscopie à force moléculaire unique (pinces optiques, pinces magnétiques, AFM). De telles expériences sont cohérentes avec l'idée que le processus de décompression peut être alimenté par une force externe, comme c'est le cas lorsque l'ADN entre dans un nanopore. Basées sur la physique des polymères, les études numériques sont un outil précieux pour tester les propriétés élastiques de l'ADN et interpréter les données expérimentales. Cependant, elles ne fournissent pas d'information précise concernant les mécanismes impliqués à l'échelle atomique.

Nous avons utilisé les Modes Statiques pour étudier comment une action mécanique peut conduire à étirer la structure de l'ADN et affecter sa conformation locale. Nous présentons une exploration dirigée des déformations d'un "27-mer" soumis à des forces imposées de l'extérieur, comme il pourrait l'être dans un nanopore. Nous montrons comment le squelette sucre-phosphate est impacté par la déformation avant que l'appariement des bases n'apparaisse, et dans quelle mesure l'initiation de la décompaction dépend de la direction de la force.

Méthodologie

Un ADNdb en structure canonique a été construit à l'aide du logiciel NAB [141]. En référence au précédent travail théorique de V. Viasnoff [31], nous avons choisi une séquence un brin de 34 bases : $5'-G_5'GCGACCTCGCGGGTTTTTCGCTATTTATGAAAAT_{3'}-3'$ sur laquelle nous n'avons apparié que les 27 bases soulignées avec le brin complémentaire. Dans la suite, nous appelons A le brin de 34 bases et B le brin de 27 bases, respectivement (Fig. 2.28, partie gauche). Cette structure a été équilibrée, à l'aide du logiciel AMBER, en solvant implicite et avec le champ de force ff99bsc0. La matrice hessienne a été extraite pour calculer les Modes Statiques. Ces derniers ont enfin été combinés pour simuler la réponse mécanique de l'ADN à différentes contraintes. Ces forces appliquées sur le squelette sucre-phosphate dans des directions définies sont représentées sur le schéma 2.28 :

cas a) forces de cisaillement sur les atomes de phosphore des extrémités 5'-terminales ; les forces sont appliquées le long de l'axe de l'hélice, selon la direction $3' \rightarrow 5'$ des brins.

cas b) forces de cisaillement sur tous les atomes de phosphore ; les forces sont toujours appliquées le long de l'axe de l'hélice, dans le sens $3' \rightarrow 5'$.

cas c) couples de forces transverses sur tous les atomes de phosphore, appliqués suivant la direction définie par l'axe reliant chaque atome de phosphore contraint à celui du nucléotide complémentaire.

cas d) forces transverses sur tous les atomes de phosphore et sur les atomes C4' des sucres.

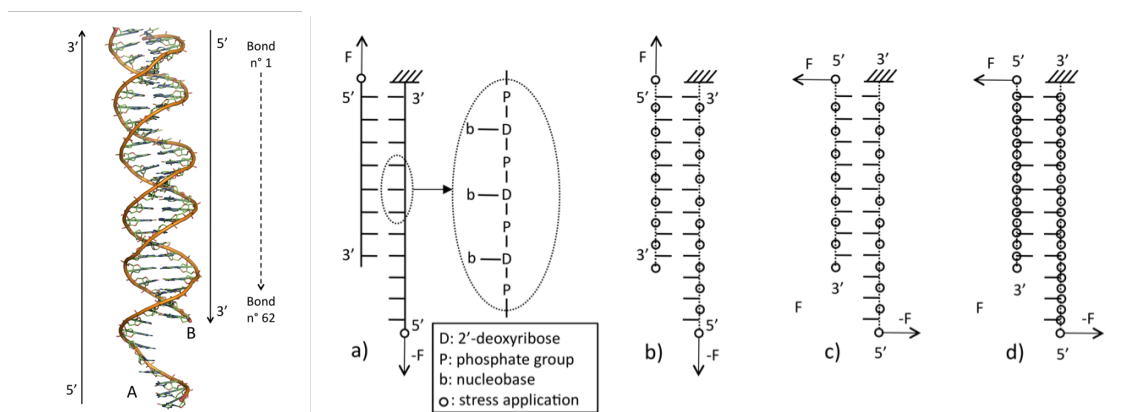


FIGURE 2.28 – A gauche: représentation de l'ADNdb, les deux brins A et B comportant respectivement 34 et 27 bases sont notés ainsi que l'ordre de la numérotation des 62 liaisons inter-brins. A droite : représentation schématique des contraintes appliquées (a) forces de cisaillement sur les atomes de phosphore des extrémités 5'-terminales ; b) forces de cisaillement sur tous les atomes de phosphore ; c) forces transverses sur tous les atomes de phosphore ; d) forces transverses sur tous les atomes de phosphore et les atomes C4' des sucres.

Comme présenté sur la figure ci-dessus, l'extrémité 3' du brin A est fixée. Les forces appliquées ont toutes une intensité fixée à $1 \text{ eV}/\text{\AA}^2$. Elles vont donc générer de petites déformations mais cette normalisation permet de comparer quantitativement et qualitativement leurs effets. Pour appliquer une force \vec{F} sur un atome N_0 , une matrice m_{N_0} ($3N \times 3$) est extraite des Modes Statiques. Elle contient donc les déplacements atomiques résultant de l'application de trois forces élémentaires appliquées selon les directions x, y, z sur l'atome N_0 . Le champ de déformation induit ΔX induit par l'application de \vec{F} est alors directement déduit de $\Delta X = m_{N_0} \vec{F}$. L'application simultanée de plusieurs forces est un simple cumul de telles combinaisons linéaires.

Les distances inter-brins sont mesurées avant et après déformation pour évaluer l'étirement des liaisons hydrogène entre bases complémentaires. Soixante deux liaisons sont dénombrées dans ce système et sont numérotées en suivant l'ordre des bases, dans le sens 5'→3' du brin B (voir Fig. 2.28). Les déformations induites permettent d'évaluer l'action mécanique induisant le plus favorablement la séparation des brins. Nous ne cherchons pas ici à désappairier les paires de bases, ce que nous ne pouvons pas faire avec cette approche, mais à évaluer l'affaiblissement des liaisons hydrogène inter-bases.

	Initial structure				Deformed structure			
	P	$\nu \text{ max}$	χ	γ	P	$\nu \text{ max}$	χ	γ
1	19.14	37.19	-156.84	58.74	26.21	40.21	-159.24	66.44
2	10.26	36.16	-154.60	58.11	2.08	30.52	-157.08	69.94
3	16.69	40.22	-150.79	57.13	17.16	40.28	-150.81	58.62
4	144.85	41.25	-98.54	57.44	144.85	41.25	-98.54	57.38

(a) Valeurs pour les quatre premiers nucléotides du brin A (5'-terminal).

	Initial structure				Deformed structure			
	P	$\nu \text{ max}$	χ	γ	P	$\nu \text{ max}$	χ	γ
1	154.65	36.68	-118.06	56.90	143.16	34.96	-119.01	54.52
2	124.60	37.60	-121.70	56.06	133.07	37.06	-124.26	61.00
3	129.17	39.79	-107.30	58.53	129.10	39.81	-107.34	59.13
4	130.37	37.13	-120.51	60.49	130.37	37.13	-120.51	60.52

(b) Valeurs pour les quatre premiers nucléotides du brin B (5'-terminal).

	Initial structure				Deformed structure			
	P	$\nu \text{ max}$	χ	γ	P	$\nu \text{ max}$	χ	γ
1	166.37	35.09	-103.25	62.86	166.37	35.09	-103.25	62.86
2	152.22	40.60	-103.47	57.66	152.22	40.60	-103.47	57.66
3	157.62	35.16	-118.18	64.17	152.62	35.16	-118.18	64.17
4	137.71	41.15	-133.83	61.06	137.71	41.15	-133.83	61.06

(c) Valeurs pour les quatre derniers nucléotides du brin A (3'-terminal).

FIGURE 2.29 – Analyse des déformations induites par l'application de forces de cisaillement sur les groupements phosphates des extrémités 5' (cas a) : on donne les variations de l'angle de phase P , de l'amplitude maximale de pseudorotation ν_{max} , et des angles de torsions χ et γ . Les valeurs sont données pour les quatre premiers nucléotides du brin A (5'-terminal) en haut, les quatre premiers nucléotides du brin B (5'-terminal) au milieu, les quatre derniers nucléotides du brin A (3'-terminal) en bas [42].

Résultats

Cas a. Lorsque les forces ne sont appliquées que sur les extrémités, seules les cinq premières liaisons, qui appartiennent aux deux premières paires de bases, sont affectées. La déformation est très locale et ne se propage pas au-delà de trois paires de bases. L'impact sur le backbone est significatif jusqu'à 15Å mais n'atteint pas les bases, la déformation étant absorbée par les sucres. Ce mécanisme d'accommodation est principalement marqué par un changement de phase des sucres [12] et des angles de torsions χ et γ qui relient respectivement le sucre à la base et le groupe phosphate au sucre [106]. Enfin, nous remarquons que ce comportement n'est pas exactement le même sur les extrémités 5' du brin A (non apparié) et du brin B (apparié). Sur ce dernier, les mêmes contraintes déstabilisent moins le backbone, rigidifié, tandis que les sucres subissent des modifications plus importantes. Ils agissent comme des pivots entre le squelette sucre-phosphate et l'empilement de paires de bases rigides, stabilisés à la fois par les forces de van der Waals et de π -stacking, ce qui se traduit par de petites variations de χ (voir Fig. 2.29). La force appliquée n'est donc pas assez forte pour prévaloir sur les interactions entre bases.

Cas b. Nous augmentons alors le nombre de forces appliquées en contraignant tous les groupes phosphates (Fig. 2.31). Ici encore, la déformation se propage le long du backbone et s'atténue en grande partie avant d'atteindre les bases. Les liaisons hydrogène présentent tout de même des variations. Les amplitudes sont hétérogènes, parfois positives ou négatives, ce qui peut s'expliquer par la structure tridimensionnelle du duplex: ce type de force peut alternativement ouvrir ou fermer les liaisons hydrogène.

Comme indiqué sur la figure 2.30, les forces de cisaillement appliquées le long de l'axe hélicoïdal induisent un étirement longitudinal. La plupart du temps, ce mécanisme étire légèrement les liaisons hydrogène, mais parfois, selon l'orientation de la base, il peut contribuer à leur contraction.

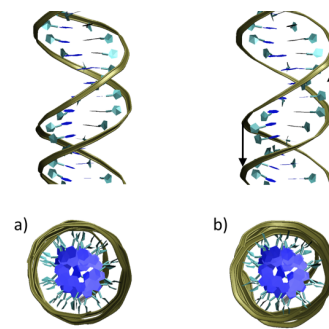


FIGURE 2.30 – A gauche, vues longitudinales et transversales de la structure initiale. A droite, vues correspondantes de la structure déformée résultant d'une contrainte appliquée sur chaque atome de phosphate, selon l'axe hélicoïdal, dans la direction 3'→5' (cas b). Le diamètre de l'hélice passe de 20 à 18.5Å [42].

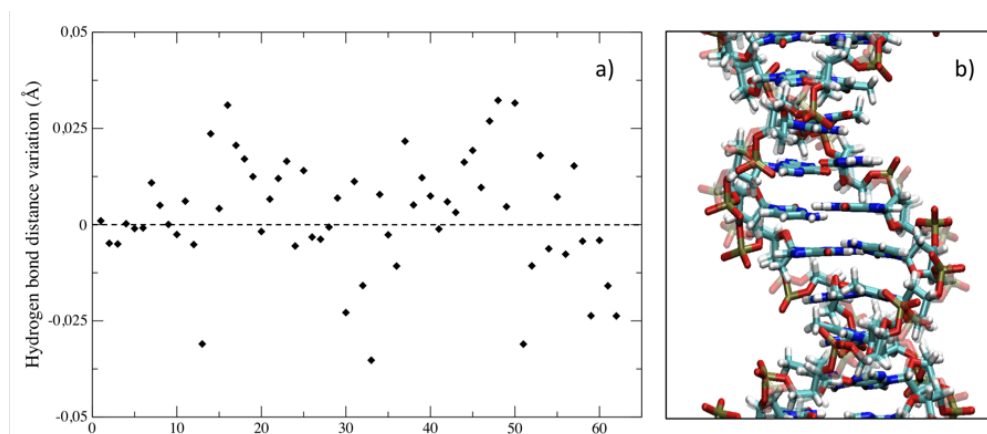


FIGURE 2.31 – Structure déformée résultant de l'application de forces sur tous les atomes de phosphate, le long de l'axe de l'hélice, dans le sens 3'→5' (cas b). Les variations de longueur des liaisons hydrogène sont données sur le graphe a). Une représentation atomique de la partie centrale de l'hélice est montrée en b) [42].

Cas c. Des couples de forces sont appliqués sur tous les atomes de phosphore, suivant la direction transversale. On crée ainsi une force d'étirement pointant vers l'extérieur du duplex, dans le plan des bases. La déformation induite est représentée sur la figure 2.33. Les amplitudes atteignent quatre fois les valeurs trouvées dans le calcul précédent et sont toutes positives. L'application d'une contrainte transversale est donc beaucoup plus efficace pour étirer les paires de bases.

Sur la figure 2.32, on observe en effet que le diamètre de l'hélice augmente avec la contrainte. Cependant, encore une fois, l'étirement inter-bases est plus petit que ce à quoi on peut s'attendre en raison de l'accommodation du sucre qui absorbe la contrainte imposée au squelette.

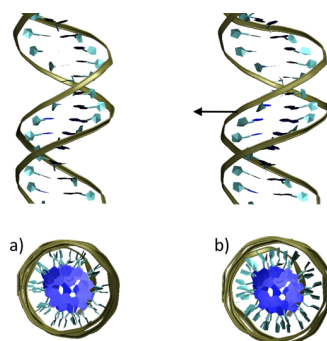


FIGURE 2.32 – A gauche, vues longitudinales et transversales de la structure initiale. A droite, vues correspondantes de la structure déformée résultant de l'application de couples de forces transverses sur les atomes de phosphore, vers l'extérieur du duplex (cas c). Le diamètre de l'hélice passe de 20 à 22Å [42].

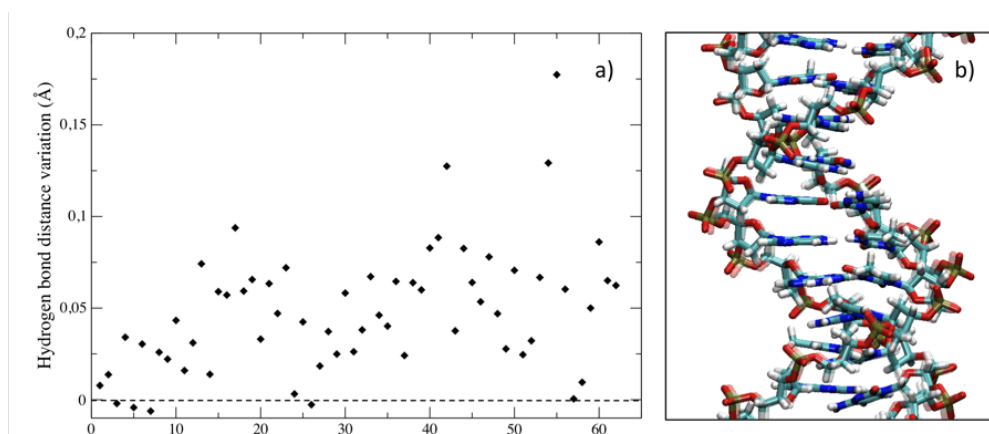


FIGURE 2.33 – Structure déformée résultant de l'application de couples de forces sur les atomes de phosphore, le long de l'axe transverse, vers l'extérieur du duplex (cas c). Les variations de longueur des liaisons hydrogène sont données sur le graphe a). Une représentation de la partie centrale de l'hélice est montrée en b) [42].

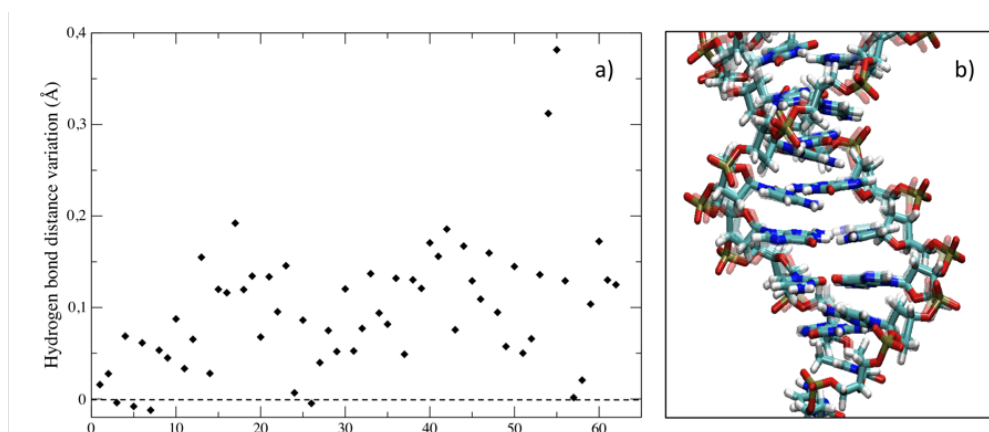


FIGURE 2.34 – Structure déformée résultant de l'application de forces transverses sur tous les atomes de phosphore et sur les atomes C4' des sucres (cas b). Les variations de longueur des liaisons hydrogène sont données sur le graphe a). Une représentation atomique de la partie centrale de l'hélice est montrée en b) [42].

Cas d. Les mêmes forces transverses que dans le cas c) sont de nouveau appliquées sur les atomes de phosphores, mais aussi sur tous les atomes C4' des sucres. Nous supposons que cette force sera trop forte pour être absorbée par les sucres et devrait atteindre les bases plus efficacement. Les déformations sont qualitativement similaires à celles obtenues précédemment, mais avec des amplitudes deux fois plus grandes qu'avec des forces appliquées uniquement sur les atomes de phosphore (voir Fig. 2.34). Il apparaît clairement que les bases ne se déplacent pas dans leur plan, l'accommodation en torsion du squelette induit un changement de phase de sucre, qui induit elle-même une rotation des bases. Ce processus explique également la disparité d'amplitudes : les bases tournant les unes par rapport aux autres, certaines liaisons hydrogène se cassent avant d'autres.

Perspectives

Dans cette étude préliminaire, **nous avons montré comment les Modes Statiques peuvent être utilisés pour concevoir des forces sur-mesure afin de déformer une molécule**, dans ce cas, pour explorer la réponse d'un ADNdb à un stress conformationnel. Nous montrons comment les forces de cisaillement le long de la direction de l'hélice sont absorbées par le squelette flexible sans atteindre les bases, et comment les forces d'étirement transversales peuvent plus favorablement conduire à affaiblir les liaisons inter-brins. Ces calculs exploratoires fournissent des **informations qualitatives** et ouvrent la voie à des investigations plus poussées, utilisant des forces plus complexes traduisant un environnement lui-même plus complexe et plus réaliste.

Sur de tels systèmes, la prise en compte des effets anharmoniques peut être particulièrement importante. Ceci peut être fait soit par l'utilisation de constantes de force empiriques dérivées des états de transitions, soit par l'utilisation de **coordonnées généralisées**, telles que les torsions. Cette seconde perspective est particulièrement intéressante car elle permet de limiter le nombre de degrés de liberté, d'accéder à des déformations de grande amplitude sur des systèmes de grande taille. Je reviendrai sur cette approche en partie 3.2, dans laquelle elle est appliquée à des moteurs moléculaires.

B. Optimiser les acides nucléiques pour la détection

Cadre de travail : Cette thématique regroupe plusieurs projets sur lesquels nous avons travaillé durant la décennie 2010. Un premier volet aptamères a été développé dans le cadre de collaborations internes au LAAS, plus particulièrement avec les travaux de thèse de A. Trapaidze (dir. A-M. Gué) sur le développement d'un capteur de thrombine. Une seconde partie porte sur l'ingénierie de séquence d'acides nucléiques pour la détection de marqueurs cancéreux, en accompagnement des travaux de thèse de R. Malbec et J. Cacheux. Ces travaux se sont appuyés sur deux stages niveau Master (E. Perraud et K. Mlih) et ont mené à la publication de 4 articles [44, 47, 222, 223].

Résumé

Ces travaux visent à proposer une modélisation prédictive englobant à la fois les aspects théoriques et technologiques liés à la problématique de l'intégration de biomolécules. Ils explorent des mécanismes fondamentaux afin de lever les barrières technologiques et d'orienter les choix expérimentaux. Je décris dans un premier temps les principaux résultats obtenus sur des aptamères utilisés au LAAS pour des applications bio-puces. Nous avons utilisé l'approche par Modes Statiques afin d'évaluer l'impact de modifications structurales imposées sur leur stabilité structurale. Nos calculs ont notamment montré qu'une fonctionnalisation par des spacers usuels (PEG, poly-T, etc) perturbe les propriétés mécaniques, suggérant un possible effet sur leurs propriétés de reconnaissance. Nous avons également montré, par des calculs de dynamique moléculaire, que les aptamères étudiés présentent plusieurs modes de liaison à leur cible, pouvant également affecter les performances attendues du capteur. Dans un second temps, je présente plusieurs codes que nous avons développés en étroite collaboration avec nos collègues expérimentateurs. Ces codes permettent de prédire des séquences d'acides nucléiques pour la conception de capteurs, séquences qui ont été testées et validées.

Contexte

Dans les deux applications développées dans cette partie, l'ADN est utilisé pour ses propriétés d'hybridation et de reconnaissance. Nous avons tout d'abord travaillé sur **les aptamères**, de courts brins d'acides nucléiques, ayant la capacité de se lier à diverses molécules avec une grande **affinité et spécificité**. Pour cette raison, ils ont été présentés pour promouvoir une **nouvelle génération de capteurs**, mais le principal verrou de leur mise en œuvre repose sur leur sélection *in vivo* (procédé SELEX [119]). Celle-ci ne tient pas compte des modifications chimiques requises pour leur intégration et de nombreuses questions peuvent être soulevées, particulièrement en ce qui concerne la conservation de leurs propriétés une fois intégrés dans un dispositif. A ce niveau, les inconnues sont nombreuses, notamment au niveau du choix des aptamères pour une fonction donnée et de leur devenir dans un environnement qui va *a priori* les modifier, voire les dénaturer (milieu, fonctionnalisation, greffage).

Dans ce cadre, notre objectif est d'améliorer la compréhension des propriétés structurales et fonctionnelles des aptamères. Très peu de calculs ont été menés avant ce travail sur ces systèmes et **aucun d'eux ne prend en compte les différents éléments qui leur sont ajoutés lors de leur fonctionnalisation. Celle-ci est pourtant essentielle pour passer d'une molécule à un dispositif de capteur**. Citons parmi les éléments les plus importants : la surface, les spacers, les systèmes de détection (fluorophores...).

Aucun n'est jamais pris en compte lors de la sélection et l'on ignore si les propriétés de l'aptamère en sont modifiées, et à quel point. Cette incertitude pourrait pourtant expliquer que l'on ne puisse reproduire la plupart du temps les propriétés de sélectivité et de spécificité des aptamères en laboratoire. **Afin de guider et d'optimiser les choix des technologues, de nouveaux outils de calculs sont indispensables pour créer une ingénierie *in silico* des acides nucléiques, afin de les modifier à volonté et de mesurer les effets de ces modifications**. Parmi les questions soulevées, nous nous sommes intéressés en particulier aux points suivants : i) influence de l'environnement sur le repliement de l'aptamère, ii) choix du dispositif de détection et effets sur l'aptamère, iii) mécanisme d'interaction aptamère/cible, iv) spécificité du capteur et la durée de vie de l'interaction.

Nous avons enfin travaillé sur la prédiction de balises moléculaires (*molecular beacons*), des courtes séquences d'acides nucléiques dont on utilise les propriétés d'hybridation (*i.e.* un ADN/ARN simple brin peut reformer spontanément une double hélice lorsqu'il est en présence d'un brin complémentaire). Ce mécanisme peut ainsi être utilisé pour détecter les ARN circulants, notamment marqueurs de cancers, ce qui a été fait au LAAS par A. Bancaud, P. Joseph, T. Leïchlé et leurs étudiants.

Méthodologie

Différentes méthodologies ont été employées pour caractériser les acides nucléiques étudiés. Nous avons tout d'abord fait appel à la mécanique moléculaire (minimisation et dynamiques), toujours avec le logiciel AMBER, afin d'obtenir les structures équilibrées des brins utilisés comme sondes, d'évaluer leur stabilité avec les Modes Statiques, mais aussi de calculer les énergies d'interaction avec la cible.

Nous avons par ailleurs développé des codes pour prédire des séquences répondant à des cahiers des charges spécifiques et fournis par nos collègues technologues. Ces codes ont notamment permis de prédire des séquences permettant la fonctionnalisation d'aptamères ou de balises moléculaires à partir de critères sur lesquels je reviendrai plus en détail par la suite. Ils ont également permis de choisir de façon optimale des séquences de balises devant s'apparier à la cible. Les différentes séquences proposées par nos approches combinatoires ont été triées par énergies ou températures de fusion calculées d'après le modèle empirique de Santa Lucia [113].

Résultats

Caractériser et optimiser la fonctionnalisation

1. Aptamères. Nous avons commencé ce travail par l'étude d'un aptamère modèle, sélectionné pour se lier à la thrombine. Il s'agit d'un 15-mer de séquence (5'-GGTTGGTGTGGTTGG-3') dont la structure et les sites de liaison sont bien connus [172]. Dans ses formes liées et non liées, il se replie par empilement de deux G-quartets connectés par deux boucles TT et une boucle TGT (Fig. 2.35(a)). Nous avons étudié la stabilité structurale de l'aptamère avec différents espaceurs usuels : poly-T, PEG, chaîne alkyle. Nous nous sommes intéressés à la boucle TGT pouvant se lier à l'exosite de thrombine 1 (Fig. 2.35(c)). Nous avons en particulier considéré la distance $d_{T7-dG10}$ pour évaluer la courbure de la boucle, en calculant les déformations induites en contraignant chaque atome grâce aux Modes Statiques. Ce critère nous permet d'anticiper un changement conformationnel pouvant affecter la liaison à la cible et reflète la rigidité des quartets et donc la stabilité globale de l'aptamère.

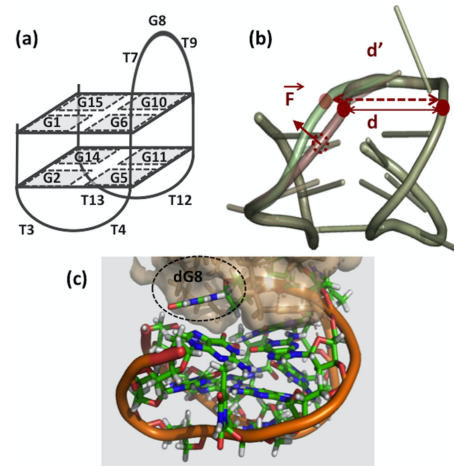


FIGURE 2.35 – (a) Représentation schématique de l'aptamère anti-thrombine. (b) Exemple de déformation locale induite. (c) Surface d'interaction entre l'aptamère et l'exosite 1 de la thrombine [44].

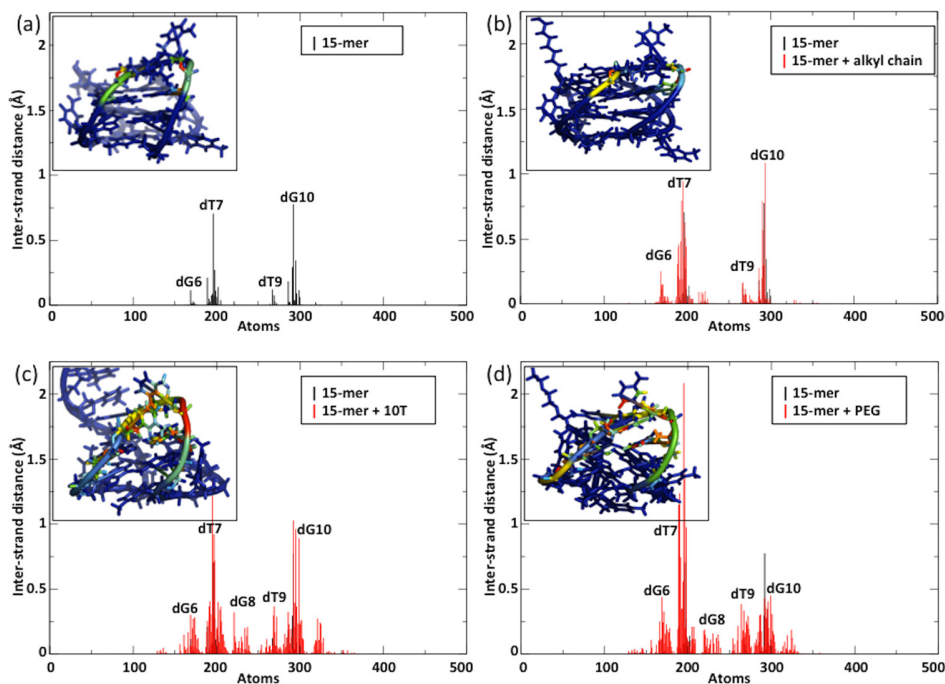


FIGURE 2.36 – Représentation des variations de distance $d_{G7-dG10}$ induites par l'application d'une force extérieure sur chaque atome (en abscisse) par la méthode des Modes Statiques. La réponse de l'aptamère canonique (a) est comparée à celle du même aptamère fonctionnalisé par une chaîne alkyle (b), une séquence de 10 thymines (c) et du PEG (d). On représente sur la structure de l'aptamère la contribution de chaque atome avec un code couleur allant du bleu vers le rouge par amplitude de déformation croissante [44].

Les forces sont d'abord appliquées sur l'aptamère non-modifié (Fig.2.36(a)) (structure PDB 1HAP, équilibrée avec AMBER en solvant implicite avec le champ de force ff99bsc0). Deux pics principaux sont détectés, correspondant aux atomes de dT7 et dT10, et plus précisément au squelette sucre-phosphate. Dans une moindre mesure, on observe également la contribution des proches voisins dG6 et dT9. Par conséquent, les forces appliquées pour déstabiliser la boucle TGT sont localisées sur cette même zone et induisent des déformations faibles, ce qui traduit la forte stabilité de l'aptamère.

Des espaceurs sont ensuite ajoutés à la structure initiale qui est soumise à la même procédure. Sur la figure 2.36(b), on observe que l'ajout d'une chaîne alkyle, neutre, n'induit pas de modification. En revanche, l'ajout d'un simple brin d'ADN (10-T) modifie drastiquement la réponse de l'aptamère (Fig.2.36(c)). Les pics correspondant à dG6 et dT9 voient leur amplitude augmenter, de nouveaux pics apparaissent au niveau de dG8 et dG11, et d'autres contributions apparaissent encore, notamment au niveau de dG5. Par rapport à la structure canonique, nous observons donc que les excitations distantes induisent une déformation de la boucle TGT et conduisent à une perte globale de stabilité. La plupart du temps, ces excitations doivent être appliquées sur le squelette sucre-phosphate, à l'exception de dG6, dT7 et dT9, dans lesquels les bases sont également impliquées dans la stabilité de la boucle. Ceci peut s'expliquer par le changement conformationnel subi par l'aptamère : la première thymine de l'espaceur est empilée avec la base de dG8. Pour atteindre cet état, l'extrémité 3' est courbée vers dG8, conduisant à la distorsion et à la déstabilisation du G-quadruplex. Un comportement similaire est obtenu après ajout de PEG. Nous notons cependant une forte asymétrie entre les pics dT7 et dG10. Les forces de répulsion électrostatique entre le PEG et l'aptamère induisent une déformation de la boucle TGT et de l'aptamère entier. Dans cette configuration, dT7 est moins contraint stériquement tandis que dG10 l'est plus fortement (tous les détails de ce travail sont présentés dans l'article [44]).

En utilisant les Modes Statiques, nous avons étudié la stabilité structurale du 15-mer anti-thrombine seul et lié à différents espaceurs usuels pouvant être utilisés pour le greffer sur une surface. Nous montrons que des modifications, comme l'ajout de PEG, peuvent drastiquement modifier les propriétés d'un aptamère qui a été sélectionné seul pour se lier à une cible. **Le choix des fonctionnalisations doit donc être étudié, et peut l'être par ce type d'approche, pour maintenir les propriétés du capteur une fois intégré dans un dispositif technologique.**

2. Intégration des aptamères dans un capteur. Dans une seconde partie de ce travail, nous avons accompagné le développement d'un capteur "sandwich", toujours pour la détection de la thrombine. L'idée est d'utiliser deux aptamères ciblant deux sites de la thrombine et de les relier l'un à l'autre pour capter la thrombine "en sandwich". Cibler deux sites d'une protéine simultanément permet d'augmenter la spécificité, la sélectivité et l'affinité, en particulier dans des milieux complexes, comme le plasma sanguin. Nous avons développé une méthode pour concevoir la séquence d'une construction d'ADN dédiée à la détection réversible de la thrombine. Cet "aptadimère" est l'assemblage de deux aptamères connectés par un linker d'ADN qui contient un couple FRET greffé sur une double hélice réversible. En l'absence de cible, l'hélice ("tige") est stable en maintenant un couple FRET à proximité, donc capable de fluorescer. Lorsque la cible vient se lier, par compétition, elle induit la déshybridation de l'hélice et donc l'arrêt de fluorescence qui peut être détectée.

Nous avons défini des règles de conception et développé un **logiciel pour prédire une séquence optimisée du linker**. Les règles ont été listées en suivant le cahier des charges suivant : i) la longueur du linker doit être suffisante pour permettre à l'aptadimère "d'envelopper" la thrombine pour atteindre les deux sites de liaison ; ii) le linker ne doit pas présenter de complémentarité avec les deux aptamères et contient plus de thymines que d'autres bases pour maximiser sa flexibilité ; iii) le linker doit présenter des complémentarités internes associées à une température de fusion pouvant être choisie (en fonction de la compétition avec la liaison des aptamères à la cible) ; iv) le quencher et le fluorophore sont greffés de sorte à se faire face lorsque l'hélice est formée et doivent être placés sur des thymines. Pour adresser la première règle (i), nous avons analysé la structure cristallographique du complexe thrombine/aptamères [122]. La distance mesurée entre les extrémités des aptamères (7,3nm) et la courbure associée au diamètre de la thrombine ($\sim 4,2$ nm) nous ont permis d'estimer une longueur du linker de l'ordre de 10nm, soit au moins 15 nucléotides [57]. La complémentarité des bases (ii) au sein de l'hélice a été sélectionnée pour garantir une température de fusion de 40°C environ.

Cette condition est obtenue pour 5 paires de bases [113], ce qui correspond à une différence d'énergie libre de $\sim 4\text{kcal/mol}$. Cette valeur est nettement en dessous des énergies que nous avons estimées pour les interactions thrombine/aptamères ($\sim 140\text{kcal/mol}$, voir paragraphe suivant), ce qui garantit la destabilisation de l'hélice en présence de la cible.

Nous avons enfin choisi de séparer par une boucle flexible de 14 thymines les deux brins complémentaires, eux-mêmes séparés des aptamères par des linkers de 8 thymines. La séquence résultante est présentée sur le figure 2.37(a). Suivant ces prédictions, l'aptadimère a ensuite été conçu et caractérisé expérimentalement par fluorimétrie différentielle à balayage. Les résultats (Fig.2.37(b)) ont démontré la pertinence de la technologie développée. Ses performances ont été évaluées avec différents titrages de thrombine, en PBS et en sérum murin, les résultats sont présentés dans l'article [223].

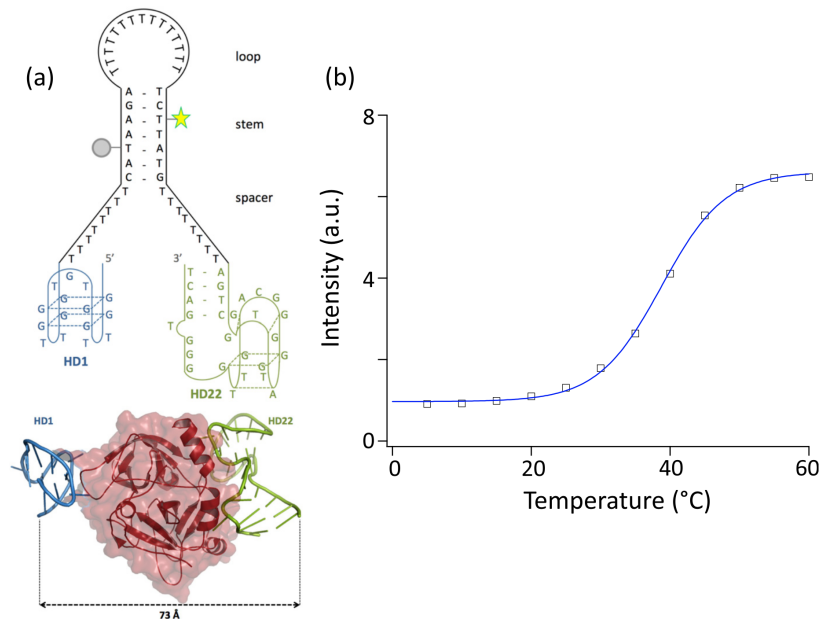


FIGURE 2.37 – (a) Séquence de l'aptadimère, composé de deux aptamères (HD1 et HD22), deux espaceurs, une tige et une boucle. Les positions du fluoroporteur et du quencher sont indiquées par une étoile et un cercle. (b) Signal de fluorescence de l'aptadimère mesuré en fonction de la température et centré à 39°C [223].

Caractériser et optimiser les interactions capteur/cible

1. Aptamères. Toujours dans le cadre du projet de développement d'un capteur "aptadimère" anti-thrombine, nous avons étudié la liaison des aptamères HD1 et HD22 à la thrombine. Nous avons pour cela utilisé des simulations de dynamique moléculaire (AMBER, solvants implicite et explicite, champ de force ff99sb) pour affiner les structures des différents complexes. HD1 est le 15-mer précédemment étudié avec les Modes Statiques ($5'$ -GGTTGGTGTGGTTGG- $3'$). Sa structure est bien connue depuis les années 1990 [30, 140]. En revanche, il a été identifié comme se liant au site fibrinogène de la thrombine, aussi appelé exosite 1, selon différentes orientations au travers des boucles TGT et TT. Il semble donc que plusieurs complexes puissent se former et coexister. Ce point est important à vérifier et à prendre en compte car certaines de ces orientations peuvent rendre impossible la liaison de HD22. En revanche, une seule structure de complexe HD22/thrombine a été identifiée. HD22 est un 29-mer ($5'$ -AGTCCGTGGTAGGGCAGGTTGGGGTGA- $3'$) se liant au site héparine, aussi appelé exosite 2, et ce avec une plus grande affinité que HD1.

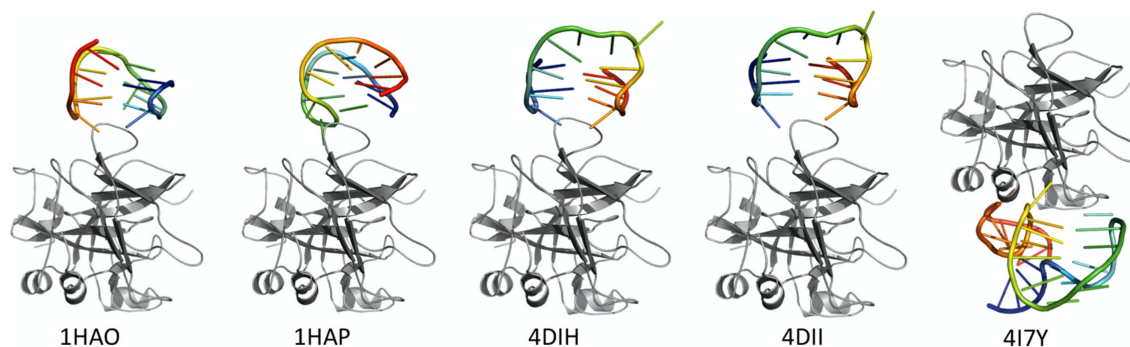


FIGURE 2.38 – Structures des quatre complexes HD1/thrombine (PDB 1HAO, 1HAP, 4DIH, and 4DII) et du complexe HD22/thrombine (4I7Y) évaluées dans les calculs. La thrombine est représentée en gris, les aptamères sont colorés en spectre rouge vers vert pour indiquer l'orientation de séquence [222].

Dans cette étude, nous avons considéré toutes les structures disponibles dans la Protein Data Bank, à savoir, quatre complexes thrombine/HD1 et un complexe thrombine/HD22 (Fig.2.38). Les principales caractéristiques structurales relatives à HD1 sont les suivantes. Dans 1HAO, les bases de dG8 et dT9 de la boucle TGT sont empilées sur le G-quadruplex, tandis que celle de dT7 est rejetée en dehors de la boucle. L'interaction avec la thrombine est établie par les deux boucles TT. Dans 1HAP, la boucle TGT est dans la même configuration mais elle interagit directement avec la thrombine.

Les structures 4DIH et 4DII ont été obtenues plus récemment, avec une meilleure résolution, et en présence d'ions Na^+ et K^+ respectivement. Dans les deux cas, l'interaction avec la thrombine se produit par les boucles TT. Cependant, la boucle TGT a une configuration différente car les bases de dT7 et dG8 sont empilées sur le G-quadruplex alors que celle de dT9 est rejetée à l'extérieur. L'orientation de HD1 sur la thrombine est par ailleurs tournée de 180° autour de l'axe de l'hélice. Enfin, l'autre différence significative vient du cation qui est inséré entre les G-quartets et augmente la distance inter-brins de 12\AA dans 1HAO et 1HAP à 21\AA dans 4DIH et 4DII.

Dans un premier temps, les quatre structures de HD1 et celle de HD22 ont été extraites et leurs énergies évaluées. Les énergies de HD1 sont toutes similaires, sur une gamme allant de -3.15kcal/mol à -3.02kcal/mol . A ce stade, on ne peut donc dire si une conformation prévaut sur les autres. L'énergie de HD22 avoisine en revanche le double, soit -5.98kcal/mol , ce qui s'explique par une chaîne plus longue, des complémentarités inter-bases plus nombreuses, et donc une plus grande stabilité conformationnelle.

Dans un second temps, les énergies de liaison aptamères/thrombine ont été estimées d'après les trajectoires de dynamique moléculaire, avec la méthode MM-PBSA (*Poisson-Boltzmann surface area*, détails donnés dans [222]). Les énergies de liaison des complexes 1HAO et 1HAP sont très proches (-47.03 et -48.81kcal/mol), mais environ deux fois plus faibles que celle du complexe HD22/thrombine (-88.37kcal/mol). En revanche, en présence de cations, ce qui est plus pertinent en conditions expérimentales, les complexes HD1/thrombine gagnent en stabilité avec des énergies de -60.29 et -66.73kcal/mol . **Les différentes structures du complexe HD1/thrombine sont énergétiquement similaires et peuvent donc coexister. Inversement, le repliement de HD22 est beaucoup plus stable et son énergie de liaison avec la thrombine est significativement plus importante que celle des complexes HD1.**

Ces résultats ont été confrontés à des expériences consistant à suivre l'agrégation conduite par la thrombine de nanoparticules d'or fonctionnalisées par des aptamères. HD1 seul, mais pas HD22, peut déclencher l'agrégation, ce qui signifie qu'il a plusieurs sites d'interactions avec la thrombine. De plus, la pré-incubation de HD22 avec la thrombine empêche l'agrégation de HD1, suggérant que HD1 et HD22 ont des affinités concurrentes pour le même site de liaison. Pour compléter cette étude, il aurait donc été pertinent d'évaluer l'interaction de HD1 sur l'exosite 2. Par ailleurs, pour être vraiment pertinents, il aurait également fallu évaluer l'interaction de la thrombine avec HD1 et HD22 dans leur forme fonctionnalisée. Ces travaux se sont terminés avec l'arrêt du développement des technologies aptamères au LAAS. Cependant, des travaux semblables ont été ensuite menés pour l'étude d'acides nucléiques utilisés dans des dispositifs de capteurs au travers du projet décrit par la suite.

2. Balises moléculaires. Les propriétés d'hybridation des acides nucléiques sont ici directement mises à profit pour détecter la présence d'ADN/ARN circulants, en particulier pour leur rôle de biomarqueurs pour le diagnostic et le suivi d'évolution de cancers par des biopsies liquides. La présence de mutations spécifiques est particulièrement intéressante d'un point de vue biologique car elle permet la détection précoce de la maladie ainsi que le suivi de l'apparition de résistances aux thérapies. Elle l'est également du point de vue de la détection car des sondes de type balises moléculaires s'avèrent particulièrement pertinentes et efficaces pour identifier ces mutations.

Les balises moléculaires sont de courtes séquences d'ADN simple brin (15-30 nucléotides) capables de s'hybrider avec une cible qui est elle-même un ADN/ARN simple brin. En l'absence de cible, la balise est repliée dans une structure de type tige-boucle, maintenant à proximité un fluorophore et un quencher, placés respectivement aux deux extrémités de la balise. En présence de la cible, la balise s'ouvre pour s'hybrider à la séquence qu'elle reconnaît et induit la fluorescence. La tige est composée de deux séquences complémentaires (environ 4-7 paires de bases) placées aux extrémités de la sonde. La boucle est une séquence complémentaire à la cible (voir Fig.2.39).

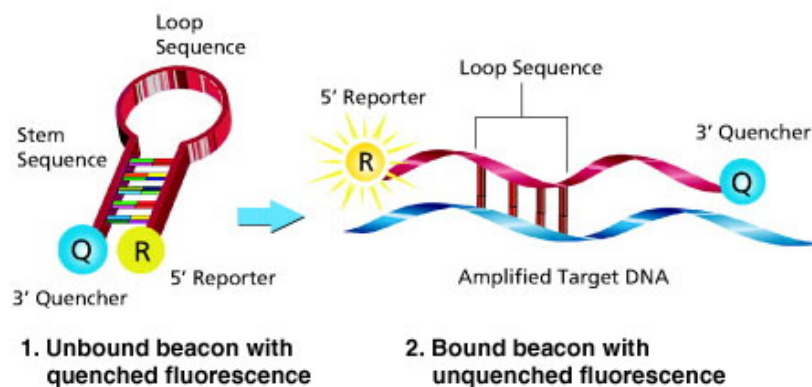


FIGURE 2.39 – Schéma de principe d'une balise moléculaire, repliée dans une structure de type tige boucle en absence de cible (à gauche). En présence de cible, elle s'ouvre par compétition pour permettre sa liaison à la boucle de séquence complémentaire (à droite). L'éloignement induit du fluorophore et du quencher déclenche alors la fluorescence de la balise (figure extraite de *sigmaaldrich.com*).

Ces sondes peuvent présenter une sélectivité à une paire de base près, sélectivité qui peut être anticipée et optimisée par la conception de séquences de sondes spécifiques. **La sélectivité des balises moléculaires, pour des séquences de complémentarité parfaite ou présentant un mismatch, repose en effet sur l'équilibre entre les formes fermées et ouvertes. Cet équilibre peut être ajusté par des changements de séquence dans les structures de tige et de boucle, améliorant la sensibilité et la spécificité de la détection du mismatch.**

Nous avons donc conçu un nouveau code, dans la lignée de celui développé pour la prédiction de séquence de linkers d'aptamères, mais avec un nouveau cahier des charges : i) la séquence de la boucle doit avoir une longueur au moins égale à celle de la cible et être complémentaire, excepté sur la position de la mutation ; ii) la séquence de la tige a une longueur choisie par l'utilisateur et ne doit présenter aucune complémentarité possible avec la boucle ou la cible ; iii) la température de fusion de la tige doit être plus basse que celle du duplex boucle/cible pour garantir l'ouverture de la sonde en présence de la cible. La complémentarité avec la séquence contenant le mismatch est énergétiquement moins stable que la correspondance complète. Cette différence d'énergie doit également être suffisante pour induire un décalage visible par fluorescence. Les séquences proposées par le code sont ainsi triées par énergies ou températures de fusion, calculées d'après le modèle empirique de Santa Lucia [113].

Le logiciel implémenté a été utilisé pour le développement d'une sonde pour la détection du microARN miARN21, ciblé comme biomarqueur surexprimé dans le cancer du pancréas. Une forme mutée sur une base a également été ciblée pour caractériser les performances de la balise moléculaire. La sonde et sa cible ont été incubées pour différentes conditions (concentration et température). Les mesures de fluorescence montrent, à 55°C, une intensité de signal 6,2 fois plus importante pour la sonde associée à miARN21 que pour l'association avec la cible mutée (Fig. 2.40). La sélectivité à une mutation près de la sonde est donc validée.

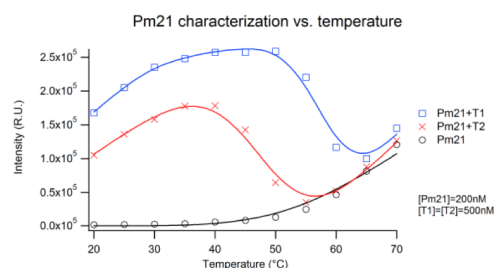


FIGURE 2.40 – Mesures de fluorescence de la balise moléculaire contre miARN21 et sa forme mutée [142].

Toujours pour accompagner le développement de ces capteurs, nous avons également mené des calculs de dynamique moléculaire pour évaluer l'énergie de liaison de complexes sonde/cible, particulier pour un autre biomarqueurs du cancer du pancréas, miARN22. L'objectif était d'évaluer les différences d'affinités entre la sonde et le micro-ARN complémentaire ("full match") ou contenant deux mutations ("DND", pour *double nucleotide difference*). Par ailleurs, la sonde contenait des bases LNA (*locked nucleic acids*) pour améliorer sa stabilité. Les calculs ont été menés en solvant explicite, avec le logiciel AMBER et le champ de force ff14SB (les détails du calcul sont donnés dans la partie *Supplementary Materials* de l'article [47]). Nous avons pris en compte les conditions de concentration et de température dans lesquelles se placent les expériences (137 mM de NaCl, et calculs menés à 308 et 318K).

Les trajectoires produites ont été analysées et traitées par méthode MM/PBSA afin d'extraire les énergies libres de liaison sonde/cibles. La comparaison des valeurs théoriques ainsi trouvées et des valeurs expérimentales donne une erreur d'environ 1%, ce qui traduit un excellent taux de confiance dans les simulations. Ce type de calculs pourrait donc être utilisé en complément des prédictions de séquence pour affiner les valeurs empiriques par lesquelles le logiciel hiérarchise les prédictions, constituant un outil de plus pour guider les expérimentateurs dans leur choix.

Perspectives

Nos activités sur les aptamères se sont *a priori* terminées avec l'arrêt de leur développement au LAAS. Les techniques que nous avons mises en œuvre pour leur étude ne leur sont cependant pas spécifiques pour la plupart (Modes Statiques, dynamique moléculaire), et celles que nous avons développées peuvent être adaptées et réutilisées pour les technologies ADN en général, comme nous l'avons fait pour les *molecular beacons* en particulier. Ce volet de nos activités a été initié pour accompagner les technologies ADN au LAAS mais peut bien sur être ouvert à des projets extérieurs le cas échéant.

Ces travaux ont par ailleurs permis de montrer l'importance des simulations pour explorer, au-delà des propriétés intrinsèques des biomolécules, l'effet des modifications qui leur sont apportées dans le cadre de leur intégration. Leurs propriétés structurales et fonctionnelles doivent en effet être maintenues, ce qui découle de notre capacité à répondre à certaines questions fondamentales comme : comment déposer les biomolécules sur une surface et/ou les fonctionnaliser sans les dénaturer ? comment diriger leur sélectivité ? comment diminuer leur complexité tout en conservant leur spécificité et leur fonction ? peut-on caractériser et optimiser leur capacité d'interaction avec leur environnement ? peut-on qualitativement et quantitativement inclure l'environnement technologique ?

Ces nombreux niveaux de compréhension ne peuvent être adressés que par une batterie d'outils de modélisation. Certains, issus des méthodes "traditionnelles" sont déjà à notre disposition, de nouveaux sont à développer, tous doivent se compléter pour guider les choix clés autour des constructions moléculaires, interpréter et prédire les performances de dispositifs qui rassemblent les mondes biologiques et non-biologiques.

3.2 Réponse mécanique et mouvements de grande amplitude des assemblages moléculaires

Cadre de travail : Ce travail a été en grande partie réalisé durant mon post-doctorat à l'Université de Stanford. Il est important d'en parler ici car la méthodologie présentée a continué d'être développée et appliquée depuis mon arrivée au LAAS. Nous présenterons les perspectives méthodologiques et applicatives qui en découlent dans la suite de ce manuscrit.

Résumé

Une variante des Modes Statiques a été développée dans l'espace des torsions et appliquée à l'aide d'un modèle énergétique travaillant aussi bien en tout-atome qu'en gros-grains. Nous décrivons ici le passage des Modes Statiques vers les Modes Statiques généralisés, ainsi que leur application à l'étude des mouvements des moteurs moléculaires. Nous avons appliqué des forces sur des moteurs protéiques pour simuler l'amarrage d'un ligand et montré comment un réarrangement local peut résulter en mouvement de grande amplitude (Ca^{2+} ATPase). Nous avons également utilisé cette approche pour décomposer et corrélérer les mouvements de domaines (myosine) afin de décrypter les mécanismes de cette "machinerie moléculaire". Ce type d'approche est applicable à tous les grands assemblages de polymères, notre seule limite est l'existence de modèles énergétiques dédiés dont nous dépendons.

Contexte

La question de savoir comment la liaison d'un ligand (souvent l'ATP dans les moteurs moléculaires) peut générer de grands changements conformationnels, c'est-à-dire **comment une force locale peut entraîner un réarrangement structural global**, a été largement étudiée au cours des dernières années. Un tel couplage est particulièrement important dans le domaine des protéines allostériques ou encore des moteurs moléculaires dont le fonctionnement est basé sur ce principe. Si ces derniers ont été largement étudiés du point de vue de la biologie [115,207], ils sont désormais également considérés d'un point de vue technologique, puisqu'ils ont émergé comme des objets d'ingénierie moléculaire. Concernant leur potentiel pour aider à concevoir et intégrer des fonctions actives dans de nouveaux dispositifs bio-hybrides [6, 91, 224], des efforts importants sont consentis pour intégrer des protéines motrices sous forme de nanomachines dans des environnements artificiels, afin d'exécuter des tâches sophistiquées à l'échelle nanométrique. Par conséquent, **comprendre et exploiter le mécanisme des moteurs biomoléculaires est aujourd'hui un défi pour des raisons biologiques et technologiques, respectivement pour comprendre les fonctions cellulaires et pour concevoir de nouveaux nanosystèmes fonctionnels.**

Les Modes Statiques sont un outil compétitif pour travailler sur des molécules ne subissant que des déformations de faible amplitude. Les premiers résultats ont prouvé la puissance de cette approche pour décrire des mécanismes avec une résolution atomique. Cependant, **les procédures deviennent fastidieuses et le temps calcul prohibitif lorsqu'il s'agit de suivre de grands changements conformationnels.** L'idée est d'abandonner les coordonnées cartésiennes et de prendre en compte le fait que les biomolécules, résultant d'un assemblage de briques élémentaires, peuvent être décrites par des angles de torsion et des encombrements stériques. Les coordonnées cartésiennes sont alors remplacées par des **coordonnées généralisées, au sens de Lagrange, associées à des "Modes Généralisés"**. Au niveau du calcul, un des intérêts est que la matrice hessienne se présente dès le départ sous une forme diagonale, puisque dans le cas des molécules biologiques, constituées d'une chaîne de base repliée par l'intermédiaire de chaînes latérales, l'essentiel des variations d'énergie aux températures ambiantes provient des variations d'angles de torsions. Les "Modes Généralisés" sont donc définis comme des mouvements de torsions, peu dépendantes les unes des autres.

Nous avons jusque-là utilisé les Modes Statiques en nous appuyant sur un modèle tout-atome, en coordonnées cartésiennes, dont nous pouvons calculer l'énergie totale à l'aide de tout logiciel de calcul classique (AMBER...) ou quantique (Gaussian...) selon la taille du système et la précision voulue. La méthode telle qu'elle est implémentée pourrait également être appliquée à des modèles gros-grains, du moment que les coordonnées des "grains" sont décrites dans un espace muni d'un repère cartésien.

En revanche, travailler dans l'espace des torsions demande de nouveaux développements. J'ai développé durant mon post-doctorat une nouvelle version basée sur un **modèle énergétique décrit dans l'espace des torsions, pouvant être de type tout-atome ou coarse-grained** (développé par le Professeur M. Levitt (ENCAD)). L'ambition est toujours de créer des déformations en réponse à l'application de forces locales, mais **l'usage des torsions doit permettre de générer une propagation suffisante pour accéder aux grands mouvements.**

Méthodologie

Nous présentons un nouvel algorithme basé sur le modèle d'énergie de torsions implémenté dans le programme ENCAD. Cet algorithme vise à prédire efficacement la déformation globale d'une macromolécule en réponse à l'application de forces locales. L'idée d'appliquer des forces statiques externes et d'induire, en un très faible temps de calcul, des mouvements à longue distance, et de grande amplitude.

Modèle énergétique. Tout comme le calcul des Modes Statiques avec le logiciel FleXible, notre algorithme nécessite un calcul préliminaire d'énergie totale, afin d'obtenir la matrice Hessienne, mais aussi la Jacobienne, toutes deux nécessaires à la simulation de l'application de forces. Dans le cas présenté, nous avons choisi de travailler avec une représentation gros-grains centrée sur les atomes C α ainsi qu'à l'aide de coordonnées généralisées, ici les angles de torsion. Le programme de simulation ENCAD (ENergy CALculation and Dynamics) [131, 132] a été utilisé à cette fin. L'équation suivante contient la fonction d'énergie potentielle ENCAD :

$$V = \sum_i^{\text{bonds}} K_{b,i} (b_i - b_{0,i})^2 + \sum_i^{\text{angles}} K_{\theta,i} (\theta_i - \theta_{0,i})^2 + \sum_i^{\text{torsions}} K_{\phi,i} (1 - \cos[n_i(\phi_i - \phi_{0,i})]) + V_{nb}$$

V est décrite comme une fonction des coordonnées internes b_i , θ_i , ϕ_i calculées à partir des coordonnées cartésiennes. Elle est exprimée en termes d'interactions intramoléculaires dues respectivement aux longueurs de liaison, aux angles de liaison et aux angles de torsion, plus un quatrième terme V_{nb} pour les énergies non liées (van der Waals et électrostatique).

Contraintes géométriques. Notre objectif dans cette étude est de comprendre l'origine des mouvements d'un point de vue mécanistique et comment ils peuvent résulter en un changement de conformation. Notre stratégie est d'induire des mouvements de séquences spécifiques, définis comme des perturbations locales, et d'en évaluer les conséquences sur l'ensemble de la molécule. A cet effet, deux algorithmes ont été écrits et testés pour déplacer des séquences données de leur position dans un état initial vers un état cible, avec et sans ligand par exemple.

1. Dans la première approche, une minimisation d'énergie est effectuée alors qu'une contrainte de distance est appliquée. Dans l'approximation quadratique, l'énergie potentielle du système $U(q)$ est fonction des termes de la matrice de Hessienne (H_{ij}) et des angles de torsion (q_{ij}) : $U(q) = 1/2 \sum_{ij} H_{ij} q_i q_j$. Si une contrainte est ajoutée pour forcer une distance d_{AB} , entre 2 atomes (ou entités) A et B , à devenir une distance cible d_{AB}^t , l'énergie potentielle devient alors :

$$U(q) = 1/2 \sum_{ij} H_{ij} q_i q_j + k(d_{AB} - d_{AB}^t)^2$$

où k est une constante de force. La dérivée première de l'énergie, donnée par :

$$\frac{\partial U(q)}{\partial q_i} = \sum_j H_{ij} q_j + 2k(d_{AB} - d_{AB}^t) \frac{\partial d_{AB}}{\partial q_i}$$

doit être égale à 0 lorsque les angles de torsion s'ajustent pour atteindre un nouvel équilibre tel que $d_{AB} = d_{AB}^t$. Un algorithme de minimisation est donc implémenté pour résoudre le système $\frac{\partial U(q)}{\partial q_i} = 0$.

Pour cela, la distance entre les atomes A et B étant donnée par :

$$(d_{AB})^2 = (x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = \sum_{l=x,y,z} (r_{l,A} - r_{l,B})^2$$

la dérivée première de l'énergie devient :

$$\frac{\partial U(q)}{\partial q_i} = \sum_j H_{ij} q_j + 2k \left(1 - \frac{d_{AB}^t}{d_{AB}}\right) \sum_{l=x,y,z} [(r_{l,A} - r_{l,B})(J_{lA,i} - J_{lB,i})]$$

où les termes $J_{x_A,i} = \frac{\partial x_A}{\partial q_i}$, ... et $J_{x_B,i} = \frac{\partial x_B}{\partial q_i}$ sont les éléments de la matrice Jacobienne J pour les atomes A et B et l'angle de torsion i .

2. L'algorithme précédent est bien adapté si les atomes A et B appartiennent à deux domaines en mouvement l'un par rapport à l'autre, ou s'ils appartiennent à un domaine qui subit un réarrangement structural important. Dans le cas où l'on veut déplacer un domaine d'une position de départ vers une position finale sans connaissance *a priori* d'autres mouvements, un deuxième algorithme a été écrit, permettant l'application de forces cartésiennes F_j sur des atomes j . Le code fait directement suite aux Modes Statiques, mais en travaillant à l'aide de coordonnées généralisées q_i (ici les torsions) et de forces généralisées f_l calculées à partir des forces imposées F_j . Ces dernières sont décrites comme des vecteurs entre les positions atomiques de l'état initial et de l'état cible. Elles sont transformées en forces généralisées f_l à l'aide de la matrice jacobienne J :

$$f_l = \sum_j F_j \frac{\partial r_j}{\partial q_l} \quad (2.4)$$

Les changements induits au niveau des coordonnées généralisées δq_l sont ensuite convertis en changements des coordonnées cartésiennes δr_i en utilisant l'équation suivante :

$$\delta r_i = \sum_l \frac{\partial r_i}{\partial q_l} \delta q_l \quad (2.5)$$

Un processus itératif (minimisation) est enfin appliqué jusqu'à ce que les forces soient nulles ($f_l(q) = 0$) et que l'on atteigne donc un nouvel équilibre. Dans les deux cas, la méthode du gradient conjugué est utilisée. Le critère de convergence est satisfait lorsque les variations du RMSD atteignent 10^{-5} . Lorsqu'une seule séquence est déplacée, le second algorithme est utilisé. Lorsque deux sous-domaines ou plus sont déplacés, les deux algorithmes ont été testés et conduisent à la même structure finale.

Résultats : Moteurs moléculaires et mouvements de grande amplitude

Les deux algorithmes sont effectifs et ont été validés sur le cas de moteurs moléculaires dont les mouvements fonctionnels peuvent atteindre plusieurs dizaines d'Ångströms. Nous considérons cette approche comme une preuve de concept des Modes Généralisés que je voudrais continuer de développer. Nous avons en particulier travaillé sur des moteurs à ATP qui hydrolysent les molécules d'adénosine-triphosphate (ATP) et convertissent cette énergie en changements conformationnels complexes permettant leur fonctionnement mécanique. Nous avons cherché à **comprendre et à reproduire les mécanismes de ces "machines moléculaires"**.

Le premier cas traité était celui de la Ca^{2+} -ATPase, une pompe à ions pour laquelle un excellent accord avec la conformation active a été obtenu en appliquant des forces locales au niveau des sites de liaison de l'ATP. Il est remarquable que chaque mouvement de domaine vers sa position cible entraîne une diminution du RMSD des autres domaines. Dans ce cas, le mouvement du domaine cytoplasmique en direction de l'ATP suffit à replacer l'ensemble de la protéine dans sa conformation active (structure E1 vers E2, voir Fig.2.41 (A)). **Dans le cas de la myosine, des mouvements locaux entraînent, par amplification, le mouvement du levier sur plus de 11Å (Fig.2.41(B)). Au-delà de ces résultats, cette méthode prouve qu'un modèle de torsions et de forces généralisées est particulièrement adapté à la modélisation des mouvements de grande amplitude des biomolécules.** Dans la suite je détaille davantage le cas de la myosine.

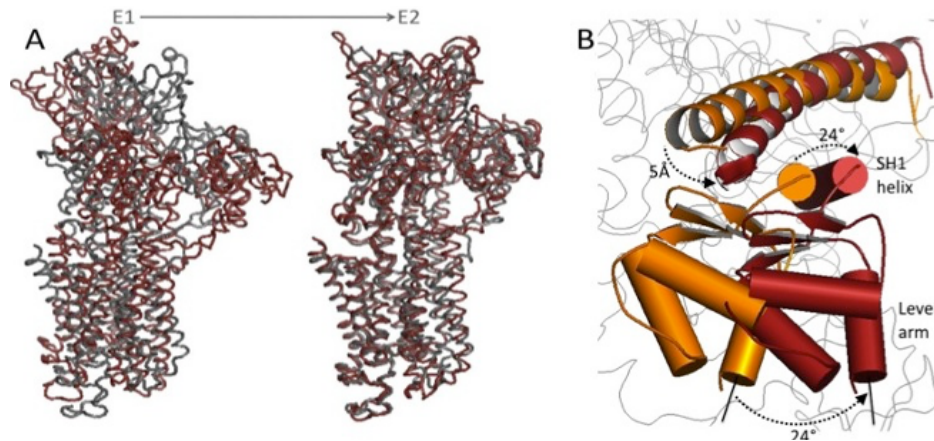


FIGURE 2.41 – A. Changement de conformation de la Ca²⁺-ATPase induit par l'application de forces locales sur les sites de liaison à l'ATP. Les structures initiales (état E1) et finales (état E2) sont représentées en rouge, la structure cible en gris. B. Mouvement du levier de la myosine (structure initiale en orange et finale en rouge) initié par un changement de courbure de l'hélice.

La myosine est l'un des moteurs à ATP les mieux étudiés [137,216]. Cette protéine joue un rôle clé dans la contraction des muscles en convertissant l'énergie chimique de l'hydrolyse de l'ATP en travail mécanique. La résolution des structures a révélé que la tête de la myosine contient deux domaines, le domaine moteur (résidus 1-770) et le bras de levier (résidus 771-837), dont le mouvement est entraîné par le moteur. Le moteur est lui-même composé de quatre sous-domaines (présentés sur la figure 2.42 et notés N, U50 L50 et C pour convertisseur). D'après les structures cristallines, il apparaît que les sous-domaines N, U50 et L50 ne subissent globalement pas de réarrangement structural important, tandis qu'un grand mouvement combiné de translation/rotation du convertisseur entraîne le bras de levier selon la liaison au nucléotide.

Pour étudier le rôle de la liaison de l'ATP dans le mouvement de la myosine, deux états sont considérés: respectivement non-lié (PDB 1KK7) et lié à l'ATP (PDB 1KK8). Etant donné que le bras du moteur se déplace avec le domaine du "convertisseur" comme un bloc rigide, seul le domaine moteur est considéré dans notre étude (sous-domaines N, U50, L50 et convertisseur, voir Fig.2.42). Au final, les résidus conservés dans les calculs sont les résidus 1 à 775.

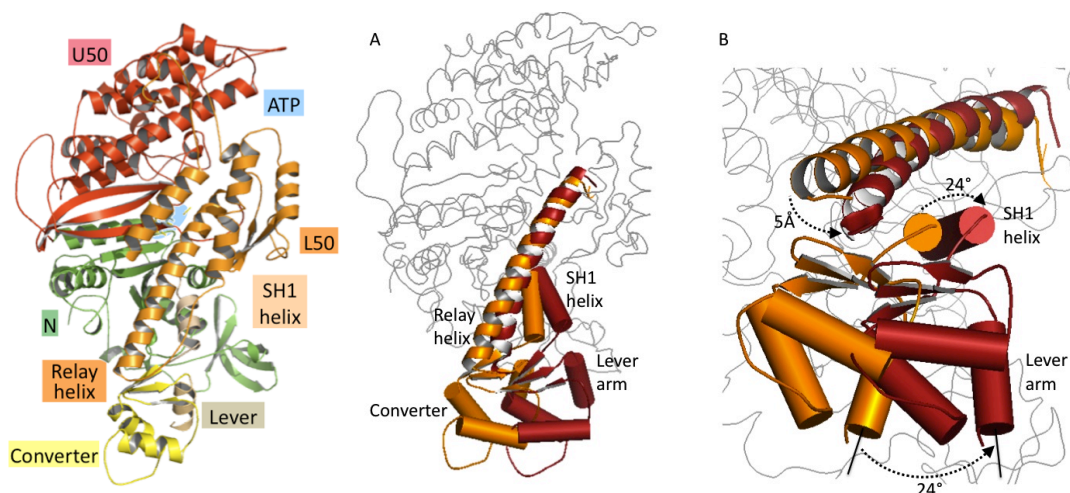


FIGURE 2.42 – A gauche : Structure du domaine moteur de la myosine présentée dans la conformation non liée. Les sous-domaines N-terminal (N), supérieur 50 kDa (U50), inférieur 50 kDa (L50) et convertisseur (C) sont respectivement coloriés en vert, rouge, orange et jaune. Le site de liaison à l'ATP est également représenté en bleu à l'interface entre les sous-domaines N et U50. A droite, en A : Structure sans ATP (orange) et avec ATP (rouge) où sont localisés : l'hélice relais, l'hélice SH1, le convertisseur et le bras de levier. B. Sur la droite, la molécule est orientée dans l'axe de l'hélice SH1 qui est également l'axe de rotation du convertisseur (24°).

Les deux structures ont été traitées à l'aide du programme ENCAD (minimisation, extraction des matrices hessienne et jacobienne). Les structures résultantes sont les points de départ de l'étape suivante: l'application des forces. La superposition des structures avec et sans ATP permet de mesurer les différences de RMS entre $C\alpha$ (1.64Å pour le sous domaine N, 1.91Å pour U50, 3.28Å pour L50 et 13Å pour C). N et U50 ne subissent pas de réarrangement notable. Les différences de RMS sont principalement dues aux mouvements des boucles suivantes: P-loop (notée N1), switch-I (N2) et switch-II (N3). N1 appartient à N, tandis que N2 et N3 appartiennent à U50. Ces boucles correspondent aux séquences impliquées dans la liaison de l'ATP, ce qui est cohérent avec les différences structurales observées entre les conformations avec et sans ATP. Enfin, L50 présente la différence la plus significative entre les états liés et non liés. Ce sous-domaine se déplace comme un bloc rigide, sauf deux séquences qui subissent un réarrangement important: l'hélice "relais" (rmsd=3.99Å) et l'hélice SH1 (rmsd=5.12Å). Enfin, le convertisseur subit un grand mouvement combiné translation/rotation, mais pas de réarrangement interne. Il semble se déplacer comme un bloc rigide articulé via l'hélice SH1 (voir Fig.2.42 (B)).

Après application de forces sur les différents sous-domaines de la myosine, les déformations induites sur l'ensemble de la protéine sont analysées. Les réponses résultant de la perturbation des sous-domaines sont présentées sur la figure 2.43. Une caractérisation systématique des perturbations des sous-domaines a été réalisée. Je présente ici les résultats obtenus lorsque des perturbations sont appliquées sur les sous-domaines N et U50 qui contiennent le site de liaison ATP.

Les figures 2.43 et 2.44 illustrent la réponse du convertisseur aux différentes perturbations :

(A) Lorsque des forces sont appliquées sur le sous-domaine N, le RMSD diminue de façon globale et le convertisseur est translaté de 2,4Å vers l'état cible, directement poussé à travers des contacts inter-domaines (entre les résidus 767/770 et 25-26 de N et C-term respectivement). Par ailleurs, en observant que N active L50, mais pas U50, nous formulons l'hypothèse que le mouvement de N active le convertisseur, qui entraîne par rétroaction SH1 puis L50 (SH1 se déplace comme un bloc avec le convertisseur et est responsable du faible mouvement de L50). Ceci est vérifié en appliquant des forces directement sur le convertisseur (son mouvement vers sa position cible entraîne une diminution notable du RMSD de L50, on retrouve également une corrélation pertinente avec les mouvements de N et des hélices relais et SH1). Ces résultats confirment un possible rétro-effet du convertisseur vers L50 et N.

(B) Les contraintes appliquées sur le sous-domaine U50 entraînent les mouvements de N et L50. On observe également une translation de l'extrémité C du relais (2,9Å) accompagnant la rotation du convertisseur. Cette translation est légèrement en dehors du plan attendu ($3,5^\circ$), ce qui peut expliquer la faible amplitude du mouvement du convertisseur.

(C) Appliquer les forces simultanément sur N et U50 mène à un résultat qui couple les observations précédentes. Ces forces induisent en effet un réarrangement global de tous les sous-domaines (RMSD total réduit de 2,6Å), et surtout du convertisseur, dont le RMSD diminue de 13 à 5,5Å. La translation de l'hélice relais (3,7Å) et la rotation du convertisseur ($22,6^\circ$ au lieu de 24° pour une rotation complète) sont presque achevées par rapport à l'état final. La translation n'est cependant pas totalement réalisée. Cela peut s'expliquer par le fait que N, qui entraîne la rotation du convertisseur, a atteint la conformation cible tandis que SH1, qui est principalement responsable de la translation, est toujours à mi-chemin entre la conformation initiale et la conformation cible. Le bras du levier est quant à lui à 1,5Å de la position cible. Il est remarquable d'observer comment la combinaison des mouvements de N et U50 améliore les comportements que nous avons trouvés en les déplaçant séparément.

Ces calculs mettent en évidence le rôle majeur joué par le relais et l'hélice SH1 qui accompagnent étroitement le mouvement du convertisseur. Afin d'analyser plus en détail comment ces éléments sont impliqués dans le mécanisme, ils sont ensuite déplacés séparément.

(D) Les forces appliquées sur l'hélice relais se traduisent par un mouvement important du convertisseur ($16,5^\circ$), accompagné de l'hélice SH1. L'extrémité C est en place, mais l'hélice est mal orientée.

(E) Lorsque les forces sont appliquées sur l'hélice SH1, le mouvement entraîne remarquablement le convertisseur vers sa position cible avec une valeur de RMSD passant de 13 à 1,5Å.

Travailler dans l'espace d'angle des torsions est donc attractif et efficace, notamment en raison du nombre modéré de degrés de liberté. Cependant, l'utilisation d'une représentation cartésienne et tout-atome présente également certains avantages. En particulier, les longueurs et angles de liaison peuvent s'écarter des valeurs standard dans un modèle cartésien, mais pas ici. Par ailleurs,

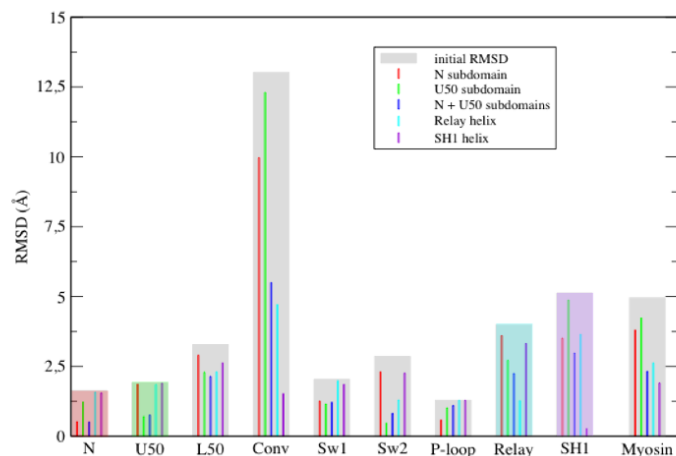


FIGURE 2.43 – Réponse des $C\alpha$ aux perturbations. En gris, sont représentées les valeurs de RMSD calculées entre l'état initial (sans ATP) et l'état cible (lié à l'ATP). Les barres colorées représentent le RMSD entre la cible et les structures obtenues en appliquant des forces sur des domaines de l'état initial vers leur position dans l'état cible. Les forces sont appliquées sur les sous-domaines N, U50, N et U50, les hélices relais et SH1.

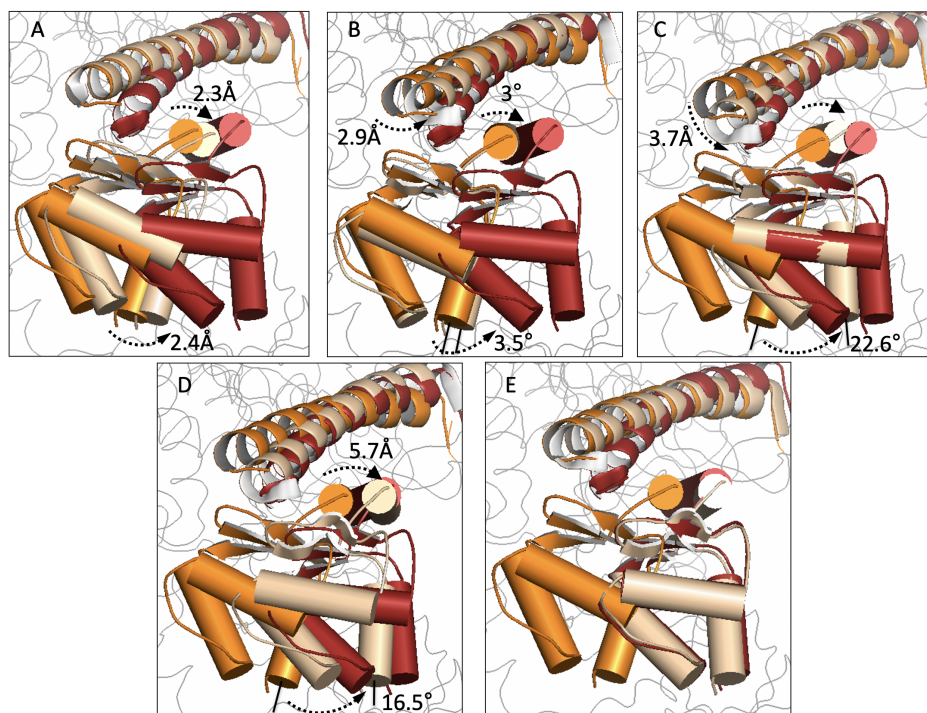


FIGURE 2.44 – Réponse du domaine convertisseur aux mouvements induits des domaines N (A), U50 (B), N et U50 (C), et des hélices relais (D) et SH1 (E).

la distribution de charges et les forces physiques, en particulier les interactions faibles, y sont décrites plus naturellement. D'un autre côté, travailler dans l'espace cartésien augmente considérablement le nombre de paramètres à prendre en compte, ce qui peut être une limitation pour traiter les grands mouvements et les grands systèmes. **Pour mesurer le bénéfice de travailler en torsions dans ce type d'application, nous avons également utilisé le programme Flexible pour évaluer les déformations de la myosine lorsqu'elle est soumise à des forces statiques externes dans l'espace cartésien et comparé les résultats avec ceux obtenus dans l'espace des torsions.**

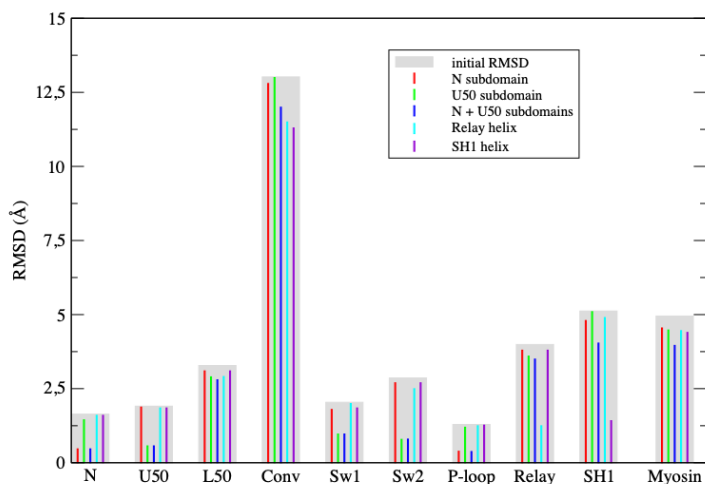


FIGURE 2.45 – Réponse des $C\alpha$ aux perturbations dans l'espace cartésien avec les Modes Statiques. En gris, sont représentées les valeurs de RMSD calculées entre l'état initial (sans ATP) et l'état cible (lié à l'ATP). Les barres colorées représentent le RMSD entre la cible et les structures obtenues en appliquant des forces sur des domaines de l'état initial vers leur position dans l'état cible. Les forces sont appliquées sur les sous-domaines N, U50, N et U50, les hélices relais et SH1.

D'un point de vue technique, la procédure habituelle a été appliquée (minimisation avec AMBER en solvant implicite et avec le champ de force ff99bsc0, puis calcul des MS et application des forces). Notons cependant une difficulté liée au calcul de la matrice hessienne qui a imposé de pouvoir allouer suffisamment d'espace mémoire RAM (soit presque 12 Go dans ce cas pour 12200 atomes, et trois fois plus pour le calcul des MS). La même procédure que précédemment utilisée avec les torsions a été appliquée, c'est-à-dire que des forces ont été appliquées sur les atomes $C\alpha$ de la structure de départ (PDB 1KK7) vers leur position cible (PDB 1KK8).

Les résultats sont présentés sur la figure 2.45. Fait intéressant, nous observons les mêmes comportements que ceux obtenus avec les torsions, mais les déplacements sont générés avec une amplitude beaucoup plus faible. **Avec ce modèle, la déformation générée par une force locale ne se propage pas significativement au-delà de 15Å. Elle n'est donc pas suffisante pour induire un mouvement important et donner accès à une nouvelle conformation stable.** Un véritable changement de conformation pourrait se construire par une succession d'étapes de déformation / minimisation (comme cela avait été fait avec le peptide amyloïde- β), ce qui serait particulièrement lourd et n'est pas l'ambition de ce travail.

Cette comparaison confirme l'efficacité de notre approche pour prédire les mouvements de grande amplitude. **Parmi ses avantages, on peut citer :** i) son nombre réduit de degrés de liberté, restreints aux angles de torsion ; ii) sa capacité à franchir les barrières énergétiques, ce qui réduit considérablement le piégeage dans des minima locaux ; iii) la validité du modèle gros-grains qui décrit très bien les mouvements fonctionnels ; iv) le très faible coût de calcul requis pour chaque perturbation puisque le paysage énergétique est totalement défini par la matrice hessienne, elle-même réduite par rapport à sa version cartésienne.

Nous montrons que **la réduction de la complexité de calcul, essentielle pour atteindre de grandes échelles, est légitime et prédictive.** Cependant, **l'approche cartésienne est nécessaire pour étudier des processus nécessitant une résolution atomique, comme l'activité chimique des protéines, dans lesquelles la géométrie covalente n'est pas préservée.** Les deux méthodes sont par conséquent complémentaires et ne s'excluent pas mutuellement.

Perspectives

Nous montrons ici comment une approche simple et purement mécanistique peut faire progresser la compréhension des mécanismes de génération de force. En appliquant des contraintes locales, nous identifions des éléments structuraux qui amplifient les mouvements et conduisent des changements conformationnels. Cette vision mécanistique, bien que simplifiée (représentation gros-grains et modèles de torsions), apparaît très efficace. Notamment, l'utilisation des coordonnées internes permet de réduire drastiquement la dimension de l'espace de recherche ainsi que le coût de calcul. Elle s'avère par ailleurs légitime, car la géométrie covalente est préservée pendant la transition conformationnelle, et particulièrement adaptée pour prédire des mouvements à grande échelle, car de petits changements dans les angles de torsion peuvent correspondre à de grands changements dans l'espace cartésien.

Ce travail est considéré comme un préliminaire à de nouveaux développements et à de nouvelles applications autres que des protéines, ce qui nécessite un modèle énergétique adapté. C'est actuellement là que se trouve notre limite. Nous avons à notre disposition de très nombreux modèles énergétiques paramétrés à différents niveaux de représentation et pour tout type de biomolécule, mais pas en torsions. ENCAD peut travailler en tout-atome et en gros-grains, il est surtout dédié aux protéines et peut également être appliqué aux acides nucléiques, mais pas à d'autres polymères sur lesquels nous projetons de travailler, comme les constituants de membranes lipidiques.

Nous avons en effet débuté un travail avec P. Joseph au LAAS qui étudie les **propriétés de bicouches lipidiques**, éléments de base des membranes cellulaires, par des manipulations microfluidiques. Caractériser leurs propriétés mécaniques est crucial à de nombreux égards, notamment pour l'étude des interactions cellule-nanoparticules (NP). Les applications adressent des questions liées à la nanotoxicité et la nanomédecine en particulier, au-delà de la connaissance fondamentale des propriétés des membranes. La modélisation moléculaire peut être un outil de choix pour reconstituer des membranes *in silico*, les déformer, les mettre en interaction avec une surface, une NP, etc. Nous avons proposé plusieurs stages pour amorcer ce travail (R. Bonhoure et V. Combette) afin de proposer un modèle de bicouche lipidique, et d'adapter nos approches pour étudier leur organisation moléculaire et leur réponse mécanique à une contrainte.

Les premiers calculs avec AMBER, en modèle tout-atome, ont rapidement été limités à des systèmes de taille trop restreinte pour pouvoir accéder à des grandeurs comparables aux données macroscopiques fournies par l'expérience, comme les modules de courbure et d'élongation [29]. Nous sommes alors passés dans un second temps à l'utilisation du champ de force Martini [146] couplé au logiciel de dynamique moléculaire Gromacs [204]. Ce travail reste à poursuivre, ce qui sera fait dans un premier temps en suivant la voie actuelle, c'est-à-dire en couplant FleXible aux sorties de Gromacs en modèle gros-grains. Nous pouvons également envisager de passer aux coordonnées généralisées et de travailler dans l'espace des torsions, ce qui nécessitera de créer un modèle énergétique adapté, ce que nous pouvons faire à partir des paramètres de Martini. La création d'un champ de force en torsions pour les protéines à partir d'un champ de force classique AMBER a déjà été réalisée dans notre équipe lors du stage de J. Teillet, ce travail peut donc être envisagé dans le futur pour d'autres biomolécules (ADN et phospholipides).

Conclusions et Perspectives

Les simulations offrent désormais la possibilité d'un aperçu unique des propriétés moléculaires avec une résolution atomique. Les développements méthodologiques et les progrès informatiques nous permettent de pouvoir choisir parmi une large gamme d'outils selon la problématique et le système traités. Ils transforment la portée des calculs qui visent une comparaison quantitative à l'expérience.

Ce manuscrit rapporte les différentes approches que nous mettons en œuvre pour adresser des applications variées, mais centrées autour des biomolécules comme principal objet d'étude. Le développement de modèles physiques prédictifs est central dans nos travaux et sert deux objectifs : i) améliorer la compréhension fondamentale des mécanismes qui régissent les propriétés des biomolécules, et ii) contrôler ces propriétés pour des applications thérapeutiques et technologiques. Ce panel d'activités nous permet d'exploiter pleinement les possibilités des modèles et de les affiner. En retour, ces développements nous permettent d'adresser de nouvelles applications sous un angle original.

La modélisation entretient ainsi un cercle vertueux entre méthodes et applications, profitant de leurs avancées et les nourrissant de questionnements et de nouveaux développements. Dans ce manuscrit, j'ai présenté nos activités en suivant ces deux voies intriquées : les deux volets applicatifs ont été successivement décrits en s'appuyant sur le panel de méthodes utilisées.

Comme précisé dans la mise en contexte de chaque projet présenté ici, les calculs sont tous menés pour répondre à un questionnement ou un besoin précis de nos partenaires expérimentateurs. La combinaison de plusieurs techniques de calculs complémentaires est souvent nécessaire. Si l'échelle atomique a été privilégiée jusque-là, la finalité des problématiques, l'accès à des grandeurs quantitativement comparables à l'expérience, nous incitent à pouvoir couvrir une plus large gamme d'échelle de taille. La nécessité de proposer de nouveaux développements méthodologiques s'impose alors à nous.

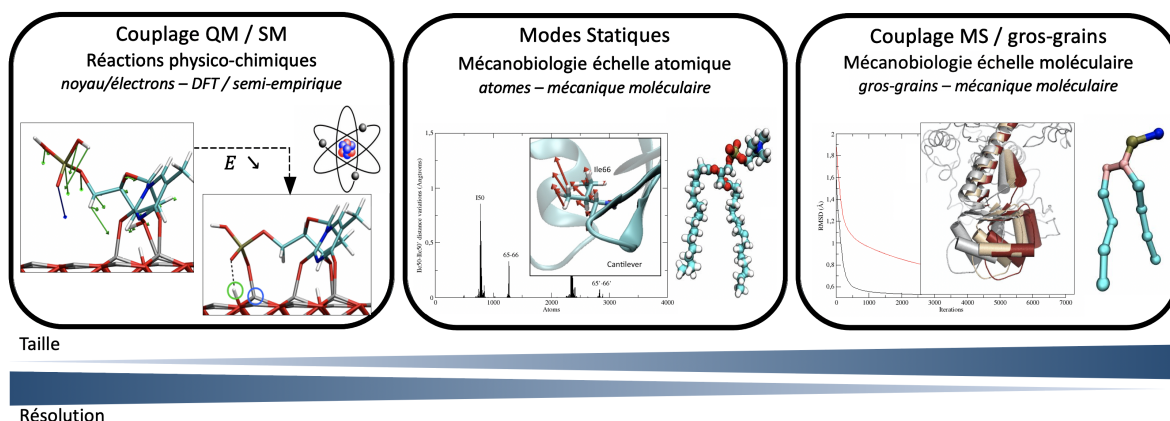
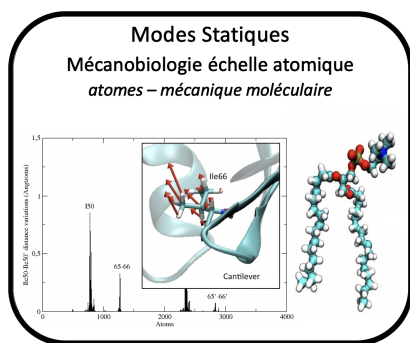


FIGURE 2.46 – Mise en perspective de l'évolution des Modes Statiques d'un point de vue applicatif et méthodologique. Au centre, notre approche telle qu'elle est implémentée et utilisée aujourd'hui, pour explorer les propriétés mécaniques de molécules à l'échelle atomique. A gauche, le couplage QM/SM, les Modes Statiques sont utilisés pour guider la DFT lors de l'exploration de chemins réactionnels. A droite, les Modes Statiques sont adaptés en terme de représentation (gros-grains) et de modèle énergétique pour explorer des déformations de grande amplitude sur des systèmes de grande taille.

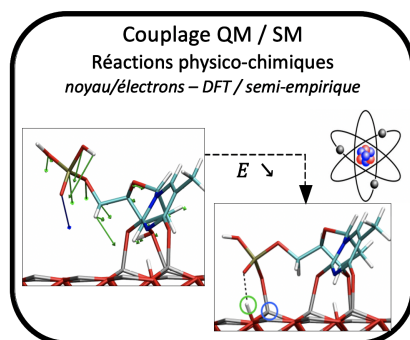
Les méthodologies que nous développons se veulent génériques et applicables à tout système. Les développements envisagés dans cette dernière partie doivent continuer de s'inscrire dans cette trajectoire pour servir de nouvelles applications. Pour cela, je propose d'**étendre l'applicabilité des Modes Statiques au travers de nouvelles approches nous donnant accès à trois niveaux de résolution**. La figure 2.46 illustre cette évolution.



1. L'approche par Modes Statiques présentée au centre est effective et constitue actuellement notre principal outil de travail. J'en ai décrit de multiples possibilités à l'aide des divers projets rapportés dans cette synthèse d'activités. J'ai particulièrement insisté sur la façon dont nous les utilisons pour explorer les propriétés de molécules biologiques sous un angle mécanistique. Le principe est simple et consiste à évaluer la réponse mécanique d'un système à l'application d'une contrainte. Le modèle énergétique est une entrée extérieure qui peut-être choisie par l'utilisateur. La résolution atomique avec laquelle nous avons principalement travaillé jusqu'ici dépend directement de ce modèle. C'est aussi le cas du caractère local des déformations calculées.

De part l'approximation harmonique dans laquelle nous travaillons, les Modes Statiques, par essence, ne peuvent proposer de nouvelle conformation stable, ni de chemin vers cette conformation. Il est cependant possible d'y accéder par une succession d'étapes de déformation/minimisation, lors desquelles les MS vont être calculés et appliqués de manière récursive. Ce procédé a déjà été mis en place et validé sur un système de petite taille en mécanique moléculaire.

Cependant, la propension du système à se déformer en réponse à une contrainte peut-elle être un guide pour explorer le paysage énergétique, trouver de nouveaux minima, et construire/suivre un chemin conformationnel ou réactionnel ?



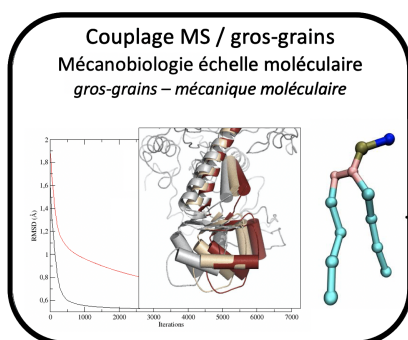
2. C'est la question à laquelle nous avons entrepris de répondre avec la thèse de L. Foulon, co-encadrée avec A. Hémercyck qui apporte son expertise dans les méthodes d'exploration des paysages énergétiques et l'étude de surfaces et interfaces pour le contrôle des propriétés de matériaux avancés. Cette thèse méthodologique a pour objectif de proposer un nouvel apport dans l'exploration du paysage énergétique en s'appuyant sur l'utilisation des Modes Statiques couplés à des calculs DFT. Elle s'inscrit dans une démarche ancrée dans les activités de l'équipe M3, visant à introduire la granularité atomique dans les outils de simulation multi-échelles, de type Monte Carlo cinétique en particulier. La principale difficulté de ce type d'approche repose sur

le besoin initial d'identifier les événements atomiques à prendre en compte dans les modèles d'échelles supérieures. Ce besoin implique d'explorer le paysage énergétique de façon exhaustive, ce qui nécessite des coûts de calcul et humains prohibitifs pour une utilisation systématique, en particulier pour l'ingénierie de matériaux avancés à visée industrielle.

Nous proposons d'**utiliser les Modes Statiques comme entrée de calculs DFT, afin de les guider et de diminuer l'intervention de l'utilisateur nécessaire pour explorer différents mécanismes**. Les MS sont ainsi exploités pour explorer systématiquement et à bas coût des déplacements atomiques induits par des contraintes spécifiques, ou pour obtenir une réponse spécifique du système. Selon cette réponse, ils sont ensuite triés pour guider le choix d'évènements significatifs pour l'évolution du système. Des calculs de type DFT sont enfin menés sur la structure déformée par les MS pour obtenir un nouveau minimum, c'est-à-dire pour identifier de façon guidée différentes configurations atomiques locales ainsi que les barrières d'énergies associées.

Ce couplage, baptisé QM/SM (*Quantum Mechanics / Static Modes*), a été mis en place durant la thèse de L. Foulon qui l'a également appliqué à différents cas d'étude dont nous avons l'expertise dans l'équipe M3. Ces calculs ont permis de valider l'approche sur des questions et systèmes hétérogènes, comme la diffusion de défauts au sein du silicium, ou le greffage d'ADN sur surface d'aluminium.

Ce dernier cas est illustré sur la partie gauche de la figure 2.46 : i) une contrainte est prédite par les MS pour optimiser le rapprochement du groupement phosphate vers la surface ; ii) cette contrainte (flèche bleue) est appliquée, toujours avec les MS, induisant le déplacement d'autres atomes (flèches vertes) ; iii) la structure déformée est relaxée par DFT, conduisant à une nouvelle configuration dans laquelle des liaisons covalentes sont formées entre l'ADN et la surface. Les calculs DFT guidés par les MS permettent de retrouver, à moindre coût, l'ensemble des configurations précédemment trouvées par implémentation DFT "manuelle et intuitive". De nouvelles configurations énergétiquement favorables ont par ailleurs pu être identifiées, ce qui est particulièrement probant quant à l'efficacité de l'approche qui s'avère compétitive dans le panorama de l'état de l'art. **Cette procédure permet de relier MS et chimie locale, son application à des systèmes biologiques comme les enzymes présente un fort potentiel que je souhaite explorer à court terme.**



3. L'approche par Modes Statiques peut tout à fait être envisagée à toute échelle, pour tout système décrit à l'aide d'atomes ou de "grains", du moment que les interactions sont elles-mêmes décrites par un modèle énergétique fournissant une structure à l'équilibre et la matrice hessienne associée. En revanche, **aller vers des systèmes de très grande taille ou simuler des mouvements de grande amplitude est difficilement atteignable avec un modèle tout-atomes.** En effet, le besoin en ressources et temps de calculs nous fait perdre l'avantage d'une méthodologie simple au résultat quasi-immédiat, tel que celui que nous avons pour des systèmes de taille allant jusqu'à quelques milliers d'atomes. Par ailleurs, la méthode des MS per-

met de déterminer les déformations moléculaires dans une approximation harmonique. Leur domaine de validité est donc restreint autour d'un équilibre et ne donne pas accès à des déformations de grande amplitude, ni à un nouvel équilibre. Les Modes Statiques nous ont cependant permis de la dépasser cette limitation pour observer des changements de conformation sous l'action de forces extérieures. Cependant, cette démarche requiert une multiplication du nombre de matrices de Modes Statiques pour tenir compte des effets non linéaires dus à l'anharmonicité.

Une démarche permettant de dépasser ces limites consiste à remplacer les coordonnées cartésiennes par des coordonnées généralisées, au sens de Lagrange. Cette approche est particulièrement intéressante dans le cas des biomolécules constituées d'une chaîne de base, repliée par l'intermédiaire de chaînes latérales. On peut alors considérer que **l'essentiel des variations d'énergie, à température ambiante, provient des variations de torsions, peu dépendantes les unes des autres.**

L'utilisation des torsions a prouvé son efficacité pour modéliser les biomolécules, comme nous l'avons discuté dans la partie 3.2 (moteurs moléculaire). Le modèle énergétique utilisé, ENCAD, a été le premier de ce type. Il a été développé dans les années 80 par M. Levitt, et continuellement amélioré depuis, dans le cadre de son utilisation en dynamique moléculaire et analyse par modes normaux. Les modèles de torsions s'avèrent plus efficaces que les modèles cartésiens pour la recherche de changements conformationnels [36, 55]. Non seulement ils permettent d'échantillonner explicitement les degrés de liberté les plus pertinents, mais ils autorisent des pas de temps d'intégration plus importants, par élimination des degrés de liberté de liaison et d'angle dont les variations nécessitent une résolution temporelle de l'ordre de la femtoseconde. Ils demeurent cependant peu représentés dans la littérature, dans laquelle on retrouve également des modèles de torsions très simplifiés (sans distinction de séquence), couplés à des approches empruntées à la robotique et basées sur la cinématique inverse [10, 83].

En fixant les longueurs et angles de liaison et en utilisant des coordonnées internes, **le nombre de degrés de liberté est drastiquement réduit. Cependant, une simplification excessive entraîne une perte d'informations physico-chimiques** et les torsions ne présentent un réel avantage pour nous que si elles remplissent simultanément ces trois conditions : i) réduire le coût de calcul, ii) traduire la propagation de déformations pertinentes à longue distance, iii) préserver une description physico-chimique du système. C'est le cas du modèle ENCAD dont les champs de force ont été développés pour la simulation de biomolécules, avec une représentation à la fois tout-atomes et gros-grains, tout en offrant la possibilité de prendre en compte le solvant.

En revanche il n'est paramétré que pour les protéines et les acides nucléiques, ce qui nous incite à une réflexion autour de l'utilisation de ce modèle avec lequel nous ne pourrions pas traiter tous les systèmes. Ainsi, à moins de paramétrer nous-mêmes le champ de force, ce qui est réalisable mais constitue un projet en soi, nous devons aussi envisager d'utiliser d'autres modèles.

Une autre solution consiste à passer à un modèle gros-grains. On conserve alors une représentation cartésienne à $3N$ degrés de liberté, mais avec un nombre N de grains nettement inférieur au nombre d'atomes du système. C'est ce que nous avons récemment entrepris dans le cadre d'une collaboration interne avec P. Joseph sur la simulation de membranes biomimétiques constituées par auto-association et auto-organisation de phospholipides. L'étude de leur structure et de leurs propriétés mécaniques est un prérequis nécessaire à la compréhension de la façon dont elles vont se déformer et se comporter lors d'interactions. Ces interactions présentent un fort intérêt applicatif, en particulier pour les problèmes de nanotoxicité (interaction avec des nanoparticules qui peuvent modifier leur perméabilité et leurs propriétés mécaniques), mais aussi en nanomédecine (pénétration de nanovecteurs et libération de molécules actives au travers de la membrane). Elles peuvent par ailleurs elles-mêmes servir de nanovecteurs, pour l'encapsulation et la délivrance de médicaments par exemple. La modélisation à l'échelle moléculaire peut apporter des éléments de compréhension quant aux propriétés intrinsèques des membranes, et quant à la façon dont elle vont être affectées lors d'une interaction, ou dont elles vont conditionner et régir cette interaction.

La représentation tout-atomes n'est pas appropriée pour ce type d'étude car elle nous restreint à des tailles de systèmes qui ne nous permettent pas d'accéder aux grandeurs caractéristiques mesurées lors des expériences microfluidiques (stage R. Bonhoure). Nous avons donc entrepris de travailler sur des systèmes gros-grains, à l'aide du champ de force Martini [37,146] et du logiciel Gromacs [2,205] (stage V. Combette). Notre objectif à court terme est d'adapter FleXible pour obtenir les Modes Statiques sur des membranes de différentes compositions, de les utiliser pour soumettre ces membranes à diverses forces (élongation, cisaillement...) et d'extraire des déformations induites des grandeurs comparables aux mesures expérimentales, le module de courbure en particulier [74]. Ce travail va notamment nous demander de revoir la procédure par laquelle nous appliquons les forces. A plus long terme, ce travail doit s'intégrer, dans le cadre d'une collaboration avec M. Abbas (LGC, toulouse), dans une approche multi-échelles qui utilisera la dynamique des particules dissipatives (DPD) pour étudier les interactions membrane-nanoparticule en prenant en compte les effets thermiques et hydrodynamique. Il s'agit pour nous d'un défi méthodologique et applicatif, dans le cadre d'une nouvelle collaboration théorie-expérience, et qui prolonge nos activités côté biotechnologies.

En parallèle, je souhaite particulièrement continuer à m'engager sur les activités "cibles thérapeutiques" qui sont au cœur de mon expertise et de ma motivation. Dans la partie 2, j'ai donné les perspectives envisagées pour chaque projet décrit et n'y reviendrai pas ici. Cependant, ces perspectives peuvent être étendues ici, chaque cas d'étude pouvant profiter des développements proposés dans cette dernière partie. Donnons pour exemple l'application du QM/SM aux réactions enzymatiques au sein des protéines. Toujours dans le cadre de la thèse de L. Foulon, nous essayons avec beaucoup de difficultés d'appliquer la procédure qu'il a développée au cas de l'hydrolyse du GTP dans la protéine NRas. L'approche DFT qu'il utilise nous contraint à tronquer le système pour ne garder que la cavité de NRas, et donc à faire des choix sur des résidus exclus des calculs ou fixés pour garantir l'intégrité de la structure du site actif. Ces biais, associés à la complexité intrinsèque du système, ne nous permettent pas pour l'instant d'avoir des résultats satisfaisants. Un autre manque important à ce type d'approche est la dynamique du système qui est nécessaire pour favoriser la réaction.

Il faut donc envisager, pour appliquer le QM/SM aux réactions enzymatiques dans les protéines, d'utiliser les Modes Statiques, non plus pour guider la DFT, mais pour **guider un QM/MM dynamique**. Cette approche pourra être appliquée à NRas, cGAS, ou d'autres systèmes sur lesquels nous avons récemment commencé de travailler. A une autre échelle, le couplage SM/gros-grains pourra être utilisé pour étudier des changements conformationnels, les mouvements de boucles, etc. La mécanique d'activation de la protéine cGAS par l'ADN est un cas particulièrement intéressant à étudier avec cette approche. Nous l'envisageons pour la suite de nos travaux.

De façon générale, les efforts sont actuellement dirigés vers la **recherche de nouvelles stratégies thérapeutiques allant au-delà du simple antagonisme** [175,235]. Ceci est vrai pour atteindre de nouvelles cibles, ou encore pour contrer les échecs thérapeutiques dus à des mutations qui rendent les médicaments inefficaces à terme. Proposer de nouvelles stratégies nécessite une connaissance détaillée de l'activité des enzymes, mais aussi la proposition d'outils permettant de solliciter les cibles et d'anticiper leur réponse, qu'elle soit chimique ou structurale. Les développements proposés peuvent répondre à ce besoin et permettent de l'aborder sous un angle original.

Au fil des projets et des besoins, nous avons rendu le logiciel FleXible compatible avec les sorties de nombreux codes couramment usités par la communauté, tant en mécanique classique (AMBER, Gromacs, etc) qu'en DFT (Gaussian, Orca, VASP, Quantum Espresso, xTB, etc). Nous avons par ailleurs développé une variété de codes d'exploitation des Modes Statiques. **Rassembler l'ensemble de ces outils et les rendre accessibles par une interface graphique dédiée, permettant le calcul et la visualisation interactive des déformations moléculaires, est un passage obligatoire pour rendre nos outils ergonomiques et utilisables par une communauté plus large**. Ce développement devra aussi intégrer, de façon incrémentale, les nouveaux développements disponibles, tels que les couplages QM/SM et SM/gros-grains. Nous avons amorcé cet effort avec le stage de P. Kobina. De façon préliminaire, il a implémenté FleXible dans SAMSOM, un logiciel de conception assistée par ordinateur de nanosystèmes (S. Redon, INRIA Grenoble⁷) et proposé un démonstrateur. Nous réfléchissons actuellement à poursuivre dans cette voie ou à proposer un outil indépendant. Au final, le résultat sera une forte plus-value qui nous permettra d'acquérir une meilleure visibilité et de proposer un véritable outil de manipulation *in silico* des molécules.

*"En essayant continuellement on finit par réussir.
Donc : plus ça rate, plus on a de chance que ça marche"* [189]

7. <https://www.samson-connect.net>

