



**HAL**  
open science

# Compréhension du mécanisme d'action de la P28GST dans le traitement des maladies inflammatoires auto-immunes

Corentin Bedart

► **To cite this version:**

Corentin Bedart. Compréhension du mécanisme d'action de la P28GST dans le traitement des maladies inflammatoires auto-immunes. Médecine humaine et pathologie. Université de Lille, 2021. Français. NNT: 2021LILUS044. tel-03891060

**HAL Id: tel-03891060**

**<https://theses.hal.science/tel-03891060>**

Submitted on 9 Dec 2022

**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.

UNIVERSITE DE LILLE  
Inserm U1286 INFINITE – UFR3S – Faculté de Pharmacie de Lille  
Par'Immune

**THESE**  
Pour l'obtention du grade de  
**DOCTEUR DE L'UNIVERSITE DE LILLE**  
*Discipline : Biomolécules, pharmacologie, thérapeutique*

*Présentée et soutenue le 08 Décembre 2021*

*Par*

**Corentin BEDART**

**COMPREHENSION DU MECANISME D'ACTION DE LA  
P28GST DANS LE TRAITEMENT DES MALADIES  
INFLAMMATOIRES AUTO-IMMUNES**

**Membres du jury :**

**Rapporteurs**

**Dr. ANDRE Gwénaëlle**, Directrice de recherche à l'INRAE

**Dr. COLLOC'H Nathalie**, Chargée de recherche à l'Université Caen-Normandie

**Examineurs**

**Pr. DAUCHEZ Manuel**, Professeur à l'Université de Reims Champagne-Ardenne

**Dr. RIVEAU Gilles**, Directeur de recherche au Centre de Recherche Biomédicale Espoir Pour La Santé de Saint Louis

**Dr. REGAD Leslie**, Maître de conférences à l'Université de Paris

**Membre invité**

**Dr. LACHGAR Abderrahim**, Par'Immune

**Directeurs de thèse**

**Pr. CAPRON Monique**, Professeur à l'Université de Lille

**Dr. FARCE Amaury**, Maître de conférences à l'Université de Lille



*« Optimiste : Qui est porté, par son caractère, à voir tout en bien, à s'estimer satisfait de ce qui arrive, à être confiant dans l'avenir. »*

Dictionnaire Larousse 2021



## REMERCIEMENTS

Je tiens tout d'abord à remercier sincèrement le Dr. Amaury FARCE et le Pr. Monique CAPRON pour m'avoir fait l'honneur d'encadrer mes travaux de thèse, de m'avoir fait confiance et de m'avoir encouragé jusqu'au bout. Les trois ans à vos côtés m'ont apporté énormément, aussi bien scientifiquement qu'humainement.

Je tiens également à remercier vivement l'entreprise Par'Immune, le Dr. Abderrahim LACHGAR, le Dr. Adeline PORCHERIE et Lionel ARNAUD pour leur accueil et leur soutien lors de ces trois années de thèse.

Merci aux membres composant mon jury de thèse : le Dr. Gwénaëlle ANDRE-LEROUX et le Dr. Nathalie COLLOC'H pour avoir accepté de juger mes travaux de thèse en tant que rapporteurs, ainsi que le Pr. Manuel DAUCHEZ, le Dr. Gilles RIVEAU et le Dr. Leslie REGAD pour avoir accepté d'examiner ces travaux le jour de ma soutenance.

Je remercie également le Pr. Matthieu MONTES, non présent dans ce jury, pour avoir accepté de suivre mes travaux ces dernières années aux côtés du Pr. DAUCHEZ et du Dr. REGAD. Vos jugements, vos conseils et vos pistes d'améliorations m'ont été très utiles et formateurs.

Je remercie le Dr. Laurent DUBUQUOY, le Pr. David LAUNAY, le Dr. Virginie CHAMEROY et Bernadette LEU pour m'avoir accueilli au sein de l'unité Inserm U1286 INFINITE. Je remercie également le Pr. Philippe CHAVATTE de m'avoir accueilli à l'Institut de Chimie Pharmaceutique Albert Lespagnol, de sa bienveillance à mon attention, et de m'avoir fait l'honneur d'encadrer ma thèse de docteur en Pharmacie.

Un grand merci au Dr. Nicolas RENAULT pour avoir supervisé au jour le jour mes travaux aux côtés du Dr. Amaury FARCE. Merci pour vos précieux conseils, votre compagnie, votre soutien, les moments mémorables, et pour avoir joué les inhibiteurs sélectifs de ma négativité.

Merci à tous mes collègues de la faculté de Pharmacie, de la faculté de Médecine et de l'ICPAL. Merci à Amélie, Annie, Claire, Frédérique, Guillaume, Laurence, Morgane, Min-Jeong, Natascha, Perrine, Raphaël, Régis, Solenne, Xavier, et tous ceux que j'oublie, pour les merveilleux moments. Je vous promets également d'être plus positif.

Merci à tous les différents acteurs de mon cursus universitaire ces 10 dernières années, à la faculté de Pharmacie de l'Université de Lille et au Master 2 In Silico Drug Design de l'Université de Paris pour m'avoir permis de développer ma double compétence entre les sciences pharmaceutiques et les méthodes *in silico*. J'adresse plus particulièrement ma gratitude au Dr. Marina PLAETEVOET, au Pr. Rebecca DEPRESZ-POULAIN, au Dr. Nicolas RENAULT, au Dr. Marc LENSINK et à l'entreprise Pierre Fabre de m'avoir encadré lors des différentes expériences professionnelles de mon cursus pharmaceutique.

Enfin, un grand merci à mes parents, ma sœur, mon frère, et plus généralement toute ma famille pour m'avoir encouragé à devenir ce que je suis aujourd'hui et soutenu dans les moments les plus complexes. Je ne le dis que très rarement, mais je vous aime sincèrement.

Enfin, merci à Pauline pour la joie et le bonheur qu'elle m'apporte au quotidien, sa patience et son support. J'ai une chance incommensurable de t'avoir à mes côtés. Je t'aime.



## TABLE DES MATIERES

<b>Introduction</b> .....	1
I. Historique de la P28GST .....	3
1. Bilharziose .....	3
2. A la recherche d'antigènes thérapeutiques - Découverte des 28GST.....	5
3. Développement clinique - De Sh28GST à Bilhvax .....	8
4. Du vaccin au rôle immunorégulateur - P28GST.....	9
5. Brevet dans les maladies inflammatoires.....	12
6. Maladies Inflammatoires Chroniques de l'Intestin (MICI) .....	12
7. Essai clinique de phase 2a - ACROHNEM .....	15
8. Valorisation industrielle – Création de Par'Immune .....	16
II. Glutathion S-Transférases .....	18
1. Classification des GST.....	18
2. Structure des GST cytosoliques .....	21
a. Domaines structuraux.....	21
b. Interface dimérique.....	23
c. Site actif.....	23
III. Caractéristiques de la P28GST .....	25
1. Place de la P28GST dans la classification des Glutathion S-Transférases .....	25
2. Glutathion S-Transférases de classe sigma.....	26
3. Mécanismes enzymatiques des GST de classe sigma .....	28
a. Activité Glutathion S-Transférase :.....	29
b. Activité Prostaglandine-D2 Synthase.....	30
c. Spécificités entre espèces .....	31
4. Spécificités structurales de la P28GST .....	32
5. Mode d'action anti-inflammatoire supposé.....	34
<b>Objectifs</b> .....	37
<b>Chapitre I - Etude structurale de la P28GST par dynamique moléculaire</b> .....	41
I. Travaux préliminaires.....	43
1. Ressources et logiciels principaux .....	43
2. Préparation des structures cristallographiques .....	44
II. Dynamique moléculaire .....	46
1. Introduction.....	46
2. Matériels & méthodes .....	48



a.	Simulations sous GROMACS .....	48
b.	Simulations sous Amber.....	48
3.	Résultats .....	50
a.	Premières dynamiques moléculaires avec Gromacs .....	50
b.	Dynamiques moléculaires de 150 ns avec Amber18.....	52
c.	Dynamiques moléculaires de 1.5 $\mu$ s avec Amber18 .....	56
4.	Alternatives à la dynamique moléculaire classique .....	57
a.	Méthode du recuit simulé .....	57
b.	Dynamique moléculaire dirigée .....	60
III.	Métadynamique .....	62
1.	Matériels et méthodes .....	62
a.	Métadynamique standard .....	62
b.	Métadynamique bien tempérée .....	64
c.	Dynamique Moléculaire à Biais Adaptatif - ABMD.....	67
2.	Résultats .....	69
a.	ABMD .....	69
b.	Métadynamique avec PLUMED .....	77
 <b>Chapitre II - Outil d'étude des réseaux d'interactions moléculaires et son application ..</b>		<b>95</b>
I.	Développement de SINAPs.....	97
1.	Introduction .....	97
2.	Article SINAPs .....	97
II.	Application de SINAPs à la recherche d'un cation divalent.....	134
1.	Rationnel sur la possible présence d'un cation divalent chez P28GST .....	134
2.	Recherche de cations divalents par ICP-MS.....	140
3.	Recherche <i>in silico</i> du positionnement du cation Zn <sup>2+</sup> .....	143
a.	Dynamiques moléculaires classiques analysées par SINAPs.....	143
b.	Dynamiques moléculaires classiques en présence de Zn <sup>2+</sup> .....	144
 <b>Chapitre III - Recherche d'inhibiteurs sélectifs de l'activité PGDS .....</b>		<b>149</b>
I.	Recherche du mode de liaison de la prostaglandine H2.....	153
1.	Matériels et méthodes .....	153
a.	Principe de l'amarrage moléculaire et du criblage virtuel .....	153
b.	Algorithme génétique & GOLD.....	153
c.	Amarrage moléculaire pour identifier le mode de liaison de la prostaglandine H2 155	
2.	Amarrages moléculaires.....	157

a.	Prostaglandine H2 .....	157
b.	Fragment de la prostaglandine H2.....	158
c.	Intermédiaire réactionnel chez P28GST.....	159
d.	Intermédiaire réactionnel chez hPGDS .....	160
e.	Discussion .....	161
3.	Dynamique moléculaire à partir d'un résultat d'amarrage .....	163
II.	Réalisation du criblage virtuel à haut débit .....	166
1.	Matériels et méthodes .....	166
a.	Choix des chimiothèques.....	166
b.	Protocole de criblage virtuel.....	170
2.	Résultats du criblage virtuel - ZINC .....	173
a.	Répartition des molécules dans le jeu de données entier .....	173
b.	Répartition des molécules dans le jeu de données ne possédant que 3 poses de criblage .....	176
c.	Jeu de 922 molécules.....	177
d.	Jeu issu du GoldScore pondéré .....	180
3.	Résultats du criblage virtuel - CNE & CN.....	184
a.	CNE.....	184
b.	CN .....	185
III.	Amarrage moléculaire des résultats du criblage virtuel .....	186
1.	Matériels et méthodes .....	186
2.	Résultats amarrage moléculaire - ZINC.....	187
a.	Critères appliqués lors du criblage virtuel.....	187
b.	Toxicité.....	187
c.	Différenciation Site H/Site L.....	188
d.	Clusterisation.....	190
e.	Disponibilité commerciale .....	190
f.	Inspection visuelle.....	191
g.	Sélection finale .....	191
3.	Résultats docking - CNE et CN .....	193
	<b>Discussion et perspectives .....</b>	<b>195</b>
1.	Dynamique moléculaire .....	197
a.	Observation structurale d'un mouvement de la boucle catalytique .....	197
b.	Dynamiques moléculaires classiques plus longues .....	199
c.	Dynamiques moléculaires avancées .....	200
2.	Métadynamique.....	201

3.	SINAPs et étude des interactions intramoléculaires .....	202
a.	Interactions de l'interface dimérique.....	203
b.	Réseaux d'interactions organisant la structure tertiaire .....	203
c.	Réseaux d'interactions intramoléculaires différentiels .....	204
4.	Cation divalent .....	205
5.	Détermination du site de liaison de la prostaglandine H2 (site L).....	208
6.	Mécanisme d'action enzymatique.....	210
a.	Initiation des activités enzymatiques.....	210
b.	Activité Glutathion S-Transférase (GST).....	211
c.	Activité Prostaglandine-D2 Synthase (PGDS).....	211
7.	Campagne exploratoire d'identification d'inhibiteurs spécifiques de l'activité PGDS213	
a.	Campagne de criblage virtuel à haut débit .....	213
b.	Validation par amarrage moléculaire et sélection des composés .....	214
	<b>Conclusion</b> .....	215
	<b>Bibliographie</b> .....	219

## LISTE DES FIGURES

Figure 1. Distribution mondiale des bilharzioses en 2017 (1) .....	3
Figure 2. Répartition géographique des différentes espèces de schistosomes (8).....	5
Figure 3. Alignement de séquences de différentes 28GST : Schistosoma bovis (Sb28GST) Schistosoma haematobium (Sh28GST), Schistosoma mansoni (Sm28GST) et Schistosoma japonicum (Sj28GST). Les épitopes majeurs sont mis en évidence par une ligne au dessus des séquences correspondantes.....	6
Figure 4. Effet sur les cytokines pro-inflammatoires (à gauche) et sur les marqueurs des cellules T et des macrophages M2 (à droite) de la prévention de la colite induite par le TNBS chez la souris par l'immunisation par la P28GST à différentes doses (45) .....	10
Figure 5. Schéma du paradigme de l'helminthothérapie (48).....	11
Figure 6. Mesure (A) de l'activité Glutathion S-Transférase par l'utilisation de CDNB (1-chloro-2,4-di-nitrobenzene), (B) de l'activité anti-oxydante par l'utilisation du radical ABTS (acide 2,2'-azino-bis), et (C) du score clinique de Wallace pour la P28GST inactivée par la chaleur (dP28GST), la P28GST intacte et la GST humaine placentaire (GST HP) (45).....	11
Figure 7. Schéma des lésions du tractus digestif causées par les MICI - Adapté de (60).....	14
Figure 8. Évolution du CDAI au cours de l'essai clinique ACROHNEM (64).....	16
Figure 9. Biotransformation des xénobiotiques chez l'Homme (67).....	18
Figure 11 : Schéma de la structure de la thiorédoxine, et des repliements apparentés à la thiorédoxine comprenant les GST cytosoliques et GST de la classe Kappa (79) .....	22
Figure 12. Arbre phylogénique non enraciné de certaines PGDS hématopoïétiques et de certaines GST de classe sigma, pi, mu et alpha. Sm28GST correspond à la catégorie "Blood fluke 28kD" (90).....	25
Figure 13. Effet de la présence d'un ion métallique sur l'activité PGDS de la GST de classe sigma humaine, HPGDS (95).....	27
Figure 14. Représentations tridimensionnelles de la coordination métallique chez la GST de classe sigma humaine, HPGDS (95). La forme liée à un ion calcium Ca <sup>2+</sup> et stabilisé par 5 molécules d'eau est représentée à gauche, et la forme liée à un ion magnésium Mg <sup>2+</sup> et stabilisé par 6 molécules d'eau est représentée à droite. L'arginine participant à l'activité PGDS est Arg14.....	27
Figure 15. Schéma mécanistique de l'activité Glutathion S-Transférase (96). X : Substrat ; GSH : glutathion sous sa forme réduite ; X-GSH : glutathion sous sa forme conjuguée au substrat.....	29
Figure 16. Schéma mécanistique de l'activité Prostaglandine-D2 Synthase retrouvée chez les GST de classe sigma (97).....	30

Figure 17. Schéma des différentes conformations de la tyrosine 10 catalytique (PDB ID : 1OE8) (80).....	32
Figure 18. Mécanisme d'action supposé de la tyrosine 10 catalytique (101). .....	33
Figure 19. Mode de liaison supposé de la prostaglandine H2 dans le site L de HPGDS (90). 34	
Figure 20. Voies biologiques supposé impactées par la présence de la P28GST via la production de prostaglandine D2. ....	35
Figure 21. Représentation tridimensionnelle de la structure cristallographique de P28GST (identifiant PDB : 1OE8) en monomère au début de la production (à gauche) et après 17 ns de production lors de la réplique 4 (à droite), avec représentés en rose la Tyrosine 10 et en cyan le glutathion.....	51
Figure 22. Graphique de l'évolution du volume du canal formé à proximité de la Tyrosine 10 lors des 4 répliques de dynamique moléculaire réalisées sous Gromacs. ....	51
Figure 23. Représentation des conformations IN (en vert) et OUT (en jaune) de la tyrosine 10 observées dans la cristallographie de P28GST (identifiant PDB : 1OE8) et de la conformation MID (en bleu) observée lors des productions de dynamique moléculaire sous AMBER18....	53
Figure 24. Distribution de l'angle dièdre Chi1 déterminé à partir de 25 490 tyrosines décrites dans la librairie de rotamères Dunbrack 2010 (132), les bornes à 5% bilatérales pour chaque rotamère, et la correspondance de chaque conformation de la Tyrosine 10 observée .....	54
Figure 25. Suivi de l'angle dièdre Chi1 (en degré) au cours des productions de dynamique moléculaire réalisées sous AMBER18 .....	55
Figure 26. Évolution de l'angle dièdre Chi1 au cours des 3 répliques de dynamique moléculaires classique de 1.5 $\mu$ s réalisées sous AMBER18 .....	56
Figure 27. Figure théorique du recuit simulé (135).....	57
Figure 28. Figures expérimentales du recuit simulé réalisé chez P28GST .....	58
Figure 29. Évolution des angles dièdres Chi1 et du RMSD des carbones alpha au cours des simulations de recuit simulé .....	59
Figure 30. Évolution de l'angle dièdre Chi1 et du RMSD des carbones alpha au cours de la simulation de dynamique moléculaire dirigée.....	61
Figure 31. Équation du potentiel de biais V amené par le logiciel construit dans l'espace des variables collectives $\mathbf{s}$ , exprimé comme étant la somme des gaussiennes déposées dans l'espace des variables collectives avec $\tau$ correspondant au pas de dépôt des gaussiennes, $\sigma_i$ la largeur des gaussiennes et $W(k\tau)$ la hauteur des gaussiennes.....	63
Figure 32. Le potentiel de biais V converge vers l'opposé de l'énergie libre F du système....	63
Figure 33. Illustration schématique d'une simulation de métadynamique (137). Le système a débuté au niveau de la conformation A, au niveau d'un minimum énergétique. Au cours de la simulation, des biais énergétiques sont ajoutés au paysage énergétique sous la forme de	

gaussiennes (ici en rouge), remplissant progressivement le puits de potentiel de la conformation de départ. Lorsque la simulation atteint un certain point où l'énergie du système est élevée, l'algorithme tend à favoriser l'exploration du paysage pour rechercher une nouvelle conformation énergétiquement plus favorable en franchissant les barrières énergétiques, ici vers la conformation B. La simulation continue ensuite jusqu'à remplir le paysage énergétique et l'aplatir totalement. La somme de tous les biais énergétiques apporté au système permettent d'estimer le négatif du paysage énergétique (en tirets gris), permettant de retrouver les minima énergétiques et les barrières énergétiques du système, comme dans cet exemple pour obtenir l'information sur la barrière énergétique C conditionnant le passage entre les conformations A et B. .... 64

Figure 34. Variation de la hauteur des gaussiennes  $W(k\tau)$  lors de la simulation, avec  $W_0$  la hauteur initiale de la gaussienne,  $\Delta T$  la différence de température amenée pour l'exploration de l'espace des variables collectives et  $k_B$  la constante de Boltzmann. .... 65

Figure 35. Le potentiel de biais  $V$  construit dans l'espace des variables collectives converge vers l'opposé d'une fraction de l'énergie libre  $F$  du système, fonction de la température  $T$  de simulation et de la température  $T+\Delta T$  des variables collectives. .... 65

Figure 36. Expression du facteur de biais. .... 65

Figure 37. Profils d'énergie libre obtenus lors des 4 répliques de 150 ns d'ABMD ayant pour variable collective l'angle dièdre  $\text{Chi1}$ . .... 70

Figure 38 Représentation de la tyrosine 10 en début de simulation (en beige et jaune - angle dièdre de  $191^\circ$  correspondant à la conformation IN) et au bout de 25 ns de simulation (en bleu - angle dièdre de  $109^\circ$  s'approchant des angles observés dans la conformation OUT) dans la réplique #2 de simulation d'ABMD. .... 71

Figure 39. Exploration de l'espace des deux variables collectives  $\text{Chi1}$  et  $\text{Chi2}$  au sein de deux répliques d'ABMD de 200ns. .... 72

Figure 40. Évolution du profil d'énergie libre dans la simulation d'ABMD de 823 ns. Une ligne correspond à 50ns de simulation, allant du rouge pour le début de la simulation au violet pour la fin de la simulation. .... 73

Figure 41. Évolution de la magnitude des potentiels de biais déposés lors des simulations d'ABMD de 823 ns avec un pas de dépôt de 100 ps (en rouge), de 120 ns avec un pas de dépôt de 1 ps (en vert) et de 10 ns avec un pas de dépôt de 0.01 ps. .... 74

Figure 42. Profils d'énergie libre en fonction de la variable collective  $\text{Chi1}$  obtenus pour les simulations d'ABMD de 823 ns avec un pas de dépôt de 100 ps (en rouge), de 120 ns avec un pas de dépôt de 1 ps (en vert) et de 10 ns avec un pas de dépôt de 0.01 ps. .... 75

Figure 43. Structures de référence pour la tyrosine 10 dans sa conformation IN (en beige), MID (en bleu) ou OUT (en rose) utilisées pour le calcul des variables collectives définies par les RMSD de ces structures avec la conformation de la tyrosine 10 au cours de la simulation. .... 79

Figure 44. Évolution des différentes variables collectives issues du RMSD de la tyrosine 10 comparé aux conformations de référence dans les 4 répliques de métadynamique réalisées. . 80

Figure 45. Évolution des potentiels de biais au cours des 4 répliques de métadynamique réalisées. ....	80
Figure 46. Évolution de l'angle dièdre Chi1 dans les 4 répliques de métadynamique réalisées en utilisant les différentes variables collectives issues du RMSD de la tyrosine 10 comparé aux conformations de référence. ....	81
Figure 47. Formule de l'intervalle de confiance IC, calculé autour d'une moyenne observée $\bar{x}$ , possédant un écart-type observé $s$ , sur un échantillon de taille $n$ , et avec $t_\alpha$ la valeur du 97.5 percentile de la distribution normale standard arrondie à 1.96. ....	83
Figure 48. Profils d'énergie libre (en kJ/mol) des 40 répliques de métadynamique réalisées avec les variables collectives RMSD (en traits fins) et profil d'énergie libre moyen (en trait épais rouge) avec son intervalle de confiance 95% bilatéral (aire rouge). ....	84
Figure 49. Profil d'énergie libre moyen (en kJ/mol) des 40 répliques de métadynamique réalisées avec les variables collectives RMSD avec son intervalle de confiance 95% bilatéral (aire rouge). ....	84
Figure 50. Différents angles dièdres Chi1 (en rad) dans les 40 répliques de métadynamique réalisées avec les variables collectives RMSD. ....	85
Figure 51. Contraintes imposées entre le glutathion et certains acides aminés de la P28GST lors de certaines simulations de métadynamique ....	86
Figure 52. Différents angles dièdres Chi1 (en rad) observés dans les 41 répliques de métadynamique réalisées avec la variables collective Chi1 démarrant de la structure présentant la tyrosine 10 en conformation IN. ....	87
Figure 53. Profils d'énergie libre (en kJ/mol) des 41 répliques de métadynamique réalisées avec la variable collective Chi1 démarrant de la structure présentant la tyrosine 10 en conformation IN (en traits fins) et profil d'énergie libre moyen (en trait épais rouge) avec son intervalle de confiance 95% bilatéral (aire rouge). ....	88
Figure 54. Profil d'énergie libre moyen (en kJ/mol) des 41 répliques de métadynamique réalisées avec la variables collective Chi1 démarrant de la structure présentant la tyrosine 10 en conformation IN, avec son intervalle de confiance 95% bilatéral (aire rouge). ....	88
Figure 55. Profils d'énergie libre moyens (en kJ/mol) en fonction de l'angle dièdre Chi1 (en rad) pour 4 conditions expérimentales mettant en jeu la Tyrosine 10 protonée, pour un minimum de 20 répliques de 20 ns de métadynamique pour chaque condition. ....	90
Figure 56. Profils d'énergie libre moyens (en kJ/mol) et intervalle de confiance à 95% en fonction de l'angle Chi1 (en rad) pour la tyrosine 10 protonée dans ses conformations IN et OUT, sans glutathion. ....	91
Figure 57. Profils d'énergie libre moyens (en kJ/mol) et intervalle de confiance à 95% en fonction de l'angle Chi1 (en rad) pour la tyrosine 10 protonée dans ses conformations IN et OUT, avec glutathion ....	91

Figure 58. Profils d'énergie libre moyens (en kJ/mol) et intervalle de confiance à 95% en fonction de l'angle Chi1 (en rad) pour la tyrosine 10 protonée dans sa conformation IN, avec et sans glutathion.....	92
Figure 59. Profils d'énergie libre moyens (en kJ/mol) et intervalle de confiance à 95% en fonction de l'angle Chi1 (en rad) pour la tyrosine 10 protonée dans sa conformation OUT, avec et sans glutathion.....	93
Figure 60. Profils d'énergie libre moyen (en kJ/mol) en fonction de la variable collective Chi1 (en rad) pour 4 conditions expérimentales mettant en jeu la tyrosine 10 déprotonée, pour un minimum de 20 répliques de 20 ns pour chaque condition.....	94
Figure 61. Représentation chez hPGDS (PDB ID : 1IYH) (95) du cation Mg <sup>2+</sup> (en vert) au centre de sa couronne de molécules d'eau, stabilisées par des acides glutamique et aspartique (en jaune) par des liaisons hydrogènes (en cyan), au niveau de l'interface dimérique et à proximité du glutathion (en bleu).....	135
Figure 62. Évolution du RMSD du cation Mg <sup>2+</sup> et de sa couronne de 6 molécules d'eau lors de la courte simulation de dynamique moléculaire de 50ns.....	137
Figure 63. Évolution du RMSD des molécules de glutathion des monomères A et B lors de la courte simulation de dynamique moléculaire de 50ns .....	137
Figure 64. Évolution du RMSD du cation Mg <sup>2+</sup> lors des 4 répliques de 1000ns de dynamique moléculaire classique. ....	138
Figure 65. Représentation de la conformation finale de la P28GST obtenue lors des dynamiques moléculaires classiques de 1000ns P28GST, avec le cation Mg <sup>2+</sup> représenté au milieu de sa couronne de 6 molécules d'eau, Glu103 en vert, Glu106 en orange et His107 en cyan. ....	139
Figure 66. Evolution du RMSD du Zn <sup>2+</sup> au cours des simulations de dynamique moléculaire classique de 750 ns.....	144
Figure 67. Positionnement du cation Zn <sup>2+</sup> à la fin des simulations de dynamique moléculaire classique. ....	145
Figure 68. Impact de la présence d'un Zn <sup>2+</sup> par comparaison de la structure cristallographique de P28GST (PDB ID : 1OE8) (en bleu) d'une image issue de dynamique moléculaire en présence de Zn <sup>2+</sup> dans une configuration octaédrique (en beige).....	146
Figure 69. Positionnement des 100 poses de la prostaglandine H2 libre. Le glutathion est représenté en bleu.....	157
Figure 70. Positionnement des 100 poses de la prostaglandine H2 sous contrainte. Le glutathion est représenté en bleu.....	157
Figure 71. Positionnement des 100 poses du fragment de la prostaglandine H2 libre. Le glutathion est représenté en bleu.....	158
Figure 72. Positionnement des 100 poses du fragment de la prostaglandine H2 sous contrainte. Le glutathion est représenté en bleu.....	158



Figure 73. Positionnement des 100 poses de l'intermédiaire réactionnel chez P28GST. Le glutathion est représenté en bleu. ....	159
Figure 74. Positionnement des meilleures poses de l'intermédiaire réactionnel chez hPGDS. Le glutathion est représenté en cyan. ....	160
Figure 75. Superposition de P28GST en beige et hPGDS en bleu (et surface) .....	161
Figure 76. Représentation de la pose d'amarrage moléculaire de PGH2 sous contrainte de distance sélectionnée comme point de départ pour une simulation de dynamique moléculaire. ....	163
Figure 77. Représentation des 4 poses sélectionnées de l'intermédiaire réactionnel dans l'image #29 .....	165
Figure 78. Sélection des molécules de la chimiothèque ZINC15 .....	167
Figure 79. Préparation des molécules de la CN .....	168
Figure 80. Préparation des molécules de la CNE .....	169
Figure 81. Zone de criblage correspondant à une sphère de 20 Å autour de l'atome d'oxygène de la tyrosine 10 (en rouge). Le glutathion est représenté en bâtonnets, avec le cation Mg <sup>2+</sup> et ses 6 molécules d'eau au niveau de l'interface dimérique. ....	171
Figure 82. Exemple des fichiers de sortie obtenus à la suite du script Python3 parcourant tous les résultats de criblage.....	172
Figure 83. Répartition des molécules en fonction de leur GoldScore maximal - Jeu global .	174
Figure 84. Répartition du nombre de liaisons hydrogènes dans le jeu global de 4 441 739 composés (gauche) et le jeu de 105 026 composés ne réalisant que 3 poses de criblage (droite) .....	175
Figure 85. Répartition des molécules en fonction de leur GoldScore maximal - Jeu 3 poses	176
Figure 86. Positionnement des 922 molécules par rapport aux poses d'amarrage des intermédiaires réactionnels représentés en bâtonnets.....	177
Figure 87. Principaux moments d'inertie des 922 molécules .....	178
Figure 88. Répartition du nombre de liaisons hydrogènes dans le jeu de 105 026 composés réalisant 3 poses (gauche) et le jeu de 922 molécules (droite).....	179
Figure 89. Répartition des poids moléculaires (en Da) en fonction du GoldScore (à gauche) et du GoldScore pondéré (à droite) .....	181
Figure 90. Répartition des molécules en fonction de leur GoldScore pondéré maximal - Jeu 3 poses .....	181
Figure 91. Répartition du nombre de liaisons hydrogènes des 898 molécules issues du jeu de données GoldScore pondéré.....	182

Figure 92. Positionnement des 898 molécules issues du jeu de données GoldScore pondéré dans le site de liaison.....	182
Figure 93. Principaux moments d'inertie des molécules composant les différents jeux de données.....	183
Figure 94. Résultats du criblage virtuel de la CNE.....	184
Figure 95. Résultats du criblage virtuel de la CN .....	185
Figure 96. Emplacement des différents points de l'espace considérés pour différencier le site H du site L.....	189
Figure 101. Carte de densité électronique (en surfaces grises) de la structure cristallographique de la P28GST (PDB ID : 1OE8) avec la représentation des molécules d'eau retrouvées dans la cristallographie (sphères rouges), et représentation du cation $Zn^{2+}$ (sphère jaune) repositionné à partir d'une conformation observée lors des dynamiques moléculaires réalisées. ....	207
Figure 102. Emplacement du site L au sein de la P28GST, conditionné par le positionnement de His169.....	209
Figure 103. Mécanisme d'action enzymatique de la P28GST. Les interactions communes sont représentées en jaune, spécifiques à la conformation IN en rouge et spécifiques à la conformation OUT en bleu, calculées et présentées avec SINAPs.....	212



## LISTE DES TABLEAUX

Tableau 1 : Classification non exhaustive des glutathion S-transférases cytosoliques (68–74) .....	19
Tableau 2. Présentation des activités enzymatiques des Glutathion S-Transférases de la classe sigma. ....	28
Tableau 3. Activités spécifiques pour certaines GST de classe sigma de mammifères et de nématodes (22,93,99) .....	31
Tableau 4. Liste des structures cristallographiques de Sh28GST disponibles (80,100) .....	44
Tableau 5. Présentation des différentes conditions expérimentales réalisées en métadynamique avec la variable collective Chi1 pour l'obtention de profils énergétiques moyens.....	89
Tableau 6. Barrières énergétiques en kJ/mol observées dans les profils d'énergie libre moyens en fonction de l'angle dièdre Chi1 pour 4 conditions expérimentales mettant en jeu la Tyrosine 10 protonée.....	90
Tableau 7. Valeurs d'activité moyennes de l'activité GST (en nmol/min/mg de protéine) pour hPGDS et P28GST en l'absence ou en présence de cation Mg <sup>2+</sup> , et augmentation de l'activité relative par comparaison avec l'enzyme sans cation. ....	136
Tableau 8. Concentrations mesurées et quantité relative des différents cations recherchés dans l'échantillon de P28GST par ICP-MS, réalisé par l'équipe du Pr. Fabrice NEIERS du Centre des Sciences du Goût et de l'Alimentation de Dijon.....	141
Tableau 9. Fréquences observées pour les liaisons ioniques entre Arg16 et Glu103 ou Glu106 en fonction de la conformation de la tyrosine catalytique en utilisant SINAPs sur les dynamiques moléculaires classiques de 1.5 µs. ....	143
Tableau 10. Fréquences de certaines liaisons ayant été impactées par la présence d'un cation Zn <sup>2+</sup> , obtenues via SINAPs. ....	147
Tableau 12. Différents ligands utilisés lors des protocoles d'amarrage moléculaire. L'oxygène 2 est représenté sous la forme d'une sphère .....	155
Tableau 13. Résultats de l'amarrage moléculaire de l'intermédiaire moléculaire dans 7 images de dynamique moléculaire.....	164
Tableau 14. Répartition des molécules en fonction du nombre de poses réalisées .....	175
Tableau 15. Répartition des molécules du jeu de données intermédiaire et du jeu de données de 922 molécules en fonction de leur poids moléculaire .....	179
Tableau 16. Structure, positionnement et informations des 5 composés de la sélection finale .....	193



# **INTRODUCTION**

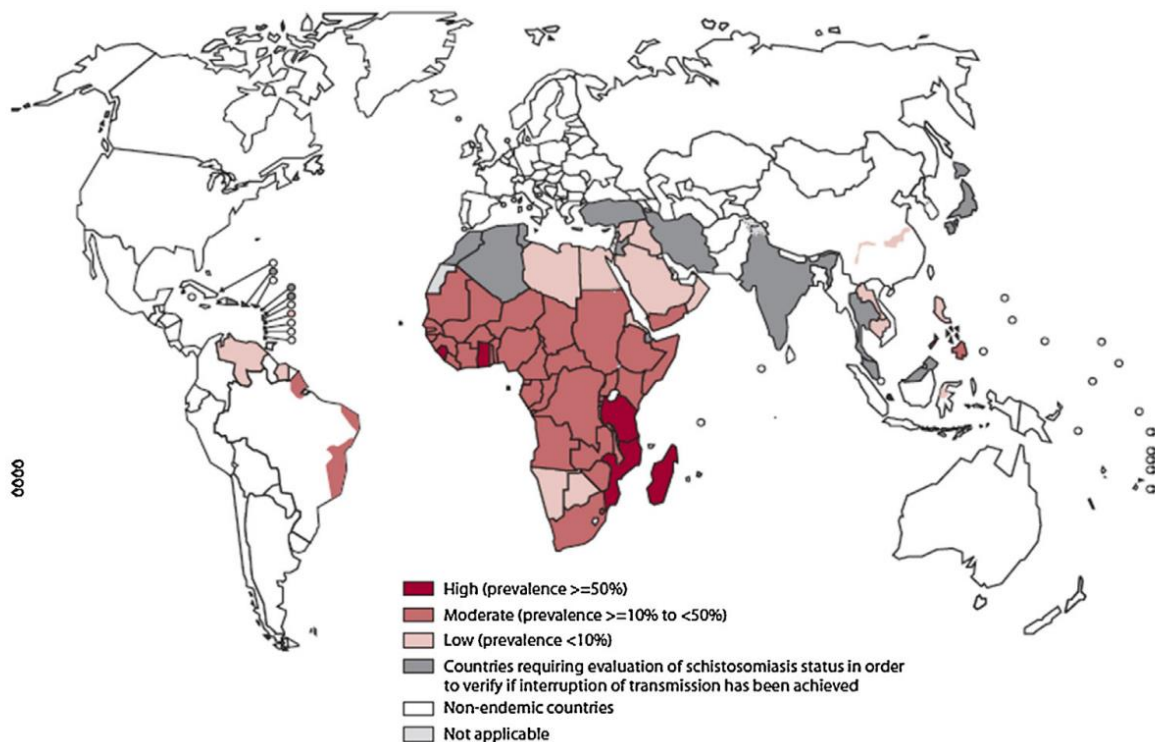


# I. Historique de la P28GST

La P28GST, ou Sh28GST, est une protéine issue de *Schistosoma haematobium*, un parasite helminthe agent de la bilharziose urinaire.

## 1. Bilharziose

La bilharziose, ou schistosomiase, est une des maladies parasitaires les plus fréquentes dans le monde, affectant plus de 240 millions de personnes dans plus de 78 pays, et menaçant plus de 800 millions de personnes (1). Retrouvée majoritairement dans les pays tropicaux et subtropicaux, elle frappe particulièrement le continent africain avec des prévalences supérieures à 50% de la population de certains pays comme le Mozambique ou Madagascar (Figure 1). La bilharziose est la 3ème maladie tropicale la plus dévastatrice au monde, et est une cause majeure de morbidité et mortalité dans ces pays.



*Figure 1. Distribution mondiale des bilharzioses en 2017 (1)*

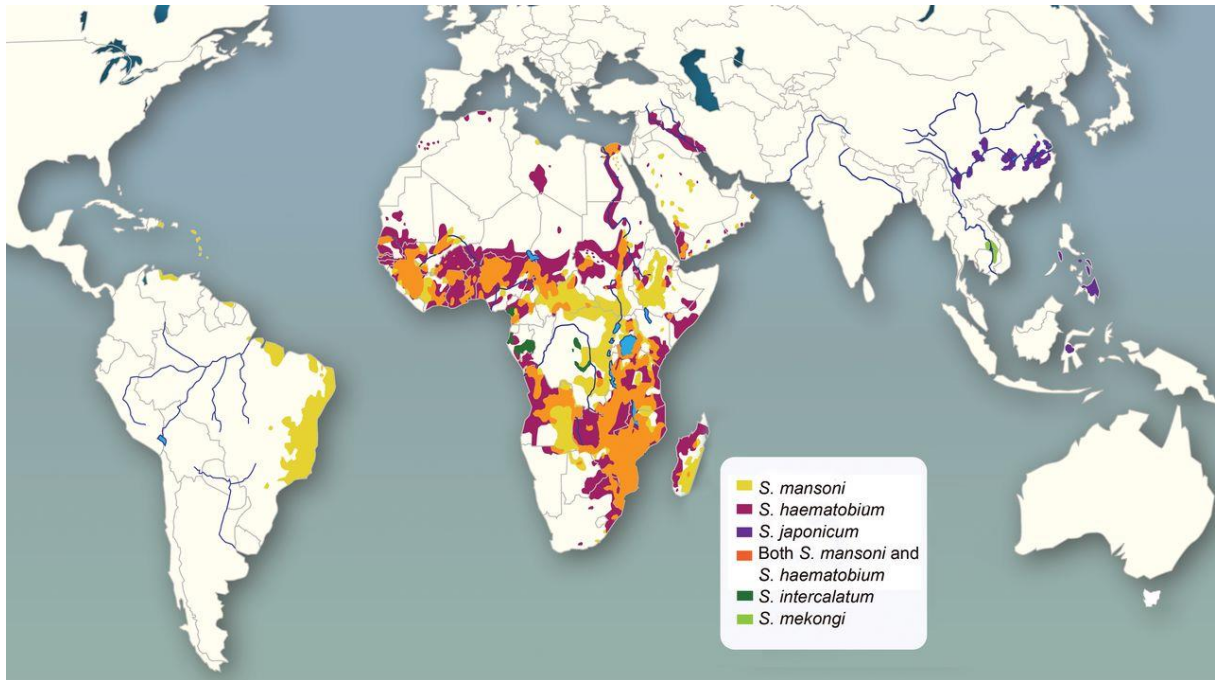
Les parasites responsables de la bilharziose sont des plathelminthes de la classe des trématodes et du genre *Schistosoma*. Les premières descriptions écrites de cette maladie remontent vers 1550 avant J.C dans le papyrus médical d'Ebers, décrivant un lien entre la



présence de vers liée à une hématurie, qui fut ensuite identifiée en 1910 par la découverte d'œufs de schistosomes dans les voies urinaires de momies égyptiennes (2–4). Le parasite a été mis en évidence pour la première fois en 1851 par le médecin Theodor Maximilian BILHARZ : il observa la présence de vers plats jusqu'alors inconnus lors d'un examen médico-légal, et leur donna le nom de *Distoma haematobium* (5). En 1858, David Friedrich WEINLAND proposa le nom *Schistosoma*, venant du grec "corps séparé", pour classer ces parasites, et nomma la maladie bilharziose en hommage aux travaux de Theodor BILHARZ (6). En 1902, le médecin Patrick MANSON a par la suite démontré l'existence de plusieurs espèces de schistosomes, avec la découverte de *Schistosoma mansoni* (7).

Parmi les espèces du parasite infectant l'Homme, *S. mansoni* et *S. haematobium* sont les deux espèces principales, causant respectivement la forme digestive et la forme génito-urinaire de la bilharziose (8). Leur transmission a lieu lors d'une exposition à des eaux douces contaminées par le parasite, avec la présence de leur hôte intermédiaire des gastéropodes aquatiques différents pour chaque espèce de schistosome (8). Le cycle parasitaire est complété par l'excrétion d'œufs par l'hôte définitif, l'Homme, dans les matières urinaires et fécales contaminant les eaux.

Ces deux parasites sont présents sur de larges zones géographiques en Afrique, et notamment au Sénégal (Figure 2) (9). La répartition géographique de cette maladie est liée à la pauvreté, avec l'accès limité aux installations sanitaires et à de l'eau potable non-contaminée, ainsi qu'à la survie de l'hôte intermédiaire obligatoire au cycle du parasite (10). Le manque d'hygiène et le contact avec de l'eau contaminée rend les enfants particulièrement vulnérables à l'infection.



**Figure 2. Répartition géographique des différentes espèces de schistosomes (8)**

Le seul traitement existant est le praziquantel (BILTRICIDE®) pour la prise en charge des bilharzioses causées par *S. haematobium*, *S. mansoni*, *S. japonicum* ou *S. intercalatum* (11). L'axe de lutte contre les bilharzioses conseillé par l'OMS est le traitement à grande échelle des populations touchées par l'administration de traitements ciblés et réguliers de praziquantel, en ciblant particulièrement les enfants, population la plus à risque (10).

## **2. A la recherche d'antigènes thérapeutiques - Découverte des 28GST**

Pour anticiper le développement de résistance au praziquantel, apparaissant de plus en plus fréquemment (12), et prévoir des alternatives thérapeutiques, les stratégies vaccinales basées sur l'utilisation de vaccins thérapeutiques représentent un espoir. Dans ce but, une stratégie de recherche d'antigènes vaccinaux a été menée pour lutter contre la bilharziose par l'équipe du Pr. André CAPRON à l'institut Pasteur de Lille. En 1987, des antigènes prometteurs ont été caractérisés chez différentes espèces de schistosomes. Ils font partie de la classe des Glutathion S-Transférases de différentes espèces de schistosomes : les 28kDa Glutathion S-

Transférases de *Schistosoma mansoni* (Sm28GST), *Schistosoma haematobium* (Sh28GST) et *Schistosoma bovis* (Sb28GST) (13).

L'administration de la Sm28GST, la première caractérisée et étudiée, chez différentes espèces animales suivie de l'infection par *S. mansoni* a permis la réduction de la charge parasitaire de 40 à 70% (14). Cet effet s'explique par la production de Sm28GST à tous les stades du développement du parasite, protéine qui est directement associée à son métabolisme (15). Suite au criblage d'une banque d'expression de *S. mansoni*, la protéine a pu être séquencée, clonée, et exprimée chez *Escherichia coli* et *Saccharomyces cerevisiae* (14,16).

Suite au clonage de Sm28GST, les autres 28GST ont pu être identifiées, séquencées et clonées (17). De fortes homologues de séquence ont pu être mises en évidence, avec 97% d'identité entre Sh28GST et Sb28GST, et 90% d'identité entre Sm28GST et Sb28GST (Figure 3). La protéine Sj28GST, identifiée chez une autre espèce de schistosome, *S. japonicum*, est bien moins homologue que les trois autres 28GST avec seulement 77% d'identité, et n'a donc pas été retenue. Ces protéines ont également démontré une activité enzymatique sous une forme dimérisée (18,19). Cette construction entraîne le fait qu'il existe des isoformes des 28GST, comme notamment Sm28GST. Cette dernière possède trois isoformes, deux homodimères SmGST-1 et SmGST-3, ainsi qu'un hétérodimère SmGST-2 composé d'un monomère SmGST-1 et d'un monomère SmGST-3 (18).

	* *		30		*		70
Sb28GST	MTGDHIKVIY	FNGRGRAESI	RMTLVAAGVN	YEDERISFQD	WPKIKPTIPG	GRLPAVKITD	NHGHVKWMLE
Sh28GST	.....	.....	.....	.....	.....	.....	.....V
Sm28GST	-A-E	-D	.....D	.....	.....	.....V	.....D
Sj28GST	V-L	.....P	.....I	.....E F	.....E	.....I	.....KR-D-T-S
		*		100	*		140
Sb28GST	SLAIARYMAK	KHHMMGETDE	EYVNVEKLI	QVEDLEHEYH	KTLMKPEEEK	QKITKEILNG	KVPVLLDIIC
Sh28GST	.....	.....G-E	.....	.....A	.....Y	.....I	.....
Sm28GST	.....	.....	.....S	.....A-V	.....Q	.....E	.....NM
Sj28GST	.....FI-R	.....N-D-D	.....II-M	.....V-S	.....I-P	.....E-S	.....I-QA
			170				211
Sb28GST	ESLKASTGKL	AVGDKVTLAD	LVLIAVIDHV	TDLDKFPLTG	KYPEIHKHRE	NLASSPRLA	KYLSDRAATP F
Sh28GST	.....	.....	.....	.....	.....	.....	.....
Sm28GST	.....G	.....	.....	.....G	.....	.....	.....N-P
Sj28GST	-T-E-N	T	.....V-S-I	.....	.....K	H-T-K	.....E-H-A

**Figure 3. Alignement de séquences de différentes 28GST : *Schistosoma bovis* (Sb28GST) *Schistosoma haematobium* (Sh28GST), *Schistosoma mansoni* (Sm28GST) et *Schistosoma japonicum* (Sj28GST). Les épitopes majeurs sont mis en évidence par une ligne au dessus des séquences correspondantes.**

L'effet immunogène des 28GST a été élucidé grâce à des peptides synthétiques issus de la séquence primaire de Sm28GST dans le but d'identifier des épitopes (20,21) et de démontrer le potentiel vaccinal de cette protéine avec la mise en évidence d'une réponse immunitaire adaptative. Trois épitopes majeurs ont pu être mis en évidence (Figure 3) :

- Un fragment N-terminal des acides aminés 24 à 43,
- Un fragment central des acides aminés 115 à 131,
- Un fragment C-terminal des acides aminés 190 à 211.

Il est à noter que les épitopes 24-43 et 190-211 sont similaires entre Sm28GST, Sh28GST et Sb28GST, mais que l'épitope 115-131 de Sh28GST n'est pas reconnu par des sérums issus d'immunisation par Sm28GST (17). De plus, malgré le fait que les régions N-terminales impliquées dans la fixation du glutathion sont homologues avec les autres GST humaines hépatiques et placentaires, aucune réaction immunologique croisée n'a été mise en évidence (22).

L'efficacité vaccinale des antigènes 28GST a été démontrée dans les études précliniques comme ayant un impact sur :

- La réduction de la charge parasitaire, qui peut atteindre 50% chez des souris et des rats (14) et jusqu'à 80% chez le singe (23). De plus, cette réduction est également croisée entre différentes espèces de *Schistosoma*, où l'administration de Sm28GST protège de *S. haematobium* (24).
- La réduction de la fécondité parasitaire par inhibition de l'activité enzymatique. Cet effet a été démontré parce que les anticorps neutralisants dirigés contre les épitopes 24-43 et 190-211, situés à proximité du site actif, ont la propriété d'empêcher l'activité enzymatique de Sm28GST (25–27).
- La prévention de la pathologie, via la diminution du nombre d'œufs pondus après immunisation par Sm28GST et l'inhibition de la synthèse de collagène hépatique limitant les atteintes pathologiques du foie après infection (28). Cette diminution est du même niveau pour Sh28GST.

### **3. Développement clinique - De Sh28GST à Bilhvax**

Suite aux essais précliniques très prometteurs, des essais cliniques ont été réalisés afin de tester le tout premier candidat vaccin contre la bilharziose au sein du programme BILHVAX. L'antigène choisi parmi ceux étudiés pour réaliser les études cliniques est Sh28GST, testé contre *Schistosoma haematobium*. Ce choix s'explique car la vaccination avec Sh28GSH ou Sm28GST entraîne une même diminution de la quantité d'œufs excrétés par le parasite, comme démontré lors des études précliniques (28). De plus, l'évaluation des lésions causées par *S. haematobium* peut être réalisée par des examens cliniques non invasifs rendant l'analyse de l'efficacité vaccinale plus facile et plus fiable (29), et l'évaluation de la quantité d'œufs présents dans l'urine des patients est plus fiable et facile que l'examen des selles (30). Enfin, la résistance à *S. haematobium* est fortement associée à l'inhibition de la fécondité du parasite, et celui-ci voit sa production d'œufs fortement réduite avec l'âge de son hôte (30,31), rendant le vaccin particulièrement utile pour les populations les plus jeunes. Il est également à noter que, du fait de l'existence d'arguments démontrant une synergie entre la réponse immune et le traitement par praziquantel, l'efficacité du vaccin a été évalué en association au traitement antihelminthique classique (praziquantel) seul, facilitant l'évaluation chez les populations les plus atteintes par bilharziose fréquemment traitées par ces traitements (30).

La forme pharmaceutique utilisée lors des essais cliniques est la protéine Sh28GST produite de manière recombinante par la levure *Saccharomyces cerevisiae* (souche TGY73-4 contenant le plasmide pTG8889) en suivant les Bonnes Pratiques de Fabrication par Eurogentec. Elle se présente sous une forme lyophilisée, à solubiliser extemporanément avec 1 mL de solution apyrogène et stérile d'alun à 0,2%. Cette formulation spécifique de Sh28GST, produite de manière recombinante et associée à de l'hydroxide d'aluminium (32,33).

Avec le concours de l'institut Pasteur de Lille, l'Institut national de la santé et de la recherche médicale (Inserm) et le centre de recherche biomédicale sénégalais "Espoir Pour La Santé" (EPLS) en 1996, un programme d'études cliniques a été lancé afin de développer le premier et unique candidat vaccin contre la bilharziose dédié en priorité aux enfants atteints infectés par *S. haematobium* (34). L'évaluation de la tolérance et de la toxicité chez des sujets sains a été réalisée de 1998 à 2000 au sein de deux études cliniques de phase 1 (Identifiant ClinicalTrials.gov : NCT00870649) (35,36). Sh28GST n'a pas induit d'effets indésirables

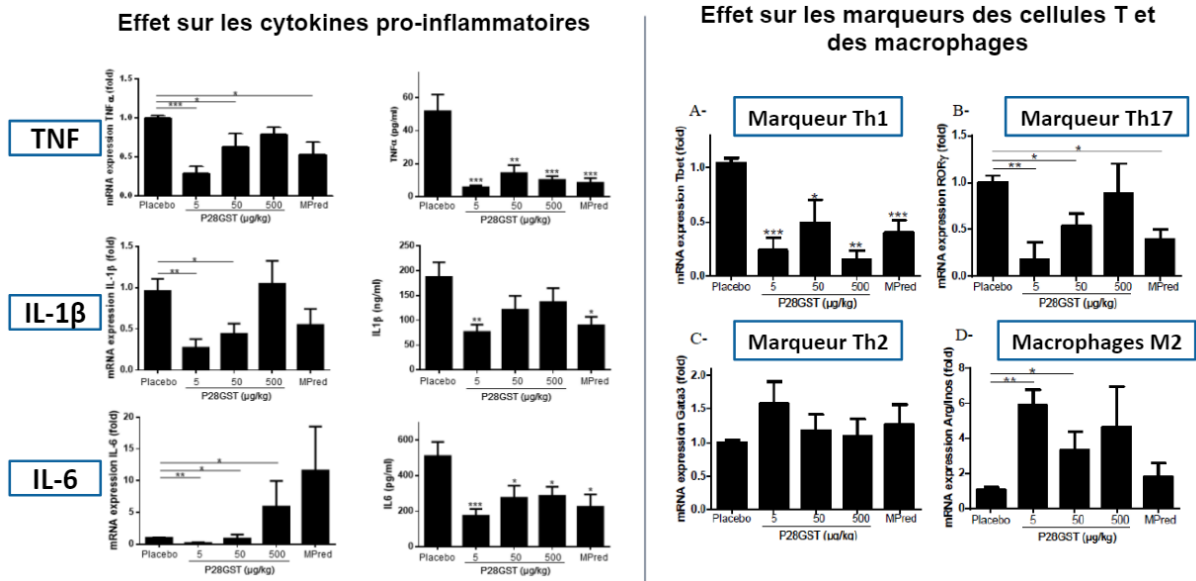
graves chez les adultes sains et a généré une réponse immunitaire de type Th2 (36). La phase 1b (Bilhvax1b) s'est intéressée à la tolérance à l'administration de Sh28GST chez des enfants africains non touchés par la bilharziose (33,37,38). Par la suite, des études cliniques de phase 2 ont démontré que plus de 80% des vaccinés ont présenté une réponse immunitaire spécifique après deux administrations à un mois d'intervalle (32,37,39).

L'évaluation de l'efficacité de Sh28GST a été réalisée dans le cadre d'un essai clinique de phase 3 (Bilhvax3) de 2009 à 2012 étudiant la diminution du risque de récurrence de bilharziose urinaire causée par *S. haematobium* pendant les 3 années suivant l'administration, ainsi que son efficacité antiparasitaire et son innocuité (Identifiant ClinicalTrials.gov : NCT00870649) (32,40). Malgré le fait que Sh28GST ait bien été immunogène et bien toléré, le critère d'évaluation principal d'efficacité n'a pas été atteint. Toutefois, l'étude a démontré que Sh28GST induisait une réponse immunitaire cohérente caractérisée par la présence d'anticorps capables d'inhiber l'activité enzymatique, donc que le candidat vaccin permet d'avoir un effet sur la fertilité de *S. haematobium* et possède un rôle protecteur.

#### **4. Du vaccin au rôle immunorégulateur - P28GST**

En 2011, dans le contexte original du rôle régulateur des parasites helminthes sur les maladies inflammatoires notamment digestives (41), l'équipe du Pr. Monique CAPRON à l'unité INSERM U995 LIRIC (*Lille Inflammation Research International Center*) devait démontrer que la Sh28GST possédait un fort potentiel anti-inflammatoire dans un modèle animal de maladie de Crohn (42,43).

En effet, il a été montré que, dans des modèles de colite chronique expérimentale induite par le TNBS (acide 2,4,6-trinitrobenzènesulfonique) chez le rat, l'immunisation par la P28GST à la dose de 25 µg/kg permet d'améliorer le score clinique de Wallace et le score histologique d'Ameho, démontrant la diminution majeure de l'inflammation locale (42,44). L'immunisation par la P28GST permet de limiter le développement de la colite aiguë induite chez le rat ou la souris en régulant l'équilibre des cytokines par la diminution des réponses pro-inflammatoires Th1 et Th17 et par l'induction de réponses anti-inflammatoire Th2 et T-régulatrice (Figure 4) (45). Toutes ces données suggéraient que la P28GST pouvait exercer un rôle immunorégulateur dans les maladies inflammatoires associées à une réponse Th1/Th17.



**Figure 4. Effet sur les cytokines pro-inflammatoires (à gauche) et sur les marqueurs des cellules T et des macrophages M2 (à droite) de la prévention de la colite induite par le TNBS chez la souris par l'immunisation par la P28GST à différentes doses (45)**

Cet effet peut également être expliqué par des arguments épidémiologiques qui suggèrent que les helminthes, dont les schistosomes, et le microbiote intestinal participent à l'hypothèse hygiéniste, qui propose un lien entre l'exposition à ces derniers lors du développement du système immunitaire et l'absence de maladies auto-immunes (46,47). Cette théorie suggère le paradigme de l'helminthothérapie, visant à réduire le déséquilibre de la réponse auto-immune et inflammatoire régie par les cellules Th1, Th17, les macrophages M1 et les cellules dendritiques conventionnelles produisant une grande quantité de cytokines pro-inflammatoires (Figure 5 - gauche) grâce à l'utilisation d'helminthes comme stratégie thérapeutique pour rétablir l'équilibre homéostatique par la diminution de la réponse inflammatoire Th1/Th17 et l'augmentation de la réponse anti-inflammatoire Th2/Treg (Figure 5 - droite) (48).

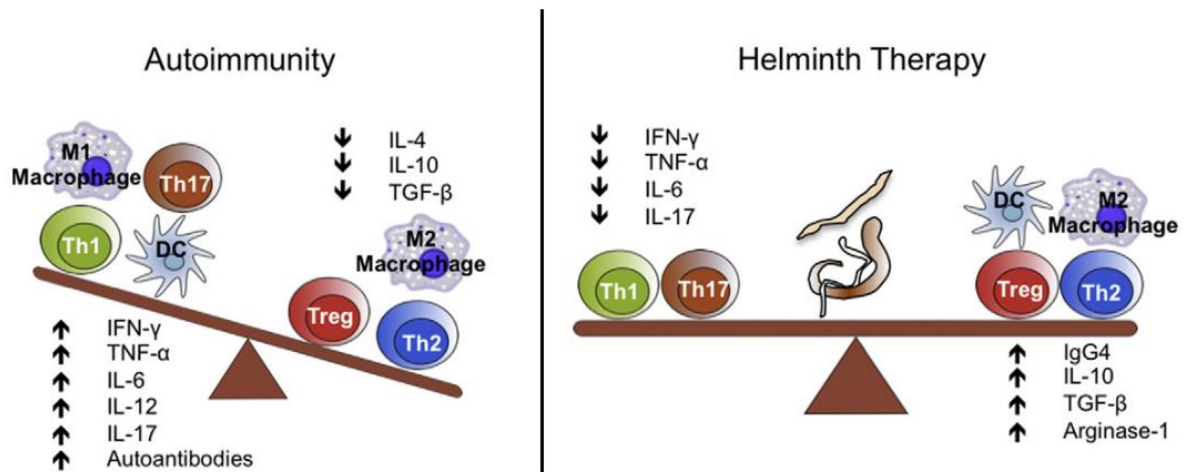


Figure 5. Schéma du paradigme de l'helminthothérapie (48)

Enfin, il a été démontré que les activités enzymatiques de la P28GST jouent un rôle majeur dans la diminution de l'inflammation causée par le TNBS, avec notamment une activité anti-oxydante. En comparant la P28GST dénaturée par inactivation thermique à la P28GST intacte et à la GST humaine placentaire, seule la P28GST intacte a été capable d'induire une réduction significative du score clinique (Figure 6) (45).

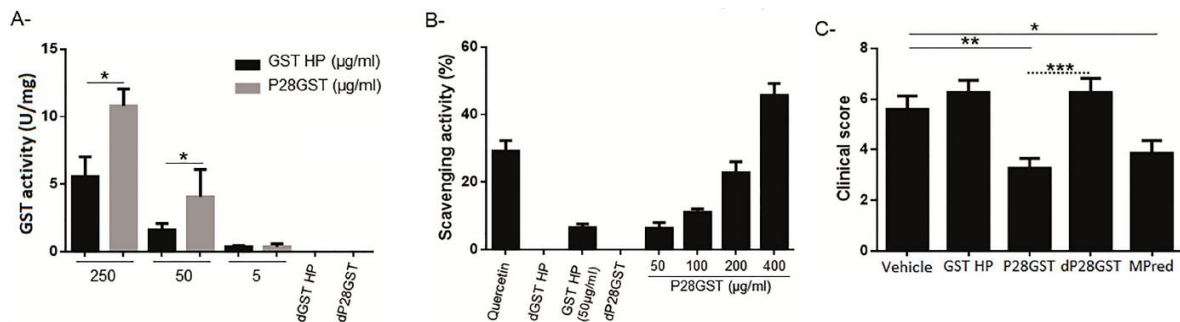


Figure 6. Mesure (A) de l'activité Glutathion S-Transférase par l'utilisation de CDNB (1-chloro-2,4-di-nitrobenzene), (B) de l'activité anti-oxydante par l'utilisation du radical ABTS (acide 2,2'-azino-bis), et (C) du score clinique de Wallace pour la P28GST inactivée par la chaleur (dP28GST), la P28GST intacte et la GST humaine placentaire (GST HP) (45).



Tous les arguments observés lors de ces essais précliniques indiquent que la P28GST représente un nouveau candidat médicament original pour le traitement des maladies inflammatoires.

## **5. Brevet dans les maladies inflammatoires**

Les propriétés anti-inflammatoires de la P28GST permettent de considérer son repositionnement pharmaceutique en tant que nouveau candidat médicament original pour le traitement des maladies inflammatoires. C'est dans cette optique qu'un brevet a été déposé le 6 février 2012 auprès de l'Institut National de la Propriété Industrielle (notice n° FR2986432) (49), puis étendu à de nombreux pays et auprès de l'Organisation Mondiale de la Propriété Intellectuelle le 6 février 2013 (notamment la notice n° WO2013117860) (50).

Ce brevet, intitulé "Protéines 28KDa GST provenant de schistosomes pour leur utilisation dans le traitement des maladies inflammatoires auto-immunes engendrant une réponse de type Th1 et/ou Th17", protège l'utilisation des glutathion S-transférases provenant de *S. haematobium*, *S. mansoni* et *S. bovis* dans les maladies inflammatoires auto-immunes. Cela concerne entre autres les maladies inflammatoires chroniques de l'intestin incluant la maladie de Crohn et la rectocolite hémorragique, le psoriasis, la polyarthrite rhumatoïde, la thyroïdite d'Hashimoto, ou encore la sclérose en plaques (49,50). Il protège également la séquence de ces protéines, tout fragment et toute séquence dérivée ou homologue, incluant notamment les 3 épitopes.

L'une des applications de la P28GST porte sur son utilisation dans le traitement des Maladies Inflammatoires Chroniques de l'Intestin, sujet sur lequel le laboratoire d'accueil (Unité Inserm U1286 INFINITE) travaille activement.

## **6. Maladies Inflammatoires Chroniques de l'Intestin (MICI)**

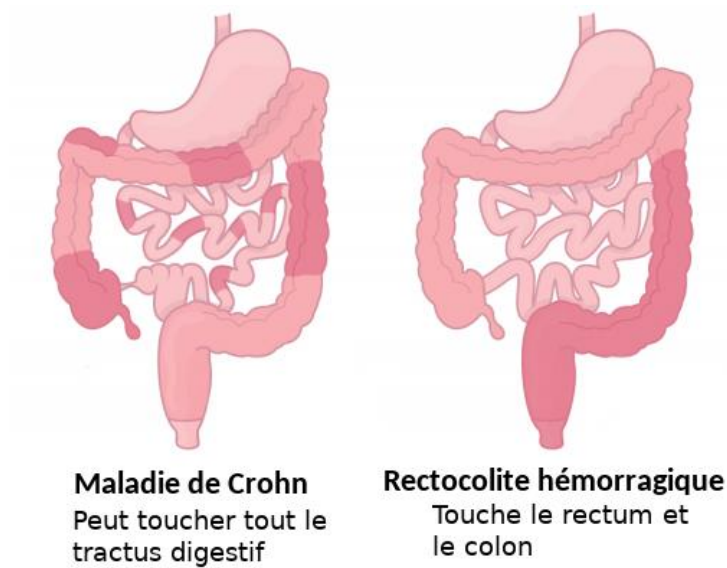
Les maladies inflammatoires chroniques de l'intestin, ou MICI, sont des maladies inflammatoires retrouvées avec une fréquence plus élevée dans la région Nord-Ouest de la France (51,52). Pour étudier ces pathologies, de nombreuses initiatives ont été menées, comme le registre EPIMAD créé au CHU de Lille permettant de recenser systématiquement

tous les nouveaux patients atteints de MICI dans 4 départements (Nord, Pas-de-Calais, Somme et Seine-Maritime) (53,54). Des équipes de recherche locales s'investissent également dans l'étude des MICI, comme la fondation DigestScience, ou encore l'unité INSERM U995 LIRIC (*Lille Inflammation Research International Center*), devenu l'unité INSERM U1286 INFINITE (*Institute for Translational Research in Inflammation*), à l'origine des travaux sur la P28GST. C'est dans ce cadre qu'un financement ANR a été obtenu en 2011 et a permis la réalisation d'une étude clinique de phase 2a, nommée ACROHNEM (*Anti CROHN Enzymatic Molecule*), permettant l'évaluation de la tolérance de la P28GST administrée à des patients atteints de la maladie de Crohn légère (55).

Les MICI seront plus fréquemment retrouvées dans les pays au mode de vie occidentalisé, où 0.3% à 0.5% de la population sont atteints de maladie de Crohn ou de rectocolite hémorragique (56,57). Cela représente plus de 5 millions de malades, dont 1.4 millions aux Etats-Unis et 3 millions en Europe (58). Avec le développement des processus d'industrialisation et l'urbanisation, les MICI deviennent de plus en plus courantes, et se répandent dans le monde entier, notamment dans les pays d'Asie de l'Est (57). Malgré le peu d'informations sur le nombre de patients dans ces pays en voie de développement, l'incidence et la prévalence des MICI augmentent dans toutes les régions du monde, indiquant que ces pathologies sont en train de devenir des fléaux de santé publique à l'échelle mondiale (56).

Les MICI sont un groupe de pathologies caractérisées par un dérèglement de la muqueuse intestinale, dans lequel l'hôte réagit de manière excessive sur le plan immunologique aux substances présentes dans le flux intestinal naturel (59). Elles évoluent dans le temps sous la forme de poussées inflammatoires entrecoupées de périodes de rémissions asymptomatiques. Les MICI sont divisées en deux sous-catégories en fonction de la localisation des lésions du tube digestif et des caractéristiques pathologiques (Figure 7) :

- La maladie de Crohn ;
- La rectocolite hémorragique.



*Figure 7. Schéma des lésions du tractus digestif causées par les MICI - Adapté de (60)*

La maladie de Crohn se caractérise par des lésions diffuses pouvant toucher n'importe quelle partie du tube digestif, de la bouche à l'anus. Toutefois, les lésions sont majoritairement localisées au niveau de l'intestin grêle et du côlon (Figure 7). Les signes cliniques sont dépendants de la localisation de l'inflammation de la muqueuse et de la forme de la maladie. La maladie de Crohn se manifeste le plus souvent par une diarrhée, des crampes et douleurs abdominales, ainsi qu'une atteinte de l'état général avec fièvre, perte de poids et fatigue. Des manifestations extra-digestives peuvent également survenir avec des atteintes ostéo-articulaires, cutanéomuqueuses et oculaires. L'activité de la maladie est mesurée à l'aide de l'indice de Best, également appelé CDAI (Crohn's Disease Activity Index) (61). Ce score permet de catégoriser l'état à partir de 8 observations incluant le bien-être général, le nombre de selles molles ou liquides, la présence de douleurs abdominales, l'observation d'autres manifestations digestives et extra-digestives, l'hématocrite, ou encore le poids. En fonction de cet indice, la maladie de Crohn sera classée comme :

- Non active avec un CDAI inférieur à 150, indiquant une période de rémission,
- Légèrement active avec un CDAI compris entre 150 et 220,
- Modérément active avec un CDAI compris entre 220 et 450,
- Sévèrement active avec un CDAI supérieur à 450.

A ce jour, il n'existe pas de traitement curatif pour les maladies inflammatoires chroniques de l'intestin (59,62,63). Les traitements médicamenteux de référence disponibles ont pour principal objectif l'amélioration de la qualité de vie des patients, en diminuant la fréquence des poussées et en retardant la formation de lésions irréversibles conduisant au recours à la chirurgie. Les objectifs à court terme sont de mettre fin le plus rapidement, complètement et durablement possible aux symptômes de la maladie en poussée. Les objectifs à moyen et long terme sont de maintenir l'absence de symptômes en réduisant la fréquence des poussées, d'obtenir la cicatrisation des lésions des muqueuses et la disparition de toute inflammation, et de limiter l'incidence des manifestations extra-digestives. Actuellement, il existe quatre catégories de traitements utilisés dans la prise en charge des patients : les dérivés de l'acide aminosalicylique, les corticoïdes, les immunosuppresseurs et les biothérapies. En cas d'échec des traitements médicamenteux, le recours à la chirurgie peut être obligatoire.

La P28GST permettrait de compléter le schéma de prise en charge classique en proposant une nouvelle alternative thérapeutique pour retarder un maximum le recours aux traitements les plus agressifs et à la chirurgie.

## **7. Essai clinique de phase 2a - ACROHNEM**

Pour s'assurer de la sécurité d'emploi de la P28GST dans le contrôle de l'inflammation dans la maladie de Crohn inactive à légère, avant ou après chirurgie de résection intestinale, une étude clinique exploratoire de phase 2a a été menée (64,65). Cette étude a pu être réalisée sur la base des connaissances des précédents essais cliniques BILHVAX, démontrant la tolérance et l'absence d'effets indésirables chez des sujets sains.

L'étude ACROHNEM a inclus un total de 10 patients présentant une valeur de base du CDAI inférieure à 220, dont 7 ont pu être inclus dans les résultats de l'étude. Le critère d'évaluation primaire a été le taux et la gravité des événements indésirables recensés au cours de l'étude, et les critères d'évaluations secondaires ont été l'apparition ou non d'une récurrence clinique évaluée par le CDAI, la mesure des principaux marqueurs immunologiques et inflammatoires sanguins et tissulaires, et l'évolution des espèces bactériennes du microbiote intestinal.

Cette étude de phase 2a a démontré la sécurité de l'utilisation de la P28GST chez des patients atteints de formes non-actives ou légères de la maladie de Crohn, sans la présence d'effets

indésirables graves. La P28GST présente également un intérêt thérapeutique par la diminution significative des scores CDAI au bout de 3 mois de traitement (Figure 8). Enfin, aucun marqueur sanguin ou tissulaire n'a été augmenté par la prise du traitement.

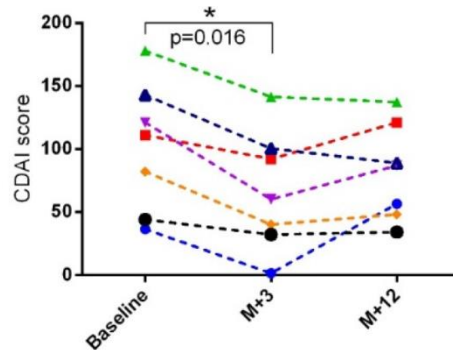


Figure 8. Évolution du CDAI au cours de l'essai clinique ACROHNEM (64)

Les suites à donner à ces résultats prometteurs sont des travaux qui concernent principalement l'optimisation du schéma thérapeutique, l'élucidation des mécanismes d'action mis en jeu, et la réalisation d'un essai clinique randomisé de phase 2b permettant de valider les premiers résultats observés.

## **8. Valorisation industrielle – Création de Par'Immune**

Le potentiel immuno-régulateur de la P28GST dans les maladies inflammatoires auto-immunes, aidé par les résultats encourageants de l'essai clinique ACROHNEM sur la maladie de Crohn, ont abouti à la création d'une startup dédiée au développement de la P28GST comme candidat médicament dans les maladies inflammatoires auto-immunes : Par'Immune. Fondée en Octobre 2017 par le Pr. Monique CAPRON et le Dr. Abderrahim LACHGAR, l'objectif principal de l'entreprise est le développement de l'utilisation de la P28GST dans les maladies inflammatoires.

Dans la perspective de la réalisation d'études de phase 2b dans la maladie de Crohn et de nouvelles études de phase 2 dans d'autres maladies inflammatoires, le développement des connaissances mécanistiques est une étape primordiale. Les précédentes études de la littérature ont été principalement focalisées sur les aspects fonctionnels et cliniques de la P28GST. Cependant, les dernières études mécanistiques ont été réalisées au début des années

2000 par des moyens techniques limitant les possibilités d'investigation, permettant de comprendre plus en détail les mécanismes enzymatiques.

Les objectifs de Par'Immune visent à une meilleure compréhension du mécanisme d'action de la P28GST et son développement comme candidat médicament. Pour cela, notre axe de recherche porte sur la compréhension des mécanismes d'action enzymatiques au niveau atomique par des approches *in silico*. Ces approches, utilisant des simulations par ordinateur, ont pour but de valider, compléter et développer les hypothèses formulées au début des années 2000 et de proposer de nouvelles hypothèses pour créer des liens entre les observations fonctionnelles et le mécanisme d'action au niveau atomique.

## II. Glutathion S-Transférases

Les Glutathion S-Transférases (GST) sont une très large famille d'enzymes jouant un rôle majeur dans le métabolisme de phase 2, permettant la détoxification de composés électrophiles, endogènes ou xénobiotiques, via l'utilisation de glutathion (66) (Figure 9).

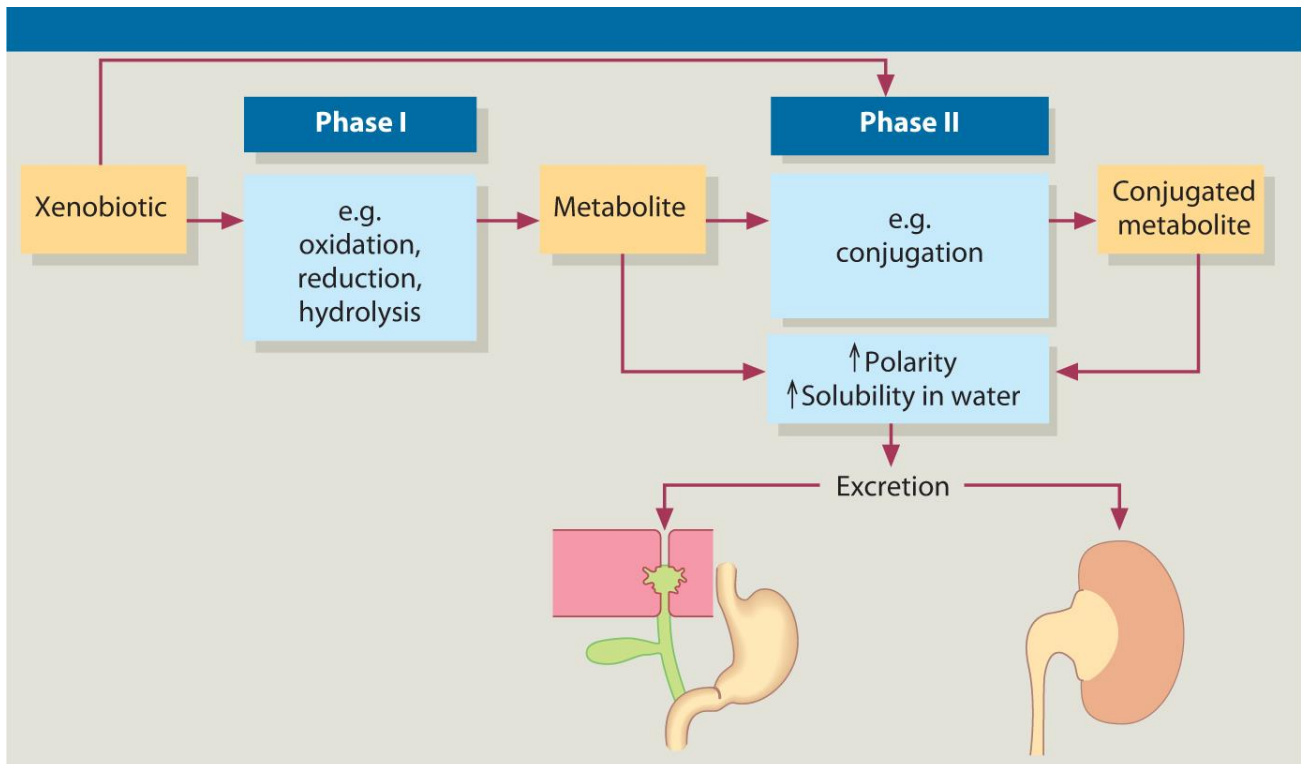


Figure 9. Biotransformation des xénobiotiques chez l'Homme (67)

### 1. Classification des GST

Les glutathion S-transférases peuvent être subdivisées en trois superfamilles en fonction de leur localisation cellulaire :

- Les GST cytosoliques, retrouvées dans le cytoplasme,
- Les GST mitochondriales, retrouvées dans les mitochondries,
- Les GST microsomales, également appelées MAPEG (*Membrane-Associated Proteins in Eicosanoid and Glutathione metabolism*), retrouvées dans le réticulum endoplasmique.

Au sein de chaque superfamille existe une subdivision en classes regroupant les isoenzymes, c'est-à-dire les GST catalysant les mêmes réactions chimiques (Tableau 1).

Classe	Organismes	Organisation (68)	Résidu catalytique (68)	Activités enzymatiques (69)
Alpha	Mammifères (70), helminthes (71)	Dimère	Tyrosine	Conjugaison du glutathion, peroxydase
Beta	Bactéries (72)	Dimère	Cystéine	Conjugaison du glutathion, catabolisme des composés organiques
Delta	Insectes (73)	Dimère	Sérine	Conjugaison du glutathion, peroxydase, résistance aux toxiques défensifs
Epsilon	Insectes (73)	Dimère	Sérine	Conjugaison du glutathion, résistance aux toxiques défensifs
Zeta	Mammifères (70), insectes (73), plantes (74)	Dimère	Sérine	Conjugaison du glutathion, métabolisme des haloacides, catabolisme de la tyrosine
Theta	Mammifères (70), insectes (73), plantes (74)	Dimère	Sérine	Conjugaison du glutathion, peroxydase, détoxification des hydroperoxydes organiques
Lambda	Plantes (74)	Monomère	Cystéine	Conjugaison du glutathion, thiol transférase
Mu	Mammifères (70), helminthes (71)	Dimère	Tyrosine	Conjugaison du glutathion, peroxydase
Pi	Mammifères (70), helminthes (71)	Dimère	Tyrosine	Conjugaison du glutathion, peroxydase
Sigma	Mammifères (70), helminthes (71), insectes (73)	Dimère	Tyrosine	Conjugaison du glutathion, synthèse de prostaglandines, dont prostaglandine D2 synthase dépendant du glutathion
Tau	Plantes (74)	Dimère	Sérine	Conjugaison du glutathion, peroxydase, catabolisme de la tyrosine
Phi	Plantes (74)	Dimère	Sérine	Conjugaison du glutathion, peroxydase, catabolisme de la tyrosine
Chi	Bactéries (72)	Dimère	?	Conjugaison du glutathion
Omega	Mammifères (70), helminthes (71), insectes (73)	Dimère	Cystéine	Conjugaison du glutathion

**Tableau 1 : Classification non exhaustive des glutathion S-transférases cytosoliques (68–74)**

Les glutathion S-transférases cytosoliques ont d'abord été classées en fonction de leur sélectivité de substrat et de leur immunité croisée (69), puis sont maintenant classées en



fonction du critère d'identité de séquence. Deux GST sont rangées dans la même classe si elles partagent plus de 40% à 60% d'identité de séquence, et n'appartiennent pas à la même classe si elles partagent une identité de séquence inférieure à 25% à 30%, limites variant en fonction des auteurs (75,76). D'autres critères peuvent entrer en jeu pour différencier les cas plus atypiques de GST d'espèces non mammifères, comme les spécificités de leur site actif ou de leur structure quaternaire (75).

Les différentes classes sont désignées par le nom de lettres de l'alphabet grec, et seront abrégées en majuscules de l'alphabet latin pour être compatibles avec les différents outils bioinformatiques (77). Pour distinguer les GST de différentes espèces, un préfixe peut être ajouté en lettres minuscules, convention qui est standard pour toutes les protéines.

Les 14 classes de GST cytosoliques les plus communes ont été répertoriées dans le tableau X. Les GST cytosoliques de mammifères sont principalement retrouvées en homodimères ou en hétérodimères au sein de 7 classes : alpha, zeta, theta, mu, pi, sigma et omega. Certaines classes de GST cytosoliques sont exclusives à certaines espèces et organismes, comme les classes beta et chi chez certaines bactéries, les classes delta et epsilon pour certains insectes, ou encore les classes tau et phi pour certaines plantes.

Il est également à noter que les GST mitochondriales suivent également cette classification, et sont regroupées dans la classe kappa.

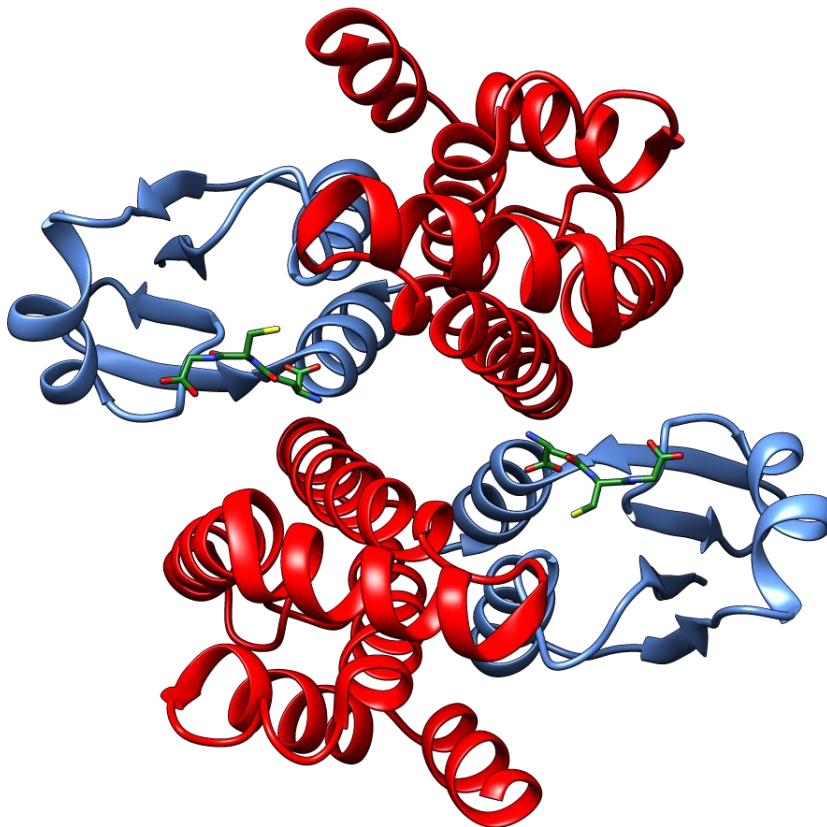
Parmi les GST cytosoliques, il existe également une subdivision en fonction de la nature du résidu catalytique permettant l'activité enzymatique (68,69,78). On peut ainsi distinguer :

- Les S/C-GST, qui utilisent une sérine ou une cystéine. Ces GST sont largement distribuées au sein de nombreuses classes et d'organismes différents.
- Les Y-GST, qui utilisent une tyrosine pour activer le glutathion. On y retrouve principalement les GST cytosoliques alpha, mu, pi et sigma.

## 2. Structure des GST cytosoliques

### *a. Domaines structuraux*

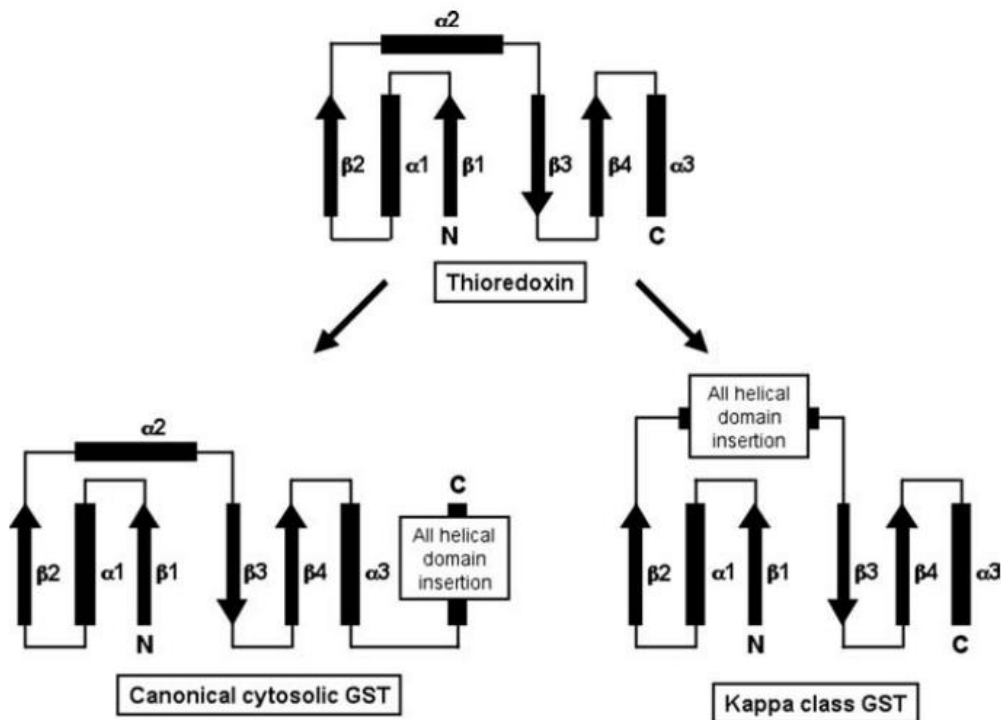
Les GST cytosoliques très majoritairement structurées sous la forme de dimères, comprennent de 199 à 244 acides aminés par monomère (79). Tous les monomères des GST cytosoliques partagent une structure tertiaire commune, composée de deux domaines distincts : le domaine N-terminal, et le domaine C-terminal (Figure 10).



**Figure 10.** Représentation tridimensionnelle d'une GST sous la forme de dimère, avec le domaine N-terminal en bleu, le domaine C-terminal en rouge, et le glutathion représenté en vert (PDB ID : 1OE8) (80)

Le domaine N-terminal est un domaine fortement conservé dans l'évolution, appelé domaine thioredoxine (76,79) (Figure 10 - en bleu). Il est constitué d'un repliement semblable aux protéines de la famille des thioredoxines ayant pour capacité de lier le glutathion, et retrouvés dans d'autres familles enzymatiques comme les glutarédoxines ou les glutathion peroxydases

(81,82). Ce domaine est constitué d'un feuillet  $\beta$  composé de 4 brins et de 3 hélices  $\alpha$ , sous la topologie  $\beta_1\alpha_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$  (Figure 11) (79). La présence d'une proline juste avant le brin  $\beta_3$  permet de maintenir les GST cytosoliques dans des conformations structurellement stables et catalytiquement fonctionnelles, sans avoir un rôle direct dans la catalyse, par la contrainte sur les angles dièdres du squelette peptidique (83). De plus, l'hélice  $\alpha_2$  est la structure secondaire possédant le plus de variations entre les classes et espèces (76). Enfin, les résidus catalytiques sont localisés au niveau de la « boucle catalytique » située entre le brin  $\beta_1$  et l'hélice  $\alpha_1$ , où le résidu catalytique des Y-GST cytosoliques est localisé à la fin du brin  $\beta_1$ , alors que le résidu catalytique des S/C-GST est localisé juste avant le début de l'hélice  $\alpha_1$  (76).



*Figure 11 : Schéma de la structure de la thioredoxine, et des repliements apparentés à la thioredoxine comprenant les GST cytosoliques et GST de la classe Kappa (79)*

Le domaine C-terminal est, à l'opposé du domaine N-terminal, un domaine fortement variable en fonction des classes et espèces, appelé domaine hélicoïdal (Figure 10 - en rouge). Il est constitué uniquement d'hélices  $\alpha$ , avec un nombre variable de 4 à 8 hélices en fonction de la classe. Un cœur d'hélices  $\alpha_4$ ,  $\alpha_5$ ,  $\alpha_6$  et  $\alpha_7$  est retrouvé systématiquement, avec les hélices  $\alpha_4$ ,

$\alpha 5$  et  $\alpha 6$  dans le même axe que les hélices  $\alpha 1$  et  $\alpha 3$ , et avec l'hélice  $\alpha 7$  perpendiculaire aux autres. Les autres hélices sont situées à la suite de l'hélice  $\alpha 7$  (75).

### ***b. Interface dimérique***

Les caractéristiques structurales de l'interface du dimère des GST permettent jusqu'à deux zones d'interaction entre les sous-unités. La première zone d'interaction est basée sur l'axe de symétrie de la protéine, avec des interactions électrostatiques fortement conservées dans l'évolution entre l'hélice  $\alpha 3$  du domaine N-terminal et l'hélice  $\alpha 4$  du domaine C-terminal du monomère opposé (84). Dans certaines classes, notamment les classes Alpha, Mu et Pi, il existe une seconde zone d'interaction sous la forme d'un mécanisme « lock and key » permettant une coopérativité positive des activités enzymatiques entre les deux monomères (75). Ce mécanisme implique une phénylalanine située entre l'hélice  $\alpha 2$  et le brin  $\beta 3$  (*key*) qui va s'insérer au niveau d'une poche hydrophobe composée de 5 acides aminés définie par l'hélice  $\alpha 4$  du premier monomère et l'hélice  $\alpha 5$  de l'autre monomère (*lock*) (75,84).

### ***c. Site actif***

Le repliement structural des glutathion S-transférases permet la formation de plusieurs sites de liaison : un site G et un site H retrouvés dans toutes les classes de GST, et un site L retrouvé moins fréquemment.

Le site G est le site permettant la fixation du glutathion. Il est principalement situé sur le domaine thiorédoxine, où la majorité des résidus impliqués dans la stabilisation du glutathion sont situés au niveau de la succession des deux feuilletts  $\beta 3$  et  $\beta 4$  et de l'hélice  $\alpha 3$  (85), avec la possibilité de stabiliser le glutathion par des résidus situés sur le domaine C-terminal du second monomère, notamment un acide aspartique fortement conservé sur l'hélice  $\alpha 4$  (75,86), avec une forte conservation de tous ces résidus dans l'évolution (86). Le thiol du glutathion sera placé à proximité du résidu catalytique, le groupement aminoacétique vers l'interface dimérique, et le groupement carboxyle vers l'extérieur de l'enzyme (Figure 10 - en vert).

Le site H est le site permettant la fixation de substrats hydrophobes (76). Il est principalement situé au niveau du domaine hélicoïdal, à proximité directe du site G, et est composé d'une grande majorité de résidus hydrophobes (86). A la différence du site G, les résidus formant le site H ne sont pas conservés dans l'évolution, permettant une diversité de structures et donc une hétérogénéité de substrats reconnus par les différentes classes de GST (86).

Enfin, le site L est un site permettant la fixation de ligands, qui n'est pas retrouvé systématiquement dans toutes les classes de GST (86). Ce site possède une localisation variable, retrouvé préférentiellement soit au niveau de l'interface dimérique comme pour les GST de classe Omega (87), soit dans une poche présente dans la continuité des sites G et H pour les GST de classe Sigma (88).

### III. Caractéristiques de la P28GST

#### 1. Place de la P28GST dans la classification des Glutathion S-Transférases

Lors de leur identification et leurs études préliminaires, les 28GST ont plutôt été classées dans les familles alpha et mu de part leur homologie avec d'autres GST de ces familles, ou classées dans les familles pi ou mu en fonction des spécificités de substrats mises en évidence (22,89). Toutefois, après la découverte de nouvelles classes de GST et des Prostaglandine-D Synthases, et sur la base de comparaisons de séquence et de la présence d'une activité PGDS, les 28GST de schistosomes ont été rangées dans la classe sigma (Figure 12) (90,91).

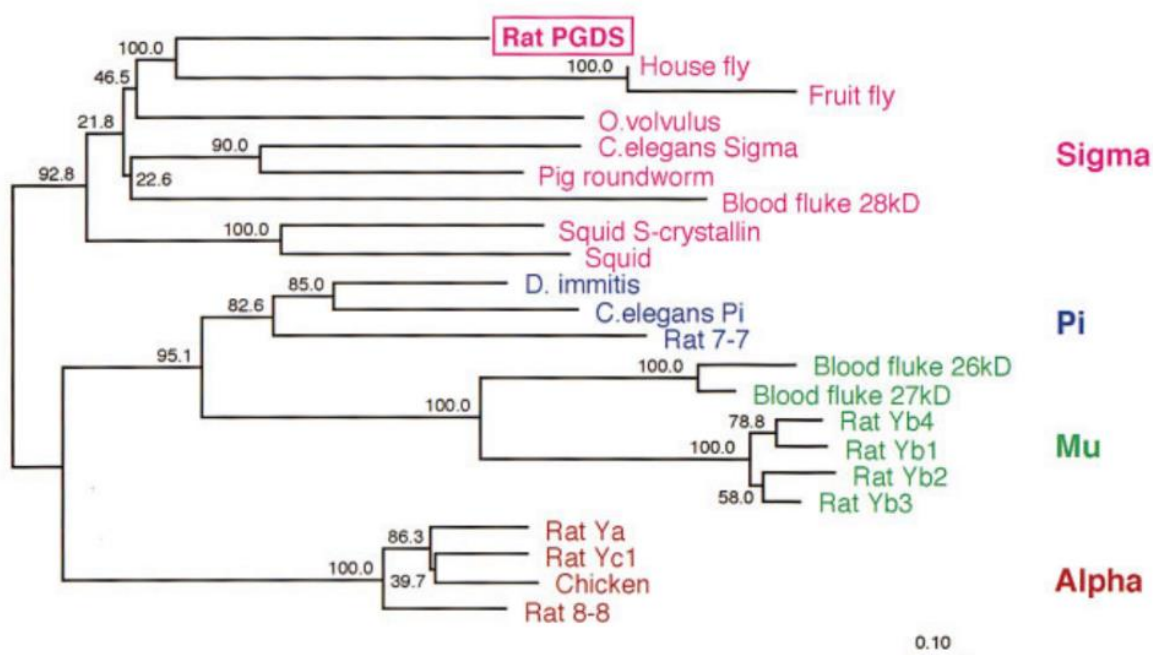


Figure 12. Arbre phylogénique non enraciné de certaines PGDS hématopoïétiques et de certaines GST de classe sigma, pi, mu et alpha. Sm28GST correspond à la catégorie "Blood fluke 28kD" (90).

## **2. Glutathion S-Transférases de classe sigma**

La classe sigma des Glutathion S-Transférases a été proposée pour la première fois chez le calamar en 1988 sur la base d'alignements de séquence et d'arbres phylogéniques (92). Depuis, de très nombreuses GST de la classe sigma ont été identifiées dans des espèces variées, comprenant l'Homme, la souris, le rat, les insectes, les helminthes et les mollusques (93).

Comme les autres classes de GST cytosoliques, les GST de classe sigma présentent un domaine thiorédoxine N-terminal permettant l'activation du glutathion. Ici, le glutathion réalise une interaction primordiale avec une tyrosine catalytique hautement conservée. Une seconde interaction est réalisée par une arginine du site actif, également conservée. Elles possèdent également un domaine hélicoïdal C-terminal des GST qui ressemble à celui des autres classes de GST cytosoliques en participant au site H, liant les ligands hydrophobes pour l'activité GST. Toutefois, des modifications spécifiques à l'interface de ces deux domaines permettent l'obtention d'un site L ayant une affinité pour la prostaglandine H<sub>2</sub>, et permettant la formation de prostaglandine D<sub>2</sub>, aboutissant à la présence d'une activité enzymatique supplémentaire de type Prostaglandine-D<sub>2</sub> Synthase (PGDS) (88).

Cette famille d'enzyme est structurée sous la forme d'homodimères, formés d'en moyenne 199 acides aminés par monomère chez les mammifères, et de 205 à 233 acides aminés par monomère chez les invertébrés (93). Toutefois, l'interface dimérique sera organisée différemment dans la classe en fonction des espèces : les GST de classe sigma des invertébrés ne possèdent pas l'interaction « lock and key » aidant à la stabilisation du dimère (80,94), alors qu'elle est retrouvée chez les GST de vertébrés comme l'Homme ou le rat (90).

Cette classe sigma est également unique car elle est activée par la présence de cations divalents au niveau de l'interface dimérique, notamment un ion magnésium Mg<sup>2+</sup> ou un ion calcium Ca<sup>2+</sup> chez l'Homme (95). La présence d'un cation divalent va augmenter l'activité enzymatique PGDS d'environ 150% par rapport à son niveau basal (Figure 13). De plus, l'augmentation de l'activité favorisée par Mg<sup>2+</sup> est associée à une diminution de 4 fois le Km pour le GSH (Figure 13 - à gauche), non observé pour le Ca<sup>2+</sup>.

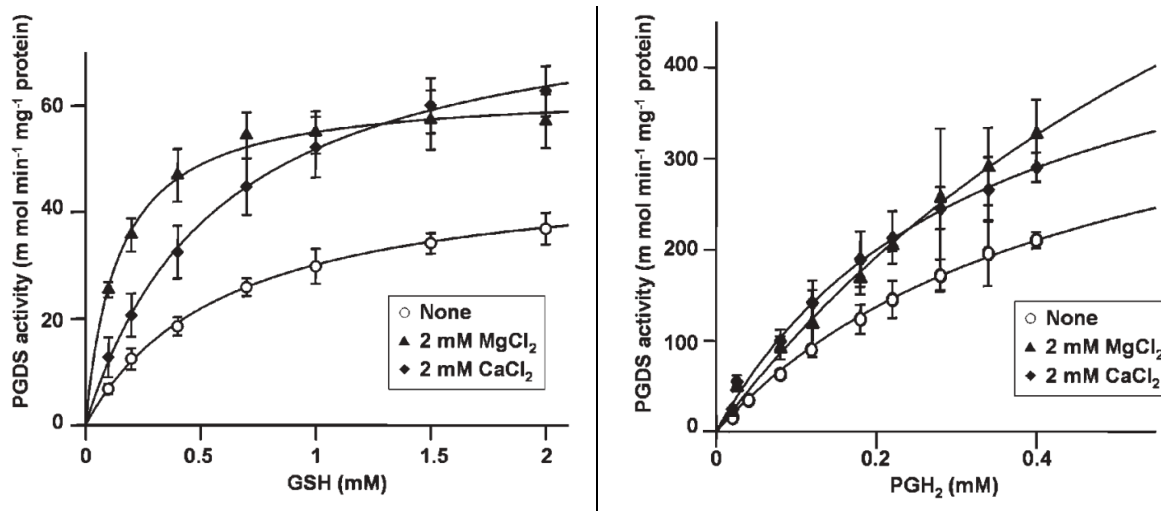


Figure 13. Effet de la présence d'un ion métallique sur l'activité PGDS de la GST de classe sigma humaine, HPGDS (95).

Ces cations sont stabilisés par la coordination de 6 molécules d'eau pour Mg<sup>2+</sup> (Figure 14 - à droite) et 5 molécules d'eau pour Ca<sup>2+</sup> (Figure 14 - à gauche), elles mêmes stabilisées par des acides aminés acides des hélices  $\alpha$ 3 et  $\alpha$ 4. Certains de ces acides participent également à la stabilisation d'autres acides aminés impliqués dans l'activité PGDS, comme l'arginine retrouvée entre les sites G et L (Figure 14 - Arg14).

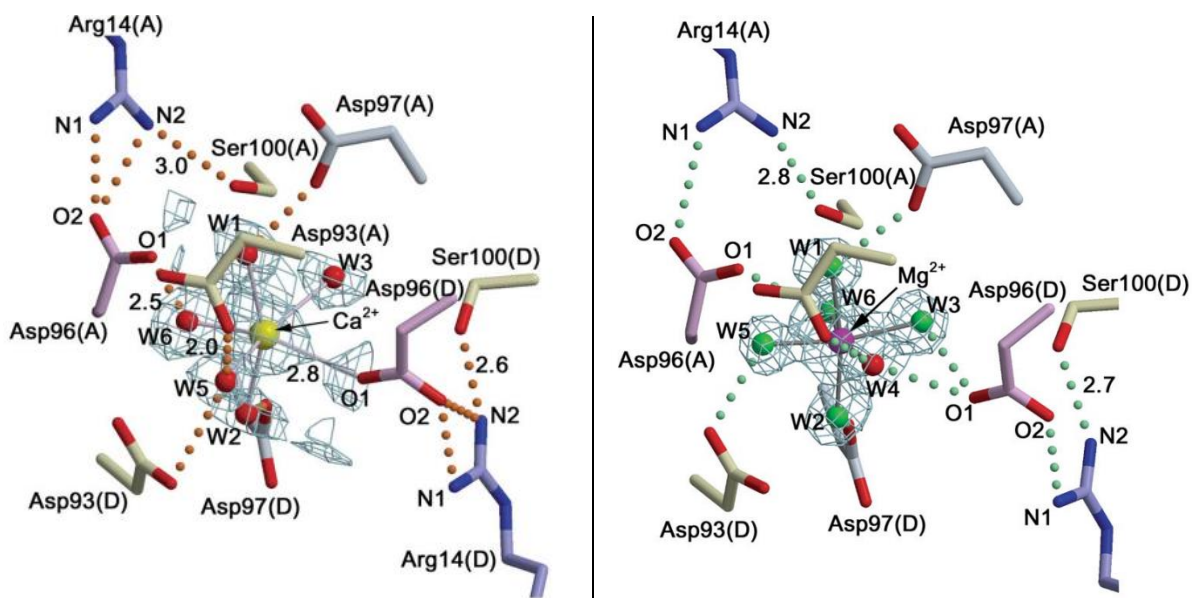


Figure 14. Représentations tridimensionnelles de la coordination métallique chez la GST de classe sigma humaine, HPGDS (95). La forme liée à un ion calcium Ca<sup>2+</sup> et stabilisé par 5 molécules d'eau est représentée à gauche, et la forme liée à un ion magnésium Mg<sup>2+</sup> et stabilisé par 6 molécules d'eau est représentée à droite. L'arginine participant à l'activité PGDS est Arg14.

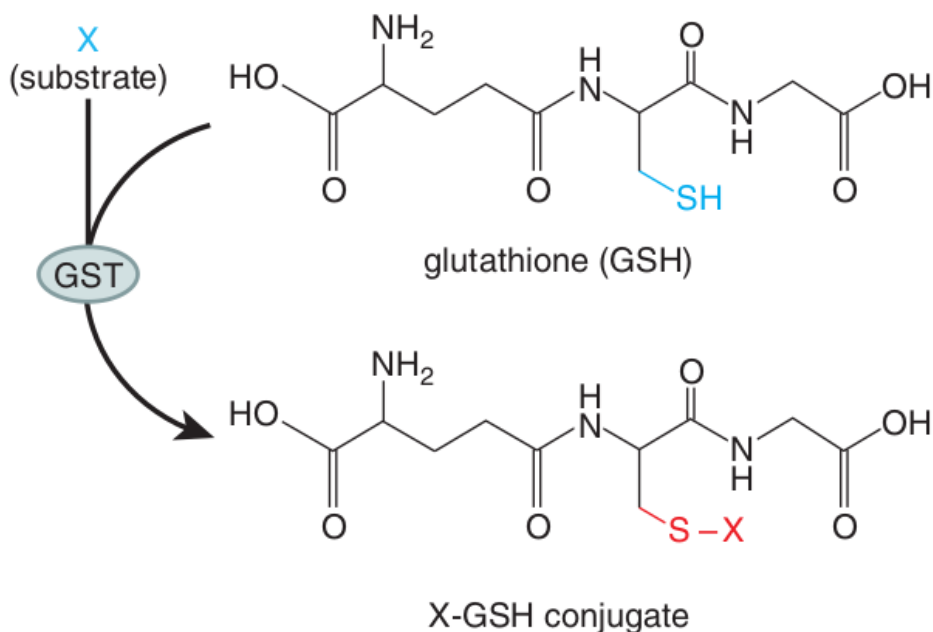




### a. *Activité Glutathion S-Transférase :*

L'activité Glutathion S-Transférase est l'activité non spécifique de toutes les GST. Cette activité catalyse le transfert du glutathion aux électrophiles réactifs permettant de protéger les macromolécules et l'environnement cellulaire (79). De manière caractéristique, tous les substrats des GST contiennent un atome électrophile (comme l'oxygène, l'azote ou le soufre) et sont hydrophobes (96). L'activité GST permet donc la défense antioxydante par neutralisation des espèces réactives de l'oxygène, et la détoxification par liaison covalente au glutathion de métabolites toxiques endogènes ou de xénobiotiques lipophiles exogènes, permettant leur élimination (79).

Le glutathion sous sa forme réduite agit dans cette activité comme cosubstrat, et est consommé lors de la réaction pour obtenir une forme conjuguée au substrat électrophile (Figure 15) (96). Lors de la formation des conjugués au glutathion, la réaction de la GST génère une liaison thioéther avec le substrat à neutraliser. Cette conjugaison passe par l'action d'un anion thiolate du glutathion obtenu par l'action de l'acide aminé catalytique, ici une tyrosine, sur la fonction thiol.

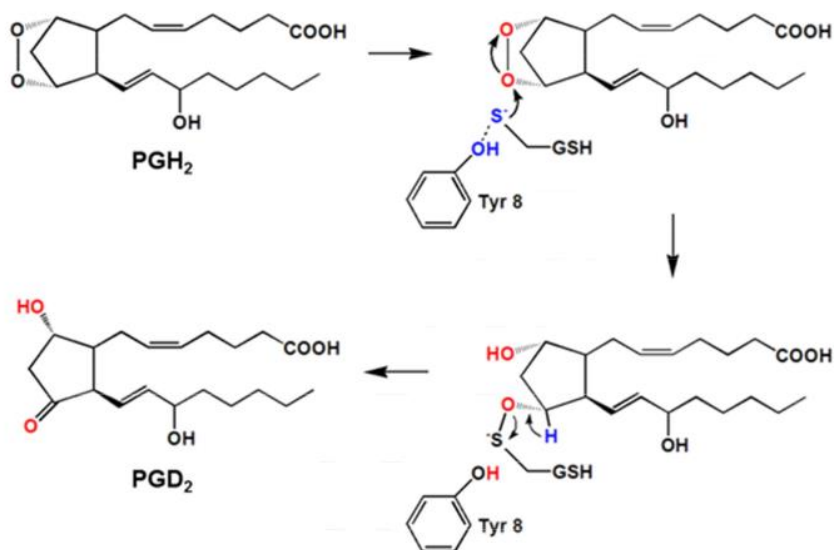


**Figure 15. Schéma mécanistique de l'activité Glutathion S-Transférase (96). X : Substrat ; GSH : glutathion sous sa forme réduite ; X-GSH : glutathion sous sa forme conjuguée au substrat.**

### *b. Activité Prostaglandine-D2 Synthase*

L'activité Prostaglandine-D2 Synthase est une activité retrouvée spécifiquement chez les GST de classe sigma permettant la conversion de la prostaglandine H2 en prostaglandine D2 (97). La prostaglandine D2 produite a ensuite un rôle dans la chimiotaxie des neutrophiles, peut se lier à ses récepteurs 1 et 2 de la prostaglandine D2 (DP1-R et DP2-R) pour déclencher des activités respectivement anti-inflammatoire et pro-inflammatoire, ou est l'intermédiaire d'autres prostaglandines anti-inflammatoires (93).

Le glutathion sous sa forme réduite agit dans cette activité comme cofacteur, permettant la transformation de la prostaglandine H2 sans être consommé (98). De manière similaire à l'activité GST, l'activité PGDS est initiée par l'anion thiolate du glutathion obtenu par l'action de la tyrosine catalytique. Cet anion thiolate va ensuite attaquer la liaison peroxyde pour former une fonction méthoxysulfanyle S-O liant de manière covalente la prostaglandine H2 et le glutathion. Enfin, au vu de la labilité de la fonction méthoxysulfanyle, cette liaison se rompt pour former une fonction carbonyle, permettant l'obtention de la prostaglandine D2 (88,95,97). (Figure 16)



*Figure 16. Schéma mécanistique de l'activité Prostaglandine-D2 Synthase retrouvée chez les GST de classe sigma (97).*

### *c. Spécificités entre espèces*

Au sein de cette classe sigma, les enzymes de différentes espèces présentent des activités spécifiques différentes (93). Les 28GST de schistosomes, représentées ici par *Sm28GST*, présentent une activité GST 12 à 40 fois supérieure, mesurée par l'activité spécifique du CDNB, ligand de référence de l'activité GST, mais présentent une activité PGDS 30 fois inférieure aux autres enzymes de mammifères (Tableau 3).

Espèce	Activité spécifique ( $\mu\text{mol}/\text{min}/\text{mg}$ de protéine)	
	CDNB	Prostaglandine H2
Schistosome - <i>Sm28GST</i>	120	1
Onchocerca - <i>OvGST1</i>	2.52	
Ascaris - <i>AsGST1</i>	38.2	
Taenia - <i>TsMσGST</i>	1.08	
Humain - <i>HPGDS</i>	5.06	30
Rat - <i>HPGDS</i>	9.2	32.5
Souris - <i>HPGDS</i>	3.0	28

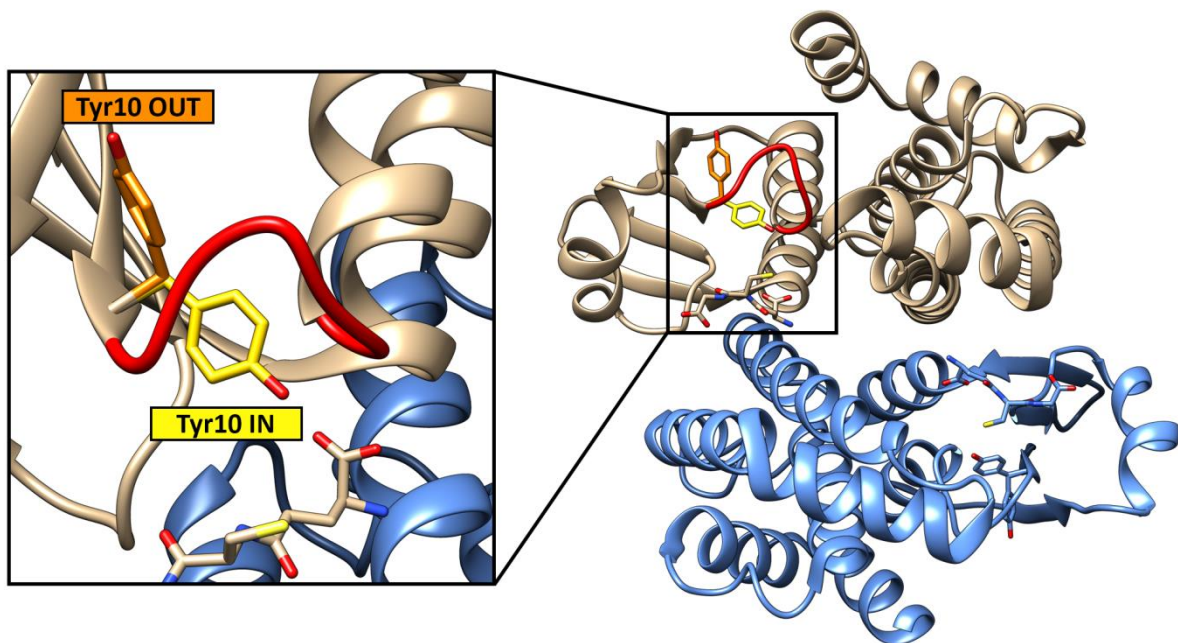
**Tableau 3. Activités spécifiques pour certaines GST de classe sigma de mammifères et de nématodes (22,93,99)**

## 4. Spécificités structurales de la P28GST

La P28GST faisant partie des GST de classe sigma, une majorité des caractéristiques de cette classe sont retrouvées. Toutefois, suite à l'obtention de premières structures cristallographiques, des spécificités structurales ont pu être mises en évidence.

Une caractéristique unique, absente de toutes les structures précédentes des GST de classe sigma, a été mise en évidence : la tyrosine catalytique, en position 10, possède deux positions alternatives (80,100). Cette tyrosine est retrouvée soit :

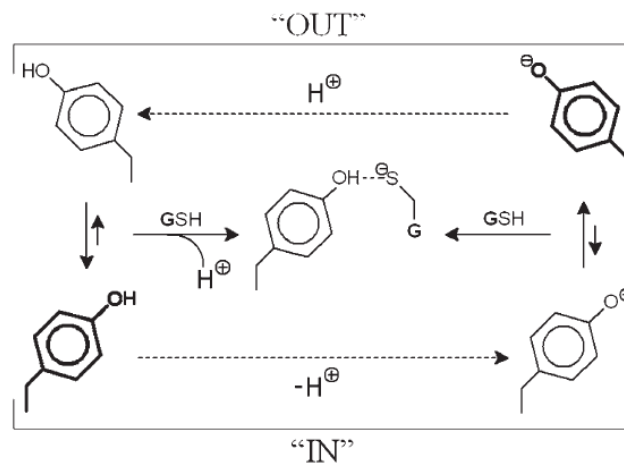
- Dans une conformation où le phénol de la tyrosine pointe en direction de la fonction thiol du glutathion, et appelée conformation "IN" (Figure 17 - en jaune);
- Dans une conformation où le phénol de la tyrosine pointe dans la direction opposée, vers l'extérieur de la protéine, et appelée conformation "OUT" (Figure 17 - en orange).



*Figure 17. Schéma des différentes conformations de la tyrosine 10 catalytique (PDB ID : 1OE8) (80).*

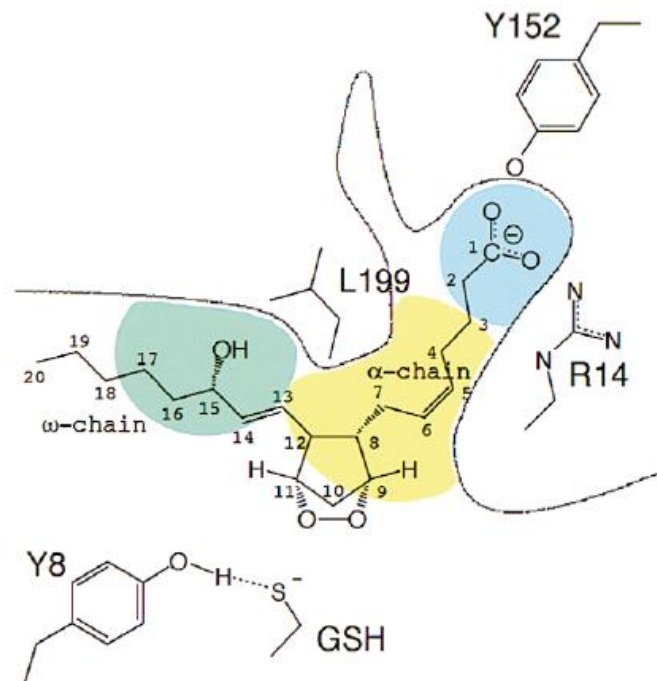
Cette tyrosine est également décrite comme étant responsable du passage de la fonction thiol à l'anion thiolate du glutathion pour l'activation des différentes réactions enzymatiques (101).

Son rôle essentiel dans les activités enzymatiques a été validé par mutagenèse dirigée, où la mutation de la tyrosine 10 en phénylalanine abolit complètement les activités GST et PGDS (99,100). De plus, l'arginine 21, participant à la stabilisation de la conformation OUT par une interaction  $\pi$ -cation, permettrait de diminuer le pKa du phénol pour obtenir un phénolate. La tyrosine déprotonée permettrait ainsi, après son retour dans sa conformation IN, l'activation du glutathion (Figure 18) (101). Ces hypothèses ont également été testées par mutagenèse dirigée : il a été démontré que l'arginine 21 ne possède pas de rôle de stabilisation de la conformation OUT par l'observation des deux conformations de la tyrosine 10, mais qu'il possède bien un rôle essentiel dans la diminution du pKa de celle-ci (100). Enfin, il a pu être mis en évidence une corrélation entre les mouvements de la tyrosine 10 et d'une arginine en position 35 (100).



**Figure 18. Mécanisme d'action supposé de la tyrosine 10 catalytique (101).**

De plus, des différences significatives existent au niveau du site de liaison de la prostaglandine H2. En effet, il est proposé dans la littérature une description du site L permettant de positionner au sein de HPGDS la prostaglandine H2, avec ses deux chaînes dans des fentes à proximité du site G et du glutathion (Figure 19) (88,90). Toutefois, chez la P28GST, les chaînes latérales de l'acide glutamique 106, de la tyrosine 110 et de l'histidine 169 interféreraient avec ce mode de liaison proposé (80). L'histidine 169 réaliserait un encombrement stérique au niveau de site, et serait maintenu en place par des liaisons avec l'arginine 14 et l'acide glutamique 106 (80).

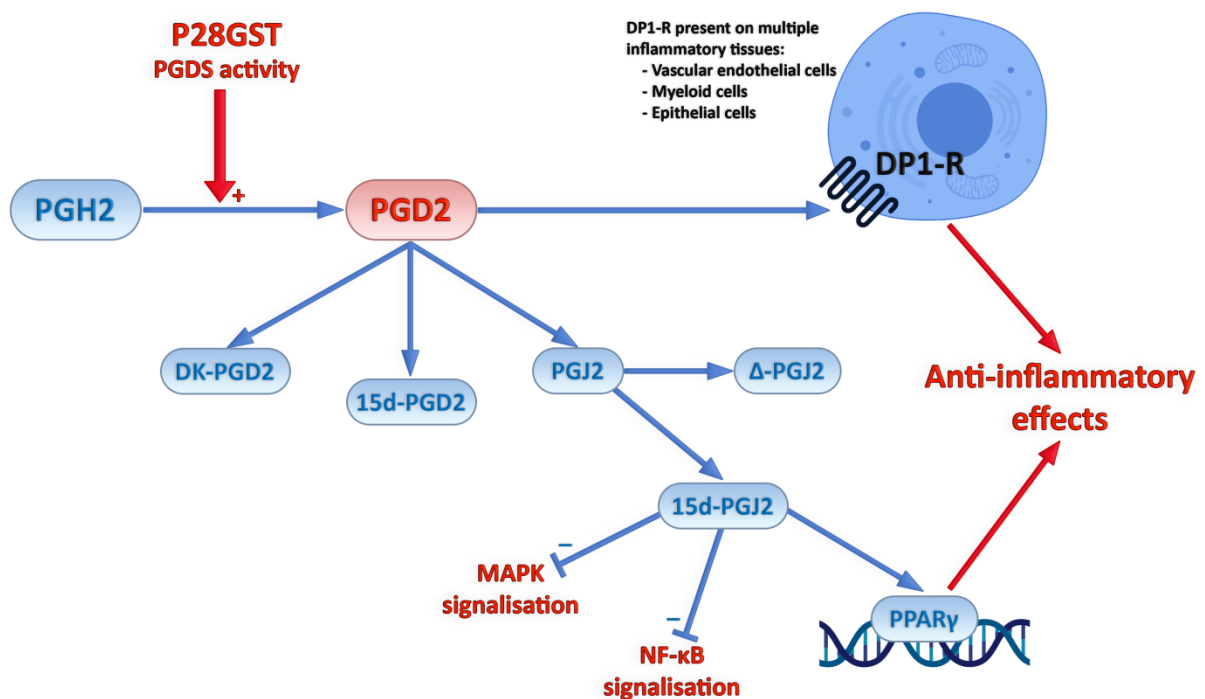


*Figure 19. Mode de liaison supposé de la prostaglandine H2 dans le site L de HPGDS (90).*

Enfin, il est à noter qu'aucun cation n'a été retrouvé dans les cristallographies réalisées, contrairement à toutes les autres GST de classe sigma (80,100).

## **5. Mode d'action anti-inflammatoire supposé**

Sur la base des différentes observations fonctionnelles et cliniques, et des premières hypothèses sur les activités enzymatiques, le mode d'action anti-inflammatoire supposé serait une dualité d'action entre les propriétés épitopiques et l'activité Prostaglandine-D2 Synthase. Cette activité PGDS serait nécessaire lors de la primo-injection de la P28GST pour permettre une cascade de réponses, comprenant une activation directe du récepteur DP1 par la prostaglandine D2 et une activation indirecte de PPAR- $\gamma$  par la production de molécules dérivées de la prostaglandine D2 (Figure 20).



*Figure 20. Voies biologiques supposé impactées par la présence de la P28GST via la production de prostaglandine D2.*

L'injection sous-cutanée de la P28GST induirait également le recrutement de cellules présentatrices d'antigène, polarisées sous leur forme régulatrice. Ces cellules permettraient la présentation des différents épitopes aux cellules T immatures, entraînant leur polarisation en cellules T régulatrices, qui pourraient migrer sur le site de l'inflammation.

Il semblerait que la présence d'épitopes et que les activités enzymatiques sont nécessaires aux propriétés immuno-régulatrices de la P28GST. La place relative des deux activités enzymatiques n'est pas connue.





# **OBJECTIFS**



En se basant sur les connaissances acquises, nous savons que la protéine P28GST combine une activité Glutathion S-Transférase, entraînant des propriétés anti-oxydantes et de détoxification, et une activité Prostaglandine-D2 Synthase, dont la fixation sur un récepteur spécifique, le DP1-R, induit une cascade anti-inflammatoire. Toutefois, malgré les potentialités thérapeutiques de la P28GST dans de nombreuses maladies inflammatoires, le mode d'action de cette protéine n'est toujours pas totalement caractérisé, notamment par le manque de connaissances fines sur les mécanismes moléculaires mis en jeu dans les deux activités enzymatiques.

**L'objectif principal** de ce travail de thèse a donc été de comprendre le rôle respectif des deux activités enzymatiques dans le mode d'action de la P28GST par des approches *in silico* modernes afin de créer un lien entre les hypothèses mécanistiques formulées au cours des 30 dernières années et la structure de la protéine. **Un second objectif** est d'exploiter ces connaissances approfondies pour la recherche et l'identification *in silico* de composés utilisables comme sondes pharmacologiques, afin de proposer des outils pour évaluer la contribution de chaque activité enzymatique dans l'effet anti-inflammatoire global en bloquant sélectivement l'une ou l'autre de ces activités.

Nous verrons dans une première partie l'étude à l'échelle moléculaire et atomique des mouvements des acides aminés du site actif et de leur participation différentielle entre les deux activités par des simulations de dynamique moléculaire classique et de métadynamique. Dans un second temps, nous verrons le développement d'un outil d'analyse en Python, validé sur des cibles biologiques fortement étudiées dans la littérature, ayant permis de mettre en évidence les réseaux d'interactions moléculaires structurant la P28GST au niveau tertiaire et quaternaire, et d'identifier les principales différences réalisées entre les différentes conformations causées par la position de la tyrosine catalytique. Enfin, nous identifierons le site L permettant la liaison de la prostaglandine H2 lors de l'activité Prostaglandine-D2 Synthase, puis nous verrons les processus mis en jeu pour la recherche d'inhibiteurs spécifiques de l'activité PGDS par criblage virtuel à haut débit et par amarrage moléculaire.

Les travaux les plus récents seront présentés dans le chapitre "discussion et perspectives".



**CHAPITRE I**

**ETUDE STRUCTURALE DE LA P28GST PAR**

**DYNAMIQUE MOLECULAIRE**



# **I. Travaux préliminaires**

## **1. Ressources et logiciels principaux**

La ressource de calcul principale mise à disposition pour les travaux de thèse est un ordinateur de bureau tournant sous Ubuntu 18.04 LTS possédant deux processeurs Intel Xeon Gold 5118 pour un total de 48 cœurs logiques, deux cartes graphiques Nvidia Quadro P2000 et 126 Go de mémoire vive.

Les principaux logiciels utilisés pour la réalisation des dynamiques moléculaires sont Gromacs 2018.8 (102), AMBER18 (103) et PLUMED 2 (104), et pour la réalisation des criblages virtuels : CCDC GOLD (105). Les différents scripts et analyses permettant de faire fonctionner ces logiciels et d'analyser leurs résultats ont été codés en R via le logiciel RStudio (106), en Python3, ou occasionnellement en Bash. La visualisation des structures cristallographiques et trajectoires de dynamique moléculaire a été réalisée par UCSF Chimera 1.13 à 1.15 (107) ainsi que VMD 1.9.3 (108). Enfin, d'autres logiciels ont été utilisés de manière ponctuelle, comme KNIME (109), le visualiseur gratuit BIOVIA Discovery Studio Visualizer (110), BIOVIA draw (111) ou encore ACD/ChemSketch (112).

Afin de permettre la reproductibilité des différents environnements de calculs lors des travaux de thèse, des conteneurs virtuels ont été créés à l'aide du logiciel Docker (113). Les conteneurs regroupent différents logiciels utiles au sein de versions d'Ubuntu 18.04 spécifiquement modifiées pour y inclure un serveur X11 et un serveur VNC. Ces serveurs permettent l'affichage graphique du système d'exploitation en utilisant un navigateur web, un client VNC ou un client X11. Les principaux conteneurs utilisés au cours de la thèse regroupaient Gromacs, Amber18, ou encore l'implantation de PLUMED à AMBER18.

Le principal avantage de ces conteneurs a été de pouvoir transférer l'environnement de calcul sur d'autres machines et sur des serveurs de calcul à distance. Cela a donc permis l'utilisation intensive du méso-centre de Calcul Scientifique Intensif de l'Université de Lille pour la réalisation des dynamiques moléculaires classiques et métadynamiques sur au maximum 4 cartes graphiques Nvidia K80, puis en fin de thèse sur au maximum 2 cartes graphiques Nvidia RTX2080 en supplément, en convertissant les conteneurs Docker vers Singularity (114).



## **2. Préparation des structures cristallographiques**

<b>Identifiant PDB</b>	<b>Espèce</b>	<b>Résolution (Å)</b>	<b>Notes</b>
1OE8	WT	1.65	Saturé en glutathion
1OE7	WT	1.8	
2C80	WT	2.3	En complexe avec S-Hexylglutathion
2C8U	R21Q	2	
2CA8	WT	2.49	
2CAQ	R21L	2	

**Tableau 4. Liste des structures cristallographiques de Sh28GST disponibles (80,100)**

Le choix de la structure cristallographique de P28GST s'est porté sur celle possédant l'identifiant PDB 1OE8 (Tableau 4). En effet, celle-ci présente la meilleure résolution du type sauvage de la protéine, et elle montre moins d'erreurs que sa structure sœur 1OE7 d'après les métriques de validation globale wwPDB (115).

Afin de préparer cette structure cristallographique pour les différentes méthodes utilisées, celle-ci a été modifiée au préalable en utilisant UCSF Chimera. Tout d'abord, la chaîne latérale de l'acide aminé Phe38 a été reconstruite du fait de son absence dans la cristallographie. Cette structure possédant des positions alternatives pour les résidus Tyr10, Phe11, Arg16, Gln39, Met78, Leu98 et Val166, seules les positions alternatives A ont été conservées pour le monomère A, et les positions alternatives B pour le monomère B. Cela permet de comparer efficacement, au sein d'une seule simulation, la condition de la position de la tyrosine catalytique, entre sa conformation "IN" au sein du monomère A, et sa conformation "OUT" au sein du monomère B. De plus, du fait de l'absence des 4 derniers acides aminés C-terminaux de la protéine à cause de leur grande flexibilité, ceux-ci ont également été rajoutés par l'outil de construction peptidique de UCSF Chimera.

Dans la très grande majorité des simulations *in silico* réalisées au cours de la thèse, la structure de la P28GST a été étudiée sous sa forme dimérique. Toutefois, lors des premiers tests de dynamique moléculaire, la structure monomérique a été étudiée en tant que contrôle négatif.

Enfin, des structures de hPGDS ont également été utilisées lors de la thèse comme point de comparaison avec la P28GST. Parmi les 20 structures cristallographiques humaines possédant l'identifiant macromolécule "Hematopoietic prostaglandin D synthase", la structure possédant l'identifiant 1IYH réalisée en 2003 avec une résolution de Å a été étudiée (95).

## II. Dynamique moléculaire

### 1. Introduction

Les protéines sont définies par leur structuration dans l'espace et le temps : ce sont des entités dynamiques, où leurs fonctions sont liées de près à leurs mouvements. Leur flexibilité permet de s'adapter à leurs ligands de manière dynamique, afin d'obtenir l'interaction la plus optimale possible pour réaliser leur activité, et pour permettre la libération des produits finaux ainsi que leur régénération.

La méthode de la dynamique moléculaire permet de rendre compte de cette flexibilité et de leurs mouvements en calculant précisément les interactions liées et non-liées entre chaque atome du système (116–119). Elle permet donc de retracer l'évolution d'un système à l'échelle de l'atome en fonction du temps, en décrivant les mouvements des atomes et des molécules qui le composent et en suivant les lois de la mécanique Newtonienne (116–119). Dans le cas de l'étude de la P28GST, la dynamique moléculaire permet d'obtenir des informations au niveau atomique sur la mécanistique de ses fonctions enzymatiques.

Toute dynamique moléculaire se décompose en cinq grandes étapes (116,118,120), comprenant :

- L'initialisation du système, permettant de préparer la structure souhaitée à partir d'une conformation de départ, obtenue à partir d'une structure cristallographique ou d'un modèle *in silico*. Celui-ci est préparé en ajoutant des modèles de molécules d'eau et/ou de contre-ions si nécessaire dans une boîte de simulation afin de se rapprocher des conditions biologiques (118,121). De plus, des conditions périodiques aux limites sont définies pour simuler un système considéré comme infini et éviter les phénomènes de tension superficielle aux limites de la boîte de simulation si celle-ci était considérée comme isolée dans l'espace, sans partenaires aux limites de la boîte (117–119).
- La minimisation de l'énergie potentielle pour optimiser le système en éliminant les contacts défavorables qui pourraient exister entre les atomes, puis en maximisant les contacts favorables permettant d'atteindre un état stable du système au fond d'un puits de basse énergie (116,118,120).
- L'application d'un thermostat spécifique, à une température de simulation définie, afin de conférer une vélocité à chaque atome du système (118,120).

- L'équilibrage de l'énergie totale du système par des itérations dynamiques du solvant et des chaînes latérales, tout en contraignant la position du squelette peptidique (118,120).
- La production d'une trajectoire dynamique de la protéine solvatée, après avoir relâché toutes les contraintes de position préalables (116,120). Le mouvement protéique sera étudié lors de cette phase.

L'analyse d'une dynamique moléculaire vise à l'étude de la trajectoire en utilisant une palette d'outils très large (122). Les premières analyses couramment réalisées permettent la simplification de la trajectoire en éliminant les molécules de solvant si elles ne sont pas indispensables. Cela permet d'optimiser l'espace disque et la quantité de mémoire vive requise pour les analyses futures, et de simplifier sa visualisation.

Des séries temporelles peuvent être calculées, afin d'étudier entre autres la variation d'angles ou de distances entre atomes présélectionnés, l'évolution des structures secondaires, la variation de surfaces ou de volumes définis, ou encore l'évolution des interactions moléculaires réalisées au cours de la simulation.

Des grandeurs théoriques peuvent également être calculées. Le RMSD (Root-Mean Square Deviation), aussi appelé écart quadratique moyen, permet de mesurer la distance moyenne entre les coordonnées de deux groupes d'atomes et donc d'évaluer les changements conformationnels au cours du temps (116,123). Le RMSF (Root-Mean Square Fluctuation), également fluctuation quadratique moyenne, mesure la variation de position moyenne de chaque élément étudié au cours du temps par rapport à une position de référence, et donc met en évidence les portions de structure qui vont fluctuer plus ou moins au cours de la simulation de dynamique moléculaire, c'est-à-dire la flexibilité conformationnelle au cours du temps.

Enfin, les trajectoires peuvent aussi être à la base d'autres utilisations, comme par exemple l'estimation de l'énergie libre de liaison d'un ligand à son récepteur à l'aide de méthodes MM/GBSA (*Molecular Mechanics Generalized Born Surface Area*) et MM/PBSA (*Molecular Mechanics Poisson-Boltzmann Surface Area*).

## **2. Matériels & méthodes**

### ***a. Simulations sous GROMACS***

Le point de départ des simulations sous Gromacs a été le monomère issu de la structure cristallographique 1OE8 contenant une molécule de glutathion (80). La topologie a été générée en considérant une boîte de simulation cubique aux conditions périodiques, avec 10 Å de marge autour de la structure protéique, remplie par des modèles d'eau TIP3P (124,125) et un contre-ion Na<sup>+</sup>.

Tous les calculs ont été réalisés en utilisant le champ de force GROMOS56 54a7 (126), un pas d'intégration de 2 fs. Le calcul des interactions électrostatiques de longue distance a été réalisé en appliquant la méthode du maillage particulier d'Ewald (*particle mesh Ewald method*) (127), avec une distance maximale fixée à 10 Å.

Dans un premier temps, des restrictions harmoniques ont été appliquées aux carbones alpha de la protéine. Puis l'énergie du système a été minimisée en appliquant l'algorithme de la plus grande pente (*Steepest descent algorithm*) lors de 5000 pas de minimisation. Le système a ensuite été équilibré pendant 300 ps dans un ensemble canonique NVT, puis dans un ensemble isotherme isobare pendant 2000 ps, en appliquant un thermostat de Berendsen modifié fixé à 300K. Enfin, les restrictions sur les carbones alpha ont été relâchées pour la réalisation des trajectoires de dynamique moléculaire.

### ***b. Simulations sous Amber***

#### **> Pourquoi passer à Amber :**

Quelques mois après le début des travaux de thèse, le logiciel principal de simulation de dynamique moléculaire a été changé de GROMACS à Amber. Ce choix a été réalisé entre autres pour préparer de potentielles simulations de dynamique quantique, MM/PBSA et MM/GBSA, qui étaient au départ considérées puis mises de côté. Le changement a également été réalisé pour permettre la compatibilité optimale des simulations avec les outils antechamber, tleap et cpptraj, fourni via AmberTools, qui présentent des avantages non négligeables. Ces outils permettent respectivement de générer les paramètres de champ de

force pour des molécules organiques, de préparer les systèmes pour une simulation sous Amber, et de proposer une large gamme d'outils d'analyses des trajectoires de dynamique moléculaire.

### > **Dynamique moléculaire classique :**

Les simulations de dynamique moléculaire classique effectuées sous Amber18 ont été réalisées avec le champ de force AMBER ff14SB (128) avec un pas d'intégration de 2 fs, à la pression constante de 1 atm maintenue par le barostat de Berendsen, à la température constante de 300 K par le thermostat de Langevin. Les structures protéiques étudiées ont été placées dans une boîte orthogonale aux conditions périodiques, avec 10 Å de marge autour de la structure protéique, remplie de modèles d'eau TIP3P et de contre-ions Na<sup>+</sup>.

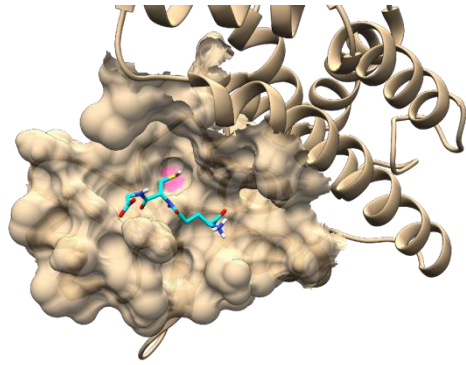
Des restrictions harmoniques ont été appliquées au préalable aux carbones alpha de la protéine. Le système simulé a été minimisé énergétiquement par 1000 pas d'algorithme de la plus grande pente (*steepest descent algorithm*) suivi de 1000 pas d'algorithme de gradient conjugué (*conjugate gradient algorithm*). Le système a ensuite été amené à la température de simulation en passant de 0 K à 300 K lors de 9000 pas de simulation, puis maintenu à 300 K pendant 1000 pas supplémentaires. Une étape d'équilibrage a été réalisée pendant 2 ns en utilisant les conditions expérimentales de la production tout en laissant les contraintes sur les carbones alpha, dans le but de permettre le bon positionnement des molécules d'eau et des chaînes latérales des différents résidus. Enfin, en relâchant toute contrainte sur les carbones alpha et en utilisant la dernière image obtenue à la fin de l'étape d'équilibrage, l'étape de production a permis l'obtention de trajectoires de dynamique moléculaire, variables en fonction du temps de simulation et du nombre de répliques souhaitées.

### **3. Résultats**

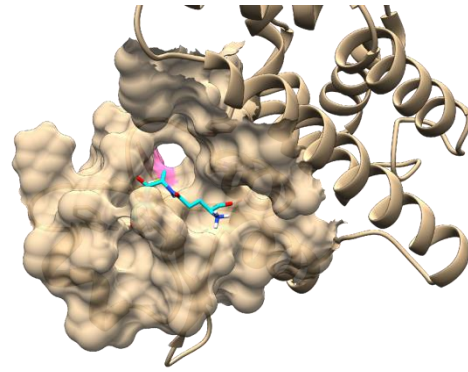
#### *a. Premières dynamiques moléculaires avec Gromacs*

Les premières dynamiques moléculaires réalisées dans le cadre des travaux de thèse ont été simulées par le logiciel Gromacs. Ce sont des dynamiques de test à but formateur, faites avec les outils immédiatement disponibles afin de comprendre les notions principales de la dynamique moléculaire. Cela a également permis de se familiariser avec la protéine, et d'essayer d'appréhender de premiers mouvements au niveau du site actif.

Lors des 4 répliques de 150 ns réalisées, l'ouverture d'un canal à proximité de la tyrosine catalytique et du thiol du glutathion a été observée de manière transitoire (Figure 21). Ce canal est majoritairement constitué d'acides aminés de la boucle catalytique, allant des acides aminés 10 à 16. Afin de vérifier cette observation sur les répliques réalisées, un script Python3 a été développé pour suivre l'évolution du volume créé par l'ouverture de ce canal. Parmi les différents logiciels disponibles aidant à calculer le volume dans une sphère, il n'existe pas d'outil fonctionnant en local permettant de prendre en entrée les trajectoires de dynamique moléculaire. Il a donc été nécessaire de travailler avec des logiciels ne prenant en entrée que des fichiers PDB, comme CAVITY de LigBuilder 2 (129), Fpocket (130) ou encore POVME 2 (131). Parmi ces logiciels, POVME 2 a été retenu car il permettait une plus grande flexibilité d'utilisation. Une image de chaque trajectoire obtenue a été extraite toutes les nanosecondes, puis converties au format PDB. Ensuite, à l'aide du logiciel POVME 2, le volume du canal formé a été mesuré dans une sphère de 3 Å de rayon centrée sur les coordonnées moyennes des carbones alphas allant des acides aminés 10 à 16, délimitant ce canal. Enfin, les résultats ont été compilés sous R (Figure 22).

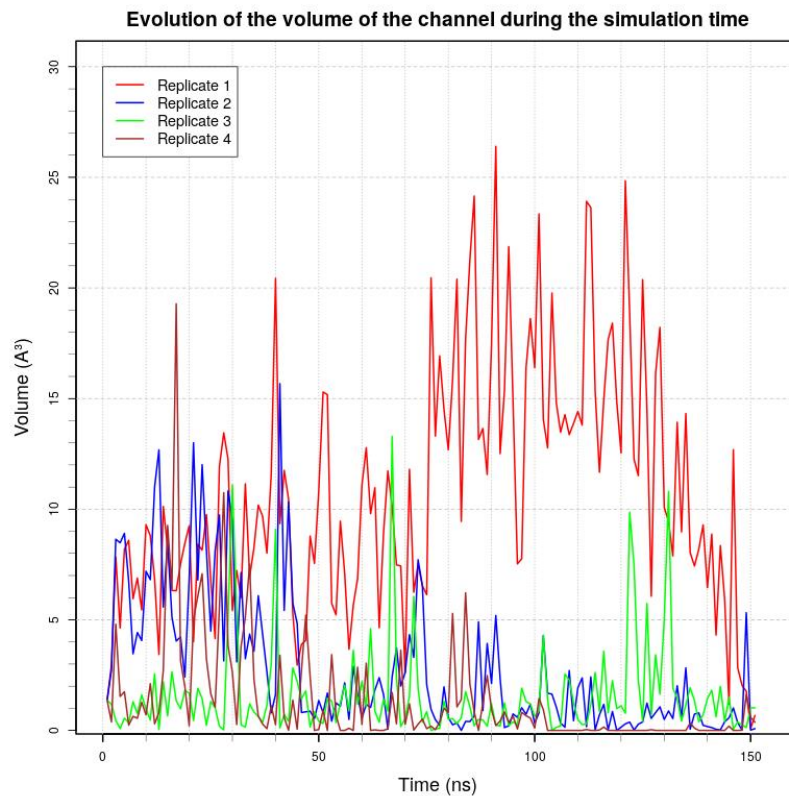


P28GST (PDB code: 1OE8) after minimization and equilibration steps



P28GST (PDB code: 1OE8) after minimization and equilibration steps, and 17ns of simulation

**Figure 21.** Représentation tridimensionnelle de la structure cristallographique de P28GST (identifiant PDB : 1OE8) en monomère au début de la production (à gauche) et après 17 ns de production lors de la réplique 4 (à droite), avec représentés en rose la Tyrosine 10 et en cyan le glutathion.



**Figure 22.** Graphique de l'évolution du volume du canal formé à proximité de la Tyrosine 10 lors des 4 répliques de dynamique moléculaire réalisées sous Gromacs.



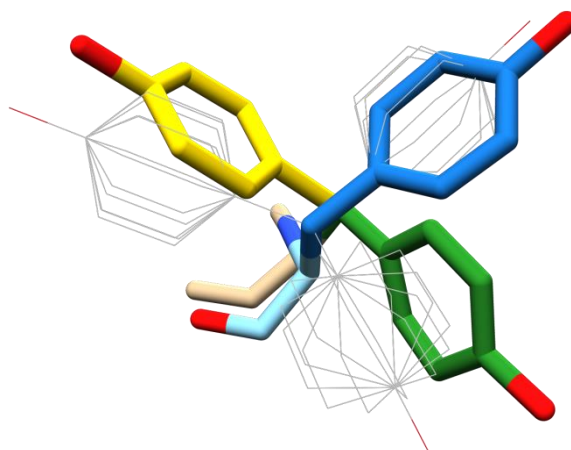
Malheureusement, l'interprétation de ce résultat est difficile, du fait de la nature du canal en elle-même. Celui-ci est variable, peut posséder plusieurs configurations, et les techniques pour mesurer un volume sont non adaptées pour un canal ni pour des dynamiques moléculaires. La mesure de l'évolution du volume au sein d'une sphère paraissait une idée pertinente, mais son application est complexe : soit la sphère à mesurer est trop grande, et vient donc soit ajouter un bruit de fond trop important ou comptabiliser la présence du glutathion à sa proximité, soit la sphère est trop petite et n'inclue pas tout le canal. D'autres aspects ont été étudiés, comme la mesure de la distance entre deux carbones alpha de la boucle catalytique, sans succès.

Toutefois, même si l'interprétation quantitative n'a pas été couronnée de succès, l'interprétation qualitative permet de conclure qu'il peut y avoir un espace suffisant pour que la Tyrosine 10 catalytique puisse basculer de part et d'autre de ce canal, afin de naviguer entre ses conformations IN et OUT.

### ***b. Dynamiques moléculaires de 150 ns avec Amber18***

A partir des conditions expérimentales proches de celles imposées sous Gromacs, et en utilisant le système dimérique pour se rapprocher au maximum de la réalité biologique, 4 répliques de 150 ns de dynamique moléculaire classique ont été réalisées sous AMBER18.

L'objectif principal de cette étude a été l'observation des mouvements de la tyrosine catalytique en partant de la conformation IN ou de la conformation OUT. Au cours de ces dynamiques, une nouvelle conformation de la tyrosine 10 non présente dans la cristallographie a pu être identifiée. Au vu de sa localisation, à mi-chemin entre les conformations IN et OUT et pointant vers l'intérieur de la protéine, celle-ci a été appelée conformation MID (Figure 23). Après avoir recherché les différences possibilités de mesure entre ces conformations, le critère principal retenu a été la mesure de l'angle dièdre Chi1, permettant d'identifier différents rotamères. Ces rotamères de l'angle dièdre chi 1 permettent de différencier les conformations IN et OUT, et d'inclure la nouvelle conformation MID.

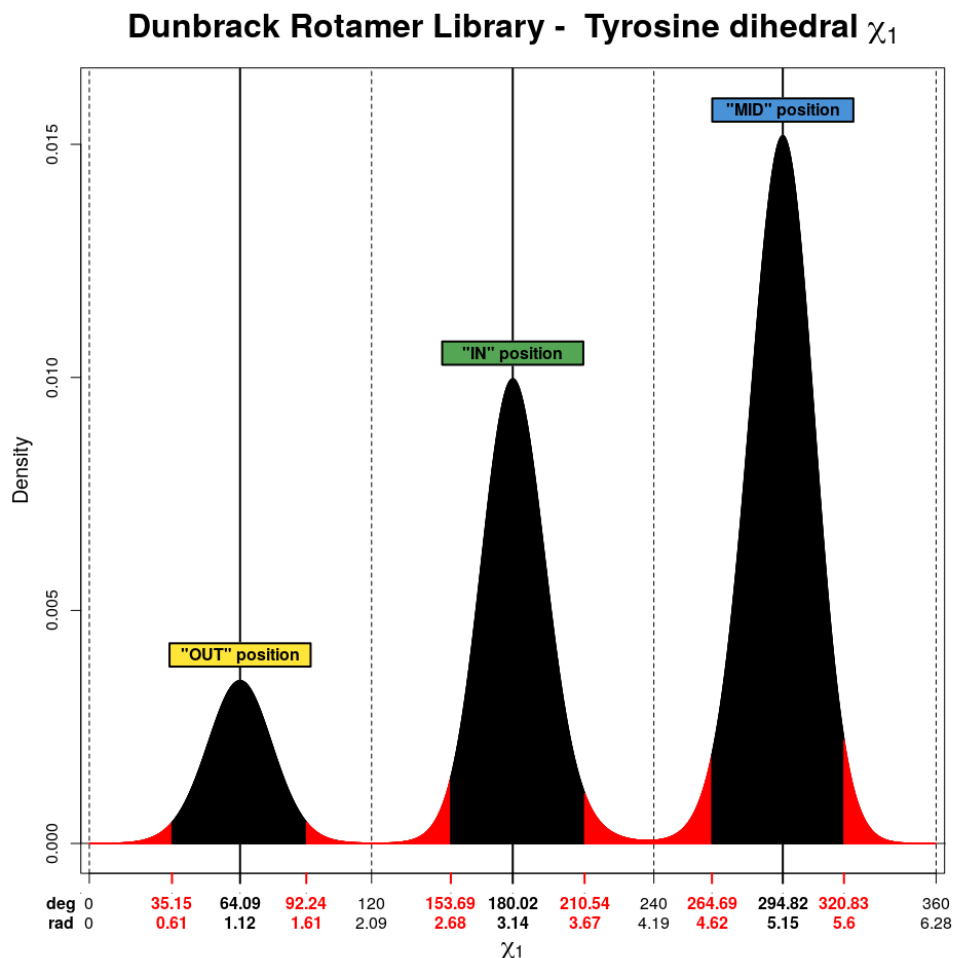


**Figure 23.** Représentation des conformations IN (en vert) et OUT (en jaune) de la tyrosine 10 observées dans la cristallographie de P28GST (identifiant PDB : 1OE8) et de la conformation MID (en bleu) observée lors des productions de dynamique moléculaire sous AMBER18

La distribution classique des angles dièdres chi1 de 25490 résidus tyrosine décrite dans la librairie de rotamères Dunbrack 2010 à partir de 3845 structures cristallographiques a permis de borner les 3 ensembles conformationnels de la tyrosine 10 (Figure 24). L'angle dièdre chi1 est déterminant pour discriminer les 3 conformations limites de la chaîne latérale de la tyrosine 10 :

- La conformation IN est caractérisée par des angles chi1 compris entre  $153.69^\circ$  et  $210.54^\circ$  (soit 2.68 rad et 3.67 rad),
- La conformation OUT est caractérisée par des angles chi1 compris entre  $35.15^\circ$  et  $92.24^\circ$  (0.61 rad et 1.61 rad),
- La conformation MID est caractérisée par des angles chi1 compris entre  $264.69^\circ$  et  $320.83^\circ$  (4.62 rad et 5.60 rad).

Bien que ces conformations dépendent de l'environnement de la tyrosine, la conformation MID est surreprésentée tandis que la conformation OUT est sous-représentée dans la librairie (Figure 24).



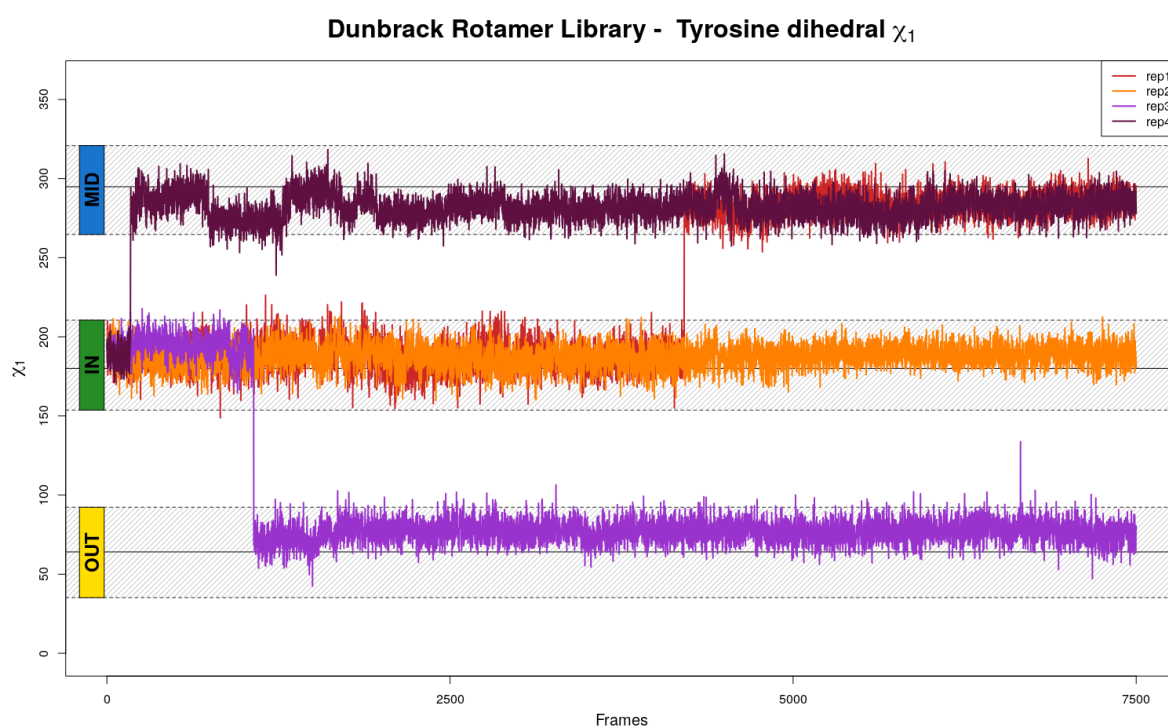
*Figure 24. Distribution de l'angle dièdre  $\chi_1$  déterminé à partir de 25 490 tyrosines décrites dans la librairie de rotamères Dunbrack 2010 (132), les bornes à 5% bilatérales pour chaque rotamère, et la correspondance de chaque conformation de la Tyrosine 10 observée*

En suivant l'évolution de l'angle dièdre  $\chi_1$  de la tyrosine catalytique du monomère A, ayant comme point de départ la conformation IN, au cours des différentes dynamiques moléculaires réalisées (Figure 25), ces trois rotamères ont pu être retrouvés. La conformation MID a pu être observée dans deux répliques et la conformation OUT dans une réplique. En ce qui concerne l'évolution de l'angle dièdre  $\chi_1$  de la tyrosine catalytique du monomère B, ayant comme point de départ la conformation OUT, cette dernière est retrouvée dans 100% des images enregistrées, indiquant un état stable de cette conformation OUT au cours de la simulation.

Afin d'évaluer la rareté de la conformation MID, et la stabilité de la conformation OUT, d'autres simulations ont été brièvement tentées en relançant de courtes dynamiques

moléculaires à partir de la première étape d'initialisation du système ou en changeant les contre-ions, mais celles-ci n'ont pas permis d'observer de changements de conformation.

Au vu de la faible occurrence du changement de conformation, il est possible que la conformation MID soit un évènement métastable, nécessitant un apport d'énergie pour l'obtenir, et que la conformation OUT soit la conformation la plus stable possible lorsque la tyrosine catalytique est protonée. Avant d'envisager d'autres techniques plus avancées, il est nécessaire de vérifier ces hypothèses avec de nouvelles dynamiques moléculaires classiques plus longues.



*Figure 25. Suivi de l'angle dièdre  $\chi_1$  (en degré) au cours des productions de dynamique moléculaire réalisées sous AMBER18*

### c. Dynamiques moléculaires de 1.5 $\mu$ s avec Amber18

De nouvelles dynamiques moléculaires classiques plus longues, avec les mêmes paramètres expérimentaux, ont été réalisées afin de confirmer l'hypothèse selon laquelle la conformation MID est une conformation métastable. Avec l'obtention d'un accès au méso-centre de Calcul Scientifique Intensif de l'Université de Lille, il a été possible de réaliser 3 répliques de 1.5  $\mu$ s de dynamique moléculaire. Il est à noter qu'avec la limite de temps maximale imposée par calcul et les performances moyennes sur les cartes graphiques du méso-centre, chaque réplique a été lancée deux fois, par lots de 750ns.

De plus, pour pouvoir étudier l'impact de tous les acides aminés présents naturellement chez la P28GST, les 4 derniers acides aminés non présents dans la cristallographie étudiée ont été reconstruits.

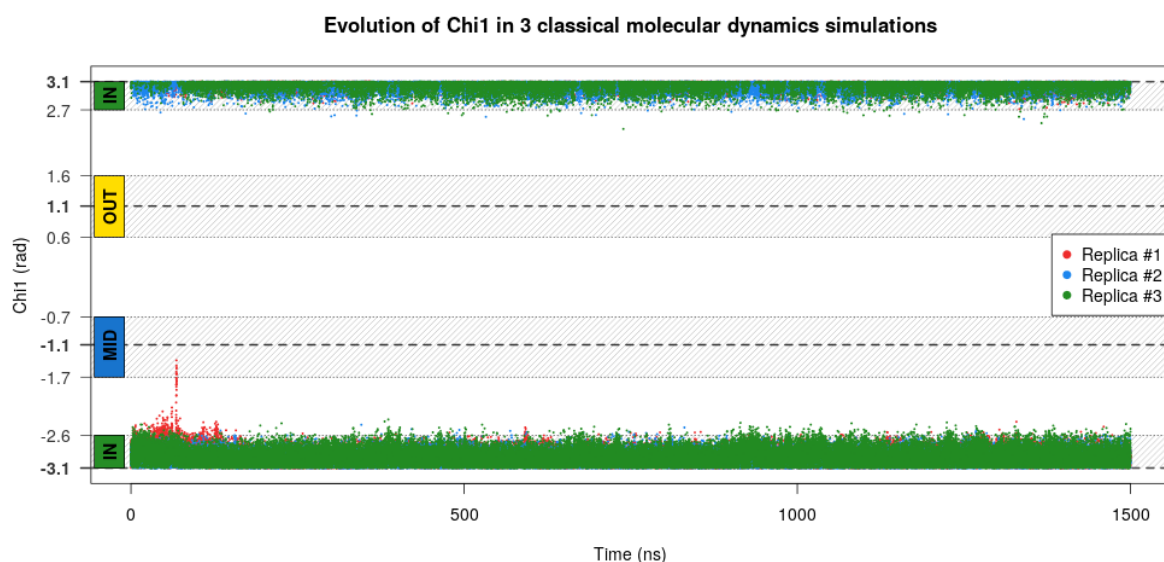


Figure 26. Évolution de l'angle dièdre Chi1 au cours des 3 répliques de dynamique moléculaire classique de 1.5 $\mu$ s réalisées sous AMBER18

De la même manière que pour les premières dynamiques moléculaires, l'angle dièdre chi1 de la tyrosine catalytique a été suivi le long des différentes trajectoires (Figure 26). Il est possible de voir que seule la conformation IN a été explorée au cours de toutes les simulations. Cette observation étaye l'hypothèse de la présence de potentielles barrières énergétiques présentes entre chaque conformation, et conditionnant les différentes étapes du processus d'activation du glutathion, et donc par conséquent les activités enzymatiques de la P28GST.

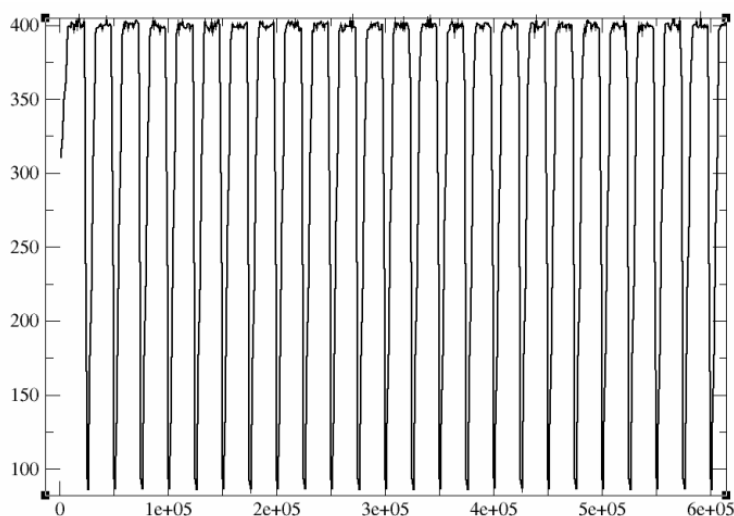
En conclusion à ces simulations, la dynamique moléculaire classique n'est pas suffisante pour étudier avec détail le phénomène d'activation du glutathion. D'autres techniques plus avancées sont requises, en amenant de l'énergie au système ou en forçant le changement de conformation par la variation de l'angle dièdre  $\chi_1$ .

## **4. Alternatives à la dynamique moléculaire classique**

### ***a. Méthode du recuit simulé***

La première technique dérivée de la dynamique moléculaire classique envisagée a été la méthode du recuit simulé (*simulated annealing*). Cette technique permet de réaliser une exploration conformationnelle d'un système pour générer des groupes de structures différentes (133).

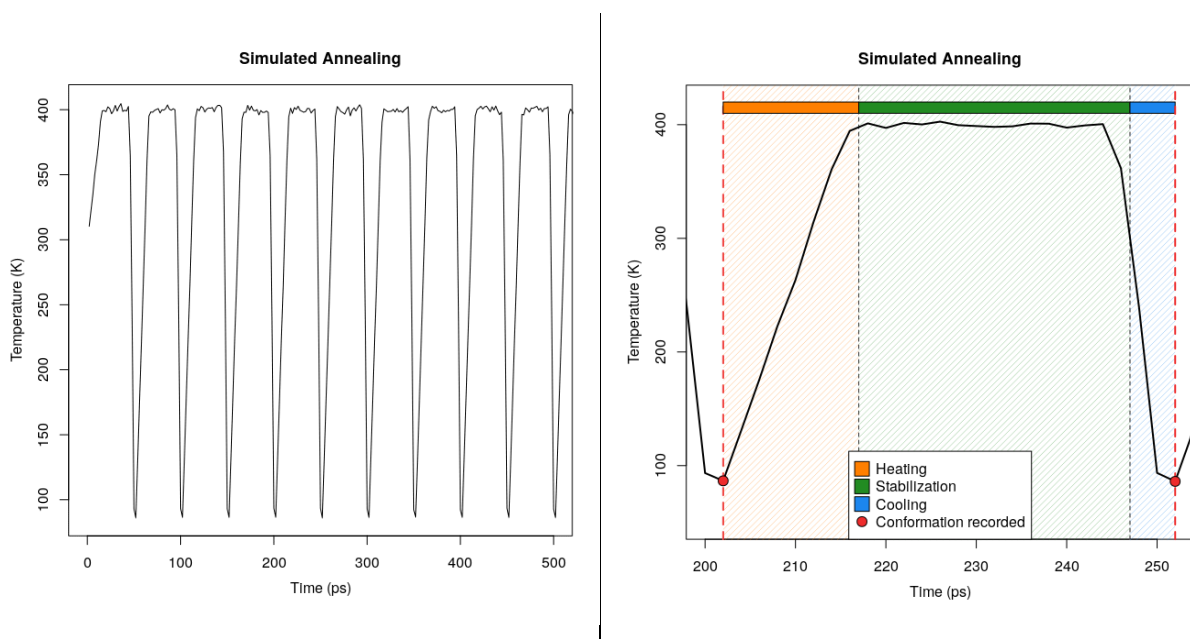
La méthode du recuit simulé est inspirée d'un processus en métallurgie, qui alterne des cycles de refroidissement lent et de réchauffage, le recuit, pour laisser le métal travaillé évoluer d'un état d'équilibre vers un autre état d'équilibre de plus faible énergie, pour au final obtenir un réseau métallique minimisé ne possédant que très peu de défauts (134). Appliqué aux simulations de dynamique moléculaire, l'augmentation cyclique de la température de simulation permet d'apporter de l'énergie au système pour l'obtention de nouveaux états d'équilibre (Figure 27). Le refroidissement permet par la suite de bloquer la conformation obtenue dans un minimum local d'énergie.



**Figure 27. Figure théorique du recuit simulé (135)**

## > Matériels et méthodes :

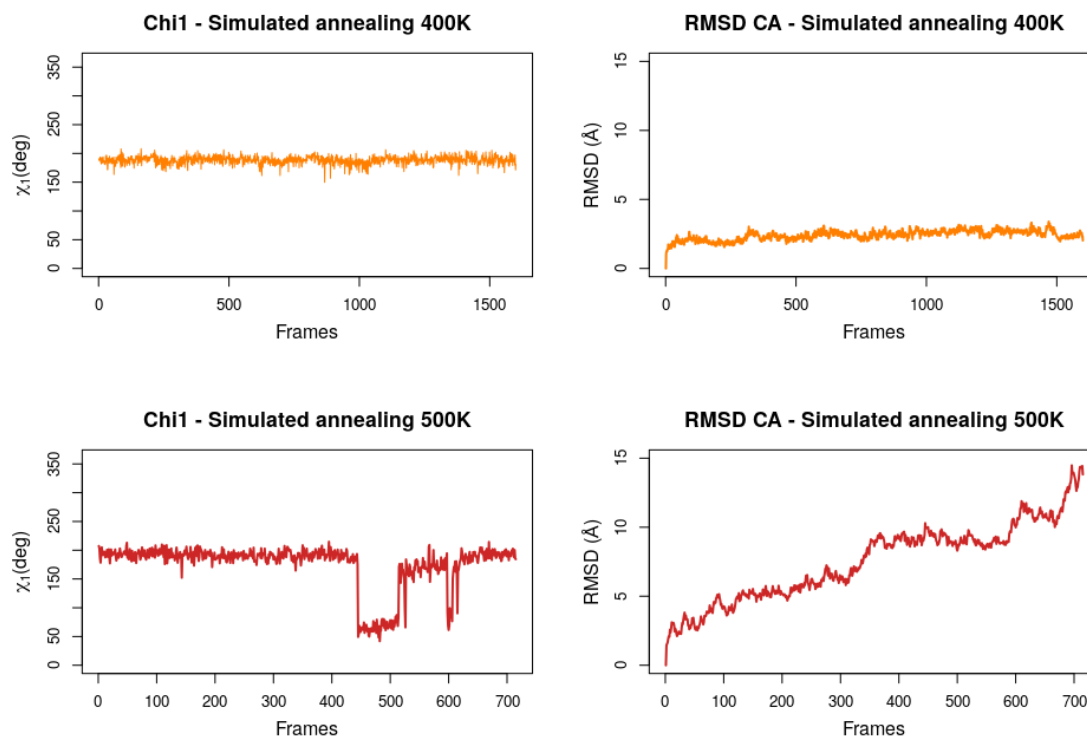
Le système protéique utilisé lors de méthode du recuit simulé a été le point de départ des dynamiques moléculaires classiques de 150 ns réalisées sous Amber18. Le protocole a été paramétré afin de réaliser un temps total de simulation de 80 ns en enchaînant des cycles de 50 ps (soit 25.000 pas d'intégration pour un cycle), comprenant dans l'ordre un chauffage d'une température basse de 50K à une température haute de 400K ou 500K en fonction de la simulation pendant 7.500 pas d'intégration, une stabilisation de la haute température pendant 15.000 pas, un refroidissement du système jusqu'à la température basse de 50K en 2.500 pas, pour enfin enregistrer la conformation observée à la fin de ce cycle (Figure 28).



**Figure 28.** *Figures expérimentales du recuit simulé réalisé chez P28GST*

## > Résultats :

Lors de la simulation ayant pour haute température 400K, aucune autre conformation de la Tyrosine 10 n'a été observée (Figure 29). Toutefois, lors de la simulation ayant pour haute température 500K, les autres conformations de la tyrosine 10 ont pu être visualisées, au détriment de la stabilité du système, où il était également possible de voir la perte de la structuration du système en dimère, la perte de certaines structures secondaires, et le crash du système au bout de 35 ns de simulation.



*Figure 29. Évolution des angles dièdres  $\chi_1$  et du RMSD des carbones alpha au cours des simulations de recuit simulé*

L'utilisation du recuit simulé n'est pas intéressante pour d'obtenir les différentes conformations de la Tyrosine 10, celle-ci privilégiant l'exploration de l'espace conformationnel de la protéine entière au lieu de l'exploration de la flexibilité d'une plus petite partie du système.



## ***b. Dynamique moléculaire dirigée***

La seconde technique dérivée de la dynamique moléculaire classique envisagée a été la dynamique moléculaire dirigée (*steered molecular dynamics*). Cette technique permet de réaliser une exploration conformationnelle d'un système en appliquant des forces externes, accélérant les processus trop lents pour être observés par modélisation *in silico* (136).

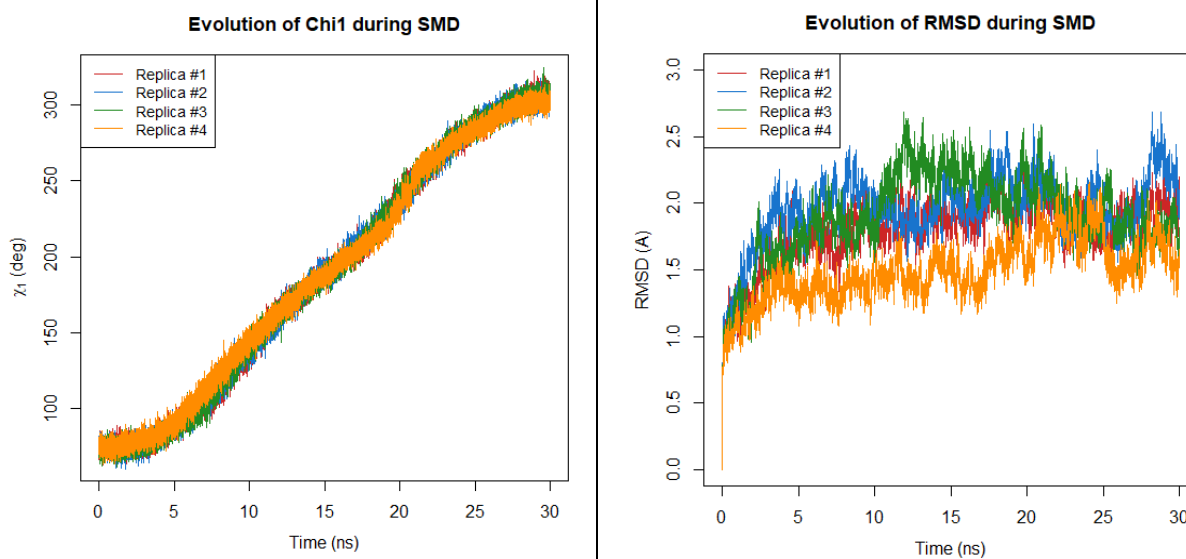
Dans le cadre des travaux de thèse, la dynamique moléculaire dirigée a été utilisée pour vérifier la stabilité du système protéique lors de la bascule de la Tyrosine 10 en imposant un mouvement des conformations IN à OUT en augmentant linéairement la valeur de l'angle dièdre Chi1.

### **> Matériels et méthodes :**

Le système protéique utilisé lors de méthode du recuit simulé a été le point de départ des dynamiques moléculaires classiques de 150 ns réalisées sous AMBER18. Le protocole a été paramétré afin de réaliser un temps total de simulation de 80 ns en enchaînant des cycles de 50 ps (soit 25.000 pas d'intégration pour un cycle), comprenant dans l'ordre un chauffage d'une température basse de 50K à une température haute de 400K ou 500K en fonction de la simulation pendant 7.500 pas d'intégration, une stabilisation de la haute température pendant 15.000 pas, un refroidissement du système jusqu'à la température basse de 50K en 2.500 pas, pour enfin enregistrer la conformation observée à la fin de ce cycle.

### **> Résultats :**

Au cours de 4 répliques de 30ns réalisées, toutes les conformations ont été observées, sans impact sur l'intégrité globale de la protéine mesurée par le RMSD sur tous les carbones alpha de la protéine (Figure 30).



**Figure 30.** Évolution de l'angle dièdre Chi1 et du RMSD des carbones alpha au cours de la simulation de dynamique moléculaire dirigée

La conformation MID étant donc une conformation qui n'entraîne pas d'impact sur l'intégrité de la protéine, et étant à mi-chemin entre les deux conformations observées dans la cristallographie, la non-observation de cette conformation MID en dynamique moléculaire classique est très certainement due à des barrières énergétiques qui n'ont pas été atteintes, à l'exception du point de départ métastable qui a permis son observation. Il est donc nécessaire de réaliser des dynamiques moléculaires mettant en jeu le calcul de l'énergie libre sans imposer le mouvement à la tyrosine 10. L'une des techniques les plus adaptées, et choisie dans le cadre de la thèse, est la métadynamique.

### **III. Métadynamique**

Le franchissement des barrières énergétiques est un événement indispensable, permettant à une protéine de naviguer entre différents états stables. Toutefois, les techniques de dynamique moléculaire classique simulent généralement les mouvements d'un seul état stable, ou de plusieurs états stables séparés par de faibles barrières énergétiques. Dans ces conditions, l'exploration du paysage conformationnel par le franchissement de barrières énergétiques est souvent le fait d'un événement rare. Par exemple, pour un événement nécessitant 1ms pour être réalisé à l'échelle biologique, la simulation requise à la vitesse moyenne de 50 ns/jour obtenue classiquement au laboratoire prendrait 20000 jours de calcul. Les échelles de temps biologiques ne sont pas compatibles avec les échelles de temps de la dynamique moléculaire tout atome : il est obligatoire d'accélérer l'obtention de ces événements. Les méthodes du recuit simulé et de la dynamique moléculaire dirigée n'ayant pas donné de résultats, la méthode qui a été retenue est la métadynamique.

#### **1. Matériels et méthodes**

##### ***a. Métadynamique standard***

La métadynamique est une technique de dynamique moléculaire supervisée permettant de cartographier l'énergie du système en fonction de certains degrés de liberté, appelés variables collectives. En introduisant un biais énergétique, l'algorithme empêche le système d'évoluer vers les conformations qu'il a déjà visitées, permettant l'exploration de toutes les conformations le long d'une ou plusieurs variables collectives (Figure 33).

Un apport énergétique pour l'exploration d'une variable collective permet de biaiser le paysage énergétique d'un système en ajoutant des gaussiennes énergétiques correspondant aux dernières conformations visitées au temps  $t$  par l'algorithme. Ces gaussiennes sont définies par une largeur et une hauteur, et se positionnent sur des valeurs échantillonnées des variables collectives (Figure 31).

$$V(\vec{s}, t) = \sum_{k\tau < t} W(k\tau) \exp\left(-\sum_{i=1}^d \frac{(s_i - s_i(q(k\tau)))^2}{2\sigma_i^2}\right)$$

*Figure 31. Équation du potentiel de biais  $V$  amené par le logiciel construit dans l'espace des variables collectives  $\vec{s}$ , exprimé comme étant la somme des gaussiennes déposées dans l'espace des variables collectives avec  $\tau$  correspondant au pas de dépôt des gaussiennes,  $\sigma_i$  la largeur des gaussiennes et  $W(k\tau)$  la hauteur des gaussiennes.*

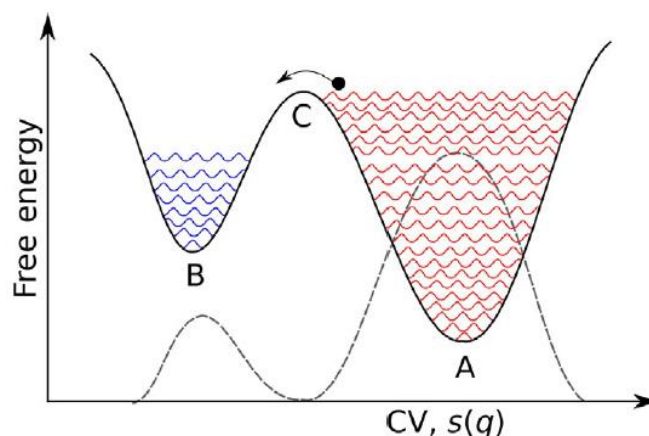
Progressivement, le paysage énergétique est rempli par la somme de toutes ces gaussiennes, encourageant la simulation à explorer des conformations moins favorables énergétiquement. En fin de simulation, le paysage énergétique peut être reconstruit à partir du négatif des biais énergétiques apportés au système (Figures 32 et 33).

$$V(\vec{s}, t \rightarrow \infty) = -F(\vec{s}) + C$$

*Figure 32. Le potentiel de biais  $V$  converge vers l'opposé de l'énergie libre  $F$  du système.*

Le choix des variables collectives est un critère primordial lors de la réalisation d'une métadynamique. Ce choix doit être effectué en fonction du phénomène étudié, et peut être réalisé parmi tous les paramètres du système, incluant des distances, des angles, des angles dièdres, des RMSD, des rayons de giration, des volumes, etc. Plusieurs conditions sont nécessaires pour que ces variables collectives soient pertinentes (138) :

- Elles doivent permettre de distinguer les états initiaux, finaux et intermédiaires.
- Elles doivent permettre de décrire les évènements rares qui sont à caractériser.
- Elles ne doivent pas être trop nombreuses, au risque de ralentir l'exploration du système et de ne jamais atteindre la convergence, et donc l'obtention du profil d'énergie libre complet.



*Figure 33. Illustration schématique d'une simulation de métadynamique (137). Le système a débuté au niveau de la conformation A, au niveau d'un minimum énergétique. Au cours de la simulation, des biais énergétiques sont ajoutés au paysage énergétique sous la forme de gaussiennes (ici en rouge), remplissant progressivement le puits de potentiel de la conformation de départ. Lorsque la simulation atteint un certain point où l'énergie du système est élevée, l'algorithme tend à favoriser l'exploration du paysage pour rechercher une nouvelle conformation énergétiquement plus favorable en franchissant les barrières énergétiques, ici vers la conformation B. La simulation continue ensuite jusqu'à remplir le paysage énergétique et l'aplatir totalement. La somme de tous les biais énergétiques apporté au système permettent d'estimer le négatif du paysage énergétique (en tirets gris), permettant de retrouver les minima énergétiques et les barrières énergétiques du système, comme dans cet exemple pour obtenir l'information sur la barrière énergétique C conditionnant le passage entre les conformations A et B.*

### ***b. Métadynamique bien tempérée***

Lors d'une métadynamique standard, les gaussiennes ajoutées durant toute la simulation sont de hauteur constante, ce qui pousse le système à explorer les régions à haute énergie libre et à obtenir des valeurs moins précises au niveau des transitions énergétiques. Pour pallier ce problème, il est possible d'utiliser la métadynamique bien tempérée (*well-tempered metadynamics* - WT), diminuant progressivement la hauteur des gaussiennes lors de la simulation (Figure 34).

$$W(k\tau) = W_0 \exp\left(-\frac{V(\vec{s}(q(k\tau)), k\tau)}{k_B \Delta T}\right)$$

*Figure 34. Variation de la hauteur des gaussiennes  $W(k\tau)$  lors de la simulation, avec  $W_0$  la hauteur initiale de la gaussienne,  $\Delta T$  la différence de température amenée pour l'exploration de l'espace des variables collectives et  $k_B$  la constante de Boltzmann.*

En appliquant cette variation de hauteur de gaussiennes, le biais énergétique converge progressivement au cours de la simulation, mais n'atteint pas le profil d'énergie libre total (Figure 35).

$$V(\vec{s}, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(\vec{s}) + C$$

*Figure 35. Le potentiel de biais  $V$  construit dans l'espace des variables collectives converge vers l'opposé d'une fraction de l'énergie libre  $F$  du système, fonction de la température  $T$  de simulation et de la température  $T+\Delta T$  des variables collectives.*

Le paramètre  $\Delta T$  est à définir pour réguler l'exploration de l'énergie libre, avec  $\Delta T=0$  correspondant à la dynamique moléculaire classique et  $\Delta T \rightarrow \infty$  correspondant à la métadynamique standard. Dans différents logiciels permettant la réalisation de métadynamiques, ce paramètre  $\Delta T$  est souvent intégré dans la notion de facteur de biais, décrit comme le rapport entre la température  $T+\Delta T$  des variables collectives et la température  $T$  du système (Figure 36).

$$\gamma = \frac{T + \Delta T}{T}$$

*Figure 36. Expression du facteur de biais.*

Les différentes variantes de métadynamique sont proposées par le plugin open-source PLUMED, ayant pour avantages le développement des différentes méthodes directement par la communauté et son implantation possible avec la grande majorité des moteurs de dynamique moléculaire. Dans le cas spécifique de la métadynamique bien tempérée utilisé lors de mes travaux de thèse, l'utilisation de PLUMED en parallèle du moteur de dynamique moléculaire pmemd.cuda.mpi proposé par Amber18 est activé par l'utilisation de la ligne de commande "plumed = 1" dans les fichiers d'entrée de la simulation de dynamique moléculaire, en renseignant un fichier d'instruction spécifique à PLUMED. Les autres paramètres de dynamique moléculaire utilisés ont été strictement identiques aux paramètres utilisés lors de la réalisation des dynamiques moléculaires classiques précédentes.

Le fichier d'instruction de PLUMED contient la définition de différentes grandeurs et/ou variables collectives qui sont utilisées, l'activation de la méthode de métadynamique par un bloc spécifique identifié par l'instruction METAD, et la génération de fichiers de sortie. Le bloc d'activation de la métadynamique prend en entrée :

- Le pas de dépôt des gaussiennes avec le paramètre PACE,
- La hauteur initiale des gaussiennes utilisées en début de simulation avec le paramètre HEIGHT,
- Le facteur de biais avec le paramètre BIASFACTOR,
- La température de simulation avec le paramètre TEMP,
- Les variables collectives utilisées avec le paramètre ARG,
- Le pas d'échantillonnage des variables collectives avec le paramètre SIGMA,
- Les valeurs minimales et maximales des variables collectives avec les paramètres GRID\_MIN et GRID\_MAX,
- Le nom du fichier de sortie renseignant les différentes gaussiennes ajoutées au cours de la simulation avec le paramètre FILE.

Dans certains cas, il est possible d'ajouter des contraintes harmoniques sur une ou plusieurs grandeurs par un autre bloc spécifique identifié par l'instruction RESTRAINT. Ce bloc prend en argument les différentes grandeurs à contraindre avec le paramètre ARG, les valeurs des constantes de force à appliquer sur chacune des grandeurs avec le paramètre KAPPA, ainsi que la position des contraintes avec le paramètre AT.

La génération des fichiers de sortie permettent l'obtention de différentes grandeurs à tous les intervalles de temps définis par le paramètre STRIDE, comprenant les différents biais énergétiques, les valeurs des variables collectives, ou encore des grandeurs calculées par le logiciel, définies en amont mais non utilisées comme variable collective.

L'obtention des profils d'énergie libre a été permis par la commande `sum_hills` proposée par PLUMED, afin de les construire à partir de la somme de toutes les gaussiennes déposées pendant la simulation et renseignées dans les fichiers de sortie de l'algorithme. Enfin, toutes les données obtenues ont été analysées par le logiciel RStudio.

### *c. Dynamique Moléculaire à Biais Adaptatif - ABMD*

Des méthodes dérivées de la métadynamique existent permettant de l'accélérer, comme la dynamique moléculaire à biais adaptatif (*Adaptively Biased Molecular Dynamics*), également appelée ABMD. Cette méthode se caractérise notamment par la simplification des paramètres de contrôle et par le remplacement des gaussiennes par des B-splines cubiques discrétisant le biais énergétique par application de la méthode d'Euler, puis les lissant avec une distribution de Dirac. De plus, à la différence de la métadynamique standard, le biais amené par la méthode ABMD varie en fonction du temps de la simulation.

De la même manière que pour la métadynamique standard, la méthode ABMD admet un algorithme additionnel, l'ABMD bien tempérée (*Well Tempered ABMD - WT-ABMD*), permettant d'améliorer la stabilité et la convergence de la méthode ABMD classique en utilisant une pseudo-température  $T'$  équivalente à la température  $T+\Delta T$  des variables collectives de la métadynamique bien tempérée.

Les simulations d'ABMD se servant directement du moteur de dynamique moléculaire `pmemd.cuda.mpi` de Amber 18, les paramètres de dynamique moléculaire qui ont été utilisés sont strictement identiques aux paramètres utilisés lors de la réalisation des dynamiques moléculaires classiques. L'activation du module d'énergie libre a été réalisée via le paramètre "`infe = 1`". Les différents paramètres de l'ABMD ont ensuite été ajoutés à la suite des fichiers d'entrée, dans une section identifiée par la balise `&ambd`.



Cette section contient :

- Le mode utilisé, ici FLOODING, avec le paramètre mode,
- Le nom et l'emplacement du fichier de suivi généré avec le paramètre monitor\_file,
- L'intervalle d'enregistrement dans le fichier de suivi avec le paramètre monitor\_freq,
- Le fichier comprenant la définition de toutes les variables collectives avec le paramètre cv\_file,
- Le pas de dépôt des biais énergétiques avec le paramètre timescale,
- La pseudo-température utilisée par la méthode bien tempérée avec le paramètre wt\_temperature,
- Le fichier de sortie comprenant les informations sur les biais énergétiques ajoutés au cours de la simulation avec le paramètre wt\_umbrella\_file.

Le fichier de définition des variables collectives contient quand à lui, pour chaque variable collective :

- Le type de grandeur utilisée avec le paramètre cv\_type,
- Le nombre d'atomes définissant cette grandeur avec le paramètre cv\_ni,
- Le numéro des atomes définissant cette grandeur avec le paramètre cv\_i,
- Le pas d'échantillonnage de la grandeur avec le paramètre resolution,
- Les valeurs minimales et maximales autorisées avec les paramètres cv\_min et cv\_max.

L'obtention des profils d'énergie libre a été réalisé par l'utilitaire nfe-umbrella-slice proposée par Amber18, celui-ci permettant d'accéder aux informations dans le fichier de sortie compressé pour générer un nouveau fichier ASCII lisible. Enfin, tous les fichiers de sortie obtenus ont été analysés à l'aide de RStudio.

## 2. Résultats

### *a. ABMD*

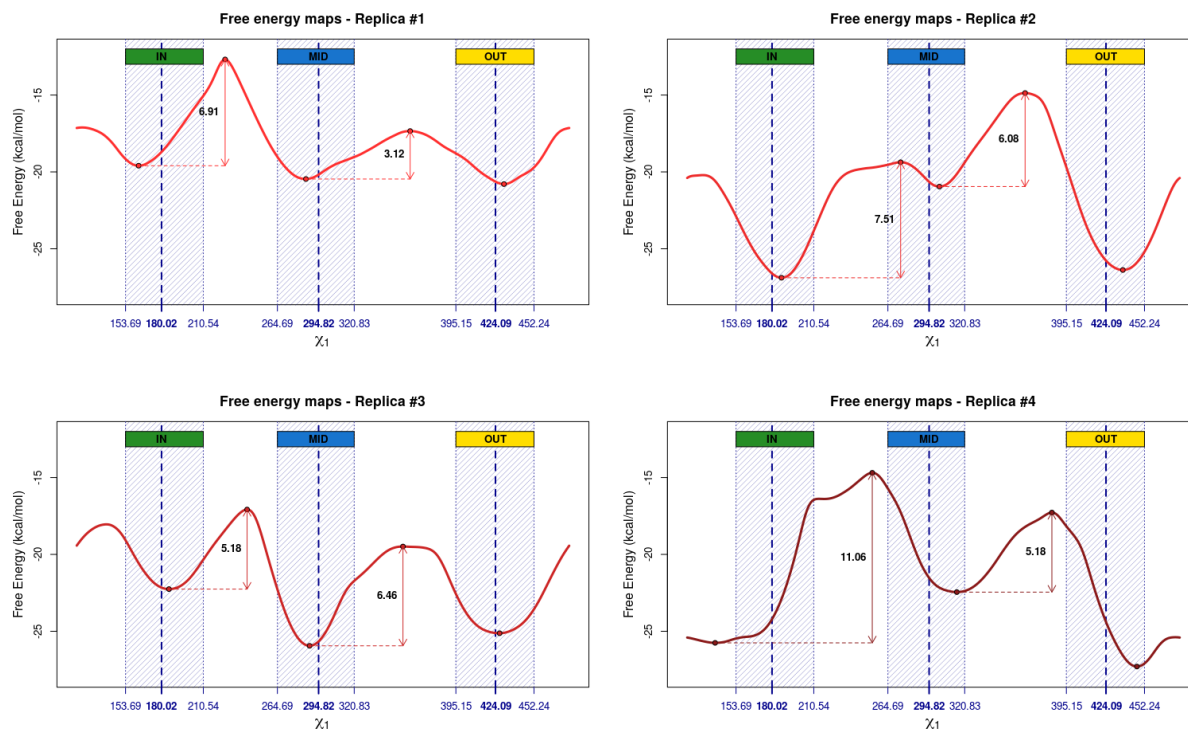
Afin d'obtenir des informations sur les différentes barrières énergétiques séparant les différentes conformations de la Tyrosine 10, des dynamiques moléculaires à biais adaptatif ont été réalisées avec le logiciel Amber18.

L'objectif principal a été l'obtention du profil énergétique des mouvements de la tyrosine 10, pour chercher à déterminer la séquence des événements entraînant la déprotonation de la tyrosine, puis l'activation du glutathion. Pour cela, la première étape a été l'identification d'une ou de plusieurs variables collectives, en prenant comme point de départ la structure de la P28GST obtenue à l'issue de l'équilibrage lors des dynamiques moléculaires classiques et les paramètres recommandés pour la réalisation d'une ABMD. Les deux conditions étudiées ont été l'angle dièdre Chi1, à l'origine des différentes conformations de la tyrosine, ainsi qu'une combinaison des angles dièdres Chi1 et Chi2.

#### **> Variable collective - angle dièdre Chi1 :**

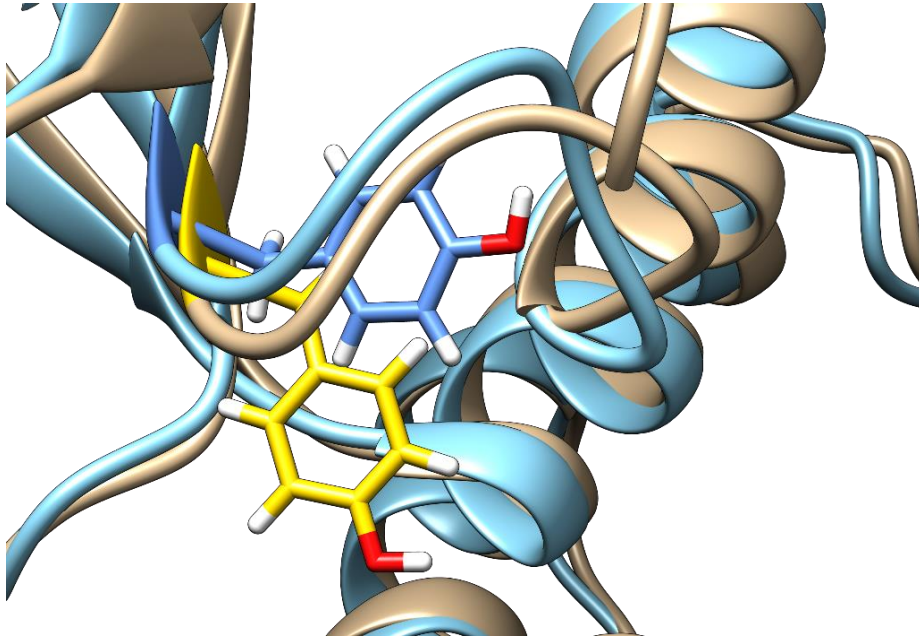
La pertinence de l'utilisation de l'angle dièdre Chi1 en tant que variable collective lors de la réalisation de simulations d'ABMD a été évaluée sur 4 répliques de 150ns, avec un pas de dépôt des biais énergétiques fixé à 100 ps et une pseudo-température fixée à 10.000 K. L'angle dièdre Chi1 a été défini sur une gamme de valeurs allant de  $-\pi$  à  $\pi$  rad, avec un pas d'échantillonnage fixé à 0.2.

Au sein de ces simulations d'ABMD, la variable collective Chi1 semble avoir été explorée efficacement. Toutefois, les profils d'énergie libre obtenus à la suite des 4 répliques ne se ressemblent pas (Figure 37). En effet, les barrières énergétiques observées ne sont pas identiques entre toutes les répliques, et les systèmes ne semblent pas avoir convergé.



**Figure 37.** Profils d'énergie libre obtenus lors des 4 répliques de 150 ns d'ABMD ayant pour variable collective l'angle dièdre  $\chi_1$ .

De plus, après l'observation des trajectoires à l'aide de VMD, il est possible de voir que les angles dièdres  $\chi_1$  observés ne correspondent pas forcément à la conformation attendue. Par exemple, au sein de la réplique #2, il est possible de voir des résidus tyrosine dans leur conformation IN, mais possédant des angles dièdres  $\chi_1$  correspondant à ceux observables dans la conformation OUT (Figure 38). De plus, au sein de toutes les répliques, les simulations n'ont exploré que partiellement les conformations IN et OUT, et très peu la conformation MID.



*Figure 38 Représentation de la tyrosine 10 en début de simulation (en beige et jaune - angle dièdre de 191° correspondant à la conformation IN) et au bout de 25 ns de simulation (en bleu - angle dièdre de 109° s'approchant des angles observés dans la conformation OUT) dans la réplique #2 de simulation d'ABMD.*

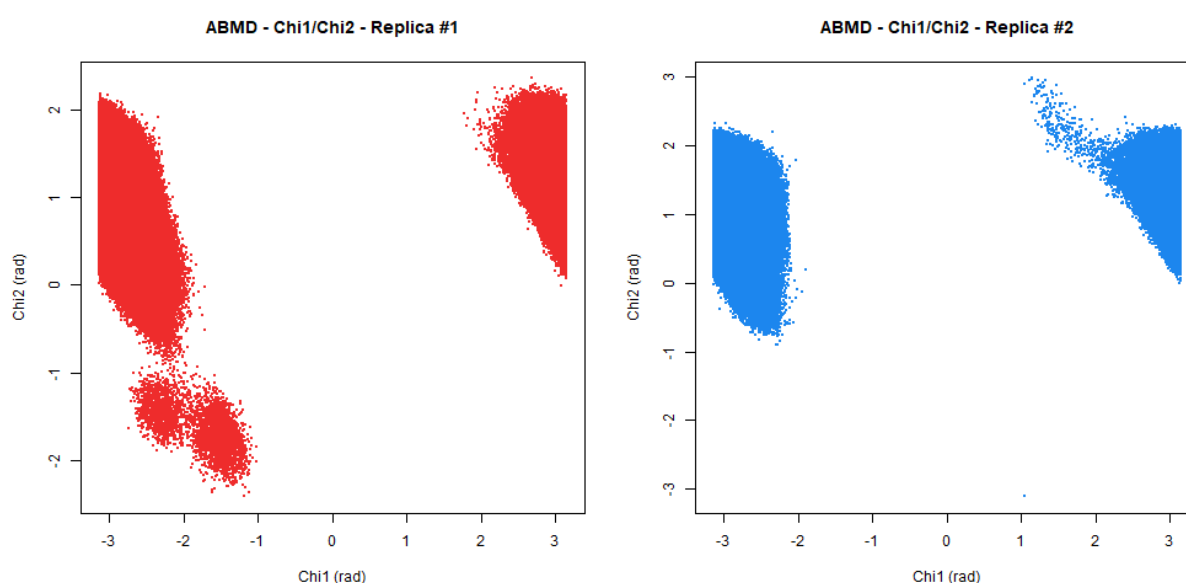
Trois possibilités peuvent expliquer ce phénomène :

- Soit l'angle dièdre  $\text{Chi1}$  est une mauvaise variable collective, auquel cas il est nécessaire de trouver une nouvelle variable collective permettant d'étudier les barrières énergétiques entre les différentes conformations.
- Soit la simulation n'a pas eu le temps de converger dans le temps de simulation alloué.
- Soit les paramètres par défaut du système ne sont pas suffisants, où le pas de dépôt est trop élevé et/ou les biais énergétiques amenés sont trop faibles.

### > Nouvelle variable collective - angles dièdres Chi1/Chi2 :

La pertinence de l'utilisation de la combinaison des variables collectives Chi1 et Chi2 lors de la réalisation de simulations d'ABMD a été évaluée sur deux répliques de 200 ns, avec un pas de dépôt des biais énergétiques fixé à 100 ps et une pseudo-température fixée à 10.000 K. Les variables collectives ont été définies sur une gamme de valeurs allant de  $-\pi$  à  $\pi$  rad, avec un pas d'échantillonnage fixé à 0.2.

Ici, l'exploration du système n'a pas été réalisée correctement, car tous les états n'ont pas pu être étudiés correctement, la simulation ayant stagné dans la conformation IN (Figure 39).

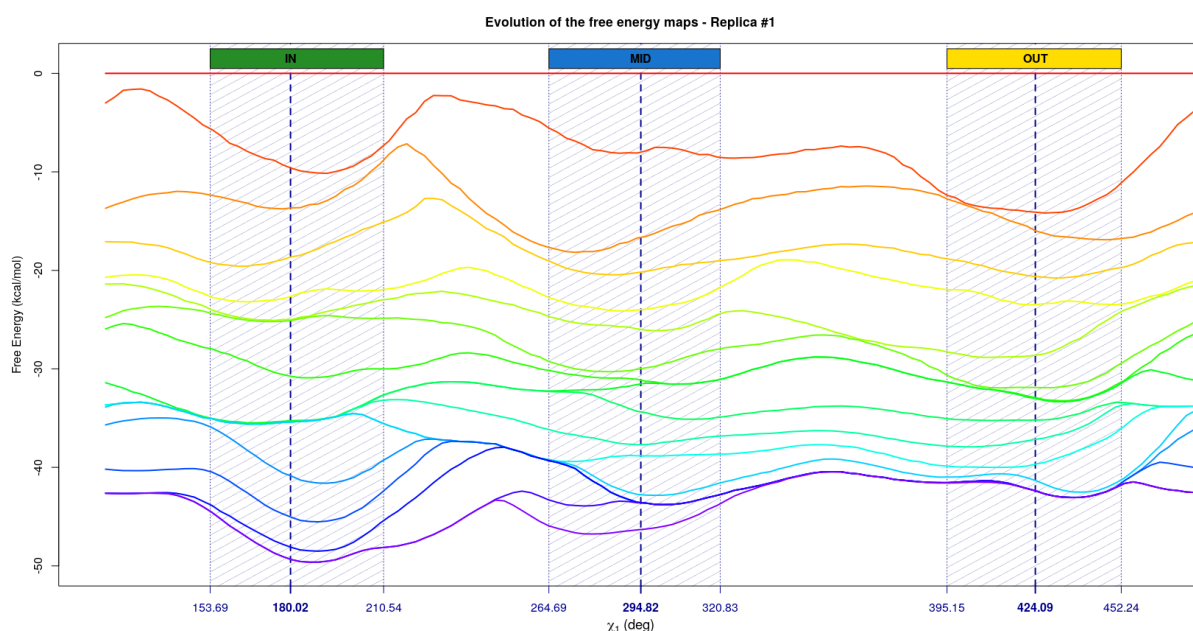


**Figure 39.** Exploration de l'espace des deux variables collectives Chi1 et Chi2 au sein de deux répliques d'ABMD de 200ns.

Au vu de l'utilisation des paramètres conseillés par défaut lors de la réalisation de simulations d'ABMD, et du temps ayant été alloué à la réalisation de ces répliques, la combinaison des variables collectives Chi1 et Chi2 n'est pas pertinente, car elle ne permet pas de décrire toutes les conditions à observer et ne permet pas l'obtention d'une convergence dans des temps raisonnables.

### > Prolongation de la simulation d'ABMD & changement de pas de dépôt :

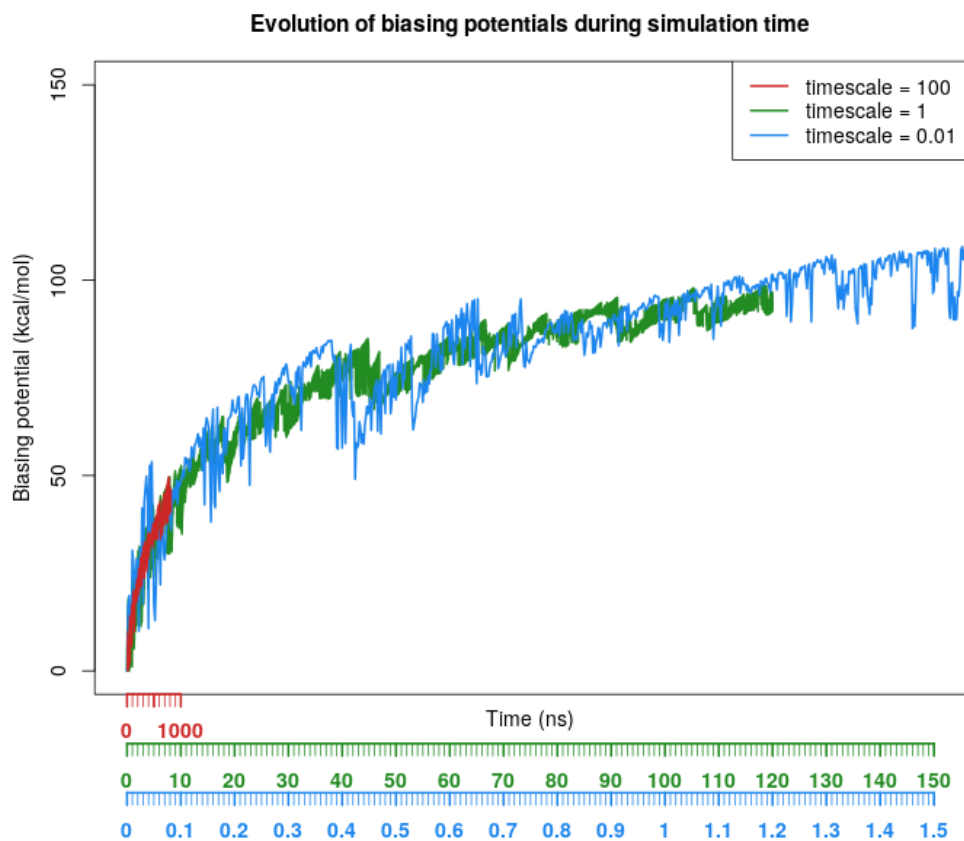
Afin d'évaluer si le temps de simulation alloué est l'une des causes de la non-convergence, la réplique #1 de la simulation d'ABMD basée sur la variable collective  $\chi_1$  a été continuée jusqu'à 823 ns de simulation, pour une vitesse moyenne de 16 ns par jour, et l'évolution du profil d'énergie libre dans le temps a été étudiée (Figure 40). Nous pouvons voir sur ce graphique que le profil d'énergie libre ne cesse d'évoluer toutes les 50 ns de simulation, indiquant que la convergence n'est toujours pas atteinte après les 823 ns de simulation, ayant demandé plus d'un mois et demi de simulation.



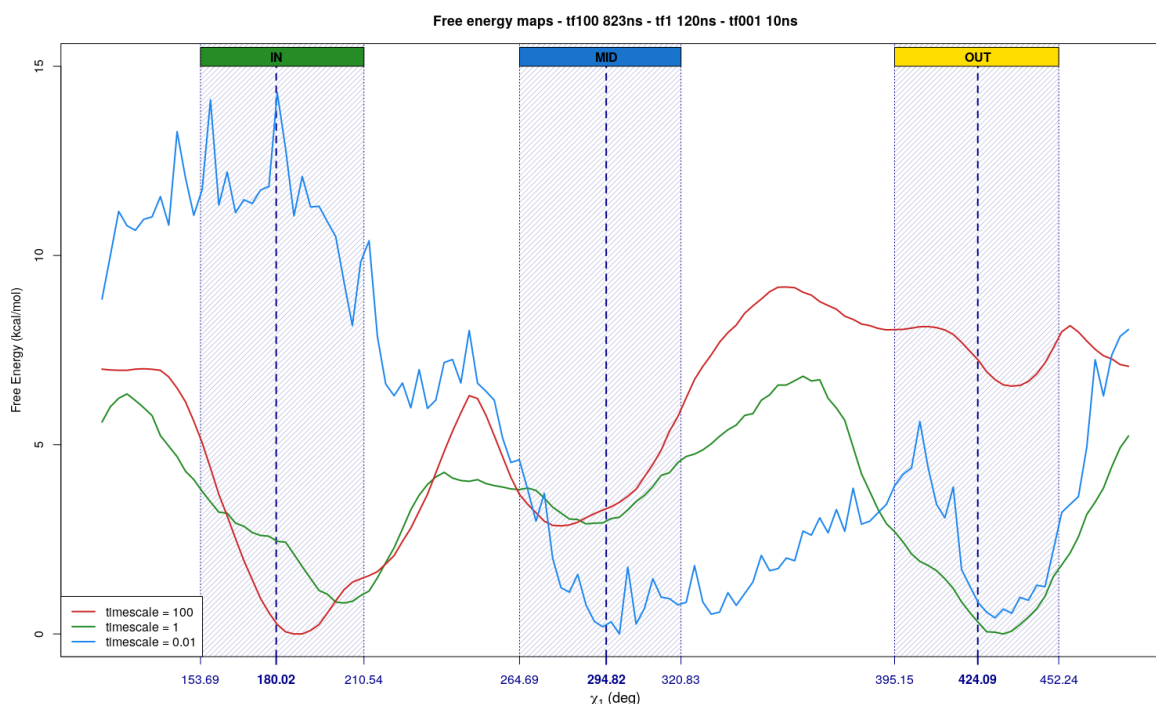
*Figure 40. Évolution du profil d'énergie libre dans la simulation d'ABMD de 823 ns. Une ligne correspond à 50ns de simulation, allant du rouge pour le début de la simulation au violet pour la fin de la simulation.*

Pour accélérer la simulation d'ABMD, des variations du pas de dépôt des biais énergétiques ont été réalisées au sein de d'une réplique de 120 ns avec un pas de 1 ps et d'une réplique de 10 ns avec un pas de 0.01 ps. En étudiant l'évolution de la magnitude des potentiels de biais lors des simulations d'ABMD (Figure 41), il est possible de voir que les biais de la simulation de 823 ns sont très loin de leur magnitude maximale, en comparaison aux données retrouvées dans les simulations possédant un pas de dépôt bien plus faible. De plus, un très faible pas de

dépôt permet d'atteindre rapidement la magnitude maximale des potentiels de biais, mais ne permet pas d'échantillonner correctement l'espace des variables collectives (Figure 42).



*Figure 41. Évolution de la magnitude des potentiels de biais déposés lors des simulations d'ABMD de 823 ns avec un pas de dépôt de 100 ps (en rouge), de 120 ns avec un pas de dépôt de 1 ps (en vert) et de 10 ns avec un pas de dépôt de 0.01 ps.*



**Figure 42.** Profils d'énergie libre en fonction de la variable collective  $\chi_1$  obtenus pour les simulations d'ABMD de 823 ns avec un pas de dépôt de 100 ps (en rouge), de 120 ns avec un pas de dépôt de 1 ps (en vert) et de 10 ns avec un pas de dépôt de 0.01 ps.

La seule modification du dépôt de biais n'est donc pas suffisante pour atteindre la convergence avec un profil d'énergie libre de qualité pour la variable collective  $\chi_1$ . Il est donc nécessaire d'évaluer l'impact du changement de la pseudo-température du système.

### > Augmentation de la pseudo-température :

D'après la théorie de l'ABMD, plus la pseudo-température de simulation est petite, plus douce et lente sera la convergence (139). Pour augmenter la vitesse de convergence, il est donc nécessaire d'augmenter la pseudo-température. Pour évaluer l'impact du changement de la pseudo-température, de courtes simulations de 20 ns d'ABMD ont été réalisées à des pseudo-températures de 30.000 K et 100.000 K en utilisant la variable collective  $\chi_1$  et un pas de dépôt des biais énergétiques de 100 ns.



Toutefois, les résultats obtenus ne sont pas concluants. Lors de la simulation comprenant la pseudo-température de 100.000 K, le logiciel a planté à de nombreuses reprises, rendant impossible la bonne réalisation de la simulation. De plus, lors des simulations comprenant la pseudo-température de 30.000 K, le système n'a que très peu évolué sur le temps de simulation, où la Tyrosine 10 n'est retrouvée que dans sa conformation IN par étude visuelle des trajectoires avec VMD.

La variation de la pseudo-température seule n'est donc pas suffisante, celle-ci n'améliorant pas la convergence du système comme espéré.

### > **Transition à PLUMED :**

Les modifications testées n'ont pas apporté de solution pour la convergence. En parallèle, la "simplification" de la métadynamique proposée par l'ABMD est difficilement exploitable via Amber18, l'algorithme fonctionnant comme une boîte noire, les fichiers principaux étant sauvegardés dans le format NetCDF compressé. Enfin, il n'existe que peu de documentations ou de retours d'expériences disponibles sur internet.

Au vu des options proposées par Amber18 lors de la réalisation d'une ABMD, et de l'impossibilité de tester d'autres méthodes comme l'algorithme de sélection multiple walkers ou l'échantillonnage parapluie (*umbrella sampling*) en raison des ressources de calcul à disposition, il a été décidé de changer de méthode pour passer à la méthode de la métadynamique bien tempérée.

Cette méthode proposée par le logiciel PLUMED, s'ajoutant au module de dynamique moléculaire d'Amber18, a permis le meilleur compromis entre la réutilisation des connaissances existantes du logiciel Amber18 et la réalisation de nouvelles simulations basées sur des algorithmes spécifiques. PLUMED est également plus transparent sur la réalisation de ses simulations de métadynamique, notamment au travers des gaussiennes ajoutées et leur

hauteur en début de simulation, et propose de nombreuses ressources documentaires au travers de didacticiels et d'exercices.

### ***b. Métadynamique avec PLUMED***

Avant d'utiliser les fonctionnalités de PLUMED, les principaux paramètres ont été convertis de l'ABMD d'Amber18 à la métadynamique de PLUMED. L'unité d'énergie libre passant du kcal/mol au kJ/mol, il est nécessaire de réaliser une conversion des résultats de PLUMED en les multipliant par 0.239. En utilisant la formule du facteur de biais de la figure 36, en se basant sur la pseudo-température de 10.000 K utilisée dans les simulations d'ABMD, le facteur de biais converti est d'environ 33.3, arrondi à 30.

La définition du pas de dépôt des biais reste identique, passant du paramètre *timescale* d'Amber18 au paramètre PACE pour PLUMED. La définition du pas d'échantillonnage des variables reste également identique, passant du paramètre résolution au paramètre SIGMA. Enfin, la hauteur des premières gaussiennes est une option spécifique à la métadynamique de PLUMED.

#### **> Premières simulations avec les paramètres utilisés dans l'ABMD :**

De premières simulations de métadynamique ont été menées avec les mêmes conditions expérimentales que les simulations d'ABMD réalisées sous Amber18, avec le même point de départ, la même variable collective Chi1 et les mêmes paramètres. Toutefois, les problèmes observés avec la méthode ABMD se sont présentés à nouveau. Plusieurs tests ont donc été mis en place afin :

- D'évaluer une nouvelle fois la pertinence de l'angle dièdre Chi1 en tant que variable collective, et d'évaluer d'autres types de grandeurs pour obtenir une alternative à cet angle,
- D'évaluer le changement du facteur de biais sur l'exploration de l'espace des variables collectives,
- D'optimiser la hauteur des premières gaussiennes.

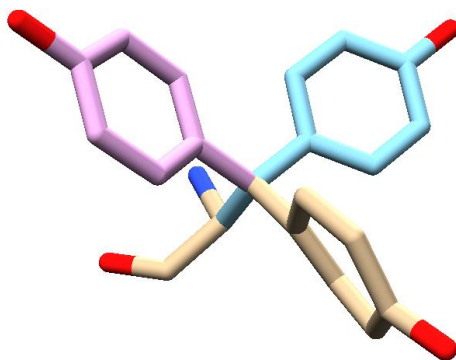
Les paramètres ont été optimisés au travers de courtes simulations de métadynamiques, allant de 10 à 50 ns. Les valeurs pour lesquelles la simulation se comportait le mieux, sans dénaturation de structures secondaires et/ou tertiaires et sans problèmes d'exploration des différentes variables collectives retenues, sont celles décrites dans les paragraphes suivants.

### > **Évaluation de nouvelles variables collectives:**

Dans un premier temps, de nouvelles variables collectives ont été étudiées.

Tout d'abord, la distance entre l'atome d'oxygène de la tyrosine 10 et le carbone alpha de l'acide aspartique 104 a été utilisée comme variable collective, et l'angle dièdre Chi1 a été mesuré au cours de la simulation pour générer le profil d'énergie libre en fonction des différentes valeurs de cet angle. L'acide aspartique 104 a été sélectionné car il se situe à distance de la tyrosine sur une structure secondaire ne bougeant que très peu au cours des simulations, et se situe sur l'axe décrit par les atomes d'oxygène des conformations IN et OUT de la tyrosine 10 dans la cristallographie. Toutefois, cette simulation n'a pas permis de séparer correctement la conformation MID de la conformation OUT. De plus, un petit excès dans les bornes minimales et maximales entraîne soit l'explosion, soit l'implosion du système, ce qui rend cette variable collective très difficile à utiliser.

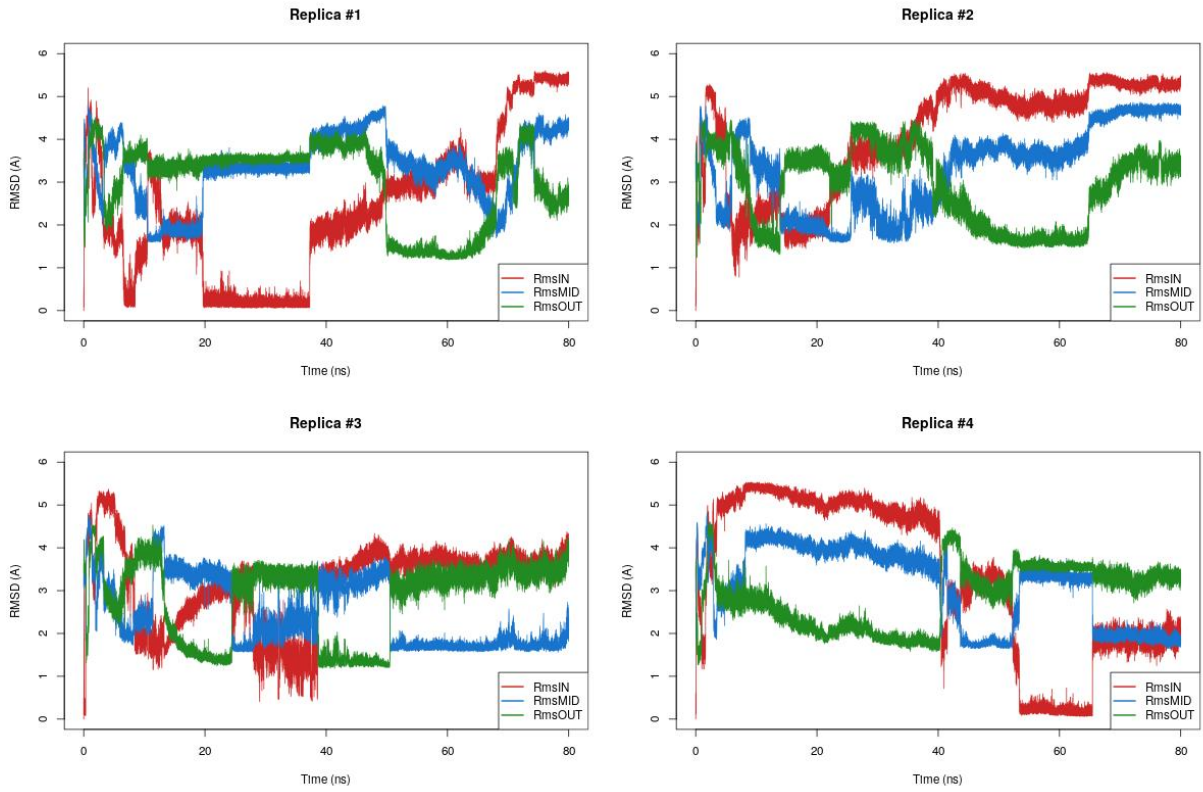
Dans un second temps, le RMSD des trois configurations IN, MID et OUT ont également été utilisés conjointement comme variables collectives, tout en suivant également l'angle dièdre Chi1 pour la génération du profil d'énergie libre. Les différents RMSD ont été pensés pour permettre une exploration plus directe des trois conformations, en évitant les biais amenés par les angles dièdres Chi1 ne correspondant pas aux conformations observées au sein des trajectoires. Pour cela, trois structures de référence ont été générées en prenant uniquement les coordonnées de la tyrosine 10 dans la structure en entrée, puis en générant les différents rotamères avec UCSF Chimera (Figure 43).



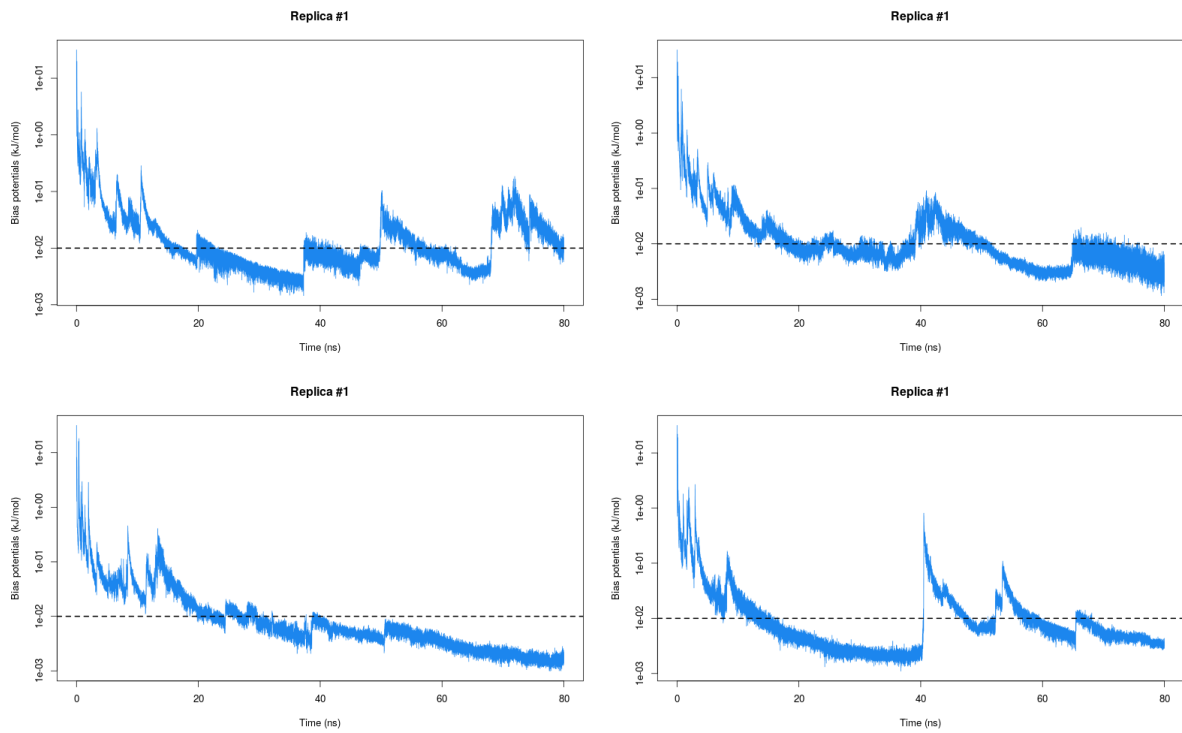
*Figure 43. Structures de référence pour la tyrosine 10 dans sa conformation IN (en beige), MID (en bleu) ou OUT (en rose) utilisées pour le calcul des variables collectives définies par les RMSD de ces structures avec la conformation de la tyrosine 10 au cours de la simulation.*

Pour optimiser la convergence du système, le facteur de biais a été diminué à 20, tout en augmentant drastiquement la hauteur des premières gaussiennes à 30 kJ/mol, soit environ 7.2 kcal/mol, correspondant approximativement aux barrières énergétiques observées jusqu'ici. Le pas de dépôt des biais est resté inchangé, et fixé à 100 ps.

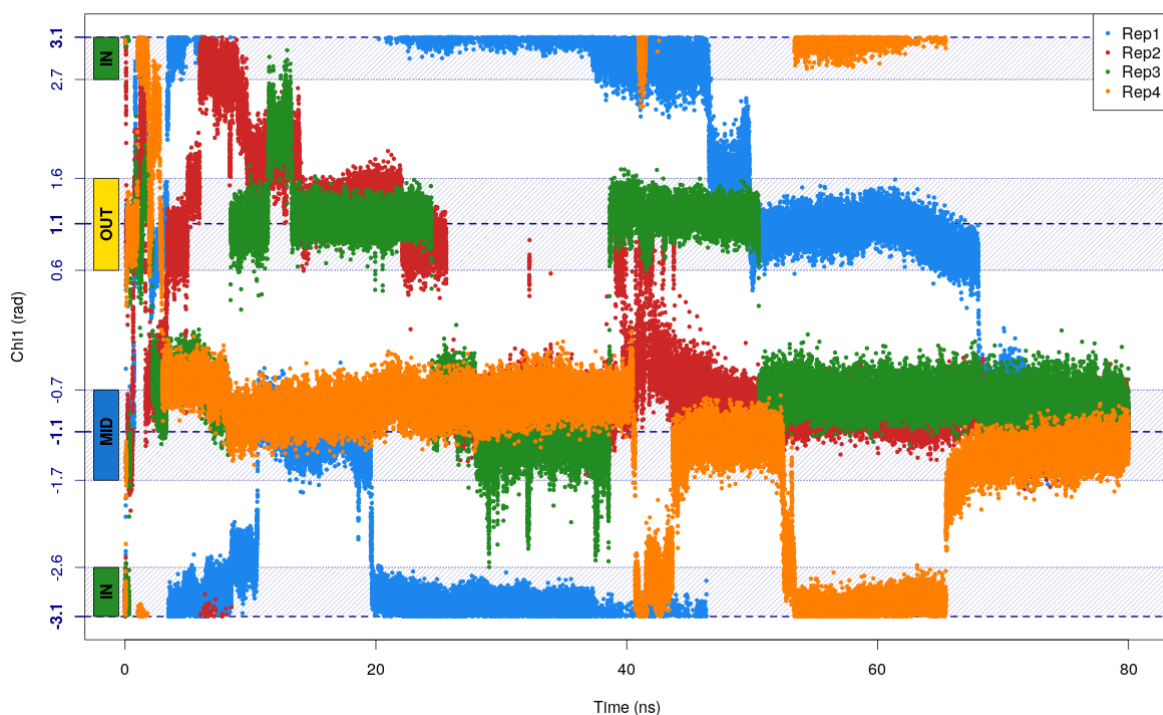
Des simulations de métadynamique avec Amber18 lancées avec les fichiers de paramètres réalisés grâce à PLUMED ont été menées en 4 répliques de 80 ns utilisant les variables collectives RMSD, définies avec un pas d'échantillonnage de ces variables collectives de 0.1 et de valeurs minimales et maximales de 0 et 15 Å. L'évolution des différentes variables collectives a été étudiée (Figure 44). Ici, les variables collectives évoluent rapidement en début de simulation, puis ralentissent progressivement leur exploration en parallèle de la diminution drastique des potentiels de biais ajoutés (Figure 45). Ces variables s'alternent au cours des 20 premières nanosecondes, où les potentiels de biais déposés sont supérieurs à une limite arbitraire fixée à 0.01 kJ/mol, indiquant que les simulations explorent bien les différentes conformations de la tyrosine 10. Toutefois, en suivant l'angle dièdre Chi1 au cours de la simulation, il est possible de voir que, même au début des simulations, toutes les conformations ne sont pas explorées (Figure 46). De plus, le RMSD ne correspond pas forcément à la bonne conformation et aux bons angles dièdres Chi1, du fait du mouvement du squelette carboné au cours de la simulation, pouvant fausser le calcul du RMSD.



*Figure 44. Évolution des différentes variables collectives issues du RMSD de la tyrosine 10 comparé aux conformations de référence dans les 4 répliques de métadynamique réalisées.*



*Figure 45. Évolution des potentiels de biais au cours des 4 répliques de métadynamique réalisées.*



*Figure 46. Évolution de l'angle dièdre Chi1 dans les 4 répliques de métadynamique réalisées en utilisant les différentes variables collectives issues du RMSD de la tyrosine 10 comparé aux conformations de référence.*

Au vu des potentiels de biais ajoutés, majoritairement inférieurs à 0.01 kJ/mol donc négligeables par rapport aux potentiels de biais ajoutés en début de simulation, il est possible de ne réaliser que des simulations de 20 ns. Toutefois, l'échantillonnage des angles Chi1 n'est pas parfait dans chaque simulation. Comme la durée des répliques a été divisée par 4, il a été décidé de réaliser au minimum 20 répliques pour s'assurer de la significativité statistique, afin d'obtenir un profil énergétique moyen. De plus, pour favoriser l'échantillonnage en début de simulation, différents points de départ ont été utilisés, correspondant aux différentes conformations de la tyrosine catalytique.

### > Préparation des points de départ MID & OUT :

Avant la réalisation de ces simulations de métadynamique, de nouveaux points de départ ont été générés.

A partir de la structure obtenue à la fin de la phase d'équilibrage des dynamiques moléculaires précédentes, possédant la tyrosine 10 dans sa conformation IN, deux nouvelles structures possédant la tyrosine 10 dans une conformation MID ou OUT ont été générées avec UCSF Chimera via l'outil *Rotamers*. Puis le même protocole que celui utilisé lors de la préparation des simulations de dynamique moléculaire classique a été appliqué à ces deux structures, comprenant l'initialisation du système, la minimisation de l'énergie potentielle, le chauffage puis l'équilibrage. La préparation du point de départ avec la tyrosine 10 dans sa conformation OUT s'est déroulée avec succès, mais la conformation MID n'a pas pu être obtenue. En effet, lors de la phase d'équilibrage, la tyrosine 10 s'est automatiquement déplacée de la conformation MID à la conformation OUT. Ce phénomène, s'étant reproduit dans 3 essais différents, empêche l'utilisation d'un point de départ MID pour les métadynamiques, et peut indiquer soit une faible barrière énergétique entre la conformation MID et la conformation OUT, soit un puits énergétique plus faible au niveau de la conformation OUT favorisant cette dernière.

### > Variables collectives RMSD - Profil énergétique moyen :

Afin d'obtenir un profil d'énergie libre moyen en fonction de l'angle dièdre  $\text{Chi}_1$ , 40 répliques de 20 ns de métadynamique ont été réalisées en utilisant les différentes variables collectives issues du RMSD de la tyrosine 10 comparé aux conformations de référence, une hauteur des premières gaussiennes fixée à 30 kJ/mol, un facteur de biais fixé à 20, et un pas de dépôt des biais fixé à 100 ps. L'angle dièdre  $\text{Chi}_1$  a été calculé au cours de la simulation, en parallèle des variables collectives. 20 répliques ont débuté avec la structure possédant la tyrosine 10 dans sa conformation IN, et 20 autres répliques ont débuté avec la structure possédant la tyrosine 10 en conformation OUT. A partir du profil d'énergie de chaque

réplique, un profil d'énergie libre moyen a pu être calculé en réalisant la moyenne et son intervalle de confiance à 95% bilatéral avec RStudio (Figure 47).

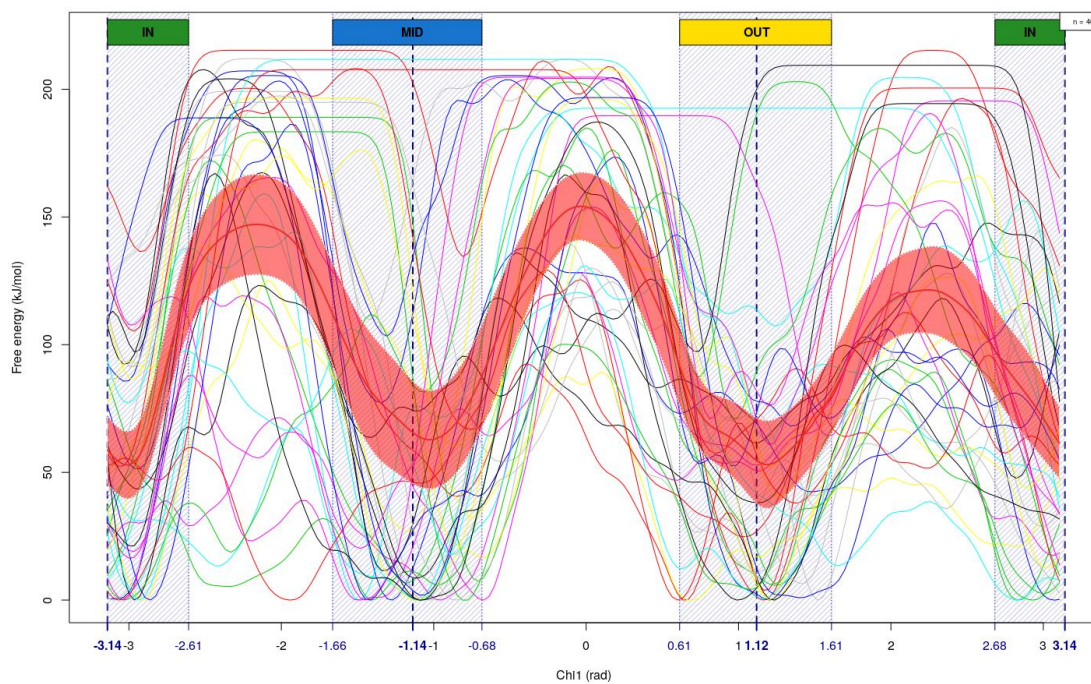
$$I_c = \left[ \bar{x} - t_\alpha \frac{s}{\sqrt{n}} ; \bar{x} + t_\alpha \frac{s}{\sqrt{n}} \right]$$

*Figure 47. Formule de l'intervalle de confiance IC, calculé autour d'une moyenne observée  $\bar{x}$ , possédant un écart-type observé  $s$ , sur un échantillon de taille  $n$ , et avec  $t_\alpha$  la valeur du 97.5 percentile de la distribution normale standard arrondie à 1.96.*

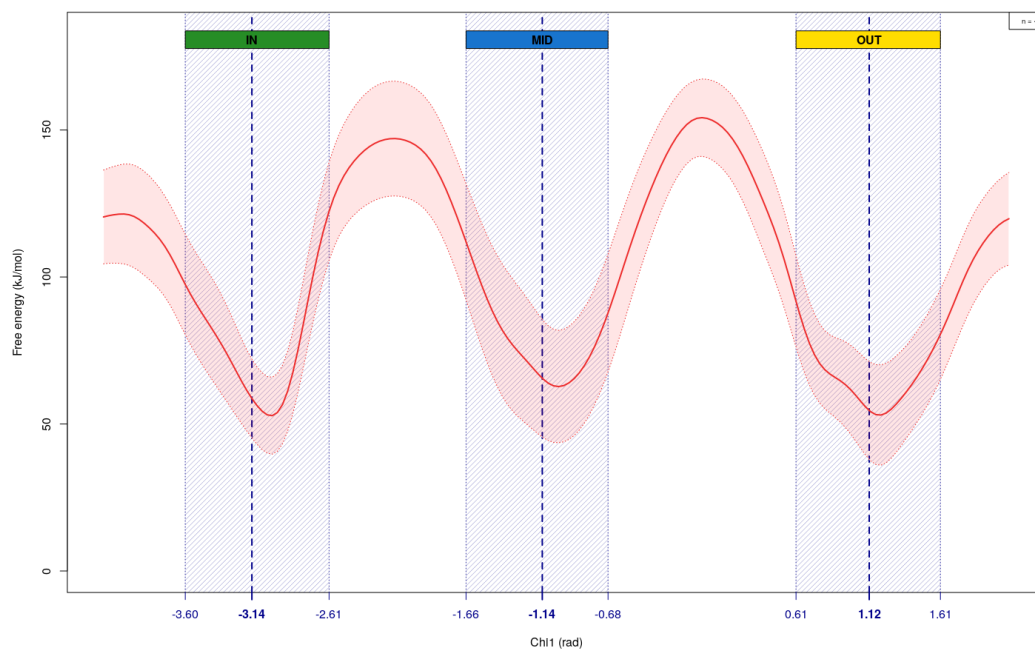
Le profil énergétique moyen obtenu à l'issue des 40 répliques de métadynamique met en évidence la présence de barrières énergétiques de 75 à 100 kJ/mol (soit 18 à 23.9 kcal/mol) entre chaque conformation de la tyrosine 10 (figures 48 et 49). Les puits énergétiques présents au niveau des différentes conformations de la tyrosine 10 sont retrouvés au même niveau, avec des barrières énergétiques du même ordre de grandeur les séparant. Il est donc nécessaire de se demander si la réalisation de la moyenne n'entraîne pas un lissage des profils énergétiques. En effet, les 40 profils d'énergie libre obtenus ne sont pas similaires, avec des zones de l'espace de l'angle dièdre Chi1 non exploré (Figure 50).

La non-exploration de certains angles dièdre Chi1 peut être causée par la complexité de la variable collective définie par les différents RMSD, mettant en jeu trois paramètres à explorer. Il est donc nécessaire de changer la variable collective pour une grandeur plus simple à évaluer et explorer, posant également moins de problème au niveau de l'interprétation des résultats : l'angle dièdre Chi1. Enfin, le glutathion ne reste pas de manière similaire au sein de son site de liaison dans les 40 répliques, souvent dès le début de la simulation. Il est donc également nécessaire de retirer au préalable les molécules de glutathion présentes dans les structures de départ pour simuler les mouvements de la tyrosine 10 sans glutathion ou d'imposer des contraintes de distance entre les molécules de glutathion et certains acides aminés pour restreindre ses mouvements.

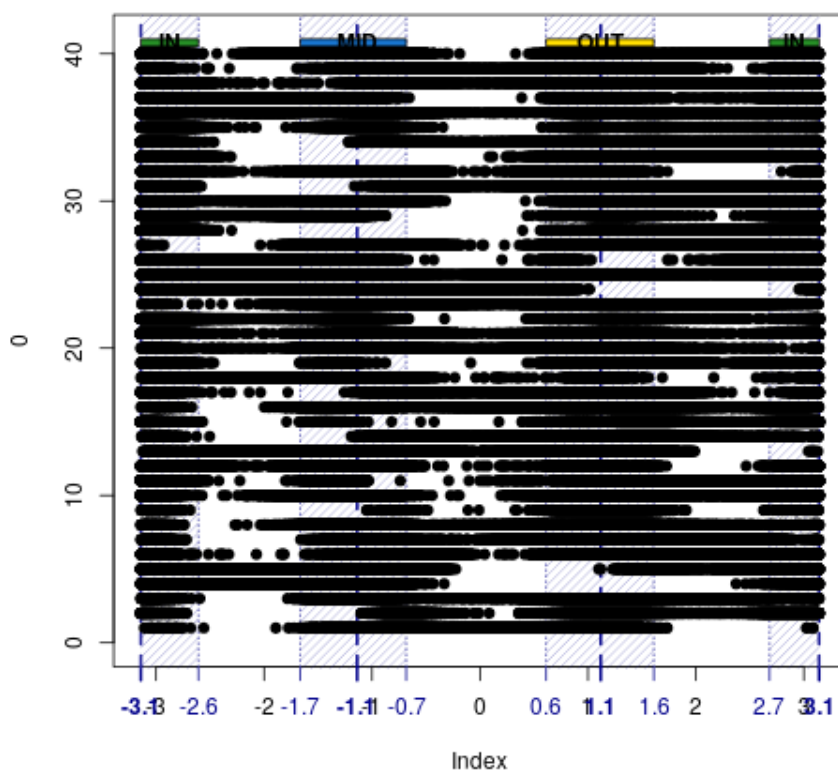




**Figure 48.** Profils d'énergie libre (en kJ/mol) des 40 répliques de métadynamique réalisées avec les variables collectives RMSD (en traits fins) et profil d'énergie libre moyen (en trait épais rouge) avec son intervalle de confiance 95% bilatéral (aire rouge).



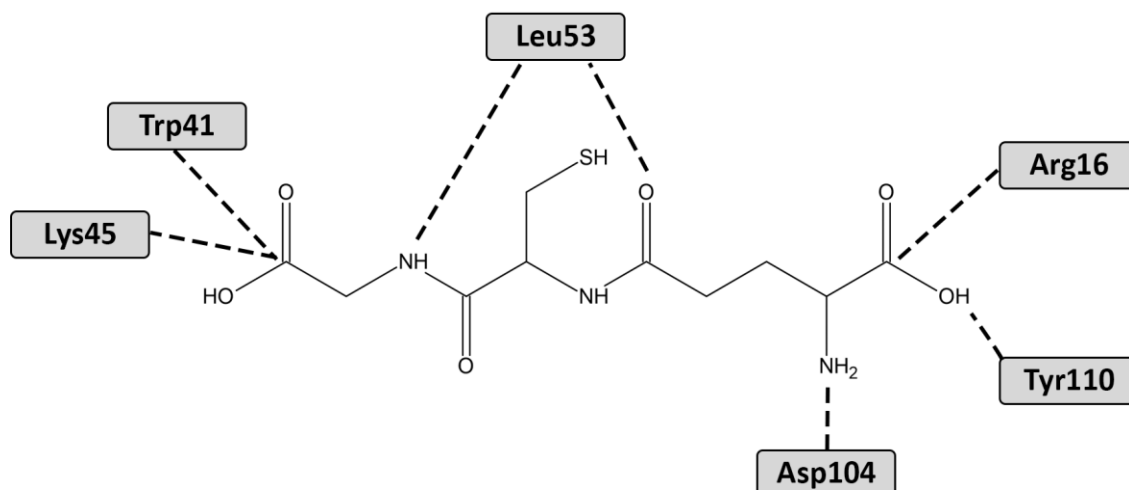
**Figure 49.** Profil d'énergie libre moyen (en kJ/mol) des 40 répliques de métadynamique réalisées avec les variables collectives RMSD avec son intervalle de confiance 95% bilatéral (aire rouge).



*Figure 50. Différents angles dièdres Chi1 (en rad) dans les 40 répliques de métadynamique réalisées avec les variables collectives RMSD.*

**> Variable collective Chi1 - Profil énergétique moyen :**

De nouvelles simulations de métadynamique ont ensuite été réalisées avec l'angle dièdre Chi1 directement comme variable collective. De plus, pour pouvoir évaluer l'impact du glutathion sur le profil d'énergie libre, celui-ci a été restreint au niveau de 7 distances, comprenant les acides aminés Arg16, Trp41, Lys45, Leu53, et Tyr110 du monomère A, et Asp104 du monomère B (Figure 51).



*Figure 51. Contraintes imposées entre le glutathion et certains acides aminés de la P28GST lors de certaines simulations de métadynamique*

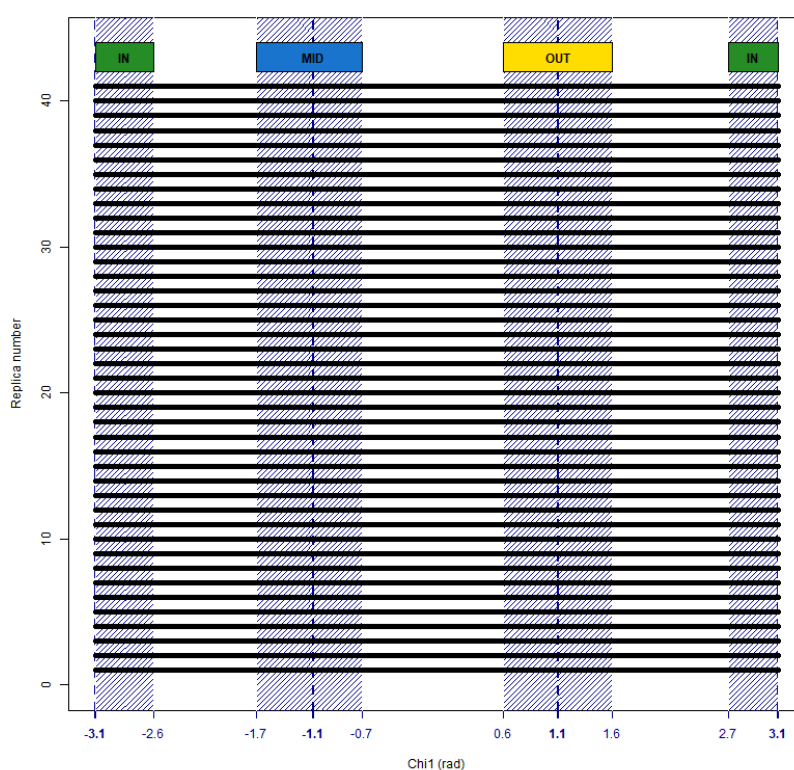
Dans ces conditions, 41 répliques de métadynamique utilisant l'angle dièdre Chi1 comme variable collective ont été réalisées à partir de la tyrosine en conformation IN. Comme précédemment, la hauteur des premières gaussiennes a été fixée à 30 kJ/mol, le facteur de biais a été fixé à 20, et le pas de dépôt des biais a été fixé à 100 ps. A partir du profil d'énergie de chaque réplique, un profil d'énergie libre moyen a pu être calculé en réalisant la moyenne et son intervalle de confiance à 95% bilatéral avec RStudio.

Ces nouvelles simulations ne présentent pas le même profil énergétique, avec des puits et barrières énergétiques qui ne sont pas égaux, et toutes les configurations de l'angle dièdre Chi1 ont pu être explorées avec succès. Ces dernières sont en outre bien corrélées avec les observations de trajectoires de métadynamiques (Figure 52). De plus, le glutathion est bien maintenu en place en suivant les différentes distances au cours de toutes les simulations.

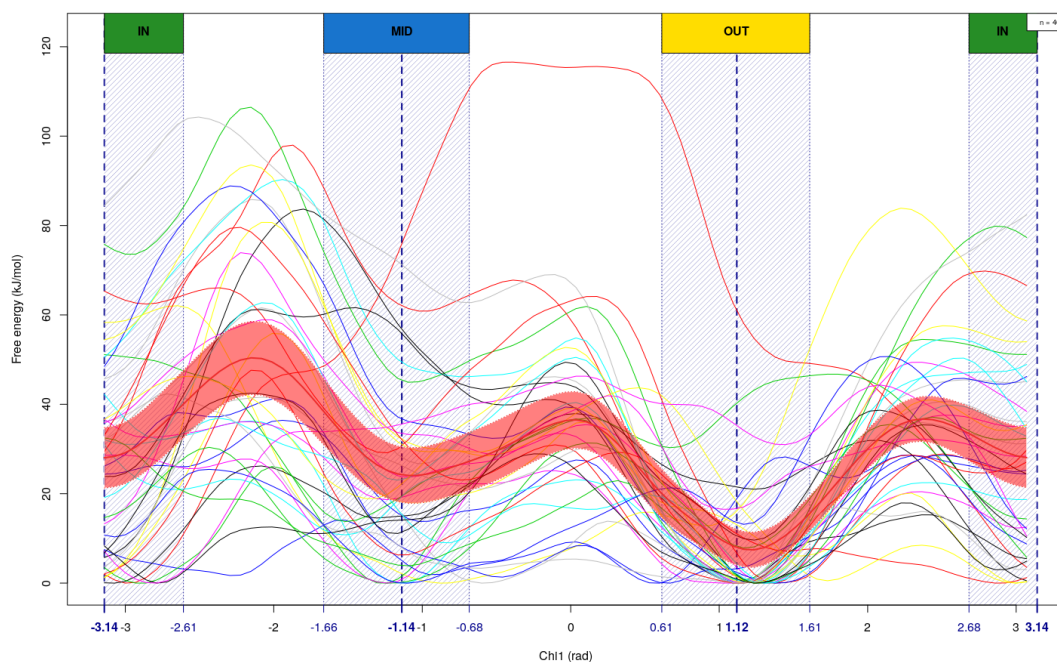
Le profil énergétique moyen observé au cours de ces simulations présente trois minima énergétiques, séparés par des barrières énergétiques de différentes grandeurs (Figures 53 et 54). Le minimum global d'énergie est observé au niveau de la conformation OUT de la tyrosine, et deux puits énergétiques minima locaux au niveau des conformations IN et MID.

Deux types de barrières énergétiques sont observés, conditionnant le passage entre ces trois configurations. De hautes barrières énergétiques sont visualisées au niveau des transitions de IN à MID, de MID à IN, de OUT à MID, et de OUT à IN, avec une différence énergétique de respectivement 21.90 kJ/mol, 25.87 kJ/mol, 29.02 kJ/mol et 28.96 kJ/mol.

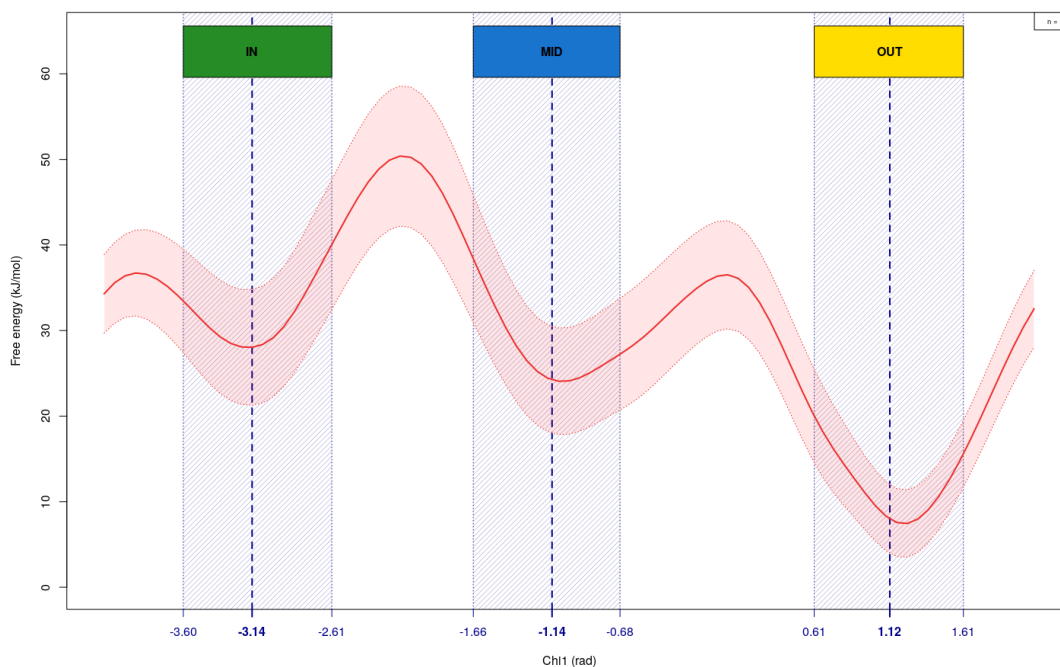
A l’opposé, il existe deux barrières énergétiques de plus faible intensité, de IN à OUT et de MID à OUT, de respectivement 8.51 kJ/mol et 12.55 kJ/mol. Leur existence peuvent expliquer la migration de la conformation MID à la conformation OUT lors de la phase d’équilibrage des différents points de départ de métadynamique, et être un précieux indice sur le fonctionnement mécanistique de la P28GST avec le passage directement de la conformation IN à OUT, sans passer par la conformation MID.



*Figure 52. Différents angles dièdres Chi1 (en rad) observés dans les 41 répliques de métadynamique réalisées avec la variables collective Chi1 démarrant de la structure présentant la tyrosine 10 en conformation IN.*



**Figure 53.** Profils d'énergie libre (en kJ/mol) des 41 répliques de métadynamique réalisées avec la variable collective  $\text{Chi1}$  démarrant de la structure présentant la tyrosine 10 en conformation IN (en traits fins) et profil d'énergie libre moyen (en trait épais rouge) avec son intervalle de confiance 95% bilatéral (aire rouge).



**Figure 54.** Profil d'énergie libre moyen (en kJ/mol) des 41 répliques de métadynamique réalisées avec la variables collective  $\text{Chi1}$  démarrant de la structure présentant la tyrosine 10 en conformation IN, avec son intervalle de confiance 95% bilatéral (aire rouge).

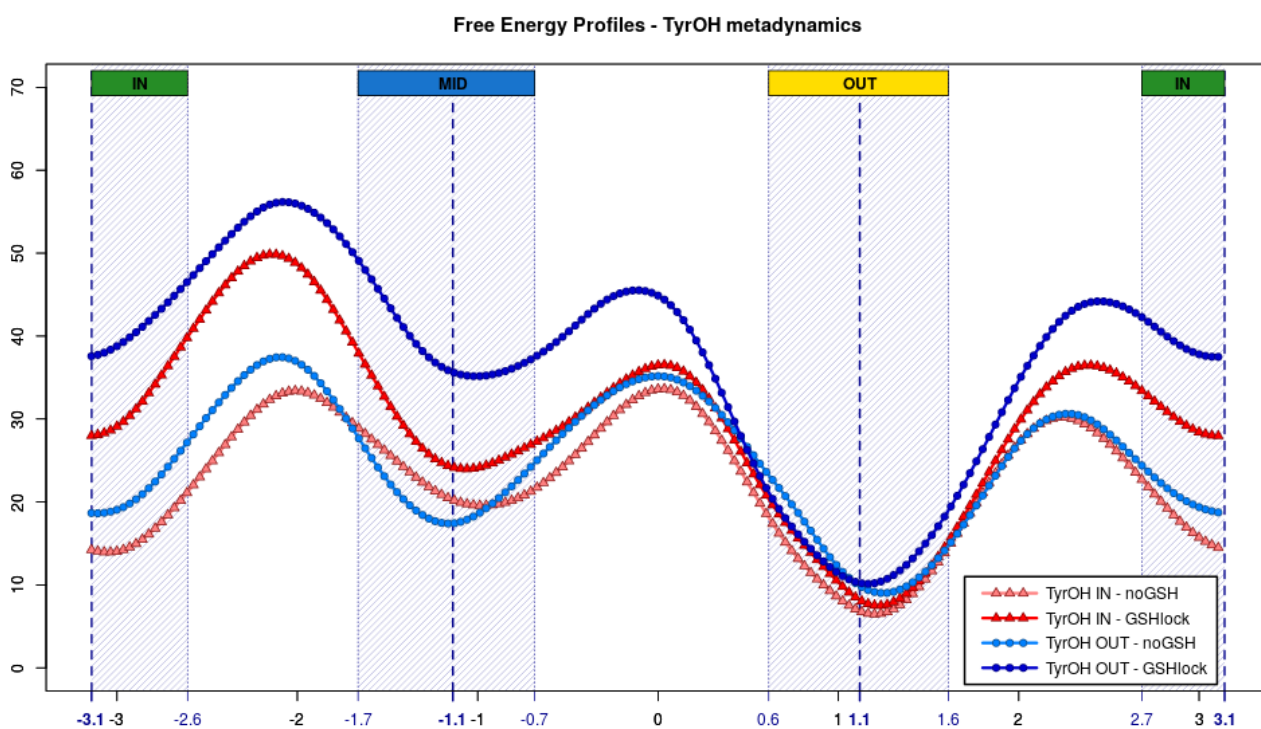
Dans le but de pouvoir comparer les observations entre toutes les combinaisons de présence du glutathion, conformation de départ et état de protonation, un total de 162 nouvelles répliques de 20 ns de métadynamique ont été réalisées (Tableau 5). Les conditions retenues sont les mêmes que pour les précédentes répliques démarrant de la tyrosine 10 protonée dans sa conformation IN avec la présence du glutathion sous contraintes. Dans le cas spécifique des métadynamiques mettant en jeu la tyrosine 10 déprotonée, les conformations de départ ont été au préalable générées par l'éditeur de fichiers de paramètres ParmEd (140), puis ont été préparées en suivant le même protocole que les points de départ protonés.

État de protonation de la tyrosine 10	Glutathion	Conformation de départ	Nombre de répliques
Protoné TyrOH	Pas de glutathion	IN	20
		OUT	20
	Glutathion sous contraintes	IN	41
		OUT	20
Déprotoné TyrO <sup>-</sup>	Pas de glutathion	IN	22
		OUT	20
	Glutathion sous contraintes	IN	30
		OUT	30

*Tableau 5. Présentation des différentes conditions expérimentales réalisées en métadynamique avec la variable collective Chi1 pour l'obtention de profils énergétiques moyens.*

### > Profils énergétiques moyens - TyrOH :

A partir des 101 répliques de métadynamique réalisées avec la variable collective Chi1 dans les conditions expérimentales considérant la tyrosine 10 sous sa forme protonée, 4 profils énergétiques moyens ont été obtenus (Figure 55), et les différentes barrières énergétiques ont été calculées (Tableau 6). Au regard des différents résultats obtenus, lorsque la tyrosine 10 est protonée, son état le plus stable énergétiquement est la conformation OUT, avec la présence de minima absolus au niveau de cette conformation. De plus, des variations significatives peuvent être observées entre les différents minima et barrières énergétiques de ces profils.

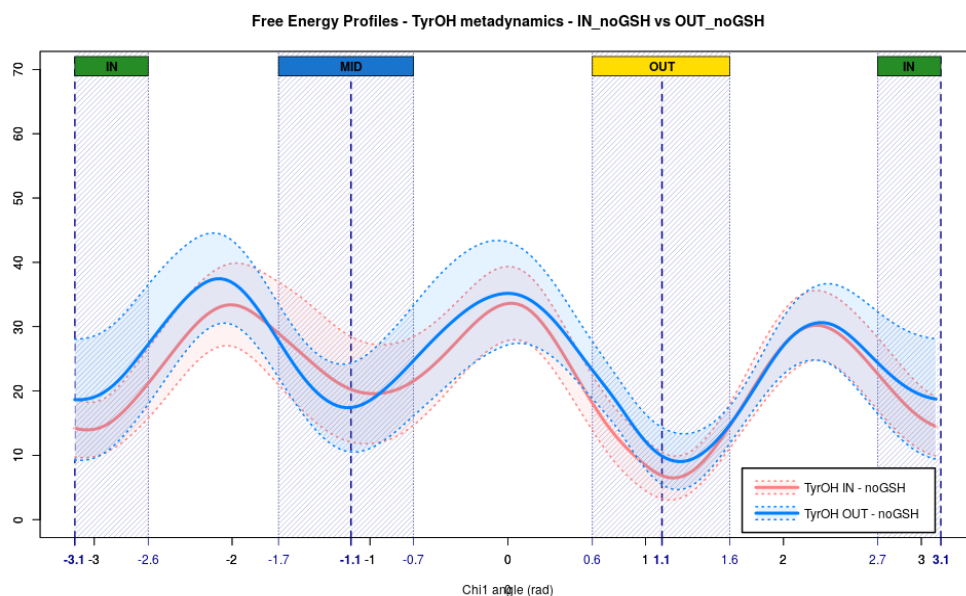


*Figure 55. Profils d'énergie libre moyens (en kJ/mol) en fonction de l'angle dièdre Chi1 (en rad) pour 4 conditions expérimentales mettant en jeu la Tyrosine 10 protonée, pour un minimum de 20 répliques de 20 ns de métadynamique pour chaque condition.*

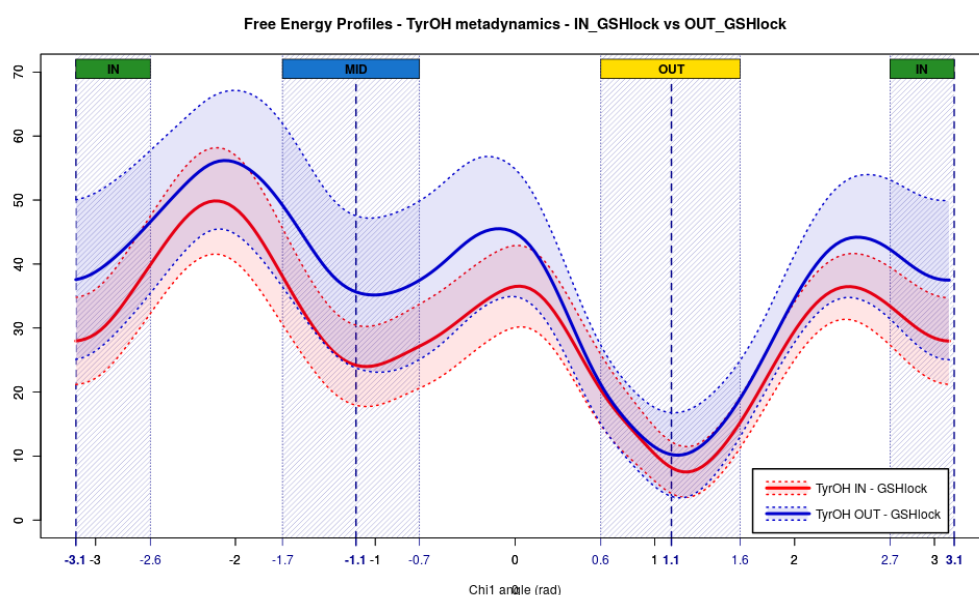
Glutathion	Point de départ	IN → MID	IN → OUT	MID → IN	MID → OUT	OUT → MID	OUT → IN
Glutathion contraint	Tyr10 IN	21.90	8.51	25.87	12.55	28.96	29.02
	Tyr10 OUT	18.68	6.69	21.00	10.35	34.05	35.39
Pas de glutathion	Tyr10 IN	19.45	16.25	13.83	14.06	23.73	27.16
	Tyr10 OUT	18.81	11.99	20.05	17.78	21.60	26.15

*Tableau 6. Barrières énergétiques (en kJ/mol) observées dans les profils d'énergie libre moyens en fonction de l'angle dièdre Chi1 pour 4 conditions expérimentales mettant en jeu la Tyrosine 10 protonée*

En comparant l'impact de la conformation de départ, soit sans glutathion (Figure 56), soit avec glutathion (Figure 57), aucune différence n'est mise en évidence. L'observation de différences sur ces graphiques aurait été le signe d'un problème inhérent aux systèmes initiaux.



**Figure 56.** Profils d'énergie libre moyens (en kJ/mol) et intervalle de confiance à 95% en fonction de l'angle Chi1 (en rad) pour la tyrosine 10 protonée dans ses conformations IN et OUT, sans glutathion



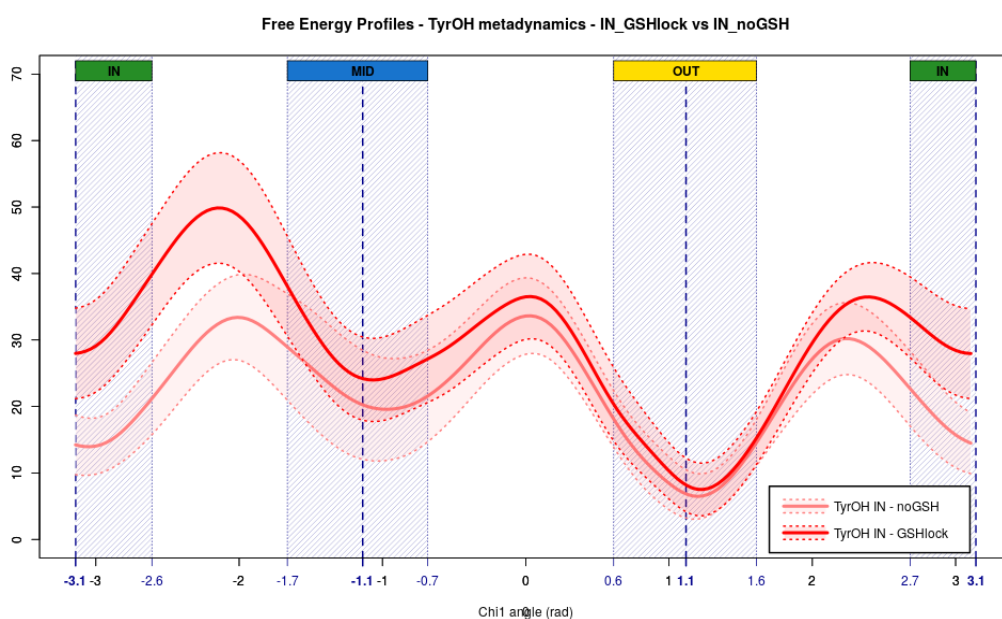
**Figure 57.** Profils d'énergie libre moyens (en kJ/mol) et intervalle de confiance à 95% en fonction de l'angle Chi1 (en rad) pour la tyrosine 10 protonée dans ses conformations IN et OUT, avec glutathion



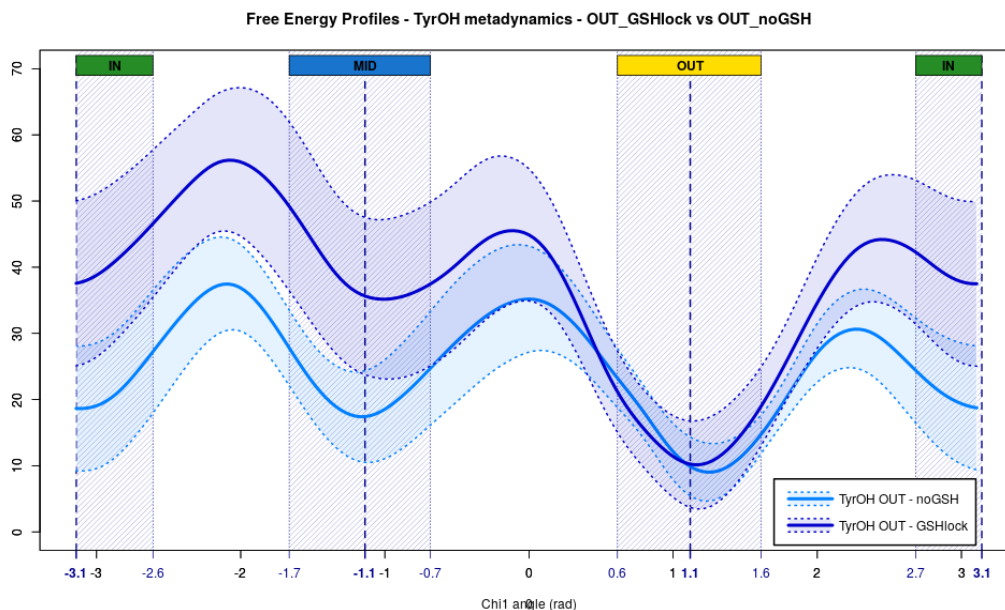
En étudiant l'impact de la présence ou non de glutathion en débutant de la conformation IN (Figure 58), il existe une augmentation significative de l'énergie libre au niveau de la conformation IN de la tyrosine 10, comme le démontre le décalage observé entre les intervalles de confiance à 95%, ayant pour conséquence le changement notable de barrières énergétiques notables :

- Une augmentation significative de MID à IN, passant de 13.83 kJ/mol sans glutathion à 25.87 kJ/mol en présence de glutathion ;
- Une diminution significative de IN à OUT, passant de 16.25 kJ/mol sans glutathion à 8.51 kJ/mol en présence de glutathion.

La diminution de la barrière énergétique de IN à OUT est confirmée par l'observation de l'impact de la présence ou non de glutathion en débutant de la conformation OUT (Figure 58), avec des énergies passant de 11.99 kJ/mol sans glutathion à 6.69 kJ/mol en présence de glutathion.



**Figure 58. Profils d'énergie libre moyens (en kJ/mol) et intervalle de confiance à 95% en fonction de l'angle Chi1 (en rad) pour la tyrosine 10 protonée dans sa conformation IN, avec et sans glutathion.**

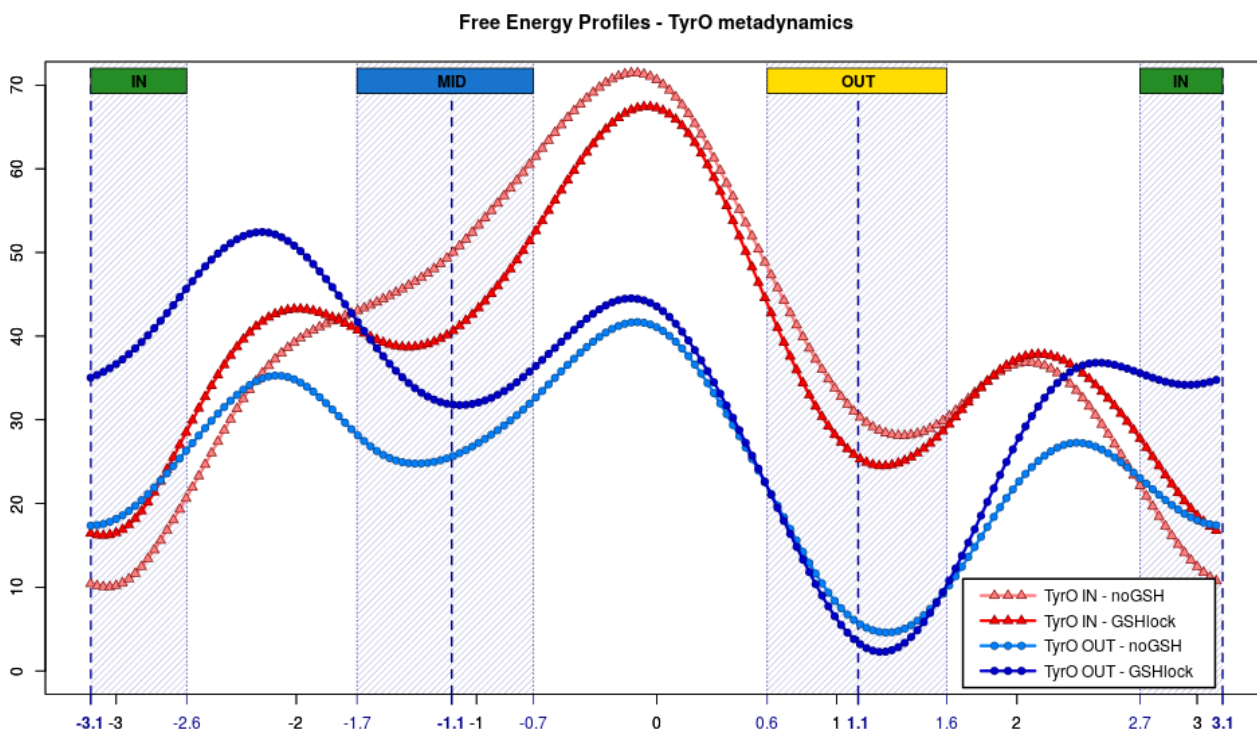


**Figure 59.** Profils d'énergie libre moyens (en kJ/mol) et intervalle de confiance à 95% en fonction de l'angle Chi1 (en rad) pour la tyrosine 10 protonée dans sa conformation OUT, avec et sans glutathion.

### > Profils énergétiques moyens - TyrO- :

A partir des 102 répliques de métadynamique réalisées avec la variable collective Chi1 dans les conditions expérimentales considérant la tyrosine 10 sous sa forme déprotonée, 4 profils énergétiques moyens ont été obtenus (Figure 60). Différents minima énergétiques globaux sont observables, avec un minimum global dans la conformation IN pour les simulations ayant comme point de départ la conformation IN, et un minimum global dans la conformation OUT en commençant à partir de la conformation OUT. De plus, en étudiant les trajectoires de métadynamique obtenues, on n'observe que peu d'évolution au sein des différentes simulations, où l'exploration du système de départ sera plutôt privilégiée, expliquant que les minima globaux sont identiques aux conformations de départ. Enfin, les minima locaux et les barrières énergétiques ne sont pas similaires, pouvant indiquer un problème lié aux structures de départ.

Ces problèmes peuvent également être inhérents au champ de force AMBER ff14SB utilisé, où les charges ajoutées sur les tyrosines 10 ont pu être mal reconnues et/ou prises en compte.



**Figure 60.** Profils d'énergie libre moyen (en kJ/mol) en fonction de la variable collective  $\chi_1$  (en rad) pour 4 conditions expérimentales mettant en jeu la tyrosine 10 déprotonée, pour un minimum de 20 répliques de 20 ns pour chaque condition.

Les simulations de métadynamique réalisées au travers de ces différents travaux ont permis la mise au point des conditions d'étude des mouvements de la tyrosine catalytique. A la suite de plus de 200 répliques de métadynamique, le paysage énergétique de la tyrosine a pu être décrit en déterminant différents profils d'énergie libre. Ces profils mettent en avant la stabilité de la conformation OUT, retrouvée au sein de structures cristallographiques, et la présence de barrières énergétiques conditionnant le changement conformationnel de la tyrosine 10. Toutefois, les conditions expérimentales utilisées ne sont pas adaptées à l'étude de l'impact de la déprotonation sur les conformations de la tyrosine, notamment par l'utilisation d'un champ de force non adapté.

# **CHAPITRE II**

## **OUTIL D'ETUDE DES RESEAUX D'INTERACTIONS**

### **MOLECULAIRES ET SON APPLICATION**



# **I. Développement de SINAPs**

## **1. Introduction**

Les principaux programmes de détermination et d'observation des réseaux d'interactions d'une protéine n'étant pas prévus pour soit l'utilisation de données de dynamique moléculaire, soit pour une représentation des résultats sur la structure tridimensionnelle, un logiciel a été développé dans le cadre de la thèse. Ce programme, codé en Python2 et Python3, est appelé SINAPs (*Structural Interaction Networks Analysis Protocols*).

Un article soumis au *Journal of Chemical Information and Modeling*, repris ci-après, présente les caractéristiques du logiciel, sa validation sur deux cibles biologiques très étudiées en retrouvant les résultats tirés de la littérature, et son application à la P28GST.

## **2. Article SINAPs**

# SINAPs: A software tool for analysis and visualization of interaction networks of molecular dynamics simulations

*Corentin BEDART<sup>a,b\*</sup>, Nicolas RENAULT<sup>a</sup>, Philippe CHAVATTE<sup>a</sup>, Adeline PORCHERIE<sup>b</sup>, Abderrahim LACHGAR<sup>b</sup>, Monique CAPRON<sup>a,b</sup>, Amaury FARCE<sup>a\*</sup>*

(a) Univ. Lille, Inserm, CHU Lille, U1286 - Infinite - Institute for Translational Research in Inflammation, F-59000 Lille, France

(b) Par'Immune, Bio-incubateur Eurasanté, 70 rue du Dr. Yersin, 59120, Loos-Lez-Lille, France

**Keywords:** Molecular dynamics; Interaction network; Analysis; Visualization; UCSF Chimera; SINAPs.

**Abstract:** As soon as the structural study of molecular mechanisms requires multiple molecular dynamics reflecting contrasted bioactive states, the subsequent analysis of molecular interaction networks remains a bottleneck to be fairly treated and requires a user-friendly 3D view of key interactions. SINAPs (Structural Interaction Networks Analysis Protocols) is a proprietary python tool, developed to i) quickly solve key interactions able to distinguish two protein states, either from two sets of molecular dynamics simulations, or from two crystallographic structures; ii) render a user-friendly 3D view of these key interactions through a plugin of UCSF Chimera, one of the most popular of open-source viewing softwares for biomolecular systems. Through two case studies, glucose transporter-1 (GLUT-1) and A2A adenosine receptor (A2AR), SINAPs easily pinpointed key interactions

observed experimentally and relevant for their bioactivities. This very effective tool was thus applied to identify the amino acids involved in the molecular enzymatic mechanisms ruling the activation of an immunomodulator drug candidate, P28 Glutathion-S-Transferase (P28GST). SINAPs is freely available at <https://github.com/ParImmune/SINAPs>.



## INTRODUCTION

Molecular dynamics (MD) simulations are advanced methods for studying the evolution of a system of particles over time. Applied to biological macromolecules, these methods allow the study of their functional mechanism, the binding kinetics of molecular modulators such as small organic or macromolecular partners as well as the identification of new druggable pockets for discovery of new active compounds through some computational drug design processes.

The recording interval of the frames of MD simulations is a parameter that must be optimized. A too long interval can lead to missing valuable information. A too short interval will increase the quality and quantity of information generated, but leading to a drastic increase of the weight of MD trajectories<sup>1,2</sup>. As these simulations generate a huge amount of data, the analysis phase is a bottleneck<sup>3,4</sup>. The use of several tools is therefore mandatory to optimize the analysis<sup>1</sup>, seeking among others to highlight differences in the interactions during the simulation.

Some tools take as input molecular dynamics trajectories, calculate the interactions, but do not have an adequate display of the results. This is the case of tools proposed directly with the molecular dynamics softwares, such as GROMACS<sup>5</sup> or AMBER via cpptraj<sup>6,7</sup>, which propose textual results. The 3D viewer VMD<sup>8</sup> provides analysis of hydrogen and ionic bonds over time with the Timeline utility, but only gives general statistics in the form of graphs. A third-party plugin called PyContact<sup>9</sup> improves VMD's flaws in the calculation of interactions but does not provide a visual representation either. RIP-MD<sup>10</sup> displays its results in Cytoscape<sup>11</sup>, but loses the information of the three-dimensional structure of the protein. MDcons<sup>12</sup> provides intermolecular contact maps from molecular dynamics trajectories of protein complexes, but does not provide information on intramolecular interactions. CONAN<sup>13</sup> produces contact

maps along trajectories as well as visual output through the creation of heatmaps images or animations, but without representations of 3D structures. G\_contacts<sup>14</sup> proposes a set-decomposition algorithm to detect all contacting atoms or residues during MD simulations, faster than traditional brute-force methods. Finally, NAPS<sup>15</sup> is an online web server that does not perform molecular interaction networks but centrality-based networks, thus does not take into account non-bonded interactions.

Other tools offer interesting graphical representations but are not adapted to molecular dynamics trajectories as they only take PDB structures as input. This is the case of the interaction calculation tools proposed by the main visualization softwares such as UCSF Chimera<sup>16</sup> or PyMOL<sup>17</sup>, or more specific to ligand-protein interactions such as LigPlot+<sup>18</sup>. In addition, some tools offer a link between Cytoscape and UCSF Chimera, such as structureViz<sup>19</sup> and RINspecter<sup>20</sup>, but only perform centrality interaction networks.

Tools hosted online on web servers are generally to be avoided, as they may disappear at any time and may cause privacy issues for confidential research subjects.

There is therefore no perfect tool available offline that performs both analysis and visualization. The SINAPs tool presented in this paper achieves a balanced compromise. The SINAPs analyzer tool evaluates the differences of interactions between two molecular systems. Then, the SINAPs visualization tool provides a simplified visualization of the results taking into account the structural context through an extension to the UCSF Chimera visualization software<sup>16</sup>.

To assess its usefulness, SINAPs was applied to several case studies covering most of its applications. The first simplest case was to compare two molecular interaction networks from two crystallographic structures exhibiting contrasted bioactivities, here the class-A A2AR GPCR complexed with agonist or antagonist ligands. The second more sophisticated case was

based on the differential study of two multi-conformational sets each carried out from MD trajectories of GLUT-1. In addition, an exploratory study was performed using classical MD of a protein drug candidate, P28GST, to assist the study of its molecular mechanism of action.

All information about SINAPs, including download information and documentation, can be found on this website: <https://github.com/ParImmune/SINAPs>.

## **MATERIALS & METHODS**

### **1. Software overview**

SINAPs includes two separate programs working together. The first component is a Python3 program analyzing the differences between two interaction networks. It takes as input two single structures (ie. PDB crystallographic structures) with the same number of residues or two multi conformational subsets from AMBER<sup>6</sup> and GROMACS<sup>5</sup> MD trajectories (Figure 1A). A MD trajectory may be divided into several parts to study the conformational changes occurring during the simulation.

The software first extracts the two structural templates, fits them according to alpha-trace and only keeps the backbone atoms to facilitate the future visualization within UCSF Chimera<sup>21</sup>. Concerning the two multi-conformational MD subsets, the structural templates are by default the first frame of both MD trajectories. The particularity of multi-conformational analysis is based on exhaustive exploration of interactions over all trajectory frames to calculate the frequency of non-bonded interactions formed over time while allowing precise control of the definition parameters of each type of interaction.

SINAPs' default settings have been set from literature, but the distance and angle criteria are easily editable. Detection of hydrogen bonds is computed from a native pytraj function, `search_hbonds`<sup>7,22</sup>. This function searches for bonds by geometric criteria according to the

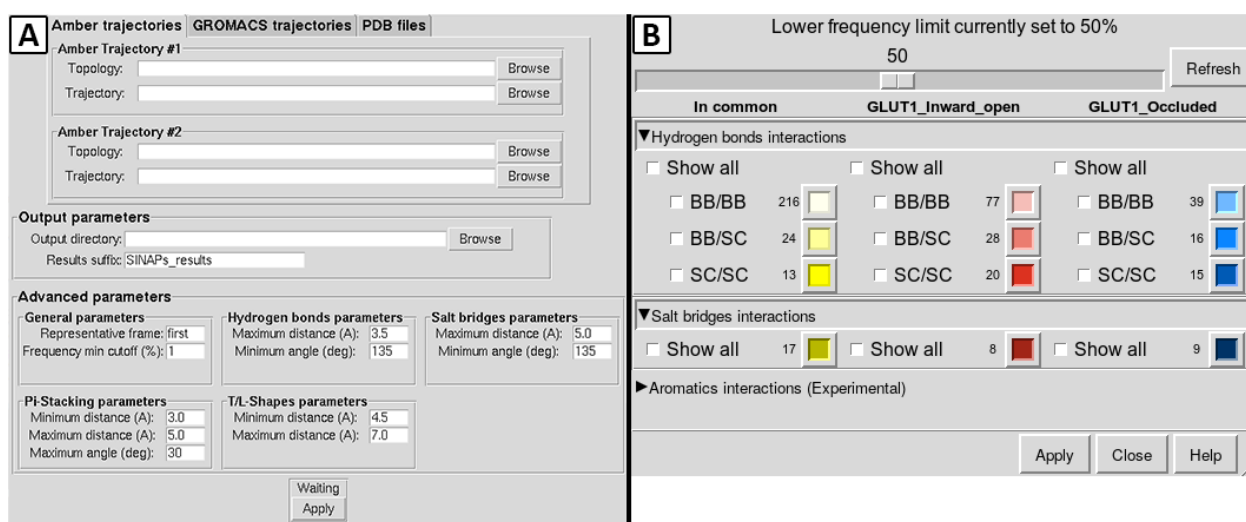
maximum distance between the donor hydrogen and the acceptor (set at 3.5 Å), as well as the minimum angle between the donor, its hydrogen, and the acceptor (set at 135°)<sup>23</sup>. Detection of salt bridges is computed with the same function, limiting donor residues to LYS and ARG, acceptor residues to GLU and ASP, and increasing the maximum bonding distance compared to hydrogen bonds from 3.5 Å to 5 Å<sup>23,24</sup>.

Detection of aromatic-aromatic interactions is computed from a homemade function based on geometric criteria for TYR, PHE, HIS, and TRP residues<sup>25,26</sup>. From the centroid and the normal vector of aromatic rings, several values are calculated: the distance between two centroids in Ångström, the planar angle describing the tilt of one aromatic ring relative to the other, and the orientation angle describing the offset of one aromatic ring relative to the other, both expressed in degrees modulo 90°<sup>26</sup>. Then, based on these parameters:

- A Pi-Stacking between two aromatic rings is defined if the distance between these two rings is within a defined range (set from 3 Å to 5 Å), if the planar angle is between 0° and 20° degrees, and if the orientation angle is between 0° and 30° (customizable value).
- A T-Shape is defined if the distance between the two rings is within a defined range (set from 4.5 Å to 7 Å), if the planar angle is between 0° and 60°, and if the orientation angle is between 0° and 30° or 60° and 90°.
- An L-Shape is defined if the distance between the two rings is within a defined range (set from 4.5 Å to 7 Å), if the planar angle is between 0° and 60°, and if the orientation angle is between 30° and 60°.

The output of the analysis tool are lists of interactions observed in the different input files, including the number of the residues involved in a given interaction, its type and its frequency in each file, as well as the chosen representative frames extracted in PDB format.

The second component is a Python2 add-on to UCSF Chimera<sup>16</sup> which allows the visualization of the results calculated by the first SINAPs Python3 tool. It takes as input the directory containing the results of the molecular interaction network analysis tool. After loading, this extension allows the observation of the representative frame of each simulation by displaying each type of non-bonded molecular interactions via a user-defined frequency lower limit and finely controlled mode of representation (Figure 1B).

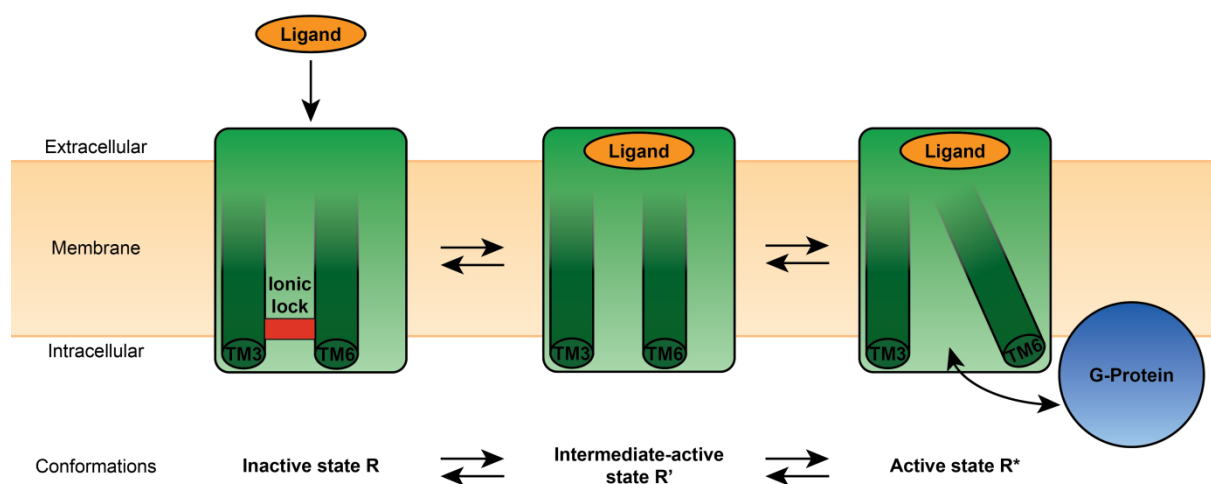


**Figure 1.** Graphical interfaces of the analysis software of SINAPs (A) and the visualization extension of SINAPs in UCSF Chimera (B). BB: Backbone chain; SC: Side chain; T-Shape: aromatic edge-to-face interaction; L-Shape: aromatic displaced edge-to-face interaction.

## 2. Case studies for depiction of critical interaction networks

### Comparison of PDB files: adenosine A2A receptor

The human adenosine A2A receptor (A2AR) belongs to the G Protein-Coupled Receptor superfamily, with four different ligand-bound states during its activation: two inactive R states with and without an “ionic lock” close to the cytoplasmic surface, an intermediate-active R’ state with the presence of a ligand, and a G protein-bound R\* active state<sup>27</sup> (Figure 2).

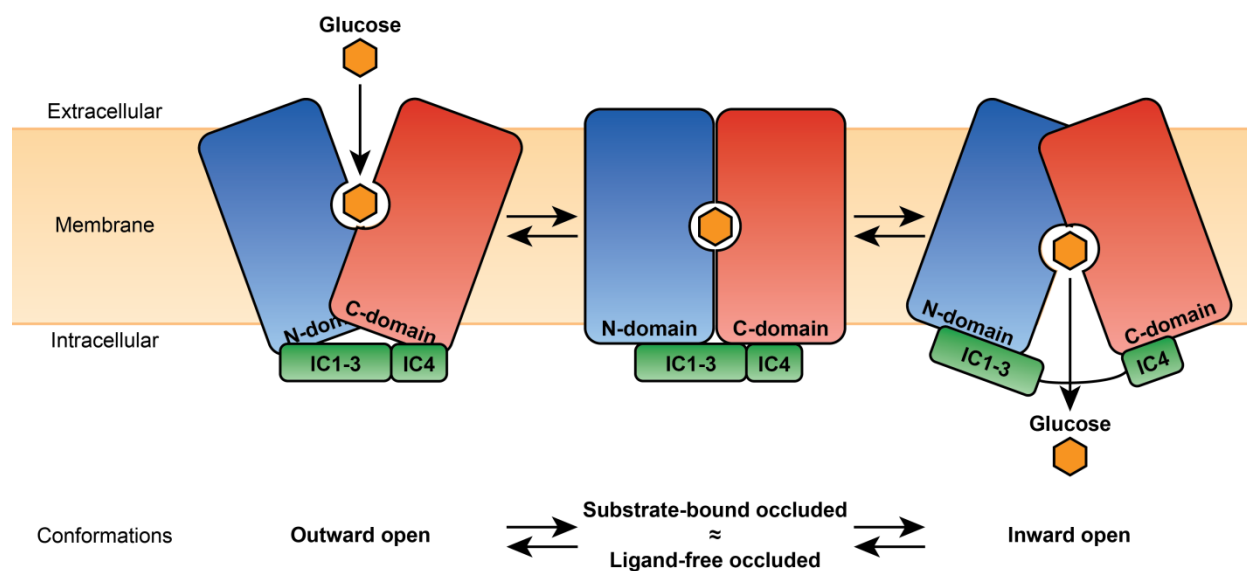


**Figure 2.** Simplified scheme of the mechanism of action of A2AR, moving from the inactive state R with the presence of the ionic lock, passing through the intermediate-active state R' with the interaction with an activating ligand and the loss of the ionic lock, then the active state R\* with interactions with an activating ligand and a G-protein.

The study of A2AR is based on crystallographic structures of the inactive conformation bound to the inverse agonist ZM241385 (PDB ID: 3PWH)<sup>28</sup>, the intermediate-active conformation bound to adenosine (PDB ID: 2YDO)<sup>29</sup> and the active conformation bound to NECA (PDB ID: 5G53)<sup>30</sup>. Using UCSF Chimera, other co-crystallized compounds and alternate positions of several residues were removed, and the hydrogen atoms were added. To ensure that SINAPs correctly considers the input structures by having the same number of residues in each structure, the residues 8 to 146, 159 to 208 and 225 to 305 were kept. Finally, the prepared structures were compared using SINAPs.

### Comparison of dynamic trajectories: glucose transporter 1

Glucose transporter 1 (GLUT-1) transfers glucose across the plasma membrane of numerous cells. During glucose transport, the receptor exhibits several different structural conformations (Figure 3). In this case, we focused particularly onto both close and open-inward conformations<sup>31</sup>.



**Figure 3.** Simplified scheme of GLUT-1 mechanism.

Before the comparative study of the multiple MD-based conformations of GLUT-1 receptor, we prepared the starting conformations: the inward-open conformation was extracted from the crystallographic structure of the human GLUT-1 (PDB ID: 4PYP)<sup>31</sup> whereas the occluded conformation was built by homology modeling using Target-Template Alignment of SWISS-MODEL<sup>32,33</sup>, starting from the human canonical amino acid sequence and the template crystallographic structure of the Major Facilitator Superfamily proton-xylose symporter Xyle, an *Escherichia coli* homologue of GLUT1 (PDB ID: 4GC0), based on the work of Sun et al.<sup>31,34</sup>. Then, using UCSF Chimera, all the ligands were removed to facilitate the preparation of MD. To have the same number of amino-acids, only residues 10 to 455 were kept. The hydrogen atoms were added via the AddH tool of UCSF Chimera using the method considering the possible hydrogen bonds.

MD simulations were performed using Amber18 software, with the ff14SB force field at constant pressure of 1 atm maintained by a Berendsen barostat, at constant temperature at 300 K with Langevin thermostat, and with periodic boundary conditions in an orthogonal box with TIP3P water and Na<sup>+</sup> counterions to neutralize the system.

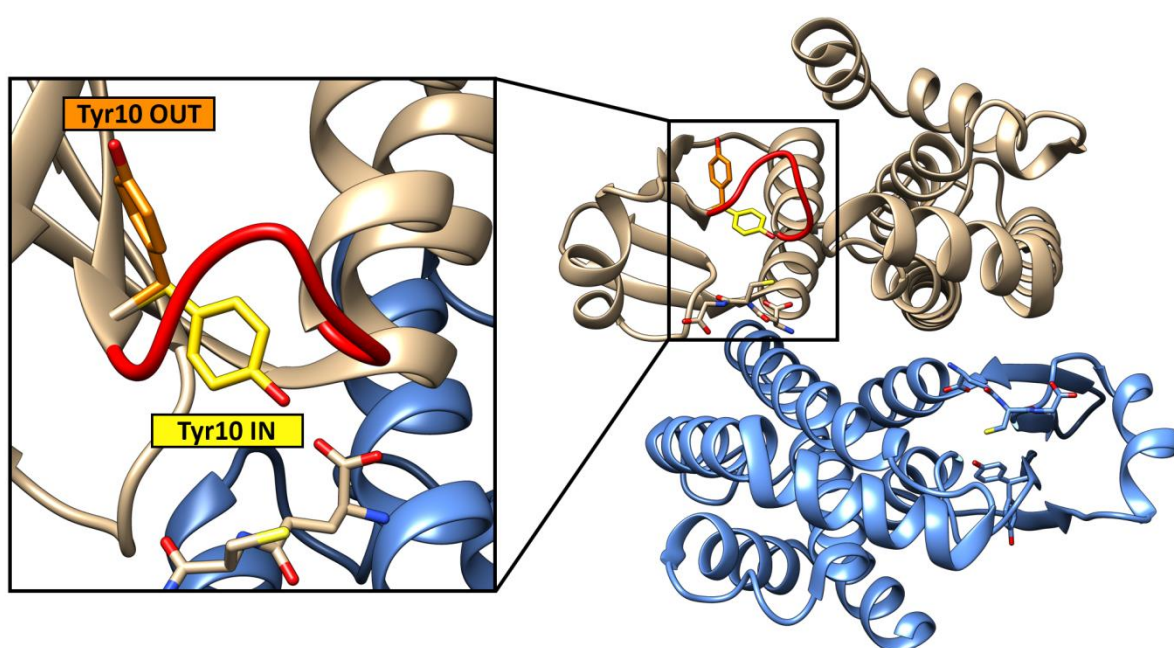
The systems were minimized by 1000 steps of steepest descent algorithm followed by 1000 steps of conjugate gradient algorithm with restraints on alpha carbons. Afterward, the systems were brought to simulation temperature with Langevin thermostat at constant volume and restrained alpha carbons, from 0 to 300 K for 9000 steps, and 300K for another 1000 steps. Then, 2ns equilibrium simulations were performed using the production conditions with restraints of alpha carbons to allow the correct positioning of the side chains of the protein and water molecules. The last frame of each equilibration step was used to launch three replicas of 50 ns of MD simulation for each system. An integration step of 2 fs was used, with the SHAKE algorithm constraining bonds involving hydrogen due to the use of TIP3P water, PME with a cutoff of 8 Å, and effective Born radii of 14 Å in the generalized Born approximation. Finally, the water molecules were removed, and then the frames were aligned on the position of alpha carbons of the first frame for each trajectory using cpptraj utility, before the trajectories were compared using SINAPs.

### **Real life application to a drug candidate: 28kDa glutathione S-transferase**

The 28kDa glutathione S-transferase protein (P28GST), a schistosome enzyme, is a drug candidate with anti-inflammatory properties developed by Par'Immune for treatment of chronic inflammatory bowel disease<sup>35</sup>. The biological activity combines immunogenic properties<sup>36</sup> with both glutathione S-transferase (GST)<sup>37,38</sup> and prostaglandin D synthase (PGDS)<sup>39</sup> enzymatic activities. Molecular mechanisms of the enzymatic activities remain to



be elucidated, in particular the involved amino acids. Up to now, only the tyrosine residue in position 10 was identified as the cornerstone of both enzymatic activities via the activation of glutathione, either into a co-substrate of GST, or into a co-factor of PGDS<sup>37,40</sup>. Upon glutathione activation, this Tyr10 gets two conformations observed from crystallography<sup>40</sup>: an IN conformation, where it points towards the catalytic site and the glutathione; an OUT conformation, where it points toward the outside of the protein and is exposed to the cellular environment (Figure 4).



**Figure 4.** Scheme of the different conformations of Tyr10. The loop following Tyr10 is represented in red.

To study the molecular interaction networks ruling P28GST function, and to investigate the main differences between these two conformations, three classical MD replicas of 1.5 $\mu$ s each were performed, starting from a crystallographic structure in its dimeric form (PDB ID: 1OE8)<sup>40</sup> and modified to exhibit the Tyr10 IN conformation in one monomer and the Tyr10 OUT conformation in the other one. In addition, to overcome the defects of the

crystallographic structure, the side chain of Phe38 was restored and the last four amino acids of each monomer were reconstructed using UCSF Chimera.

MD simulations were performed using Amber18 following the protocol described previously. After removing the water molecules and fitting the protein to the t0 alpha trace, the three replicas were merged as 225,000 trajectory frames. One tenth of this trajectory was sampled and submitted as a reasonable input set to SINAPs program. The first SINAPs analysis compared the whole trajectory against itself to study the intramolecular interaction network ruling the quaternary protein state. After separating the two monomers to generate separate sets of trajectories, the second SINAPs analysis related to the differential interaction network occurring in each monomer and depending on the conformational state of Tyr10.

### **3. Applying SINAPs**

After obtaining the main input files of all the cases under investigation, these were studied using SINAPs. Crystallographic structures of A2AR (in PDB format) were studied by changing several parameters. To compensate for resolution of the crystallographic structures, the hydrogen bonds and salt bridges parameters were upped by a 15% tolerance around the usual distance and angle limits. The lower limit angle taken into consideration was set at 115°, and the maximum distance for a hydrogen bond was set at 4 Å, and 6 Å for a salt bridge<sup>23,24</sup>. Regarding the interactions made by ligands, only the study of hydrogen bonds is currently possible.

Trajectory files (in nc format) and topology files (in parm7 format) of the GLUT-1 and P28GST cases were analyzed using the default settings of the software. The hydrogen bonds and salt bridges were searched with a lower limit angle set at 135°, and with a maximum distance between donor and acceptor atoms set at 3.5 Å and 5 Å respectively<sup>23,24</sup>. Parallel pi-

stackings were searched among aromatic amino-acids with a distance between the centers of mass of the aromatic rings between 3 and 5 Å and a maximal planar angle set at 30°<sup>25,26</sup>. T-shaped and L-shaped pi-stackings were defined among aromatic amino-acids with a distance between 4.5 and 7 Å<sup>25,26</sup>. To limit the number of interactions taken into consideration, and therefore to reduce the calculation time, a minimum frequency cutoff was set at 1%. In order to highlight the key interactions of our case studies, the minimum frequency cutoff was set at 50% for the analysis of network interactions from MD simulations.

All the figures presented in this article are screenshots from SINAPs' UCSF Chimera extension.

## **RESULTS & DISCUSSION**

### **1. Adenosine A2A receptor**

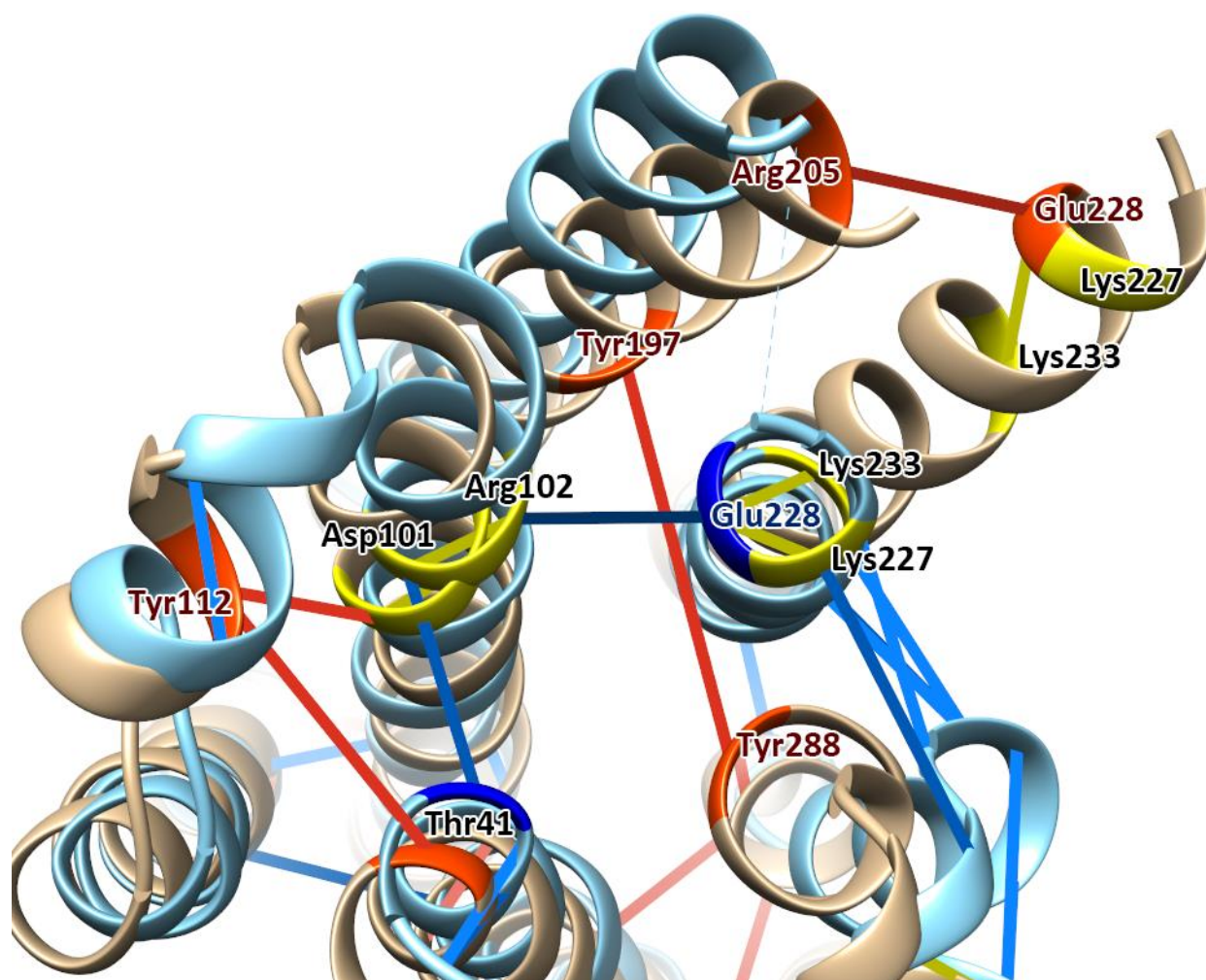
#### *Differential intramolecular interactions between A2AR conformational states*

The first observation carried out on the crystallographic structures of A2AR was to compare the main interaction networks formed between transmembrane helices depending on the conformation (Figure 5).

An interaction network can be detected with stable salt bridges between Asp101 and Arg102, between Lys227 and Glu228, and between Glu228 and Lys233, found in all conformations. However, a salt bridge can be observed between Arg102 and Glu228 only in the inactive conformation and replaced by another salt bridge between Arg205 and Glu228 only in the active conformation, through the TM6 tilt.

A second interaction network made of hydrogen bonds was highlighted between TM6 and TM7 helices in R and R' conformations. This network involves Lys227, His230, Ser234, and Phe242 on TM6, and Asn280, Arg291, and Arg293 on TM7. This network also involves interactions between TM7 and TM3/TM5 helices depending on the studied conformation. A

hydrogen bond between Tyr288 from TM7 and Arg102 from TM3 is observed in the R' conformation, replaced by a bond between Tyr288 and Tyr197 from TM5 in R\*.



**Figure 5.** Interaction network at cytoplasmic surface of A2AR (cyan) and A2AR\* (beige) A2AR states. Common salt bridges are shown in dark yellow whereas R\* exclusive salt bridges and hydrogen bonds are displayed respectively in dark and light red. Exclusive R salt bridges, side chain-side chain hydrogen bonds and side chain-backbone hydrogen bonds are shown respectively from dark to pale blue.

In the literature, the amino acids described in this study are part of structural pattern preserved in the vast majority of GPCRs. Asp101, Arg102, and Glu228 are described as part

of the E/DRY motif<sup>41</sup>, and amino acids ranging from Asp284 to Tyr288 as part of the NPxxY motif<sup>42</sup>.

The E/DRY motif plays a pivotal role in regulating GPCR conformational states<sup>41</sup>. The salt bridge between Arg102 and Glu228 is described as forming an “ionic lock” that stabilizes the receptor in an inactive form. This ionic lock allows the intracellular end of TM6 to be bent towards TM3 in the inactive conformation<sup>29,43</sup>.

In the intermediate-active conformation, Arg102 of the E/DRY motif no longer participates in the ionic lock, but interacts with Tyr288 of the NPxxY motif, bringing TM3 closer to TM7<sup>44</sup>. TM6 is still held in place near TM7 by a network of hydrogen bonds involving residues on either side of the NPxxY motif, but it is no longer interacting with TM3 with the deflection of the ionic lock.

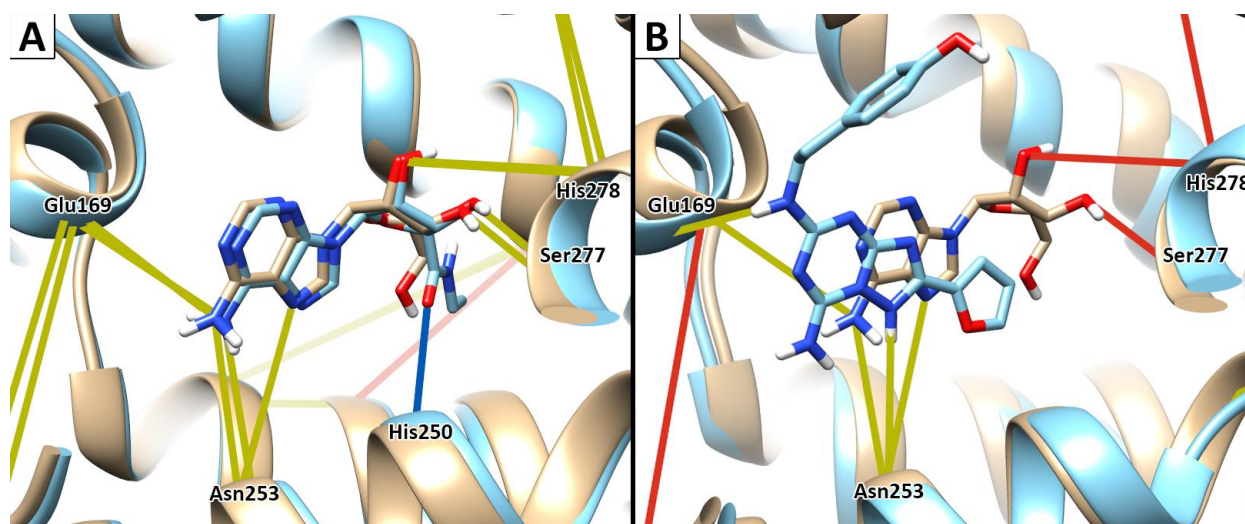
Finally, the interaction found only in the active conformation between Tyr288 of the NPxxY motif and Tyr197 is described as critical for forming the active state, as it operates as an activation-associated micro-switch<sup>27,42,45</sup>, which enables the tilting of TM6 to allow a G-protein to interact<sup>46</sup>. Since water molecules were removed, it would be relevant to evaluate their involvement in this interaction<sup>44</sup>. Moreover, after tilting, TM6 is maintained against TM5 by a salt bridge between Glu228 of the E/DRY pattern with Arg205<sup>45</sup>.

It is also worth mentioning that other observed phenomena are not covered in depth in this paper. These include the involvement of Asp52, Ser91, and Ser281 in the formation of a dense hydrogen bonds network in the intermediate-active and active conformations, supposed to be significant in the signal transduction mechanism of class-A GPCR<sup>47,48</sup>. Differences in hydrogen bonds network of the E/DRY motif with Ser41 and Tyr112 are also observed and may have direct implications in G-proteins activation<sup>49</sup>.

### ***Differential ligand binding modes between A2AR conformational states***

The second observation made was to compare the binding mode of studied crystallized ligands within their active site. The ligand of the active form, NECA, as well as that of the intermediate-active form, adenosine, spread identically into binding site and interact by hydrogen bonds with Glu169, Asn253, Ser277, and His278. Due to the additional methylamide group of NECA, this allows an additional hydrogen bond with His250 (Figure 6A). The ligand of the inactive form, ZM241385, only binds by two hydrogen bonds with Glu169 and Asn253 (Figure 6B).

The interactions found by SINAPs for all these ligands are found in the literature<sup>50</sup>. Moreover, Ser277 and His278 are highly conserved in evolution and play a key role in the activation of the receptor<sup>29</sup>, and the implication of all residues mentioned above in the binding of an agonist and/or an antagonist is verified by mutational analysis<sup>51</sup>.



**Figure 6.** Receptor-ligand interaction networks in A2AR conformations. (A) Interaction network between adenosine and R' (beige) then between NECA and R\* state (cyan), where common hydrogen bonds are shown in yellow, adenosine-exclusive hydrogen bonds in red, and NECA-exclusive hydrogen bonds in blue. (B) Interaction network between adenosine and

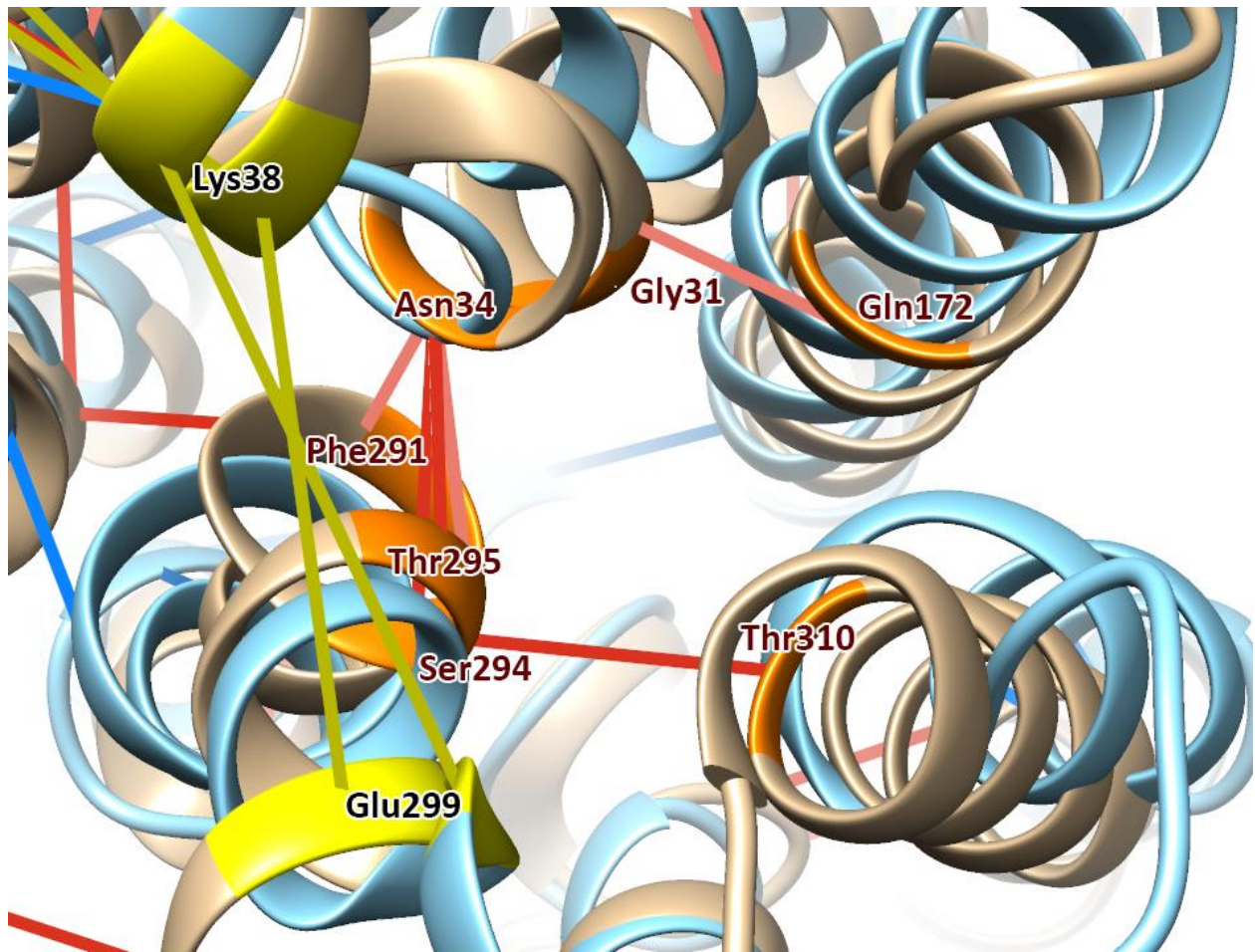
R' state (beige) then between ZM241385 and R state (cyan), where common hydrogen bonds are shown in yellow, adenosine-exclusive hydrogen bonds in red, and ZM241385-exclusive hydrogen bonds in blue.

## **2. Glucose transporter 1**

Two aspects of the conformational changes of GLUT-1 were examined. The first one was the extracellular gate, which occludes the ligand-binding site from the extracellular environment and promotes opening to the intracellular side in the inward-open conformation. The second aspect was the intracellular gate, with a dense network of salt bridges and hydrogen bonds at the intracellular side working as a door closer.

Both examined conformations of GLUT-1 exhibit alternatively the open extracellular ligand binding gate (EG) with the close intracellular gate (IG) and then the close EG with the open IG (Figure 3).

The EG is defined by amino acids located on the extracellular ends of transmembrane helices TM1, TM5, TM7 and TM8 (Figure 7). A common salt bridge is found between Lys38 on TM1 helix and Glu299 on TM7 helix. A clear difference exists at the level of interaction network between open and close EG. In open EG, hydrogen bonds exist between Gly31 and Asn34 on TM1 helix, Gln172 on TM5 helix, Phe291, Ser294 and Thr295 on TM7 helix, and Thr310 on TM8 helix whereas no interaction is observable in close EG.



**Figure 7.** Interaction network in open (beige) and close (cyan) EG of GLUT-1. Common salt bridges are shown in yellow whereas exclusive hydrogen bonds between open EG amino acids are displayed in red and exclusive hydrogen bonds between close EG are in light red.

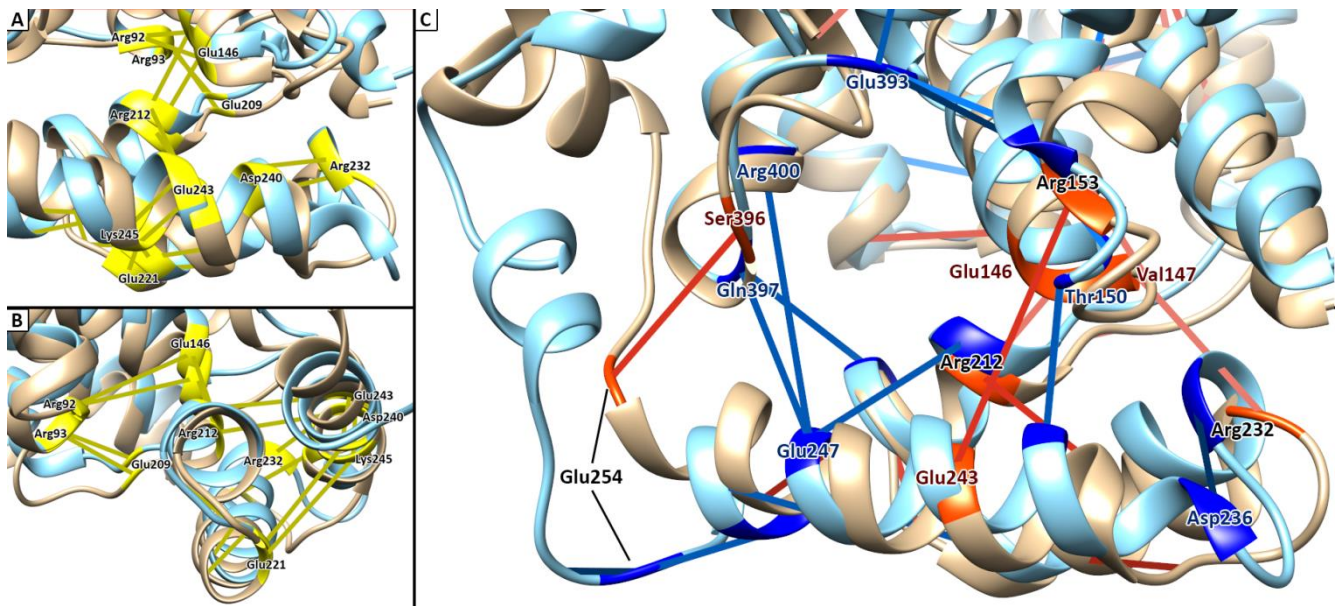
The IG is defined by amino acids located on the intracellular helices IC1 to IC4, and on the intracellular ends of transmembrane helices TM3, TM4, TM9, and TM10 (Figure 3). A common dense network of salt bridges is observed between Arg212, Glu221, Arg232, Asp240, Glu243, and Lys245, allowing to stabilize the IC1 to IC3 helices together. This helices bundle is then connected by other salt bridges to TM3 with the participation of Arg93 and Glu209, and to TM4 by Glu146 and Arg212. Finally, there are other notable common salt bridges between Arg92 of TM3 and Glu146 of TM4, between Arg333 of TM9 and Glu393 of TM10, and between Arg334 of TM9 and Glu454 of TM12 (Figure 8A-B).



Furthermore, notable differences were identified between the two studied conformations.

In the open IG conformation, the exclusive interactions bring the helices IC1 to IC3 closer to the N-domain, with a salt bridge between Glu243 of IC3 and Arg153 of TM4 and with two backbone-side chain hydrogen bonds, between Arg212 of IC1 and Glu146 of TM4, and between Arg232 of IC2 and Val147 of TM4. Interactions also bring the IC4 helix closer to the C-domain, with a hydrogen bond between the side chains of Ser396 of TM10 and Glu254 located on the coil before IC4. Glu254 also interacts by a backbone-backbone hydrogen bond with Gln250 of IC3, lengthening IC3 helix and positioning the IC4 coil in the direction of the cell membrane (Figure 8C).

In the close IG, the exclusive interactions have the opposite effect, as they bring the helices IC1 to IC3 closer to the C-domain, with a salt bridge between Glu247 of IC3 and Arg400 of TM10, and two hydrogen bonds connecting Glu247 of IC3 and Asn217 of IC1 with Gln397 of TM10. This helix complex is still linked to the C-domain with a hydrogen bond between Asp240 of IC3 and Thr150 of TM4. The two domains are also close together, with a dense interaction network between Ser148 of TM4, Arg153 and Gly154 of TM5, Glu393 of TM10, and Arg333 of TM9. Finally, the interactions between IC4 and TM10 are no longer present, as they are replaced by a strong interaction between Arg249 of IC3 and Glu254 located on the coil before IC4, positioning the IC4 coil in the direction of the intracellular medium (Figure 8C).



**Figure 8.** Interaction network in open (beige) and close (cyan) IG of GLUT-1 from two orthogonal points of view (A-B). In the open IG, exclusive salt bridges are shown in dark red, exclusive side chain-side chain hydrogen bonds in red, and exclusive side chain-backbone hydrogen bonds in light red. In the close IG, exclusive salt bridges are shown in dark blue, exclusive side chain-side chain hydrogen bonds in blue, and exclusive side chain-backbone hydrogen bonds in light blue (C).

GLUT-1 has several interaction networks found to be stable and unchanged regardless of the state or evolving during the passage between the different conformations of the protein during glucose transport.

A first interaction network mainly composed of hydrogen bonds is only found at the open EG. This network allows transmembrane helices TM3, TM4, TM9 and TM10 to be brought together close to the extracellular side in the open IG. This hydrogen bonds network is also described in the literature as allowing to tag the ligand-binding site in the open IG<sup>31,52</sup>.

A second interaction network of stable salt bridges, found in the two oEG/cIG and cEG/oIG studied conformations, stabilize helices IC1 to IC3 together, as well as these helices with the

N-domain through interactions with TM3 and TM4. However, additional interactions, hydrogen bonds and salt bridges, are observed depending on the conformation of the protein, allowing varying the distance between the helices IC1 to IC3 and the C-domain, as well as the change in geometry of the coil containing IC4. The inward-open conformation sees IC1 to IC3 helices interacting with the N-domain, when they rather interact with the C-domain in the occluded conformation. These differences are found in the literature, where all the cited substructures form a “door closer” which restrains the opening degree of the two domains towards the intracellular side<sup>53,54</sup>.

However, nothing in the literature refers to the involvement of Glu254 observed in this study. Differences in its interactions can explain the variation in the positioning of the IC4 helix coil in the studied conformations. Further studies are needed to minimize the biases present in this study, to confirm the presence of these interactions, and to assess the contribution of Glu254 in the change of geometry of IC4 during glucose transfer.

Furthermore, as the E329Q mutation has not been restored, its involvement has not been measured. However, its proximity to the Gly154 and Lys155 residues allows hypothesizing the involvement of this amino acid in the dense interaction network bringing together the two domains in the occluded conformation<sup>31</sup>.

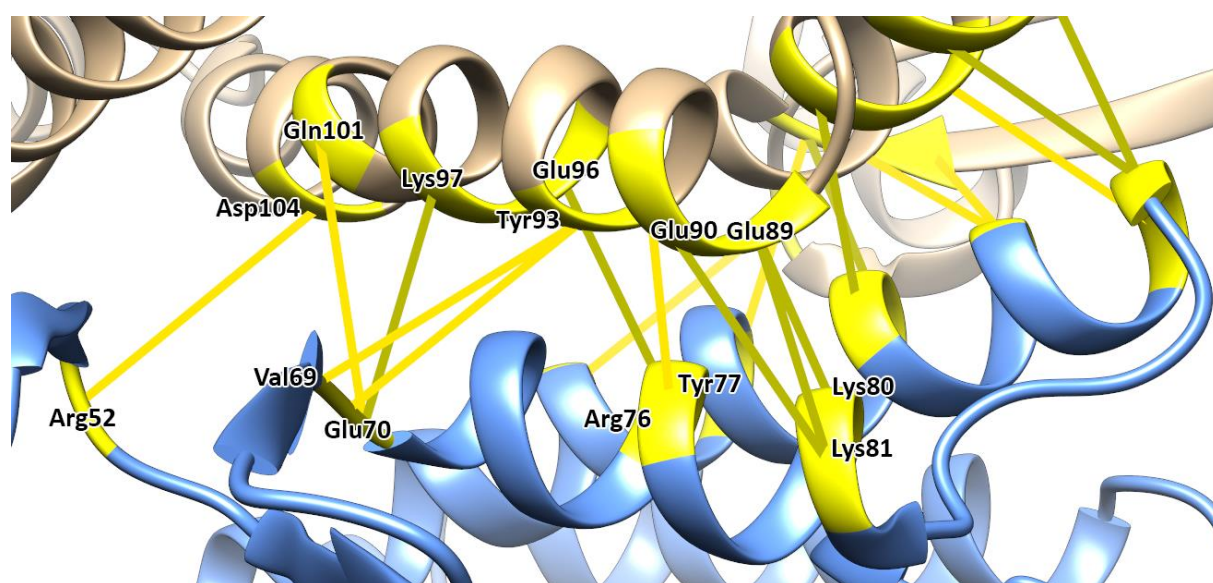
Despite the simplifications made in this study, it is possible to retrieve the very specific interaction networks of GLUT-1 described in the literature. From advanced MD simulations such as steered MD by simulating the passage of glucose from the extracellular to the intracellular side, it is conceivable to analyze the different conformations obtained by creating sub-trajectories for each conformation and analyze them with SINAPs tools. This analysis would easily identify from a simplified tridimensional representation the major differences in

the interaction networks, thus accelerating the studies conducted in the framework of a research project.

### 3. 28 kDa glutathione S-transferase

#### *Quaternary structure of P28GST – Dimer interface*

The quaternary organization of P28GST is based on axial symmetrical salt bridges and hydrogen bonds between dimers.



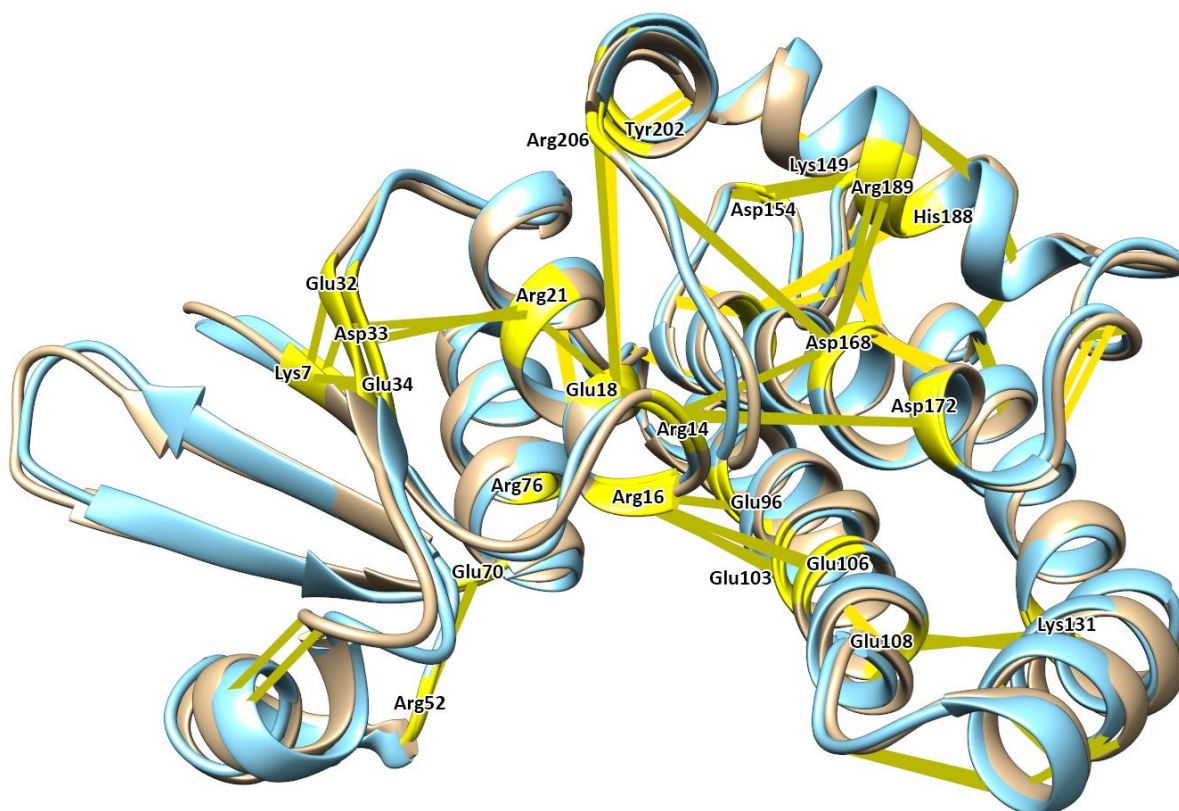
**Figure 9.** Interaction network at the dimer interface of P28GST. Common salt bridges are shown in dark yellow and common hydrogen bonds are shown in bright yellow.

A total of 6 salt bridges were found between Glu70 and Lys97, Glu89 and Lys80, Glu89 and Lys81, Glu96 and Arg76, Glu90 and Lys81, and between Asp104 and Arg52. Interestingly the 4 first mentioned salt bridges occurred in at least 90% of MD frames. Hydrogen bonds are also observed between Val69 and Tyr93, between Glu70 and Gln101, and between Gly90 and Tyr77 (Figure 9).

Finally, since water molecules were not considered, interactions may exist between several water molecules and the acidic amino acids Glu70, Glu96, Glu103 and Asp 104, all exposed in the direction of the other monomer. The involvement of a divalent cation with its stabilizing water molecules at the dimer interface may also exist, as suggested by other proteins of the same enzyme family<sup>55</sup>, but this cation has not been described in the crystal structures of P28GST.

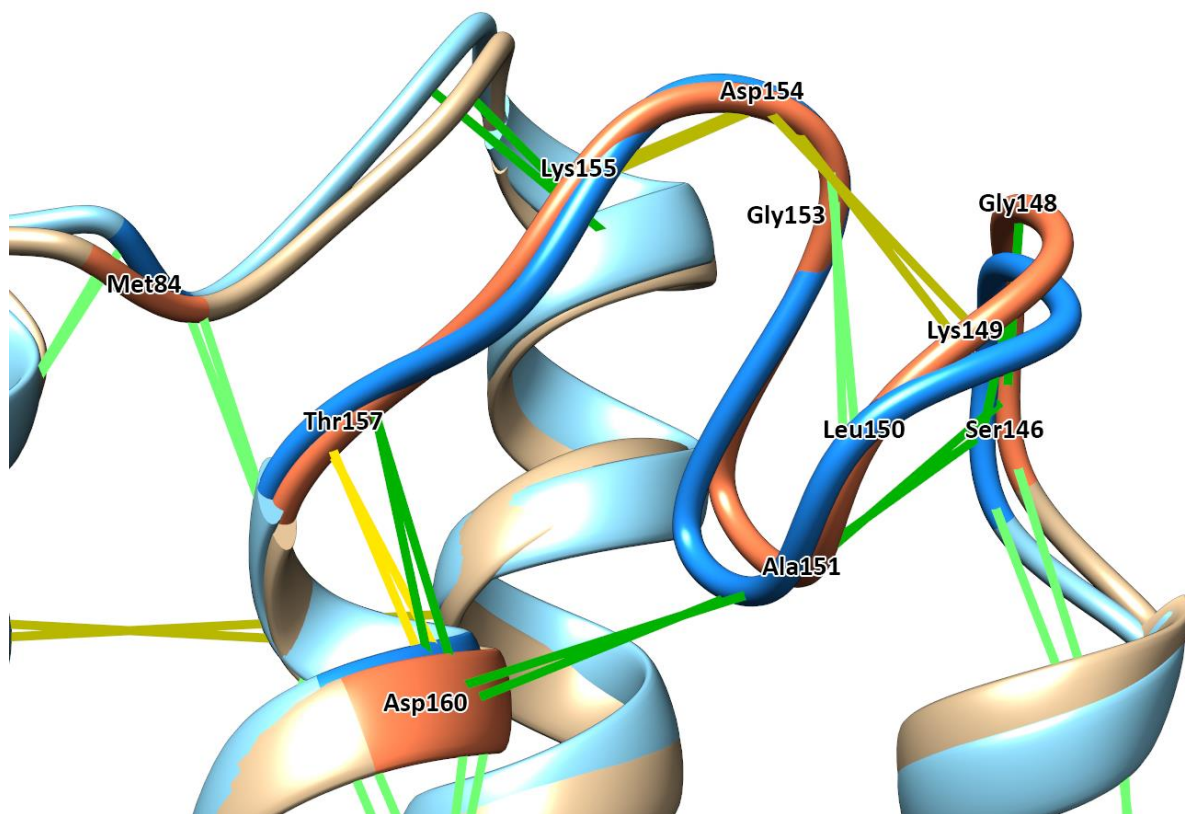
### ***Tertiary structure of P28GST – Common interactions***

The tertiary organization of P28GST is ruled by a very high number of salt bridges stabilizing the different secondary structures and coils together (Figure 10). At a frequency higher than 90%, 9 salt bridges covering the whole structure are observed, with the contribution of Lys7, Glu18, Arg21, Glu32, Asp33, Glu34, Arg76, Glu96, Glu108, Lys131, Lys149, Asp154, Asp168 and Arg206. At a frequency higher than 50%, 7 other salt bridges are found, with the contribution of Arg14, Arg16, Glu18, Arg52, Glu70, Glu103, Glu106, Asp168, Asp172, and Arg189. In addition, a few hydrogen bonds supplement the dense salt bridges interaction network with the participation of Glu18, Asp160, His188 and Tyr202.



**Figure 10.** Common interactions network stabilizing the tertiary organization of P28GST. Common salt bridges are shown in dark yellow and common hydrogen bonds are shown in bright yellow.

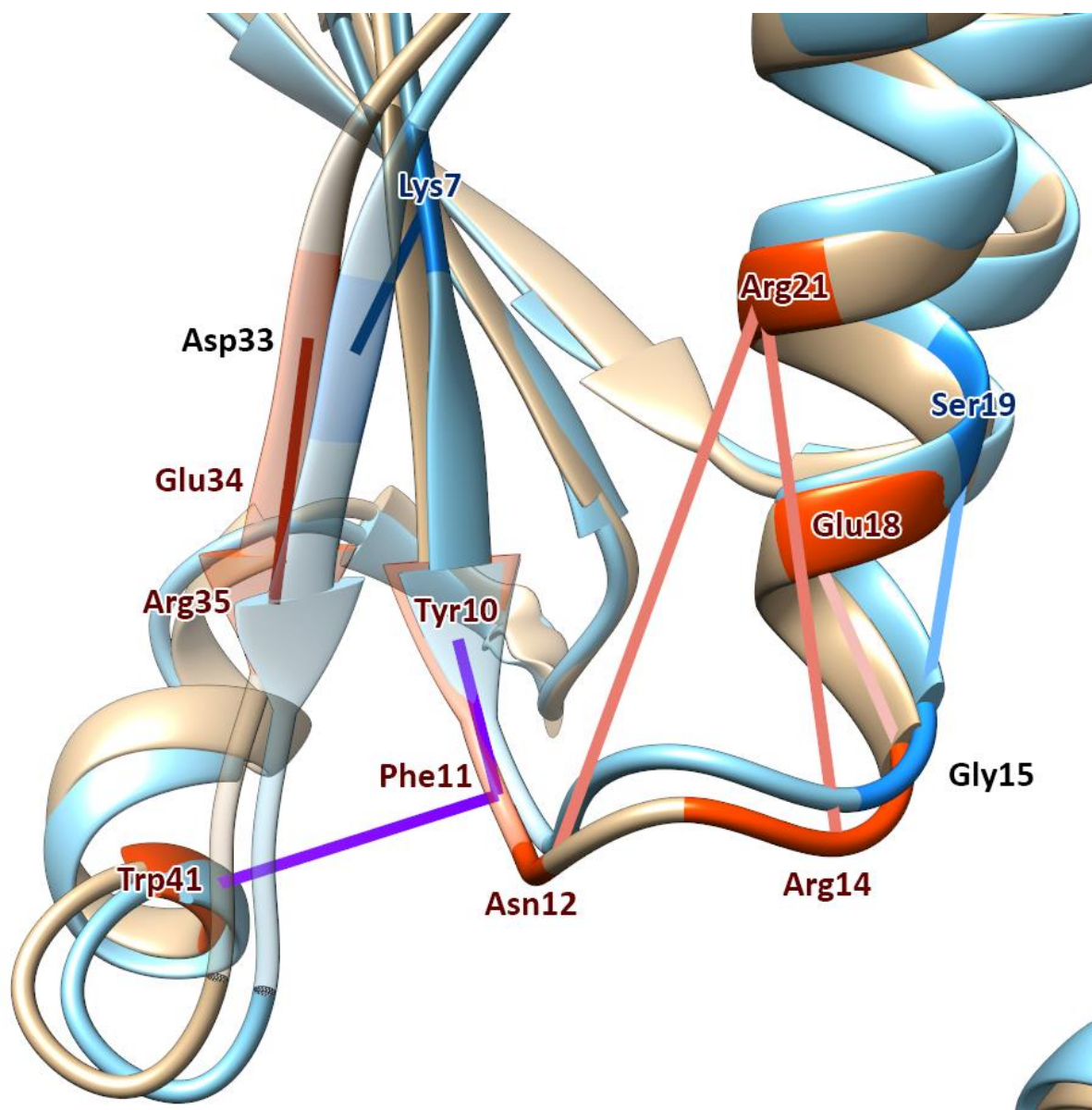
Moreover, some hydrogen bonds stabilize a coil from amino acids Ser146 to Thr157 in an unconventional shape with 3 turns (Figure 11). A side chain-side chain hydrogen bond is found between Asp160 and Thr157. Five backbone-side chain bonds are found between Thr157 and Asp160, Asp160 and 151, Leu150 and Ser146, Ser146 and Gly148, and between Lys149 and Asp154. Finally, two backbone-backbone hydrogen bonds are found between Leu150 and Gly153, and between Leu158 and Met84.



**Figure 11.** Interaction network stabilizing a coil between Ser146 and Thr157 within P28GST. Common salt bridges are shown in dark yellow, common side-chain side-chain hydrogen bonds in bright yellow, common side-chain backbone hydrogen bonds in dark green, and common backbone backbone hydrogen bonds in bright green.

### *Tertiary structure of P28GST - Tyr10 conformation-specific interactions*

When comparing monomers with each other, specific interactions are highlighted, arranging the environment around Tyr10 in a different way depending on its conformation (Figure 12).



**Figure 12.** Differences in interaction network around the catalytic Tyr10 between IN conformation (in beige and orange) and OUT conformation (in cyan and blue). IN exclusive salt bridges is displayed in dark red, sidechain-backbone hydrogen bonds in light red, backbone hydrogen bonds in pink, and aromatic interactions in purple. OUT exclusive salt bridges is displayed in dark blue and backbone hydrogen bonds in cyan.

In the Tyr10 IN conformation, Tyr10 is stabilized by an aromatic bond with Phe11, which is itself stabilized by another aromatic bond with Trp41. Salt bridges hold Asp33, Glu34 and



Arg35 together. Lastly, the loop following Tyr10 is maintained by hydrogen bonds between Asn12 and Arg21, Arg14 and Arg21, and between Gly15 and Glu18.

In the Tyr10 OUT conformation, due to its position and close neighbors, Tyr10 is no longer stabilized. The salt bridges maintaining the polar amino acid chunk in place are no longer present, allowing Asp33 to perform a salt bridge with Lys7. Finally, the loop is no longer held in place, and there is a backbone hydrogen bond between Gly15 and Ser19 allowing the subsequent helix-alpha to be lengthened.

### ***Discussion***

Numerous intramolecular and intermolecular interaction networks are found in P28GST, organizing its structure.

The quaternary structure is composed of a symmetrical interaction network, including very stable salt bridges and hydrogen bonds. These numerous interactions at the interface suggest a biological stability of P28GST as a dimer, corroborating the physical stability observable with crystallographic methods. As the SINAPs tool currently does not consider water molecules and ions, it would be worthwhile to study their impact in the formation of a hypothetical interaction network of protein-water, or protein-water-cation.

The tertiary structure of P28GST is mostly supported by a dense salt bridge interaction network that holds the architecture of the protein. Further interaction networks are specific to activation states of the protein conditioning its activity.

In the literature, no information was found on the involvement of the coil organization of amino acids Ser146 to Thr154 in proteins of the same family. In addition, no information is found when searching for specific interactions at long range induced by the different conformations of Tyr10 down to the minimum possible frequency of 1%. The purpose of this structure, including its potential role in the mechanism of action of P28GST and in the movement of Tyr10, remains to be elucidated.

Depending on the IN or OUT conformations of Tyr10, specific interactions are found in its surroundings, governing its activation state. In the literature, site-directed mutagenesis and X-ray crystallography have highlighted the importance of the amino acids Arg21 and Arg35 in the enzymatic function of P28GST<sup>37</sup>.

Arg21 was identified by SINAPs as being essential for the global architecture of P28GST through salt bridges with Glu18 and Asp33. When it is mutated to leucine, thus no longer allowing the realization of salt bridges and hydrogen bonds, only the OUT conformation of Tyr10 is populated in crystallography<sup>37</sup>. This observation is explained by the results of SINAPs, which show the presence of specific side chain-backbone hydrogen bonds with Asn12 and Arg14 in the loop after Tyr10 in its IN conformation, so having a possible stabilizing role of this loop.

The combined observations point to the importance of Arg21 in the motion of Tyr10 by playing a switch role, conditioning the IN and OUT conformations.

Asp33, maintained by Arg35 in the IN conformation but interacting with Lys7 in the OUT conformation, was described as a proton acceptor and could be involved in the deprotonation

of Tyr10, that is then able to perform its glutathione deprotonation function after switching to its IN conformation.

Arg35, found with SINAPs to perform salt bridges with Asp33 and Glu34 only when Tyr10 is in its IN conformation, was described in the literature as performing concerted movements with the latter, allowing the formation of a "gate" conditioning the conformational switch of Tyr10.

Through this example, the SINAPs tool helps to find molecular interactions described in the literature within classical MD, allowing the validation of the biological reality of the simulation, and to support the hypotheses on the enzymatic mechanism of glutathione activation in P28GST. SINAPs also permits the validation of the method: as the features found here are proven by experimental evidence described in the literature, we can assume that new observations made in the future will also be valid.

## **CONCLUSION**

We presented here SINAPs, a dual analysis-visualization series of tools coded in Python that allows a user-friendly study of the similarities and differences in non-bonded interaction networks between molecular systems. These systems may be drawn from crystallographic datasets or come from MD trajectories, therefore helping greatly in the analysis of the large amount of data generated. The visualization part is offered via an add-on to UCSF Chimera and permits a high level of user customization of the representation. Its most important advantage is to link the interactions and the structure of the proteins in a single view, while keeping the information about the type and the frequency of the interactions. It currently takes

into account intramolecular salt bridge, hydrogen bonds and aromatic interactions as well as intermolecular hydrogen bonds, which permits to study both protein internal interaction networks throughout their conformational changes and ligand-receptor interactions, as exemplified with GLUT-1 dynamics and A2AR crystallography. The SINAPs software also enables the validation of classical MD simulations by facilitating the search for very specific interaction networks, as described here with examples from the literature, to facilitate the understanding of the mechanism of action of a target of interest, and to characterize new structures, as seen with the exploratory study of P28GST.

Its integration as an easy-to-implement plug-in to one of the most widely used free 3D molecular viewer opens it to both chemoinformatics experts and less experienced users from different backgrounds. This software is also scalable, with the future addition of the detection of intermolecular protein-small molecules salt bridges and aromatic interactions. Other improvements are also considered, such as hydrophobic interactions, the detection of intermolecular protein-ligand interactions in addition to the detection of hydrogen bonds already implemented, the optimization of the calculation algorithm, the current one running only on one CPU, or the improvement of the display of the results, in particular in the form of a list of interactions in UCSF Chimera, to enhance its running speed and improve the friendliness and depth of the representation.

## **ASSOCIATED CONTENT**

SINAPs\_SI (in XLSX format): List of interactions observed when comparing the opened and closed intracellular gate conformations in GLUT-1 molecular dynamics simulations, within P28GST in its dimeric form, specifically at the dimeric interface of P28GST, and when comparing the IN and OUT conformations of P28GST after splitting of its monomers. The different tables, provided in different sheets, list the residue ID of acceptors and donors, the observed frequencies, and the type of bond in the specific case of hydrogen bonds.

## **AUTHOR INFORMATION**

### **Corresponding Authors**

**Corentin BEDART** - Par'Immune, Bio-incubateur Eurasanté, 70 rue du Dr. Yersin, 59120, Loos-Lez-Lille, France ; Email : [c.bedart@parimmune.com](mailto:c.bedart@parimmune.com)

**Amaury FARCE** - INFINITE UMR1286 Inserm | Université de Lille | Institut de Chimie Pharmaceutique Albert Lespagnol, 3 rue du Professeur Laguesse, BP83, F-59006 Lille Cedex ; Email: [amaury.farce@univ-lille.fr](mailto:amaury.farce@univ-lille.fr)

## **ACKNOWLEDGMENT**

CB acknowledges a CIFRE PhD fellowship from both Par'Immune (Loos, France) and the French Association Nationale de la Recherche et de la Technologie (ANRT) (CIFRE fellowship n°2018/1313).

A part of molecular dynamics simulations presented were performed using high performance computing capabilities of Lille University datacenter with the use of GPU acceleration (<http://hpc.univ-lille.fr/cluster-hpc-htc>).

## REFERENCES

- (1) Mures, O. A.; Padrón, E. J.; Raffin, B. Leveraging the Power of Big Data Tools for Large Scale Molecular Dynamics Analysis. In *JP2016-XXVII Jornadas de Paralelismo*; 2016; pp 1–7.
- (2) Bernetti, M.; Bertazzo, M.; Masetti, M. Data-Driven Molecular Dynamics: A Multifaceted Challenge. *Pharmaceuticals* **2020**, *13* (9), 253. <https://doi.org/10.3390/ph13090253>.
- (3) Yu, I.; Feig, M.; Sugita, Y. High-Performance Data Analysis on the Big Trajectory Data of Cellular Scale All-Atom Molecular Dynamics Simulations. *Journal of Physics: Conference Series* **2018**, *1036*, 012009. <https://doi.org/10.1088/1742-6596/1036/1/012009>.
- (4) Liu, S. Compression of Molecular Dynamics Simulation Data. *Electrical Engineering and Computer Sciences University of California at Berkeley* **2019**.
- (5) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (6) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cruzeiro, V. W. D. AMBER 2018. *University of California, San Francisco* **2018**.
- (7) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. <https://doi.org/10.1021/ct400341p>.
- (8) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (9) Scheurer, M.; Rodenkirch, P.; Siggel, M.; Bernardi, R. C.; Schulten, K.; Tajkhorshid, E.; Rudack, T. PyContact: Rapid, Customizable, and Visual Analysis of Noncovalent Interactions in MD Simulations. *Biophysical Journal* **2018**, *114* (3), 577–583. <https://doi.org/10.1016/j.bpj.2017.12.003>.

- (10) Contreras-Riquelme, S.; Garate, J.-A.; Perez-Acle, T.; Martin, A. J. M. RIP-MD: A Tool to Study Residue Interaction Networks in Protein Molecular Dynamics. *PeerJ* **2018**, *6*, e5998. <https://doi.org/10.7717/peerj.5998>.
- (11) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- (12) Abdel-Azeim, S.; Chermak, E.; Vangone, A.; Oliva, R.; Cavallo, L. MDcons: Intermolecular Contact Maps as a Tool to Analyze the Interface of Protein Complexes from Molecular Dynamics Trajectories. *BMC Bioinformatics* **2014**, *15* (S5), S1. <https://doi.org/10.1186/1471-2105-15-S5-S1>.
- (13) Mercadante, D.; Gräter, F.; Daday, C. CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps. *Biophysical Journal* **2018**, *114* (6), 1267–1273. <https://doi.org/10.1016/j.bpj.2018.01.033>.
- (14) Blau, C.; Grubmuller, H. G\_contacts: Fast Contact Search in Bio-Molecular Ensemble Data. *Computer Physics Communications* **2013**, *184* (12), 2856–2859. <https://doi.org/10.1016/j.cpc.2013.07.018>.
- (15) Chakrabarty, B.; Parekh, N. NAPS: Network Analysis of Protein Structures. *Nucleic Acids Res* **2016**, *44* (W1), W375–W382. <https://doi.org/10.1093/nar/gkw383>.
- (16) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry* **2004**, *25* (13), 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- (17) Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8, 2015.
- (18) Laskowski, R. A.; Swindells, M. B. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **2011**, *51* (10), 2778–2786. <https://doi.org/10.1021/ci200227u>.
- (19) Morris, J. H.; Huang, C. C.; Babbitt, P. C.; Ferrin, T. E. StructureViz: Linking Cytoscape and UCSF Chimera. *Bioinformatics* **2007**, *23* (17), 2345–2347. <https://doi.org/10.1093/bioinformatics/btm329>.
- (20) Brysbaert, G.; Mauri, T.; Lensink, M. F. Comparing Protein Structures with RINspecter Automation in Cytoscape. *F1000Res* **2018**, *7*, 563. <https://doi.org/10.12688/f1000research.14298.2>.
- (21) Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25* (11), 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
- (22) Nguyen, H.; Roe, D. R.; Swails, J.; Case, D. A. PYTRAJ: Interactive Data Analysis for Molecular Dynamics Simulations. **2016**.

- (23) Sticke, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. Hydrogen Bonding in Globular Proteins. *Journal of Molecular Biology* **1992**, 226 (4), 1143–1159. [https://doi.org/10.1016/0022-2836\(92\)91058-W](https://doi.org/10.1016/0022-2836(92)91058-W).
- (24) Kumar, S.; Nussinov, R. Relationship between Ion Pair Geometries and Electrostatic Strengths in Proteins. *Biophysical Journal* **2002**, 83 (3), 1595–1612. [https://doi.org/10.1016/S0006-3495\(02\)73929-5](https://doi.org/10.1016/S0006-3495(02)73929-5).
- (25) Banerjee, A.; Saha, A.; Saha, B. K. Understanding the Behavior of  $\pi$ - $\pi$  Interactions in Crystal Structures in Light of Geometry Corrected Statistical Analysis: Similarities and Differences with the Theoretical Models. *Crystal Growth & Design* **2019**, 19 (4), 2245–2252. <https://doi.org/10.1021/acs.cgd.8b01857>.
- (26) Lanzarotti, E.; Biekofsky, R. R.; Estrin, D. A.; Marti, M. A.; Turjanski, A. G. Aromatic–Aromatic Interactions in Proteins: Beyond the Dimer. *J. Chem. Inf. Model.* **2011**, 51 (7), 1623–1633. <https://doi.org/10.1021/ci200062e>.
- (27) Weis, W. I.; Kobilka, B. K. The Molecular Basis of G Protein–Coupled Receptor Activation. *Annu. Rev. Biochem.* **2018**, 87 (1), 897–919. <https://doi.org/10.1146/annurev-biochem-060614-033910>.
- (28) Doré, A. S.; Robertson, N.; Errey, J. C.; Ng, I.; Hollenstein, K.; Tehan, B.; Hurrell, E.; Bennett, K.; Congreve, M.; Magnani, F.; Tate, C. G.; Weir, M.; Marshall, F. H. Structure of the Adenosine A2A Receptor in Complex with ZM241385 and the Xanthines XAC and Caffeine. *Structure* **2011**, 19 (9), 1283–1293. <https://doi.org/10.1016/j.str.2011.06.014>.
- (29) Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G. W.; Tate, C. G. Agonist-Bound Adenosine A2A Receptor Structures Reveal Common Features of GPCR Activation. *Nature* **2011**, 474 (7352), 521–525. <https://doi.org/10.1038/nature10136>.
- (30) Cheng, R. K. Y.; Segala, E.; Robertson, N.; Deflorian, F.; Doré, A. S.; Errey, J. C.; Fiez-Vandal, C.; Marshall, F. H.; Cooke, R. M. Structures of Human A 1 and A 2A Adenosine Receptors with Xanthines Reveal Determinants of Selectivity. *Structure* **2017**, 25 (8), 1275–1285.e4. <https://doi.org/10.1016/j.str.2017.06.012>.
- (31) Deng, D.; Xu, C.; Sun, P.; Wu, J.; Yan, C.; Hu, M.; Yan, N. Crystal Structure of the Human Glucose Transporter GLUT1. *Nature* **2014**, 510 (7503), 121–125. <https://doi.org/10.1038/nature13306>.
- (32) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics* **2006**, 22 (2), 195–201. <https://doi.org/10.1093/bioinformatics/bti770>.
- (33) Studer, G.; Tauriello, G.; Bienert, S.; Biasini, M.; Johner, N.; Schwede, T. ProMod3—A Versatile Homology Modelling Toolbox. *PLoS Comput Biol* **2021**, 17 (1), e1008667. <https://doi.org/10.1371/journal.pcbi.1008667>.
- (34) Sun, L.; Zeng, X.; Yan, C.; Sun, X.; Gong, X.; Rao, Y.; Yan, N. Crystal Structure of a Bacterial Homologue of Glucose Transporters GLUT1–4. *Nature* **2012**, 490 (7420), 361–366. <https://doi.org/10.1038/nature11524>.



- (35) Capron, M.; Béghin, L.; Leclercq, C.; Labreuche, J.; Dendooven, A.; Standaert, A.; Delbeke, M.; Porcherie, A.; Nachury, M.; Boruchowicz, A.; Dupas, J.-L.; Fumery, M.; Paupard, T.; Catteau, S.; Deplanque, D.; Colombel, J.-F.; Desreumaux, P. Safety of P28GST, a Protein Derived from a Schistosome Helminth Parasite, in Patients with Crohn's Disease: A Pilot Study (ACROHNEM). *JCM* **2019**, *9* (1), 41. <https://doi.org/10.3390/jcm9010041>.
- (36) Driss, V.; El Nady, M.; Delbeke, M.; Rousseaux, C.; Dubuquoy, C.; Sarazin, A.; Gatault, S.; Dendooven, A.; Riveau, G.; Colombel, J. F.; Desreumaux, P.; Dubuquoy, L.; Capron, M. The Schistosome Glutathione S-Transferase P28GST, a Unique Helminth Protein, Prevents Intestinal Inflammation in Experimental Colitis through a Th2-Type Response with Mucosal Eosinophils. *Mucosal Immunol* **2016**, *9* (2), 322–335. <https://doi.org/10.1038/mi.2015.62>.
- (37) Baiocco, P.; Gourlay, L. J.; Angelucci, F.; Fontaine, J.; Hervé, M.; Miele, A. E.; Trottein, F.; Brunori, M.; Bellelli, A. Probing the Mechanism of GSH Activation in *Schistosoma Haematobium* Glutathione-S-Transferase by Site-Directed Mutagenesis and X-Ray Crystallography. *Journal of Molecular Biology* **2006**, *360* (3), 678–689. <https://doi.org/10.1016/j.jmb.2006.05.040>.
- (38) Sarazin, A.; Dendooven, A.; Delbeke, M.; Gatault, S.; Pagny, A.; Standaert, A.; Rousseaux, C.; Desreumaux, P.; Dubuquoy, L.; Capron, M. Treatment with P28GST, a Schistosome-Derived Enzyme, after Acute Colitis Induction in Mice: Decrease of Intestinal Inflammation Associated with a down Regulation of Th1/Th17 Responses. *PLoS ONE* **2018**, *13* (12), e0209681. <https://doi.org/10.1371/journal.pone.0209681>.
- (39) Hervé, M.; Angeli, V.; Pinzar, E.; Wintjens, R.; Faveeuw, C.; Narumiya, S.; Capron, A.; Urade, Y.; Capron, M.; Riveau, G.; Trottein, F. Pivotal Roles of the Parasite PGD2 Synthase and of the Host D Prostanoid Receptor 1 in Schistosome Immune Evasion. *Eur. J. Immunol.* **2003**, *33* (10), 2764–2772. <https://doi.org/10.1002/eji.200324143>.
- (40) Johnson, K. A.; Angelucci, F.; Bellelli, A.; Hervé, M.; Fontaine, J.; Tsernoglou, D.; Capron, A.; Trottein, F.; Brunori, M. Crystal Structure of the 28 KDa Glutathione S-Transferase from *Schistosoma Haematobium*. *Biochemistry* **2003**, *42* (34), 10084–10094. <https://doi.org/10.1021/bi034449r>.
- (41) Rovati, G. E.; Capra, V.; Neubig, R. R. The Highly Conserved DRY Motif of Class A G Protein-Coupled Receptors: Beyond the Ground State. *Mol Pharmacol* **2007**, *71* (4), 959–964. <https://doi.org/10.1124/mol.106.029470>.
- (42) Katritch, V.; Cherezov, V.; Stevens, R. C. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu. Rev. Pharmacol. Toxicol.* **2013**, *53* (1), 531–556. <https://doi.org/10.1146/annurev-pharmtox-032112-135923>.
- (43) Latorraca, N. R.; Venkatakrishnan, A. J.; Dror, R. O. GPCR Dynamics: Structures in Motion. *Chem. Rev.* **2017**, *117* (1), 139–155. <https://doi.org/10.1021/acs.chemrev.6b00177>.
- (44) Bruzzese, A.; Dalton, J. A. R.; Giraldo, J. Insights into Adenosine A2A Receptor Activation through Cooperative Modulation of Agonist and Allosteric Lipid Interactions.

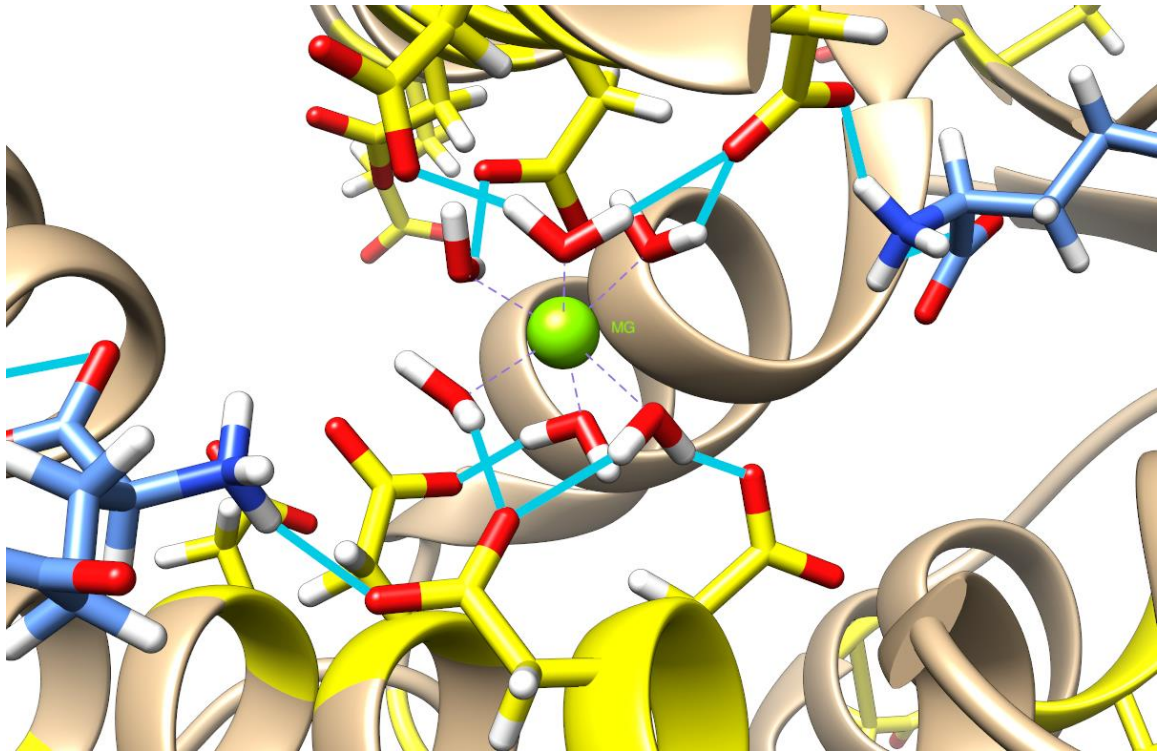
- (45) Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. The Structure and Function of G-Protein-Coupled Receptors. *Nature* **2009**, *459* (7245), 356–363. <https://doi.org/10.1038/nature08144>.
- (46) Millar, R. P.; Newton, C. L. The Year In G Protein-Coupled Receptor Research. *Molecular Endocrinology* **2010**, *24* (1), 261–274. <https://doi.org/10.1210/me.2009-0473>.
- (47) Zhang, X. C.; Cao, C.; Zhou, Y.; Zhao, Y. Proton Transfer-Mediated GPCR Activation. *Protein Cell* **2015**, *6* (1), 12–17. <https://doi.org/10.1007/s13238-014-0106-4>.
- (48) White, K. L.; Eddy, M. T.; Gao, Z.-G.; Han, G. W.; Lian, T.; Deary, A.; Patel, N.; Jacobson, K. A.; Katritch, V.; Stevens, R. C. Structural Connection between Activation Microswitch and Allosteric Sodium Site in GPCR Signaling. *Structure* **2018**, *26* (2), 259–269.e5. <https://doi.org/10.1016/j.str.2017.12.013>.
- (49) Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; IJzerman, A. P.; Stevens, R. C. The 2.6 Angstrom Crystal Structure of a Human A<sub>2A</sub> Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322* (5905), 1211–1217. <https://doi.org/10.1126/science.1164772>.
- (50) Carpenter, B.; Lebon, G. Human Adenosine A<sub>2A</sub> Receptor: Molecular Mechanism of Ligand Binding and Activation. *Front. Pharmacol.* **2017**, *8*, 898. <https://doi.org/10.3389/fphar.2017.00898>.
- (51) Kim, S.-K.; Gao, Z.-G.; Van Rompaey, P.; Gross, A. S.; Chen, A.; Van Calenbergh, S.; Jacobson, K. A. Modeling the Adenosine Receptors: Comparison of the Binding Domains of A<sub>2A</sub> Agonists and Antagonists. *J. Med. Chem.* **2003**, *46* (23), 4847–4859. <https://doi.org/10.1021/jm0300431>.
- (52) Chen, L. Y.; Phelix, C. F. Extracellular Gating of Glucose Transport through GLUT 1. *Biochemical and Biophysical Research Communications* **2019**, *511* (3), 573–578. <https://doi.org/10.1016/j.bbrc.2019.02.067>.
- (53) Deng, D.; Sun, P.; Yan, C.; Ke, M.; Jiang, X.; Xiong, L.; Ren, W.; Hirata, K.; Yamamoto, M.; Fan, S.; Yan, N. Molecular Basis of Ligand Recognition and Transport by Glucose Transporters. *Nature* **2015**, *526* (7573), 391–396. <https://doi.org/10.1038/nature14655>.
- (54) Galochkina, T.; Ng Fuk Chong, M.; Challali, L.; Abbar, S.; Etchebest, C. New Insights into GluT1 Mechanics during Glucose Transfer. *Sci Rep* **2019**, *9* (1), 998. <https://doi.org/10.1038/s41598-018-37367-z>.
- (55) Seo, M.-J.; Oh, D.-K. Prostaglandin Synthases: Molecular Characterization and Involvement in Prostaglandin Biosynthesis. *Progress in Lipid Research* **2017**, *66*, 50–68. <https://doi.org/10.1016/j.plipres.2017.04.003>.

## **II. Application de SINAPs à la recherche d'un cation divalent**

Lors de l'analyse des différentes dynamiques moléculaires par SINAPs, certains réseaux semblaient incomplets, notamment au niveau de l'interface dimérique. Dans la littérature et les structures cristallographiques, aucun cation n'a été retrouvé dans les cristallographies réalisées, contrairement à toutes les autres GST de classe sigma (80,100). Un travail sur la potentielle présence d'un ion métallique à l'interface a donc été mené, notamment au travers de résultats biologiques et de nouvelles dynamiques moléculaires.

### **1. Rationnel sur la possible présence d'un cation divalent chez P28GST**

Dans les structures cristallographiques d'enzymes de la même famille que la P28GST, il est possible d'apercevoir la présence de cations divalents stabilisés à l'interface du dimère, principalement des ions  $Mg^{2+}$  ou  $Ca^{2+}$ . Ces cations sont présents dans une large cavité, stabilisés par une couronne de molécules d'eau, qui sont elles-mêmes stabilisées par un grand nombre d'acides aminés acides (acides glutamique et aspartique). C'est notamment le cas de la structure cristallisée de l'enzyme homologue à P28GST chez l'Homme, la hPGDS (PDB ID : 1IYH) (95).



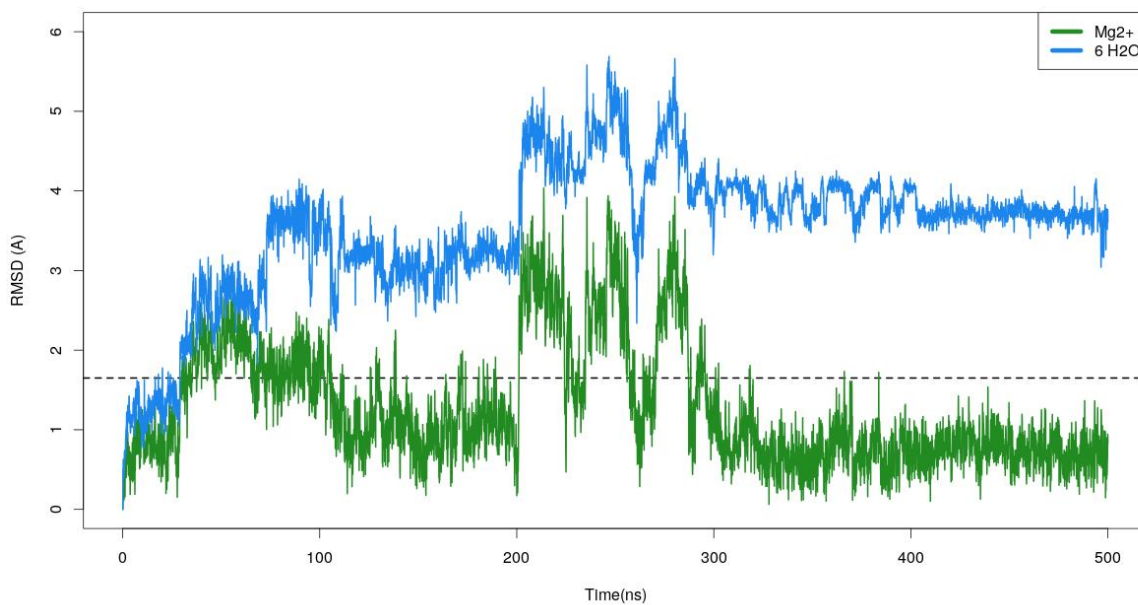
*Figure 61. Représentation chez hPGDS (PDB ID : 1IYH) (95) du cation Mg<sup>2+</sup> (en vert) au centre de sa couronne de molécules d'eau, stabilisées par des acides glutamique et aspartique (en jaune) par des liaisons hydrogènes (en cyan), au niveau de l'interface dimérique et à proximité du glutathion (en bleu).*

Il a été prouvé que la présence de Mg<sup>2+</sup> ou de Ca<sup>2+</sup> permet d'augmenter l'activité générale de hPGDS jusqu'à +150% de son niveau basal de manière concentration dépendante, avec une EC50 (concentration efficace médiane) de 400µM pour Ca<sup>2+</sup> et de 50µM pour Mg<sup>2+</sup>, et que la présence de Mg<sup>2+</sup> augmente l'affinité pour le glutathion en diminuant le Km par 4 (95). Ce phénomène a pu être mesuré au laboratoire lors d'une manipulation en duplicat visant à mesurer l'impact de l'ajout de Mg<sup>2+</sup> sur l'activité GST de hPGDS et de P28GST, en utilisant le protocole décrit précédemment (Tableau 7). On peut observer une augmentation aux alentours de 150% de l'activité sans cation pour hPGDS. Toutefois, cet effet est présent mais limité chez P28GST, avec une augmentation de l'activité enzymatique aux alentours de 20%.

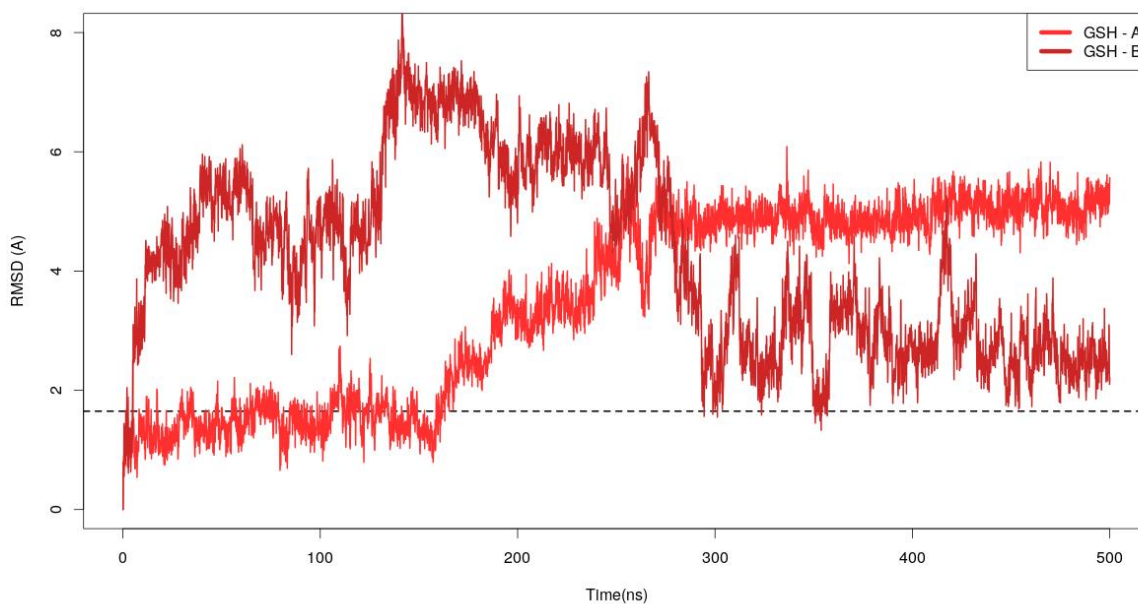
Condition	Activité moyenne GST nmol/min/mg protéines	% relatif
hPGDS seul	3,17	
hPGDS + 1mM Mg <sup>2+</sup>	7,62	+140,4%
hPGDS + 2mM Mg <sup>2+</sup>	8,22	+159,3%
P28GST seule	16,92	
P28GST + 1mM Mg <sup>2+</sup>	20,92	+23,6%
P28GST + 2mM Mg <sup>2+</sup>	19,60	+15,8%

*Tableau 7. Valeurs d'activité moyennes de l'activité GST (en nmol/min/mg de protéine) pour hPGDS et P28GST en l'absence ou en présence de cation Mg<sup>2+</sup>, et augmentation de l'activité relative par comparaison avec l'enzyme sans cation.*

Afin d'évaluer la possibilité de la présence d'un ion Mg<sup>2+</sup>, une première simulation par dynamique moléculaire classique a été réalisée pour mesurer sa stabilité. A partir d'une structure reconstruite et minimisée de P28GST, un ion Mg<sup>2+</sup> avec ses 6 molécules d'eau a été ajouté virtuellement par superposition de cette structure à la structure cristallographique de hPGDS (Identifiant PDB : 1IYH) (95). Après minimisation, mise en température et équilibration du système, en utilisant le protocole expérimental des dynamiques moléculaires classiques, une phase de production a été réalisée afin d'obtenir une trajectoire de 50ns. Il a pu être observé par la mesure de son RMSD au cours de cette simulation que l'ion Mg<sup>2+</sup> était stable, pouvant indiquer sa présence (Figure 62). De plus, cette simulation a également pu mettre en évidence dans les 50ns de simulation l'ouverture de manière transitoire du site L, permettant d'accueillir le substrat PGH2, ainsi qu'un potentiel effet stabilisant sur le glutathion, celui-ci restant au contact de la protéine pendant toute la simulation (Figure 63).

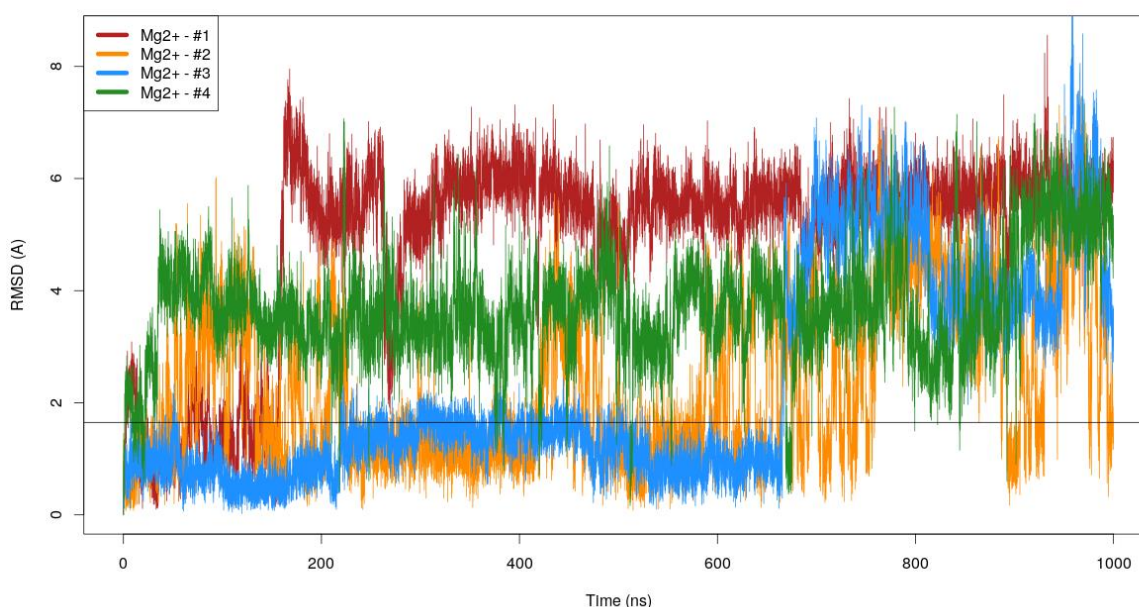


**Figure 62.** *Évolution du RMSD du cation  $Mg^{2+}$  et de sa couronne de 6 molécules d'eau lors de la courte simulation de dynamique moléculaire de 50ns.*

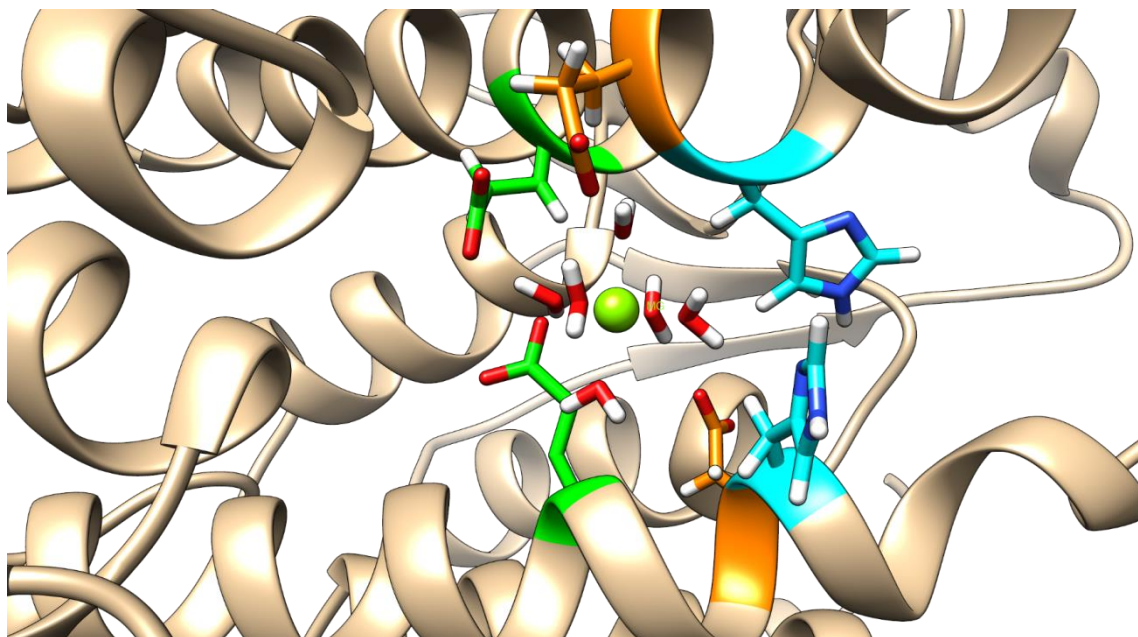


**Figure 63.** *Évolution du RMSD des molécules de glutathion des monomères A et B lors de la courte simulation de dynamique moléculaire de 50ns*

Afin de confirmer les observations réalisées, 4 répliques de simulation de dynamique moléculaire classique de 1000ns ont été réalisées, en utilisant le protocole expérimental des dynamiques moléculaires classiques. De la même manière, en étudiant le RMSD du  $Mg^{2+}$  au cours des simulations réalisées, celui-ci reste stable lors de la simulation, avec trois conformations possibles : la conformation de départ, une conformation intermédiaire avec un RMSD de 4 Å, et une conformation finale avec un RMSD de 6 Å (Figure 64). A la fin des 1000ns de simulation, toutes les répliques convergent vers la conformation finale. Cette conformation est située au milieu de l'interface dimérique, et positionnée autour des acides aminés Glu103, Glu106 et His107 (Figure 65).



**Figure 64.** *Évolution du RMSD du cation  $Mg^{2+}$  lors des 4 répliques de 1000ns de dynamique moléculaire classique.*



**Figure 65.** Représentation de la conformation finale de la P28GST obtenue lors des dynamiques moléculaires classiques de 1000ns P28GST, avec le cation  $Mg^{2+}$  représenté au milieu de sa couronne de 6 molécules d'eau, Glu103 en vert, Glu106 en orange et His107 en cyan.

La combinaison des caractéristiques des enzymes de la même famille que P28GST, des observations de l'effet *in vitro* d'un ajout de  $Mg^{2+}$  et de la stabilité *in silico* de cet ion mesuré par dynamique moléculaire indique la possibilité d'un cation divalent présent au niveau de l'interface dimérique ayant un effet activateur sur les activités enzymatiques. Une validation expérimentale est donc nécessaire afin de mesurer sa présence dans les échantillons de P28GST. Nous avons ici envisagé deux méthodes.



## **2. Recherche de cations divalents par ICP-MS**

La recherche d'un cation divalent au sein de la P28GST a été réalisée par l'équipe du Pr. Fabrice NEIERS du Centre des Sciences du Goût et de l'Alimentation de Dijon, à l'aide d'une analyse élémentaire par ICP-MS (spectrométrie de masse par plasma à couplage inductif) permettant de doser les métaux même à des concentrations très faibles (avec une limite de détection de l'ordre du femtomolaire dans les conditions idéales).

La préparation de l'échantillon de P28GST ayant servi à démontrer une différence d'activité en présence ou en l'absence de cation a été produite par *Saccharomyces cerevisiae* dans un milieu Sacer III complet, contenant les cations divalents  $Mg^{2+}$ ,  $Ca^{2+}$ ,  $Zn^{2+}$ ,  $Mn^{2+}$  et  $Cu^{2+}$ . Nous supposons qu'un de ces cations ait été piégé lors de la production de la molécule, et puisse avoir un effet sur l'activité enzymatique de la P28GST.

La préparation de cet échantillon a été réalisée sur place, avec dans un premier temps une dialyse de l'échantillon avec deux bains de 2 litres de tampon Tris-HCl 50 mM à pH 8.0, puis la mesure de la concentration de l'échantillon à partir du coefficient d'extinction molaire déduit de la séquence, pour enfin ajustement de la concentration à 30 $\mu$ M. Lors de cette analyse, les cations divalents  $Cr^{2+}$ ,  $Mn^{2+}$ ,  $Zn^{2+}$ ,  $Se^{2+}$ ,  $Cu^{2+}$ ,  $Ni^{2+}$  et  $Mg^{2+}$  ont été recherchés dans l'échantillon, avec au préalable le passage d'une gamme de calibration, et pour vérification l'analyse des tampons de dialyse et différents contrôles qualité. La méthode étant basée sur un protocole d'analyses médicales, l'ion  $Ca^{2+}$  n'a pas pu être recherché. Les résultats obtenus sont présentés dans le tableau 8 ci-dessous.

Cation dosé	Type d'échantillon	Concentration ( $\mu\text{M}$ )	% relatif
$\text{Mg}^{2+}$	P28GST 30 $\mu\text{M}$ - A	<0,000	0,00
	P28GST 30 $\mu\text{M}$ - B	<0,000	0,00
	Tampon - A	<0,000	0,00
	Tampon - B	<0,000	0,00
$\text{Ni}^{2+}$	P28GST 30 $\mu\text{M}$ - A	0,20	0,67
	P28GST 30 $\mu\text{M}$ - B	0,24	0,80
	Tampon - A	0,02	0,07
	Tampon - B	0,05	0,17
$\text{Cu}^{2+}$	P28GST 30 $\mu\text{M}$ - A	0,09	0,30
	P28GST 30 $\mu\text{M}$ - B	0,09	0,30
	Tampon - A	<0,000	0,00
	Tampon - B	<0,000	0,00
$\text{Se}^{2+}$	P28GST 30 $\mu\text{M}$ - A	0,00	0,00
	P28GST 30 $\mu\text{M}$ - B	0,01	0,03
	Tampon - A	<0,000	0,00
	Tampon - B	0,01	0,03
$\text{Zn}^{2+}$	P28GST 30 $\mu\text{M}$ - A	1,64	5,47
	P28GST 30 $\mu\text{M}$ - B	1,64	5,47
	Tampon - A	0,01	0,03
	Tampon - B	0,01	0,03
$\text{Mn}^{2+}$	P28GST 30 $\mu\text{M}$ - A	0,02	0,07
	P28GST 30 $\mu\text{M}$ - B	0,02	0,07
	Tampon - A	<0,000	0,00
	Tampon - B	<0,000	0,00
$\text{Cr}^{2+}$	P28GST 30 $\mu\text{M}$ - A	0,10	0,33
	P28GST 30 $\mu\text{M}$ - B	0,09	0,30
	Tampon - A	0,12	0,40
	Tampon - B	0,14	0,47

*Tableau 8. Concentrations mesurées et quantité relative des différents cations recherchés dans l'échantillon de P28GST par ICP-MS, réalisé par l'équipe du Pr. Fabrice NEIERS du Centre des Sciences du Goût et de l'Alimentation de Dijon.*

Parmi les cations divalents dosés, on ne trouve pas de  $\text{Mg}^{2+}$  comme présent chez d'autres enzymes de la même famille. Toutefois, on retrouve ici 5% de cation  $\text{Zn}^{2+}$ . Après étude bibliographique, le cation  $\text{Zn}^{2+}$  peut être stabilisé au sein d'une protéine par des structures en "doigts de zinc", qui sont des motifs structuraux retrouvés pour ordonner en complexe un ou plusieurs  $\text{Zn}^{2+}$  (141,142). Ici, une possibilité de positionnement de ce cation serait au niveau de l'interface dimérique, au même endroit que le cation dans la conformation finale de  $\text{Mg}^{2+}$  observée lors des dynamiques moléculaires de 1000ns, avec la participation de His107 et

Glu103 afin de former un homologue de doigt de zinc (142). De plus, la présence de seulement 5% de  $Zn^{2+}$  peut s'expliquer par le fait que le motif structural présent chez P28GST n'est pas le plus optimal pour stabiliser ce genre de cation. Enfin, ce motif n'est pas retrouvé dans la structure cristallographique, car l'ion n'est pas présent dans le milieu à une concentration suffisante permettant l'observation de celui-ci.

Il est également à noter que chez hPGDS, les histidines sont remplacées par des sérines, participant à la stabilisation de la coordination des 6 molécules d'eau autour de l'ion  $Mg^{2+}$ , et les acides glutamiques sont conservés.

### **3. Recherche *in silico* du positionnement du cation $Zn^{2+}$**

#### ***a. Dynamiques moléculaires classiques analysées par SINAPs***

En réutilisant les données obtenues par l'utilisation de SINAPs sur les 3 répliques de dynamiques moléculaires classiques de 1.5 $\mu$ s, permettant de comparer les réseaux d'interactions moléculaires réalisées en fonction de la position de la tyrosine catalytique, deux liaisons ioniques sont retrouvées entre Arg16 et Glu103, ainsi que Arg16 et Glu106 (Figure 10 de l'article SINAPs) (Tableau 9).

<b>Liaison ionique</b>	<b>Tyr10 - IN</b>	<b>Tyr10 - OUT</b>
<b>Arg16 - Glu103</b>	75,0%	73,4%
<b>Arg16 - Glu106</b>	86,2%	70,1%

***Tableau 9. Fréquences observées pour les liaisons ioniques entre Arg16 et Glu103 ou Glu106 en fonction de la conformation de la tyrosine catalytique en utilisant SINAPs sur les dynamiques moléculaires classiques de 1.5  $\mu$ s.***

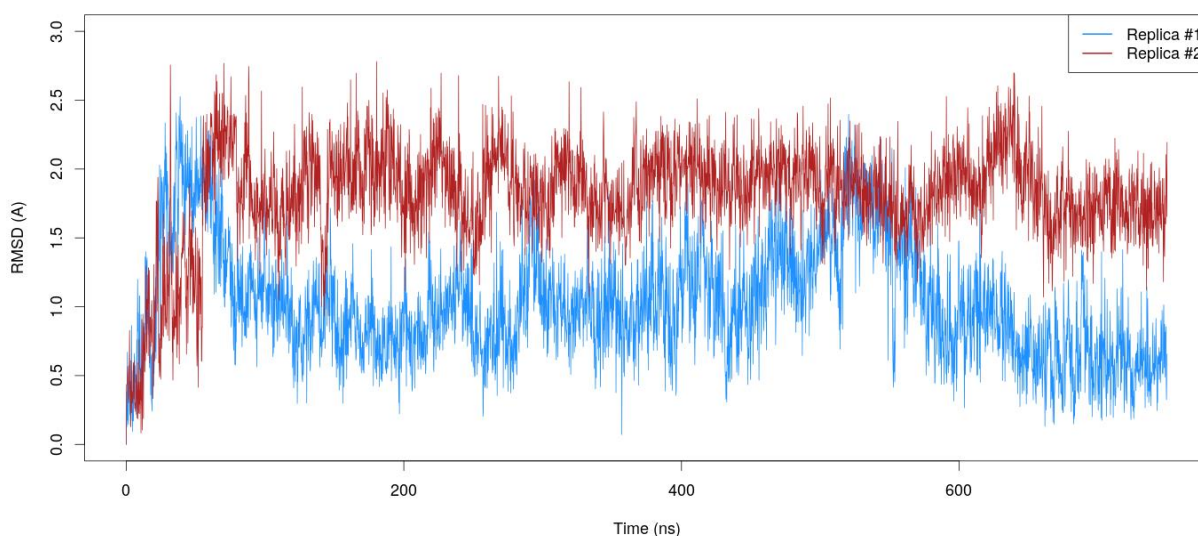
A ce niveau, on peut observer une légère différence au niveau du réseau d'interaction, avec une différence d'occurrence de ces liaisons en fonction de la position de la Tyr10 catalytique. La présence d'un cation  $Zn^{2+}$  stabilisé dans cet environnement permettrait de diminuer la fréquence de ces interactions en réalisant des interactions avec les acides glutamique. L'arginine étant moins stabilisée, la boucle catalytique serait plus flexible, ce qui augmenterait par conséquent les mouvements de la tyrosine catalytique, et donc les activités enzymatiques.

De plus, ces deux interactions ioniques vont également former le "sol" de la poche acceptant la PGH2, et il existe une liaison hydrogène entre Glu106 et His169 bloquant l'ouverture de la poche ayant une fréquence de 12% d'après les résultats de SINAPs. De la même manière, la dérivation de Glu103 pour stabiliser  $Zn^{2+}$  peut également causer l'augmentation de la fréquence d'ouverture de la poche en augmentant la fréquence d'observation de l'interaction entre Glu106 et His169.

### ***b. Dynamiques moléculaires classiques en présence de $Zn^{2+}$***

Afin d'étudier les modalités de liaison du cation  $Zn^{2+}$ , de nouvelles dynamiques moléculaires classiques ont été réalisées. Deux simulations de 750 ns ont été menées en utilisant le même protocole que précédemment. La structure de départ a été obtenue à partir de l'image équilibrée des dynamiques moléculaires classiques réalisées dans le chapitre 1, à laquelle a été ajouté un cation  $Zn^{2+}$  au niveau de l'interface dimérique, aux mêmes coordonnées que le cation  $Mg^{2+}$  retrouvé dans hPGDS après superposition avec la structure de P28GST.

Dans les systèmes simulés, la stabilité du cation  $Zn^{2+}$  a été évaluée par l'étude de son RMSD au cours du temps (Figure 66). Ici, deux configurations semblent se distinguer, avec une première stabilisation du  $Zn^{2+}$  aux alentours de 1 Å de sa conformation d'origine, et une seconde stabilisation aux alentours de 2 Å, mais la stabilisation du cation est similaire, avec les mêmes partenaires protéiques.



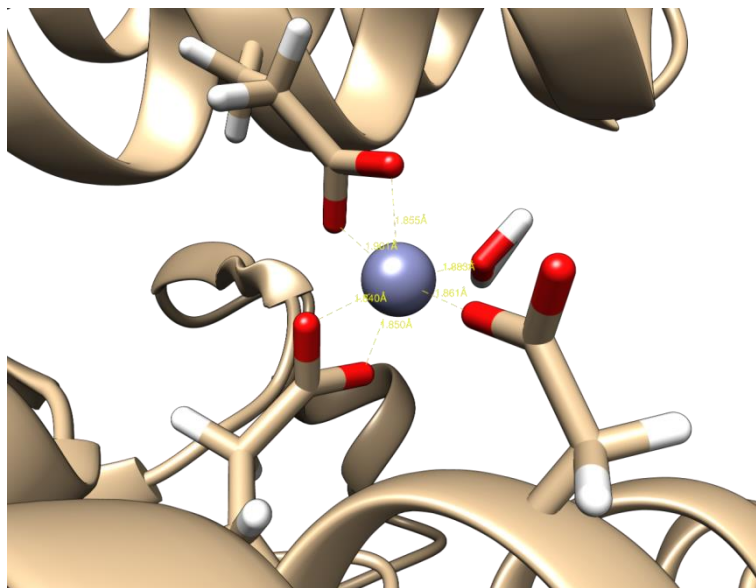
***Figure 66. Evolution du RMSD du  $Zn^{2+}$  au cours des simulations de dynamique moléculaire classique de 750 ns.***

Le positionnement du cation  $Zn^{2+}$  au niveau de l'interface dimérique évolue au cours du temps des simulations, en passant d'une stabilisation par une configuration tétraédrique avec 4 points de contacts, puis une configuration bipyramidale trigonale avec 5 points de contacts,

puis enfin par une configuration octaédrique avec 6 points de contact à partir de 100 ns de simulation. A la différence de ce qui était attendu, avec la participation des Glu103 et His107 des deux monomères pour former un homologue de doigt de zinc, le cation  $Zn^{2+}$  n'est ici stabilisé que par les acides Glu103 et Asp104 des deux monomères, avec la participation d'une seule molécule d'eau. Dans la configuration octaédrique observée dans les deux répliques, le cation est stabilisé par (Figure 67) :

- Les deux atomes d'oxygène de Glu103 du monomère A,
- Les deux atomes d'oxygène de Glu103 du monomère B,
- Un atome d'oxygène de Asp104 du monomère A,
- Un atome d'oxygène d'une molécule d'eau.

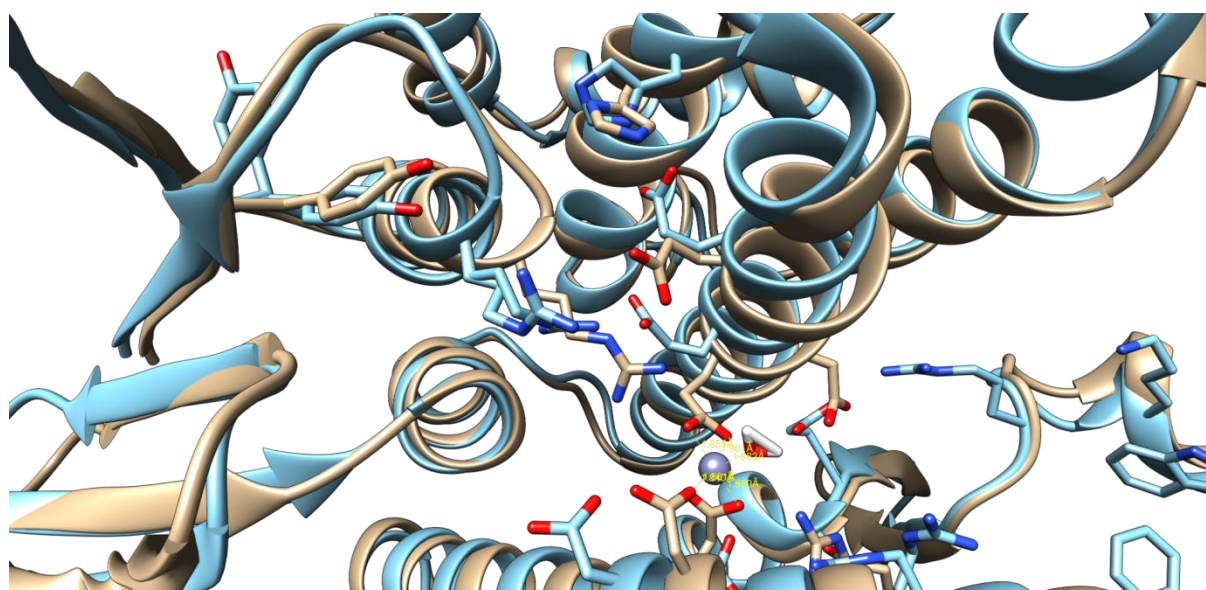
Ces trois configurations observées sont décrites dans la littérature comme des géométries majeures de coordination du cation  $Zn^{2+}$  retrouvées dans une majorité de metalloprotéines à zinc (143).



**Figure 67.** *Positionnement du cation  $Zn^{2+}$  à la fin des simulations de dynamique moléculaire classique.*

Le positionnement asymétrique du cation  $Zn^{2+}$  a des conséquences sur le monomère A, notamment au niveau des structures secondaires et du positionnement des chaînes latérales composant le site actif. En étudiant le RMSF des carbones alpha des différents résidus au sein des deux répliques de simulation, la présence du cation  $Zn^{2+}$  n'entraîne pas de différence notable de flexibilité du squelette carboné entre les deux monomères, ni entre les simulations avec et sans cation. L'architecture globale de la P28GST n'est donc pas impactée par la

présence de ce cation. Toutefois, des différences notables peuvent être observées en comparant les trajectoires de dynamique moléculaire avec la structure cristallographique (Figure 68). Au niveau du monomère A, le résidu Arg16 de la boucle catalytique a tendance à être attiré par le cluster de résidus acides Glu103 et Asp104 stabilisant le cation divalent à l'interface. Cette Arg16 interagit également avec Glu106, permettant de détourner l'interaction que celle-ci réalise généralement avec His169, permettant à ce dernier résidu de pouvoir changer de configuration, pour obtenir une rotation du cycle aromatique proche de celle observée lors de la recherche du site L. Toutefois, la poche du site L n'est pas visible, car la boucle catalytique est déportée vers le domaine hélicoïdal avec le déplacement de Arg16 au niveau de l'interface dimérique ainsi que le déplacement de Arg11 vers le domaine hélicoïdal. La non-visibilité de ce site peut également être expliquée par la non-stabilité du glutathion lors des simulations.



*Figure 68. Impact de la présence d'un  $Zn^{2+}$  par comparaison de la structure cristallographique de P28GST (PDB ID : 1OE8) (en bleu) d'une image issue de dynamique moléculaire en présence de  $Zn^{2+}$  dans une configuration octaédrique (en beige).*

SINAPs a également été utilisé pour comparer les réseaux d'interactions moléculaires entre une réplique réalisée en présence de  $Zn^{2+}$  et une réplique de dynamique moléculaire classique de 1.5 $\mu$ s sans cation (Tableau 10).

Interactions	Sans Zn <sup>2+</sup> (MD 1.5 µs)	Avec Zn <sup>2+</sup>
Arg16 - Glu103.A	75,0%	83,3%
Arg16 - Glu103.B	5,1%	15,1%
Arg16 - Asp104.B	6,6%	82,4%
Arg16 - Glu106	86,2%	99,8%
Glu106 - His169	12,9%	6,0%

*Tableau 10. Fréquences de certaines liaisons ayant été impactées par la présence d'un cation Zn<sup>2+</sup>, obtenues via SINAPs.*

Du fait de l'interaction des acides aminés Glu103 et Asp104 avec l'ion Zn<sup>2+</sup> au niveau de l'interface dimérique, les fréquences des liaisons ioniques réalisées entre ces partenaires et Arg16 sont augmentées significativement, passant respectivement de 75.0%, 5.1% et 6.6% à 83.3%, 15.1% et 82.4% avec Glu103 du monomère A, Glu103 du monomère B et Asp104 du monomère B. Ces interactions vont avoir pour conséquence de privilégier une interaction très stable entre Arg16 et Glu106, pour laquelle une fréquence de 99.8% est observée. Par conséquence, Glu106 interagit moins avec His169, avec une fréquence d'interaction diminuée de moitié en présence de Zn<sup>2+</sup>, lui permettant de changer de configuration plus librement. His169 ayant été décrit comme un acide aminé clé pour l'ouverture du site L permettant la liaison de la prostaglandine H2, la présence de ce cation Zn<sup>2+</sup> pourrait entraîner une augmentation de la fréquence d'ouverture du site L, et une potentielle augmentation de l'activité spécifique PGDS.





**CHAPITRE III**

**RECHERCHE D'INHIBITEURS SELECTIFS DE**

**L'ACTIVITE PGDS**



Un des angles d'attaque permettant de vérifier *in vitro* les observations réalisées lors des différentes dynamiques moléculaires est l'utilisation de sondes pharmacologiques spécifiques des activités enzymatiques de la P28GST. Du fait du fonctionnement enzymatique des glutathion S-transférases de la classe sigma, il est impossible de dissocier l'activité glutathion S-transférase des autres activités. Les inhibiteurs ciblant cette activité seront donc des inhibiteurs non-sélectifs. Toutefois, il est possible d'agir sur l'activité prostaglandine D synthase en ciblant le site de liaison de la prostaglandine H2. Les inhibiteurs ciblant l'activité PGDS seront donc des inhibiteurs sélectifs. En identifiant un inhibiteur non-sélectif et un inhibiteur sélectif, il est donc possible de fournir des outils pharmacologiques permettant d'identifier l'implication de chaque activité et des propriétés épitopiques dans l'effet anti-inflammatoire global à l'aide d'un plan d'expérience (Tableau 11). L'utilisation d'un inhibiteur non sélectif permet d'identifier l'implication des propriétés épitopiques seules, l'utilisation d'un inhibiteur sélectif l'implication de l'activité PGDS seule, et la différence entre les deux permet de déterminer l'implication de l'activité GST.

		P28GST	P28GST & Non-selective inhibitor	P28GST & Selective inhibitor
Epitopic activity		✓	✓	✓
Enzymatic activities	GST	✓	✗	✓
	PGDS	✓	✗	✗

**Tableau 11. Effets des inhibiteurs non sélectifs et sélectifs sur les activités enzymatiques de la P28GST.**

Toutefois, les conditions nécessaires à la réalisation d'un criblage virtuel à haut débit ne sont pas remplies. Afin de trouver un inhibiteur sélectif de l'activité PGDS, il est nécessaire d'identifier le site de liaison de la prostaglandine H2 et d'identifier une conformation de la P28GST dans laquelle sera réalisé le criblage virtuel. Ensuite, plusieurs chimiothèques devront être sélectionnées et filtrées en fonction de critères physicochimiques et pharmacocinétiques. Enfin, après la réalisation des criblages virtuels, les résultats seront analysés et filtrés en fonction de nouveaux critères jusqu'à l'identification de molécules susceptibles d'inhiber l'activité PGDS.

Les différentes étapes préalables à la réalisation de ce criblage virtuel sont donc :

1. L'identification du site de liaison de la prostaglandine H<sub>2</sub>,
2. L'identification d'une conformation de la P28GST,
3. La sélection des chimiothèques de composés qui seront criblée virtuellement.

# **I. Recherche du mode de liaison de la prostaglandine**

## **H2**

### **1. Matériels et méthodes**

#### ***a. Principe de l'amarrage moléculaire et du criblage virtuel***

L'amarrage moléculaire, couramment appelé “*docking*”, est une technique permettant de prédire *in silico* les interactions entre une protéine cible et un ligand d'intérêt au sein d'un site de liaison lors de la formation d'un complexe (144–146). L'application de cette technique à plus grande échelle avec l'utilisation de chimiothèques virtuelles permet la réalisation de criblages virtuels à haut débit, ou “*High-Throughput Virtual Screening*” (144,147). A partir d'une conformation prédéterminée d'une protéine, en connaissant au préalable le site de liaison de ligands de référence, et en utilisant de grandes bases de données de petites molécules pouvant comprendre des milliers à des millions de composés différents, un sous-ensemble de molécules pertinentes peut être mis en évidence. La détermination de ce sous-ensemble permet de réduire drastiquement le nombre de composés à tester expérimentalement, permettant de réduire le temps et le coût des expérimentations *in vitro*.

#### ***b. Algorithme génétique & GOLD***

Au cours de l'amarrage moléculaire, le logiciel cherche à optimiser la fonction de score en utilisant une méthode d'optimisation. Dans le cadre du logiciel CCDC GOLD, un algorithme génétique est utilisé.

L'algorithme génétique va au préalable définir aléatoirement une ou plusieurs populations d'orientations possibles du ligand étudié. Chaque orientation, appelée “chromosome”, contiendra comme information le mappage des atomes des liaisons hydrogènes réalisées entre le ligand et la protéine, le mappage des points hydrophobes du ligand contre ceux de la protéine, et la conformation autour des liaisons flexibles du ligand et des groupes OH de la protéine (148). Chaque chromosome est associé à un score calculé à partir de la fonction de score utilisée, permettant de classer tous les chromosomes en fonction de leur affinité de liaison prédite.

Par la suite, de manière itérative pendant un nombre de pas prédéterminé, la population de chromosomes est optimisée de manière itérative par la réalisation de mutations ponctuelles sur un chromosome, de croisements entre deux chromosomes parents, ou de migrations de chromosomes d'autres populations. La sélection des chromosomes parents pour ces opérations est orientée vers ceux possédant un meilleur score. Lors du processus, une pression de sélection est appliquée, afin de permettre une convergence des populations vers des solutions optimales. La pression de sélection est définie comme le rapport entre la probabilité qu'un chromosome possédant un grand score soit sélectionné comme parent et la probabilité qu'un chromosome possédant un score moyen voire faible soit sélectionné comme parent. De plus, même si la sélection de l'opération génétique est aléatoire, des poids sont attribués pour réguler les fréquences relatives des trois types d'opérations pouvant se produire.

Pour préserver une diversité au sein des populations, il est possible de définir des niches regroupant un maximum défini de chromosomes proches. Deux chromosomes partagent la même niche si le RMSD entre les coordonnées de leurs atomes donneurs et accepteurs est inférieur à 1 Å.

A la fin de chaque opération génétique, les chromosomes enfants générés remplacent les plus mauvais membres de la population. S'il existe des niches au sein de la population, et que la niche d'un chromosome enfant a déjà atteint son maximum défini, celui-ci remplacera le pire élément de la niche au lieu du pire élément de la population.

Par défaut, le logiciel réalise 100 000 opérations génétiques dans 5 populations de 100 chromosomes. Ces populations sont composées de niches ne comprenant pas plus de 2 éléments par niche. La pression de sélection est fixée à 1.1 pour éviter une convergence trop rapide. Enfin, les poids des croisements, mutations et migrations sont respectivement fixés à 95, 95 et 10, permettant d'avoir autant de croisements et de mutations, mais 10 fois moins de migrations.

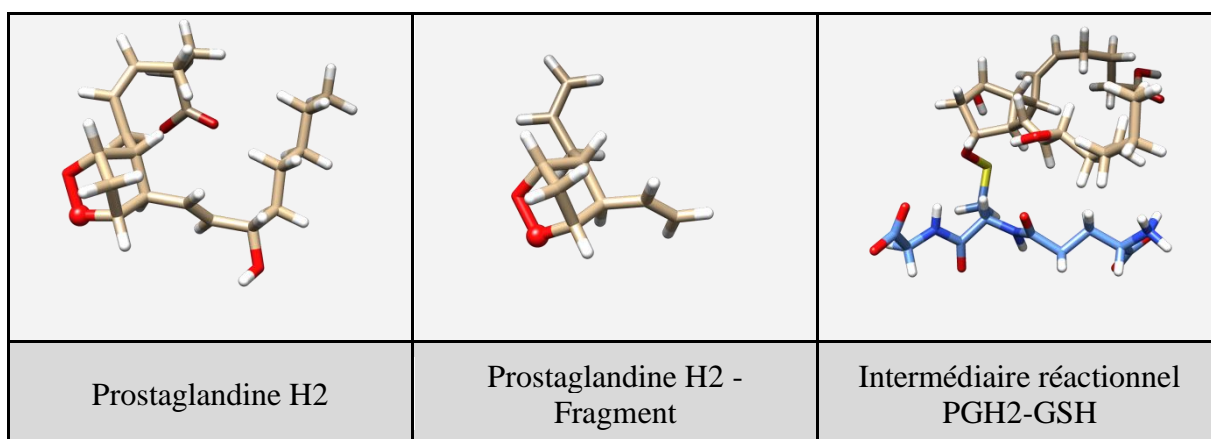
A la fin du processus global, le chromosome présentant le meilleur score parmi toutes les populations est gardé, devenant une pose d'amarrage moléculaire. L'obtention de plusieurs poses d'amarrage moléculaire est ainsi la conséquence de plusieurs algorithmes génétiques lancés indépendamment.

***c. Amarrage moléculaire pour identifier le mode de liaison de la prostaglandine H2***

Afin d'étudier le mode de liaison de la prostaglandine H2 dans la P28GST, des amarrages moléculaires ont été réalisés pour 3 molécules :

- La prostaglandine H2 (Tableau 12 - gauche), en amarrage libre ou avec une contrainte de distance imposée entre l'atome de soufre du glutathion et l'oxygène 2 du peroxyde de la prostaglandine H2,
- Un fragment de la prostaglandine H2, où les deux chaînes carbonées de celle-ci ont été coupées pour ne laisser que deux carbones de chaque côté (Tableau 12 - milieu), en amarrage libre ou avec une contrainte de distance imposée entre l'atome de soufre du glutathion et l'oxygène 2 du peroxyde du fragment de la prostaglandine H2,
- L'intermédiaire réactionnel entre le glutathion et la prostaglandine H2, où ces composés sont liés covalentement (Tableau 12 - droite), en amarrage libre en ayant enlevé au préalable le glutathion du site de liaison.

En comparaison, l'amarrage moléculaire de l'intermédiaire réactionnel a également été réalisé chez hPGDS.



***Tableau 12. Différents ligands utilisés lors des protocoles d'amarrage moléculaire. L'oxygène 2 est représenté sous la forme d'une sphère***



Les amarrages moléculaires ont été réalisés à l'aide du logiciel CCDC GOLD (105) :

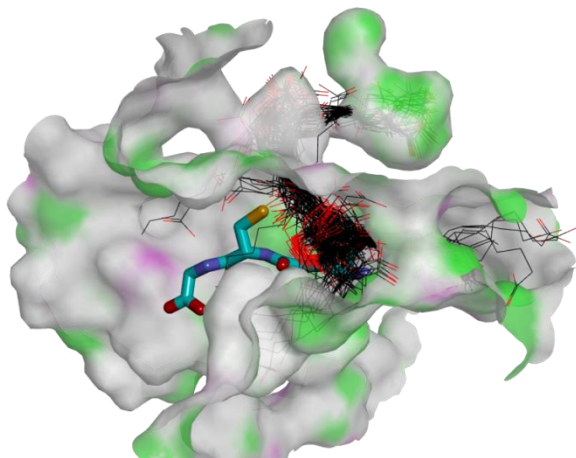
- Soit librement dans une sphère de 20 Å autour de l'atome d'oxygène de la tyrosine 10,
- Soit en imposant une contrainte de distance entre l'atome de soufre du glutathion et l'atome d'oxygène 2 de la prostaglandine H2 ou du fragment de la prostaglandine H2 (Tableau 12 - Sphères rouges).

La structure de départ utilisée a été la structure cristallographique de P28GST (identifiant PDB : 1OE8) ayant été préparée lors de la réalisation des dynamiques moléculaires classiques.

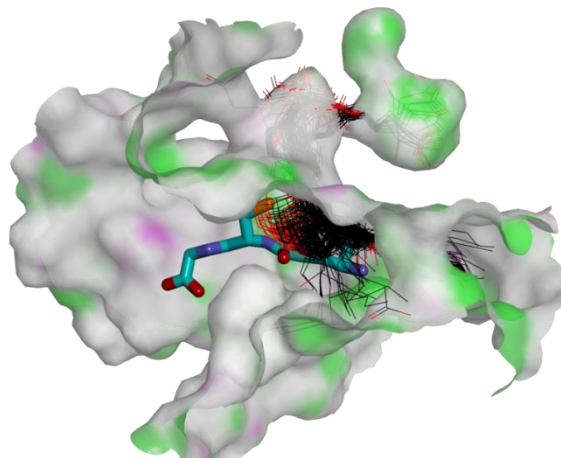
Pour chaque condition, 100 poses d'amarrage ont été générées par le logiciel CCDC GOLD en utilisant la fonction de score GoldScore et l'algorithme génétique réglé automatiquement, permettant un bon échantillonnage conformationnel. Les scores des différentes poses ont ensuite été recalculés en utilisant la fonction de score CHEMPLP.

## 2. Amarrages moléculaires

### *a. Prostaglandine H2*



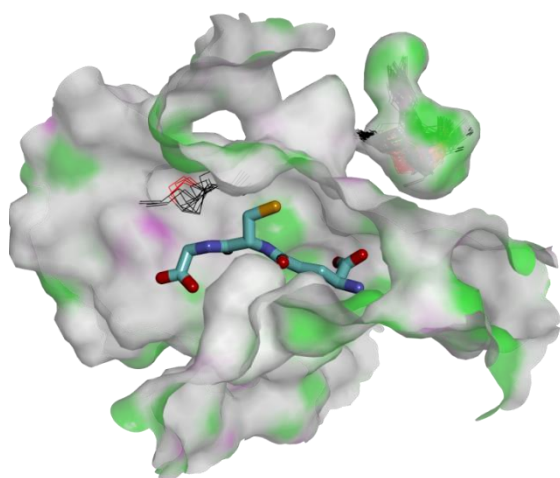
*Figure 69. Positionnement des 100 poses de la prostaglandine H2 libre. Le glutathion est représenté en bleu.*



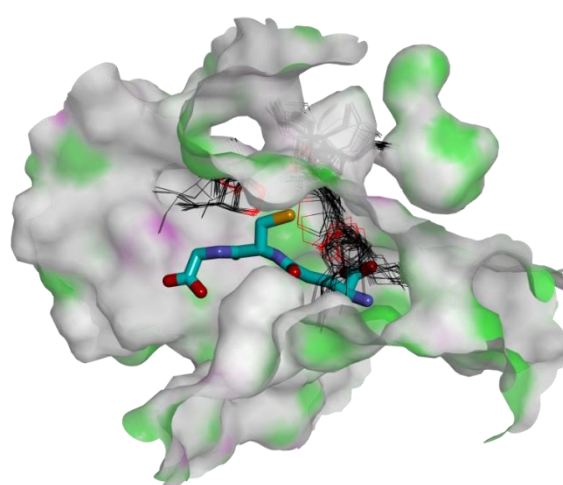
*Figure 70. Positionnement des 100 poses de la prostaglandine H2 sous contrainte. Le glutathion est représenté en bleu.*

Lors de l'amarrage de la prostaglandine H2 avec et sans contraintes de distance, il est possible de voir que certaines poses placent une des chaînes carbonées dans une petite cavité, entre les domaines thiorédoxine et hélicoïdal, malgré le fait qu'il n'existe qu'une très petite ouverture visible lors de l'affichage de la surface accessible au solvant (Figures 69 et 70). De plus, la prostaglandine H2 a également tendance à se placer au niveau de l'interface dimérique.

### *b. Fragment de la prostaglandine H2*



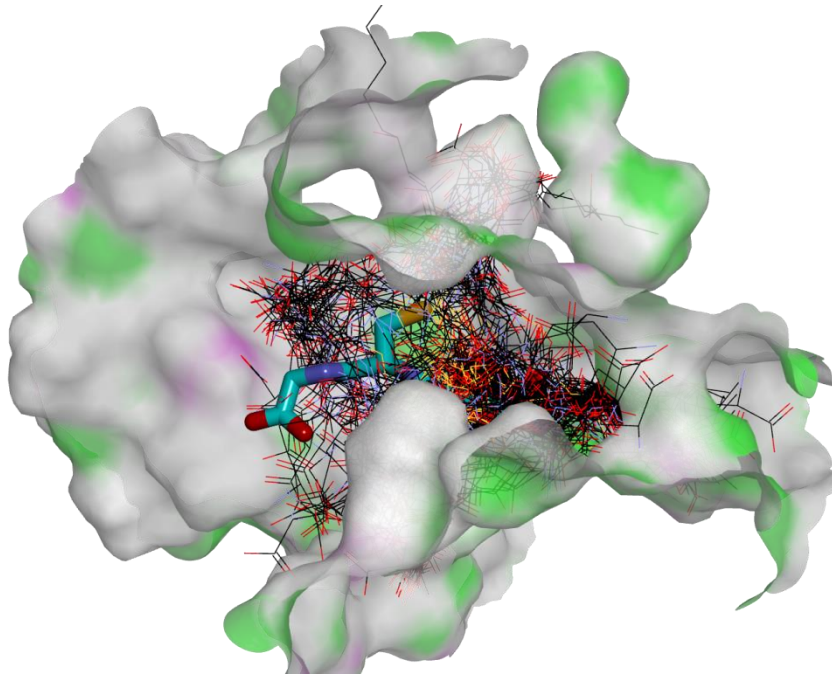
*Figure 71. Positionnement des 100 poses du fragment de la prostaglandine H2 libre. Le glutathion est représenté en bleu.*



*Figure 72. Positionnement des 100 poses du fragment de la prostaglandine H2 sous contrainte. Le glutathion est représenté en bleu.*

Lors de l'amarrage du fragment de la prostaglandine H2 sans contrainte de distance, une très grande majorité des poses seront retrouvées dans la petite cavité observée précédemment (Figure 71). De plus, lors de l'amarrage du fragment de la prostaglandine H2 avec une contrainte de distance, il est possible de trouver un cluster de poses intéressantes visuellement, plaçant un de ses débuts de chaîne latérale en direction de la petite cavité, et l'autre début de chaîne de l'autre côté, vers l'extérieur de la protéine, sans toucher à l'interface dimérique (Figure 72).

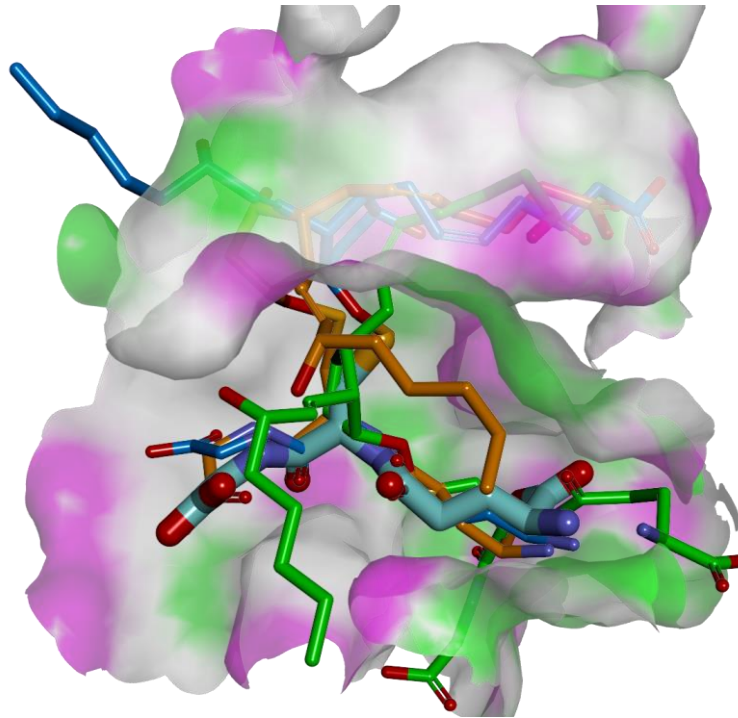
*c. Intermédiaire réactionnel chez P28GST*



*Figure 73. Positionnement des 100 poses de l'intermédiaire réactionnel chez P28GST. Le glutathion est représenté en bleu.*

La qualité du positionnement de l'intermédiaire réactionnel a été évaluée visuellement en recherchant la superposition de la partie glutathion de l'intermédiaire avec la position du glutathion de la structure cristallographique. Ici, aucune pose valide n'a pu être retrouvée dans le site catalytique, et de manière identique à l'amarrage moléculaire de la prostaglandine H2 sans contrainte, l'intermédiaire réactionnel a tendance à se positionner au niveau de l'interface dimérique de la protéine (Figure 73).

*d. Intermédiaire réactionnel chez hPGDS*



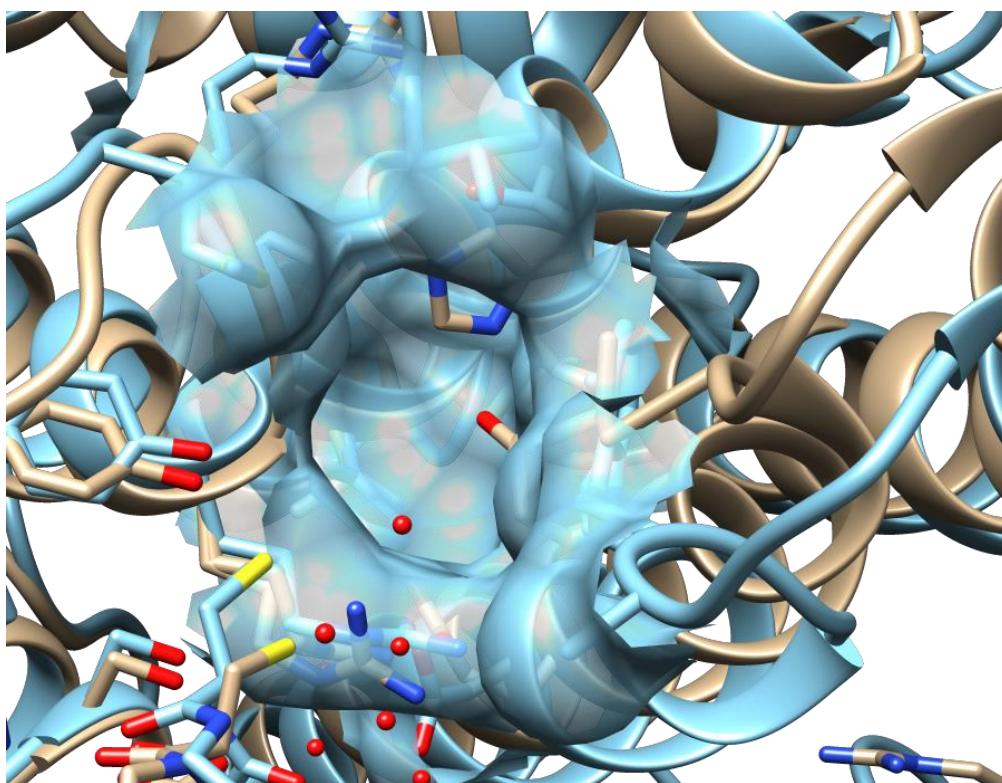
*Figure 74. Positionnement des meilleures poses de l'intermédiaire réactionnel chez hPGDS. Le glutathion est représenté en cyan.*

En comparaison, le protocole d'amarrage moléculaire de l'intermédiaire réactionnel sans contrainte de distance a également été réalisé chez hPGDS. Dans ces conditions, il est possible de placer l'intermédiaire réactionnel dans le site actif de la cristallographie, où 3 poses se superposant avec le glutathion ont pu être retrouvées. De plus, à la place d'une petite cavité, on retrouve ici une poche profonde avec une large ouverture entre les domaines thiorédoxine et hélicoïdal. Enfin, les poses retrouvées ne pénètrent pas l'interface dimérique, mais vont plutôt être stabilisées par la présence de l'ion magnésium et de sa couronne de 6 molécules d'eau agissant sur l'une des extrémités du glutathion.

### *e. Discussion*

Lors des différents amarrages moléculaires de la prostaglandine H2 et de ses dérivés dans la structure cristallographique de P28GST, le site L n'a pas pu être identifié. D'autres amarrages réalisés sur des structures après différentes phases d'équilibrage des dynamiques moléculaires et non présentés ici amènent également à la même conclusion.

Toutefois, des hypothèses existent sur la localisation de ce site L, notamment par la présence d'une petite cavité située entre les domaines thiorédoxine et hélicoïdal. En effet, lors de certaines conditions chez P28GST, il est possible d'y placer une des chaînes latérales de la prostaglandine H2, et chez hPGDS, l'emplacement de cette cavité est remplacé par une large poche décrite dans la littérature comme étant le site L (90) (Figure 75).



*Figure 75. Superposition de P28GST en beige et hPGDS en bleu (et surface)*

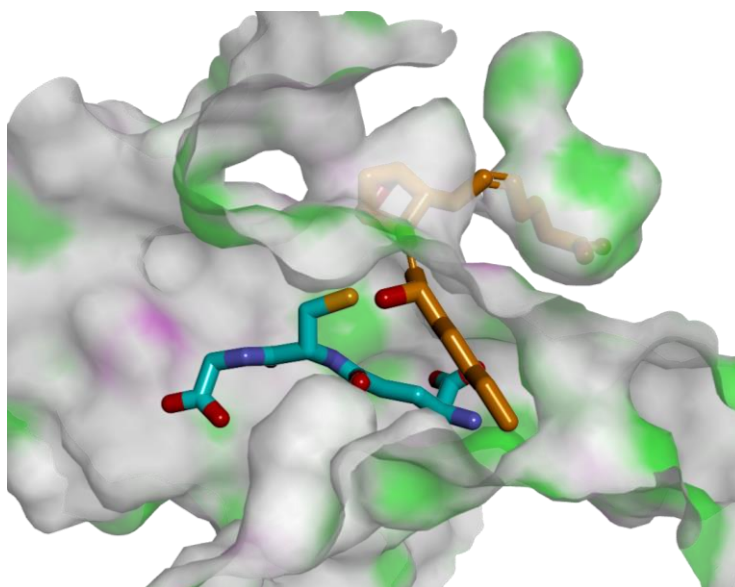
Il est donc nécessaire d'étudier l'existence éventuelle de cette poche chez P28GST. Pour cela, une courte dynamique moléculaire a été réalisée en partant d'une pose d'amarrage moléculaire de la prostaglandine H2 sous contrainte de distance possédant une chaîne

carbonée dans la petite cavité, en ayant pour objectif l'obtention d'une conformation de la poche ouverte, favorable au positionnement de la prostaglandine H2.

En parallèle, l'interface dimérique de la structure cristallographique de P28GST d'être problématique lors du futur criblage virtuel à haut débit. En effet, cette interface étant présente dans la structure cristallographique sous la forme d'une très grande poche vide, et étant accessible dans la sphère de 20 Å définie autour de l'atome d'oxygène de la tyrosine 10, un grand nombre de ligands pourraient avoir tendance à se positionner à cet endroit et donner des faux positifs, comme observé lors des amarrages moléculaires de la prostaglandine H2 et de l'intermédiaire réactionnel chez P28GST. Enfin, l'interface dimérique a déjà été ciblée lors de précédents travaux réalisés dans le cadre de travaux de Master 2, et les molécules qui ont pu être identifiées ne possédaient pas l'effet attendu. La diminution de la taille de la sphère n'étant pas possible, du fait de la présence assez éloignée de la cavité entre les deux domaines, le cation  $Mg^{2+}$  et les 6 molécules d'eau de hPGDS seront placés virtuellement au niveau de l'interface dimérique de P28GST après superposition des structures par UCSF Chimera. La présence de ce cation devrait limiter le nombre de faux positifs en concentrant la recherche des poses d'amarrage dans les zones présumées former le site actif.

### **3. Dynamique moléculaire à partir d'un résultat d'amarrage**

A partir d'une pose d'amarrage de la prostaglandine H2 sous contrainte de distance (Figure 76), une dynamique moléculaire classique de 50 ns a été réalisée en utilisant les mêmes conditions que les dynamiques moléculaires classiques. La topologie de la prostaglandine H2 a été obtenue en utilisant le logiciel ANTECHAMBER. A partir de la trajectoire générée, 50 images ont été extraites, représentant la conformation au bout de chaque nanoseconde. Par sélection visuelle à partir de la surface accessible au solvant proposée par UCSF Chimera, 7 images ont été présélectionnées car elles présentaient une poche ouverte. Enfin, un nouvel amarrage moléculaire de l'intermédiaire réactionnel a été réalisé dans ces images pour obtenir 50 poses d'amarrage, en ayant au préalable ajouté l'ion magnésium et sa couronne de molécules d'eau par superposition de la structure cristallographique de hPGDS.



*Figure 76. Représentation de la pose d'amarrage moléculaire de PGH2 sous contrainte de distance sélectionnée comme point de départ pour une simulation de dynamique moléculaire.*

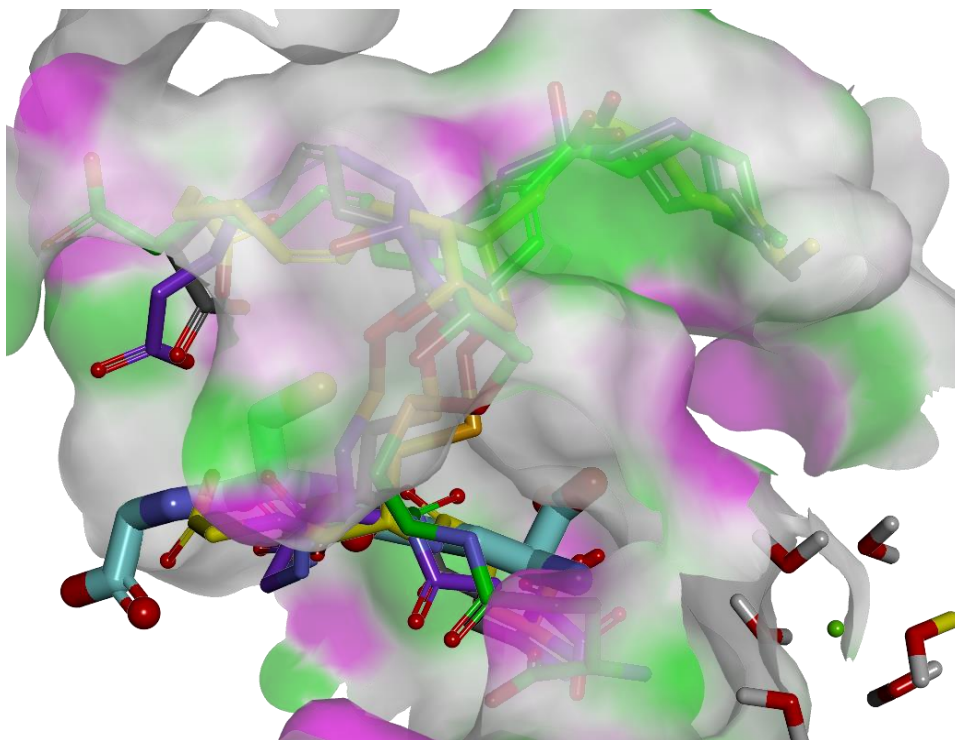
Dans ces conditions, l'intermédiaire réactionnel a pu être positionné de manière convaincante dans 5 des 7 images présélectionnées (Tableau 13). Une image s'est particulièrement distinguée, en permettant de trouver 4 poses satisfaisantes sur les 50 calculées, avec 3 poses possédant les 1ère, 2ème et 4ème plus hauts GoldScore : l'image #29 (Figure 77). Les poses retrouvées sont en adéquation avec les observations réalisées chez hPGDS et avec le fragment de la prostaglandine H2 avec contrainte de distance, et permettent de valider l'hypothèse



formulée pour le site L de la P28GST. Ici, la chaîne possédant la fonction cétone se place dans la poche située entre les domaines thiorédoxine et hélicoïdal, en réalisant une liaison hydrogène conventionnelle ou une liaison hydrogène donneur-pi avec l’histidine 169, tout en étant stabilisé par des acides aminés hydrophobes. La chaîne possédant la fonction acide se place de l’autre côté, vers l’extérieur de la protéine et parallèle au glutathion, en étant stabilisée par des liaisons hydrogènes avec le squelette de l’arginine 14 et de la glycine 15, et par des acides aminés hydrophobes. Enfin, la tête 2,3-dioxabicyclo[2.2.1]heptane de la prostaglandine H2 est maintenue en place par des interactions pi-alkyl formées avec les acides aminés aromatiques à proximité incluant la tyrosine 110 et l’histidine 169.

Frames	Interesting poses	GOLD ranking
Frame #16	0/50 poses	∅
Frame #17	3/50 poses	#1, #37, #46
Frame #22	5/50 poses	#32, #35, #37, #38, #50
Frame #26	3/50 poses	#6, #27, #34
Frame #29	4/50 poses	#1, #2, #4, #25
Frame #31	0/50 poses	∅
Frame #36	2/50 poses	#15, #35

**Tableau 13. Résultats de l’amarrage moléculaire de l’intermédiaire moléculaire dans 7 images de dynamique moléculaire.**



*Figure 77. Représentation des 4 poses sélectionnées de l'intermédiaire réactionnel dans l'image #29*

En parallèle, la pertinence de la fonction de score ChemPLP a été évaluée, et celui-ci présentait de moins bons résultats que le GoldScore, malgré le fait qu'il soit plus rapide. Le GoldScore a donc été choisi comme fonction de score de référence pour la suite des travaux. Enfin, la présence de l'ion magnésium et de ses molécules d'eau a bien limité le nombre de faux positifs à l'interface dimérique, où très peu de poses ont été retrouvées.

En conclusion, la conformation de la P28GST de l'image #29 est utilisable dans le cadre de campagnes de criblage virtuel à haut débit. Ce qui permet la recherche d'inhibiteurs sélectifs de l'activité prostaglandine D2 synthase, en recherchant des molécules pouvant se placer au niveau du site L observé, rendu accessible dans cette conformation.

## II. Réalisation du criblage virtuel à haut débit

### 1. Matériels et méthodes

#### *a. Choix des chimiothèques*

Lors du criblage virtuel, plusieurs chimiothèques ont été étudiées. La chimiothèque principale ayant servi de base aux travaux a été la chimiothèque ZINC15 (149). Celle-ci contient plus de 230 millions de composés prêts à être amarrés et disponibles commercialement. Elle contient également plus de 750 millions de composés permettant de rechercher des composés analogues aux composés testés lors des différentes campagnes d'amarrage moléculaire et de criblage virtuel. L'avantage de la chimiothèque ZINC15 est sa grande diversité structurale, mais son défaut principal est le manque d'informations à jour sur la disponibilité commerciale des composés.

Parmi les 250 millions de composés, une présélection de la chimiothèque a été réalisée pour ne garder dans un premier temps que 12 187 069 composés :

- Possédant une configuration 3D, disponibles dans les formats mol2, sdf et pdbqt, évitant donc la génération des conformations 3D de chaque composé ;
- Etant dans un sous ensemble considéré comme propre, en fournissant un sous-ensemble filtré de l'espace chimique moins enclin à réaliser des artefacts lors de tests *in vitro*, éliminant notamment les PAINS (150,151) ;
- Disponibles commercialement, et étant indiqués comme étant en stock.

Une seconde sélection a ensuite été réalisée en se basant sur des normes proposées au sein des règles de Lipinski décrivant les caractéristiques idéales d'un composé biodisponible. Tous les composés possédant un poids moléculaire compris strictement entre 250 et 500 Da, et possédant un coefficient de partage logP de 3 au maximum, ont été sélectionnés, pour un total de 4 620 118 composés répartis au sein de 2568 tranches de chimiothèque (Figure 78).

Molecular Weight (up to, Daltons)													Totals, by LogP
	200	250	300	325	350	375	400	425	450	500	>500		
-1	6,875	3,540	6,834	3,707	3,443	2,372	1,419	1,097	845	1,234	2,291	20,951	
0	24,646	19,875	31,347	20,032	20,526	12,446	6,375	3,933	2,500	2,533	1,372	99,692	
1	67,433	79,202	145,585	105,750	109,105	69,976	36,733	19,334	12,360	12,083	5,245	510,926	
2	84,930	163,289	316,268	297,320	217,630	219,857	119,412	78,103	55,598	49,825	14,033	1,354,013	
2.5	29,927	91,421	252,499	231,306	198,287	194,090	121,135	89,851	66,816	60,882	17,078	1,214,866	
3	16,833	72,137	232,985	225,703	190,117	265,380	169,649	131,186	102,856	101,794	31,225	1,419,670	
3.5	7,266	44,128	172,333	108,345	155,950	251,907	144,803	169,504	139,360	145,895	54,288	0	
4	2,946	19,883	100,808	123,577	212,223	221,165	235,247	198,142	165,975	107,195	84,037	0	
4.5	419	6,326	35,369	69,990	124,037	158,772	179,311	186,846	186,438	204,311	99,638	0	
5	54	1,842	16,462	28,632	66,374	64,642	107,159	134,119	141,400	193,984	145,864	0	
>5	8	510	5,861	12,500	25,546	78,718	101,280	151,247	168,777	395,008	645,667	0	
Totals, by Weight	0	0	985,518	883,818	739,108	764,121	454,723	323,504	240,975	228,351	0	4,620,118 Protomers 2.5K Tranches	

**Figure 78. Sélection des molécules de la chimiothèque ZINC15**

Afin d'optimiser les calculs du criblage virtuel, GOLD ne fonctionnant que sur un seul processeur à la fois, 41 instances ont été lancées en parallèle pour 41 groupes de molécules. La répartition de l'ensemble des composés au sein des 41 groupes a été réalisée à l'aide d'un script Python3 prenant en compte le nombre de molécules par tranches, afin de répartir ces 2568 tranches au sein de groupes équilibrés possédant en moyenne 110 000 molécules par groupe.

En parallèle, la Chimiothèque Nationale, gérée par ChemBioFrance, rattachée au CNRS, a également été considérée (152). Cette chimiothèque contient plus de 70 000 composés et plus de 15 000 extraits naturels originaux, conditionnés sous une forme standardisée et pouvant être sollicités dans le cadre de collaborations scientifiques. Deux bases de données sont disponibles :

- La collection complète de la Chimiothèque Nationale (CN) comprenant 81 374 composés dont 62 281 composés en plaque au 12 juin 2021, disponibles à des concentrations de 5mM dans du DMSO pour des volumes de 2,5 à 300 nL à des prix très raisonnables.

- La collection de la Chimiothèque Nationale Essentielle (CNE) comprenant 1040 composés reflétant la diversité chimique de la CN, et également disponible dans les mêmes conditions.

L'avantage de la CN et la CNE est la transparence de la disponibilité des produits à des prix avantageux, mais le principal défaut est le manque de diversité structurale.

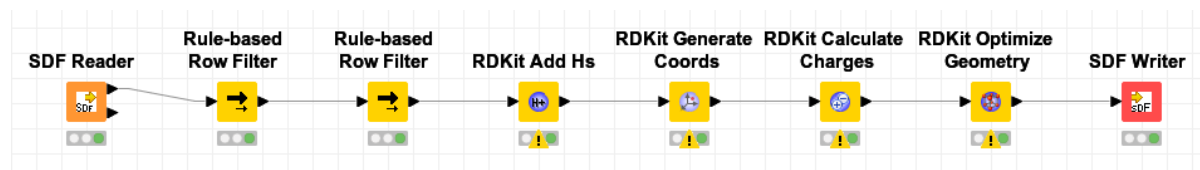
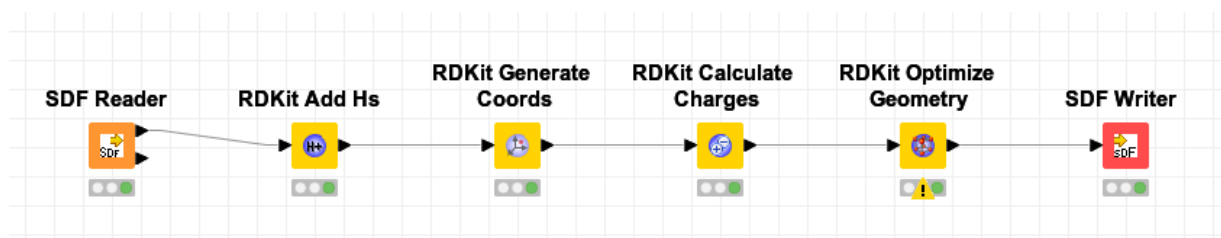


Figure 79. Préparation des molécules de la CN

Les conformations tridimensionnelles des molécules composant la CN ont été obtenues à partir des conformations 2D au format SDF de 78 371 composés proposés par ChemBioFrance en utilisant le logiciel KNIME et les plugins RDKit implantés (Figure 79). Dans un premier temps, l'ensemble des composés a été filtré en fonction de leur poids moléculaire, pour ne garder que les composés compris entre 250 et 500 Da. Puis les composés ont de nouveau été sélectionnés en fonction de leur disponibilité immédiate en utilisant les informations de disponibilité fournies par la chimiothèque. Ensuite, pour chaque molécule, les hydrogènes ont été ajoutés, les coordonnées 3D ont été générées, les charges partielles ont été calculées avec la méthode de calcul de Gasteiger (153), puis la géométrie a été optimisée en utilisant le champ de force *Merck Molecular Force Field* MMFF94 (154) pendant 1000 itérations d'optimisation. Enfin, les 46 420 composés ayant atteint la fin du processus sans erreur ont été enregistrés au format SDF.

De la même manière que pour la chimiothèque ZINC15, l'ensemble des molécules ont été réparties au sein de 21 groupes de criblage virtuel, avec 2200 molécules par groupe.



*Figure 80. Préparation des molécules de la CNE*

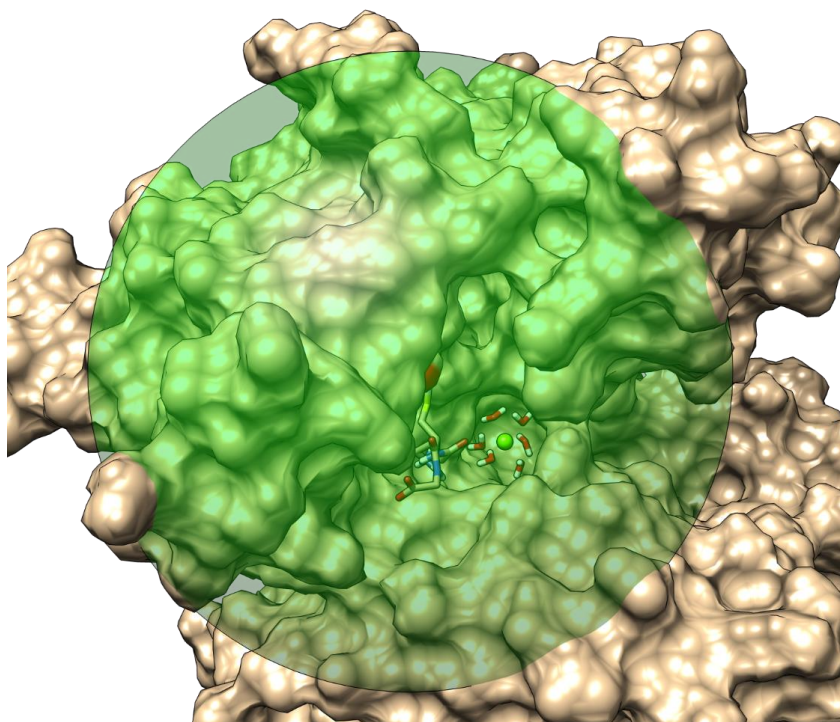
Les conformations tridimensionnelles des molécules composant la chimiothèque CNE ont été obtenues de la même manière que pour la chimiothèque CN, mais sans avoir filtré au préalable les composés (Figure 80). En fin de processus, 1026 molécules ont été obtenues.

### ***b. Protocole de criblage virtuel***

La structure utilisée lors du criblage a été la frame 29 identifiée précédemment, en ayant au préalable ajouté par superposition de structure avec hPGDS le cation magnésium et sa couronne de 6 molécules d'eau au niveau de l'interface dimérique. La zone de criblage a été définie par une sphère de 20 Å autour de l'atome d'oxygène de la tyrosine 10, et la fonction de score utilisée a été le GoldScore, comme définis lors des précédents amarrages moléculaires (Figure 81).

Chaque instance de criblage virtuel réalisée avec le logiciel CCDC GOLD a été lancée en ligne de commande. Pour chaque molécule criblée, le logiciel a réalisé un maximum de 10 000 opérations génétiques dans une seule population de 100 chromosomes, composée de niches ne comprenant pas plus de 2 éléments par niche. La pression de sélection a été fixée à 1.1, le poids des croisements et des mutations ont été fixés à 100, et le poids des migrations a été fixé à 0 du fait de la présence d'une seule population. Un maximum de 5 poses de criblage a été gardé pour chaque composé, mais avec un arrêt précoce du criblage si les 3 meilleures poses de criblage possèdent un RMSD inférieur à 1.5 Å, indiquant la convergence du positionnement du composé.

Ces paramètres ont été définis pour permettre le calcul de 3 à 5 poses de criblage virtuel pour un temps moyen de 20 secondes par molécule criblée, tout en obtenant des résultats d'une qualité satisfaisante.



*Figure 81. Zone de criblage correspondant à une sphère de 20 Å autour de l'atome d'oxygène de la tyrosine 10 (en rouge). Le glutathion est représenté en bâtonnets, avec le cation  $Mg^{2+}$  et ses 6 molécules d'eau au niveau de l'interface dimérique.*

Pour chaque groupe de molécules criblées, plusieurs fichiers de sortie notables sont obtenus, incluant le fichier répertoriant toutes les poses de criblage au format sdf, ainsi qu'un fichier "bestranking.lst" incluant pour chaque molécule le meilleur score obtenu ainsi que le temps de criblage pour chaque molécule.

A l'aide d'un script Python3, le GoldScore maximum pour chaque molécule a été extrait du fichier "bestranking.lst" des 41 groupes de criblage. L'analyse de cette information a ensuite été réalisée à l'aide d'un script R. De manière similaire, avec un second script Python3, tous les résultats de criblage au format sdf ont été parcourus afin d'extraire pour chaque pose (Figure 82) :

- L'identifiant ZINC de la molécule,
- La tranche ZINC dans laquelle est classée la molécule,
- L'identifiant intra-groupe de criblage,
- Le numéro de la pose de criblage, de dock1 à dock5,



- Un identifiant unique créé spécifiquement pour identifier les stéréoisomères séparément (possédant le même identifiant ZINC), en concaténant la tranche ZINC, l'identifiant intra-groupe et l'identifiant ZINC,
- Le score de docking GoldScore et les différentes contributions calculées définissant le GoldScore,
- Le nombre de liaisons hydrogène déterminé par le logiciel de criblage.

A partir de ces nouveaux fichiers de sortie, les différents résultats sont également analysés avec des scripts R.

```
ZINC000091663138 FFAARN.xaa 148 dock1 FFAARN.xaa_148_ZINC000091663138 56.4810 -6.8573 -9.0621 41.0860 0.0000 0.0000 -15.9071 0
ZINC000091663138 FFAARN.xaa 148 dock3 FFAARN.xaa_148_ZINC000091663138 54.3754 -7.9699 -9.1244 37.0257 0.0000 4.6523 -15.9071 1
ZINC000091663138 FFAARN.xaa 148 dock2 FFAARN.xaa_148_ZINC000091663138 53.5485 -9.0775 -8.6588 37.6595 0.0000 3.5958 -15.9071 1
ZINC000091663138 FFAARN.xaa 148 dock4 FFAARN.xaa_148_ZINC000091663138 49.1737 -7.3255 -8.6036 35.3425 0.0000 0.7684 -15.7385 2
ZINC000091663138 FFAARN.xaa 148 dock5 FFAARN.xaa_148_ZINC000091663138 44.3525 -15.1947 -7.6081 36.9415 0.0000 0.6222 -15.7385 2
ZINC000058248447 FFAARN.xaa 149 dock5 FFAARN.xaa_149_ZINC000058248447 60.2801 -28.0642 -10.3014 43.7850 0.0000 0.3341 -38.1072 1
ZINC000058248447 FFAARN.xaa 149 dock4 FFAARN.xaa_149_ZINC000058248447 53.1909 -35.8946 -11.4647 43.0264 0.0000 3.2817 -38.1072 1
ZINC000058248447 FFAARN.xaa 149 dock2 FFAARN.xaa_149_ZINC000058248447 46.4114 -30.0864 -9.6736 34.9558 0.0000 0.0000 -38.1072 0
ZINC000058248447 FFAARN.xaa 149 dock3 FFAARN.xaa_149_ZINC000058248447 42.2085 -37.3755 -9.0610 36.7375 0.0000 0.0237 -38.1072 1
ZINC000058248447 FFAARN.xaa 149 dock1 FFAARN.xaa_149_ZINC000058248447 41.3844 -35.7601 -12.6667 37.6029 0.0000 0.0000 -38.1072 0
ZINC000036002253 FFAARN.xaa 150 dock5 FFAARN.xaa_150_ZINC000036002253 52.4160 -5.6146 -9.6588 39.3659 0.0000 0.0000 -13.5613 0
ZINC000036002253 FFAARN.xaa 150 dock2 FFAARN.xaa_150_ZINC000036002253 49.5027 -3.4377 -11.0166 34.9421 0.0000 1.7977 -14.1139 1
ZINC000036002253 FFAARN.xaa 150 dock1 FFAARN.xaa_150_ZINC000036002253 47.4314 -2.6405 -15.9489 35.2927 0.0000 1.5669 -15.9265 1
ZINC000036002253 FFAARN.xaa 150 dock4 FFAARN.xaa_150_ZINC000036002253 47.3083 -4.8155 -9.4482 34.9169 0.0000 0.0000 -13.5613 0
ZINC000036002253 FFAARN.xaa 150 dock3 FFAARN.xaa_150_ZINC000036002253 44.2470 -7.4567 -12.3725 35.7336 0.0000 0.8286 -14.1139 1
```

**Figure 82. Exemple des fichiers de sortie obtenus à la suite du script Python3 parcourant tous les résultats de criblage**

Enfin, l'étude de la forme des différents composés sélectionnés a été réalisée à partir du calcul des principaux moments d'inertie à l'aide de l'extension Vernalis implémentée dans KNIME.

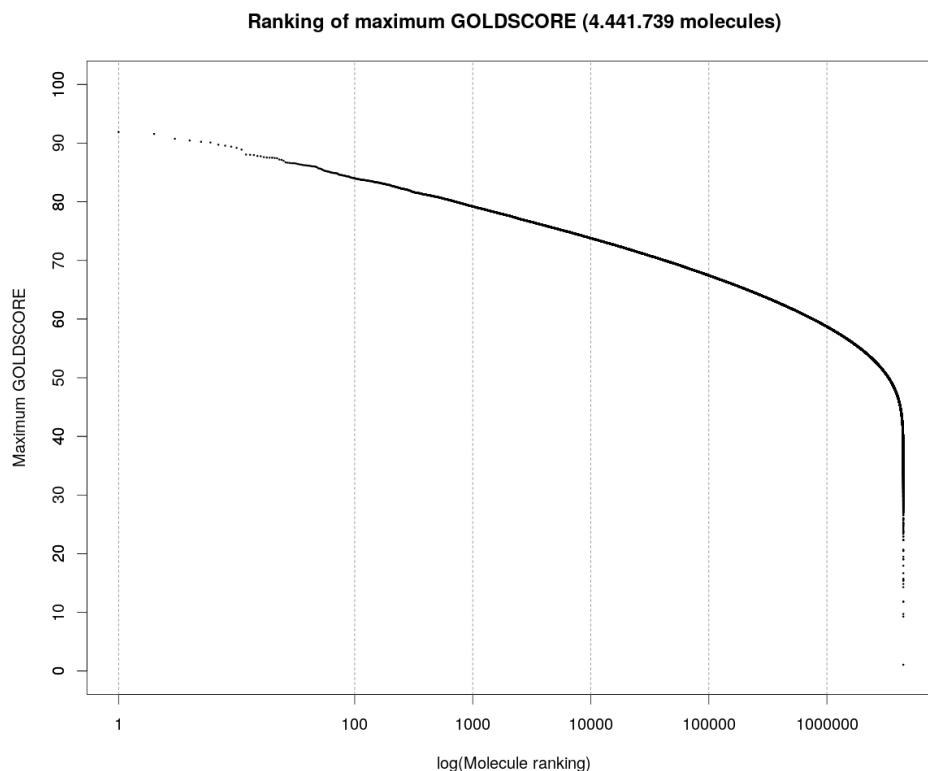
## **2. Résultats du criblage virtuel - ZINC**

Le criblage de la chimiothèque ZINC a été réalisé pour les 4 620 118 molécules sélectionnées précédemment. Au total, 4 441 739 molécules ont pu être criblées avec succès, avec l'obtention de 3, 4 ou 5 poses de criblage, pour un total de 21 845 430 poses de criblage. Le différentiel obtenu entre le nombre théorique et le nombre final de molécules criblées est dû à la présence de doublons et d'erreurs lors du criblage. En effet, quelques molécules ont échappé à la sélection à cause d'un incident technique lié à un manque de place sur le disque dur. Au vu du faible nombre de molécules manquantes, 178 379 molécules soit moins de 4% du total, il a été décidé de ne pas les récupérer et de ne pas les considérer lors de la suite du criblage.

Le principal objectif de ce criblage est de diminuer drastiquement le nombre de molécules, pour passer aux alentours du millier de molécules et faciliter ainsi l'étape d'amarrage moléculaire suivant le criblage virtuel.

### ***a. Répartition des molécules dans le jeu de données entier***

Au sein du jeu de données entier, nous pouvons voir un grand nombre de molécules possédant un haut GoldScore (Figure 83). Afin de diminuer la quantité de molécules, une limite de GoldScore peut être utilisée. En plaçant cette limite pour ne garder que les molécules possédant un GoldScore supérieur à 60, un total de 759 988 molécules peut être séparé du jeu de données global, mais ce nombre est encore trop important pour la suite de l'analyse. Si l'on choisit d'être plus strict et que l'on décide de ne garder que les molécules possédant un GoldScore supérieur à 70, ce nombre diminue à 42 181 molécules, ce qui reste bien trop.



**Répartition des molécules en fonction de leur score de docking maximal**

> 80	80 - 70	70 - 60	60 - 50	50 - 40	< 40
709	41 472	717 807	2 568 060	1 086 964	26 727

**Figure 83. Répartition des molécules en fonction de leur GoldScore maximal - Jeu global**

Le moyen de diminuer dans un premier temps le nombre total de molécules a été de regarder la répartition des molécules en fonction du nombre de poses réalisées (Tableau 14). Ici, 105 026 molécules ne possèdent que 3 poses de criblage, indiquant que l'algorithme a convergé autour d'une unique position pour l'ensemble de ces composés. En étudiant également le nombre de liaisons hydrogène réalisées entre le jeu global et le jeu intermédiaire des composés ne réalisant que 3 poses de criblage, on peut également observer une répartition similaire et un nombre moyen de liaisons hydrogène réalisées par composé proche, passant de 1,50 liaisons hydrogène par composé dans le jeu global à 1,40 liaisons hydrogène par composé dans le jeu de données intermédiaire (Figure 84).

Cette sélection permet donc de ne garder qu'un sous-ensemble comprenant 2,36% du jeu de données global, sans perdre en diversité.

## Répartition des molécules en fonction du nombre de poses réalisées

3 poses	4 poses	5 poses
105 026	102 708	4 223 904

Tableau 14. Répartition des molécules en fonction du nombre de poses réalisées

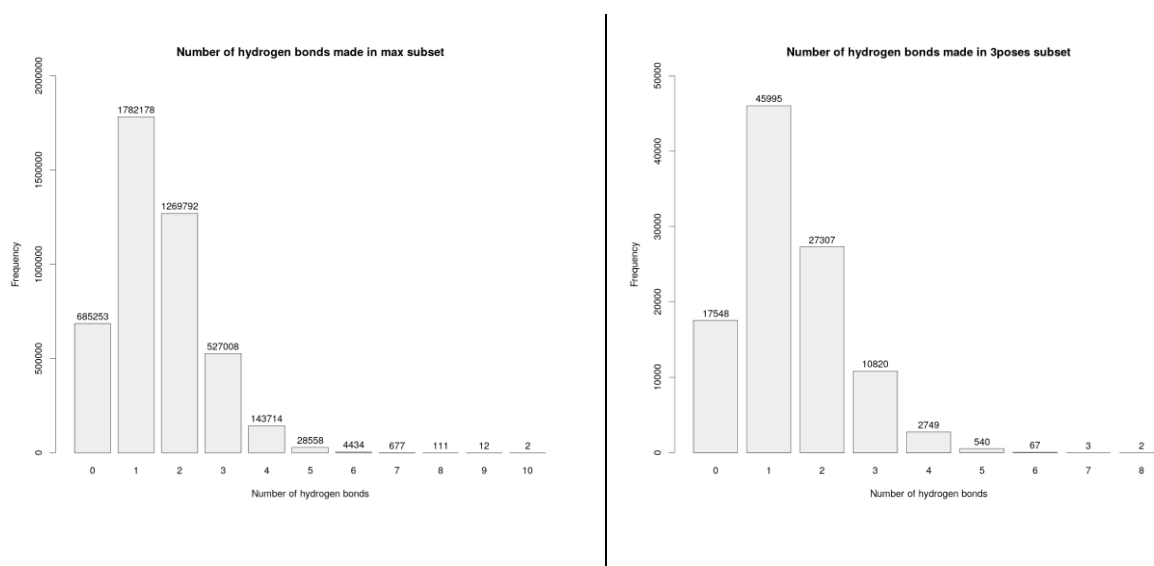
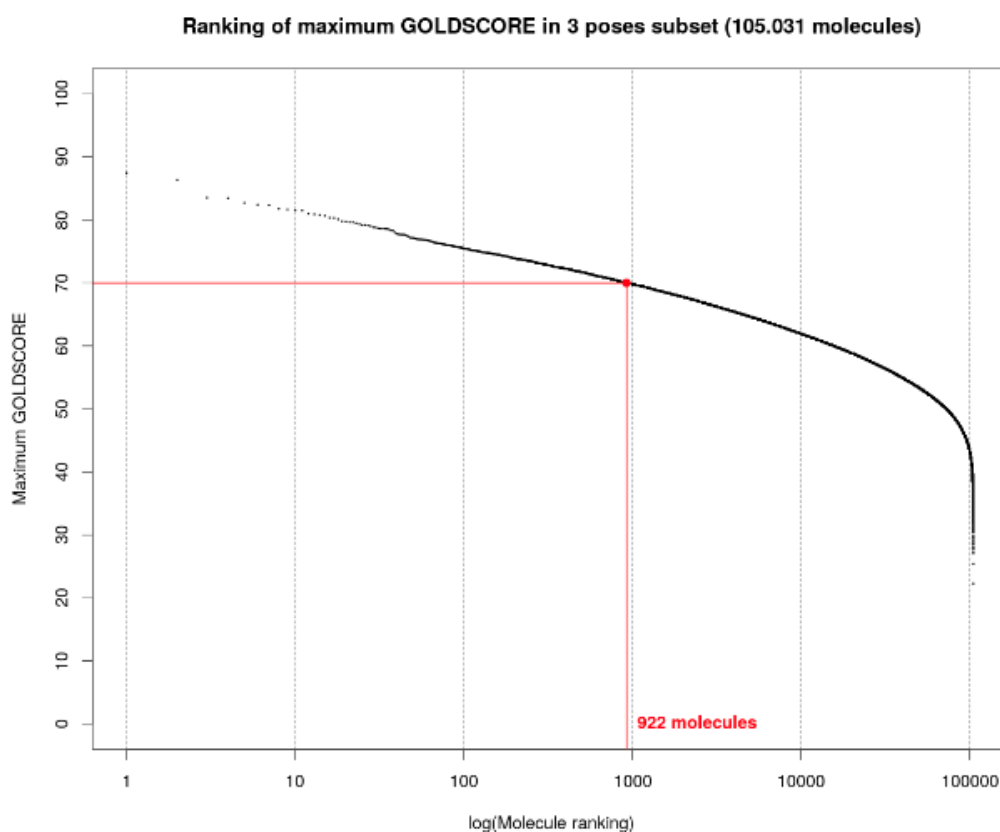


Figure 84. Répartition du nombre de liaisons hydrogènes dans le jeu global de 4 441 739 composés (gauche) et le jeu de 105 026 composés ne réalisant que 3 poses de criblage (droite)

***b. Répartition des molécules dans le jeu de données ne possédant que 3 poses de criblage***

Les 105 026 composés sélectionnés précédemment ont ensuite été soumis à l'analyse de leur GoldScore, comme indiqué plus haut (Figure 85). Ainsi, si l'on ne souhaite garder que les molécules possédant un GoldScore maximal supérieur à 60, un total de 15 610 molécules peut être mis en évidence. De la même manière, pour un GoldScore maximal supérieur à 70, un total de 922 molécules peuvent être mises en évidence. Par rapport à l'objectif initial qui visait à ne garder qu'un millier de composés qui continueront à l'étape d'amarrage moléculaire, les 922 molécules ont été gardées pour la suite du processus d'identification de composés inhibiteurs spécifiques de l'activité PGDS.



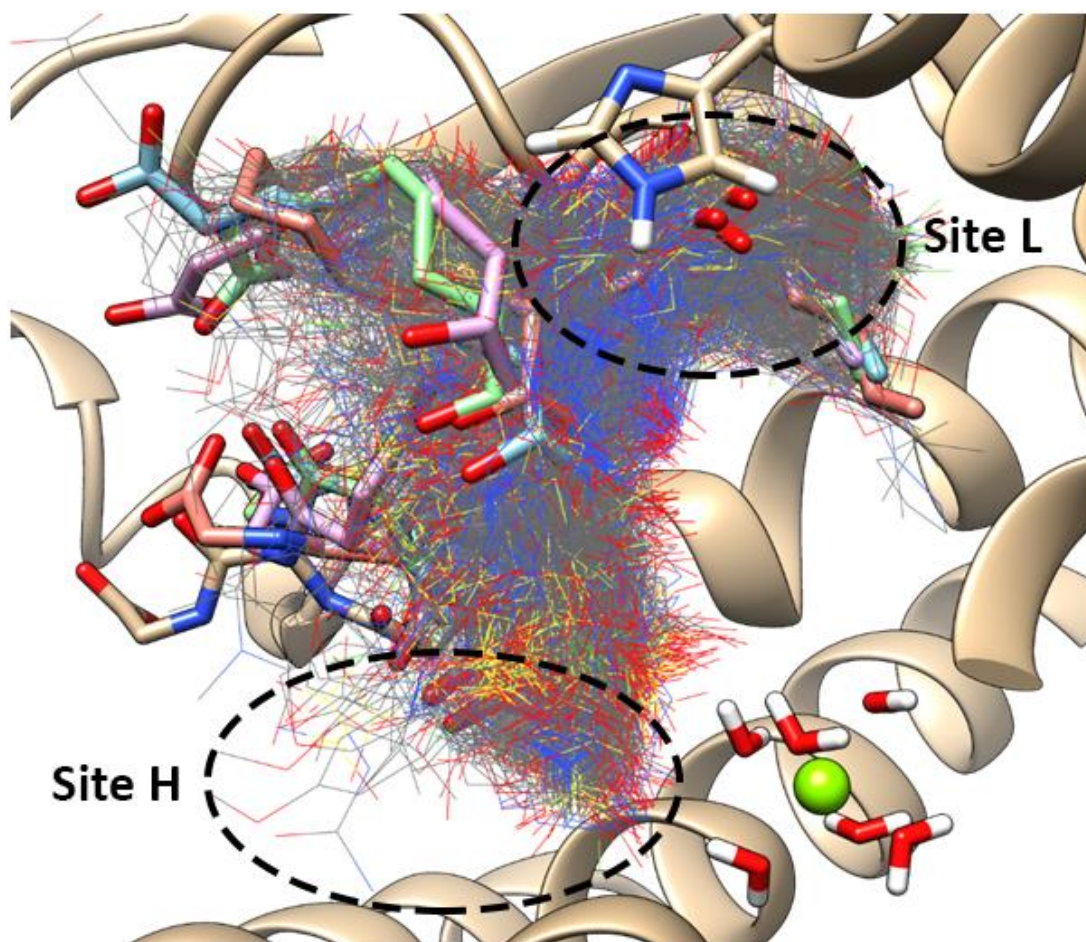
**Répartition des molécules "3 poses" en fonction de leur score de docking maximal**

> 80	80 - 70	70 - 60	60 - 50	50 - 40	< 40
18	904	14 688	57 034	30 978	1 409

**Figure 85. Répartition des molécules en fonction de leur GoldScore maximal - Jeu 3 poses**

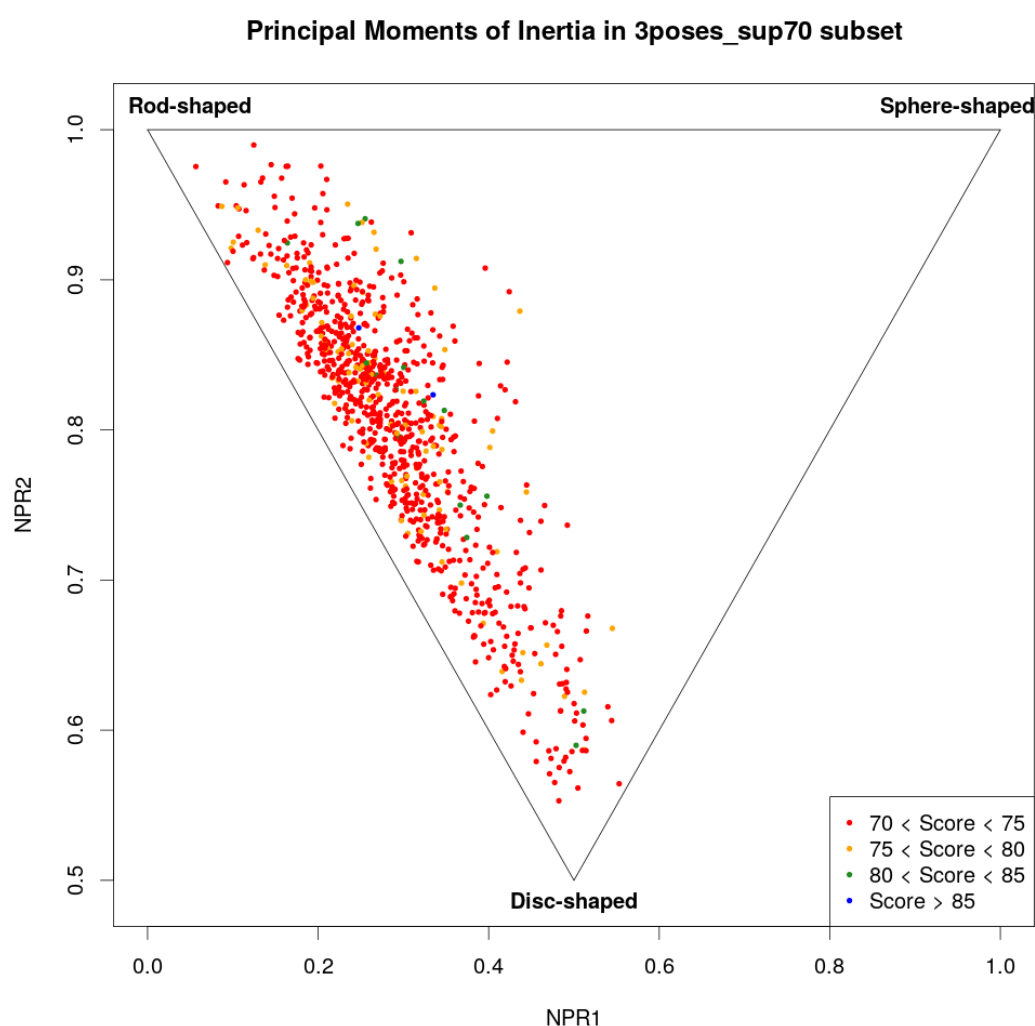
### *c. Jeu de 922 molécules*

Les 922 molécules mises en évidence sont réparties dans 34 groupes sur 41. Leur positionnement au sein du site de liaison est très diversifié (Figure 86). Tout d'abord, les composés ne se positionnent pas au niveau de l'interface dimérique du fait de la présence du cation, qui a bien eu l'effet escompté. De plus, beaucoup de molécules sont retrouvées dans le site L hypothétique, site de liaison privilégié de la prostaglandine H2. Ces composés ont également une bonne superposition avec la partie prostaglandine des poses d'amarrage identifiées précédemment pour l'intermédiaire réactionnel entre la prostaglandine H2 et le glutathion. Toutefois, un cluster de molécules a tendance à se positionner plus bas, au niveau du site H, le site de liaison non spécifique de l'enzyme. Ce positionnement est à prendre en compte lors de processus de sélection plus avancés.



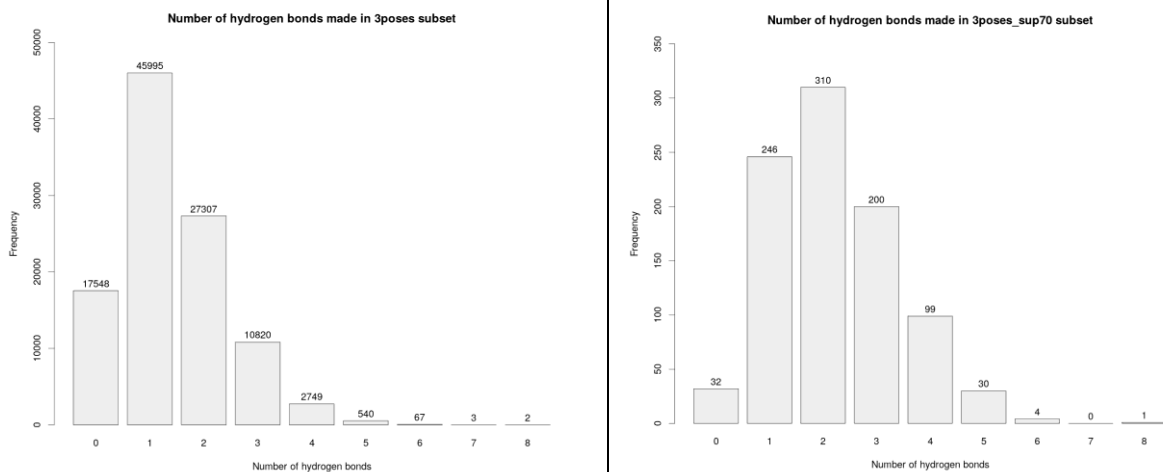
*Figure 86. Positionnement des 922 molécules par rapport aux poses d'amarrage des intermédiaires réactionnels représentés en bâtonnets.*

L'analyse de la forme des molécules a été réalisée par la mesure des principaux moments d'inertie, permettant de placer chaque composé entre une forme linéaire, une forme plane et une forme sphérique (Figure 87). Les 922 composés mis en évidence vont avoir des formes principalement situées entre linéarité et planéité. Il est également à noter que les molécules possédant des scores plus élevés auront tendance à être plus linéaires, et à se placer préférentiellement dans le coin supérieur gauche du graphique des principaux moments d'inertie.



**Figure 87. Principaux moments d'inertie des 922 molécules**

De plus, la répartition et le nombre moyen de liaisons hydrogènes réalisées entre le jeu intermédiaire de 105 026 composés et le jeu de 922 molécules est différent, avec un passage de 1,40 liaisons hydrogènes réalisées pour les molécules du jeu intermédiaire à 2,22 liaisons hydrogènes pour les 922 composés.



**Figure 88. Répartition du nombre de liaisons hydrogènes dans le jeu de 105 026 composés réalisant 3 poses (gauche) et le jeu de 922 molécules (droite)**

Toutefois, en analysant la répartition des gammes de poids moléculaires, obtenues à partir des tranches de la chimiothèque ZINC, pour le jeu intermédiaire et le jeu ayant été sélectionné par une limite de GoldScore, nous pouvons voir que la sélection par le GoldScore a entraîné un biais favorisant les molécules les plus lourdes. En effet, le GoldScore prenant en compte la force de van der Waals, celui-ci est proportionnel au nombre d'atomes composant la molécule qui sont en contact avec la surface de la poche. De plus, le GoldScore ayant été développé à la base pour des poches fermées et non pas pour des interactions de surface, le score de certaines molécules peut être biaisé si celles-ci se positionnent plus vers l'extérieur de la protéine, notamment au niveau du site H (105). Pour compenser ce phénomène, le score a été divisé par le poids moléculaire afin d'obtenir un GoldScore normalisé par unité de masse, appelé dans la suite GoldScore pondéré.

MW (Da)	250-300	300-325	325-350	350-375	375-400	400-425	425-450	450-500
<b>Jeu de départ</b>	933 068	866 501	722 054	720 149	424 482	315 481	237 707	222 297
<b>3 poses</b> 105 031 molécules	32201	16727	22770	15686	7901	4953	2877	1916
	3.451 %	2.628 %	2.317 %	2.178 %	1.861 %	1.570 %	1.210 %	0.862 %
<b>3 poses + score &gt; 70</b> 922 molécules	1	4	18	81	118	222	242	236
	0.003 %	0.018 %	0.108 %	0.516 %	1.493 %	4.482 %	8.412 %	12.317 %

**Tableau 15. Répartition des molécules du jeu de données intermédiaire et du jeu de données de 922 molécules en fonction de leur poids moléculaire**

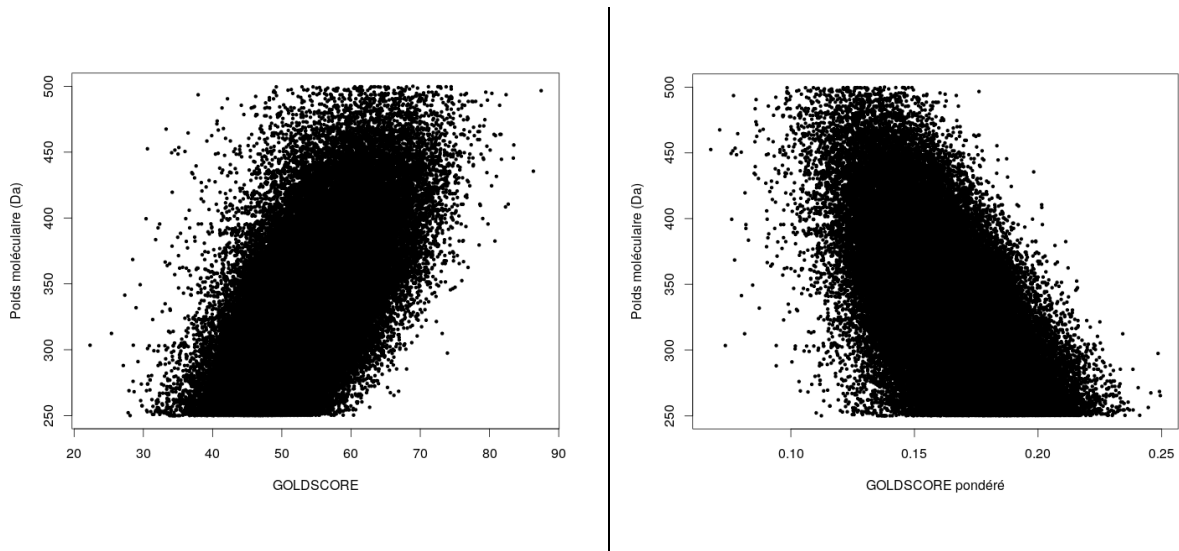


#### *d. Jeu issu du GoldScore pondéré*

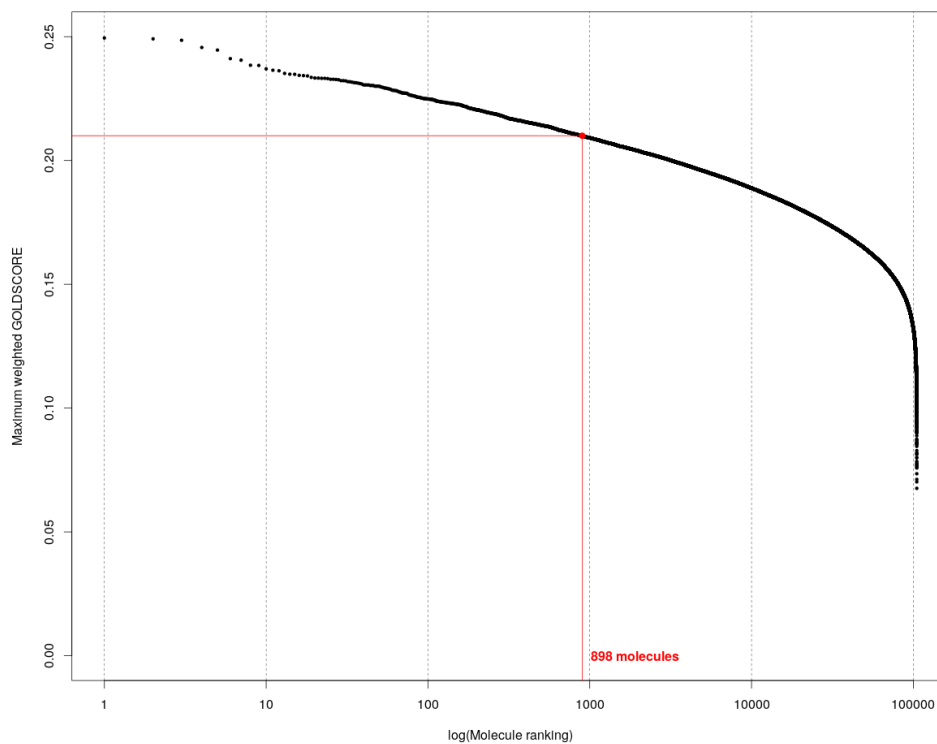
Afin de pouvoir calculer le GoldScore pondéré, il a été nécessaire d'obtenir le poids moléculaire pour les 4 441 739 composés du jeu de données entier. Du fait du très grand nombre de données en entrée dépassant les 20 Go, il est très difficile voire impossible de calculer dans des temps raisonnables le poids moléculaire pour chaque composé à partir de logiciels tiers comme KNIME, ou à l'aide de scripts Python3 ou R utilisant des packages dédiés. La solution qui a été retenue est d'utiliser la requête de la base de données ZINC à partir de l'API, permettant d'obtenir pour chaque composé son poids moléculaire à partir de son identifiant ZINC en utilisant une syntaxe spécifique de l'URL du site internet de la base de données. Une fois les poids moléculaires obtenus, le GoldScore pondéré a été calculé à l'aide d'un script R.

En étudiant le poids moléculaire de chaque composé en fonction du GoldScore, nous pouvons voir que, comme vu précédemment, le GoldScore va bien avoir tendance à favoriser les composés les plus lourds (Figure 89 - gauche). En réalisant la même analyse en étudiant le poids moléculaire de chaque composé en fonction du GoldScore pondéré, nous pouvons voir l'effet inverse, où le GoldScore pondéré va avoir tendance à favoriser des composés plus légers (Figure 89 - droite). Une sélection supplémentaire de molécules sélectionnées à partir du GoldScore pondéré a donc été envisagée pour obtenir une sélection complémentaire au jeu de données déterminé à partir du GoldScore brut, permettant de multiplier les chances de récupérer des composés pertinents. De la même manière que pour le premier groupe, l'objectif qui a été fixé est de récupérer environ un millier de molécules pour ce nouveau groupe.

En s'intéressant à la répartition du GoldScore pondéré en fonction du classement des composés du jeu de données intermédiaire comprenant 105 026 composés, nous pouvons voir qu'il est possible de réaliser la sélection d'un groupe en fonction d'une limite de GoldScore pondéré (Figure 90). En étudiant le nombre de composés pour chaque limite au centième près, la limite permettant d'obtenir aux alentours d'un millier de composés est la limite permettant de ne garder que les composés ayant un GoldScore pondéré supérieur à 0,21, comprenant 898 composés.



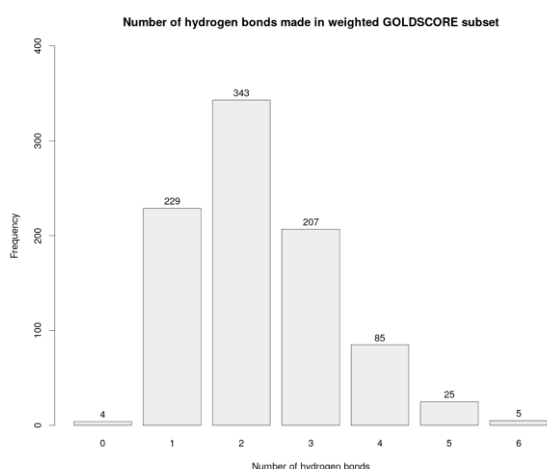
**Figure 89.** Répartition des poids moléculaires (en Da) en fonction du GoldScore (à gauche) et du GoldScore pondéré (à droite)



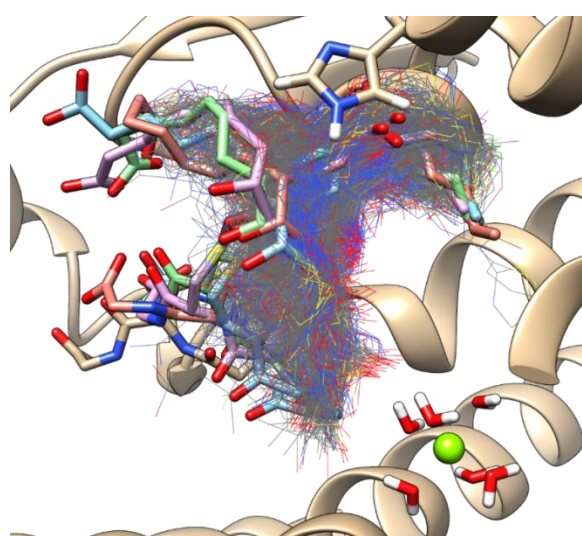
<b>GoldScore pondéré &gt; 0.23</b>	<b>GoldScore pondéré &gt; 0.22</b>	<b>GoldScore pondéré &gt; 0.21</b>	<b>GoldScore pondéré &gt; 0.20</b>	<b>GoldScore pondéré &gt; 0.19</b>
48	215	898	3189	9024

**Figure 90.** Répartition des molécules en fonction de leur GoldScore pondéré maximal - Jeu 3 poses

Ces 898 composés formant le jeu de données GoldScore pondéré vont posséder des caractéristiques similaires aux 922 composés du jeu de données GoldScore. La répartition et le nombre de liaisons hydrogènes réalisées par composé est identique au jeu de données GoldScore, avec une moyenne de 2,26 liaisons hydrogène par molécule (Figure 91). De plus, le positionnement des molécules au sein du site de liaison est également similaire (Figure 92).

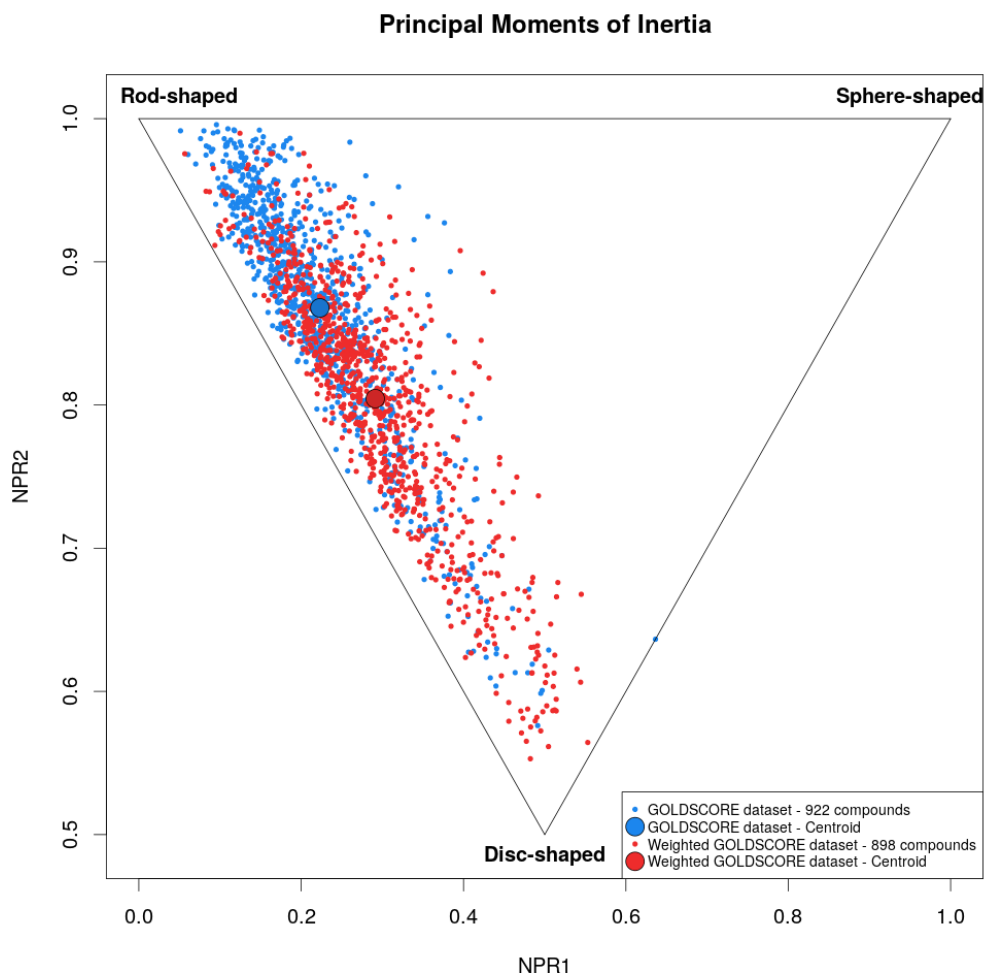


**Figure 91.** Répartition du nombre de liaisons hydrogènes des 898 molécules issues du jeu de données GoldScore pondéré



**Figure 92.** Positionnement des 898 molécules issues du jeu de données GoldScore pondéré dans le site de liaison

Toutefois, il existe une différence notable au niveau de la forme des molécules composant ces différents jeux de données lors de l'analyse des principaux moments d'inertie (Figure 93). En observant la dispersion du nuage de points obtenu, ainsi que la position des barycentres, les molécules présentes dans le jeu de données GoldScore pondéré sont plus linéaires que le premier jeu de données.



**Figure 93. Principaux moments d'inertie des molécules composant les différents jeux de données**

En conclusion, les jeux de données GoldScore, comprenant 922 molécules, et GoldScore pondéré, comprenant 898 molécules, sont deux sélections complémentaires issues de l'ensemble des 105.026 molécules ayant convergées lors de l'étape de criblage virtuel à haut débit. Parmi cet ensemble, 17 sont retrouvées en doublon, et ont été soustraites au jeu de données GoldScore pondéré, pour obtenir un total de 881 molécules. Ces deux jeux de données ont été traités de manière similaire dans la suite des travaux.

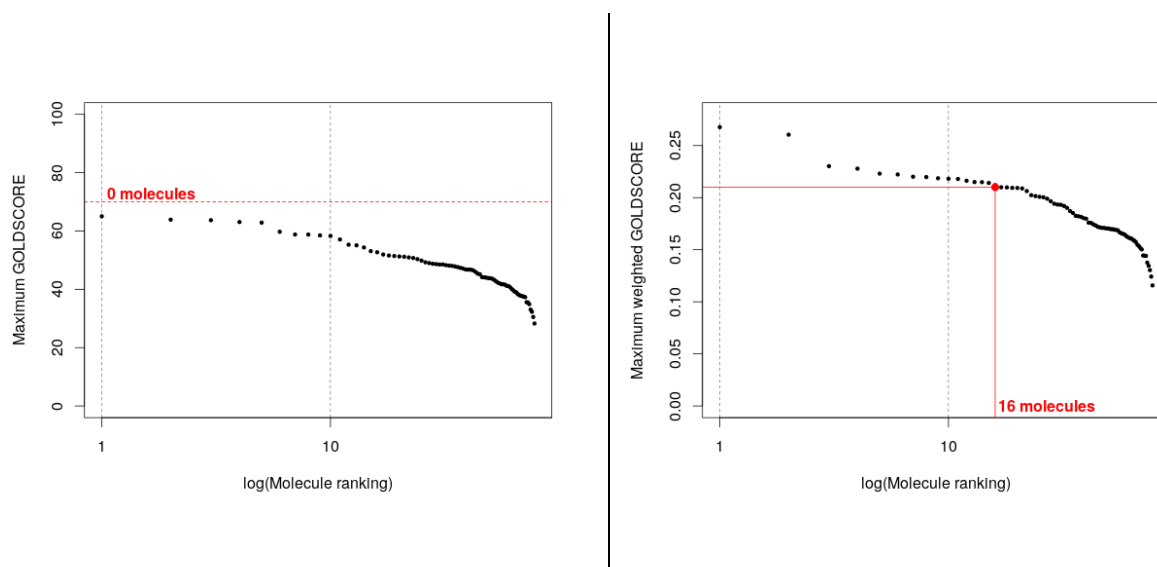
### 3. Résultats du criblage virtuel - CNE & CN

En parallèle, les mêmes analyses ont été réalisées pour la chimiothèque nationale essentielle CNE, comprenant 1026 molécules, et pour la chimiothèque nationale CN, comprenant 78.371 composés.

#### a. CNE

Comme indiqué dans la partie des présélections des composés d'intérêt, par rapport au faible nombre de composés que la CNE contient, il a été décidé que tous les composés seraient analysés quelques soient leurs poids moléculaires. A partir des 1026 molécules, 78 ne possédaient que 3 poses de criblage. Parmi ces 78 composés, en utilisant les mêmes bornes de GoldScore et de GoldScore pondéré définies pour la chimiothèque ZINC :

- Aucune molécule n'a été sélectionnée sur la base du GoldScore (Figure 94 - gauche),
- 16 molécules ont été sélectionnées sur la base du GoldScore pondéré (Figure 94 - droite).



**Figure 94. Résultats du criblage virtuel de la CNE**

Il est également à noter que les 16 molécules sélectionnées ici possèdent un poids moléculaire inférieur à 250 Da. Toutefois, en suivant le même argument du faible nombre de molécules de départ, ces 16 molécules ont été gardées pour l'étape suivante.

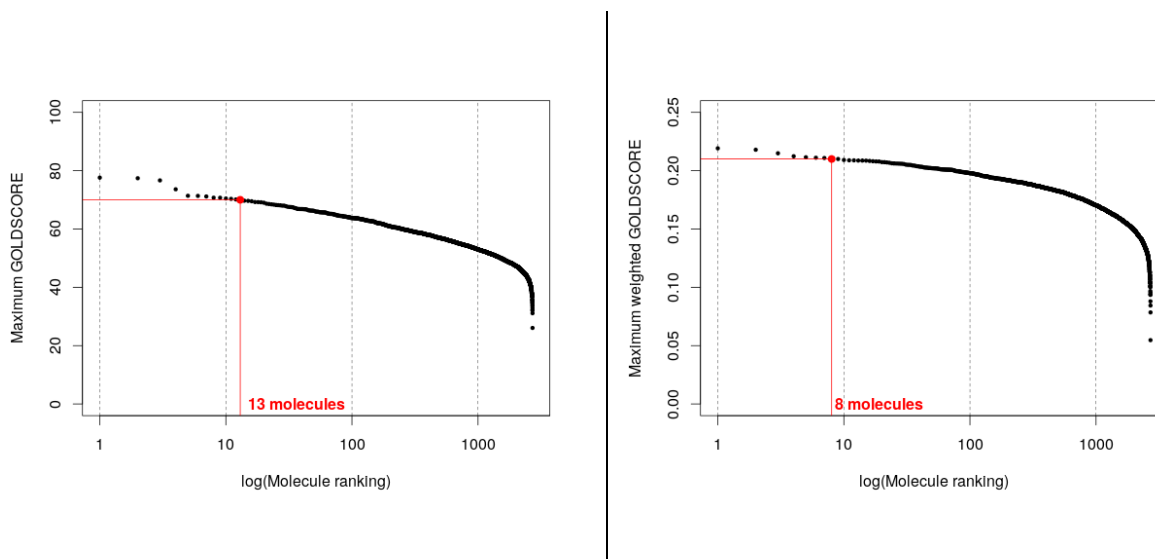
Toutefois, au vu du nombre de molécules très atypiques dans la CNE, avec des cycles aromatiques ou des atomes peu communs comme du sélénium, et des résultats obtenus à la suite du criblage virtuel, il a été décidé de réaliser le criblage virtuel sur la CN directement, pour obtenir des molécules dans la gamme de poids moléculaire souhaitée.

### ***b. CN***

A partir des 46 420 composés présélectionnés de la CN répartis en 21 groupes de criblage, 2705 molécules ne possédaient que 3 poses de criblage. De la même manière que pour ZINC et CNE, en utilisant les mêmes bornes de GoldScore et de GoldScore pondéré définies pour ZINC :

- 13 molécules ont été sélectionnées sur la base de leur GoldScore maximal (Figure 95 - gauche),
- 8 molécules ont été sélectionnées sur la base de leur GoldScore pondéré maximal (Figure 95 - droite).

Pour la chimiothèque CN, un total de 21 composés a donc été sélectionné pour l'étape suivante.



***Figure 95. Résultats du criblage virtuel de la CN***

### **III. Amarrage moléculaire des résultats du criblage virtuel**

Afin d'affiner la sélection de molécules, une étape d'amarrage moléculaire est nécessaire. Cette étape permet de positionner avec plus de précision toutes les molécules au sein du site de liaison défini, et d'obtenir de nouveaux scores plus précis.

#### **1. Matériels et méthodes**

Lors de l'amarrage moléculaire de tous les composés sélectionnés lors du criblage virtuel, la même conformation de la P28GST a été utilisée, la même zone de criblage a été définie, et le GoldScore a encore été choisi comme fonction de score, en calculant également en parallèle le GoldScore pondéré en utilisant le poids moléculaire de chaque composé. Pour le cas spécifique des groupes de molécules issues du criblage de la chimiothèque ZINC, 20 et 21 groupes de 45 composés au maximum ont été respectivement définis pour les groupes GoldScore pondéré de 881 molécules et GoldScore de 922 molécules.

Chaque instance d'amarrage moléculaire réalisée avec le logiciel GOLD a été lancée en ligne de commande. Pour chaque molécule criblée, le logiciel a réalisé un maximum de 100 000 opérations génétiques dans 5 populations de 100 chromosomes, composée de niches ne comprenant pas plus de 2 éléments par niche. La pression de sélection a été fixée à 1.1, le poids des croisements et des mutations ont été fixés à 95, et le poids des migrations a été fixé à 10, pour obtenir environ 10 fois moins de migrations que de croisements et mutations. De la même manière que pour le criblage virtuel, un maximum de 5 poses de criblage a été gardé pour chaque composé, avec un arrêt précoce de l'amarrage moléculaire si les 3 meilleures poses possèdent un RMSD inférieur à 1,5 Å, indiquant la convergence du positionnement.

Ces paramètres ont été définis afin d'obtenir un positionnement précis de chaque composé au sein de 3 à 5 poses d'amarrage moléculaire, pour un temps moyen de 120 secondes par molécule criblée.

## **2. Résultats amarrage moléculaire - ZINC**

### ***a. Critères appliqués lors du criblage virtuel***

L'amarrage moléculaire des résultats de criblage de la chimiothèque ZINC ont été réalisés à partir des 922 molécules du jeu de données GoldScore, et des 881 molécules du jeu de données GoldScore pondéré.

Dans un premier temps, les critères utilisés lors du criblage virtuel ont été appliqués. Parmi les 1803 molécules amarrées, seules 1263 ne convergent pas, et 1150 vont être retrouvées dans la gamme de GoldScore ou de GoldScore pondéré déterminée précédemment, afin de sélectionner :

- 571 molécules du jeu de données GoldScore,
- 579 molécules du jeu de données GoldScore pondéré

### ***b. Toxicité***

Afin de diminuer le nombre de composés, le premier filtre appliqué a été l'évaluation de la toxicité des composés, pour éviter les molécules mutagènes, tumorigènes, ayant un effet sur les fonctions reproductrices, ou encore les fonctions chimiques réactives. Après évaluation de techniques différentes, la solution retenue a été le calcul de ces critères à l'aide du logiciel DataWarrior. En appliquant les filtres "*Mutagenic*", "*Tumorigenic*", "*Reproductive Effective*", et "*Nasty Functions*" :

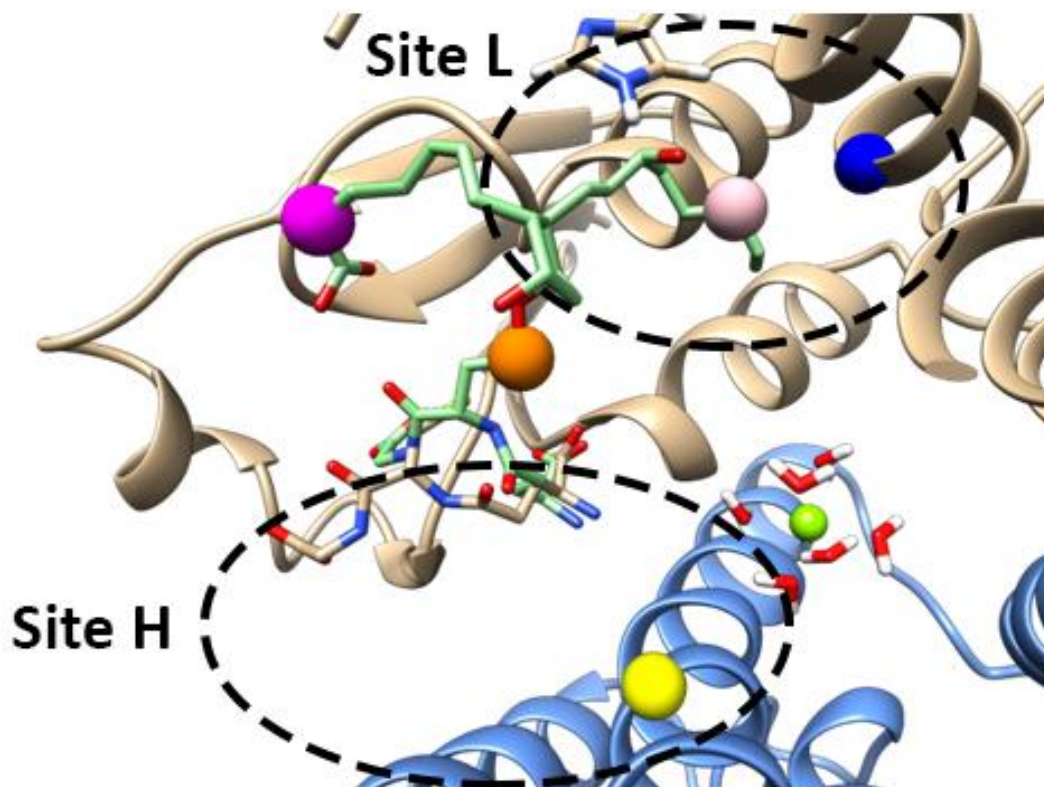
- 190 composés ont pu être retirés du jeu de données GoldScore, pour passer à un total de 381 composés,
- 105 composés ont pu être retirés du jeu de données GoldScore pondéré, pour un total de 474 composés.



### *c. Différenciation Site H/Site L*

Le second filtre qui a été appliqué aux 855 composés ayant passé avec succès le filtre de toxicité est un filtre qui prend en compte la position de chaque composé dans le site de liaison. Certains composés se plaçant au niveau du site H, le site de liaison non spécifique de l'enzyme, il est nécessaire d'éliminer ces composés pour ne garder que ceux se plaçant correctement dans le site L, le site de liaison spécifique de la prostaglandine H2. Pour cela, un script en Python3 a été créé pour évaluer la distance minimale de chaque molécule par rapport à certains points de l'espace, incluant :

- Les coordonnées du carbone alpha de la Valine en position 162, présent au fond de la poche du site L (Figure 96 - sphère bleue),
- Les coordonnées du carbone alpha de l'Asparagine en position 104 sur l'autre monomère, présent au niveau du site H (Figure 96 - sphère jaune),
- Les coordonnées de l'atome de Soufre du Glutathion d'une pose d'amarrage moléculaire de l'intermédiaire réactionnel Glutathion-Prostaglandine H2 obtenu précédemment (Figure 96 - sphère orange),
- Les coordonnées du Carbone 15 de cette même pose, correspondant au dernier carbone sp<sup>3</sup> de l'extrémité acide de la partie prostaglandine (Figure 96 - sphère magenta),
- Les coordonnées du Carbone 18 de cette même pose, correspondant à un carbone de l'extrémité cétonique de la partie prostaglandine situé au milieu de la poche du site L (Figure 96 - sphère rose).



*Figure 96. Emplacement des différents points de l'espace considérés pour différencier le site H du site L*

En étudiant la répartition de ces distances au sein du jeu de molécules global, et leur impact sur le positionnement des molécules retenues, le critère le plus pertinent qui a été sélectionné a été la distance entre chaque molécule et le carbone alpha de Asp104B. Une distance de plus de 10 Å par rapport à ce dernier permet d'éliminer efficacement toutes les molécules se positionnant au niveau du site H, ainsi que toutes les molécules possédant une chaîne latérale vers cette direction. Au total :

- Seules 6 molécules ont été gardées du jeu de molécules GoldScore,
- 117 molécules ont été gardées du jeu de molécules GoldScore pondéré.

#### *d. Clusterisation*

Pour diminuer le nombre de molécules présentes dans le jeu de molécules GoldScore pondéré, une clusterisation a été réalisée en utilisant DataWarrior. Ce logiciel réalise une classification ascendante hiérarchique à partir d'empreintes moléculaires calculées à partir de la structure de chaque composé, pour regrouper au sein de différents clusters toutes les molécules ayant une similarité supérieure à 0.6, dans le but d'obtenir une diversité chimique optimale. Ici, à partir des 117 molécules, 57 clusters ont pu être formés. Les 57 composés issus du jeu de molécules GoldScore pondéré représentatives de chaque cluster ont ensuite été gardés, en plus des 6 composés du jeu de molécules GoldScore.

#### *e. Disponibilité commerciale*

Pour prévoir l'achat des molécules, il est nécessaire de connaître précisément leur disponibilité commerciale, notamment le(s) différent(s) vendeur(s), les quantités disponibles, et les prix associés. A partir des deux jeux de données, de 6 molécules du jeu GoldScore et de 57 molécules du jeu GoldScore pondéré, un script en python a été créé permettant d'obtenir, à partir du site de la chimiothèque ZINC, les disponibilités commerciales et les vendeurs renseignés dans la base de données. Parmi l'ensemble des vendeurs, trois vendeurs principaux ont été sélectionnés, en fonction de la confiance accordée par le laboratoire et du nombre de molécules disponibles :

- Enamine, regroupant 15 molécules,
- Chembridge (hit2lead.com), regroupant 6 molécules,
- ChemDiv, regroupant 5 molécules.

Il est à noter qu'il existe une molécule retrouvée chez Chembridge et Enamine, une autre molécule retrouvée chez ChemDiv et Enamine, et que le total des 24 molécules sélectionnées par ce biais sont situées dans le jeu de données GoldScore pondéré.

### *f. Inspection visuelle*

Le dernier critère utilisé est l'inspection visuelle de chaque meilleure pose de criblage virtuel et d'amarrage moléculaire pour apprécier leur positionnement et leur mode de liaison. Les principaux critères ont été :

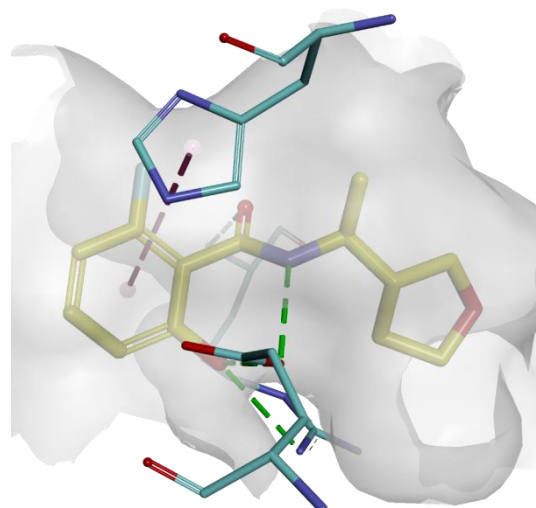
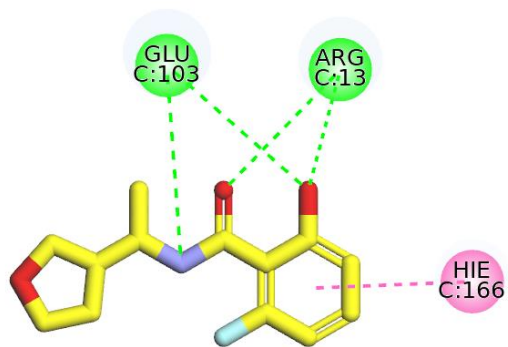
- Leur localisation au sein de la protéine, préférentiellement les molécules positionnées dans le site L et superposant la partie prostaglandine de l'intermédiaire réactionnel glutathion-prostaglandine H2, tout en étant à distance de la partie glutathion,
- Le nombre et le type des interactions réalisées,
- Et si les interactions ne vont pas interférer avec le mécanisme d'activation du glutathion, en réalisant des liaisons hydrogène avec la tyrosine catalytique.

Suite à cette étape d'inspection visuelle, 11 molécules ont été retenues, dont 7 du jeu Enamine, 4 du jeu ChemDiv, et 2 du jeu Chembridge.

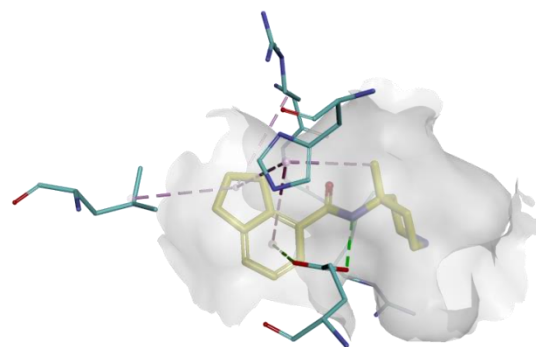
### *g. Sélection finale*

La sélection finale des molécules retenues pour la réalisation de validations expérimentales s'est basée sur le choix d'un seul vendeur ayant à disposition un maximum de molécules à des prix abordables. Ici, le vendeur Enamine a été retenu, et 5 molécules ont été sélectionnées (Tableau 16).

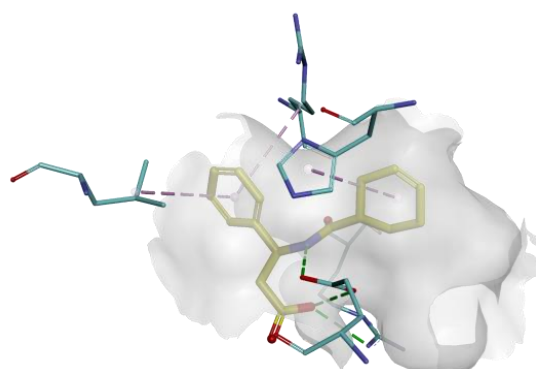
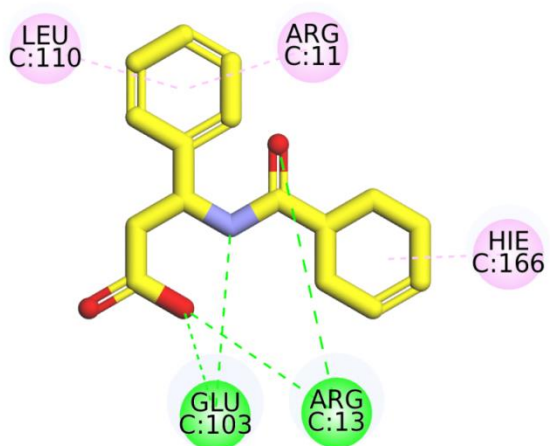
Ces 5 molécules vont posséder des fonctions chimiques et des modes de liaison différents. Au vu de la configuration du site L retrouvé chez P28GST, ces molécules sont de manière générale linéaire, avec la présence d'au minimum un cycle aromatique et/ou d'une fonction amide au centre de la molécule. Le cycle aromatique permet d'interagir avec l'histidine 169, acide aminé qui a été décrit précédemment comme décrivant l'ouverture du site L, via une liaison aromatique. Il permet également d'interagir dans certains cas avec l'Arginine 14 par une liaison pi-cation, ou avec la Phénylalanine 211 par une liaison aromatique. La liaison amide permet la formation d'une ou de plusieurs liaisons hydrogène avec les acides aminés Arginine 16 et Acide glutamique 106, situés de part et d'autre du site L. Enfin, des interactions hydrophobes sont généralement retrouvées pour les molécules se positionnant au fond du site L, avec la participation des résidus Valine 162, Alanine 165 et Valine 166.



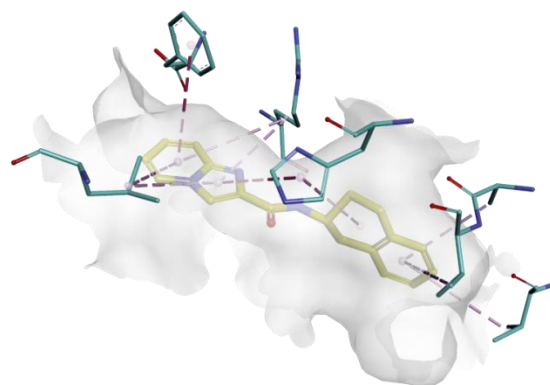
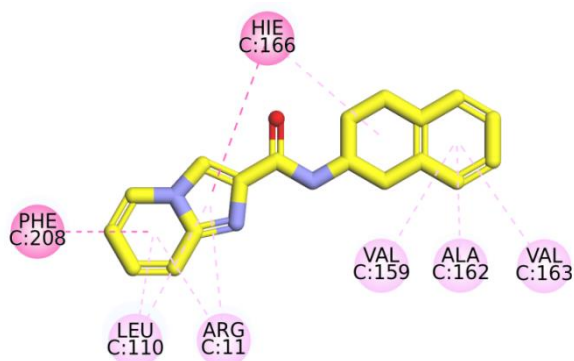
ZINC ID : ZINC000126323335  
Enamine ID : Z1724524224



ZINC ID : ZINC000262398518  
Enamine ID : Z2394390793

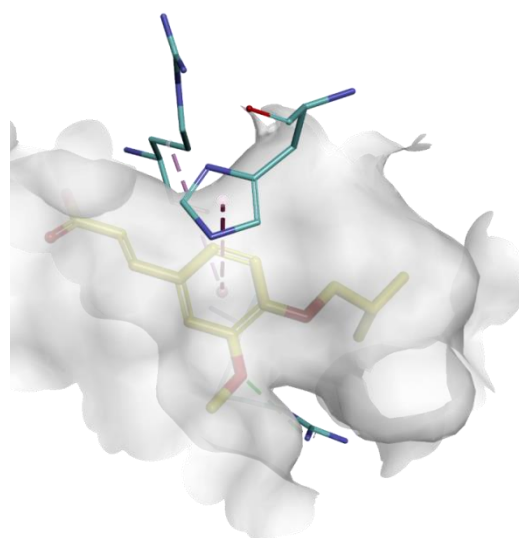
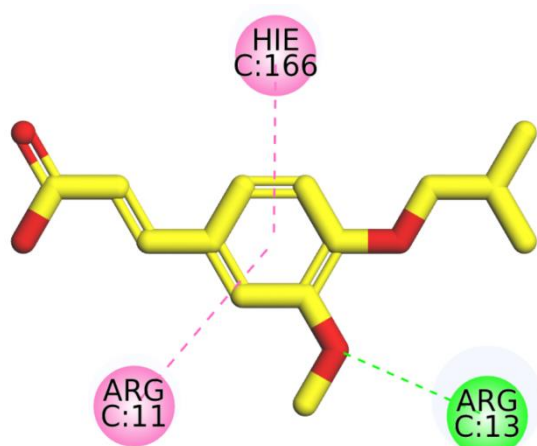


ZINC ID : ZINC000013225021  
Enamine ID : Z146597916



ZINC ID : ZINC000070753883

Enamine ID : Z29213909



ZINC ID : ZINC000002635407

Enamine ID : EN300-832884

*Tableau 16. Structure, positionnement et informations des 5 composés de la sélection finale*

### **3. Résultats docking - CNE et CN**

Pour les chimiothèques CNE et CN, l'amarrage moléculaire des 11 molécules et 21 molécules sélectionnées ont été réalisées avec les mêmes conditions expérimentales que pour celui réalisé avec la chimiothèque ZINC. Toutefois, après une inspection visuelle des meilleures poses obtenues et l'application des différents filtres précédemment utilisés, aucune molécule intéressante n'a pu être retenue.

Les étapes de criblage virtuel à haut débit, d'amarrage moléculaire et de sélection de composés d'intérêt ont eu pour but principal l'identification de molécules potentiellement sélectives de l'activité PGDS. Ici, l'application de nombreux filtres ont permis l'identification de 5 composés possédant des fonctions chimiques et des modes de liaison différents au sein du site de liaison, en ayant tous pour point commun une interaction avec His169, décrit précédemment comme acide aminé clé pour la conformation du site L permettant le bon positionnement de la prostaglandine H2. La validation expérimentale de ces 5 composés a été initiée.

# **DISCUSSION ET PERSPECTIVES**





Dans le cadre de ces travaux de thèse, en mettant en œuvre une palette de méthodes *in silico*, notre but était de comprendre le fonctionnement des deux activités enzymatiques GST et PGDS au niveau le plus fin possible, et d'exploiter ces connaissances pour la recherche de composés utilisables comme sondes pharmacologiques afin de proposer des outils d'évaluation de la contribution de chaque activité enzymatique dans l'effet anti-inflammatoire global.

Dans le premier temps de la thèse, les mouvements à l'échelle moléculaire et atomique des mouvements des acides aminés du site actif, et de leur participation différentielle entre les deux activités, ont été menés par des simulations de dynamique moléculaire classique et de métadynamique.

## **1. Dynamique moléculaire**

### ***a. Observation structurale d'un mouvement de la boucle catalytique***

Les premières dynamiques moléculaires réalisées avec le logiciel GROMACS ont permis de se familiariser avec cette technique, en s'intéressant aux différents algorithmes, méthodes et techniques proposées par ce genre de logiciel de simulation. Les toutes premières simulations ont été réalisées sur la P28GST sous sa forme de monomère. Même si l'objectif était la réalisation de premières dynamiques sous la forme d'une preuve de concept, l'observation de la formation d'une ouverture à proximité de la tyrosine catalytique a été un fait très intéressant. Son observation a permis d'envisager une première hypothèse sur la bascule de la tyrosine 10 entre ses conformations IN et OUT en passant par cette ouverture, cette dernière étant délimitée par la boucle catalytique, très flexible car composée de nombreuses glycines.

Du fait du manque de logiciels permettant de mesurer dynamiquement un volume déterminé entre différents atomes dans une dynamique moléculaire, un script a été créé en dérivant un autre outil, POVME 2, prenant en fichier d'entrée des fichiers au format PDB. Ce logiciel a été utilisé pour sa facilité d'utilisation, en utilisant une simple commande précisant le fichier de configuration. Ce fichier a pu être généré progressivement au sein du script python créé, s'adaptant aux coordonnées de la simulation. Même si ce script fonctionnait bien, car il a pu être appliqué avec succès dans le cadre d'autres travaux au laboratoire, il n'a pas donné de

résultats satisfaisants. La présence du glutathion ainsi que le changement de configuration de la boucle, déplaçant de quelques angströms les coordonnées moyennes définie comme étant le centre de la sphère de mesure, entraînait de trop grandes variations de volume. Cette piste n'a pas été creusée du fait du changement de logiciel, passant de GROMACS à Amber18.

L'utilisation d'Amber a permis d'amener un nouvel outil dans la palette des logiciels du laboratoire. Grâce à du logiciel Docker, permettant de créer des conteneurs logiciels, une solution clé en main sous la forme d'une image Docker a été réalisée, facilitant ses futures utilisations.

De la même manière que pour GROMACS, les premières simulations réalisées ont été réalisées sous la forme d'une preuve de concept, mais ont amené de nombreuses informations. A partir de ces simulations, la P28GST a été simulée en dimère. L'observation principale a mis en évidence une nouvelle conformation de la tyrosine catalytique jamais observée auparavant, ni décrite ou supposée dans la littérature. Au vu de la position de cette nouvelle conformation à mi-chemin entre les conformations IN et OUT, cette dernière a été nommée conformation "MID". Lors de la recherche d'un moyen de différencier ces conformations, l'utilisation de la banque de données des rotamères de Dunbrack a été un atout majeur, permettant de relier les différentes conformations aux angles dièdres Chi1.

En se basant sur la distribution de l'angle dièdre Chi1 déterminé à partir de tyrosines de structures cristallographiques de cette librairie, il a pu être vu que la plus grande probabilité de retrouver l'angle dièdre Chi1 correspondait à la conformation MID, allant à l'encontre des observations réalisées. Toutefois, sa faible occurrence dans les 4 répliques réalisées a laissé supposer qu'il s'agissait d'un évènement rare. Pour l'observer plus facilement, soit il est nécessaire d'augmenter le temps de simulation, soit d'apporter de l'énergie au système, ce qui est difficile dans une dynamique moléculaire classique, d'où l'utilisation de techniques plus avancées. Ces techniques devaient également permettre d'étudier le changement de conformations de la tyrosine catalytique, qui n'a pas pu être étudié avec la dynamique moléculaire classique, comme nous venons de le voir.

### *b. Dynamiques moléculaires classiques plus longues*

Dans un premier temps, de plus longues simulations de dynamique moléculaire ont été réalisées. A ce niveau d'avancement, il a été décidé de reconstruire entièrement la protéine, la structure cristallographique ayant quelques acides aminés en moins dans sa partie C-terminale, ceux-ci étant situés à proximité du site actif, et faisant partie de l'épitope 3. De la même manière, pour étudier les mouvements de la tyrosine dans ses conformations IN et OUT au sein d'une seule simulation, il a été décidé de ne garder que les coordonnées alternatives A, correspondant à la conformation IN, pour le monomère A, et les coordonnées alternatives B, correspondant à la conformation OUT, pour le monomère B.

Les nouvelles simulations de dynamique moléculaire classique de 1.5  $\mu$ s ont été réalisées sur le méso-centre de Calcul Scientifique Intensif de l'Université de Lille, en passant par l'image Docker de Amber18, permettant un environnement logiciel contrôlé et comparable aux simulations réalisées en local. Toutefois, du fait des limites matérielles et du temps maximal de calcul fixé à 16 jours, les simulations réalisées ont dû être programmées pour être réalisées en deux lots de 750 ns pour un total de 40 jours en incluant l'attente des ressources.

Lors de ces simulations, en étudiant le monomère A, il a été observé un profil totalement différent des simulations précédentes, avec la tyrosine catalytique restant dans sa conformation IN au sein de toutes les trajectoires de dynamique moléculaire classique. Toutefois, la reconstruction des derniers acides aminés de la partie C-terminale a dû jouer un rôle dans ce résultat, en stabilisant les structures secondaires alentours, notamment la partie débutante de l'hélice  $\alpha 1$ , située juste après la boucle catalytique. Cet effet n'a pas pu être étudié, notamment du fait de l'utilisation des ressources locales et du cluster de calcul sur les techniques avancées de dynamique moléculaire. Toutefois, il serait pertinent de réaliser de nouvelles répliques en appliquant le même protocole, mais en ne reconstruisant pas les derniers acides aminés de la protéine, pour mesurer l'impact de la présence de cette partie sur la flexibilité de la boucle catalytique et les différents résidus de l'hélice  $\alpha 1$ .

Les observations réalisées au sein de ces dynamiques moléculaires ont permis de valider l'hypothèse de la présence de barrières énergétiques, confirmant l'intérêt de réaliser des techniques avancées de dynamique moléculaire. La dynamique moléculaire classique a montré ici ses limites techniques, ne permettant pas l'apport d'énergie donc l'exploration de différents minima énergétiques.

### *c. Dynamiques moléculaires avancées*

En parallèle de la réalisation des dynamiques moléculaires classiques de 1.5  $\mu$ s, trois méthodes avancées de dynamique moléculaire ont été évaluées pour étudier la pertinence de la conformation MID et observer les transitions entre les conformations IN et OUT.

La première technique évaluée a été le recuit simulé dans le but de réaliser une exploration conformationnelle de la P28GST. Toutefois, aucune autre conformation de la tyrosine n'a pu être observée, et il a été décidé de passer à d'autres techniques plus prometteuses. Une solution qui aurait pu être explorée aurait été d'adapter la simulation en utilisant d'autres champs de force, en ajoutant des contraintes sur certains atomes pour éviter l'explosion du système causée par l'apport d'une trop grande quantité d'énergie, au risque d'empêcher les mouvements ou de les biaiser, ou encore l'utilisation de solvant implicite, permettant de continuer les difficultés techniques induites par la topologie du solvant. De plus, le protocole aurait pu être adapté en ajustant les temps utilisés, en réalisant notamment une phase de refroidissement plus longue, et donc plus progressive, permettant d'atteindre potentiellement différents minima énergétiques.

La seconde technique évaluée a été la dynamique moléculaire dirigée, permettant d'appliquer des biais de force sur une variable, ici l'angle dièdre  $\text{Chi}_1$  de la tyrosine catalytique, pour la faire passer de la conformation IN à la conformation OUT, en passant par la conformation MID. A la suite de cette simulation, toutes les conformations de la tyrosine ont pu être retrouvées sans impact sur l'intégrité de la protéine. Cette observation permet de proposer un nouvel argument de l'existence d'une conformation MID : le changement de conformation de la tyrosine en passant par le canal formé par la boucle catalytique n'entraîne pas de dénaturation de la protéine. Les barrières énergétiques entre les différentes conformations pourraient aussi présenter des hauteurs différentes conditionnant le passage d'une conformation à l'autre en fonction de l'état des interactions stabilisant la tyrosine. D'autres simulations simulant le passage inverse, de la conformation OUT à IN en passant par la conformation MID, ou encore étudiant la transition directe entre les conformations IN et OUT auraient pu être envisagées. De plus, la mesure du profil énergétique de ces variables n'avait pas été considérée, mais aurait pu être une information pertinente à mettre en relation avec les métadynamiques décrites ci-après.

La troisième technique, qui a été la base des travaux suivants, a été la métadynamique.

## **2. Métadynamique**

Les simulations de métadynamique réalisées ont eu pour but de construire le paysage conformationnel de la tyrosine catalytique en fonction de certains degrés de liberté à partir de gaussiennes énergétiques. La première méthode appliquée a été la dynamique moléculaire à biais adaptatif (ABMD) ; elle n'a pas donné les résultats attendus. Malgré de nombreux essais, la convergence de la simulation n'a jamais pu être atteinte. L'implémentation de cet algorithme est une boîte noire, avec peu d'informations sur les résultats intermédiaires, ni aucun contrôle sur le dépôt des biais énergétiques.

Le passage à la métadynamique proposée par PLUMED sous la forme d'un plugin à Amber18 a permis un contrôle plus fin des paramètres de simulation. Toutefois, de la même manière, la convergence de la simulation n'a pas pu être atteinte. De plus, malgré le fait que l'angle dièdre Chi1 semblait avoir été correctement exploré, la position spatiale de la tyrosine catalytique ne correspondait pas forcément à la conformation définie par l'angle Chi1. Cet angle étant un élément propre à la tyrosine et que les conformations IN, MID et OUT sont les positionnements de cette tyrosine par rapport à la structure de la P28GST, ces éléments ne sont pas nécessairement corrélés. La solution expérimentée a été le changement de la variable collective utilisée au cours de la simulation, mais soit les profils énergétiques étaient inexploitable, soit les profils étaient exploitables mais ne convergeaient pas. La position de tyrosine par rapport à la boucle catalytique a été envisagée, mais entraînait des difficultés dans la distinction entre les différentes conformations. D'autres variables collectives externes ont été évaluées, comme des variables collectives issues du RMSD des conformations de référence ou des points d'ancrage avec d'autres atomes à distance, mais aucune de ces variables n'a permis l'obtention de résultats interprétables et satisfaisants.

L'angle Chi1 devant nécessairement être sélectionné comme variable collective, les paramètres expérimentaux ont été optimisés pour limiter ce phénomène tout en se rapprochant d'une convergence. De plus, il a été décidé de réaliser un grand nombre de répliques permettant l'obtention d'un profil énergétique moyen permettant de compenser la non-convergence.

Les simulations de métadynamique impliquant la tyrosine déprotonée n'ont pas permis l'obtention de résultats pertinents, par l'utilisation du champ de force ff14SB non adapté pour

la simulation d'effets stéréo-électroniques. L'utilisation d'un champ de force polarisable permettrait de prendre en considération toutes les charges du système pour permettre l'obtention d'un profil énergétique fiable. De plus, l'utilisation de l'approche hybride de la dynamique quantique et de la mécanique moléculaire (QM/MM) permettrait d'étudier les processus chimiques mis en jeu lors de la déprotonation de la tyrosine dans sa conformation OUT, ainsi que lors de la déprotonation du glutathion lorsque la tyrosine est dans sa conformation IN.

Enfin, la contribution de la conformation MID, n'ayant pas forcément de contribution au niveau mécanistique lorsque la tyrosine catalytique est protonée, doit être évaluée lorsque la tyrosine est déprotonée. Cette conformation peut faire partie du chemin retour de la conformation OUT à la conformation IN. Le chemin de OUT à IN en passant par MID présente moins d'acides aminés polaires, et donc un franchissement plus facile de la boucle catalytique. Cette contribution peut également être étudiée lors de l'utilisation de champs de force polarisables.

### **3. SINAPs et étude des interactions intramoléculaires**

La mesure des différentes interactions moléculaires réalisées différemment entre chaque conformation a été envisagée dès la réalisation des premières dynamiques moléculaires sous Amber. Le manque de logiciels permettant une représentation de ces interactions dans l'architecture de la protéine nous a décidés à développer un script permettant de mettre en évidence les réseaux d'interactions moléculaires réalisés au cours de chaque simulation en dérivant un logiciel existant, suivant la même logique que le script de mesure de volumes. Le logiciel de calcul de réseaux d'interactions retenu a été RIP-MD (156), fonctionnant en local, permettant l'obtention de fichiers de sorties facilement utilisables. Des premières versions du script permettant de comparer les résultats de RIP-MD ont été développées pour une utilisation en ligne de commande, mais étaient difficilement utilisables à cause du manque de représentation visuelle. Une solution a été trouvée avec UCSF Chimera, en chargeant les résultats pour former des liaisons entre tous les atomes décrits dans un fichier obtenu à la fin du script. De plus, du fait de la faible flexibilité de RIP-MD au niveau du format des fichiers d'entrée, il a été décidé de directement le remplacer par une librairie python basée sur les

outils d'analyse d'Amber : pytraj. Le développement de cette script a posé les bases du logiciel SINAPs.

Ce logiciel a été validé sur deux sujets fortement étudiés dans la littérature, permettant de retrouver les grands réseaux d'interactions les organisant et conditionnant leur activité : GLUT-1 et A2AR. Par la suite, SINAPs a été appliqué à toutes les trajectoires obtenues au cours des travaux de thèse, notamment aux dynamiques moléculaires classiques de 1.5 $\mu$ s, pour mettre en évidence les différents réseaux d'interactions de la P28GST au niveau de l'interface dimérique et intramoléculaires.

### ***a. Interactions de l'interface dimérique***

Dans un premier temps, l'interface dimérique a été étudiée, permettant de mettre en évidence tous les acides aminés participant à la stabilisation du dimère. L'interface est formée de nombreuses liaisons ioniques, mettant en jeu la participation à plus de 90% de Glu70, Arg76, Lys80, Lys81, Glu89, Glu96 et Lys97, et à plus de 50% de Arg52, Glu90 et Asp104. Des liaisons hydrogènes sont également retrouvées, avec la contribution de Val69, Glu70, Tyr77, Gly90, Tyr93 et Gln101. Tous ces acides aminés sont situés sur les hélices  $\alpha$ 3 et  $\alpha$ 4 des deux monomères, ce qui confirme les descriptions de l'interface dimérique retrouvées dans la grande majorité des protéines de la famille des GST (84). Le nombre important de résidus polaires, notamment acides, laissent penser à une contribution importante de molécules d'eau au niveau de l'interface dimérique.

### ***b. Réseaux d'interactions organisant la structure tertiaire***

Dans un second temps, les deux monomères ont été séparés puis comparés entre eux, permettant de comparer les interactions en fonction de la position de la tyrosine. Lors de l'étude des interactions organisant la structure tertiaire, de très nombreuses liaisons ioniques communes ont été retrouvées, stabilisant les structures secondaires et boucles ensemble. Elles sont formées à partir de très nombreux résidus polaires : Lys7, Arg14, Arg16, Glu18, Arg21, Glu32, Asp33, Glu34, Arg52, Glu70, Arg76, Glu96, Glu103, Glu106, Glu108, Lys131, Lys149, Asp154, Asp168, Asp172, Arg189 et Arg206. Quelques liaisons hydrogènes communes complètent le réseau, avec la participation de Glu18, Asp160, His188 et Tyr202.



Lors de l'étude des réseaux d'interaction communs, la boucle située entre les acides aminés Ser146 et Thr157 se différencie particulièrement par son organisation structurale composée de 3 coudes. Un dense réseau de liaisons hydrogène est retrouvé, avec notamment 5 liaisons entre chaîne latérale et chaîne carbonée et 2 liaisons entre chaînes carbonées. Dans la littérature, aucune information n'a été retrouvée sur l'implication de cette boucle spécifique dans les différentes activités enzymatiques d'autres protéines de la famille des GST. Au vu de l'implication non évidente de cette structure sur les activités enzymatiques, aucune interaction directe n'étant réalisée avec des acides aminés clés du site actif, il a été décidé de ne pas poursuivre les investigations sur son rôle.

### *c. Réseaux d'interactions intramoléculaires différentiels*

Enfin, les réseaux d'interactions différentiels au niveau de la tyrosine et de la boucle catalytique ont été observés. L'acide aminé Arg21 a été observé comme étant essentiel par la réalisation de liaisons ioniques communes avec Glu18 et Asp33 dans l'architecture globale, et comme interagissant avec la boucle catalytique avec Asn12 et Arg14 lorsque la tyrosine est dans sa conformation IN uniquement. Des liens peuvent être réalisés avec des observations de la littérature, notamment avec les études de mutagenèse dirigée. Lorsque Arg21 est mutée en leucine, seule la tyrosine dans sa conformation OUT est observée dans une structure cristallographique (100). Ce résultat permet de démontrer que Arg21 possède un rôle stabilisant sur la boucle catalytique, permettant de contraindre ses mouvements par la réalisation de deux liaisons hydrogènes, confortant donc nos observations.

De la même manière, Asp33 possède une interaction commune avec Arg21, mais interagit avec soit Arg35 lorsque la tyrosine est dans sa conformation IN, soit Lys7 lorsque la tyrosine est dans sa conformation OUT. Cet acide aminé est décrit dans la littérature comme pouvant être impliqué dans la déprotonation de la tyrosine catalytique (100). Cette description permet d'expliquer les observations de SINAPs, où Asp33 est préférentiellement positionnée à proximité de la tyrosine dans sa conformation OUT de part sa liaison ionique avec Lys7, encerclant ainsi la fonction alcool de la tyrosine de nombreuses chaînes latérales polaires. De plus, Arg35 réalise des liaisons ioniques avec Asp33 et Glu34 préférentiellement lorsque la tyrosine est dans sa conformation IN, validant des descriptions de la littérature reliant les mouvements de la tyrosine avec ceux de Arg35 (100). Arg35 ne réalisant aucune interaction commune dans la stabilisation des conformations tertiaires et quaternaires, et étant à proximité directe de la tyrosine catalytique, cet acide aminé peut participer au phénomène de bascule

entre les conformations IN et OUT, et semble être l'un des éléments initiant le mouvement de l'une à l'autre. Lors de la perte des interactions de Arg21 avec la boucle catalytique, Asp33 viendrait préférentiellement interagir avec Arg21, ce qui libérerait les mouvements de Arg35, et offrirait plus de place à la Tyr10 catalytique de basculer en passant de la conformation IN à la conformation OUT directement.

L'implication clé des résidus Asp33 et Arg35 dans l'initiation des activités enzymatiques peut être validé indirectement par les observations réalisées lors du développement préclinique de la P28GST. En effet, les anticorps reconnaissant l'épitope 1, localisé des acides aminés 24 à 43, ont la particularité d'empêcher l'activité Glutathion S-Transférase (25–27). Cet effet peut être expliqué par l'apport d'une gêne stérique au niveau des résidus Asp33 et Arg35, limitant leurs mouvements et entravant la chaîne d'activation permettant la bascule de la Tyr10 catalytique entre ses conformations IN et OUT.

Le développement du logiciel SINAPs a permis d'automatiser la recherche des différents d'interactions pour étudier leurs différences et leurs ressemblances entre les deux états conformationnels identifiés chez la P28GST et conditionnés par la position de la tyrosine catalytique. Les différences observées sont liées aux mouvements de la tyrosine, elle-même impliquée dans les activités enzymatiques. Donc les différents résidus mis en évidence par SINAPs ont un rôle dans l'initiation du cycle enzymatique, aboutissant à l'activation du glutathion et aux deux activités enzymatiques GST et PGDS. Ce logiciel, utilisable à partir de différents types de fichier d'entrée et validé à partir de résultats de la littérature, est à la disposition de la communauté scientifique.

#### **4. Cation divalent**

L'étude de l'interface dimérique de la P28GST a mis en évidence un volume important, bordé par des acides aminés polaires. Au vu des caractéristiques structurales des autres protéines de la classe sigma des GST, la présence d'un cation divalent peut être envisagée. Ce cation peut être un ion magnésium  $Mg^{2+}$  ou  $Ca^{2+}$ , comme retrouvé chez HPGDS (95), avec leur couronne de molécules d'eau stabilisée par ces nombreux résidus acides. Cette possibilité a été étudiée dans la suite des travaux.

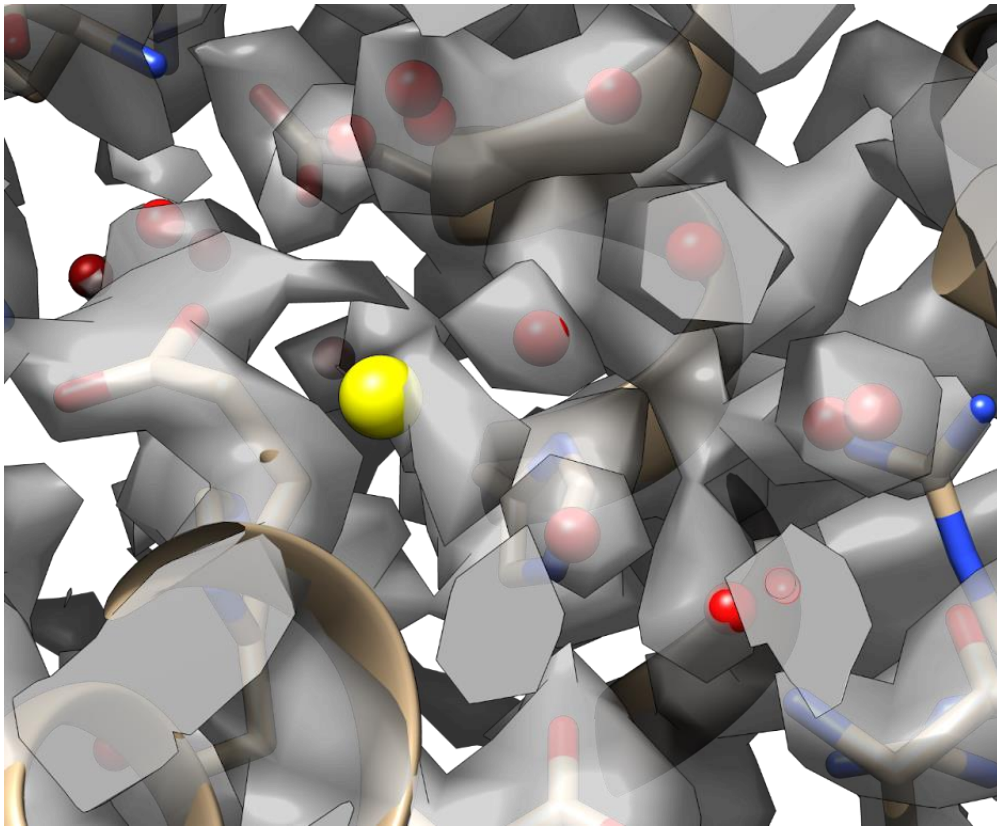
De premières pistes obtenues par dynamique moléculaire ont pu mettre en évidence la stabilité d'un cation  $Mg^{2+}$  et de ses molécules d'eau positionné au niveau de l'interface dimérique, entre des acides aspartique et glutamique. Le choix du cation s'est porté sur  $Mg^{2+}$ , car cet ion augmente l'activité PGDS chez HPGDS tout en diminuant le  $K_m$  pour le glutathion (95), et que la présence d'une quantité saturante de  $Mg^{2+}$  entraîne une augmentation de 15 à 25% de l'activité GST chez la P28GST. De plus, la mesure expérimentale de la présence de cations divalents par ICP-MS n'a pas démontré la présence de  $Mg^{2+}$ , mais a permis de mettre en évidence la présence d'un ion  $Zn^{2+}$  à hauteur de 5% dans un échantillon de P28GST.

Au vu des différents résultats obtenus par les travaux de différentes disciplines comprenant la chimie analytique, la biochimie et la chémoinformatique, et tirés de résultats de la littérature et des travaux de thèse, il est possible de conclure sur la présence d'un cation divalent  $Zn^{2+}$  présent à l'interface entre les deux monomères de la P28GST. Ce cation semble être stabilisé par les acides aminés Glu103 et Asp104, et pourrait avoir un impact majeur sur l'activité PGDS en favorisant l'ouverture du site L permettant la liaison de la prostaglandine H2 par des modifications au niveau des réseaux d'interactions moléculaires.

De nouvelles simulations de dynamique moléculaire classique en présence de  $Zn^{2+}$  ont confirmé le positionnement et la stabilité de ce cation au niveau de l'interface dimérique. Ce cation est stabilisé par les acides aminés Glu103 et Asp104. De plus, la présence du cation  $Zn^{2+}$  entraîne un changement au niveau des réseaux interactions intramoléculaires, en augmentant la fréquence des interactions entre Glu103, Asp104, Arg16 et Glu106, tout en diminuant la fréquence de l'interaction entre Glu106 et His169.

D'autres pistes indirectes, issues de la littérature, permettent d'étayer *a posteriori* ces hypothèses. En remplaçant le cation  $Zn^{2+}$  par superposition des conformations identifiées avec la structure cristallographique, et en analysant les cartes de densité électronique obtenues lors de la réalisation de cette cristallographie à rayons X, il est possible de mettre en évidence une densité électronique non prise en compte à la position du  $Zn^{2+}$  retrouvée lors des simulations de dynamique moléculaire (Figure 101). Cette zone de densité est de forme atypique, plus petite qu'une molécule d'eau. En outre, un cation divalent n'était pas recherché à l'époque de

sa réalisation, et qu'il n'y a que 5% de cation  $Zn^{2+}$  au sein de l'ensemble des P28GST. Il a donc pu être pensé que cette densité était un artefact de cristallisation.



**Figure 101.** Carte de densité électronique (en surfaces grises) de la structure cristallographique de la P28GST (PDB ID : 1OE8) avec la représentation des molécules d'eau retrouvées dans la cristallographie (sphères rouges), et représentation du cation  $Zn^{2+}$  (sphère jaune) repositionné à partir d'une conformation observée lors des dynamiques moléculaires réalisées.

De plus, le zinc est un élément chimique clé dans le métabolisme des schistosomes. Chez ces parasites, l'homéostasie du zinc est particulièrement développée. Leur absorption de zinc est relativement élevée à la fois par les vers et les œufs (157). En parallèle, chez *Schistosoma mansoni*, un phénomène d'hormèse peut être observé. De faibles concentrations de zinc, égales à  $10 \mu\text{g/L}$  soit  $0.07 \mu\text{M}$ , augmentent la prolifération et la survie du parasite, alors que de plus fortes concentrations, supérieures à  $1000 \mu\text{g/L}$  soit  $7 \mu\text{M}$ , sont très toxiques, inhibant significativement l'éclosion des œufs et la survie des jeunes parasites (158). La P28GST jouant également un rôle dans l'éclosion des œufs et la survie, un lien peut être imaginé entre cette protéine et le métabolisme du parasite en présence de zinc. On peut supposer un effet dose-réponse sur l'activité enzymatique en présence d'un cation  $Zn^{2+}$ , et le relier à ces

observations macroscopiques. La concentration de  $Zn^{2+}$  égale à 5.47% dans l'échantillon de P28GST par des techniques analytiques, semble être la concentration optimale pour la survie du parasite et sa prolifération, et être la cause d'une pression de sélection.

Toutefois, ces hypothèses *in silico* n'ont pas pu être testées sur les activités enzymatiques GST et PGDS de la P28GST. Il est à ce jour impossible de conclure sur l'impact réel de ce cation : est-ce que sa présence est requise pour l'obtention des activités enzymatiques comme pour d'autres espèces de GST avec la présence obligatoire d'un cation  $Mg^{2+}$ , en agissant comme un cofacteur ? Est-ce qu'il permet plutôt de potentialiser l'activité enzymatique ? Si oui, est-ce que son effet est spécifique d'une activité, ou est-ce qu'il joue sur les deux activités enzymatiques simultanément ?

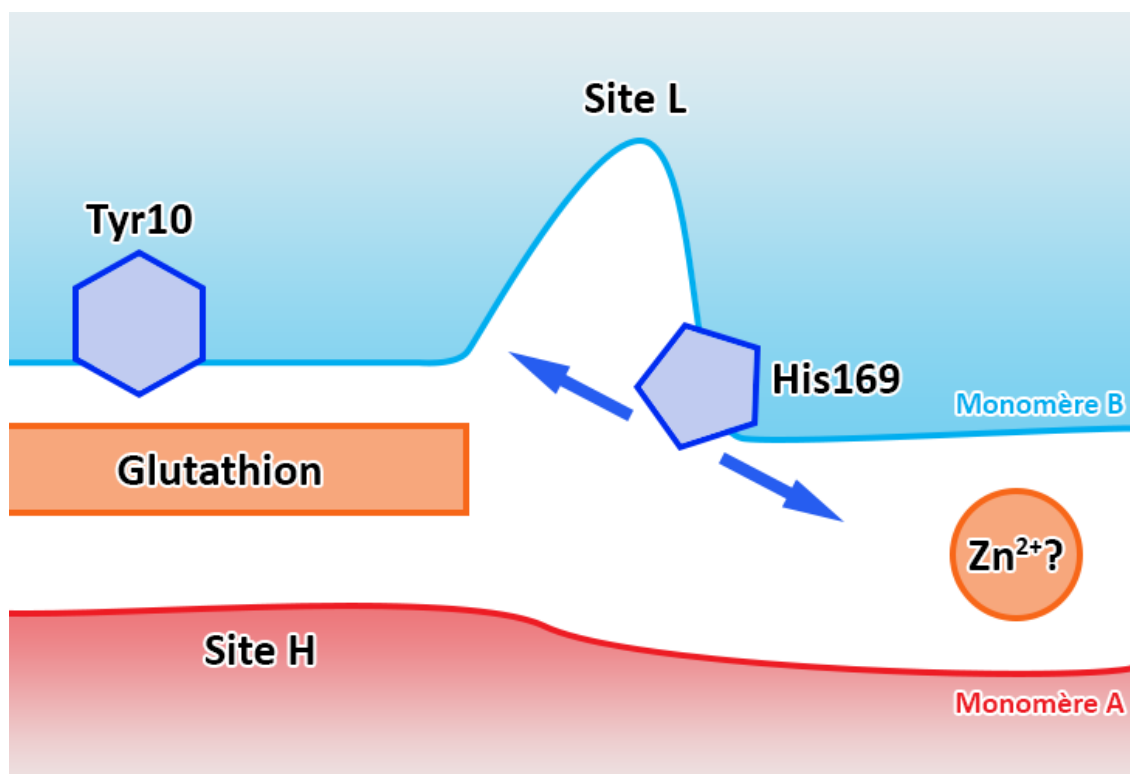
La réalisation d'une nouvelle production de P28GST en s'assurant de l'absence de  $Zn^{2+}$  permettrait d'évaluer l'impact de ce cation sur les différentes activités enzymatiques en réalisant des tests activités GST et PGDS. De plus, la réalisation d'une cristallographie en présence d'une quantité saturante de  $Zn^{2+}$  permettrait de connaître le positionnement exact de ce cation, et peut être combiné avec la co-cristallisation d'un inhibiteur spécifique de l'activité PGDS.

## **5. Détermination du site de liaison de la prostaglandine H2 (site L)**

Dans le but de trouver des sondes pharmacologiques ciblant spécifiquement l'activité PGDS, une campagne de criblage virtuelle a été menée. Toutefois, la condition obligatoire pour pouvoir mener cette campagne est l'identification du site de liaison de la prostaglandine H2, qui n'a jamais été déterminé chez P28GST. L'amarrage moléculaire de la prostaglandine H2 n'a pas abouti. Par superposition des structures cristallographiques de P28GST et hPGDS, chez qui le site L est identifié, il est possible de voir que le site retrouvé chez hPGDS est difficilement identifiable chez P28GST. La différence au niveau de ces sites peut être expliquée par la mutation d'un acide aminé en position 169 chez P28GST, passant d'une thréonine chez hPGDS à une histidine. Cette His169 formerait une porte conditionnant l'entrée du site L. Le positionnement du fragment de la prostaglandine H2 que nous avons tenté se fait dans une petite cavité. Au vu de la conformation des sites L de hPGDS et P28GST, l'absence de résultat correct de la prostaglandine H2 ne serait donc pas un problème

d'algorithme mais bien une possible mauvaise conformation du site L de la P28GST dans la cristallographie. La recherche d'une nouvelle conformation était donc nécessaire.

Lors de la recherche d'une conformation ouverte de la cavité par la potentielle bascule de His169, aucune n'a pu être identifiée au sein des différentes simulations de dynamique moléculaire classique, et un début de poche n'a pu être observé que de manière très transitoire dans un très faible nombre de simulations de métadynamique. En conséquence, pour forcer une bonne configuration du site L, une simulation de dynamique moléculaire classique à partir d'un résultat d'amarrage moléculaire de la prostaglandine H2 prépositionnée avec une extrémité dans cette petite cavité. Une conformation de la P28GST avec un site L accessible a pu être mis en évidence, et confirmé par l'amarrage moléculaire de l'intermédiaire réactionnel prostaglandine H2 - glutathion en retrouvant une bonne superposition avec le glutathion dans son site. L'étude des réseaux d'interactions intramoléculaires avec SINAPs en présence de  $Zn^{2+}$  a montré que His169 était moins stabilisée par Glu106. Ceci suggère une augmentation de la probabilité d'ouverture du site L induite par la présence d'un cation divalent  $Zn^{2+}$ . Cet ion interviendrait donc pour préférentiellement augmenter l'activité PGDS. Ces résultats in silico sont à confirmer biologiquement. (Figure 102)



*Figure 102. Emplacement du site L au sein de la P28GST, conditionné par le positionnement de His169*

## **6. Mécanisme d'action enzymatique**

Avec tous les éléments à notre disposition, issus des travaux de thèse réalisés ici, ainsi que les descriptions de la littérature, nous pouvons poser l'hypothèse mécanistique suivante au niveau atomique (Figure 103).

### ***a. Initiation des activités enzymatiques***

Les activités enzymatiques GST et PGDS sont toutes deux conditionnées par l'activation du glutathion sous sa forme thiolate. La déprotonation du thiol est réalisée par la tyrosine 10, l'acide aminé catalytique de la P28GST. Toutefois, pour réaliser cette action, la tyrosine doit également être activée sous sa forme phénolate.

La tyrosine possède différentes conformations en fonction de son orientation spatiale, définies par son angle dièdre Chi1. La conformation IN est la position où le phénol de la tyrosine pointe en direction de la fonction thiol du glutathion. De nombreuses interactions sont retrouvées aux alentours, notamment la réalisation de liaisons hydrogènes entre Arg21 et deux résidus Asn12 et Arg14 de la boucle catalytique, située à la suite de la tyrosine 10, la stabilisant. D'autres interactions sont retrouvées, comme des liaisons ioniques entre Arg35 et Asp33, ainsi qu'entre Arg35 et Glu34.

Au vu des différents profils énergétiques, la tyrosine est préférentiellement retrouvée dans une autre conformation, plus stable : la conformation OUT, où le phénol de la tyrosine pointe vers l'extérieur de la protéine. Le passage entre la conformation IN et la conformation OUT semble être réalisé directement, sans passer par la dernière conformation qui peut être déduite à partir des rotamères des résidus tyrosine, la conformation MID, pointant vers l'intérieur de la protéine.

La conformation OUT voit également différentes interactions, notamment la perte des liaisons ioniques réalisées avec Arg35 pour favoriser une liaison ionique entre Asp33 et Lys7, et la perte des interactions de Arg21 avec la boucle catalytique, augmentant sa flexibilité. L'environnement chimique de la tyrosine dans sa conformation OUT est favorable à sa déprotonation par la diminution de son pKa, avec la présence de nombreux acides aminés polaires.

La tyrosine 10 dans son état déprotonée retourne ensuite dans sa conformation IN pour réaliser l'activation du glutathion. La transition faisant appel à des charges non étudiées faute

d'outils adaptés, les hypothèses qui peuvent être formulées sont soit un retour de la tyrosine déprotonée de la conformation OUT à IN suivant le même chemin que le passage aller, soit en passant par la conformation MID, proposant un environnement chimique plus favorable, puis en passant de la conformation MID à IN au niveau d'une petite cavité décrite par la boucle catalytique rendue plus flexible.

Une fois le glutathion déprotoné par la tyrosine catalytique, deux activités enzymatiques sont retrouvées en fonction du ligand et de son positionnement.

### ***b. Activité Glutathion S-Transférase (GST)***

L'activité GST est permise par la présence d'un ligand hydrophobe non spécifique retrouvé au niveau du site H. Ce site H est défini par un grand nombre d'acides aminés par la coopération des deux monomères. L'activité n'étant pas spécifique de la P28GST, et son inhibition touchant l'autre activité enzymatique, celle-ci n'a pas été étudiée dans le cadre de ces travaux de thèse.

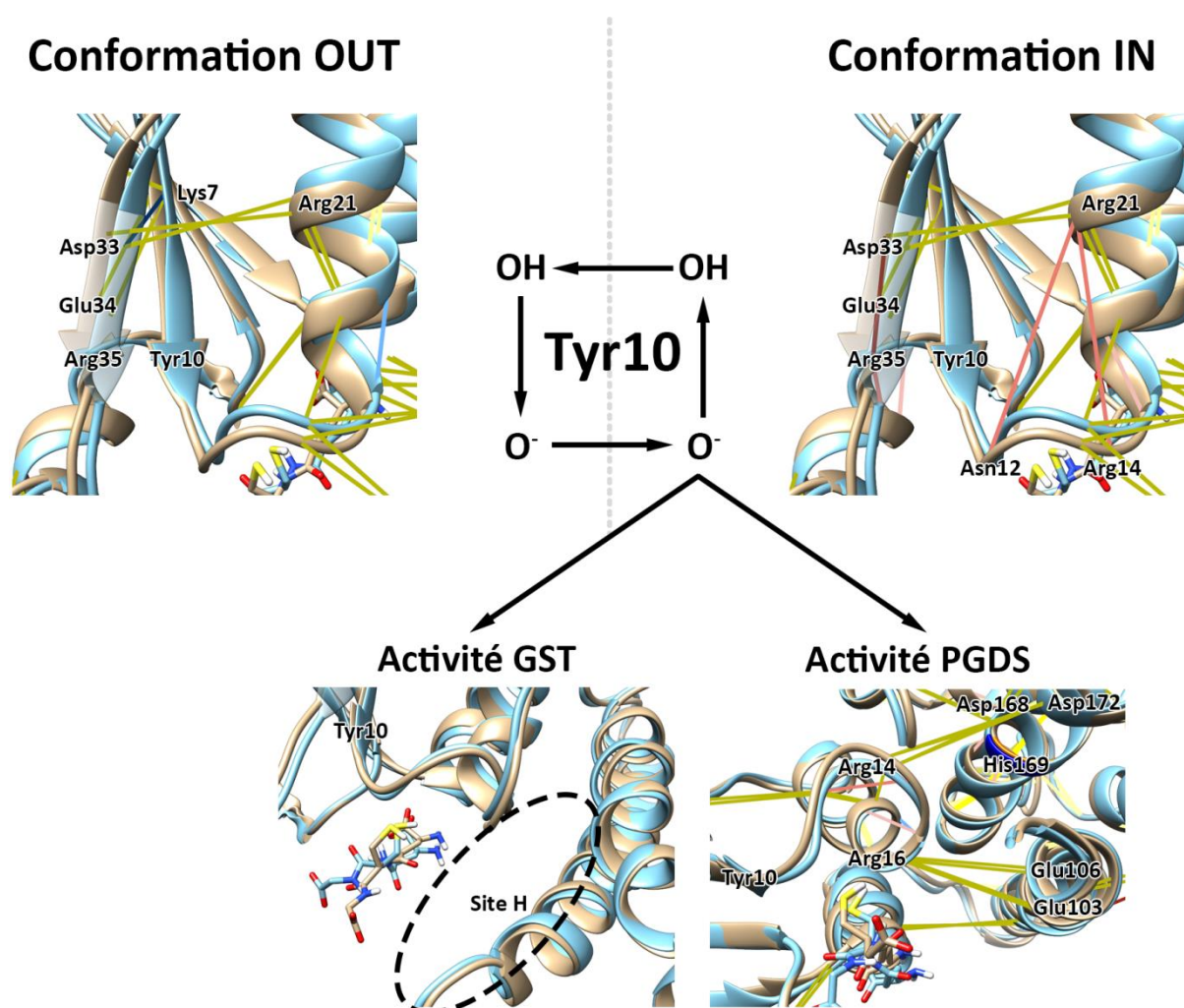
### ***c. Activité Prostaglandine-D2 Synthase (PGDS)***

L'activité PGDS est une activité permettant la conversion de la prostaglandine H2 en prostaglandine D2. L'activité PGDS est également initiée par l'action de l'anion thiolate, permettant l'obtention d'un intermédiaire réactionnel liant de manière covalente le glutathion avec la prostaglandine H2 par l'attaque de la liaison peroxyde pour former une fonction méthoxysulfanyle. Au vu de la labilité de la liaison réalisée, celle-ci se rompt pour donner la prostaglandine D2 et restaurer le glutathion sous sa forme réduite.

Pour permettre la réalisation de cette activité, la prostaglandine H2 doit être retrouvée au niveau de son site de liaison spécifique, appelé site L. Ce site est défini de part et d'autre par des liaisons ioniques stabilisant l'organisation tridimensionnelle de la P28GST, entre Arg14 et Asp168 ou Asp172 d'une part, ainsi qu'entre Arg16 et Glu103 ou Glu106 d'autre part. L'ouverture du site L est conditionnée par la présence de His169, possédant sa chaîne latérale de part et d'autre de ces liaisons ioniques.



De plus, la contribution spécifique d'un cation  $Zn^{2+}$  sur cette activité est possible, en favorisant à distance l'ouverture du site L. Ce cation peut être retrouvé au niveau de l'interface dimérique, stabilisé par Glu103, Asp104, et une molécule d'eau. La présence de ce cation, favorisant la contribution de ces acides aminés polaires, peut modifier les réseaux d'interactions intramoléculaires en favorisant la réalisation de liaisons ioniques entre Asp104 et Arg16 ainsi qu'entre Glu106 et Arg16, impliqués dans le site L. De plus, His169 pouvant être stabilisé par Glu106, la présence d'un cation divalent peut diminuer la fréquence de cette interaction augmentant ainsi ses mouvements, donc la fréquence d'ouverture du site L et l'activité enzymatique PGDS.



*Figure 103. Mécanisme d'action enzymatique de la P28GST. Les interactions communes sont représentées en jaune, spécifiques à la conformation IN en rouge et spécifiques à la conformation OUT en bleu, calculées et présentées avec SINAPs.*

## **7. Campagne exploratoire d'identification d'inhibiteurs spécifiques de l'activité PGDS**

### ***a. Campagne de criblage virtuel à haut débit***

Une fois que la conformation du site L de la P28GST a été identifiée, une campagne de criblage virtuel a pu être menée pour identifier des composés ciblant spécifiquement ce site de liaison. Il est à noter qu'il n'existe aucun ligand connu spécifique de l'activité PGDS chez P28GST, ce qui rend impossible le *benchmarking*. Les objectifs ont été définis, le site de liaison de la prostaglandine H2 est supposé correct, mais l'absence de ligands de références fait que le criblage virtuel réalisé est exploratoire.

Le principal problème identifié lors de la réalisation du criblage virtuel a été au niveau de la sélection des molécules à considérer. Le GOLDScore étant plutôt additif, donc basé sur le nombre d'atomes du composé étudié en contact avec la protéine, celui-ci tend à favoriser les plus grands composés au détriment des plus petits. C'est pour cette raison que nous avons testé différentes alternatives de la fonction de score pour améliorer la couverture de l'espace chimique.

Nous avons sélectionné 1803 composés de deux sélections différentes, à partir des 4.441.739 composés de la chimiothèque ZINC.

En parallèle, les chimiothèques "Chimiothèque Nationale" et son sous-ensemble représentatif "Chimiothèque Nationale Essentielle" ont été également criblées. Toutefois, peu de composés pertinents ont été mis en évidence, avec la sélection de respectivement 21 et 16 composés à l'issue du criblage virtuel. Ces composés n'ont pas été confirmés dans les étapes suivantes. La CN, et encore plus la CNE qui est une sous-sélection de la CN, n'échantillonnent qu'une fraction de l'espace chimique, essentiellement biaisée car développée à l'origine pour un usage en chimie thérapeutique. L'espace chimique ciblé par la poche du site L n'est pas forcément représenté dans ces chimiothèques.

### ***b. Validation par amarrage moléculaire et sélection des composés***

Pour affiner le positionnement des composés, un amarrage moléculaire plus précis en augmentant le nombre d'opérations génétiques a été réalisé. L'arrêt précoce a été conservé pour avoir une mesure de la convergence de l'algorithme, et procéder à la confirmation de la sélection en ré-appliquant les mêmes critères qu'à l'étape précédente, en incluant les limites de scores définies préalablement. A l'issue de l'amarrage moléculaire des 1803 composés issus de la chimiothèque ZINC, de nouveaux critères ont été appliqués pour diminuer progressivement ce nombre vers une sélection finale de 5 composés.

Ces critères ont notamment concerné la différenciation du site L spécifique, avec le site H non spécifique. La méthode retenue, mesurant la distance minimale entre un atome situé dans le site H et la position du composé, n'est peut être pas optimale. Du fait du manque de ligands de référence, il a été impossible de partir sur un principe de *drug design* indirect, sans pharmacophore, sans similarité de positionnement, et aucun élément d'interaction ne devait être conservé systématiquement. Ce qui nous a conduit à adopter ce principe de distance minimale.

De plus, l'inspection visuelle des composés aurait pu être réalisée plus en amont, en éliminant les composés en fonction de l'appréciation de leur positionnement. Toutefois, il a été décidé de privilégier les méthodes automatiques aux sélections empiriques, non reproductibles.

En perspective, les composés identifiés sont à valider biologiquement, pour confirmer que les hypothèses de sélection des molécules dans l'espace chimique criblé soient bonnes. J'ai contribué à l'initiation des validations expérimentales en mettant en place et en appliquant de premières méthodes biochimiques.

# CONCLUSION



Les objectifs de ce travail de thèse étaient d'utiliser les méthodes *in silico* modernes pour comprendre le mode d'action de la P28GST par l'identification du fonctionnement moléculaire et atomique des activités enzymatiques GST et PGDS, ainsi que pour identifier des composés utilisables comme sondes pharmacologiques, servant d'outils pour évaluer la contribution de chaque activité enzymatique dans l'effet anti-inflammatoire global de la P28GST.

Ce travail a permis d'étudier la structuration de la P28GST ainsi que l'identification des principales interactions dans les deux activités enzymatiques, et les barrières énergétiques conditionnant les mouvements de la tyrosine catalytique, par dynamique moléculaire classique et métadynamique. Cette étude a été menée grâce au développement d'un outil d'analyse et de visualisation des réseaux d'interactions, appelé SINAPs, et mis à la disposition de la communauté scientifique. Le site L permettant la liaison spécifique de la prostaglandine H2 a pu être identifié, et sa connaissance a permis la réalisation d'une campagne de criblage virtuel à haut débit sur plus de 4 millions de composés, affinée par amarrage moléculaire et l'application d'une succession de filtres, permettant l'identification de 5 composés ciblant spécifiquement l'activité PGDS.

Le fonctionnement enzymatique de la P28GST a donc pu être élucidé. Nous sommes passés d'une hypothèse de mécanisme d'action où le rôle essentiel d'un seul acide aminé, la tyrosine catalytique, était confirmé expérimentalement, à un schéma réactionnel dont la contribution de tous les acides aminés est connue, de l'activation du glutathion par la tyrosine catalytique à l'activité spécifique Prostaglandine-D2 Synthase.

Les hypothèses formulées tout au long de ce travail doivent être confirmées biologiquement, étape déjà initiée par la validation expérimentale des composés sélectionnés à l'étape de criblage virtuel. D'autres validations sont envisageables, comme la réalisation de mutagenèse dirigée pour valider les hypothèses mécanistiques, la réalisation de cristallographies en présence d'inhibiteurs s'ils sont confirmés, ou encore la contribution de  $Zn^{2+}$  sur les activités enzymatiques, plus spécifiquement l'activité PGDS, par des techniques biochimiques et par cristallographie. Enfin, des méthodes *in silico* plus adaptées, comme l'utilisation d'un champ de force polarisable ou de la dynamique quantique QM/MM, permettraient de prendre en compte les charges retrouvées dans le mécanisme d'action.



# **BIBLIOGRAPHIE**





1. Ross AGP, Chau TN, Inobaya MT, Olveda RM, Li Y, Harn DA. A new global strategy for the elimination of schistosomiasis. *International Journal of Infectious Diseases*. janv 2017;54:130-7.
2. Bitschai J, Brodny ML. A history of urology in Egypt. In: *A history of urology in Egypt*. 1956. p. vi-122.
3. Bryan CP. *The papyrus Ebers*. 1930.
4. BOUCHUT C. *La Bilharziose en France : État des lieux et perspectives de vaccin*. 2020.
5. Bilharz T. Fernere Beobachtungun Uber Das Pfortader Des Menschen Bewohnende Distomum Haematobium Und Sein Verhaltniss Zu Gewissen Pathologischen Bildungen Aus Brieflichen Mittheilungen an Professor v. Sielbold vom 29 Marz 1852. *Zeitschrift fur Wissenschaftliche Zoologie Leipzig*. 1852;4:72-6.
6. Weinland DF. *Human Cestoides: An Essay on the Tapeworms of Man...* Metcalf; 1858.
7. Manson P. Report of a case of bilharzia from the West Indies. *British medical journal*. 1902;2(2190):1894.
8. Weerakoon KGAD, Gobert GN, Cai P, McManus DP. Advances in the Diagnosis of Human Schistosomiasis. *Clin Microbiol Rev*. oct 2015;28(4):939-67.
9. Ndir O. *Situation des schistosomoses au Sénégal*. 2000;
10. Organisation mondiale de la Santé. *Schistosomiase*. 18 mai 2021; Disponible sur: <https://www.who.int/fr/news-room/fact-sheets/detail/schistosomiasis>
11. VIDAL Recos. *Parasitoses* [Internet]. 2021. Disponible sur: <https://evidal.vidal.fr/recos/details/4059/parasitoses/traitements>
12. Wang W, Wang L, Liang Y-S. Susceptibility or resistance of praziquantel in human schistosomiasis: a review. *Parasitol Res*. nov 2012;111(5):1871-7.
13. Balloul JM, Grzych JM, Pierce RJ, Capron A. A purified 28,000 dalton protein from *Schistosoma mansoni* adult worms protects rats and mice against experimental schistosomiasis. *J Immunol*. 15 mai 1987;138(10):3448-53.
14. Balloul JM, Sondermeyer P, Dreyer D, Capron M, Grzych JM, Pierce RJ, et al. Molecular cloning of a protective antigen of schistosomes. *Nature*. 12 mars 1987;326(6109):149-53.
15. Liu JL, Fontaine J, Capron A, Grzych JM. Ultrastructural localization of Sm28 GST protective antigen in *Schistosoma mansoni* adult worms. *Parasitology*. oct 1996;113 ( Pt 4):377-91.
16. Loison G, Vidal A, Findeli A, Roitsch C, Balloul J, Lemoine Y. High level of expression of a protective antigen of schistosomes in *Saccharomyces cerevisiae*. *Yeast*. 1989;5(6):497-507.

17. Trottein F, Godin C, Pierce RJ, Sellin B, Taylor MG, Gorillot I, et al. Inter-species variation of schistosome 28-kDa glutathione S-transferases. *Molecular and Biochemical Parasitology*. août 1992;54(1):63-72.
18. Pierce RJ, Khalife J, Williams DL, Kanno R, Trottein F, Lepresle T, et al. *Schistosoma mansoni*: Characterization of Sequence Variants of the 28-kDa Glutathione S-Transferase. *Experimental Parasitology*. août 1994;79(1):81-4.
19. Vandewaa EA, Campbell CK, Oleary KA, Tracy JW. Induction of *Schistosoma mansoni* Glutathione S-Transferase by Xenobiotics. *Archives of Biochemistry and Biophysics*. mai 1993;303(1):15-21.
20. Auriault C, Gras-Masse H, Wolowczuk I, Pierce RJ, Balloul JM, Neyrinck JL, et al. Analysis of T and B cell epitopes of the *Schistosoma mansoni* P28 antigen in the rat model by using synthetic peptides. *J Immunol*. 1 sept 1988;141(5):1687-94.
21. Auriault C, Wolowczuk I, Gras-Masse H, Marguerite M, Boulanger D, Capron A, et al. Epitopic characterization and vaccinal potential of peptides derived from a major antigen of *Schistosoma mansoni* (Sm28 GST). *Pept Res*. févr 1991;4(1):6-11.
22. Taylor JB, Vidal A, Torpier G, Meyer DJ, Roitsch C, Balloul JM, et al. The glutathione transferase activity and tissue distribution of a cloned Mr28K protective antigen of *Schistosoma mansoni*. *EMBO J*. févr 1988;7(2):465-72.
23. Boulanger D, Reid GDF, Sturrock RF, Wolowczuk I, Balloul JM, Grezel D, et al. Immunization of mice and baboons with the recombinant Sm28GST affects both worm viability and fecundity after experimental infection with *Schistosoma mansoni*. *Parasite Immunol*. sept 1991;13(5):473-90.
24. Boulanger D, Warter A, Trottein F, Mauny F, Bremond P, Audibert F, et al. Vaccination of patas monkeys experimentally infected with *Schistosoma haematobium* using a recombinant glutathione S-transferase cloned from *S. mansoni*. *Parasite Immunol*. juill 1995;17(7):361-9.
25. Xu CB, Verwaerde C, Grzych JM, Fontaine J, Capron A. A monoclonal antibody blocking the *Schistosoma mansoni* 28-kDa glutathione S-transferase activity reduces female worm fecundity and egg viability. *Eur J Immunol*. août 1991;21(8):1801-7.
26. Grzych JM, Grezel D, Xu CB, Neyrinck JL, Capron M, Ouma JH, et al. IgA antibodies to a protective antigen in human Schistosomiasis mansoni. *J Immunol*. 15 janv 1993;150(2):527-35.
27. Xu CB, Verwaerde C, Gras-Masse H, Fontaine J, Bossus M, Trottein F, et al. *Schistosoma mansoni* 28-kDa glutathione S-transferase and immunity against parasite fecundity and egg viability. Role of the amino- and carboxyl-terminal domains. *J Immunol*. 1 févr 1993;150(3):940-9.
28. Pancre V, Wolowczuk I, Guerret S, Copin MC, Delanoye A, Capron A, et al. Protective effect of rSm28GST-specific T cells in schistosomiasis: role of gamma interferon. *Infect Immun*. sept 1994;62(9):3723-30.
29. Gryseels B. Schistosomiasis. *Infectious Disease Clinics of North America*. juin 2012;26(2):383-97.

30. Capron A, Riveau GJ, Bartley PB, McManus DP. Prospects for A Schistosome Vaccine. *curr drug targets immune endocr metabol disord.* 1 oct 2002;2(3):281-90.
31. Agnew A, Fulford AJ, Mwanje MT, Gachuhi K, Gutschmann V, Krijger FW, et al. Age-dependent reduction of schistosome fecundity in *Schistosoma haematobium* but not *Schistosoma mansoni* infections in humans. *Am J Trop Med Hyg.* sept 1996;55(3):338-43.
32. Riveau G, Schacht A-M, Dompnier J-P, Deplanque D, Seck M, Waucquier N, et al. Safety and efficacy of the rSh28GST urinary schistosomiasis vaccine: A phase 3 randomized, controlled trial in Senegalese children. Dalton JP, éditeur. *PLoS Negl Trop Dis.* 7 déc 2018;12(12):e0006968.
33. Riveau G. Bilhvax : Une approche vaccinale contre la bilharziose urinaire [Internet]. 2010 mai; Institut Pasteur Paris. Disponible sur: <http://www.infectiologie.com/UserFiles/File/medias/diaporamas/CEMI/2010/CEMI-2010-RIVEAU.pdf>
34. Nau J-Y. «Bilhvax»: un premier vaccin contre la bilharziose ? *Revue Médicale Suisse* [Internet]. Disponible sur: <https://www.revmed.ch/revue-medicale-suisse/2009/revue-medicale-suisse-229/bilhvax-un-premier-vaccin-contre-la-bilharziose>
35. Riveau G, Deplanque D, Remoué F, Schacht A-M, Vodougnon H, Capron M, et al. Safety and Immunogenicity of rSh28GST Antigen in Humans: Phase 1 Randomized Clinical Study of a Vaccine Candidate against Urinary Schistosomiasis. Diemert DJ, éditeur. *PLoS Negl Trop Dis.* 3 juill 2012;6(7):e1704.
36. Clinical Trial of Bilhvax, a Vaccine Candidate Against Schistosomiasis (Bilhvax1a) [Internet]. Disponible sur: <https://clinicaltrials.gov/ct2/show/NCT01512277>
37. Espoir Pour La Santé - Études réalisées [Internet]. Disponible sur: <https://espoir-sante.org/etudes/realisees.html>
38. Boulanger D. Vaccins contre les schistosomes. Institut de Recherche pour le Développement [Internet]. janv 2000; Disponible sur: [https://horizon.documentation.ird.fr/exl-doc/pleins\\_textes/divers07/010024021.pdf](https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers07/010024021.pdf)
39. Gordon L, Hall BF, Mo AX, Walson JL, Agosti JM. Schistosomiasis Elimination Strategies and Potential Role of a Vaccine in Achieving Global Health Goals. *The American Journal of Tropical Medicine and Hygiene.* 8 janv 2014;90(1):54-60.
40. Efficacy of Bilhvax in Association With Praziquantel for Prevention of Clinical Recurrences of *Schistosoma Haematobium* (Bilhvax3) [Internet]. Disponible sur: <https://clinicaltrials.gov/ct2/show/NCT00870649>
41. Sipahi AM, Baptista DM. Helminths as an alternative therapy for intestinal diseases. *WJG.* 7 sept 2017;23(33):6009-15.
42. Driss V, El Nady M, Delbeke M, Rousseaux C, Dubuquoy C, Sarazin A, et al. The schistosome glutathione S-transferase P28GST, a unique helminth protein, prevents intestinal inflammation in experimental colitis through a Th2-type response with mucosal eosinophils. *Mucosal Immunol.* mars 2016;9(2):322-35.
43. El Nady M. Rôle immunomodulateur de la P28GST, une enzyme recombinant du parasite helminthe *Schistosoma* dans la prévention de la colite expérimentale. 2012.

44. Antoniou E, Margonis GA, Angelou A, Pikouli A, Argiri P, Karavokyros I, et al. The TNBS-induced colitis animal model: An overview. *Annals of Medicine and Surgery*. nov 2016;11:9-15.
45. Sarazin A, Dendooven A, Delbeke M, Gatault S, Pagny A, Standaert A, et al. Treatment with P28GST, a schistosome-derived enzyme, after acute colitis induction in mice: Decrease of intestinal inflammation associated with a down regulation of Th1/Th17 responses. *Arafa HM, éditeur. PLoS ONE*. 28 déc 2018;13(12):e0209681.
46. Wammes LJ, Mpairwe H, Elliott AM, Yazdanbakhsh M. Helminth therapy or elimination: epidemiological, immunological, and clinical considerations. *The Lancet Infectious Diseases*. nov 2014;14(11):1150-62.
47. Rook GAW, Martinelli R, Brunet LR. Innate immune responses to mycobacteria and the downregulation of atopic responses: Current Opinion in Allergy and Clinical Immunology. oct 2003;3(5):337-42.
48. Khan AR, Fallon PG. Helminth therapies: Translating the unknown unknowns to known knowns. *International Journal for Parasitology*. mars 2013;43(3-4):293-9.
49. Capron M, NADY ME, COLOMBEL J-F, Riveau G. Proteines 28kda gst provenant de schistosomes pour leur utilisation dans le traitement des maladies inflammatoires auto-immunes engendrant une reponse de type th1 et/ou th17 [Internet]. WO2013117860A1, 201. Disponible sur: <https://patents.google.com/patent/WO2013117860A1/fr?oq=WO2013117860>
50. Capron M, El NM, Colombel J-F, Riveau G. Proteines 28kda gst provenant de schistosomes pour leur utilisation dans le traitement des maladies inflammatoires auto-immunes engendrant une reponse de type th1 et/ou th17 [Internet]. FR2986432B1, 2014. Disponible sur: <https://patents.google.com/patent/FR2986432B1/fr?oq=WO2013117860>
51. Gower-Rousseau C, Salomez JL, Dupas JL, Marti R, Nuttens MC, Votte A, et al. Incidence of inflammatory bowel disease in northern France (1988-1990). *Gut*. 1 oct 1994;35(10):1433-8.
52. Chouraki V, Savoye G, Dauchet L, Vernier-Massouille G, Dupas J-L, Merle V, et al. The changing pattern of Crohn's disease incidence in northern France: a continuing increase in the 10- to 19-year-old age bracket (1988-2007): Epidemiology of inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics*. mai 2011;33(10):1133-42.
53. EPIMAD, l'étude prospective d'incidence la plus importante au monde concernant les maladies inflammatoires du tube digestif [Internet]. Disponible sur: <https://epimad.chru-lille.fr/index.html>
54. Gower-Rousseau C, Vasseur F, Fumery M, Savoye G, Salleron J, Dauchet L, et al. Epidemiology of inflammatory bowel diseases: New insights from a French population-based registry (EPIMAD). *Digestive and Liver Disease*. févr 2013;45(2):89-94.
55. Agence Nationale de la Recherche. ANR-11-RPIB-0021 - Une enzyme parasitaire comme nouvel outil thérapeutique des maladies inflammatoires chroniques de l'intestin : ACROHNEM [Internet]. Disponible sur: <https://anr.fr/Projet-ANR-11-RPIB-0021>
56. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases With Time, Based on Systematic Review. *Gastroenterology*. janv 2012;142(1):46-54.e42.

57. Kaplan GG, Ng SC. Understanding and Preventing the Global Increase of Inflammatory Bowel Disease. *Gastroenterology*. janv 2017;152(2):313-321.e2.
58. Burisch J, Munkholm P. The epidemiology of inflammatory bowel disease. *Scandinavian Journal of Gastroenterology*. 3 août 2015;50(8):942-51.
59. Rose N, éditeur. *The autoimmune diseases*. 6th ed. Waltham: Elsevier; 2019.
60. Malnutrition and IBD - Crohn's & Colitis Foundation [Internet]. Disponible sur: <https://www.crohnscolitisfoundation.org/diet-and-nutrition/malnutrition-and-ibd>
61. Best WR, Beckett JM, Singleton JW, Kern F. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology*. mars 1976;70(3):439-44.
62. Maladie de Crohn - VIDAL Recos [Internet]. 2021. Disponible sur: <https://www.vidal.fr/maladies/recommandations/crohn-maladie-de-3751.html>
63. Rectocolite hémorragique - VIDAL Recos [Internet]. 2021. Disponible sur: <https://www.vidal.fr/maladies/recommandations/rectocolite-hemorragique-4021.html>
64. Capron M, Béghin L, Leclercq C, Labreuche J, Dendooven A, Standaert A, et al. Safety of P28GST, a Protein Derived from a Schistosome Helminth Parasite, in Patients with Crohn's Disease: A Pilot Study (ACROHNEM). *JCM*. 24 déc 2019;9(1):41.
65. University Hospital, Lille. Safety Study of P28GST Treatment in Crohn's Disease Patients, a Multicenter Phase 2 Clinical Trial [Internet]. [clinicaltrials.gov](https://clinicaltrials.gov); 2018 juin. Report No.: NCT02281916. Disponible sur: <https://clinicaltrials.gov/ct2/show/NCT02281916>
66. Vos RME, Van Bladeren PJ. Glutathione S-transferases in relation to their role in the biotransformation of xenobiotics. *Chemico-Biological Interactions*. 1990;75(3):241-65.
67. Dancygier H. Hepatic Biotransformation. In: *Clinical Hepatology* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 127-30. Disponible sur: [http://link.springer.com/10.1007/978-3-540-93842-2\\_8](http://link.springer.com/10.1007/978-3-540-93842-2_8)
68. Jacquot JP, Dietz KJ, Rouhier N, Meux E, Lallement PA, Selles B, et al. Redox Regulation in Plants: Glutathione and "Redoxin" Related Families. In: Jakob U, Reichmann D, éditeurs. *Oxidative Stress and Redox Regulation* [Internet]. Dordrecht: Springer Netherlands; 2013. p. 213-31. Disponible sur: [http://link.springer.com/10.1007/978-94-007-5787-5\\_8](http://link.springer.com/10.1007/978-94-007-5787-5_8)
69. Awasthi YC, éditeur. *Toxicology of glutathione transferases*. Boca Raton, FL: CRC Taylor & Francis; 2007. 375 p.
70. Pryszyzhnyuk V, Sydoruk L, Sydoruk R, Pryszyzhniuk I, Bobkovich K, Buzdugan I, et al. Glutathione-S-transferases genes-promising predictors of hepatic dysfunction. *WJH*. 27 juin 2021;13(6):620-33.
71. Torres-Rivera A, Landa A. Glutathione transferases from parasites: A biochemical view. *Acta Tropica*. févr 2008;105(2):99-112.
72. Allocati N, Federici L, Masulli M, Di Ilio C. Glutathione transferases in bacteria: Bacterial GSTs. *FEBS Journal*. janv 2009;276(1):58-75.

73. Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. *Insect Mol Biol.* janv 2005;14(1):3-8.
74. Dixon DP, Laphorn A, Edwards R. Plant glutathione transferases. *Genome Biol.* 2002;3(3):REVIEWS3004.
75. Sheehan D, Meade G, Foley VM, Dowd CA. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal.* 15 nov 2001;360(1):1-16.
76. Oakley A. Glutathione transferases: a structural perspective. *Drug Metabolism Reviews.* mai 2011;43(2):138-51.
77. Mannervik B, Board PG, Hayes JD, Listowsky I, Pearson WR. Nomenclature for Mammalian Soluble Glutathione Transferases. In: *Methods in Enzymology* [Internet]. Elsevier; 2005. p. 1-8. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S0076687905010013>
78. Atkinson HJ, Babbitt PC. Glutathione Transferases Are Structural and Functional Outliers in the Thioredoxin Fold. *Biochemistry.* 24 nov 2009;48(46):11108-16.
79. Hayes JD, Flanagan JU, Jowsey IR. Glutathione Transferases. *Annu Rev Pharmacol Toxicol.* 22 sept 2005;45(1):51-88.
80. Johnson KA, Angelucci F, Bellelli A, Hervé M, Fontaine J, Tsernoglou D, et al. Crystal Structure of the 28 kDa Glutathione S-Transferase from *Schistosoma haematobium*. *Biochemistry.* sept 2003;42(34):10084-94.
81. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology.* avr 1995;247(4):536-40.
82. Martin JL. Thioredoxin —a fold for all reasons. *Structure.* mars 1995;3(3):245-50.
83. Nathaniel C, Wallace LA, Burke J, Dirr HW. The role of an evolutionarily conserved cis-proline in the thioredoxin-like domain of human class Alpha glutathione transferase A1-1. *Biochemical Journal.* 15 mai 2003;372(1):241-6.
84. Piromjitpong J, Wongsantichon J, Ketterman AJ. Differences in the subunit interface residues of alternatively spliced glutathione transferases affects catalytic and structural functions. *Biochemical Journal.* 1 févr 2007;401(3):635-44.
85. Armstrong RN. Structure, Catalytic Mechanism, and Evolution of the Glutathione Transferases. *Chem Res Toxicol.* 1 janv 1997;10(1):2-18.
86. Board PG, Menon D. Glutathione transferases, regulators of cellular metabolism and physiology. *Biochimica et Biophysica Acta (BBA) - General Subjects.* mai 2013;1830(5):3267-88.
87. Brock J, Board PG, Oakley AJ. Structural Insights into Omega-Class Glutathione Transferases: A Snapshot of Enzyme Reduction and Identification of a Non-Catalytic Ligandin Site. Kobe B, éditeur. *PLoS ONE.* 9 avr 2013;8(4):e60324.

88. Pinzar E, Miyano M, Kanaoka Y, Urade Y, Hayaishi O. Structural Basis of Hematopoietic Prostaglandin D Synthase Activity Elucidated by Site-directed Mutagenesis. *Journal of Biological Chemistry*. oct 2000;275(40):31239-44.
89. Walker J, Crowley P, Moreman AD, Barrett J. Biochemical properties of cloned glutathione S-transferases from *Schistosoma mansoni* and *Schistosoma japonicum*. *Molecular and Biochemical Parasitology*. oct 1993;61(2):255-64.
90. Kanaoka Y, Ago H, Inagaki E, Nanayama T, Miyano M, Kikuno R, et al. Cloning and Crystal Structure of Hematopoietic Prostaglandin D Synthase. *Cell*. sept 1997;90(6):1085-95.
91. Meyer DJ, Thomas M. Characterization of rat spleen prostaglandin H d-isomerase as a sigma-class GSH transferase. *Biochemical Journal*. 1 nov 1995;311(3):739-42.
92. Tomarev SI, Zinovieva RD. Squid major lens polypeptides are homologous to glutathione S-transferases subunits. *Nature*. nov 1988;336(6194):86-8.
93. Flanagan JU, Smythe ML. Sigma-class glutathione transferases. *Drug Metabolism Reviews*. mai 2011;43(2):194-214.
94. Ji X, von Rosenvinge EC, Johnson WW, Tomarev SI, Piatigorsky J, Armstrong RN, et al. Three-dimensional structure, catalytic properties, and evolution of a sigma class glutathione transferase from squid, a progenitor of the lens S-crystallins of cephalopods. *Biochemistry*. 25 avr 1995;34(16):5317-28.
95. Inoue T, Irikura D, Okazaki N, Kinugasa S, Matsumura H, Uodome N, et al. Mechanism of metal activation of human hematopoietic prostaglandin D synthase. *Nat Struct Mol Biol*. 1 avr 2003;10(4):291-6.
96. Brunton LL, Knollmann BC, Hilal-Dandan R, éditeurs. *Goodman & Gilman's the pharmacological basis of therapeutics*. Thirteenth edition. New York: McGraw Hill Medical; 2018. 1419 p.
97. Seo M-J, Oh D-K. Prostaglandin synthases: Molecular characterization and involvement in prostaglandin biosynthesis. *Progress in Lipid Research*. avr 2017;66:50-68.
98. Urade Y, Fujimoto N, Ujihara M, Hayaishi O. Biochemical and immunological characterization of rat spleen prostaglandin D synthetase. *J Biol Chem*. 15 mars 1987;262(8):3820-5.
99. Hervé M, Angeli V, Pinzar E, Wintjens R, Faveeuw C, Narumiya S, et al. Pivotal roles of the parasite PGD2 synthase and of the host D prostanoid receptor 1 in schistosome immune evasion. *Eur J Immunol*. oct 2003;33(10):2764-72.
100. Baiocco P, Gourlay LJ, Angelucci F, Fontaine J, Hervé M, Miele AE, et al. Probing the Mechanism of GSH Activation in *Schistosoma haematobium* Glutathione-S-transferase by Site-directed Mutagenesis and X-ray Crystallography. *Journal of Molecular Biology*. juill 2006;360(3):678-89.
101. Angelucci F, Baiocco P, Brunori M, Gourlay L, Morea V, Bellelli A. Insights into the Catalytic Mechanism of Glutathione S-Transferase: The Lesson from *Schistosoma haematobium*. *Structure*. sept 2005;13(9):1241-6.



102. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. sept 2015;1-2:19-25.
103. Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham, III TE, Cruzeiro VWD. AMBER 2018. University of California, San Francisco. 2018;
104. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*. févr 2014;185(2):604-13.
105. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. *Proteins*. 1 août 2003;52(4):609-23.
106. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. Boston, MA: RStudio, PBC.; 2020. Disponible sur: <http://www.rstudio.com/>
107. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. oct 2004;25(13):1605-12.
108. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. févr 1996;14(1):33-8.
109. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME: The Konstanz Information Miner. In: *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer; 2007.
110. Dassault Systèmes. BIOVIA, Dassault Systèmes, Discovery Studio Visualizer, San Diego. 2021.
111. Dassault Systèmes. BIOVIA, Dassault Systèmes, BIOVIA Draw, San Diego. 2021.
112. Advanced Chemistry Development, Inc., Toronto, ON, Canada. ACD/ChemSketch, version 2021.1.1 [Internet]. 2021. Disponible sur: [www.acdlabs.com](http://www.acdlabs.com)
113. Docker CE. Software available from [docker.com](https://www.docker.com).
114. Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. Guroy A, éditeur. *PLoS ONE*. 11 mai 2017;12(5):e0177459.
115. worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data | *Nucleic Acids Research* | Oxford Academic [Internet]. Disponible sur: [https://academic.oup.com/nar/article/35/suppl\\_1/D301/1115659?login=true](https://academic.oup.com/nar/article/35/suppl_1/D301/1115659?login=true)
116. Leach AR. *Molecular modelling: principles and applications*. 2nd ed. Harlow, England ; New York: Prentice Hall; 2001. 744 p.
117. Tuckerman ME, Martyna GJ. *Understanding Modern Molecular Dynamics: Techniques and Applications*. *J Phys Chem B*. janv 2000;104(2):159-78.
118. Schlick T. *Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide* [Internet]. New York, NY: Springer New York; 2010. (Interdisciplinary Applied Mathematics; vol. 21). Disponible sur: <http://link.springer.com/10.1007/978-1-4419-6351-2>

119. Vlugt TJH, Eerden JPJM van der, Dijkstra M, Smit B, Frenkel D. Introduction to molecular simulation and statistical thermodynamics [Internet]. 2009. Disponible sur: <http://homepage.tudelft.nl/v9k6y/imsst/book-15-6-2009.pdf>
120. Madej BD, Walker R. An Introduction to Molecular Dynamics Simulations using AMBER [Internet]. 2015. Disponible sur: <https://ambermd.org/tutorials/basic/tutorial0/index.htm>
121. Ganesan A, Coote ML, Barakat K. Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today*. févr 2017;22(2):249-69.
122. Duneau J-P. Simulation de dynamique moléculaire [Internet]. 2004. Disponible sur: [https://lism.cnrs-mrs.fr/JS\\_files/Page\\_JP/Biomolec/DM5DynTheo.pdf](https://lism.cnrs-mrs.fr/JS_files/Page_JP/Biomolec/DM5DynTheo.pdf)
123. Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. 2nd ed. San Diego: Academic Press; 2002. 638 p. (Computational science series).
124. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 15 juill 1983;79(2):926-35.
125. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. *J Chem Phys*. 1 août 1996;105(5):1902-21.
126. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J*. juill 2011;40(7):843-56.
127. Ewald PP. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der physik*. 1921;369(3):253-87.
128. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 11 août 2015;11(8):3696-713.
129. Yuan Y, Pei J, Lai L. LigBuilder 2: A Practical *de Novo* Drug Design Approach. *J Chem Inf Model*. 23 mai 2011;51(5):1083-91.
130. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*. 2009;10(1):168.
131. Durrant JD, Votapka L, Sørensen J, Amaro RE. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J Chem Theory Comput*. 11 nov 2014;10(11):5047-56.
132. Shapovalov MV, Dunbrack RL. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure*. juin 2011;19(6):844-58.
133. Hao G-F, Xu W-F, Yang S-G, Yang G-F. Multiple Simulated Annealing-Molecular Dynamics (MSA-MD) for Conformational Space Search of Peptide and Miniprotein. *Sci Rep*. déc 2015;5(1):15568.

134. Lefebvre D. L'algorithme du recuit simulé [Internet]. Disponible sur: <http://www.tangentex.com/RecuitSimule.htm>
135. The Bowers Group. Theory/Analysis - Simulated Annealing [Internet]. Disponible sur: [https://labs.chem.ucsb.edu/bowers/michael/theory\\_analysis/mm/annealing.shtml](https://labs.chem.ucsb.edu/bowers/michael/theory_analysis/mm/annealing.shtml)
136. Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*. avr 2001;11(2):224-30.
137. Liao Q. Enhanced sampling and free energy calculations for protein simulations. In: *Progress in Molecular Biology and Translational Science* [Internet]. Elsevier; 2020. p. 177-213. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S187711732030017X>
138. Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep Prog Phys*. 1 déc 2008;71(12):126601.
139. Pan F, Moradi M, Roland C, Celeste S. The Nonequilibrium Free Energy (NFE) Toolkit for PMEMD AMBER: Capabilities, Examples, and Collective Variables [Internet]. 2017. Disponible sur: <http://ambermd.org/tutorials/advanced/tutorial31/>
140. Shirts MR, Klein C, Swails JM, Yin J, Gilson MK, Mobley DL, et al. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J Comput Aided Mol Des*. janv 2017;31(1):147-61.
141. Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology*. févr 2001;11(1):39-46.
142. Kluska K, Adamczyk J, Krężel A. Metal binding properties, stability and reactivity of zinc fingers. *Coordination Chemistry Reviews*. juill 2018;367:18-64.
143. Yao S, Flight RM, Rouchka EC, Moseley HNB. A less-biased analysis of metalloproteins reveals novel zinc coordination geometries. *Proteins*. août 2015;83(8):1470-87.
144. Lill MA, Future Science Ltd. In silico drug discovery and design [Internet]. 2013. Disponible sur: <http://www.futuremedicine.com/doi/pdf/10.4155/9781909453012>
145. Lengauer T, Rarey M. Computational methods for biomolecular docking. *Current Opinion in Structural Biology*. juin 1996;6(3):402-6.
146. Stouten PFW, Kroemer RT. Docking and Scoring. In: *Comprehensive Medicinal Chemistry II* [Internet]. Elsevier; 2007. p. 255-81. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B008045044X002534>
147. Leach AR, Gillet VJ. An introduction to chemoinformatics [Internet]. Dordrecht; London: Springer; 2007. Disponible sur: <http://www.springerlink.com/openurl.asp?genre=book&isbn=978-1-4020-6290-2>
148. Cambridge Crystallographic Data Centre. GOLD User Guide - A Component of the CSD-Discovery Suite [Internet]. 2019. Disponible sur: [https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/GOLD\\_User\\_Guide.pdf](https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/GOLD_User_Guide.pdf)

149. Sterling T, Irwin JJ. ZINC 15 – Ligand Discovery for Everyone. *J Chem Inf Model*. 23 nov 2015;55(11):2324-37.
150. Baell JB, Holloway GA. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *Journal of Medicinal Chemistry*. 8 avr 2010;53(7):2719-40.
151. Baell JB, Nissink JWM. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017—Utility and Limitations. *ACS Chemical Biology*. 19 janv 2018;13(1):36-44.
152. ChemBioFrance - Chimiothèques [Internet]. 2021. Disponible sur: <https://chembiofrance.cn.cnrs.fr/fr/composante/chimiotheque>
153. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*. janv 1980;36(22):3219-28.
154. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem*. avr 1996;17(5-6):490-519.
155. Salazar A, Velázquez-Soto H, Ayala-Balboa J, Jiménez-Martínez MC. Allergen-Based Diagnostic: Novel and Old Methodologies with New Approaches. In: Athari SS, éditeur. Allergen [Internet]. InTech; 2017. Disponible sur: <http://www.intechopen.com/books/allergen/allergen-based-diagnostic-novel-and-old-methodologies-with-new-approaches>
156. Contreras-Riquelme S, Garate J-A, Perez-Acle T, Martin AJM. RIP-MD: a tool to study residue interaction networks in protein molecular dynamics. *PeerJ*. 7 déc 2018;6:e5998.
157. Booth GH, Schulert AR. Zinc Metabolism in Schistosomes. *Experimental Biology and Medicine*. 1 mars 1968;127(3):700-4.
158. Morley NJ, Crane M, Lewis JW. Toxicity of cadmium and zinc to miracidia of *Schistosoma mansoni*. *Parasitology*. janv 2001;122(1):81-5.