

UNIVERSITE DE NICE SOPHIA ANTIPOLIS - UFR Sciences
Ecole Doctorale des Sciences Fondamentales et Appliquées

T H E S E

pour l'obtention du titre de
Docteur en Sciences
De l'Université de Nice Sophia Antipolis

Discipline : Chimie

Présentée et soutenue par
Julien DIHARCE

Etude par modélisation moléculaire de systèmes multienzymatiques impliqués dans la biosynthèse des flavonoïdes

Thèse dirigée par le Pr. Serge ANTONCZAK
Soutenue le 4 Décembre 2014

Jury :

-Pr. Gérard Monard	Université de Lorraine, Vandœuvre-lès-Nancy	Rapporteur
-Dr. Daniel Borgis	Ecole Normale Supérieure, Paris	Rapporteur
-Pr. Stéphane Quideau	Université de Bordeaux, Bordeaux	Examineur
-Dr. Sylvain Antoniotti	Université Nice Sophia Antipolis, Nice	Examineur
-Dr. Sébastien Fiorucci	Université Nice Sophia Antipolis, Nice	Co-directeur
-Pr. Serge Antonczak	Université Nice Sophia Antipolis, Nice	Directeur de thèse

Pour vous, Rose et Antoine.

Puissiez-vous être fiers de moi et de ce que je suis devenu grâce à vous.

Remerciements

La réalisation d'une thèse est une formation permettant la réalisation d'un projet professionnel, mais également une aventure humaine assez incroyable. De ce fait, la réalisation et l'aboutissement de ce manuscrit, et de cette étude en général, n'auraient pu se faire sans l'aide de plusieurs personnes à qui je tiens à rendre hommage à travers ces lignes.

Je remercie tout d'abord mon directeur de thèse, le professeur Serge Antonczak, pour tout. Tu n'as pas hésité à me tendre la main pour mon premier stage dans la recherche, il y a de cela 6 ans maintenant! Tu m'as formé, m'as appris énormément de choses durant ces trois ans (et même 6 en fait) et a permis de faire de moi la personne que je suis maintenant. Toujours à l'écoute, tu n'as pas non plus hésité à me recadrer quand il le fallait. Sans toi, je ne serai pas ce que je suis aujourd'hui, donc merci pour tout. Je retiendrai quand même nos différences de point de vue sur la visualisation des membranes!

Je remercie ensuite le docteur Sébastien Fiorucci qui m'a formé de manière complète au docking protéine-protéine. Sans toi, une bonne partie de ce travail n'aurait pas pu voir le jour. Merci également pour ces discussions acharnées sur les différentes méthodes de simulation de dynamique des systèmes motorisés, et ce dans divers paysages et climats.

Je remercie chaleureusement le professeur Gérard Monard et le docteur Daniel Borgis qui m'ont fait l'honneur de juger ce travail et de partager leur expertise. Pour cela, je leur adresse ma gratitude et mon respect les plus sincères.

Je remercie également le professeur Stéphane Quideau et le docteur Sylvain Antoniotti pour faire également parti de ce jury, et d'avoir accepté d'examiner ce travail.

Je tiens à remercier le docteur Jérôme Golebiowski, sans qui ce manuscrit n'aurait pu être écrit. Merci également pour ton regard critique sur l'aspect scientifique de ce travail, qui n'a jamais été relégué au second plan dans nos nombreuses conversations.

D'un point de vue plus personnel :

Je tiens à remercier une seconde fois les membres de l'équipe : Serge, pour toutes nos discussions : les films, séries, ou même nos très bons repas ensemble (à quand une seconde descente au O'neil?). Jérôme, merci pour nos autres discussions, beaucoup plus légères, et ton humour omniprésent. Comment pourrai-je également oublier ton incroyable sens de la compétition et du fair-play? Seb, je me souviendrai toujours de ton comportement sportif lors des matches de badminton!

Merci également au docteur Martine Adrian-Scotto, une personne de bon conseil, ainsi qu'au docteur Fabien Fontaine-Vive pour les bonnes anecdotes mais également les intéressantes discussions que nous avons eu. Dommage qu'il puisse encore croire que Nice est toujours rattaché au comté de Savoie !

Merci à Claire, communément appelée "sulkoon", pour le quotidien de ces deux ans, pour la gentillesse, l'écoute et le partage dont tu fais preuve, pour notre complicité, nos délires, nos soirées ainsi que tes compétences uniques en danse et en descente ! Dire que maintenant, tu programmes en python... Merci pour tout, je suis chanceux d'avoir gagné une amie telle que toi durant ce laps de temps et espère te rendre ce que tu m'as donné.

Mon cher Jean-Mathieu "pucetino", comment t'oublier? On commence enfin à y arriver hein! Merci d'être mon ami, de nos délires, de ta présence, notamment durant les derniers temps. Bientôt ton tour de soutenir ! En attendant je commence à regarder les beaux paysages australiens... et je réfléchis aux Pinsdiharces !

A ma chère Emmanuelle, "ma manue", merci pour ta bonne humeur, ta folie au quotidien. Travailler avec toi fut enchanteur, mais être ami avec toi l'est encore plus. J'espère que Serge et moi avons réussi à faire de toi une future scientifique exemplaire. A quand la formation finale de notre air-band ? Puis je n'oublie pas ton projet annexe...

Je tiens à remercier grandement le docteur Jérémie Jacques Daniel Grigoire Evariste Topin, ainsi que son épouse, le docteur Cornélia "Conny" Topin Meinert, tous deux chercheurs de profession. Jérémie, je n'oublierai jamais ton langage "châtié" ainsi que tes bagarres formidables mais perdues d'avance. Merci mon pote pour tes différentes aides et ton sens de l'esthétique. Conny, tes compétences en danse sont inoubliables, ainsi que ton rire sonore et unique, un peu comme moi ! Puis je te remercie également pour nos discussions, qui ont été très fructueuses à tous les niveaux.

Un grand merci pour mes collègues de l'étage : Xavier et son cynisme avéré, Alex et ses sublimes caleçons, Emilie pour son langage d'église, Audrey et ses cheveux, Maxime et son incroyable swing, Jean-Jacques et nos bonnes lasagnes de kangourou ou de cheval, Nico pour son humour toujours piquant mais efficace, ainsi que Thomas, le nouveau venu breton qui cuisine bien ! Ces trois années n'auraient pas eu la même saveur sans vous !

Tir groupé pour les doctorants du 1er étage : ma "petite" Mélody, qui m'a permis de voir Fréjus sous un autre jour et ma chère Marie-Aude, pour notre plongée sous-marine et nos conseils de labo mouvementés. Je n'oublie pas les collègues du LPMC : Amandine, Guillaume, Sébastien, Nathanaël et Matthieu! Evidemment, je n'oublie pas Alice avec qui nous avons eu de superbes discussions. Merci à vous tous pour ces petits moments forts sympathiques.

Merci à tous mes autres collègues du bâtiment, notamment : Momo-san et ses vannes au quotidien, Babeth et sa touche d'humour irrésistible ainsi que tous les petits stagiaires que j'ai eu le plaisir de côtoyer et/ou de former (petite pensée pour toi Bichon !). Puis mes chers LARdons du dessus : Vincent, Pierrick et Antoine. Je ne pourrai pas oublier notre escapade à Turin, qui restera dans les annales !

Ces remerciements ne sont pas complets si je ne citai pas les personnes extérieures au laboratoire, toujours là pour faire du bien : Merci à Stéphane et à Alex pour les gros délires et la touche si particulière qu'ils apportent, à Nico et ses expressions corrosives inégalées, à Sylvia et Jérôme les pontaults-combalusiens.

Je tenais à remercier ma famille, notamment mes parents qui m'ont toujours soutenu dans mes décisions (parfois discutables, certes) et mes envies, et qui ont toujours été là pour moi. Vous êtes juste formidables. Merci à mes grands parents, ma chère cousine qui compte beaucoup, tonton et tatie. Jessica, Odette, Leslie et Didier, merci pour tous ces moments, trop rares !

Enfin, je tenais à te remercier particulièrement ma formidable puce. Nous avons fêté récemment nos 10 ans ensemble et tu n'es pas encore saturée de moi! Merci pour tout, d'être là, de m'avoir soutenu, de m'avoir supporté, de m'avoir écouté, conseillé, alors que je fus difficile à vivre dans les derniers temps. Je ne te le dis pas souvent, mais je n'aurai pas pu et ne pourrai pas bien aller loin sans toi et toutes tes petites attentions. On repart pour 10 ans !

Table des matières

Remerciements	3
Liste des figures	11
Liste des Abréviations	17
Introduction	19
I) Position du sujet	20
II) Les flavonoïdes dans Vitis vinifera	22
II.1) Organisme d'étude : Vitis vinifera.	22
II.2) Schéma de biosynthèse des flavonoïdes	24
II.3) Les flavonoïdes	25
II.4) Structure des flavonoïdes.....	26
II.4.a) Les flavonols et les 2,3-dihydroflavonols.....	28
II.4.b) Les flavanols et les flavan-3,4-diols.....	29
II.4.c) Les anthocyanidines et anthocyanes	30
II.4.d) Les proanthocyanidines.....	31
II.5) Description du processus de biosynthèse des flavonoïdes.....	32
II.6) Complexes multienzymatiques	36
III) Description des enzymes choisies pour l'étude	40
III.1) La dihydroflavonol-4-réductase (DFR)	40
III.2) La leucocanthocyanidine réductase (LAR)	42
III.3) La flavonoïde-3'-hydroxylase (F3'H)	44
IV) Apport de précédentes études théoriques à la compréhension de la biosynthèse des flavonoïdes.	45
Problématique générale	51
Méthodologie	57
I) Modélisation Moléculaire	58
I.1) Minimisation d'énergie	60
I.1.a) Méthode de la plus grande pente / Steepest-descent	61
I.1.b) Gradient conjugué	61
I.2) Dynamique Moléculaire	62
I.2.a) Algorithme de Verlet.....	63
I.2.b) Algorithme Leapfrog.....	64
I.2.c) Pas d'intégration.....	65

I.2.a) Thermostat de Langevin	65
I.2.b) Conditions périodiques et Maillage particulier d'Ewald	66
I.3) Simulation de type RAMD (Random Acceleration Molecular Dynamics)	68
I.4) Potentiel de force moyenne et Umbrella Sampling.....	69
I.5) Méthode de docking protéine protéine : protocole ATTRACT.	71
I.6) Protocole MM-GBSA	73
II) Mécanique quantique.....	74
II.1) Méthode Hartree-Fock	75
II.2) Méthode de la fonctionnelle de la densité (DFT)	76
Chapitre 1 : Préparation des systèmes	79
I) Préparation des structures enzymatiques	80
I.1) DFR	80
I.2) LAR.....	80
I.3) F3'H	85
II) Paramètres du champ de force	87
II.1) Représentation des enzymes	87
II.2) Représentation des partenaires biologiques	88
II.2.a) Cofacteurs	88
II.2.b) Substrat.....	88
II.3) Préparation des systèmes avec membrane	89
III) Protocole de simulation de dynamique moléculaire.....	91
Chapitre 2 : Etude de la réactivité de DFR.....	93
I) Méthode QM/MM : Théorie	95
I.1) Principe.....	95
I.2) Délimitation entre partie quantique et partie classique	96
II) Protocole des simulations hybrides.....	97
II.1) Choix de la partie quantique	97
II.2) Apport du protocole d'Umbrella Sampling et choix des contraintes	98
III) Interactions entre les différents substrats et l'enzyme - calculs US-DM QM/MM à une dimension	101
III.1) Simulation sans contraintes DM classique et hybrides QM/MM.....	101
III.2) Calculs DM-QM/MM US à une dimension	103
IV) Calculs DM-QM/MM US 2D	105
IV.1) Substrat n°1 : DHQ (dihydroquercétine).....	106
IV.1.a) Analyse énergétique du profil	106
IV.1.b) Analyse structurale.....	109

IV.2) Substrat n°2 : DHK (dihydrokaempferol)	112
IV.2.a) Analyse énergétique du profil	112
IV.2.b) Analyse structurale du mécanisme	113
V) Complexe pré-réactif (Near-attack conformation)	116
V.1) DHQ (Dihydroquercétine)	116
V.2) DHK (Dihydrokaempferol)	120
VI) Conclusion	122
Chapitre 3 : Modèles DFR-LAR et "Substrate channeling"	125
I) Identification des chemins d'accès aux sites actifs des enzymes.....	127
I.1) Protocoles de simulation	127
I.2) Résultats pour l'enzyme DFR.....	128
I.2.a) Simulations de type RAMD du complexe enzyme-cofacteur-substrat	128
I.2.b) Simulations de DM du complexe enzyme-cofacteur-substrat.....	129
I.2.c) Simulations DM sur le complexe enzyme-cofacteur-produit.....	130
I.3) Résultats pour l'enzyme LAR.....	131
I.3.a) Simulations de type RAMD.....	131
I.3.b) Simulation DM sur le complexe LAR-cofacteur-substrat.....	132
II) Prédiction de la structure du complexe protéine-protéine DFR-LAR.....	133
II.1) Protocole de simulation.....	133
II.2) Résultats du docking DFR-LAR.....	133
II.3) Analyse structurale des complexes DFR-LAR retenus	134
III) Simulations de dynamique moléculaire des complexes DFR-LAR.....	137
III.1) Protocoles de simulation.....	137
III.2) Résultats des simulations pour le complexe S1.....	137
III.2.a) Analyse structurale	137
III.2.b) Diffusion du substrat.....	139
III.3) Résultats des simulations pour le complexe S2.....	140
III.3.a) Analyse structurale	140
III.3.b) Libération du produit et diffusion vers la seconde enzyme.....	142
III.4) Rôle de la solvation dans le phénomène de "substrate channeling"	148
IV) Conclusion	149
Chapitre 4 : Interactions F3'H/membrane et premiers modèles de métabolisme	153
I) Simulations préparatoires de dynamique moléculaire de systèmes F3'H/membrane.	155
I.1) Identification des chemins d'accès au site actif de F3'H	155
I.1.a) Canaux déjà identifiés dans les cytochromes P450.....	155
I.1.b) Protocole de simulation.....	157

I.1.c) Résultat pour l'enzyme F3'H.....	158
I.2) Evolution de l'interaction F3'H/membrane lors des simulations de DM.....	159
I.2.a) Protocole de simulation.....	159
I.2.b) Evolution du mode d'interaction F3'H/membrane	159
I.3) Changements structuraux induits par l'interaction avec la membrane	163
II) Premier modèle de métabolon.....	165
II.1) Modèle avec sortie par le canal S	167
II.2) Modèle avec sortie par canal 2e.....	169
II.3) Simulation de dynamique moléculaire.....	171
III) Conclusion	172
Conclusion Générale	175
Chapitre Annexe : Modèle d'inhibition de DFR par excès de substrat	185
I) Les différents modèles d'inhibition possibles	187
II) Etude de l'association DFR-DHQ	189
II.1) Protocole de simulation.....	189
II.2) Résultats	189
III) Calculs de docking sur l'enzyme seule.....	190
III.1) Protocole de simulation	190
III.2) Résultats.....	190
IV) Simulation de dynamique moléculaire sur les différents systèmes	194
IV.1) Protocole de simulation	194
IV.2) Résultats des simulations de DM	194
V) Conclusion.....	197
Bibliographie.....	199
Résumé.....	214
Abstract	214

Liste des figures

- Figure 1** : Schéma d'une coupe de baie de raisin ainsi que la présentation des différentes parties : la pellicule, qui recouvre le fruit et préserve la chair. Elle est riche en anthocyanes permettant la pigmentation de la pellicule; la pulpe, qui contient en concentration élevée du sucre et des flavonoïdes; les pépins, qui peuvent être considérés comme le cœur de la baie. Ils contiennent des tannins condensés (proanthocyanidines) en grande quantité. 23
- Figure 2** : Schéma d'une cellule végétale de *Vitis vinifera*. Sont représentées les différentes parties de la cellule. 4CL : 4-coumarate CoA ligase; PAL : phénylalanine ammonia-lyase; C4H : cinnamate-4-hydroxylase; CHI : chalcone isomérase; CHS : chalcone synthase; ANR : anthocyanidine synthase; LAR : leucoanthocyanidine réductase; . Tiré de la ref [8]..... 24
- Figure 3**: Schéma partiel de la biosynthèse des flavonoïdes. CHS : chalcone synthase, CHI : chalcone isomerase, IFS : isoflavone synthase, F3H : Flavonoïde-3-hydroxylase, FLS : Flavonol synthase, F3'H : Flavonoïde-3'-hydroxylase, DFR : dihydroflavonol-4-reductase, ANS : Anthocyanidine Synthase, LAR : leucoanthocyanidine reductase, ANR : Anthocyanidine reductase, 3-GT : 3-Glucosyle transferase..... 25
- Figure 4** : Nomenclature et numérotation des atomes dans un composé de type flavonoïde.. 27
- Figure 5** : Schéma général de la structure type d'un composé de type flavonol, ainsi que le nom de quelques composés selon le nombre et la position de groupements hydroxyles..... 28
- Figure 6** : Structure commune des composés de type flavanonols, avec quelques exemples de noms de molécules selon l'état d'oxydation du cycle B. 28
- Figure 7**: Structure de base des composés de type leucoanthocyanidines, avec différents exemples de noms de molécules selon le degré d'oxydation du cycle B. 29
- Figure 8**: Structure des composés de type 2R,3S-flavanols (en haut) et des 2R,3R-flavanols (en bas), avec différents exemples de noms de molécules selon l'état d'oxydation du cycle B. 30
- Figure 9** : Schéma des structures de composés de type anthocyanidines (en haut) et anthocyanes (en bas), après action de la 3-O glucosyle transférase (3-GT). 31
- Figure 10**: Structure des trois types de proanthocyanidines. De gauche à droite : type A, type B, type C..... 32
- Figure 11**: Schéma général de la biosynthèse des flavonoïdes donné par le site KEGG (Kyoto encyclopedia of gene and genomes)[34]. 33
- Figure 12**: Premières étapes de la formation du précurseur des flavonoïdes. 34
- Figure 13**: Schéma d'un métabolite impliqué dans la biosynthèse des flavonols (en haut) et des flavonoïdes (en bas). Chaque acronyme désigne une enzyme, et cet ensemble serait en interaction avec la membrane cellulaire. Tiré de la ref n°[48]. 37
- Figure 14** : Représentation d'un métabolite qui serait à la fois impliqué dans la biosynthèse des flavonoïdes et dans celle des flavonols, d'après l'article de Ralston *et al.*[58] 38
- Figure 15** : à gauche, structure tridimensionnelle de l'enzyme DFR obtenu par Petit *et al.*(PDB ID:2C29). A droite, détail des résidus du site actif interagissant avec le substrat et représentation de la cavité de liaison..... 41

Figure 16 : Proposition de mécanisme enzymatique réalisée par la DFR.....	41
Figure 17 : Structure de LAR correspondante au fichier PDB 3I52. A droite, un zoom sur le site actif et les différents résidus impliqués dans l'interaction avec le substrat, ainsi que la représentation de la cavité de liaison.....	43
Figure 18 : Mécanisme proposé par Maugé <i>et al.</i> pour la transformation de leucocyanidine en catéchine. Tiré de la ref [74].	43
Figure 19 : Structure cristallographique de l'enzyme humaine Cytochrome P450 1A2. La partie N-terminale, transmembranaire, n'est pas représentée car non présente dans la structure, mais devrait être reliée à la partie entourée en rouge.	44
Figure 20 : Structure de la molécule de Dhurrin.	49
Figure 21 : Schéma récapitulatif l'ensemble du travail effectué sur cette portion de la biosynthèse.	55
Figure 22 : Schéma du principe de minimisation.....	60
Figure 23 : Représentation du fonctionnement des conditions limites périodiques.....	66
Figure 24 : Schéma de la représentation d'un acide aminé en gros grain.....	72
Figure 25 : Potentiel de Lennard-Jones modifié pour le protocole ATTRACT.....	72
Figure 26 : Schéma récapitulatif du protocole ATTRACT.	73
Figure 27 : Prédiction de la structure secondaire de LAR par le serveur PSI-PRED. Sont encadrées en rouge les zones manquantes dont nous voulons prédire la structure secondaire.....	82
Figure 28 : Grille définie pour le docking du peptide (325-330) sur la structure cristallisée de LAR. En rouge se situe l'extrémité du résidu 317, le dernier cristallisé, et les points dorés correspondent aux différents points de la grille. Enfin, le cercle rouge montre les résidus 42 à 54 qui n'étaient pas présents dans la structure cristallographique de 3I52.....	83
Figure 29 : à gauche, interaction entre l'hélice (résidu 325-330), représentée en noir, et la structure de LAR, après relaxation. En rouge est représenté l'atome de carbone CA du squelette du résidu 317. A droite, structure complète de LAR reconstruite avec cette position de l'hélice 325-330.	84
Figure 30 : Alignement de la séquence de F3'H avec celle des quatre modèles choisis.....	86
Figure 31 : Structure de l'enzyme F3'H obtenue par reconstruction par homologie. En <i>balls and sticks</i> sont représentés le groupement hème, cofacteur de la réaction, et la cystéine qui interagit directement avec lui.	86
Figure 32 : Pose du dihydrokaempferol retenue pour les simulations de dynamique moléculaire de F3'H.	89
Figure 33 : à gauche, représentation de l'interaction F3'H/membrane prédite par le serveur OPM. A droite, détail de l'interaction de surface et des chaînes secondaires des résidus impliqués.	90
Figure 34 : Représentation de la partie QM. En <i>licorice</i> sont représentés les différents atomes considérés dans la partie quantique. Les atomes oranges désignent les quatre "link atom" pour les résidus S128, Y163, K167 et le cofacteur NADPH.....	98
Figure 35 : Mécanisme enzymatique de la réduction du substrat par l'enzyme DFR. L'hydrure est fourni par le cofacteur NADPH. Les résidus Y163 et K167 participent à la	

cascade de transferts de protons. Pour rappel, DHQ : Dihydroquercétine, DHK : Dihydrokaempferol. 99

Figure 36 : Schéma des distances impliquées dans la définition des coordonnées de réaction. Le transfert d'hydrure est modélisé par la contrainte CR1 (égale à d1-d2), tandis que le transfert de proton venant de la tyrosine 163 est représentée par le biais CR 2 (égale à d3 - d4)..... 100

Figure 37 : Représentations des différentes interactions du substrat dans le site actif. L'interaction entre l'arginine et la dihydroquercétine, qui n'existe pas dans le cas du dihydrokaempferol, est entourée en rouge. 102

Figure 38 : Potentiel de force moyenne obtenu pour le transfert d'hydrure. 103

Figure 39 : Surface d'énergie potentielle obtenue pour le système DFR-NADPH-DHQ. L'axe x représente la contrainte CR1 et l'axe y CR2. A droite, sur la projection sur le plan horizontal, les cercles blancs désignent les réactifs (en bas à droite), l'état de transition (au centre) et les produits de la réaction (en haut à droite). Le cercle rouge représente un état intermédiaire lié au rapprochement NADPH-DHQ. 106

Figure 40 : Chemin réactionnel suivi par le mécanisme enzymatique. Le point rouge correspond à la position de l'état de transition. 107

Figure 41 : Evolution des distances impliquées dans la définition des contraintes de calculs DM-QM/MM US 2D. A gauche sont représentées les distances composant la contrainte CR1, à droite celles de CR2. Sont ajoutées aussi les distances entre les atomes lourds concernés. 109

Figure 42 : Evolution de la distance entre les atomes de carbone du cofacteur et du substrat, le long de la surface d'énergie potentielle. A droite est représenté une vue du dessus de la surface présentée. 110

Figure 43 : Evolution de la distance OTyr-ODHQ durant toutes les fenêtres de simulation. 111

Figure 44 : Evolution des distances représentant les transferts de protons se déroulant lors de la réaction de DHQ dans DFR..... 111

Figure 45 : Profil réactionnel obtenu pour le complexe ternaire enzyme/NADPH/DHK. La courbe noire désigne le chemin de plus basse énergie suivie par le mécanisme réactionnel. Le point rouge désigne l'état de transition. 112

Figure 46 : Evolution de la distance C-C lors des différentes fenêtres calculées pour le cas du DHK. A gauche est représentée la vue de dessus de cette surface..... 114

Figure 47 : Evolution des distances impliquant la cascade de transferts de protons..... 114

Figure 48 : Evolution du RMSd des substrats par rapport à la pose cristallographique. 115

Figure 49 : Représentation de l'évolution des distances composant la contrainte du transfert d'hydrure durant la réaction. Le trait noir marque la fin de la première montée en énergie observée..... 117

Figure 50 : Evolution de l'angle de pyramidalisation du carbonyle le long du chemin réactionnel 117

Figure 51 : Distribution des valeurs échantillonnées durant la simulation DM-QM/MM pour l'angle de pyramidalisation (à gauche) et la distance hydrure-DHQ (à droite). 119

Figure 52 : projection des points échantillonnés en simulation de DM-QM/MM sur le profil réactionnel de DHQ. Ne sont représentés que les points représentatifs de tous les états observés. L'étoile rouge représente le point décrivant l'état initial de la réaction. 120

Figure 53 : Distribution des valeurs échantillonnées lors de la simulation de type DM-QM/MM de la distance hydrure-DHK.	121
Figure 54 : Points échantillonnés dans les simulations DM-QM/MM avec le DHK, et projetés sur le profil	122
Figure 55 : Détail des 3 voies d'accès détectées par les simulations de type RAMD. La sortie 1 est représentée en rouge, la sortie 2 en vert et la sortie 3 en bleu. Les résidus impliqués dans chacun des chemins sont également représentés de la même couleur.	129
Figure 56 : Position finale échantillonnée lors des simulations de dynamique moléculaire longues avec 5 ligands. La pose obtenue par les calculs est représentée en orange. En bleu clair est représentée la position cristallographique.	130
Figure 57 : Représentation du chemin suivi lors de la libération du produit de l'enzyme DFR. Les atomes du cofacteur sont représentés ainsi que le LCC en orange. Les résidus définissant une porte menant à la cavité (M88 et P204) sont également mis en avant.	131
Figure 58 : Sortie du site actif identifiée par les simulations de type RAMD. En orange est représenté le ligand, et en <i>licorice</i> sont représentés le cofacteur, et les résidus définissant l'entrée de la cavité. Le chemin de diffusion est également montré.....	132
Figure 59 : Structures protéine-protéine obtenues par le protocole ATTRACT. LAR est représentée en blanc tandis que DFR est en noir. A gauche est représenté le cluster 4 et à droite le cluster 7 (appelé respectivement S1 et S2 par la suite).....	134
Figure 60 : Détail d'une interface protéine-protéine transitoire. Le cœur de l'interaction est caractérisé par une compacité élevée des résidus, avec des interactions majoritairement hydrophobes. Les résidus sont notés 1 sur le schéma. La périphérie de l'interface est définie par des interactions plus hydrophiles. Elle est composée des résidus de type 2 (interagissant avec le cœur mais partiellement solvatés) et de type 3 (interaction avec l'autre protéine uniquement grâce aux molécules d'eau pontantes, identifiée par la lettre w).	135
Figure 61 : Schéma représentant la topologie de l'interface protéine-protéine pour les deux clusters sélectionnés. A gauche, la topologie du cluster S2. A droite le cluster S1, avec une zone partiellement solvatée entre deux zones d'interactions.	136
Figure 62 : Evolution de la structure du complexe DFR (en gris)-LAR (en noir). A gauche, structure sortie du protocole ATTRACT; à droite, après simulation de dynamique moléculaire, dans laquelle on observe une dissociation partielle.....	139
Figure 63 : Chemin de diffusion caractérisé par simulation de dynamique moléculaire dans le cas du complexe S1, observé durant la simulation MD 2. En rouge est représentée la position initiale, et en bleu l'état final obtenu pendant le calcul. DFR est représentée en gris et LAR en noir.	140
Figure 64 : Réarrangement du complexe S2 observé en simulation de dynamique moléculaire. DFR est en gris, LAR en noir. La flèche rouge montre la formation du pont salin lors du réarrangement, et la flèche bleue l'autre zone d'interaction impliquée.	142
Figure 65 : Chemin de diffusion suivi par le substrat observé lors de la simulation 2 du complexe S2 et évolution du RMSd durant le transfert. La position initiale du substrat est représentée en rouge, tandis que la position finale en bleu.	143
Figure 66 : A gauche, évolution de l'énergie d'interaction LCC/complexe le long du chemin de diffusion. A droite, projection de cette énergie en fonction de la distance entre les centres de masse de DFR et du ligand, et celle entre les centre de masse de LAR et de LCC. Les trois couleurs représentent les trois étapes précédemment citées.....	144

Figure 67 : Position intermédiaire rencontrée lors du chemin de diffusion du LCC. Ne sont affichées ici que les interactions de type liaison hydrogène à la surface de LAR.	146
Figure 68 : Interaction jouant le rôle de "porte" de la cavité de LAR, impliquant le cofacteur, la lysine catalytique K140 et la méthionine M136.....	147
Figure 69 : Evolution du nombre de molécules d'eau autour du substrat lors de sa diffusion du site actif de DFR vers celui de LAR (sphère de solvation prise entre 1 et 3,4 Å autour du solvant). En rouge est représentée l'évolution, et en noire les barres d'erreurs associées. La ligne rouge montre le nombre moyen de molécules d'eau pour un substrat complètement solvaté.....	148
Figure 70 : Nomenclature des différentes structures secondaires des cytochromes P450. Les cercles désignent les différents accès au site actifs identifiés. A gauche, vue de côté; à droite, vue de dessus. Les cercles désignent la totalité des canaux identifiés comme étant des voies d'accès au site actif. Tiré de la ref [173].....	156
Figure 71 : Identification des hélices de la structure reconstruite de F3'H. La boucle B-C et la boucle F-G sont respectivement représentées en rouge et vert.	157
Figure 72 : Représentation de l'évolution de la position de l'hélice transmembranaire au sein de la membrane. A gauche, la position initiale de l'hélice au sein de la membrane est représentée en bleu et la position finale en violet. A droite, détail de l'interaction entre le résidu en N-terminal et les groupements phosphates des lipides	160
Figure 73 : Représentation des interactions cation- π observées durant les simulations de dynamique moléculaire.	162
Figure 74 : Evolution des distances entre le résidu F69 et les trois lipides impliqués dans les interactions cation- π . Le trait rouge définit la limite au delà de laquelle on ne considère plus d'interaction.	163
Figure 75 : Ouverture de la porte du canal S après 300 ns, par rotation de la valine 205. Le DQH et le groupement Hème sont représentés en <i>balls and sticks</i>	164
Figure 76 : Ouverture du canal 2e après 200 ns de simulation. Le phénomène implique la rotation des résidus Y108 et F113. La molécule de DHQ et le cofacteur Hème sont représentés en <i>balls and sticks</i>	164
Figure 77 : Grille définie autour de F3'H pour les calculs de docking protéine-protéine du complexe DFR-LAR sur F3'H. La délimitation a été réalisée de sorte que la position de la membrane soit prise en compte.	166
Figure 78 : Mesure de la distance entre la sortie 1 de DFR et la sortie du canal 2e de F3'H pour chaque cluster obtenu. La ligne rouge délimite le critère de sélection et d'analyse des modèles.....	167
Figure 79 : Modèle du complexe F3'H-DFR-LAR obtenu par les calculs de docking protéine/protéine, dans le cas d'une sortie par le canal S. En violet est représentée la sortie du canal S de l'enzyme F3'H. En orange est représentée F3'H, en gris LAR et en marron-vert, DFR.	168
Figure 80 : Valeurs des distances entre la sortie du canal 2e de F3'H et la voie d'accès 1 de DFR pour tous les clusters. La ligne rouge désigne la limite de sélection des modèles, à une distance de 30 Å.	169
Figure 81 : Modèle du complexe F3'H-DFR-LAR, avec une sortie des composés par le canal 2e. En vert sont représentés les résidus définissant la porte du canal 2e.	170

Figure 82 : Représentation du modèle de métabolon avec sortie par le canal S, ainsi que l'interaction avec la membrane. F3'H est représentée en orange, DFR en marron-vert et LAR en gris.	171
Figure 83 : Equation bilan de la réaction enzymatique se déroulant au sein de l'enzyme DFR. En italique est précisée la nomenclature officielle des flavonoïdes.	187
Figure 84 : Identification de la "porte" de la voie d'accès 1. A gauche, au début de la simulation, on observe que la porte est ouverte. A droite, nous voyons au cours du temps la création de cette interaction hydrophobe provoquant la fermeture de la cavité.....	189
Figure 85 : Coupe transversale de l'enzyme montrant ainsi les 3 clusters obtenus par calcul de docking. Les sites A, B et C représentent respectivement le site du cofacteur, la position cristallographique du DHQ dans le complexe ternaire et la pose associée à l'entrée de la cavité par la voie d'accès 2. Pour le site B, deux positions sont représentées.	191
Figure 86 : Détails des interactions dans le cas des trois clusters identifiés durant les calculs de docking. a) et b) montrent les deux meilleures poses du cluster de la pose cristallographique, de la même manière que la Figure 85, c) représente la meilleure pose du cluster associé au site de liaison du cofacteur et d) présente la pose associée à la sortie 2. ..	192
Figure 87 : Evolution du potentiel électrostatique le long de la cavité de DFR. A gauche est représenté le sens de la coordonnée de réaction. A droite est montrée l'évolution du potentiel électrostatique en fonction de la position du substrat dans la cavité. L'origine a été choisie à la position cristallographique de DHQ dans le complexe ternaire enzyme-cofacteur-substrat, dans le site B.	193

Liste des Abréviations

3-GT :	3-Glucosyle transferase
4CL :	4-coumarate CoA ligase
Å :	Angströms (10^{-10} m)
aMD :	"Accelerated Molecular Dynamics"
ANR :	Anthocyanidine reductase
ANS :	Anthocyanidine Synthase
C4H :	Cinnamate-4-hydroxylase
CHS :	Chalcone synthase
CHI :	Chalcone isomerase
CYP :	Cytochrome P450
DFR :	Dihydroflavonol-4-reductase
DKH :	Dihydrokamepferol
DM :	Dynamique moléculaire
DMH :	Dihydromyricétine
DQH :	Dihydroquercétine
F3H :	Flavonoïde-3-hydroxylase
F3'H :	Flavonoïde-3'-hydroxylase
ff03.r1 :	Force Field 03, révision 1
FLS :	Flavonol synthase
GAFF :	General Amber Force Field
IFS :	Isoflavone synthase
LAR :	Leucoanthocyanidine reductase
LCC :	Leucocyanidine
MM :	Mécanique Moléculaire
MM-GBSA :	"Molecular Mechanics-Generalized Born Surface Area"
NAC :	"Near-Attack Conformer"
NADPH :	Nicotinamide adénine dinucléotide phosphate

ns :	Nanoseconde
OPM :	Orientation of Protein in Membranes
PDB :	Protein Data Bank
PRP :	Protéines riches en proline
PAL :	Phénylalanine ammonia-lyase
POPC :	Palmitoyl-2-oleoylphosphatidylcholine
QM :	Mécanique quantique (Quantum mechanics)
QM/MM :	Mécanique quantique/mécanique moléculaire
RAMD :	"Random Acceleration Molecular Dynamics"
REMD :	"Replica Exchange Molecular Dynamics"
RCPG :	Récepteur couplé à la protéine G
RMSd :	Déviatiion de l'écart quadratique moyen (Root Mean Square déviation)
SDR :	Short-chain dehydrogénases/réductases
US :	Umbrella Sampling
μs :	Microseconde
WHAM :	"Weight Average Histogram Analysis"

Introduction

1) Position du sujet

Les plantes constituent une ressource très riche en molécules d'intérêt biologique. De nombreuses molécules bioactives découvertes sont d'origine naturelle, ou inspirées de celle-ci. Au delà de la caractérisation de nouveaux composés, la communauté scientifique est maintenant en mesure de modifier génétiquement des organismes vivants tels que les plantes, dans le but d'obtenir un nouveau mode de production de métabolites. Dans cette optique, il est nécessaire de mieux comprendre les voies de biosynthèse de composés naturels.

Au sein de la matrice végétale, les synthèses sont assurées par des enzymes. Elles vont permettre de catalyser en cascade, et de manière plus ou moins sélective, les transformations de précurseurs en différents métabolites, dont certains pourraient montrer une activité biologique d'intérêt pour l'Homme.

Au delà des métabolismes qui impliquent une production ou une utilisation de molécules nécessaires au développement de l'organisme, (nous pouvons penser aux métabolismes impliquant les sucres ou les lipides par exemple) de nombreux autres, impliqués dans la défense ou la reproduction, mènent à la biosynthèse de familles de molécules appelées métabolites secondaires. A titre d'exemple, nous pouvons mettre en avant les métabolismes menant aux vitamines, aux terpènes ou aux flavonoïdes. Toutes les voies de biosynthèse mènent à des molécules de structures chimiques extrêmement diverses dont les propriétés et les caractéristiques offrent un large éventail d'utilisations potentielles.

Les raisons pour lesquelles elles sont produites au sein de la plante peuvent alors être mises à profit pour l'Homme sans modifier leur propriétés intrinsèques. Au sein de la famille des flavonoïdes, quelques milliers de composés ont été caractérisés. Il est maintenant reconnu que nombre d'entre eux présentent des activités antioxydantes et antiradicalaires bénéfiques pour l'Homme. Pour revenir à leurs voies de production, aboutir à une richesse de composés nécessite des voies de biosynthèse impliquant de nombreuses enzymes. Ce schéma général de biosynthèse commence à être bien connu, même si des zones d'ombre subsistent encore : certaines enzymes n'ont pas encore été cristallisées ou leur mécanisme réactionnel reste sujet à controverse. L'enchaînement des réactions enzymatiques est toujours discuté car plusieurs séquences de transformations chimiques peuvent être envisageables; de même la spécificité

des enzymes pour un substrat reste pour certains cas sujet à débat. Pour entre autre répondre à ces inconnues, de nombreux groupes s'intéressent à la possibilité que des enzymes d'une même voie de biosynthèse s'associent temporairement, sous la forme d'une structure supramoléculaire complexe qui permettrait d'atteindre plus efficacement les produits finaux de la cascade biochimique. Cet assemblage d'enzymes présenterait de nombreux avantages à la conversion de composés. Le principal atout serait le phénomène de "substrate channeling", qui optimise le processus de biosynthèse en évitant la diffusion de composés dans le milieu cytosolique.

Par ailleurs, l'étude de ces entités biologiques constitue un autre des enjeux majeurs pour la recherche actuelle : le développement de nouvelles méthodes de synthèse plus efficaces et plus respectueuses de l'environnement, qui s'inscrit dans le cadre de développement d'une chimie "verte", c'est-à-dire moins polluante. L'idée est de s'appuyer sur l'ingénierie végétale pour la fabrication de molécules naturelles à l'échelle du laboratoire pour, à terme, transférer ces connaissances dans le domaine industriel. Le bio-mimétisme est ainsi devenu un secteur de recherche incontournable [1] dans la thématique de la chimie "verte". Par ailleurs, les récentes investigations sur l'utilisation d'enzymes dans les synthèses [2, 3] attestent de l'impact de ce pan de la recherche. De plus, la nature est source d'inspiration, pas uniquement pour la synthèse mais également pour le développement de nouveaux matériaux [3, 4].

Pour décrypter la complexité de la biosynthèse des composés naturels, les méthodes analytiques, telles que les méthodes de spectrométrie d'absorption, peuvent nous aider à identifier des composés intermédiaires et déduire ainsi les réactions mises en jeu. De par les propriétés chromophores des flavonoïdes et des anthocyanes (qui participent à la pigmentation des fruits et des fleurs), la méthode d'identification de choix est la spectrophotométrie UV-Visible. D'autres méthodes, telles que la cristallographie par rayons X, nous permettent de résoudre la structure tridimensionnelle des enzymes, sous réserve de trouver les bonnes conditions de cristallisation. Cette dernière demeure une méthode de choix. En cas de co-cristallisation avec un composé, les interactions de ce dernier avec l'enzyme sont ainsi accessibles, permettant d'avoir des données pour l'étude mécanistique dans le cas d'un substrat, ou de l'inhibition dans le cas d'un inhibiteur connu.

La chimie théorique fait partie des méthodes pouvant apporter un nouvel éclairage sur la compréhension de cette ingénierie végétale. La diversité des méthodes existantes permet

une étude multi-échelle des différents aspects de la recherche sur ces composés et leur biosynthèse par la plante. Les méthodes hybrides QM/MM, mélangeant les deux théories et tirant les avantages de chacune d'entre elles, permettent d'améliorer notre compréhension des mécanismes enzymatiques. Ces dernières, récompensées en 2013 par le prix Nobel de chimie décerné à Martin Karplus, Arieh Warshel et Michael Levitt, constituent l'état de l'art de l'étude des catalyses enzymatiques. La simulation de dynamique moléculaire est un outil de choix pour l'étude du comportement de grands systèmes biologiques (protéines, enzymes, ADN, complexes protéine-protéine...) sur des échelles de temps longues, allant de la nano à la micro, voire la milliseconde, ce qui constitue à ce jour l'état de l'art dans l'étude du comportement dynamique des systèmes moléculaires ainsi que l'accès à des grandeurs macroscopiques d'intérêt. D'autre part, la mécanique quantique permet l'étude des propriétés structurales et réactives des molécules composant la famille des flavonoïdes.

Le travail présenté ici apporte un modèle de complexe supraenzymatique, impliqué dans la biosynthèse des flavonoïdes par l'utilisation de méthodes théoriques. Nous avons choisi l'organisme *Vitis vinifera* (vigne) pour notre étude sur la biosynthèse des flavonoïdes.

II) Les flavonoïdes dans Vitis vinifera

II.1) Organisme d'étude : Vitis vinifera.

Vitis vinifera est un organisme végétal que l'on appelle communément la vigne. C'est une plante cultivée depuis longtemps en Europe. Elle produit durant la saison d'été (Août à Septembre) des grappes de baies de raisin. L'organisme *Vitis* désigne de manière générale la vigne, tandis que *Vitis vinifera* est une sous-espèce très étendue en Europe et au Proche-Orient. C'est à partir des baies de *Vitis vinifera* que sont conçus les grands cépages et les vins associés.

Baggiolini propose en 1952 [5] une échelle décomposant le cycle phénologique de *Vitis vinifera* en 8 étapes.

- i) Mi-Mars : Débourrement. Apparition des premiers bourgeons et début de croissance des sarments.
- ii) Mai : Floraison et fécondation des plants. Etape importante qui conditionne la future récolte.

- iii) Juin : Nouaison, c'est à dire développement des fruits au stade "petit pois"
- iv) Juillet : véraison. Fin de croissance végétative et début de la phase de croissance des baies, jusqu'au changement de couleur à la fin du mois.
- v) Août : Murissement. Augmentation de la concentration en sucre dans les baies.
- vi) Fin septembre : Vendanges. Fin de la maturation et récoltes des grappes de fruits.
- vii) Octobre-Novembre : Défeuillaison. Perte du feuillage et préparation pour l'hiver.
- viii) Novembre-Mars : Repos hivernal.

Durant ce cycle, la plante va progressivement exprimer les différentes protéines qui permettront la biosynthèse des composés nécessaires à sa croissance et à son développement [6]. Par exemple, les anthocyanes ne seront synthétisées que lors du murissement du fruit.

La baie de raisin est constituée de 3 parties principales [7] que sont la pellicule, les pépins et la pulpe (représentée par le schéma de la Figure 1) :

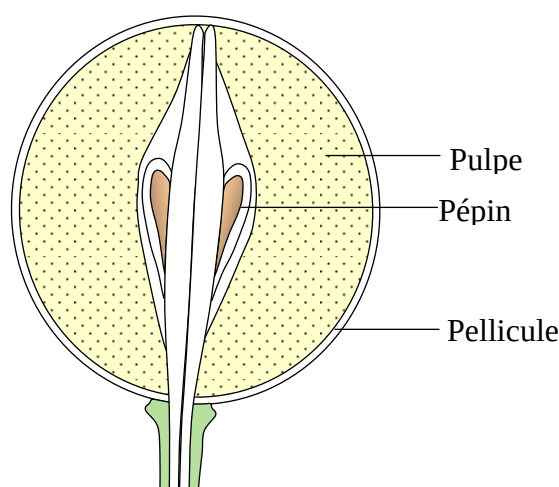


Figure 1 : Schéma d'une coupe de baie de raisin ainsi que la présentation des différentes parties : la pellicule, qui recouvre le fruit et préserve la chair. Elle est riche en anthocyanes permettant la pigmentation de la pellicule; la pulpe, qui contient en concentration élevée du sucre et des flavonoïdes; les pépins, qui peuvent être considérés comme le cœur de la baie. Ils contiennent des tannins condensés (proanthocyanidines) en grande quantité.

Les molécules sont généralement synthétisées au sein de la cellule végétale. En particulier, les enzymes responsables de la biosynthèse sont distribuées dans toute la cellule. Le transport des flavonoïdes se fait par le biais de vésicules membranaires, comme représenté sur la Figure 2. Le travail réalisé par Lacampagne durant sa thèse a permis de montrer que

plusieurs enzymes sont co-localisées à certains endroits de la cellule, pouvant permettre localement la création de complexes multienzymatiques [8].

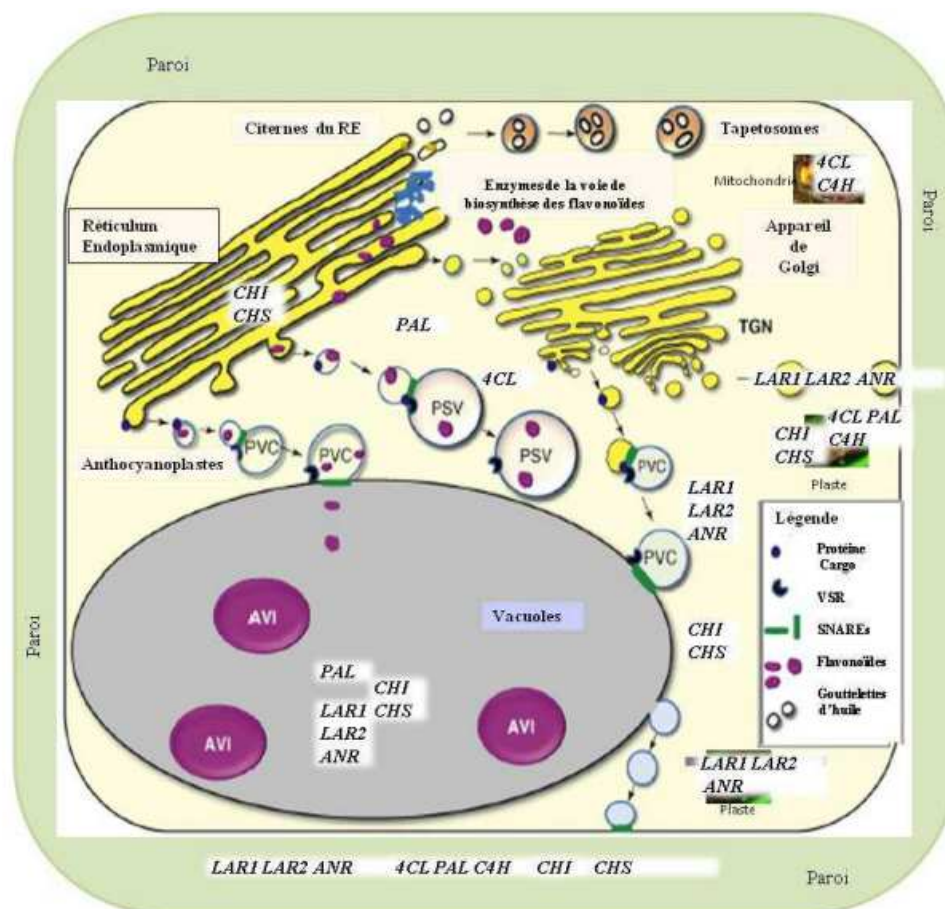


Figure 2 : Schéma d'une cellule végétale de *Vitis vinifera*. Sont représentées les différentes parties de la cellule. 4CL : 4-coumarate CoA ligase; PAL : phénylalanine ammonia-lyase; C4H : cinnamate-4-hydroxylase; CHI : chalcone isomérase; CHS : chalcone synthase; ANR : anthocyanidine synthase; LAR : leucoanthocyanidine réductase; . Tiré de la ref [8]

II.2) Schéma de biosynthèse des flavonoïdes

La Figure 3 résume de manière partielle l'implication des enzymes successives permettant, à partir des premiers précurseurs, de mener aux différentes classes de flavonoïdes au sein de *Vitis vinifera*. Chaque acronyme définit alors une enzyme qui catalyse une réaction particulière. La diversité d'enzymes présentes dans la plante induit un grand nombre de réactions.

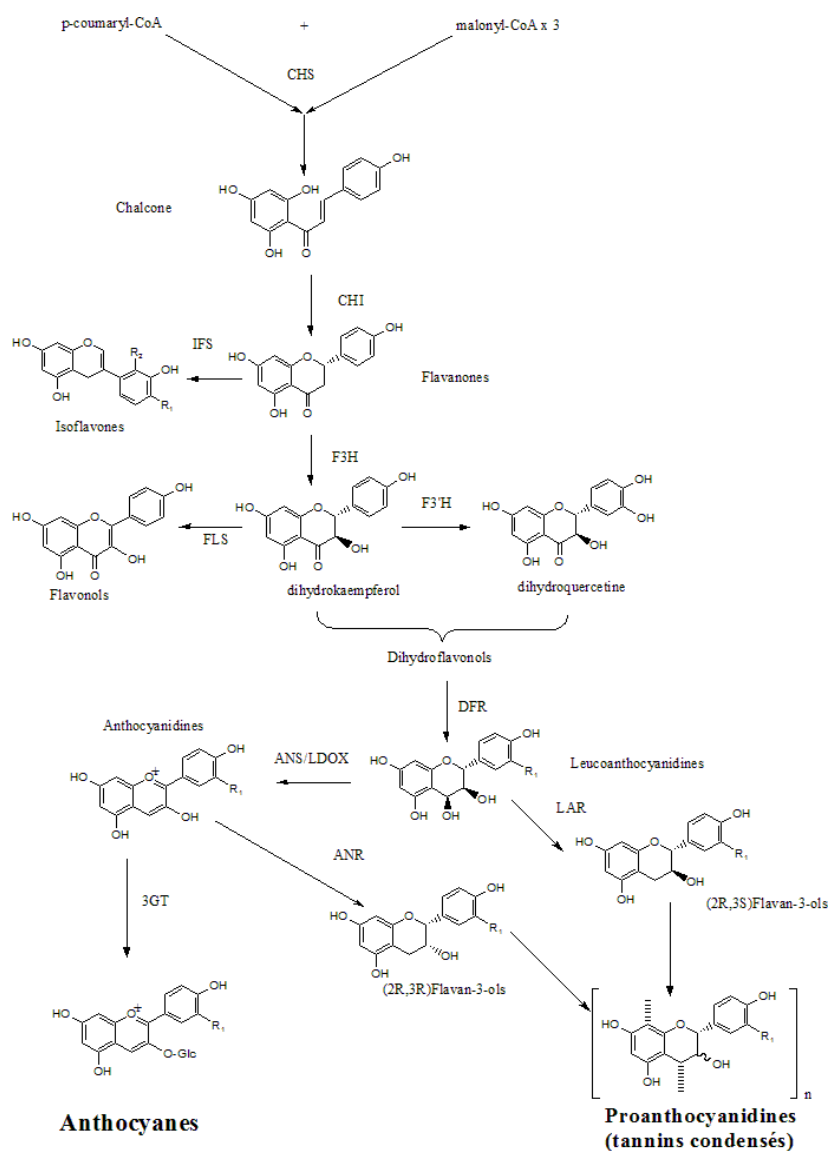


Figure 3: Schéma partiel de la biosynthèse des flavonoïdes. CHS : chalcone synthase, CHI : chalcone isomérase, IFS : isoflavone synthase, F3H : Flavonoïde-3-hydroxylase, FLS : Flavonol synthase, F3'H : Flavonoïde-3'-hydroxylase, DFR : dihydroflavonol-4-réductase, ANS : Anthocyanidine Synthase, LAR : leucoanthocyanidine réductase, ANR : Anthocyanidine réductase, 3-GT : 3-Glucosyle transférase.

II.3) Les flavonoïdes

Les flavonoïdes constituent une des familles de composés naturels les plus étudiées et appartiennent à la grande famille des polyphénols. Les flavonoïdes sont classés en plusieurs grandes sous-familles, parmi lesquelles nous pouvons citer les anthocyanes, les tannins condensés, les flavanols... [9]. Cette famille présente une grande diversité moléculaire [10-12]. Leurs propriétés antioxydantes et antiradicalaires sont les raisons principales pour lesquelles ils sont grandement étudiés.

Ces composés se retrouvent dans la plupart des aliments que l'on consomme, que ce soient les légumes, les fruits, les boissons (café, thé, vin) et les céréales (riz, soja), et ce même après leur transformation [10, 13, 14]. L'ingestion de ces molécules entraîne des effets bénéfiques pour la santé, par piégeage de radicaux libres générés dans notre organisme lors d'agressions extérieures par exemple. De nombreuses études ont démontré ces dernières années [15] que les flavonoïdes peuvent participer activement à la prévention de plusieurs cancers ou d'autres pathologies telles que les maladies cardiovasculaires. Ainsi, la consommation modérée de vin permet de diminuer les risques de crises cardiaques, phénomène que l'on ne retrouve pas dans les autres boissons. Cela démontre ainsi que ce sont les tannins condensés uniquement présents dans le vin, qui induisent ces effets [16]. Ce phénomène est appelé le "french paradox". Ces composés sont également responsables de la sensation d'astringence que l'on ressent lorsque nous dégustons une boisson tannique [10]. Il s'agirait d'une complexation de ces molécules aux PRP (protéines riches en proline), présentes dans la salive, qui induirait une précipitation et une agrégation de ces dernières, conduisant ainsi à cette sensation d'assèchement [17]. Certains sites d'interactions préférentiels de ces tannins avec ces protéines peu structurées ont d'ailleurs été identifiés [18]. Ces résultats ont donc tendance à soutenir cette hypothèse.

Dans la plante, la famille des flavonoïdes se trouve être une catégorie de composés de grande importance car les molécules participent à la protection de la plante contre des agents pathogènes externes, comme les champignons et les bactéries [19-21], ou contre le rayonnement solaire par absorption de rayonnement UV. Ils participent également à la pigmentation des fruits et des fleurs par le biais de l'accumulation d'anthocyanes dans les cellules de la peau des fruits [22]. Enfin, ils permettent de produire un goût amer et très astringent dissuadant les prédateurs, notamment les insectes, de manger le végétal considéré [23]. Les différentes propriétés énoncées attestent de l'intérêt porté par la communauté scientifique mondiale pour l'étude de ces composés, dans différents domaines comme la biologie, la santé ou même l'alimentation.

II.4) Structure des flavonoïdes

Les flavonoïdes possèdent une structure commune composée d'un squelette carboné de type 2-phénylchromane, appelée aussi noyau flavane. Il est composé de deux cycles aromatiques, et d'un cycle central (nommé respectivement A, B et C dans la nomenclature

initialement proposée par Hemingway en 1982 et améliorée par Porter en 1988 [24, 25]). La figure 4 montre la nomenclature associée aux flavonoïdes.

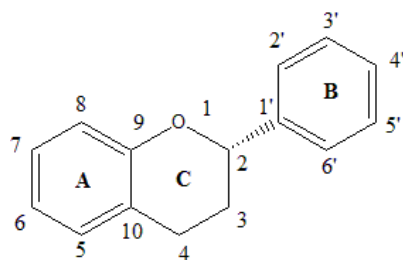


Figure 4 : Nomenclature et numérotation des atomes dans un composé de type flavonoïde.

Le cycle A peut être de type phloroglucinol ou résorcinol. Quand au cycle B, il peut généralement être mono-, di- ou tri-hydroxylés sur les positions 3', 4' et 5'. Les positions 2' et 6' sont possibles mais beaucoup plus rare. Enfin, le cycle C est rencontré généralement sous 3 formes : pyrane, pyrillium ou γ -pyrone. Il existe cependant des cas spéciaux. Les chalcones possèdent un cycle C ouvert, tandis que les isoflavonoïdes possèdent un cycle B en position 3 et non en 2.

L'appartenance d'une molécule à une sous famille dépend de deux points principaux :

- le degré d'oxydation de l'hétérocycle central
- le degré d'hydroxylation des différents cycles

Dans la nature, la plupart des flavonoïdes se retrouvent sous forme glycosylée, c'est-à-dire liés à des sucres. L'hydroxylation de certaines positions est nécessaire pour que la glycosylation puisse se faire. Au niveau de la nomenclature, la forme glycosylée est appelée différemment de la forme phénolique simple. Par exemple, le terme anthocyane désigne toutes les formes glycosylées dérivant de la famille des anthocyanidines, nom uniquement utilisé pour les molécules phénoliques. Le niveau de glycosylation d'une molécule génère un grand nombre de dérivés à partir d'une structure de base. Par exemple, pour la seule molécule de quercétine, pas moins de 80 formes glycosylées différentes existent [22].

Au sein de l'organisme *Vitis vinifera*, les principales familles de flavonoïdes sont les flavonols, les flavan-3,4-diols, les flavanols, les anthocyanes et les proanthocyanidines [26, 27].

II.4.a) Les flavonols et les 2,3-dihydroflavonols

Les flavonols participent à la protection de la plante contre le rayonnement UV [9], et jouent aussi le rôle de pigments dans les baies, en complément des anthocyanes [28]. D'un point de vue structural, ils se caractérisent par un cycle C sous forme γ -pyrone, présentant une hydroxylation en position 3. La diversité de composés dans cette famille provient majoritairement du degré d'hydroxylation du cycle B.

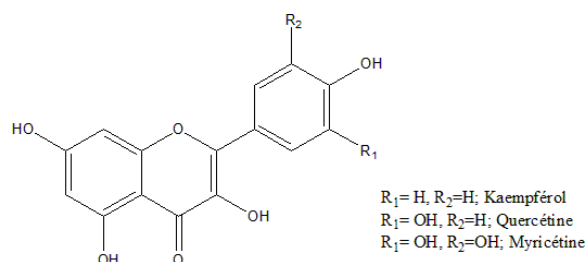


Figure 5 : Schéma général de la structure type d'un composé de type flavonol, ainsi que le nom de quelques composés selon le nombre et la position de groupements hydroxyles.

Les flavonols sont obtenus à partir d'une autre sous famille : les 2,3-dihydroflavonols ou encore appelés flavanonols. Il s'agit d'un flavonol ayant subi une hydrogénation de la fonction alcène en position 2 et 3, induisant ainsi la création de deux centres stéréogènes. La transformation d'un dihydroflavonol en flavonol se fait par l'action de l'enzyme FLS (Flavonol synthase, voir Figure 3). Dans la nature, on retrouve ainsi les quatre stéréoisomères possibles (deux énantiomères et deux diastéréoisomères), mais les composés présentant une configuration (2R, 3R) sont les majoritaires et les plus étudiés. Nous parlons ainsi de composés *trans*, car le cycle B, lié à l'atome de carbone C2, est à l'opposé du groupement hydroxyle lié à l'atome de carbone C3. La structure générale de ce composé est représentée sur la Figure 6.

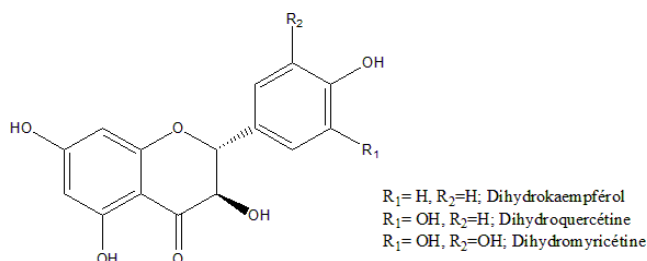


Figure 6 : Structure commune des composés de type flavanonols, avec quelques exemples de noms de molécules selon l'état d'oxydation du cycle B.

II.4.b) Les flavanols et les flavan-3,4-diols

Les flavan-3,4-diols, appelés également leucoanthocyanidines, sont des composés obtenus à partir des dihydroflavonols grâce à l'action de l'enzyme DFR (Dihydroflavonol-4-réductase, voir figure 3). Cette transformation constitue la première étape vers la biosynthèse des proanthocyanidines. L'enzyme catalyse la réduction du groupement cétone présent sur l'atome de carbone C3 en groupement hydroxyle.

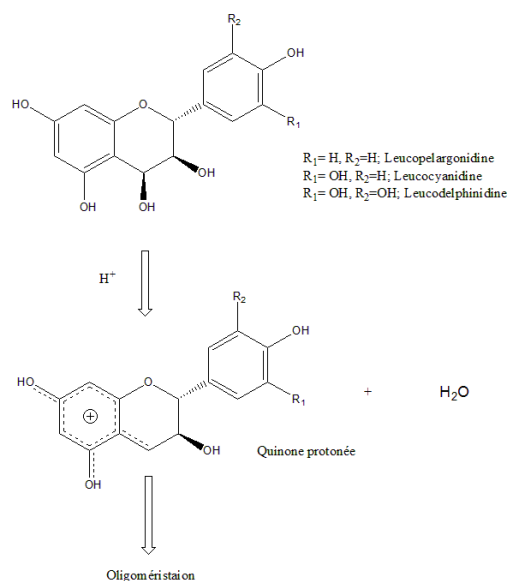


Figure 7: Structure de base des composés de type leucoanthocyanidines, avec différents exemples de noms de molécules selon le degré d'oxydation du cycle B.

Expérimentalement, quelques exemples de leucoanthocyanidines ont pu être caractérisés[29]. Cependant, ceux possédant un groupement hydroxyle en position 5 et 7 n'ont jamais été isolés et extraits en raison de leur instabilité [30]. En effet, en milieu légèrement acide, ces molécules ont grandement tendance à se déshydrater et à s'ioniser en quinones protonées qui s'oligomérisent rapidement. La Figure 7 présente la structure des leucoanthocyanidines ainsi que leur forme déshydratée. Même s'ils sont présents en quantité moindre que les flavanols dans la plante, ces composés sont cruciaux car ils peuvent réagir selon deux chemins biosynthétiques: le premier passant par les anthocyanidines qui mène à la formation des anthocyanes et l'autre qui amène à la synthèse des tannins condensés en passant par les flavanols.

Quant aux flavanols, ils dérivent des flavonols et constituent une des sous-familles menant à la biosynthèse des anthocyanes. D'un point de vue structural, la différence avec les dihydroflavonols vient de l'absence de groupement hydroxyle sur la position 4 de la molécule.

Au sein de l'organisme *Vitis vinifera*, on distingue deux stéréoisomères majoritaires, les 2R,3S et 2R,3R. Ces derniers sont représentés sur la Figure 8.

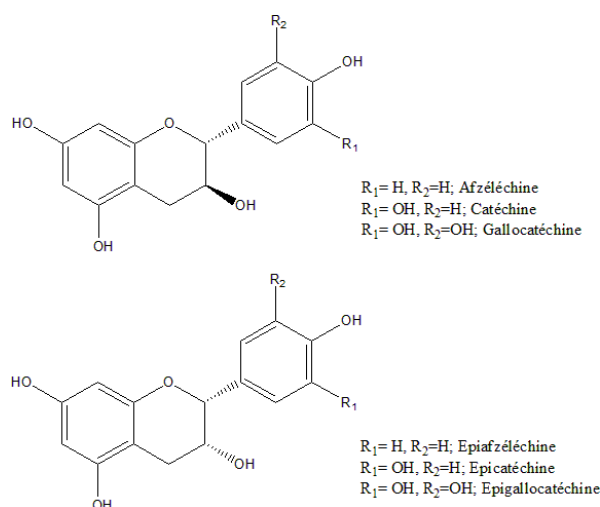


Figure 8: Structure des composés de type 2R,3S-flavanols (en haut) et des 2R,3R-flavanols (en bas), avec différents exemples de noms de molécules selon l'état d'oxydation du cycle B.

II.4.c) Les anthocyanidines et anthocyanes

En suivant la biosynthèse des flavonoïdes, les anthocyanes constituent une des dernières familles connues. Elles sont responsables de la pigmentation des fruits et des fleurs des plantes [9] et sont obtenues à partir des leucoanthocyanidines par action de l'enzyme ANS (anthocyanidine synthase, voir Figure 3). Cette dernière catalyse une réaction lors de laquelle le cycle C, sous forme pyrane, subit une double déshydratation menant à la formation d'un ion pyrilium. Au sein des plantes, les anthocyanidines sont retrouvées principalement sous leur forme glycosylée (anthocyanes), qui se trouve être plus stable que leur forme phénolique. La glycosylation se fait par l'intermédiaire de l'enzyme 3-GT (3-O glucosyle transférase, voir Figure 3). La structure générale des anthocyanidines et des anthocyanes est représentée sur la Figure 9.

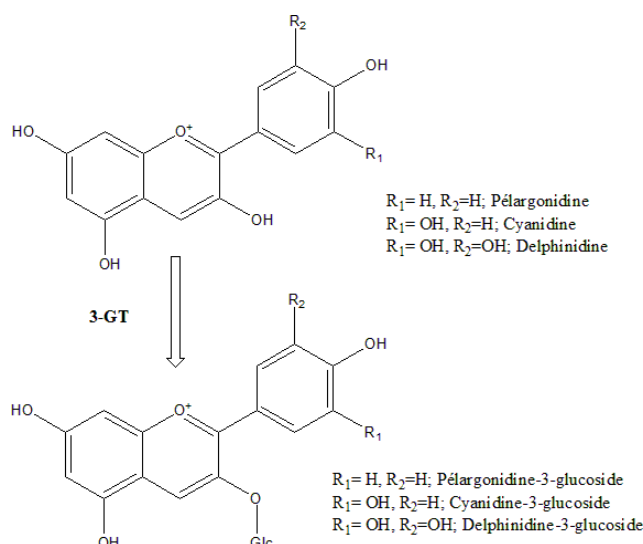


Figure 9 : Schéma des structures de composés de type anthocyanidines (en haut) et anthocyanes (en bas), après action de la 3-O glucosyle transférase (3-GT).

II.4.d) Les proanthocyanidines

En parallèle à la biosynthèse des anthocyanes, les leucoanthocyanidines peuvent également donner les flavanols, qui sont les précurseurs des proanthocyanidines. Ces composés, également appelés tannins condensés, se présentent sous la forme de polymères ayant pour monomère initiaux les (2R,3R) et les (2R,3S) flavanols (principalement la catéchine, l'épicatéchine, la gallocatéchine et l'épigallocatechine, voir Figure 8). La polymérisation induit la formation de molécules complexes, avec un poids moléculaire variant entre 500 et 2000 Da [31, 32]. Les principaux composés de cette famille retrouvés dans l'organisme *Vitis vinifera* sont les proanthocyanidines de type B, qui se caractérisent par des liaisons C-C entre les atomes de carbone C4 et C8 de deux unités monomériques distinctes.

Nous pouvons noter qu'il existe deux autres types de proanthocyanidines que l'on retrouve dans d'autres végétaux [33]. Les proanthocyanidines de type A se caractérisent par une liaison entre les atomes de carbone C4 et C8 mais également par une liaison de type éther entre l'atome de carbone C2 et l'oxygène de l'atome de carbone C7 entre chaque monomère. Enfin, les proanthocyanidines de type C se démarquent par la présence de liaisons interflavanes entre les atomes de carbone C4 et C8, mais aussi entre les atomes C4 et C6, sur trois unités monomériques. Les trois types de proanthocyanidines sont représentés sur la Figure 10.

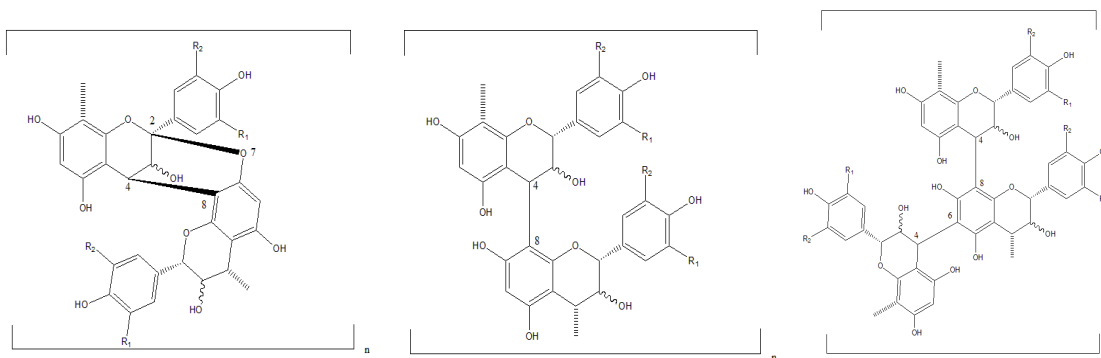


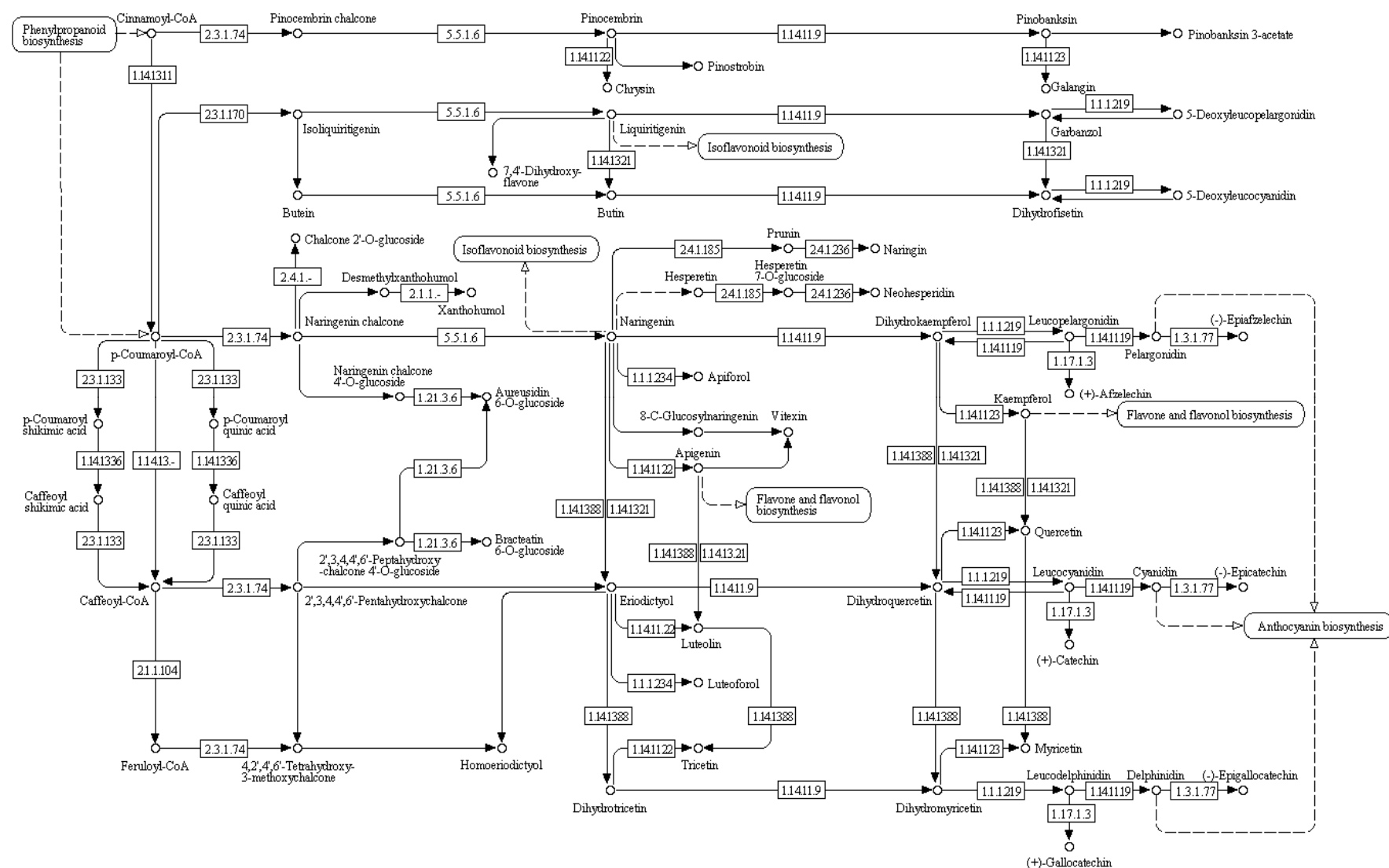
Figure 10: Structure des trois types de proanthocyanidines. De gauche à droite : type A, type B, type C.

II.5) Description du processus de biosynthèse des flavonoïdes

L'intérêt des chercheurs réside dans les moyens mis en œuvre par la plante pour atteindre cette diversité moléculaire. Nous savons que la biosynthèse associée aux flavonoïdes fait intervenir un nombre important d'enzymes, chacune catalysant une réaction particulière. Cependant, certains points restent encore obscurs, notamment la polymérisation des proanthocyanidines.

La Figure 3 présente de manière partielle les différentes étapes de la biosynthèse. Si on considère la totalité de la biosynthèse, avec toutes les différentes ramifications possibles, le schéma se densifie rapidement, comme le montre la Figure 11.

FLAVONOID BIOSYNTHESIS



00941 5/13/13
(c) Kanehisa Laboratories

Figure 11: Schéma général de la biosynthèse des flavonoïdes donné par le site KEGG (Kyoto encyclopedia of gene and genomes)[34].

La Figure 3 montre que les précurseurs de cette biosynthèse sont la p-coumaryl-CoA, obtenue par la voie shikimate [35], et le malonyl-CoA, obtenu par le biais de la voie des polycétides [36]. Le processus de transformation menant de la phénylalanine, précurseur de la voie shikimate, au p-coumaryl-CoA est présenté sur la Figure 12.

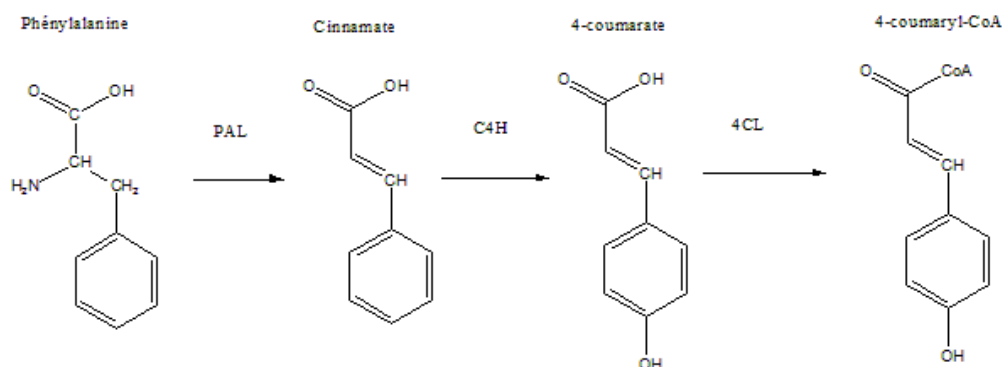


Figure 12: Premières étapes de la formation du précurseur des flavonoïdes.

La phénylalanine subit une déamination par la phénylalanine ammonia-lyase (PAL) pour former le cinnamate. Ce dernier subit par la cinnamate 4-hydroxylase (C4H) une hydroxylation en position 4 du cycle aromatique pour donner le 4-coumarate. Enfin, le 4-coumaryl-CoA est obtenu par action de la 4-coumarate CoA ligase (4CL).

Les deux précurseurs 4-coumaryl-CoA et le malonyl-CoA vont réagir ensemble grâce à l'action l'enzyme CHS (chalcone synthase, EC:2.3.1.74). Cette dernière catalyse la condensation de trois molécules de malonyl-CoA avec la 4-coumaryl-CoA, permettant la synthèse des chalcones. Etant l'entité qui catalyse la première transformation de la biosynthèse des flavonoïdes, la CHS a fait l'objet de nombreuses études qui ont permis de démontrer que seule la forme dimérique est active [34, 37, 38]. La réaction, catalysée par la CHS, permet la formation des cycles A et B du squelette des flavonoïdes. L'enzyme suivante, la CHI (chalcone isomérase, EC:5.5.1.6), catalyse la fermeture du cycle C et la finalisation du squelette carboné des flavonoïdes. L'isomérisation (basée sur un mécanisme de réarrangement de type Claisen) permet de créer, à partir des chalcones, les flavanones avec une stéréospécificité élevée (99%) pour les molécules de configuration 2S [39, 40]. A ce point, le chemin biosynthétique se sépare en deux embranchements.

D'une part, les flavanones peuvent réagir avec l'enzyme IFS (2-hydroxyisoflavone synthase) pour être transformées en isoflavones. Cette enzyme catalyse une isomérisation qui permet au cycle B de se retrouver lié à l'atome de carbone C3 au lieu de l'atome de carbone

C2. Par ailleurs, les isoflavones, produits de l'enzyme IFS, suivent un chemin biosynthétique propre menant à l'ensemble des composés de type isoflavonoïde.

D'autre part, dans la voie de biosynthèse des flavonoïdes, les flavanones sont les précurseurs des dihydroflavonols par l'action de l'enzyme F3H (flavonoid-3-hydroxylase, EC:1.14.11.9). Cette dernière permet l'hydroxylation de la position 3 du squelette 2-phenylchromane. Cette réaction est stéréospécifique et permet d'obtenir uniquement des composés de configuration absolue (2S, 3R).

La F3'H (flavonoïde-3'-hydroxylase, EC:1.14.13.21) induit le changement de degré d'oxydation du cycle B de la molécule considérée en oxydant la position 3'. Les dihydroflavonols sont ses principaux substrats (transformation du dihydrokaempferol en dihydroquercétine, comme montré sur la Figure 3), mais elle peut agir également sur les flavanones dans certains organismes [41]. Nous pouvons noter que l'obtention des composés trihydroxylés sur le cycle B se fait par l'intermédiaire d'une autre enzyme appelée F3'5'H (flavonoïde-3',5'-hydroxylase, EC:1.14.13.88). Elle catalyse de manière simultanée l'hydroxylation des positions 3' et 5'.

Les dihydroflavonols sont les précurseurs d'une famille apparentée aux flavonoïdes : les flavonols. La réaction, catalysée par l'enzyme FLS (flavonol synthase, EC:1.14.11.22), consiste en la création d'une double liaison entre les atomes de carbone C2 et C3. Cette transformation constitue la première étape vers la biosynthèse des flavonols. Parallèlement à cela, les dihydroflavonols peuvent subir une réduction du groupement cétone lié à l'atome de carbone C3 par l'action de l'enzyme la DFR (dihydroflavonol-4-reductase, EC:1.1.1.234) [42]. Nous pouvons remarquer que la réaction impliquant DFR est la dernière étape commune entre la biosynthèse des anthocyanidines et des tanins condensés.

Les leucoanthocyanidines produites par l'enzyme DFR peuvent :

- soit subir l'action de LAR (leucoanthocyanidine reductase, EC:1.17.1.3), qui conduit à la formation des flavan-3-ol, précurseurs des proanthocyanidines.
- soit être les substrats de l'enzyme ANS (anthocyanidine synthase, EC:1.14.11.19), menant ainsi à l'obtention des anthocyanidines.

La réaction impliquant LAR catalyse la déshydratation du composé, par perte du groupement hydroxyle porté par l'atome de carbone C3. D'autre part, l'ANS induit une

déshydratation et une oxydation du cycle C permettant l'obtention des anthocyanidines qui, par l'action de la 3-GT (3-O glucosyl transferase), se transforment en anthocyanes. D'après la figure 3, nous pouvons remarquer que les anthocyanidines peuvent aussi réagir avec l'ANR (anthocyanidine reductase, EC:1.3.1.77) pour créer des (2S,3S) flavan-3-ols, dans lesquelles la configuration du carbone C3 a été modifiée par le passage à l'état d'anthocyanidine.

Concernant les proanthocyanidines, très peu de connaissances sont disponibles sur leur mécanisme de formation. La transformation peut passer par une voie enzymatique ou se faire dans le milieu intracellulaire. De nombreux points sont encore à éclaircir sur cet aspect de la biosynthèse [43]. Les seules données expérimentales que nous possédons à ce jour sont des études génétiques permettant de réguler le taux de tanins contenus dans les fruits [44-46]. Cela donne quelques indications concernant la biosynthèse de ces composés.

D'un point de vue fonctionnel, les enzymes impliquées dans la biosynthèse des flavonoïdes peuvent être séparées en trois familles distinctes [36] :

- les réductases NADPH dépendantes, appelées également short-chain dehydrogenase/reductase (SDR) dont font partie DFR, ANR et LAR.
- les dioxygénases 2-oxoglutarate dépendantes dont font partie ANS, FLS et F3H.
- les cytochromes P450 dont font partie F3'H, IFS, C4H et F3'5'H.

CHI et CHS sont les deux seules exceptions qui ne peuvent rentrer dans ces trois catégories.

II.6) Complexes multienzymatiques

La richesse de la biosynthèse des flavonoïdes réside dans le grand nombre d'enzymes impliquées. Dans plusieurs cas, certaines enzymes partagent les mêmes substrats et donc sont en compétition pour leur réaction. Par ailleurs, sachant que les molécules de type leucoanthocyanidine sont instables, comment expliquer que nous observions malgré tout les produits de réaction les utilisant comme substrats? C'est pour répondre à ces questions que, dès 1969, Stafford émet l'hypothèse qu'il se formerait un complexe multienzymatique lors de la réalisation de cette biosynthèse [47]. Cette superstructure, appelée métabolon, est définie comme l'assemblage de plusieurs enzymes qui interagissent entre elles [48], par l'intermédiaire d'interactions faibles, et forment ainsi une structure fonctionnelle transitoire qui serait ancré dans la membrane du réticulum endoplasmique. La Figure 13 montre de manière schématique l'aspect d'un métabolon.

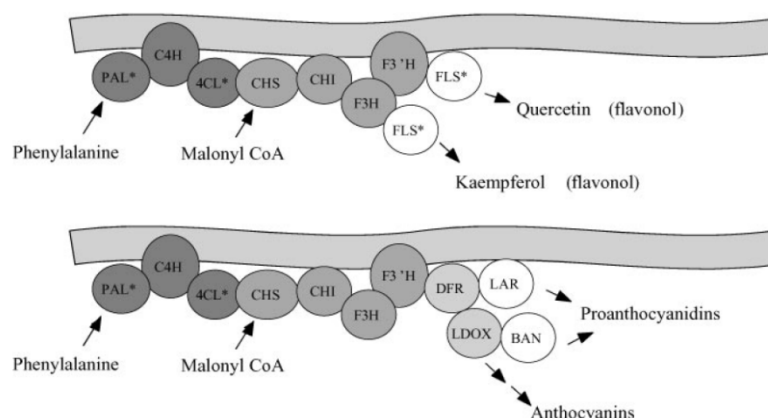


Figure 13: Schéma d'un métabolon impliqué dans la biosynthèse des flavonols (en haut) et des flavonoïdes (en bas). Chaque acronyme désigne une enzyme, et cet ensemble serait en interaction avec la membrane cellulaire. Tiré de la ref n°[48].

L'hypothèse émise par Stafford en 1969 est appuyée par ses observations qui montrent que les intermédiaires associés à des réactions enzymatiques ne sont pas observés (ou en très faibles quantités) dans les cellules de la plante. Cela implique donc que ces derniers ne seraient pas libérés dans le milieu intracellulaire car retenus dans le complexe supramoléculaire. Ainsi, la structure présenterait une connexion des sites actifs des différentes enzymes impliquées dans sa formation, et les intermédiaires transiteraient d'un site actif à l'autre sans diffuser dans l'espace cytosolique. Ce phénomène est connu sous le nom de "substrate channeling" et est l'un des principaux avantages offerts par la formation d'un métabolon. Par ailleurs, une récente étude de Crosby *et al.* [49] met en évidence une enveloppe de ce complexe dans la biosynthèse des flavonoïdes au sein de l'organisme *Arabidopsis thaliana*.

Plusieurs études sur la cinétique associée au phénomène de "substrate channeling" ont été proposées ces dernières années [50, 51]. Le phénomène de co-localisation de ces différentes enzymes, et donc des molécules pouvant être substrats et/ou produits, permet une optimisation du processus de biosynthèse par différents points : i) pas de perte de substrat dans le cytoplasme par diffusion jusqu'à la prochaine protéine ii) protection des molécules instables et/ou toxiques iii) temps de diffusion grandement diminué, iv) sélection d'un chemin unique de biosynthèse, réduisant ainsi les compétitions entre protéines. Cependant, concernant la catalyse enzymatique, une hypothèse a été proposée : l'agrégation d'enzymes les unes aux autres a un impact sur la réactivité propre de chaque protéine. Les réactions dans les enzymes deviennent moins efficaces et l'énergie d'activation s'en retrouve modifiée [52], à cause des effets allostériques induits par les interactions protéine-protéine. Toutefois, la diffusion de

substrats d'une enzyme à l'autre constitue l'étape limitante sur la biosynthèse totale. Ainsi, et même si les réactions de chaque enzyme sont moins efficaces, le gain de temps sur la diffusion fait que l'association des enzymes en métabolon optimise de manière significative la catalyse globale de la biosynthèse des flavonoïdes.

En dehors de cette biosynthèse particulière, plusieurs études montrent que des structures de ce type se formeraient dans d'autres organismes [51, 53]. Dans certains cas, des équipes ont réussi à obtenir un cristal de ce type de structure [54, 55]. Ainsi Lomakin *et al.* ont cristallisé la "fatty acid synthase" au sein d'une levure, une méga-enzyme responsable de la synthèse des acides gras correspondant à une machinerie moléculaire composée de pas moins de 8 sites actifs! Dans ce cas précis, le complexe n'est plus de nature transitoire car les enzymes permettant la formation des acides gras ont évolué dans certains organismes (par exemple les champignons) en un complexe multienzymatique permanent d'une masse de 2,6 mégadalton. Alors que dans les plantes, ces enzymes restent indépendantes mais peuvent s'assembler de manière transitoire [56].

Néanmoins, l'organisation spatiale des enzymes en interaction reste sujet à débat. Deux théories existent : la première défendue par l'équipe de Jørgensen *et al.* [57] consiste en un rassemblement linéaire d'enzymes, impliquant qu'une seule voie de biosynthèse serait alors possible dans ce complexe. Un métabolon se formerait de façon à favoriser la biosynthèse d'une famille de composé, par exemple vers le développement d'anthocyanes lors de la phase de maturation de la plante. La seconde théorie, défendue par Ralston *et al.* [58] propose un modèle où les différentes enzymes, participant à plusieurs biosynthèses, sont empaquetées ensemble. Par exemple, ils émettent l'hypothèse d'un métabolon commun entre ces deux voies, comme le montre la Figure 14.

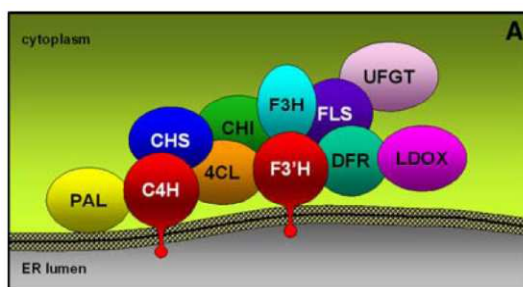


Figure 14 : Représentation d'un métabolon qui serait à la fois impliqué dans la biosynthèse des flavonoïdes et dans celle des flavonols, d'après l'article de Ralston *et al.*[58]

Il faut toutefois noter que malgré cette légère différence, les auteurs des deux revues sont en accord sur la plupart des points caractérisant ce type de complexe.

Le rôle des enzymes de type cytochrome P450 est double dans ce type de complexe. En effet, au delà de leur rôle catalytique dans la biosynthèse des flavonoïdes (exemple de la F3'H qui gère le degré d'hydroxylation du cycle B), elles servent aussi d'ancre à la membrane [58] pour le complexe multienzymatique par leur partie N-terminale qui est transmembranaire [59, 60]. Cette propriété des enzymes de type cytochrome P450 induit un phénomène de co-localisation des protéines autour de la membrane (voir Figure 13 et Figure 14), propice à la formation des complexes supramoléculaires. De par cet ancrage, de nouvelles interactions de surface enzyme-membrane vont être créées, permettant d'apporter de la stabilité à ce complexe [61]. Ainsi, cette ancre serait donc la clé permettant au complexe d'exister suffisamment longtemps pour permettre la cascade de réactions chimiques.

La compréhension des mécanismes associés à la formation de ce métabolon, ainsi que de ses propriétés, passe par l'étude de 4 points:

- Interactions entre enzymes
- Mécanismes d'entrée et de sortie des composés
- Interactions protéines/membrane
- Réactivité enzymatique

Dans ce travail, nous avons choisi d'étudier le métabolon au travers d'un complexe de trois enzymes, qui regroupent l'ensemble des points clés cités. Les résultats pourront ainsi être extrapolés sur un complexe multienzymatique aussi important que la Figure 13, dont l'étude totale n'est pas réalisable. Nous avons donc choisi des enzymes qui nous permettront d'étudier tous les aspects de ce type de complexe. L'enzyme Flavonoïde-3'-hydroxylase (F3'H) a été choisie pour permettre l'étude des interactions protéine/membrane. Nous avons ensuite décidé de choisir la dihydroflavonol-4-reductase (DFR), très étudiée car elle constitue la dernière étape commune entre la biosynthèse des anthocyanes et des proanthocyanidines. Enfin, la troisième enzyme choisie est la leucoanthocyanidine reductase (LAR), première étape réactionnelle vers la formation des tannins condensés.

III) Description des enzymes choisies pour l'étude

III.1) La dihydroflavonol-4-réductase (DFR)

La dihydroflavonol-4-réductase est une enzyme composée de 329 acides aminés et appartenant à la classe des SDR (Short-chain dehydrogenases/reductase). Ce sont des enzymes qui catalysent plusieurs types de réactions (réduction, oxydation, isomérisation...), mais qui nécessitent la participation d'un cofacteur NADPH ou NADH [62, 63]. Même si le pourcentage d'identité sur toutes les SDR ne dépasse pas les 30%, les structures tridimensionnelles connues de ces enzymes montrent une haute conservation [64]. Par exemple, nous pouvons citer le site d'interaction du cofacteur, composé d'un motif riche en glycine (Gly-X3-Gly-X-Gly) et d'un motif structural de type "Rossmann Fold" (succession de 6 feuillets bêta parallèles couplés à deux paires d'hélices alpha [65]) conservé dans plus de 80% des SDR cristallisés. Le site actif est grandement conservé, notamment trois acides aminés : une sérine, une tyrosine et une lysine, qui constituent la triade catalytique essentielle à la réactivité des SDR. Le motif est de type Ser.....Tyr-X3-Lys et est conservé dans plus de 80% des SDR, mais l'absence de l'acide aminé tyrosine n'est pas forcément synonyme de non-fonctionnement de l'enzyme. En effet, elle peut être remplacée par une histidine dans certains organismes [66]. Cette triade intervient dans le mécanisme par une cascade de transfert de protons, impliquant la tyrosine, le cofacteur et la lysine [67]. La sérine permet quand à elle la création d'une liaison hydrogène qui stabilise la fonction réactive du substrat. Dans certaines SDR, on peut même parler de tétrade catalytique avec la présence d'une asparagine aidant à la reprotonation de la lysine après réaction, mais ce cas reste rare [63]. Les connaissances sur les différents protagonistes de la réaction sont ainsi bien renseignées, mais pour le moment nous ne savons pas si le mécanisme est par étape ou concerté.

En 2007, Petit *et al.* ont cristallisé l'enzyme DFR présente dans l'organisme *Vitis vinifera* [68] sous la forme d'un complexe ternaire (c'est à dire co-cristallisé avec le cofacteur, NADP⁺ et un substrat connu, la dihydroquercétine, voir Figure 15). Le site actif a pu être ainsi caractérisé, mais également les domaines de liaison du cofacteur et du substrat. La triade catalytique est composée de la sérine 128, la tyrosine 163 et la lysine 167. Les résidus de reconnaissance de la dihydroquercétine sont la glutamine 227 et l'asparagine 133. Des études de mutagénèse dirigée montrent que la mutation du résidu 133 hautement conservé (en aspartate ou leucine) modifie grandement la sélectivité de DFR vis-à-vis de ses substrats. L'asparagine permettrait une sélectivité de la dihydroquercétine par rapport au

dihydrokaempferol (voir Figure 6) par l'interaction avec l'hydroxyle du cycle B supplémentaire sur la dihydroquercétine. Si le résidu est muté, on observe soit une chute de la conversion des deux substrats (d'après le travail de Petit *et al.*[68]) dans la dihydroflavonol-4-réductase de *Vitis vinifera*, soit une inversion de la sélectivité en faveur du dihydrokaempferol dans l'enzyme présente dans l'organisme *Gerbera* [69]. Ces deux points montrent l'importance que revêt ce résidu dans la reconnaissance du substrat par l'enzyme.

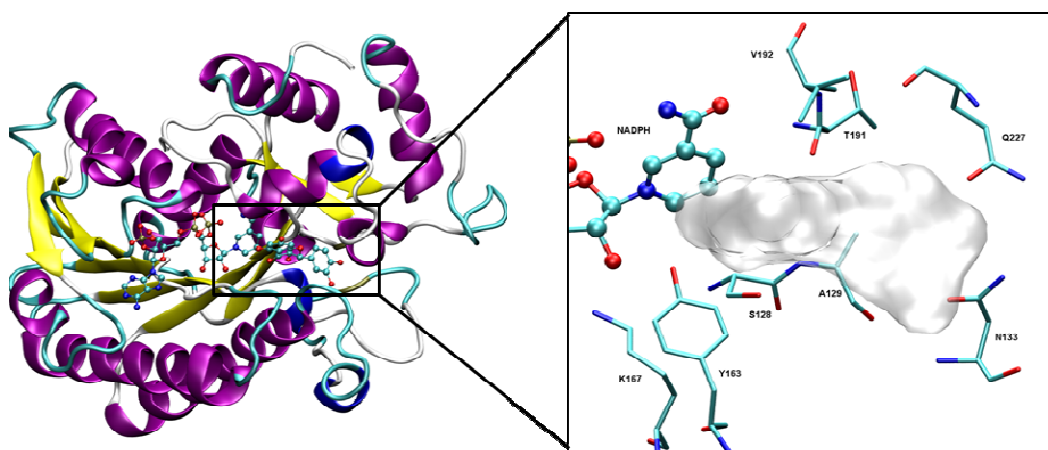


Figure 15 : à gauche, structure tridimensionnelle de l'enzyme DFR obtenu par Petit *et al.*(PDB ID:2C29). A droite, détail des résidus du site actif interagissant avec le substrat et représentation de la cavité de liaison.

L'identification des résidus permet de faire une proposition du mécanisme de réduction de la dihydroquercétine en leucocyanidine, comme montrée sur la Figure 16.

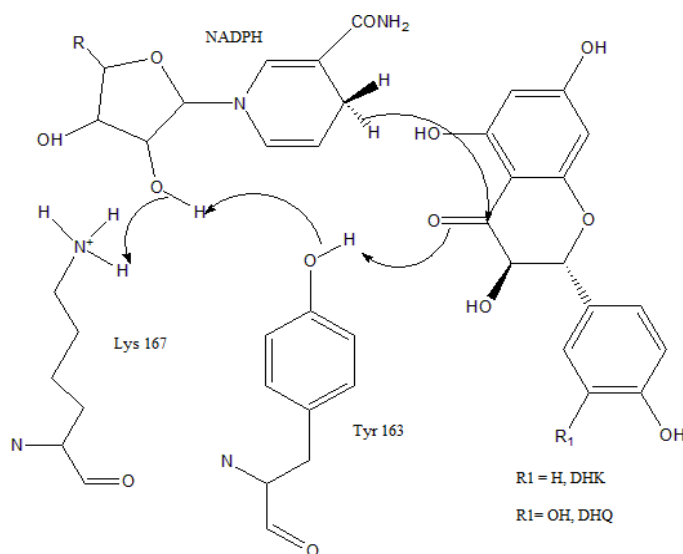


Figure 16 : Proposition de mécanisme enzymatique réalisée par la DFR.

Le cofacteur NADPH transfère un hydrogène sous forme d'hydrure de manière stéréospécifique (configuration de type pro-S) sur le groupement cétone, ce qui implique sa réduction en alcoolate. Ce groupement sera ensuite protonné en hydroxyle selon un mécanisme basé sur une cascade de transfert de protons impliquant la tyrosine 163, le NADPH une seconde fois et enfin la lysine 167. La sérine 123 de la triade interagit avec le groupement cétone afin de le préparer à la réaction.

D'autres cristaux de DFR ont également été obtenus (PDB id : 3C1T et 3BXX). Basée sur leur structure, une inhibition de l'activité de l'enzyme peut être envisagée par interaction avec des flavonols [70]. Enfin, une étude cinétique récente montre que l'excès de dihydroquercétine peut être également un facteur d'inhibition de l'enzyme, par formation d'un complexe enzyme-dihydroquercétine abortif ou d'un complexe ternaire non productif [71]. Ce dernier point sera abordé dans ce document au sein d'une partie annexe.

III.2) La leucoanthocyanidine réductase (LAR)

La leucoanthocyanidine réductase est une enzyme composée de 346 résidus, et qui appartient à une sous famille de réductase dérivée des SDR, les PIP (Pinoresinol-lariciresinol reductase, Isoflavone reductase, et Phenylcoumaran benzylic ether reductase *like*) [72]. Le point commun principal des trois enzymes qui composent cette famille est leur capacité de clivage d'une liaison carbone-oxygène liée à un para-phénol. Cette réaction est assez éloignée des réactions classiques des SDR, qui consistent en une réduction d'une double liaison C-C ou C-O en simple liaison. Une autre caractéristique de cette sous famille d'enzymes est l'absence de la triade catalytique, contrairement aux autres SDR. La tyrosine et la lysine sont les seuls résidus présents de la triade, la lysine possédant un rôle catalytique important [73].

Dans leur étude, Maugé *et al.* [74] ont cristallisé LAR sous trois états :

- une forme apo de l'enzyme, c'est à dire vide de tout partenaire biologique (PDB ID : 3I5M)
- une structure avec seulement le cofacteur (PDB ID : 3I6I).
- un complexe ternaire enzyme/NADPH/catéchine (le produit de la réaction) (PDB ID : 3I52). Elle est représentée sur la Figure 17.

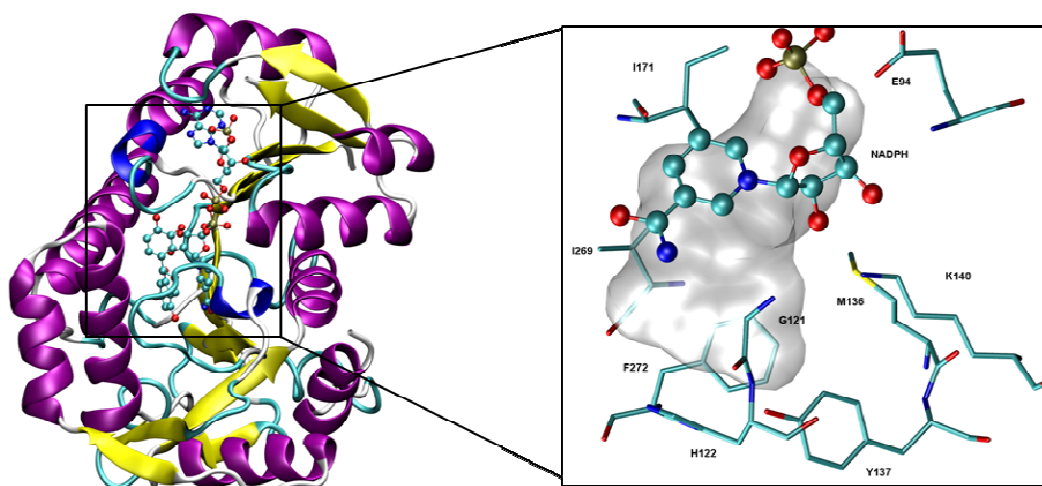


Figure 17 : Structure de LAR correspondante au fichier PDB 3I52. A droite, un zoom sur le site actif et les différents résidus impliqués dans l'interaction avec le substrat, ainsi que la représentation de la cavité de liaison.

Le site de fixation du cofacteur est très similaire aux SDR, avec un motif "Rossman Fold". L'histidine 122, la lysine 140 et la tyrosine 137 interagissent avec le produit, et donc jouent probablement un rôle dans la reconnaissance du substrat (leucocyanidine). L'obtention du complexe tertiaire permet de comprendre comment le site actif est organisé autour du cofacteur et du substrat. Maugé *et al.* ont proposé une hypothèse de mécanisme qui est très différente de ce que l'on peut observer dans une SDR. Ce mécanisme, représenté sur la Figure 18, implique que la lysine 140, considérée comme étant la lysine catalytique des PIP, soit déprotonée pour jouer le rôle de base et qu'une histidine protonée qui jouerait le rôle d'acide. Ces deux conditions seraient nécessaires à l'initiation de la réaction.

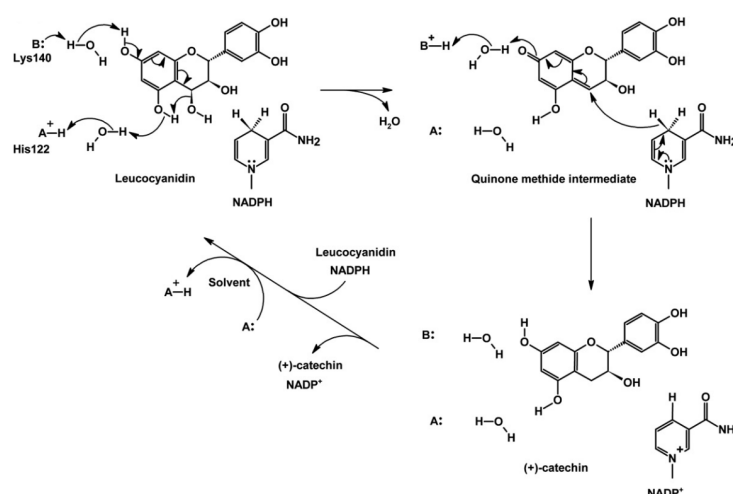


Figure 18 : Mécanisme proposé par Maugé *et al.* pour la transformation de leucocyanidine en catéchine. Tiré de la ref [74].

III.3) La flavonoïde-3'-hydroxylase (F3'H)

L'enzyme flavonoïde-3'-hydroxylase est composée de 509 acides aminés et appartient à la famille des cytochromes P450 (CYP). C'est une famille de protéines très répandue dans le vivant et qui catalyse généralement des réactions d'oxydation. Elle permet la synthèse de molécules vitales pour l'Homme (Vitamine D, œstrogène, testostérone...). Dans le métabolisme de drogues, 75% des réactions mises en jeu impliquent une enzyme de type cytochrome P450 [75]. A ce jour, plus de 18 classes de cytochromes P450, basées sur la nature de leurs substrats, ont été identifiées dans le corps humain, et pas moins de 18000 protéines différentes réparties dans tout le monde du vivant, y compris les virus, ont pu être découvertes [76]. Tous les membres de cette famille ont une identité de 40% sur leur séquence primaire, ainsi que des caractéristiques structurales très proches. Dans les organismes eucaryotes, les enzymes de type CYP sont toutes liées à la membrane par leur partie N-terminale [59, 77]. Seules les cytochromes bactériennes n'interagissent pas avec la membrane. La Figure 19 représente la structure obtenue par diffraction de rayons X d'une enzyme de type cytochrome P450 (la cytochrome humaine P450 1A2 [78]). Nous pouvons noter que la partie N-terminale est tronquée (voir Figure 19 en rouge) car elle empêche de manière systématique la cristallisation des enzymes de cette catégorie.

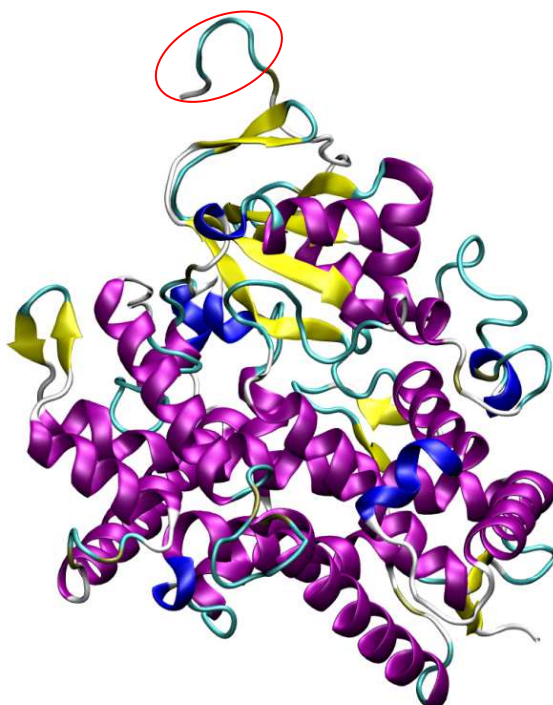


Figure 19 : Structure cristallographique de l'enzyme humaine Cytochrome P450 1A2. La partie N-terminale, transmembranaire, n'est pas représentée car non présente dans la structure, mais devrait être reliée à la partie entourée en rouge.

Ces enzymes possèdent toutes un groupement hème dans leur site actif qui va être le principal donneur d'électrons (grâce à son atome de fer) dans les réactions d'oxydations. Ce cofacteur est lié à l'enzyme par une liaison de coordination avec un atome de soufre d'une cystéine, ainsi que par des interactions ioniques. Toutes les enzymes de type cytochrome P450 utilisent de manière générale un donneur d'électrons externe pour finaliser la réaction (cytochrome P450 réductase possédant en son sein une molécule de NADPH réalisant le transfert d'électrons [58]). La réaction fait intervenir l'atome de fer au centre du groupement hème.

A ce jour, plus de 60 structures de cytochromes P450 ont été cristallisées et résolues, la plupart issues du corps humain et des bactéries [79]. Malheureusement, la structure cristallographique associée à la flavonoïde-3'-hydroxylase est manquante. Son rôle n'est pourtant pas à négliger : au delà de la catalyse de la réaction d'hydroxylation du cycle B des flavonoïdes, elle participerait à la stabilisation d'un possible métabolon en jouant le rôle d'ancre à la membrane.

IV) Apport de précédentes études théoriques à la compréhension de la biosynthèse des flavonoïdes.

Le travail présenté ici consiste donc en l'étude de ces trois enzymes (DFR, LAR et F3'H), afin de proposer un premier modèle d'un métabolon réduit. L'objectif est de comprendre quels sont les mécanismes mis en jeu lors des phénomènes associés, comme par exemple le processus de "substrate channeling". Pour cela, l'utilisation de méthodes théoriques permet d'apporter un nouvel éclairage sur ces points difficiles à appréhender expérimentalement. L'application de méthodes de chimie et de biochimie théoriques à par ailleurs déjà permis l'explication de plusieurs phénomènes associés à la biosynthèse des flavonoïdes.

La diversité des méthodes de chimie théorique existantes (chimie quantique, modélisation moléculaire, méthodes hybrides...) permet de traiter et de comprendre de nombreux aspects de la biosynthèse, que ce soit l'étude d'intermédiaires biosynthétiques, l'étude des enzymes ou encore leur réactivité. Dans le cadre de la biosynthèse des flavonoïdes, certaines enzymes ont fait l'objet d'études théoriques sur lesquelles nous nous sommes appuyées pour établir nos hypothèses de travail. Les méthodes les plus utilisées sont les protocoles de reconstruction par homologie, afin d'obtenir des structures non cristallisées et

les calculs de chimie quantique et/ou de type hybride QM/MM afin d'étudier la réactivité et les mécanismes enzymatiques.

En ce qui concerne CHI, de nombreuses études mécanistiques ont été réalisées sur la réaction enzymatique. L'enzyme, cristallisée en 2000 par l'équipe de Jez *et al.* [80], catalyse un réarrangement interne stéréospécifique de la chalcone pour former les 2S-flavanones (cf Figure 3). Ce réarrangement constitue un mécanisme simple à étudier car il ne fait pas intervenir de cofacteur, seulement l'enzyme et la molécule [81]. Ainsi, ce type de réaction est une étude de choix pour l'application de différentes théories associées aux mécanismes enzymatiques : la théorie du NAC (Near-attack conformer [82]) et la théorie de l'état de transition. Dans l'étude de Ruiz-Pernia *et al.* [83], les auteurs ont montré, par l'utilisation de protocoles hybrides QM/MM, que l'enzyme favorisait de manière intéressante uniquement l'isomérisation des composés S-trans des chalcones, qui sont les moins stables en milieu aqueux d'après leurs calculs. Ils démontrent également que l'enzyme diminue de manière importante la barrière énergétique associée au réarrangement du substrat par la stabilisation de l'atome de carbone C2 qui, portant une charge δ^- durant la réaction, se verra être stabilisé par interaction avec la lysine 97 du site actif. Par la suite, cette équipe a publié une nouvelle étude sur le mécanisme et notamment sur les effets électrostatiques de l'environnement sur la réaction [84]. La nouveauté provient de la prise en compte du potentiel électrostatique créé par l'environnement sur le nucléophile en tant que coordonnée de réaction. Cela a permis de démontrer que l'enzyme possède des caractéristiques électrostatiques optimales pour le déroulement du réarrangement de Claisen.

Le mécanisme des enzymes 2-oxoglutarate dépendante, comme l'ANS, a été grandement étudié dans le domaine de la chimie quantique. Le mécanisme de la réaction, faisant intervenir un oxoglutarate mais également un ion Fe^{2+} , reste un sujet de débat [85, 86]. Les possibilités sont nombreuses et l'oxydation de la molécule peut être réalisée sur l'atome de carbone C3 ou C4. En 2006, une étude théorique de Nakajima *et al.* [87] montre à l'aide de méthodes de chimie quantique (niveau de calcul B3LYP/3-21G**) que l'oxydation se fera préférentiellement sur l'atome de carbone C4, même si les deux hypothèses sont possibles. Par ailleurs, l'état de spin du fer sera quintet et non triplet. L'oxydation sur l'atome de carbone C4 induit la formation de la dihydroquercétine à partir de la leucocyanidine (le substrat de l'ANS). Cela constitue une réaction différente de ce à quoi on s'attendait. Il est ainsi démontré que la formation des pigments des plantes, les anthocyanidines, ne sont qu'une réaction secondaire de l'ANS qui catalyse la formation de DHQ de manière majoritaire. Ces derniers

peuvent à nouveau réagir, selon un deuxième cycle catalytique, pour donner des flavonols et notamment la quercétine.

La quercétine est le membre le plus prédominant de la famille des flavonols, ce qui en fait un parfait cas d'étude de la solubilité de cette famille. Chebil *et al.* [88] ont calculé les propriétés de solubilité de la molécule dans différents solvants. La méthodologie développée, utilisant des calculs d'énergie libre avec prise en compte des effets d'empilement des molécules, permet de déterminer de manière précise les solubilités dans différents solvants. Ici, les valeurs obtenues par ces calculs sont très proches des valeurs expérimentales, validant ainsi le protocole et permettant de pouvoir l'utiliser dans le but de prédire la solubilité dans des cas plus complexes.

Le mécanisme d'oxygénolyse de la quercétine est également une réaction d'intérêt de par la nature de la dioxygénase utilisée. En effet, la quercétine 2,3-dioxygénase, cristallisée en 2002, est la seule représentante de cette famille qui possède un atome de cuivre comme cofacteur [89, 90]. Généralement, les dioxygénases possèdent un atome de fer, parfois un atome de manganèse ou de magnésium. Deux études utilisant les méthodes de chimie quantique proposent une description du mécanisme. Fiorucci *et al.* proposent [91], via l'utilisation d'un modèle simplifié de l'enzyme, un mécanisme passant par une cycloaddition de type 1,3, même si la réaction de type 1,2 est apparemment facilitée. Cependant, la position nécessaire à la réalisation de cette dernière n'est obtenue ni par le docking réalisé, ni expérimentalement. Siegbahn poursuit ce travail par l'étude du mécanisme complet de la quercétine 2,3-dioxygénase par le biais de méthodes de type DFT, permettant ainsi la résolution du mécanisme enzymatique [92].

Les propriétés antioxydantes et antiradicalaires étant l'intérêt principal des flavonoïdes et des polyphénols en général, la recherche sur leur mécanisme d'action est très active [34, 93-95]. Ces travaux démontrent le potentiel anti radicalaire de différentes molécules (daidzein comme isoflavone, quercétine pour les flavonols et isoorientine pour les flavones), par l'utilisation de diverses méthodes comme l'étude d'orbitales frontières, le potentiel électrostatique, ainsi que la réaction avec des radicaux OH $\cdot$. Concernant le daidzein, l'étude utilise la dynamique réactionnelle par méthode QM/MM afin de prendre en compte les effets du solvant sur les propriétés anti-radicalaires.

Pour la chalcone synthase, de nombreux travaux ont été réalisés afin d'en étudier tous les aspects. Au delà des études cristallographiques déjà citées auparavant [37, 38], qui ont

permis une proposition de mécanisme associé à la formation des chalcones, une étude de modélisation par homologie de différents mutants de l'enzyme CHS de l'organisme *Arabidopsis thaliana* a été réalisée [96]. Les auteurs tentent de proposer une explication rationnelle sur l'impact des mutations sur la structure de l'enzyme par la modélisation de ces différents mutants. Même si elles se situent loin du site actif, elles ont un rôle important sur la structuration de ce dernier. Dans certains cas, la mutation (par exemple W174P; tryptophane en proline) permet d'éviter une accumulation de flavonoïdes dans le site actif, alors que cette mutation ne concerne pas directement cette partie de l'enzyme. Cela montre que le réarrangement structural, induit par cette mutation, en est la cause. A ce jour, rien n'a été réalisé sur l'étude mécanistique de l'enzyme à partir des propositions données par la structure cristallographique de CHS.

Au vu de l'importance de la famille des SDR, de nombreuses études ont été réalisées, notamment sur la triade catalytique de cette famille. Des études de reconstruction par homologie couplées à de la mutagenèse dirigée démontrent le rôle important de la triade catalytique pour l'interaction du substrat avec l'enzyme mais également dans la réactivité de celle-ci [97]. Une autre étude de Gani *et al.* [67] montre par des simulations de dynamique moléculaire couplées à des calculs de docking que la lysine reste de manière constante sous forme protonnée. La tyrosine peut, quand à elle, exister sous la forme tyrosinate (pKa diminué par la présence de la lysine), permettant de proposer de nouvelles hypothèses pour la conception de nouveaux inhibiteurs pour les SDR. Cependant, la sérine ne joue que le rôle de stabilisation du substrat par liaison hydrogène.

Les flavonols sont obtenus grâce à l'action de l'enzyme flavonol synthase (FLS). Les laboratoires de génétique se sont beaucoup intéressés à cette enzyme. Notamment, Fujimoto *et al.* utilisent, en complément des analyses des gènes codant la FLS, des méthodes de modélisation par homologie afin de reconstruire cette enzyme à partir de la structure cristallographique de l'anthocyanidine synthase d'*Arabidopsis thaliana* [98]. Il apparaît que le génome de cet organisme peut coder 5 types de FLS différentes, mais seulement une seule d'entre elles possède des propriétés catalytiques. En se basant sur les séquences des différentes FLS, la partie en N-terminale constituée de 30 acides aminés est retrouvée dans la FLS catalytique et dans toutes les dioxygénases présentes dans *Arabidopsis thaliana*, mais n'est pas identifiée dans les autres FLS. La reconstruction par homologie montre que la structure globale des cinq types de FLS est pratiquement identique (RMSd faible) mais des différences substantielles apparaissent, notamment sur les résidus coordonnant le métal nécessaire à la

réaction (ASP 222 et HIS 220). Nous pouvons remarquer également que la partie N-terminale doit jouer un rôle important dans la réactivité car si elle est tronquée, la FLS catalytique perd ses propriétés réactives. Ceci peut expliquer pourquoi les autres types de FLS ne peuvent pas être réactives, ne possédant pas cette partie de 30 acides aminés.

Les études théoriques permettent donc d'aborder et d'apporter des précisions, voire des réponses sur tous les aspects des flavonoïdes, de leurs propriétés intrinsèques à leur voie de biosynthèse. Cependant, l'étude des complexes supramoléculaires reste encore à effectuer. Une étude récente propose un premier modèle de métabolon qui est impliqué dans la biosynthèse du dhurrin, un membre des glucosides cyanogéniques (représenté sur la Figure 20).

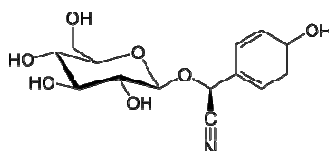


Figure 20: Structure de la molécule de Dhurrin.

Ces molécules sont impliquées dans les mécanismes de défense de la plante lorsque cette dernière est attaquée [99] et plusieurs preuves expérimentales montrent que la biosynthèse de cette molécule passe par la formation d'un métabolon [100]. Jensen *et al.* ont ainsi proposé un premier modèle d'un métabolon impliqué dans cette biosynthèse par l'utilisation des méthodes de reconstruction par homologie [101]. Trois enzymes, deux de types cytochromes P450 et une réductase ont été reconstruites suivant des structures similaires déjà résolues. En identifiant les résidus interagissant avec la membrane, ils ont pu positionner les deux enzymes de type CYP. Puis, en se basant sur des analyses de surface et de potentiel électrostatique, les auteurs proposent un modèle d'interaction entre les cytochromes et la réductase. Cela constitue une première incursion dans l'étude structurale d'un complexe supramoléculaire.

Cela étant, aucune étude à ce jour n'a été publiée sur le métabolon associé à la biosynthèse des flavonoïdes. L'utilisation de méthodes théoriques, comme les simulations de dynamique moléculaire, n'a jamais été réalisée à ce jour sur ce type de complexe. Le travail présenté ici consiste donc en la proposition d'un premier modèle de métabolon impliqué dans la biosynthèse de ces composés antioxydants au sein de l'organisme *Vitis vinifera*. Ce dernier impliquera les enzymes DFR-LAR-F3'H et la membrane du réticulum endoplasmique, permettant ainsi l'étude de tous les avantages offerts par sa formation. Les différents aspects et

propriétés intrinsèques à ce complexe seront étudiés par le biais de l'application de différentes méthodes avancées constituant l'état de l'art dans les domaines abordés (réactivité enzymatique, interaction protéine-protéine, "substrate channeling", interaction protéine-membrane).

Problématique générale

De par sa constitution impliquant de nombreux partenaires biologiques tels que substrats, cofacteurs, enzymes, membrane cellulaire... un métabolisme se révèle être une machinerie moléculaire complexe ne pouvant être modélisée par un seul type de protocole théorique. Dans cette première étude de ce type de structure, qui se veut avant tout être exploratoire, nous allons analyser les différentes facettes de la reconstruction et du comportement dynamique d'un tel système multienzymatique impliqué dans la biosynthèse des flavonoïdes. Les protocoles employés et les résultats analysés seront développés dans les différents chapitres de ce manuscrit. Afin de pouvoir appréhender de manière complète ce complexe, nous devons analyser plusieurs points cruciaux, notamment :

1. les interactions entre protéines et membrane
2. les interactions entre différentes protéines
3. les mécanismes d'entrée-sorties des composés
4. la réactivité associée aux enzymes

1) L'ancrage à la membrane de ce système multienzymatique est fortement dépendant de l'interaction de l'un de ses composants avec les molécules constituant la double couche lipidique. Cet ancrage permet de définir une place fixe à une des protéines. Ainsi, l'association d'une autre enzyme ne sera fonction que de la diffusion de cette dernière, en considérant en première approximation une membrane fixe. Lors de l'association des autres enzymes, un deuxième effet d'intérêt pourrait intervenir. En effet, il est possible que les enzymes non-ancrées puissent montrer une zone d'affinité forte avec la membrane lipidique, par des interactions en surface. L'association pourrait alors être considérée selon une interaction à trois corps et plus simplement comme un "simple" arrimage protéine/protéine. Nous devons donc prédire si ces entités se lient à la surface lipidique. Pour cela, le serveur OPM (orientation of protein in membrane) peut nous donner quelques clés afin d'anticiper la position de l'enzyme considérée (en surface ou au cœur de la membrane en cas de protéine transmembranaire) [102]. Bien que la structure tridimensionnelle de l'enzyme F3'H ne soit pas connue, il est possible d'en proposer une à l'aide de protocole de reconstruction par homologie. En se basant sur les structures des cytochromes P450 déjà existantes, et dont l'alignement entre séquences primaires présente des taux d'identité et d'homologie satisfaisants, le logiciel MODELLER [103] permet de reconstruire la structure de l'enzyme considérée. Après vérification de la pertinence de la structure, celle-ci est validée comme étant utilisable pour la prédiction de l'interaction avec la membrane. Ces modèles en

interaction devront être soumis à des protocoles de dynamique moléculaire afin d'en estimer la stabilité et de pouvoir en tirer un mode d'interaction privilégié.

2) Les interactions entre protéines constituent le point central de l'étude du métabolisme. La prédiction des structures multiprotéiques est en effet cruciale pour la pertinence de l'analyse des phénomènes de "substrate channeling", qui implique que les molécules suivent un chemin établi entre les sites actifs sans avoir à subir de long processus diffusifs dans l'espace cytosolique. Ces complexes multienzymatiques ont été obtenus par l'application du protocole ATTRACT [104-106]. Dans ce protocole, un champ de force de type knowledge-based potential est utilisé. Il a été développé sur la base d'une représentation gros grain des acides aminés des protéines et, en plus de permettre une diminution du temps de calcul, mène à des familles de solutions clairement différenciées en fonction de leurs caractéristiques d'interaction. En résumé, il permet de rejeter plus clairement les solutions menant à de faibles énergies de stabilisation ou à des interactions défavorables. Plus de 10000 structures peuvent être ainsi générées. Malgré tout ici, nous ne pouvons baser notre recherche que sur les résultats obtenus par ce protocole. En effet, si nous partons du postulat que les enzymes en interaction doivent permettre une diffusion aisée des substrats/produits d'un site actif au suivant, il est alors nécessaire d'inclure cette contrainte dans le choix final de la structure protéine/protéine. Afin de biaiser au minimum nos choix, deux types de procédures ont été utilisées et comparées : une analyse énergétique, c'est à dire les meilleures poses de docking, suivie par un raffinement structural (recherche de solutions préservant une proximité spatiale des sites actifs) ou, à l'inverse, la production de structure de docking en excluant certaines zones d'interaction, suivie cette fois-ci, d'un choix basé sur une analyse énergétique. L'application de protocole de dynamique moléculaire sur ces structures hypothétiques a permis d'en tester la stabilité et, au final, de converger vers une structure compatible à la fois avec une interaction forte et des probabilités de diffusion de molécules.

3) Dès lors, la caractérisation des voies d'accès aux sites actifs devient un élément central de la définition de ces interactions protéine/protéine. Elle a été réalisée par l'application de la méthode du RAMD (Random acceleration molecular dynamic [107]). Le principe est d'appliquer sur le substrat une accélération aléatoire dans une direction de l'espace, le but étant de le forcer à quitter sa position initiale pour explorer de nouvelles conformations. L'intérêt principal est de partir d'une structure dans laquelle le substrat est dans le site actif, puis de lui appliquer l'accélération afin de détecter quel chemin il suit afin de sortir du cœur de l'enzyme. Il est alors possible d'étudier les voies de diffusion des

produits/substrats d'un site actif à l'autre. Pour cela, la méthode de choix serait d'échantillonner le comportement de ceux-ci le long d'une trajectoire de dynamique moléculaire suffisamment longue pour que le transfert soit effectif. Hélas, ces simulations sont relativement coûteuses en temps de calcul pour un temps simulé effectif "relativement" court. Ces diffusions sont dépendantes de nombreux facteurs. On peut noter parmi eux l'implication de molécules d'eau qui viendraient perturber les interactions substrat/enzyme mais aussi, à une autre échelle, les processus allostériques appliqués à chaque enzyme et découlant de leurs interactions réciproques. Cela permet d'obtenir des informations sur le profil énergétique lié au chemin considéré.

4) Enfin, il est intéressant de se pencher sur la réactivité de chacune de ces enzymes. Chacune d'entre elles catalyse une réaction qui lui est propre et cette réaction peut être modulée par son interaction avec les autres enzymes. On parle alors d'effets allostériques. En effet, l'interaction d'une protéine avec une autre peut provoquer des changements structuraux globaux et locaux, légers ou importants, et qui peuvent alors modifier l'architecture du site actif de l'enzyme considérée.

Comparé à une réaction se déroulant dans un milieu aqueux par exemple, le site actif enzymatique constitue un milieu propice aux réactions chimiques. En effet, la mise en contact des différents protagonistes, ainsi que les effets électrostatiques, dynamiques et stériques de l'environnement de la protéine, induisent de grandes modifications sur le déroulement de la réaction. C'est pourquoi la prise en compte de la totalité de ces effets est nécessaire pour réaliser une étude pertinente d'une réaction enzymatique. A ce titre, la mécanique quantique ne peut prétendre à une étude pertinente de ce genre de réaction, car la prise en compte d'un très grand nombre d'atomes rendra impossible la réalisation du calcul dans des échelles de temps acceptables. La mécanique et la dynamique moléculaire, quant à elles, sont capables de permettre la modélisation de systèmes biologiques entiers, mais ne décrivent pas la densité électronique associée, et donc ne permettent pas non plus l'étude de la réactivité de l'enzyme. Nous devons donc utiliser des méthodes hybrides de type DM-QM/MM, qui sont des méthodes à l'état de l'art pour l'analyse de la réactivité enzymatique. Ces dernières allient la prise en compte d'une partie quantique pour permettre la réaction, entourée d'un environnement traité en mécanique classique afin de considérer l'influence du milieu enzymatique ainsi que le caractère dynamique associé au système [108].

Dans notre étude, la dihydroflavonol-4-reductase joue le rôle de pivot dans notre métabolisme réduit à 3 enzymes. Nous allons donc observer la diffusion de molécules à partir de F3'H vers DFR, puis la diffusion du produit de la réaction de cette dernière vers LAR. L'étude du mécanisme enzymatique lié à DFR est donc primordiale pour assurer la continuité de l'étude et de l'histoire associée à cette portion de la biosynthèse des flavonoïdes. L'ensemble de ce travail permet ainsi d'appréhender tous les phénomènes associés à la formation d'un métabolite. La Figure 21 résume de manière schématique l'ensemble des études effectuées sur nos trois systèmes.

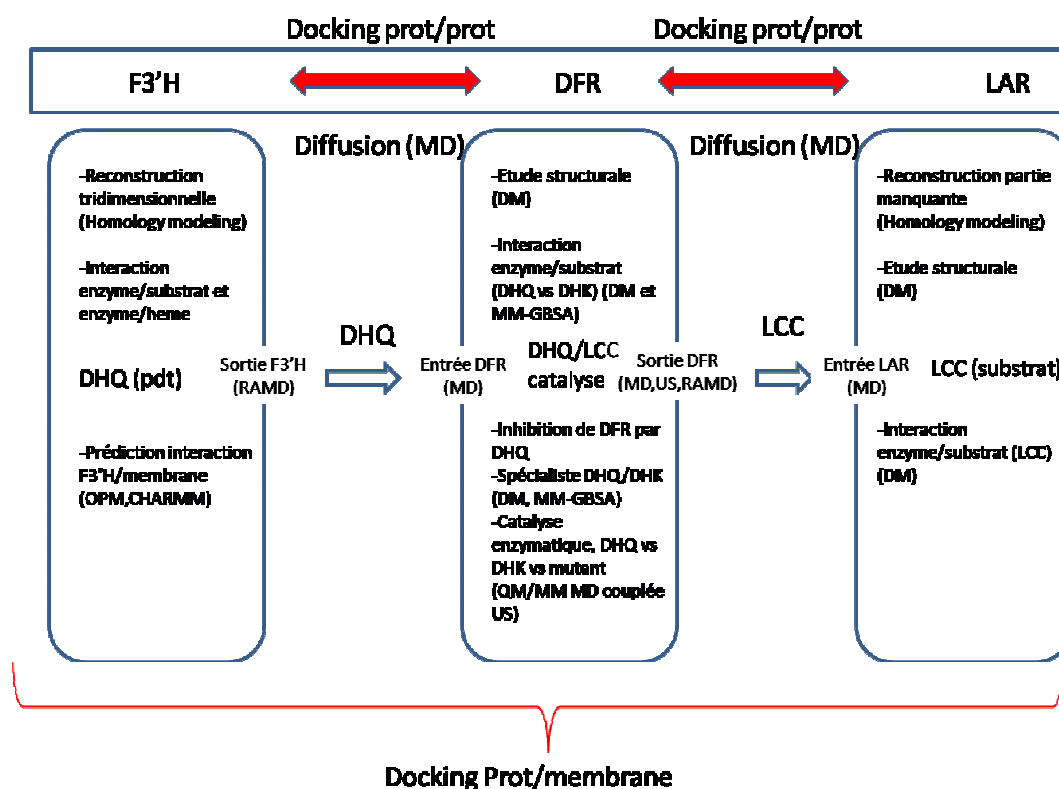


Figure 21 : Schéma récapitulant l'ensemble du travail effectué sur cette portion de la biosynthèse.

Ce document résume l'intégralité des travaux réalisés pour l'étude du métabolite composé des enzymes DFR, LAR et F3'H. Après une présentation des différentes méthodes utilisées pour la réalisation de ce projet, une première partie décrivant la préparation des différentes enzymes sera présentée. Les résultats concernant la réactivité de l'enzyme centrale de notre métabolite, DFR, seront ensuite mis en valeur dans une deuxième partie. Le troisième chapitre de ce manuscrit sera consacré à la présentation des premiers complexes DFR-LAR, ainsi que la caractérisation du phénomène de "substrate channeling". Dans une dernière partie

seront présentés nos résultats sur les interactions entre F3'H et la membrane cytosolique, ainsi que nos premiers modèles de métabolisme à 3 corps. Enfin une dernière partie annexe présentera un modèle expliquant un phénomène d'inhibition par excès de substrat associée à l'enzyme DFR.

Méthodologie

Comme indiqué précédemment, l'étude de tous les aspects caractérisant une structure telle que le métabolon requiert l'utilisation d'un large éventail de méthodes théoriques. Nous aborderons ici une description des méthodes utilisées pour la réalisation de ce travail.

1) Modélisation Moléculaire

Basée sur les principes de la mécanique classique de Newton, l'application de la mécanique moléculaire permet de représenter un système à l'échelle atomique et d'en calculer l'énergie. Réaliser ce calcul pour différentes conformations de la structure considérée permet de caractériser différents points de la surface d'énergie potentielle associée au système. Chaque point de la surface est donc associé à une probabilité d'échantillonnage, ce qui induit la création d'une statistique de rencontre des conformations du système. Cette méthode de calcul se base sur la sommation de différents termes et est très utilisée pour la description de systèmes de taille importante, excédant le millier d'atomes [109, 110]. Dans ce paradigme, l'énergie potentielle d'un système est donnée par la relation suivante :

$$E_{potentielle} = E_{liée} + E_{non\ liée} \quad (\text{Équation 1})$$

avec

$$E_{liée} = E_{liaison} + E_{angle} + E_{dièdre} \quad (\text{Équation 2})$$

et

$$E_{non\ liée} = E_{électrostatique} + E_{vdW} \quad (\text{Équation 3})$$

Les termes d'énergie "liée" correspondent aux contributions des liaisons, angles et angles dièdres. Les atomes sont considérés comme des sphères non déformables, possédant une charge et un rayon de van der Waals. Les liaisons qui relient les atomes, ainsi que les angles associés, sont représentés par des potentiels harmoniques. Les angles dièdres sont quand à eux modélisés par une fonction sinusoïdale, qui permet la caractérisation de tous les états pouvant être échantillonnés par une torsion de ce type d'angle.

$$E_{liaison} = k (x - x_{eq})^2 \quad (\text{Équation 4})$$

$$E_{angle} = k(\theta - \theta_{eq})^2 \quad (\text{Équation 5})$$

$$E_{dièdre} = \frac{V_n}{2} (1 + \cos[i\phi - \gamma]) \quad (\text{Équation 6})$$

E_{liaison} , E_{angle} et $E_{\text{dièdre}}$ représentent les énergies de déformations des liaisons et angles considérés et de la rotation de l'angle dièdre correspondant. x_{eq} et θ_{eq} sont respectivement la distance à l'équilibre et l'angle à l'équilibre impliquant les atomes considérés. La constante de force du potentiel est donnée par la valeur de k (Vn dans le cas de l'angle dièdre). Plus celle-ci est élevée, plus la liaison (ou l'angle) que représente ce potentiel sera forte, impliquant une augmentation drastique de l'énergie en cas d'éloignement de la valeur d'équilibre. Pour l'angle dièdre, ϕ représente la valeur de l'angle dièdre tandis que γ représente la phase de l'angle.

Pour maîtriser l'ensemble de la physique du système, la prise en compte de termes d'énergie "non liée" est nécessaire afin de représenter les interactions entre atomes éloignés. Il s'agit de l'interaction électrostatique et les interactions de Van der Waals.

$$E_{\text{électrostatique}} = \sum_{j>i}^i \frac{q_i q_j}{\epsilon d_{ij}}$$

(Équation 7)

L'énergie électrostatique est calculée selon un potentiel de Coulomb, et est reliée à la constante diélectrique du milieu. Dans cette équation, i et j représentent deux atomes de charge respective q_i et q_j , séparés par une distance d_{ij} .

$$E_{vdw} = \sum_{i<j} \left(\frac{\sigma_{ij}}{d_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{d_{ij}} \right)^6$$

(Équation 8)

Cette relation prend en compte les interactions entre atomes séparés de 3 liaisons minimum, ainsi que les interactions entre atomes non liés. Il permet de prendre en compte les interactions faibles dipôles-dipôles permanents (London), dipôles-dipôles induits (Debye) et dipôles induits-dipôles induits (Keesom), qui sont attractives et représentées par le terme en puissance 6, ainsi que les termes répulsifs entre atomes, décrits par le terme en puissance 12. σ_{ij} représente la distance à l'équilibre entre les deux atomes i et j .

Les paramètres nécessaires à l'utilisation de cette méthode ont été déterminés expérimentalement (par spectroscopie RX par exemple) ou par des calculs de chimie quantique, et sont répertoriés dans un champ de force. Il existe un grand nombre de champs de force (AMBER, CHARMM, Martini...), qui sont utilisables selon la nature du système que l'on étudie. Dans notre cas de figure, le champ de force AMBER (plus particulièrement le

champ de force ff03.r1) [111, 112], qui s'applique aux molécules d'intérêts biologiques, a été utilisé pour cette étude.

I.1) Minimisation d'énergie

L'étape de minimisation de l'énergie potentielle d'un système est cruciale dans le cadre de la préparation du modèle pour les simulations de dynamique moléculaire. Elle permet de corriger les défauts des structures initiales, mais également d'adapter le système au champ de force, et ainsi de trouver le minimum de la nouvelle surface d'énergie potentielle associée à la méthode utilisée. Le point de départ d'un modèle repose sur l'existence de structures cristallographiques (obtenues avec la cristallographie par rayons X ou par RMN), ou obtenue par des protocoles de reconstruction par homologie. La minimisation étant un processus différent de la simulation de dynamique moléculaire, dans lequel chaque point est décorrélié du suivant, aucune notion temporelle ou de température ne sont ici considérées.

La surface d'énergie potentielle représente le lien entre l'énergie du système et sa géométrie. Le principe général d'une minimisation est de modifier la géométrie du système (par déplacement d'atomes par exemple) afin que ce dernier puisse se déplacer sur la surface d'énergie et se retrouver dans un puits de potentiel. Pour cela, la dérivée de l'énergie est calculée afin de déterminer la pente de la surface. Ainsi, nous pouvons trouver quelle direction suivre pour diminuer l'énergie potentielle du système, de façon à se retrouver dans le puits d'énergie le plus proche. La Figure 22 représente de manière schématique le principe de la minimisation. Différents algorithmes existent pour permettre la minimisation de l'énergie : la méthode de la plus grande pente (Steepest-descent) et la méthode du gradient conjugué.

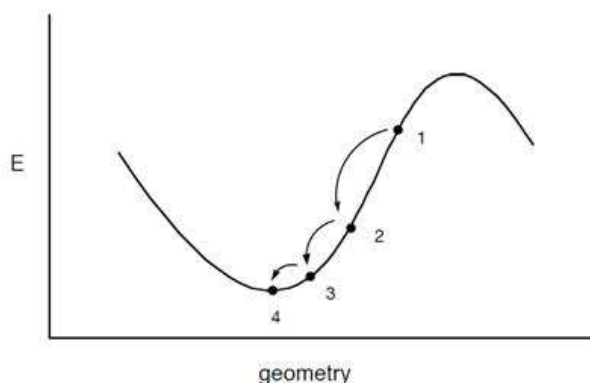


Figure 22: Schéma du principe de minimisation.

I.1.a) Méthode de la plus grande pente / Steepest-descent

Cette méthode consiste à rechercher la direction de plus grande pente selon laquelle la fonction $E(X)$ représentant l'énergie décroît la plus fortement. La direction de recherche D_k est donnée par l'opposé du signe du gradient de l'énergie :

$$\vec{D}_k = -\nabla E(r_k) \quad (\text{Equation 9})$$

Le pas α , définissant l'écart entre deux calculs, est ajusté pour chaque itération, et sera augmenté si la direction est accompagnée d'une diminution de l'énergie, diminué dans le cas contraire. Chaque direction calculée pour s'approcher du minimum d'énergie est orthogonale à la précédente. Le nouveau vecteur coordonnée $\vec{r}_{k+1}$ est calculé en utilisant le vecteur coordonnée $\vec{r}_k$ de l'étape précédente, de α_k ainsi que de la direction de recherche $\vec{D}_k$:

$$\vec{r}_{k+1} = \vec{r}_k + \alpha_k \vec{D}_k \quad (\text{Equation 10})$$

Cet algorithme a le mérite d'être rapide et peu coûteux en mémoire. Néanmoins, il rencontre des problèmes de convergence lors de l'approche d'un minimum où des oscillations de géométrie autour du minimum d'énergie peuvent alors apparaître.

I.1.b) Gradients conjugués

Cette autre méthode prend en compte les étapes précédentes afin de définir plus précisément la direction de recherche $\vec{D}_k$. Cette direction est calculée en tenant compte de la direction $\vec{D}_{k-1}$ en utilisant la relation suivante :

$$\vec{D}_k = -\nabla E(r_k) + \frac{\nabla E(r_k)^T \cdot \nabla E(r_k)}{\nabla E(r_{k-1})^T \cdot \nabla E(r_{k-1})} \vec{D}_{k-1} \quad (\text{Equation 11})$$

L'intérêt de cet algorithme (algorithme de Fletcher-Reeves qui est utilisé dans AMBER[113]) est d'éviter une oscillation de la géométrie du système autour de celle associée au minimum d'énergie. Il permet également d'accélérer la convergence puisque la direction est ajustée à chaque cycle, afin d'optimiser la recherche du minimum d'énergie. Cet algorithme présente deux principaux inconvénients. Tout d'abord, il est peu efficace pour éliminer les défauts structuraux comme des contacts entre atomes. D'autre part, la méthode des gradients conjugués possède un effet mémoire important (chaque nouvelle direction

dépend de toutes les anciennes), ce qui implique que si la structure initiale possède beaucoup de défauts, le résultat n'en sera que moins fiable. Il est préférable donc de ne pas utiliser uniquement cette technique pour la recherche d'un minimum. De manière générale, une combinaison de la méthode du Steepest-Descent, afin de déterminer rapidement la direction à prendre, avec la méthode du gradient conjugué, est utilisée pour trouver de manière plus précise la structure correspondant au minimum d'énergie local le plus proche de la structure initiale de la surface d'énergie potentielle.

I.2) Dynamique Moléculaire

La simulation de dynamique moléculaire est un outil qui permet l'étude de l'évolution du comportement dynamique de systèmes en fonction du temps. Cette méthode peut jouer un rôle prédictif sur de nombreux phénomènes (interaction ligand-récepteur, diffusion de molécules, formation de structures biologiques complexes, etc...). Elle obéit à la loi de Boltzmann, ce qui implique que les états de haute énergie sont difficilement échantillonnés. Cela signifie que les changements de conformation (le passage d'un puits de potentiel à un autre) sont rarement observés (ils sont d'ailleurs appelés "événements rares"). Le seul moyen d'augmenter la probabilité d'observer ces transitions est d'allonger autant que faire se peut le calcul, d'où l'intérêt de réaliser de longues simulations.

Le caractère probabilistique des calculs de dynamique moléculaire permet l'application de l'hypothèse ergodique. Cette hypothèse stipule qu'une grandeur obtenue par simulation de dynamique moléculaire (découlant donc d'une moyenne et d'un échantillonnage suffisamment long des conformations d'un seul représentant du système) est comparable à une valeur moyennée de cette grandeur sur un grand nombre d'individu obtenue de manière instantanée. Il nous est donc possible de déterminer, en plus du comportement dynamique, des critères supplémentaires comme des énergies d'interaction dans le cas d'un complexe ligand-récepteur, et de les comparer directement à l'expérience.

Le système étant défini par la mécanique classique, la simulation de son comportement dynamique est assurée par l'application de la deuxième loi de Newton pour le déplacement des atomes. La seconde loi de Newton relie la somme des forces appliquées sur le système à son accélération.

$$\sum \vec{F} = m \vec{a} \quad (\text{Équation 12})$$

Sachant que les forces appliquées sur le système sont directement liées au gradient de l'énergie calculé précédemment :

$$\vec{F} = -\vec{\nabla} E_{\text{système}} \quad (\text{Équation 13})$$

Dans les simulations de dynamique moléculaire, l'énergie du système est la somme de l'énergie potentielle (calculée par les équations précédemment citées de modélisation moléculaire) et de l'énergie cinétique des atomes.

$$E_{\text{système}} = E_{\text{potentielle}} + E_{\text{cinétique}} \quad (\text{Équation 14})$$

On peut ainsi réécrire l'Equation 12 comme suit:

$$-\frac{dE_{\text{système}}}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (\text{Équation 15})$$

Cette transformation de la seconde loi permet de relier la dérivée de l'énergie du système à ses coordonnées. Cette équation peut être intégrée afin de calculer en chaque point de la simulation les nouvelles coordonnées et construire ainsi une évolution du comportement dynamique du système. L'intégration de la loi de Newton repose sur l'utilisation du théorème de Taylor, qui permet d'approximer la dérivée d'une fonction à un polynôme.

$$\begin{aligned} r(t + \delta t) &= r(t) + v(t)\delta t + a(t)\frac{\delta t^2}{2} + \dots \\ v(t + \delta t) &= v(t) + a(t)\delta t + b(t)\frac{\delta t^2}{2} + \dots \\ a(t + \delta t) &= a(t) + b(t)\delta t + c(t)\frac{\delta t^2}{2} + \dots \end{aligned} \quad (\text{Equation 16})$$

L'intégration nécessitant la définition d'un intervalle entre deux valeurs, on définit le pas d'intégration δt comme étant l'écart entre la valeur initiale au temps t et la valeur au temps t_1 (égale à $t + \delta t$). Le principe est donc de calculer les coordonnées au temps t , puis d'utiliser un algorithme pour calculer la nouvelle position au temps $t + \delta t$, ainsi que les vitesses des différents atomes, jusqu'à l'obtention du temps de simulation voulu. Différents types d'algorithmes peuvent être utilisés afin d'automatiser l'intégration de la seconde loi de Newton. Les deux plus utilisés sont présentés ci-après.

1.2.a) Algorithme de Verlet

Cet algorithme prend en compte les positions et accélérations aux temps $t - \delta t$ et t pour calculer les positions au temps $t + \delta t$.

$$\begin{aligned} r(t + \delta t) &= r(t) + v(t)\delta t + a(t)\frac{\delta t^2}{2} \\ r(t - \delta t) &= r(t) - v(t)\delta t + a(t)\frac{\delta t^2}{2} \end{aligned} \quad \text{(Equation 17)}$$

En additionnant les deux équations, on obtient alors la position au temps $t + \delta t$:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 \quad \text{(Equation 18)}$$

L'algorithme de Verlet n'utilise donc pas les vitesses pour calculer les nouvelles positions. C'est une méthode extrêmement stable, qui est réversible, qui ne nécessite pas d'importantes ressources de mémoire et procure un temps de calcul rapide.

1.2.b) Algorithme Leapfrog

Le principe de cet algorithme est de diviser le pas d'intégration par deux, cela afin d'être plus précis dans le calcul des nouvelles coordonnées. Les vitesses sont calculées au temps $t + \frac{1}{2}\delta t$ et utilisées pour calculer les positions au temps $t + \delta t$.

$$\begin{aligned} r(t) &= r\left(t + \frac{\delta t}{2} - \frac{\delta t}{2}\right) = r\left(t + \frac{\delta t}{2}\right) - v\left(t + \frac{\delta t}{2}\right)\frac{\delta t}{2} \\ r(t + \delta t) &= r\left(t + \frac{\delta t}{2} + \frac{\delta t}{2}\right) = r\left(t + \frac{\delta t}{2}\right) + v\left(t + \frac{\delta t}{2}\right)\frac{\delta t}{2} \end{aligned} \quad \text{(Equation 19)}$$

En soustrayant les deux équations on obtient alors l'expression de la position au temps $t + \delta t$:

$$r(t + \delta t) = r(t) + v\left(t + \frac{\delta t}{2}\right)\delta t \quad \text{(Equation 20)}$$

Les vitesses au temps $t + \frac{1}{2}\delta t$ sont calculées selon là partir des valeurs des vitesses au temps $t - \frac{1}{2}\delta t$ et des accélérations au temps t :

$$v\left(t + \frac{\delta t}{2}\right) = v\left(t - \frac{\delta t}{2}\right) + a(t)\delta t \quad \text{(Equation 21)}$$

De la même manière que l'algorithme de Verlet, l'algorithme Leapfrog présente les propriétés de réversibilité et de stabilité. De plus, il permet d'apporter de la précision dans le calcul des positions par l'utilisation de pas d'intégration plus petits mais peut entraîner en contrepartie une perte de la corrélation entre les positions et les vitesses, donc des divergences dans le calcul. Cet algorithme est celui que nous avons utilisé avec la suite de logiciel AMBER.

I.2.c) Pas d'intégration

Les calculs ne pouvant être effectués de manière continue, il faut intégrer l'équation du mouvement de manière régulière durant un intervalle de temps δt . Cette variable s'appelle le pas d'intégration. Si cet intervalle de temps est trop grand, le temps de simulation final sera très grand, mais l'accumulation d'erreurs sur l'intégration sera proportionnelle au nombre total d'étapes d'intégration de la loi de Newton. La mauvaise prédiction des nouvelles positions peut être telle que deux atomes peuvent occuper la même position au même moment. Si l'intervalle est trop petit, la précision sera grande sur chaque calcul et donc sur chaque nouvelle position, mais le nombre de calculs sera beaucoup plus important pour atteindre un temps de simulation donné, ce qui ne nous permettra pas d'échantillonner des événements d'intérêts nécessitant une simulation de longue durée. Nous devons donc trouver le bon équilibre entre précision et efficacité.

Si nous voulons pouvoir comprendre le comportement dynamique du système, le pas d'intégration doit être choisi de façon à ce que toute la physique du système puisse être captée. Cela implique de pouvoir prendre en compte le mouvement de tous les degrés de liberté du système (angles, liaisons...), notamment les vibrations de toutes les liaisons présentes dans le système. Sachant que la fréquence de vibration de liaison la plus rapide, celle de la liaison C-H, est de 1 fs, il faudrait prendre un pas d'intégration de 0,5 fs afin de pouvoir capter le plus fidèlement possible le comportement de toutes les liaisons du système. Dans le cas de systèmes protéiques, un pas de 0,5 fs impliquerait un nombre d'étape trop grand pour échantillonner des événements rares, qui peuvent intervenir au bout de plusieurs centaines de nanosecondes. Afin de pouvoir augmenter le pas d'intégration pour réaliser des simulations plus longues, nous approximations que les mouvements du système biologique ne sont pas liés aux vibrations des liaisons X-H (X étant un hétéroatome). Pour pouvoir faire ceci, l'algorithme SHAKE est utilisé afin de figer ces liaisons. De manière générale, le pas d'intégration que l'on choisi est de 2 fs, qui constitue un bon compromis entre un nombre d'étapes de calcul raisonnable pour obtenir des simulations de longues durées et bonne précision des étapes d'intégration.

I.2.d) Thermostat de Langevin

Lors de la préparation du système, nous avons vu que les étapes de minimisation du système, à partir de la structure cristallographique, était nécessaire pour atteindre le puits de potentiel de la surface d'énergie. Cependant, la minimisation ne prend pas en compte la notion

de température, alors qu'elle est nécessaire dans les simulations de dynamique moléculaire. Dans nos calculs, la température est gérée par le thermostat de Langevin. Le principe est d'insérer des atomes fantômes dans la simulation, qui induisent des collisions qui vont permettre la gestion de l'énergie cinétique des atomes, donc par conséquent de la température du système. La friction et la viscosité du milieu sont pris en compte dans ce type de simulation et sont contrôlés par le coefficient de friction γ .

I.2.e) Conditions périodiques et Maillage particulaire d'Ewald

Les simulations de dynamique moléculaire ont pour but de proposer un modèle permettant d'apporter à des grandeurs expérimentales une explication au niveau atomique. Pour ce faire, nous devons reproduire aussi fidèlement que possible les conditions biologiques. Cela signifie que la phase solvant doit être traitée de manière explicite car elle joue un rôle majeur dans la structuration des enzymes (molécules d'eau cristallographiques qui permettent la cohésion d'une structure tertiaire), mais également dans les interactions ligand-récepteur (par exemple des molécules d'eau pontantes). Afin de prendre en compte ces effets, nous pouvons utiliser les conditions périodiques. Le système biologique est placé dans une boîte de solvant qui est répliquée à l'infini dans les trois directions de l'espace, simulant ainsi un milieu biologique. Cela permet d'éviter les effets de bords, ainsi que la dispersion des molécules d'eau. Si une molécule sort de la boîte, elle réapparaît du côté opposé, permettant ainsi un nombre de particules constant dans la maille unitaire. Ces concepts sont schématisés sur la Figure 23.

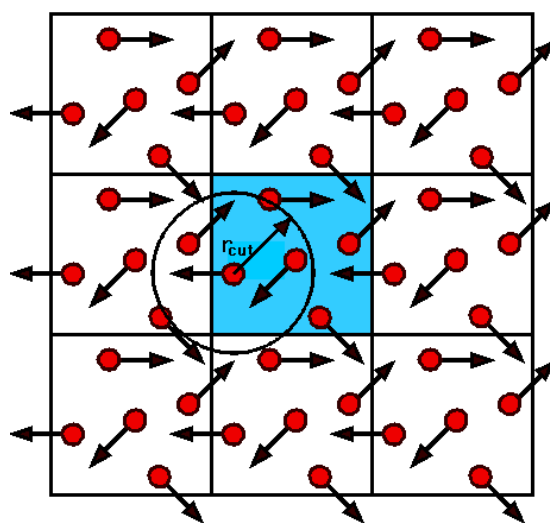


Figure 23: Représentation du fonctionnement des conditions limites périodiques.

Pour la considération des interactions interatomiques, et au vu du nombre d'atomes présents dans les systèmes qui ne cesse de s'accroître, le calcul des interactions électrostatiques doit être repensé si nous voulons optimiser les temps de calculs. Pour cela, on utilise la méthode du Maillage Particulaire d'Ewald (Particle Mesh Ewald, PME) qui différencie les interactions électrostatiques à longue et à courte distance[114] par l'utilisation de la sommation d'Ewald. A courte distance (définie comme inférieure au rayon de troncature r_{cut} , représenté sur la Figure 23), le calcul de l'interaction électrostatique est réalisé de manière classique par l'utilisation d'un potentiel de Coulomb. A longue distance (supérieure à r_{cut}), le calcul de cette interaction se fait par une transformée de Fourier du potentiel électrostatique (espace réciproque). Dans cet espace, chaque charge ponctuelle est désormais vue comme une distribution de charge de forme gaussienne, permettant ainsi une convergence plus rapide du calcul de l'interaction à longue distance.

$$E_{\text{électrostatique tot}} = E_{\text{courte distance}} + E_{\text{longue distance}} \quad (\text{Équation 22})$$

avec

$$E_{\text{courte distance}} = \sum_{j>i}^i \frac{q_i q_j}{\epsilon d_{ij}} \quad (\text{Équation 23})$$

et

$$E_{\text{longue distance}} = \sum_k \tilde{V}_{lr}(k) |\tilde{u}(k)|^2 \quad (\text{Équation 24})$$

$\tilde{V}_{lr}$ et $\tilde{u}$ représentent respectivement la transformée de Fourier du potentiel électrostatique et de la densité de charge. Le choix de ce calcul hybride permet d'optimiser au maximum le calcul de cette interaction, qui représente une part des plus chronophages dans un calcul. D'une façon générale, le rayon de troncature est fixé à 8 Angströms, ce qui constitue un bon compromis entre la description la plus précise possible de l'environnement proche et un maximum d'interactions traitées dans l'espace réciproque.

Toutes les méthodes présentées précédemment ont été appliquées pour la réalisation de ce travail par le biais de la suite de logiciel AMBER[115], et notamment la version AMBER12, pour les calculs sur GPU[116].

I.3) Simulation de type RAMD (Random Acceleration Molecular Dynamics)

En simulation de dynamique moléculaire, nous devons simuler des échelles de temps assez longues si nous voulons espérer échantillonner des "événements rares", tels que la sortie d'un ligand d'une enzyme par exemple. Afin d'augmenter la probabilité d'échantillonner de tels phénomènes, nous pouvons diminuer les barrières associées en introduisant un biais (accelerated MD, par exemple) ou en forçant le phénomène à se faire, en suivant un biais représentant une coordonnée de réaction (Umbrella Sampling). Cela implique de débiaiser la simulation en fin de calcul afin d'obtenir des résultats pouvant être exploités, ce qui peut être délicat selon la contrainte choisie.

Pour essayer d'introduire un biais le plus petit possible, Ludemann *et al.* [107] ont développé la méthode de REMD (Random Expulsion Molecular Dynamics), appelée par la suite RAMD (Random Acceleration Molecular Dynamics), qui permet d'introduire une force artificielle supplémentaire dans le calcul. La force est appliquée sur le centre de masse de la molécule choisie, et sa direction est choisie aléatoirement. Cette direction est maintenue pendant N pas de simulation, permettant ainsi de calculer de manière théorique la position que la molécule doit atteindre en un temps dt , connaissant sa vitesse moyenne.

Si le substrat rencontre une partie rigide de la cavité, et ne peut atteindre cette position, alors la vitesse dans cette direction diminuera, et une nouvelle direction du vecteur sera choisie de manière aléatoire. Si, au contraire, la distance prévue est atteinte, la direction sera maintenue pendant encore un temps dt . La simulation continue ainsi jusqu'à ce qu'un critère final (ici pris comme étant une distance entre le centre de masse du ligand et celui de la protéine appelée $r_{\max}$) sera atteint. Si la distance est assez élevée, des événements tels que la sortie du ligand hors d'une cavité enzymatique peuvent être échantillonnés, et ainsi permettre la détection de voies d'accès au site actifs des composés.

La force aléatoire appliquée pendant le calcul est définie simplement par l'équation suivante :

$$\vec{F} = k\vec{r}_0 \quad (\text{Équation 25})$$

dans laquelle F est la force supplémentaire appliquée, k la constante de force et r_0 un vecteur unitaire de direction aléatoire. Trois paramètres doivent ainsi être ajustés pendant la

simulation : la constante de force k , qui conserve la même valeur pendant la totalité de la simulation, le nombre d'étapes N au bout desquelles la direction doit être réévaluée et $r_{\max}$ qui est la distance maximale déterminant la fin de la simulation.

I.4) Potentiel de force moyenne et Umbrella Sampling

Le potentiel de force moyenne a été développé par Kirkwood en 1935.[117] Pour un système composé de N atomes, cette grandeur est la résultante de la moyenne des forces de toutes les configurations $n+1 \dots N$ particules agissant sur la particule j en gardant l'ensemble des particules $i=1..n$ fixes :

$$-\nabla_j w^n = \frac{\int \exp(-\beta V) (-\nabla_j V) d\mathbf{q}_{n+1} \dots d\mathbf{q}_N}{\int \exp(-\beta V) d\mathbf{q}_{n+1} \dots d\mathbf{q}_N}, j=1, 2, \dots, n \quad (\text{Equation 26})$$

Avec, $\beta = 1/k_b T$, $-\nabla_j w^n$ la moyenne des forces et w^n le potentiel de force moyenne.

Le protocole d'Umbrella Sampling (US) vise à forcer le système étudié à échantillonner les états de haute énergie et donc de faible probabilité en introduisant un terme énergétique supplémentaire ou biais. Cette méthode a été développée par Torrie et Valleau au cours des années 70.[118, 119] Sa mise en œuvre a été étudiée par Roux dans les années 90.[120]

La coordonnée de réaction ξ est divisée en plusieurs fenêtres au sein desquelles le biais est un terme énergétique w_i ajouté selon :

$$E^b(r) = E^u(r) + w_i(\xi) \quad (\text{Equation 27})$$

Le terme portant l'exposant b se réfère au terme biaisé de la simulation et ceux portant u au terme non biaisé.

En divisant la coordonnée de réaction en plusieurs fenêtres, il est alors possible de réaliser ces simulations indépendamment les unes des autres. Ce protocole autorise alors d'échantillonner toute la coordonnée de réaction rapidement en utilisant des grilles de calcul.

Afin d'obtenir l'énergie libre associée à la fenêtre i , $A_i(\xi)$, il est nécessaire de connaître la distribution non biaisée du système pour cette fenêtre :

$$P_i^u(\xi) = \frac{\int \exp[-\beta E(r)] \delta[\xi'(r) - \xi] d^N r}{\int \exp[-\beta E(r)] d^N r} \quad (\text{Equation 28})$$

Dans le cadre de l'hypothèse ergodique, la distribution biaisée selon la coordonnée de réaction est donnée par :

$$P_i^b(\xi) = \frac{\int \exp\{-\beta[E(r) + w_i(\xi'(r))]\} \delta[\xi'(r) - \xi] d^N r}{\int \exp\{-\beta[E(r) + w_i(\xi'(r))]\} d^N r} \quad (\text{Equation 29})$$

Le biais dépend de uniquement de la coordonnée de réaction, et l'intégration du numérateur est réalisé par rapport à tous les degrés de liberté excepté la coordonnée de réaction, on peut alors écrire :

$$P_i^b(\xi) = \exp[-\beta w_i(\xi)] \times \frac{\int \exp[-\beta E(r)] \delta[\xi'(r) - \xi] d^N r}{\int \exp\{-\beta[E(r) + w_i(\xi'(r))]\} d^N r} \quad (\text{Equation 30})$$

En utilisant l'expression de la probabilité non biaisée P^u il s'ensuit que :

$$\begin{aligned} P_i^u(\xi) &= P_i^b(\xi) \exp[-\beta w_i(\xi)] \times \frac{\int \exp\{-\beta[E(r) + w_i(\xi(r))]\} d^N r}{\int \exp[-\beta E(r)] d^N r} \\ P_i^u(\xi) &= P_i^b(\xi) \exp[-\beta w_i(\xi)] \times \frac{\int \exp[-\beta E(r)] \exp\{-\beta w_i[\xi(\vec{r})]\} d^N r}{\int \exp[-\beta E(r)] d^N r} \\ P_i^u(\xi) &= P_i^b(\xi) \exp[-\beta w_i(\xi)] \langle \exp[\beta w_i(\xi)] \rangle \end{aligned} \quad (\text{Equation 31})$$

Cette dernière équation permet de directement calculer $A_i(\xi)$:

$$A_i(\xi) = -\frac{1}{\beta} \ln P_i^b(\xi) - w_i(\xi) + F_i \quad (\text{Equation 32})$$

$$\text{Avec } F_i = -\frac{1}{\beta} \ln \langle \exp[-\beta w_i(\xi)] \rangle \quad (\text{Equation 33})$$

La grandeur $A_i(\xi)$ est exacte si l'échantillonnage dans la fenêtre est suffisant. Si le calcul est réalisé dans une fenêtre unique $A(\xi)$ est définie à une constante près et F_i peut-être choisie arbitrairement. Par contre lorsqu'il est nécessaire de combiner plusieurs fenêtres afin de reconstruire le profil de A le long de la coordonnée ξ , la quantité F_i doit être déterminée.

$$\begin{aligned} \exp(-\beta F_i) &= \langle \exp[-\beta w_i(\xi)] \rangle \\ \exp(-\beta F_i) &= \int P^u(\xi) \exp[-\beta w_i(\xi)] d\xi \quad (\text{Equation 34}) \\ \exp(-\beta F_i) &= \int \exp\{-\beta[A(\xi) + w_i(\xi)]\} d\xi \end{aligned}$$

Avec P^u la distribution non biaisée totale. L'échantillonnage ne permet pas d'obtenir directement les grandeurs F_i . Il est alors nécessaire de recourir à des méthodes de résolutions numériques comme WHAM (Weighted Histogram Analysis Method) afin de retrouver le potentiel de force moyenne associé à notre simulation[121, 122].

L'introduction du biais dans la simulation a pour but d'obtenir une distribution uniforme de l'échantillonnage le long de la coordonnée de réaction. Idéalement, il serait nécessaire de satisfaire la relation suivante : $w_{idéal} = -A(\xi)$. Ce qui est impossible en pratique puisque $A(\xi)$ est la grandeur recherchée.

Pour réaliser un échantillonnage de toutes les valeurs de ξ , il est possible de découper la coordonnée de réaction en plusieurs fenêtres. Dans chacune des fenêtres le biais est introduit pour garder le système proche de la valeur ξ_{ref} . Pour les calculs effectués à l'aide du logiciel AMBER, le biais est une fonction harmonique de force K :

$$w_i(\xi) = \frac{K}{2} (\xi - \xi_{ref})^2 \quad \text{(Equation 35)}$$

Il a été montré qu'une augmentation du nombre de fenêtre de calcul entraînait une diminution de l'erreur statistique sur la valeur de $A(\xi)$ [123]. La valeur de la constante de force K a une influence capitale sur le résultat final. En effet le biais doit être suffisamment important pour permettre au système de franchir les barrières énergétiques. Une valeur trop grande de cette constante de force induira un échantillonnage très resserré autour de la valeur ξ_{ref} et donc nécessitera de multiplier les fenêtres pour échantillonner l'ensemble des valeurs de ξ .

I.5) Méthode de docking protéine protéine : protocole ATTRACT.

Le protocole ATTRACT [104, 106] se base sur l'utilisation d'un modèle gros grain pour définir nos protéines. Concrètement, cela implique qu'un atome ou un groupement d'atomes sera remplacé par un grain qui va reproduire les propriétés associées à l'ensemble. Le nombre de particules considérées baisse alors, induisant une forte diminution des calculs d'interactions (termes liés et non liés) à traiter. Cela nous permettra d'échantillonner de manière plus rapide les différentes possibilités d'interaction entre protéines, afin d'être le plus exhaustif possible. Le concept du modèle gros grain est représenté sur la Figure 24.

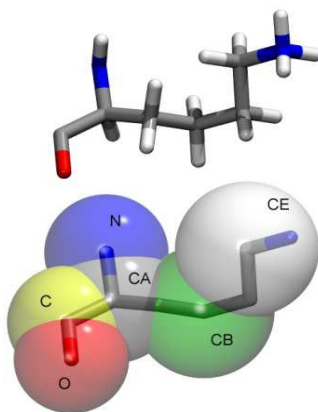


Figure 24: Schéma de la représentation d'un acide aminé en gros grain.

La première étape est de définir quelle protéine sera le récepteur et le ligand. Autour de la protéine que l'on considérera comme le récepteur, on définit une grille dans laquelle chaque point sera une position de départ pour notre enzyme ligand. Et, de manière à être encore plus complet dans l'échantillonnage des différentes interactions possibles, 258 rotamères, c'est-à-dire différentes orientations autour du point de départ, seront utilisés pour chaque point de la grille. Chaque pose subit un calcul de minimisation d'énergie afin de déterminer une structure de complexe protéine-protéine ainsi que l'énergie d'interaction associée au moyen d'une fonction de score. Cette fonction est constituée d'une part électrostatique, représentée par un potentiel de Coulomb, et d'une composante d'interaction de type van der Waals, représentée par un potentiel Lennard-Jones modifié. En effet, pour prendre en compte la modification de la taille des particules, le terme répulsif est légèrement corrigé afin d'avoir moins de pénalité d'énergie sur un rapprochement entre deux grains. L'avantage important que propose ATTRACT est l'utilisation des différents grains et de leurs propriétés (il s'agit d'un "knowledge-based potential"). Ainsi, lorsque nous savons que l'interaction est défavorable, le potentiel est modifié (voir U^{rep} , cf Figure 25) de façon à écarter rapidement les solutions les moins bonnes.

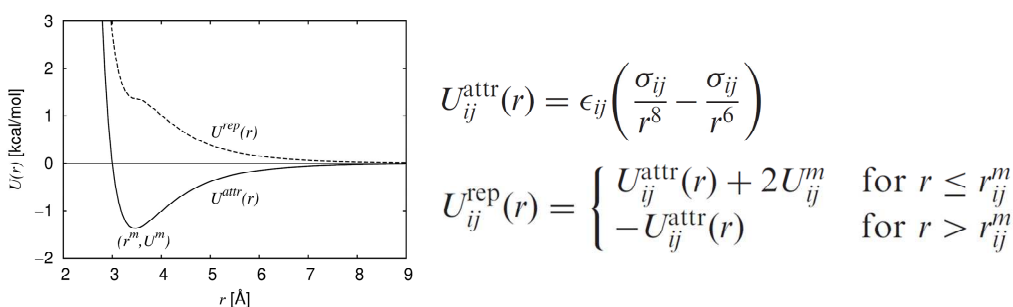


Figure 25 : Potentiel de Lennard-Jones modifié pour le protocole ATTRACT.

Le calcul des différentes interactions donne des milliers de possibilités, qu'il est possible de mettre en commun pour l'analyse. La Figure 26 résume les différentes étapes du protocole ATTRACT. Ce dernier peut être amélioré en prenant en compte les pénalités de désolvatation des différents acides aminés en surface des deux protéines dans le calcul. Plus le grain aura une pénalité de désolvatation importante, plus son poids devrait être grand dans l'interaction protéine-protéine[105].

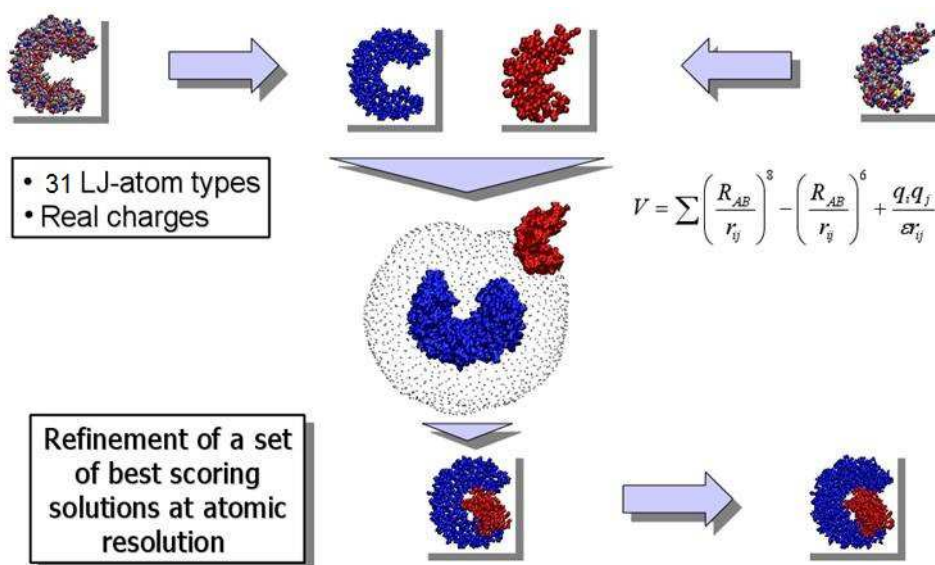


Figure 26 : Schéma récapitulatif du protocole ATTRACT.

I.6) Protocole MM-GBSA

La méthode MM-GBSA (Molecular Modeling-Generalised Born/ Surface Area) est une méthode end-point (c'est à dire qu'elle ne dépend que de l'évaluation des états initiaux et finaux, et non du chemin suivi) qui permet de prédire l'énergie d'interaction entre ligand et récepteur[124]. L'énergie libre de Gibbs de fixation d'une molécule avec son récepteur est donnée par la relation :

$$\Delta G_{bind} = G_{complexe} - (G_{récepteur} + G_{ligand}) \quad (\text{Équation 36})$$

Chaque composante peut être approximée selon l'équation suivante:

$$G = E_{MM} + G_{solv} + H_{rot/trans} - TS \quad (\text{Équation 37})$$

E_{MM} correspond à la somme des contributions de l'énergie calculée par le potentiel de mécanique moléculaire (somme des termes liés et non liés). G_{solv} correspond à l'enthalpie libre de solvation des différents protagonistes tandis que $H_{rot/trans}$ représente la contribution

énergétique des 6 degrés de liberté de rotation et de translation, tirée de la théorie cinétique des gaz ($3kT = 1,8 \text{ kcal.mol}^{-1}$ à 300 K). Enfin, le terme TS représente la contribution entropique à l'énergie libre, et est calculé à partir de l'analyse des modes normaux de vibration.

L'estimation de l'énergie de solvation est divisée en deux contributions : polaire et non polaire. La contribution non polaire associée à cette enthalpie libre est calculée par la définition de la surface accessible au solvant, par interaction de van der Waals.

$$G_{SA} = \gamma A + B \quad (\text{Équation 38})$$

A correspond à la surface du système, obtenue en faisant rouler une bille représentant la surface d'une molécule de solvant sur la molécule considérée. B et γ sont des constantes empiriques. La partie polaire est la plus difficile à calculer. Deux méthodes peuvent être utilisées : la résolution de l'équation de Poisson-Boltzmann par méthode itérative, ou l'approximation de la contribution par l'équation de "Born Généralisé" :

$$G_{sol} = \left(1 - \frac{1}{\epsilon}\right) \sum_{i < j}^i \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j} e^{-\left(\frac{r_{ij}^2}{4\alpha_i \alpha_j}\right)}} \quad (\text{Équation 39})$$

Lorsque les ligands possèdent une structure similaire, le terme TS peut être négligé (car équivalent dans chacun des cas). Ainsi seul le calcul de l'enthalpie ΔH est réalisé et est assez pertinent pour réaliser un classement selon la valeur de l'énergie d'interaction.

II) Mécanique quantique

Développée au début du XXème siècle par un ensemble de scientifique (Heisenberg, de Broglie, Einstein, Bohr,...), la mécanique quantique est à ce jour la théorie la plus aboutie sur l'infiniment petit. Son application à la chimie est basée sur l'équation de Schrödinger, ici écrite dans l'approximation de Born et Oppenheimer (les noyaux atomiques sont considérés fixes comparés aux mouvements des électrons), pour un système de N électrons et A noyaux :

$$H. \Psi = E. \Psi \quad (\text{Équation 40})$$

$$\hat{H} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} - \sum_{A=1}^M \sum_{i=1}^N \frac{Z_A}{r_{iA}}$$

(Équation 41)

Dans l' (Équation 40, H désigne l'opérateur Hamiltonien, E l'énergie électronique totale du système considéré, et Ψ la fonction d'onde associée, décrivant de manière exacte le comportement des électrons. Le premier terme de l'hamiltonien correspond à l'hamiltonien mono électronique, calculé pour chaque électron du système. Le second terme correspond à l'hamiltonien bi-électronique, qui permet de déterminer les interactions entre électrons. Le dernier terme représente l'interaction électrostatique entre noyaux et électrons. Malheureusement, le terme bi-électronique ne peut être calculé de manière exacte dans la majorité des cas, ce qui rend l'équation impossible à résoudre pour les systèmes non-hydrogènoïdes. Ainsi, nous avons recours à des approximations pour pouvoir résoudre cette équation : il s'agit des méthodes *ab initio* (HF, post-HF) ou les méthodes basées sur la densité électronique (DFT).

II.1) Méthode Hartree-Fock

La méthode Hartree-Fock[125, 126] permet de résoudre approximativement l'équation de Schrödinger, en assimilant les intégrales bi-électroniques en une intégrale d'interaction coulombienne (notée J) et une intégrale d'échange (notée K). On définit ainsi le nouvel opérateur F (opérateur de Fock) :

$$F. \varphi = E. \varphi \quad (\text{Équation 42})$$

$$\hat{F} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \sum_{i=1}^N \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N J_{ij} - K_{ij}$$

(Équation 43)

Les formes des intégrales sont connues, donc l'équation peut être résolue. Cependant, la fonction d'onde dépend directement des paramètres J et K, qui dépendent aussi de la fonction d'onde directement. Ainsi, une résolution linéaire de l'équation est à exclure. La méthode utilisée est donc celle du *self-consistent field* (champ auto-cohérent), qui consiste en la résolution de l'équation à partir d'un jeu d'orbitales de départ, appelée *basis set*. Ce jeu de départ permet de calculer une première fois les paramètres J et K, qui sont réutilisés pour

recalculer les orbitales, et ainsi de suite jusqu'à convergence des divers paramètres (énergie, RMSd...).

Ce jeu d'orbitales de départ peut être plus ou moins développé selon nos attentes. Plus la base est grande, plus le calcul sera précis, et par conséquent plus long. Les orbitales sont soit représentées par des orbitales de type Slater (STO), soit par des fonctions gaussiennes ($e^{-\alpha r^2}$) représentant de manière différente les orbitales de cœur et les orbitales de valence.

Cette approximation et la méthode du champ auto-cohérent permettent au protocole Hartree-Fock de résoudre de manière rapide l'équation de Schrödinger. Cependant, les approximations faites précédemment font que l'énergie calculée sera toujours supérieure à l'énergie des solutions exactes (la différence est appelée énergie de corrélation). D'un point de vue physique, la position d'un électron à un temps donné n'est en réalité pas indépendante des autres. On dit alors que les électrons sont corrélés, et ceci peut induire une différence très importante entre l'énergie HF et l'énergie exacte d'un système. Certaines méthodes *ab initio*, appelées post-Hartree-Fock, permettent une meilleure estimation de l'énergie, et donc de se rapprocher au maximum de l'énergie des solutions exactes, mais sont plus coûteuses en temps de calcul et utilisent beaucoup plus de ressources informatiques (Interaction de Configuration, méthode perturbative Møller-Plesset, coupled-cluster...).

II.2) Méthode de la fonctionnelle de la densité (DFT)

Cette méthode constitue une bonne alternative aux méthodes *ab initio*. L'idée de base est d'utiliser la densité électronique (mesurable par rayons X par exemple) de notre système comme point de départ afin de retrouver l'énergie exacte du système, et donc la fonction d'onde associée. Les premiers théorèmes ont été développés par Kohn et Hohenberg [127]. Ils ont réussi à démontrer que :

- Pour un potentiel externe donné V_{ext} , un système à particules interactives possédera une seule et unique densité électronique.
- Une fonctionnelle universelle de l'énergie peut être définie en terme de densité. L'énergie exacte d'un système correspond au minimum global de la fonctionnelle de densité.

En 1965, Kohn et Sham [128] développent les équations auto-cohérentes pour décrire l'énergie d'un système en fonction de sa densité, en présence d'un potentiel externe.

$$E[\rho(r)] = T[\rho(r)] + \int \rho(r)v(r)dr + V_{ee}[\rho(r)] + E_{xc}[\rho(r)]$$

(Équation 44)

Dans laquelle le terme $T[\rho(r)]$ représente l'énergie cinétique d'un système de particules sans interaction et possédant la même densité que le système réel. Les deux termes suivants décrivent respectivement l'énergie d'attraction électron-noyau et le terme répulsif bi-électronique électron-électron. Le dernier terme, E_{xc} , est défini comme l'énergie d'échange-corrélation. Il caractérise les contributions d'échange et de corrélation (directement pris en compte en DFT, contrairement aux méthodes HF). Malheureusement, ce dernier terme est inconnu, et il convient de l'approximer afin de pouvoir résoudre l'équation de Kohn-Sham.

Différents types de fonctionnelles ont été développés afin de représenter au mieux les contributions d'échange et de corrélation. Nous pouvons citer :

- Les fonctionnelles LDA (*Local density approximation*) : basées sur un gaz uniforme d'électron, elles négligent toute inhomogénéité du système. L'approximation est correcte si la densité électronique varie très lentement, mais ces fonctionnelles ne sont plus utilisées de nos jours.
- Les fonctionnelles GGA (*Generalized gradient approximation*) : elles introduisent dans leur expression un gradient de densité, ce qui permet de considérer l'inhomogénéité de la densité électronique de la plupart des systèmes. Les fonctionnelles de ce type sont : Becke88, PW91, P86 et LYP.
- Les fonctionnelles méta-GGA font intervenir le laplacien de la densité électronique en place du gradient, afin d'augmenter la précision de la méthode. Nous pouvons citer B95 comme fonctionnelle de ce type.
- Les fonctionnelles hybrides méta-GGA : dans ce cas, l'énergie d'échange est calculée avec un pourcentage de l'énergie Hartree-Fock (qui calcule de manière exacte la contribution d'échange), tandis que l'énergie de corrélation est calculée par DFT. Ces fonctionnelles, dont la plus connues est B3LYP sont les plus robustes et les plus fiables. Notamment, de nouvelles fonctionnelles de ce type ont été développées par Zhao et Truhlar [129] (M06-2X, M06-HF...) et montrent une grande précision dans le calcul de l'énergie.

Chapitre 1 : Préparation des systèmes

I) Préparation des structures enzymatiques

I.1) DFR

Concernant l'enzyme dihydroflavonol-4-réductase (DFR), deux structures cristallographiques sont disponibles sur le site de la Protein Data Bank sous les identifiants 2IOD et 2C29 [68]. Hormis les cinq premiers acides aminés, ces structures sont complètes. La première est un complexe impliquant l'enzyme, le cofacteur et deux substrats. Dans leurs publications, les auteurs de cette structure rapportent que la géométrie du site actif est fortement perturbée dans cette configuration. Quand à la seconde, il s'agit d'un complexe ternaire impliquant l'enzyme, le substrat naturel de l'enzyme (la molécule de dihydroquercétine, DHQ) ainsi que le cofacteur NADP⁺ provenant de l'oxydation de NADPH. Cette structure est un complexe abortif, ce qui signifie qu'en l'absence de régénération du cofacteur, i.e réduction du NADP⁺ en NADPH, la réaction de réduction du substrat ne peut avoir lieu. Malgré tout, il est admis que les sites de reconnaissances du cofacteur et du substrat ne sont que peu modifiés. Cette structure a été retenue pour les études menées sur DFR, et la molécule de NADP⁺ sera remplacée par le cofacteur non oxydé NADPH(ajout d'un hydrure).

I.2) LAR

Dans la PDB, cette enzyme apparaît sous 3 formes : apoenzyme (PDB ID : 3I5M), enzyme en interaction avec le cofacteur NADPH (PDB ID : 3I6I et 3I6Q), et en tant que complexe ternaire, l'enzyme étant co-cristallisée avec le cofacteur et le substrat (PDB ID 3I52) [74]. Etant données les fortes réorganisations possibles du site actif lors de l'introduction du substrat, c'est cette dernière structure qui est retenue pour nos études. Cependant, la structure tridimensionnelle de nombreuses zones de cette enzyme n'ont pu être élucidées expérimentalement. Les positions des 10 premiers et des 30 derniers acides aminés ne sont présentes dans aucune des structures PDB. Cela provient du protocole de production hétérologue de cette protéine qui a entraîné des coupures à ces positions. Notons également que deux autres zones (42-54 et 161-166) n'apparaissent pas. Il est fort probable que, ces zones étant au cœur de la séquence, ce problème provienne d'une labilité trop importante de cette portion de la structure. Cela étant, la définition de ces zones manquantes est nécessaire, pour une description correcte du système, sachant que la portion 42-54 manquante est directement en interaction avec le cofacteur NADPH. Cette portion présente ainsi une

arginine et une lysine (R42 et K50) qui interagissent de manière forte avec le groupement phosphate du cofacteur. Dans ce dernier cas, les résidus impliqués dans ces zones ont été reconstruits à l'aide de la structure de l'apoenzyme (PDB ID : 3I5M) dans laquelle la position spatiale de ces acides aminés a pu être identifiée.

Les parties N-terminale et C-terminale de l'enzyme sont manquantes dans toutes les structures obtenues à ce jour. Nous avons reconstruit ces parties et avons appliqué plusieurs protocoles afin d'aboutir à une structure pour laquelle le positionnement de ces acides aminés correspond à une structure stable dans le temps. Dans un premier temps, nous avons reconstruits ces parties de séquence sans structuration particulière et nous avons appliqué plusieurs tests de simulations de dynamique moléculaire, de type "brut force" ou REMD afin d'échantillonner un repliement naturel. Ces simulations n'ont mené à aucun résultats probants. Dans un second temps, à l'aide d'une prédiction de structure secondaire réalisée avec le serveur PSI-PREDD [130], nous avons pu estimer que la partie en N-terminale ne présentait aucune structuration particulière, tandis que la partie en C-terminale possédait 5 acides aminés qui seraient potentiellement structurés en hélice alpha. La prédiction est présentée sur la Figure 27.

D'après cette prédiction, les résidus 325 à 330 sont structurés en hélice alpha. Cela implique que cette hélice doit probablement interagir avec le reste de la structure protéique. Nous pouvons considérer que nous nous trouvons ici face à une recherche de la meilleure interaction entre deux chaînes polypeptidiques structurées. Pour cela, nous avons utilisé des méthodes de docking protéine-protéine, au moyen du protocole ATTRACT (Page 74, partie méthodologie). Néanmoins, cette recherche est ici contrainte par le fait que ce positionnement de l'hélice n'est pas indépendant de la structure protéique. En effet, celle-ci est reliée au dernier acide aminé de la structure résolue par un enchaînement de 7 acides aminés pour lesquels aucune structuration tridimensionnelle particulière n'a été prédite. Nous pouvons donc considérer ces 7 acides aminés comme un "linker" qui va définir la zone possible d'interaction enzyme/hélice. La définition de la grille définissant les positions de départ du docking a été ajustée afin de contraindre nos recherches par rapport à cette contrainte. La grille a ainsi été limitée à une distance maximale de 30 Angströms autour du résidu 317, comme représenté sur la Figure 28.



Figure 27: Prédiction de la structure secondaire de LAR par le serveur PSI-PRED. Sont encadrées en rouge les zones manquantes dont nous voulons prédire la structure secondaire.

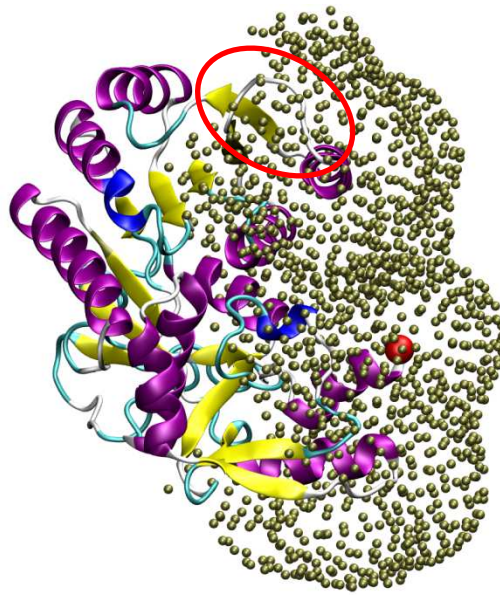


Figure 28 : Grille définie pour le docking du peptide (325-330) sur la structure cristallisée de LAR. En rouge se situe l'extrémité du résidu 317, le dernier cristallisé, et les points dorés correspondent aux différents points de la grille. Enfin, le cercle rouge montre les résidus 42 à 54 qui n'étaient pas présents dans la structure cristallographique de 3I52.

Une première campagne de calculs de docking a été réalisée et a mené à un millier de structures enzyme/hélice en interaction. Trois positions particulières ont retenu notre attention. En effet, la plupart des positions n'étaient pas compatibles avec la liaison entre le résidu 317 et l'hélice, via les sept acides aminés intermédiaires. La boucle reliant les deux parties était en effet trop tendue ou s'interpénétrait de manière trop importante avec la structure. Afin de valider ces possibilités, il a été nécessaire de reconstruire les parties encore manquantes, à savoir les 7 acides aminés constituant le "linker" et les derniers acides aminés des parties N-terminale et C-terminale. Ceci a été réalisé au moyen du logiciel MODELLER [103].

La grille englobe les résidus 42 à 54 (voir Figure 28) et nous avons observé que deux des poses retenues montrent une interaction avec cette zone. Il s'avère que cette partie, manquante dans la structure initiale et reconstruite grâce à MODELLER, interagit de manière importante avec le cofacteur. Nous avons donc décidé de relaxer dans un premier temps cette portion d'enzyme, afin de pouvoir obtenir de potentielles interactions hélice-enzyme pertinentes avec cette partie de la séquence. Afin de pouvoir considérer toutes les possibilités, la zone 42-54 sera équilibrée avec et sans le cofacteur en interaction.

Les structures, avec et sans la molécule de NADPH, ont donc été relaxées par simulation de dynamique moléculaire (100 ns), puis le docking de l'hélice a été de nouveau effectué sur ces deux systèmes. Dans les deux cas, trois poses ont pu être sélectionnées. Parmi celles-ci, il existe une pose en commun entre le système LAR-NADPH et l'apoenzyme, qui correspond de surcroît à la troisième pose isolée dans le premier calcul de docking réalisé, sans relaxation de la structure. Cette dernière est représentée sur la Figure 29 et interagit avec deux hélices alpha de l'enzyme.

Les six poses obtenues ont été soumises à des simulations de dynamique moléculaire. Les cinq autres poses (qui diffèrent par l'orientation de l'hélice et de l'emplacement de l'interaction) ont tendance à être instables sur des simulations de durée assez courte (100ns ou moins). Elles évoluent vers une structure non conforme, soit vers la position représentée sur la Figure 29. La position interagissant avec les deux hélices présente, quant à elle, une grande stabilité sur une simulation d'une durée de 330 ns. Et les critères d'énergie obtenus par méthode MM-GBSA (protocole décrit en partie méthodologie) montrent également que ce placement est de loin le plus stable par rapport aux autres. Nous avons donc choisi de retenir cette position de l'hélice. La structure entière de LAR a donc pu ainsi être reconstruite, et cette dernière est représentée sur la Figure 29.

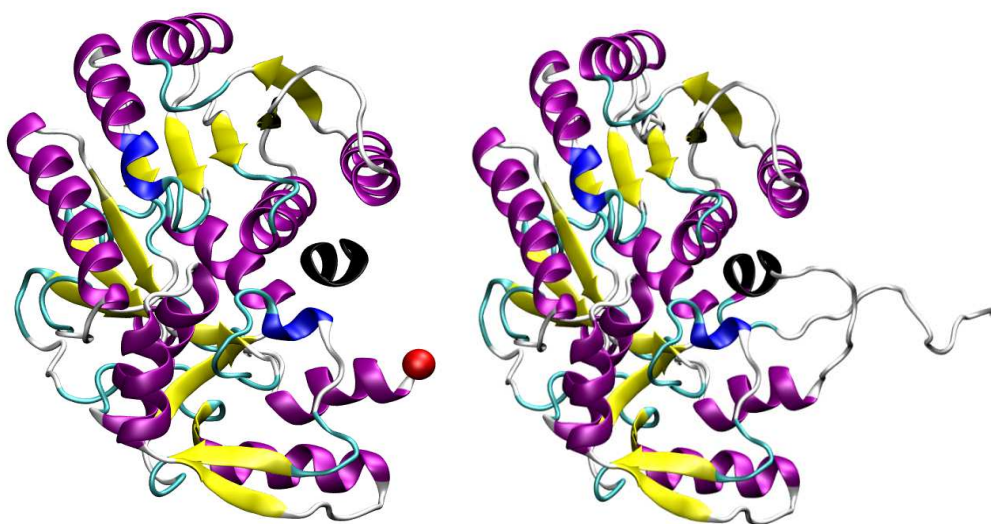


Figure 29 : à gauche, interaction entre l'hélice (résidu 325-330), représentée en noir, et la structure de LAR, après relaxation. En rouge est représenté l'atome de carbone CA du squelette du résidu 317. A droite, structure complète de LAR reconstruite avec cette position de l'hélice 325-330.

I.3) F3'H

L'enzyme F3'H appartient à la famille des cytochromes P450, une des familles d'enzymes les plus étudiées. Les membres de cette famille se caractérisent par une homologie très importante au niveau de leur structure tertiaire, notamment au niveau du nombre d'hélices alpha. Cependant, la structure cristallographique de F3'H n'a pas été résolue. Nous devons donc l'obtenir par un protocole de reconstruction par homologie. L'alignement de séquence et le calcul du pourcentage d'homologie représente un critère de sélection des modèles à utiliser pour la reconstruction. Un alignement préliminaire réalisé sur le serveur BLAST montre que quatre enzymes, représentées par les numéros de PDB 2HI4, 4IV8 et 3QZ1, présentent une séquence primaire très similaire à celle de F3'H. En effet, les pourcentages d'homologie (c'est à dire des résidus présentant les mêmes caractéristiques physico-chimiques) et d'identité des acides aminés de chaque enzymes avec F3'H sont respectivement de 91%/30%, 92%/30% et 93%/29%. De plus, les trois modèles déjà cristallisés ont une structure tridimensionnelle très similaire. Il est à noter que dans les trois cas, la structure de l'hélice transmembranaire n'est pas présente. En effet, les expérimentateurs doivent faire en sorte de ne pas exprimer cette portion, ou de la couper afin de permettre la cristallisation de l'enzyme. Or cette partie est extrêmement importante de par son caractère transmembranaire. Cette hélice serait l'ancre permettant au métabolite de se former autour de F3'H. Nous devons donc faire en sorte de proposer un modèle pertinent de cette portion de séquence. Dans une première approche, nous avons donc aligné l'hélice transmembranaire en N-terminale seule, afin de trouver un autre modèle pour sa reconstruction. La structure représentée par le fichier PDB 3PMN a ainsi été choisie. La Figure 30 présente de manière plus précise l'alignement des quatre séquences avec celle de F3'H, réalisé avec le logiciel Jalview [131].

Le logiciel MODELLER a été utilisé de concert avec l'alignement pour proposer un premier modèle de la structure de F3'H. Les analyses de structures (de type Ramachandran) ont été réalisées par le logiciel PROCHECK [132] pour évaluer la qualité du modèle. Il est apparu que de nombreux résidus se situaient dans des zones non favorables de la carte de Ramachandran, impliquant que les angles phi et psi, caractérisant la structure de l'enzyme, ne possédaient pas des valeurs optimales. De plus, l'hélice alpha en position N-terminale est mal représentée, avec une structuration uniquement partielle des 25 acides aminés impliqués dans sa formation. Cela est problématique, sachant que la structuration en hélice est importante si nous voulons avoir une bonne description des interactions entre la partie N-terminale et la membrane.

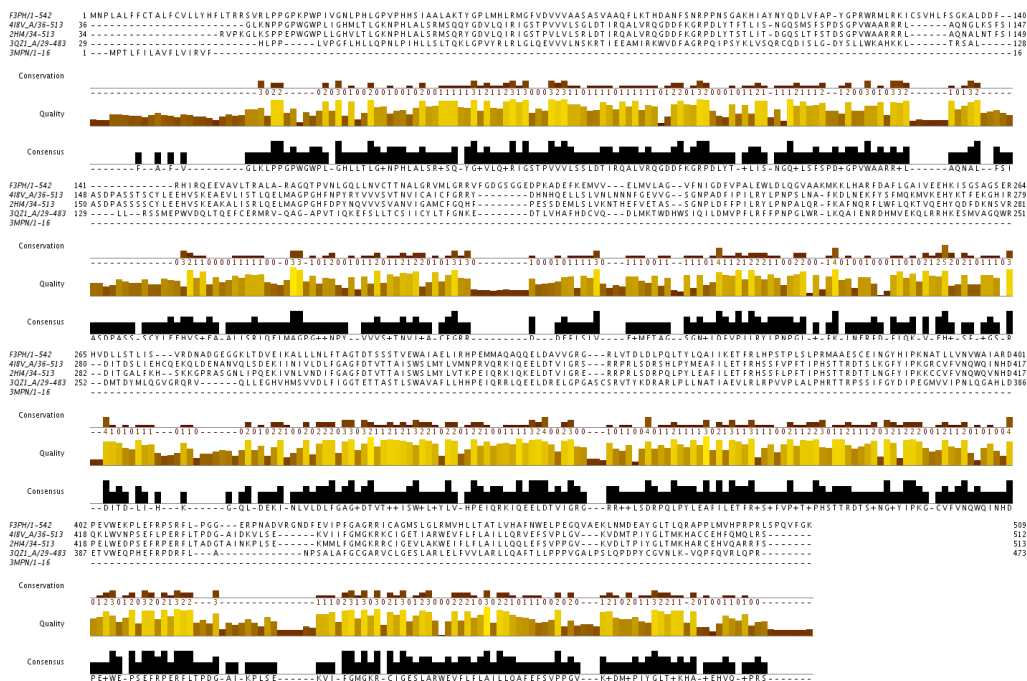


Figure 30 : Alignement de la séquence de F3'H avec celle des quatre modèles choisis.

Ainsi, une analyse de prédiction de structure secondaire (réalisée à nouveau avec le serveur PSI-PREDD) nous propose une structuration des 25 premiers acides aminés en hélice alpha. Nous avons donc décidé d'imposer la structuration en hélice alpha à ces résidus avec l'aide de MODELLER, en éliminant la structure représentée par le fichier 3MPN de l'alignement. Le deuxième modèle de la structure de F3'H obtenu est représenté sur la Figure 31.

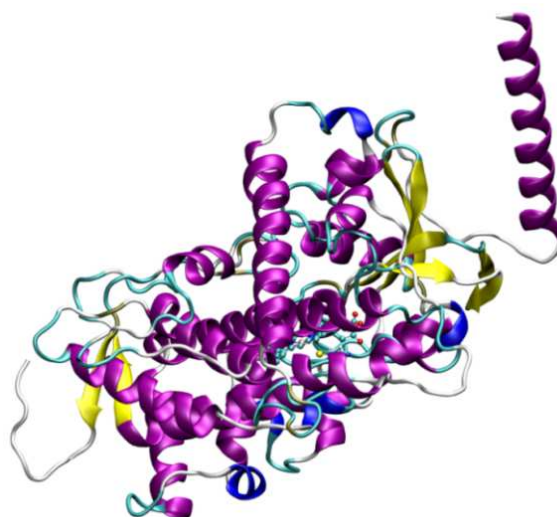


Figure 31: Structure de l'enzyme F3'H obtenue par reconstruction par homologie. En *balls and sticks* sont représentés le groupement hème, cofacteur de la réaction, et la cystéine qui interagit directement avec lui.

Les analyses données par Procheck sur ce modèle montrent que 99% des résidus de l'enzyme se situent dans les zones favorables de la carte de Ramachandran. Cela atteste de la bonne qualité du modèle reconstruit, nous avons donc choisi cette structure comme point de départ de nos simulations concernant F3'H. Le groupement hème a été placé à l'intérieur de l'enzyme grâce à la superposition de l'enzyme avec la structure la plus homologue (ici 3QZ1), comme représenté sur la Figure 31 et la Figure 32. Les interactions importantes entre l'hème et l'enzyme déjà identifiées (interaction des deux groupements carboxylates de l'hème avec quatre arginines) dans le fichier 3QZ1 sont retrouvées lors du placement par superposition (R95, R124, R372, et R442). La cystéine permettant la coordination de l'atome de fer à l'enzyme (distance initiale Fer-Soufre = 3,67 Å) a été identifiée comme étant la cystéine 444. La structure ainsi obtenue est celle que nous avons gardé pour les simulations de dynamique moléculaire.

II) Paramètres du champ de force

II.1) Représentation des enzymes

Le champ de force ff03.r1 [111,112] de la suite AMBER a été utilisé pour modéliser les atomes de nos différentes enzymes. Afin de reproduire les conditions biologiques, les différents résidus des trois protéines ont été protonés au pH 6,5 grâce au serveur H++ [133]. Un résumé des différents états de protonation des résidus d'intérêts (c'est-à-dire ceux impliqués dans les interactions entre la protéines et les différents partenaires biologiques, et ceux impliqués dans les réactions enzymatiques) pour chaque enzyme est présenté dans le Tableau 1.

On pourra noter dans le cas de LAR que la lysine catalytique, identifiée comme étant K140 par l'expérience, est considérée comme étant protonée à ce pH et possède un pKa encore très élevé. Ce résultat va à l'encontre des observations de Mauge *et al.* qui basait le mécanisme enzymatique de LAR sur une lysine catalytique déprotonée qui initierait la réaction. Les contre-ions ont été rajoutés pour garantir la neutralité électronique du milieu (DFR : 10 contre-ions Na⁺; LAR : 8 contre-ions Na⁺; F3'H : 7 contre-ions Cl⁻). Chacun de ces systèmes a été ensuite placé dans une boîte de solvant avec conditions périodiques et avec une distance minimale du placement du solvant autour du soluté de 8 Angströms.

DFR	LAR
K39 (Lysine protonée)	K140 (Lysine protonée)
K167 (Lysine protonée)	H122 (Histidine protonée en δ)
H244 (Histidine protonée en ϵ)	H172 (Histidine doublement protonée)
/	E94 (Glutamate déprotoné)
/	D98 (Aspartate déprotoné)
/	K50 (Lysine protonée)

Tableau 1 : Etat de protonation de résidus d'intérêts au sein des enzymes DFR et LAR. F3'H possédant un site actif très hydrophobe, il ne présente pas de résidus titrables.

II.2) Représentation des partenaires biologiques

II.2.a) Cofacteurs

En ce qui concerne le cofacteur NADPH et son état oxydé NADP⁺, les paramètres ont été repris de l'étude de Ryde *et al.*[134, 135] et sont utilisés dans les simulations impliquant DFR et LAR.

Le groupement hème est un cofacteur dans les cytochromes P450. De nombreuses études et de nombreux paramètres ont été publiés[136, 137], mais une publication récente de Shahrokh *et al.*[138] met en lumière l'importance de la cystéine coordinante dans l'interaction globale de l'hème avec F3'H. De nouveaux paramètres ont donc été calculés de manière spécifique pour cette cystéine et pour le groupement hème impliquant différents états d'oxydation de l'atome métallique. Ici, ces paramètres ont donc été choisis pour un état d'oxydation II de l'atome de fer (état initial avant réaction).

II.2.b) Substrat

Dans ces différentes études, nous avons choisi de considérer trois substrats différents : DHQ (produit de F3'H et substrat de DFR), DHK (substrat de F3'H, mais également de DFR), et LCC (produit de DFR et substrat de LAR). Les paramètres de ces différentes molécules ont été assignés depuis le champ de force GAFF (Général Amber Force Field[139]). Les charges ont été obtenues à partir de calculs de chimie quantique avec le logiciel Gaussian03[140] au

niveau de théorie HF/6-31G*. Le logiciel Antechamber a ensuite été utilisé pour calculer les charges de type RESP sur chaque atome de nos substrats.

Dans le cas particulier de F3'H, dont la structure est issue d'une reconstruction par homologie, le placement du cofacteur et du substrat à l'intérieur de la cavité ne sont pas connus et nécessite alors l'utilisation de méthodes de docking pour pouvoir prédire leur position. Autodock Vina[141] a été utilisé pour réaliser le calcul de docking du substrat (DHK) et du produit (DHQ). La grille a été définie au centre de l'enzyme et est de dimension $16,5 \times 15 \times 16,5 \text{ \AA}^3$, ce qui permet de limiter l'espace de recherche d'interactions. Une seule position du DHK a été retenue pour la réalisation des simulations de dynamique moléculaire. En effet, les autres solutions proposées par le logiciel n'étaient pas cohérentes avec le mécanisme de l'enzyme, qui consiste en une oxydation du cycle B, par addition d'un groupement hydroxyle[142, 143]. Seule la position, représentée sur la Figure 32, était compatible avec le mécanisme enzymatique envisagé.

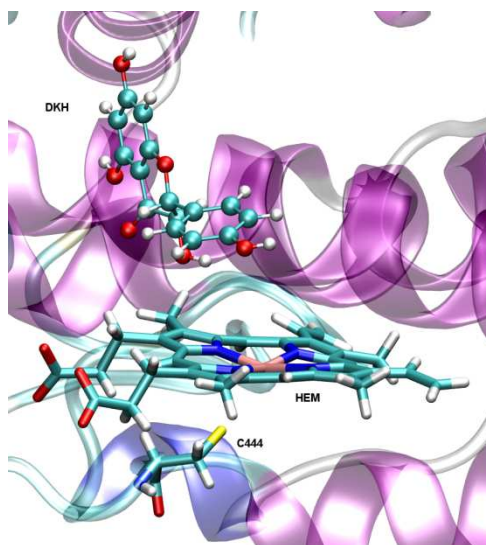


Figure 32 : Pose du dihydrokaempferol retenue pour les simulations de dynamique moléculaire de F3'H.

Concernant DHQ, une seule position a été retenue pour les simulations et est équivalente à la pose du DHK.

II.3) Préparation des systèmes avec membrane

Dans le cadre de l'étude du métabolisme, la description des interactions entre enzyme et la membrane est cruciale, plus particulièrement pour F3'H qui joue le rôle d'ancre à la membrane. Il est donc indispensable de pouvoir prédire correctement les interactions entre

F3'H et la membrane, via son hélice transmembranaire uniquement ou également par une interaction de surface. Pour obtenir une première estimation de cette interaction, la structure de F3'H a été soumise à l'analyse du serveur OPM, qui permet de prédire les interactions entre protéines et membrane par comparaison avec une base de données préexistante.

Les 25 acides aminés structurés en hélice alpha, dans la partie N-terminale, ont été prédits comme étant transmembranaires. L'angle de "tilt" de l'hélice par rapport à la surface de la membrane est de 30°. Au delà de cette caractérisation, qui confirme les données expérimentales sur les cytochromes P450, d'autres résidus de l'enzyme ont été identifiés comme interagissant à la surface de la membrane. Ces derniers sont pour la plupart hydrophobes (tryptophane 226, alanine 223, proline 222, isoleucine 37, proline 36, phénylalanine 69). Afin de vérifier si cette prédiction n'est pas un artefact dû à la présence de l'hélice transmembranaire, nous avons décidé de prédire l'interaction avec la membrane de F3'H en enlevant la partie N-terminale. Il est intéressant de noter que les mêmes acides aminés sont retrouvés en interaction avec la membrane, laissant supposer un unique mode d'interaction F3'H/membrane. Les différentes portions impliquées dans l'interaction sont donc l'hélice transmembranaire, jouant le rôle d'ancre membranaire, ainsi les résidus 222-227, 35-37 et 69, interagissant à la surface. Ces derniers peuvent empêcher des mouvements trop importants de l'enzyme lorsque cette dernière est liée. L'interaction est représentée sur la Figure 33.

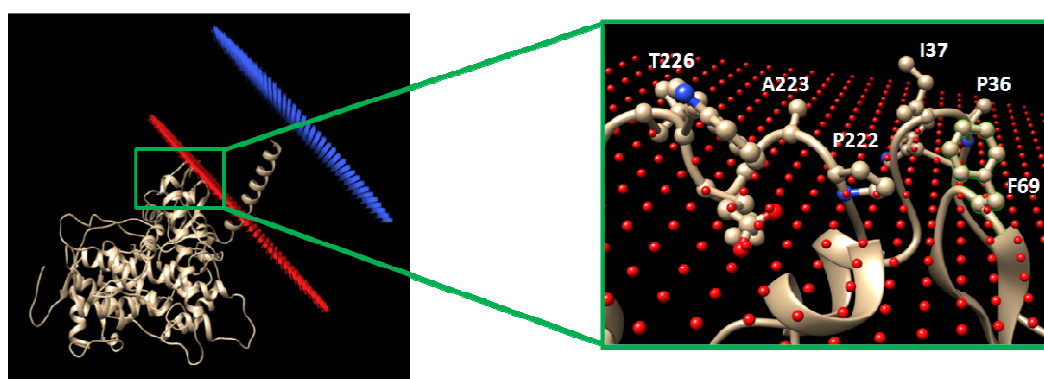


Figure 33 : à gauche, représentation de l'interaction F3'H/membrane prédite par le serveur OPM. A droite, détail de l'interaction de surface et des chaînes secondaires des résidus impliqués.

En se basant sur ce positionnement, le logiciel Maestro[144], de la suite Schrödinger, a été utilisé pour placer la membrane explicite autour de la partie N-terminale. En plaçant la membrane de telle façon à avoir la position prédite par le serveur OPM, nous avons observé

effectivement une interaction de surface avec les zones 35-36 et 222-227, sans provoquer de phénomènes d'encombrement stériques, qui entraînerait un déséquilibre trop grand dans le système. La membrane est traitée par le champ de force GAFF lipids, et est représentée par un modèle POPC pour les lipides utilisés. 332 lipides de type POPC ont donc été placés pour former la couche lipidique, puis le solvant est ajouté, pour un total de 118000 atomes.

III) Protocole de simulation de dynamique moléculaire

Avant de réaliser les simulations de dynamique moléculaire, nous devons d'abord adapter nos systèmes au champ de force que nous utilisons pour le représenter. Les lipides (dans le cas de calculs avec membrane) et surtout les molécules d'eau sont disposés autour de notre structure sans rationnel particulier. Leur placement doit donc être optimisé autour de notre entité biologique. Pour cela, nous avons développé un protocole permettant de minimiser l'énergie totale de notre système, de manière à relaxer le système initial généré par le logiciel Xleap. Se faisant, les informations que nous pourrions tirer des simulations seront ainsi pertinentes. Le protocole comprend 4 étapes :

- 1) Etape de minimisation : l'énergie du système est minimisée afin de le faire atteindre un minimum de la surface d'énergie potentielle. La minimisation (combinaison de Gradient conjugué et de "steepest descent") se fait avec contrainte imposée sur l'entité biologique (contrainte fixée initialement à $20 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$). La structure de cette dernière est donc figée, et seules la boîte de solvant et la membrane (si présente) seront minimisées. La minimisation est couplée à de courtes simulations de dynamiques moléculaires (30 à 50 ps selon la taille des systèmes) pour relaxer le système. Plusieurs étapes de ce type seront appliquées, avec une diminution de la contrainte imposée sur la structure (diminution de $5 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ par étapes). Cela permet d'équilibrer de manière douce l'environnement du système. La dernière minimisation se fait sur le système entier, sans contrainte.
- 2) Thermalisation : Cette étape permet l'injection d'énergie cinétique dans le système, afin de simuler les conditions d'une température de 300K. Le thermostat de Langevin (plus de détails dans la partie méthodologique, page 65) est utilisé pour la réalisation pour cette étape, et par la suite pour réguler la température.
- 3) Equilibration : Cette étape permet au système de se relaxer et d'atteindre l'équilibre après que la température soit montée jusqu'à 300K. Elle est réalisée dans l'ensemble isobare-isotherme (NTP) et d'une durée variable entre 2 et 10 ns selon les systèmes.

- 4) Production : Cela consiste en la simulation de dynamique moléculaire à proprement parler, dans laquelle les résultats et informations observés seront jugés comme étant pertinents. Les étapes de simulation sont enregistrées tous les 20 ps.

Chapitre 2 : Etude de la réactivité de DFR

Au sein des organismes vivants, la biosynthèse des composés de la famille des flavonoïdes implique un nombre important d'enzymes (voir Figure 3 et Figure 11) [22]. L'étude de ces chemins de biosynthèse nécessite des connaissances et compréhensions précises des étapes qui les composent. Ainsi, il est ici important d'obtenir des descriptions précises concernant les transformations chimiques catalysées par ces enzymes, à la fois d'un point de vue structural mais aussi d'un point de vue énergétique.

Dans le cadre de notre étude sur un complexe de trois enzymes, l'enzyme DFR joue le rôle de pivot central. Il s'agit de l'enzyme qui reçoit son substrat par diffusion après réaction au sein de F3'H. Elle le transforme par une réaction de réduction impliquant un cofacteur et qui va finalement relarguer le produit pour qu'il diffuse vers le site actif de l'enzyme suivante, LAR (leucoanthocyanidine réductase). Nous avons choisi de porter tout particulièrement nos efforts sur une description du mécanisme catalysée par cette enzyme centrale [83, 145, 146]. Le site actif induit des effets très importants sur la stabilisation et la réactivité des composés, permettant la réalisation de réactions optimisées et souvent stéréosélectives. De ce fait, la prise en compte des effets du milieu enzymatique, qu'ils soient électrostatiques, stériques ou dynamiques, est absolument nécessaire afin de caractériser correctement la catalyse de l'enzyme. Pour cela, nous avons choisi de mettre en place une campagne de calculs utilisant des méthodes hybrides de type QM/MM-MD, couplées à des protocoles d'Umbrella Sampling. Cette méthodologie nous assure de prendre en compte l'influence de l'environnement sur la réaction et, si les contraintes sont bien choisies, de représenter au mieux un chemin réactionnel le long d'une coordonnée de réaction approchée. La partie réactionnelle sera donc traitée par des méthodes de chimie quantique afin de permettre la réaction, tandis que l'environnement sera modélisé par un champ de force, et le tout sera étudié par application de simulation de dynamique moléculaire. Cependant, ces dernières ne permettent pas l'échantillonnage de conformation haute en énergie, donc défavorable selon la statistique de Boltzmann. Le franchissement de barrières d'activation fortes est très difficile à observer dans des simulations de dynamique moléculaire, nous ne pourrions donc pas échantillonner l'état de transition d'une réaction chimique de manière naturelle. Pour réaliser cela, nous appliquons le protocole d'Umbrella Sampling [147], qui consiste en l'introduction d'un biais dans le calcul, généralement sous forme de potentiel harmonique, selon une coordonnée préalablement définie (distance, angle, combinaison linéaire...). Cette contrainte devra être appliquée de façon à représenter au mieux la coordonnée de réaction associée à notre transformation. Le calcul réalisé est ensuite retraité par le logiciel WHAM [148] qui va

nous permettre d'obtenir les variations d'énergie associées à la modification de la contrainte choisie. Si le biais que nous avons choisi représente de manière correcte la coordonnée de réaction associée à une réaction chimique, nous pouvons ainsi obtenir la surface d'énergie potentielle correspondant à la réaction. Dans le cas de réactions complexes faisant intervenir plusieurs partenaires, une seule contrainte n'est pas suffisante pour décrire de manière exhaustive la réaction considérée. Nous pouvons dans ce cas prendre plusieurs contraintes, afin d'obtenir une analyse plus complète.

Dans ce chapitre seront présentés les résultats concernant les interactions entre deux substrats de DFR, la dihydroquercétine et le dihydrokaempferol (DHQ et DHK voir figure 5). La réactivité de l'enzyme sera étudiée dans le cas de complexes ternaires DFR-NADPH-DHQ et DFR-NADPH-DHK. Des études de réactivité ont été réalisées en faisant varier le nombre de contraintes utilisées : tout d'abord une seule contrainte qui décrit le transfert d'hydrure provenant du cofacteur, puis l'ajout d'une seconde contrainte, représentant le transfert de proton provenant de la tyrosine163 de l'enzyme.

I) Méthode QM/MM : Théorie

I.1) Principe

L'utilisation de méthodes hybrides de type QM/MM nécessite la définition d'une partie traitée par des méthodes de chimie quantique, englobée dans une partie modélisée par un champ de force[108]. L'énergie du système peut s'écrire :

$$E_{sys} = E_{QM} + E_{MM} + E_{QM/MM}$$

(Équation 45)

E_{QM} correspond à l'énergie du système quantique obtenue par résolution de l'hamiltonien correspondant à la méthode utilisée, E_{MM} représente l'énergie de la partie traitée en mécanique moléculaire (selon la somme des termes liés et non liés, voir partie méthodologique page 60), et $E_{QM/MM}$ l'interaction entre les deux interfaces. Cette dernière est séparée en plusieurs termes, détaillés dans l'équation ci-dessous.

$$E_{QM/MM} = \sum_{j < i}^i \left[\frac{q_i q_j}{r_{ij} \epsilon} + \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right) \right]$$

(Équation 46)

Cette interface est gérée par le procédé "mechanical embedding", c'est à dire que l'interaction est traitée comme si la partie quantique était considérée en mécanique moléculaire, avec des rayons de van der Waals attribués selon un type d'atome. Les charges ponctuelles sont attribuées aux atomes de la zone QM qui interagissent avec l'environnement. Même si cette approche permet d'appliquer des contraintes stériques sur la partie réactive, aucune polarisation de la densité électronique n'est cependant prise en compte, ce qui est assez problématique si nous voulons décrire de manière précise le comportement de la partie quantique. Il existe un procédé plus élaboré, appelé "electronical embedding" qui implique que les charges de l'environnement MM soient directement intégrées dans l'hamiltonien de la partie quantique, sous forme d'une perturbation. Par conséquent, l'interface QM/MM de l'équation 46 ne représente plus que l'interaction de type van der Waals. La prise en compte de l'effet électrostatique sur notre zone réactive est ainsi plus précise. Cette méthode est celle que nous avons choisi dans nos calculs.

I.2) Délimitation entre partie quantique et partie classique

La délimitation de la frontière entre les parties QM et MM représente le point le plus délicat à traiter. En effet, il faut prendre les atomes nécessaires pour la réaction dans la partie réactive pour une bonne description du mécanisme. En contrepartie, il ne faut pas trop alourdir le calcul dans la zone QM en insérant trop d'atomes. Il est possible que lors de la démarcation des deux parties, une liaison covalente doit être rompue. Afin de ne pas avoir de problème d'hybridation et de valence partielle, il faut qu'elle soit présente dans notre système quantique. Plusieurs méthodes existent afin de combler la valence manquante. Nous pouvons citer la méthode du "link atom", des orbitales gelées ou du potentiel cappé[149]. Ici, la méthode du "link atom" a été utilisée. Il s'agit d'insérer dans la partie quantique un atome (souvent un atome d'hydrogène) afin de compléter la valence de l'atome qui a perdu une liaison[150]. La position de cet atome est superposée sur celle qu'il remplace dans la partie MM le long de la simulation de dynamique moléculaire. La distance entre le link atom et celui avec lequel il est lié dans la partie QM est fixée à 1,09 Å (distance classique C-H). Au niveau des interactions avec l'environnement, le link atom ne possède pas de rayon de van der Waals. Cependant, il possèdera une charge, calculée par la résolution de l'hamiltonien en QM, qui sera associée à sa paire dans la zone MM.

Afin que cette approximation du link atom ne perturbe que peu le système, il est nécessaire de bien choisir la délimitation entre la partie QM et la zone MM. Dans ce cadre, il

est préférable de choisir une liaison peu polarisée comme les liaisons simples entre atomes de carbone. Le rayon de troncature délimite les interactions à courte distance, telles que calculées par le potentiel coulombien standard. Cette limite de distance est généralisée pour tous les atomes de la partie QM. Cette définition induit la délimitation d'une zone tout autour de la partie quantique associée aux interactions à courte portée. Au delà du rayon de troncature (choisi classiquement à 8 Å), on passe sur des interactions à longue portée, traitées comme en simulation de dynamique moléculaire classique par la méthode du Particle Mesh Ewald (PME).

II) Protocole des simulations hybrides

II.1) Choix de la partie quantique

Dans notre système, la partie quantique est composée d'un total de 96 atomes et est traitée par une variante plus élaborée de la méthode semi-empirique "tight-binding" DFT, appelée SCC-DFTB (Self-Consistent Charge DFTB)[151, 152]. Dans celle-ci, les charges des atomes de la partie quantique sont optimisées au travers d'un cycle auto-cohérent qui prend en compte les charges de la partie MM lors de la résolution de ces cycles.

La molécule de NADPH étant composée d'un nombre important d'atomes (74 atomes), nous avons choisi de ne pas la prendre en compte en totalité dans la partie quantique. Dans le cas contraire, des interactions susceptibles de perturber l'équilibre du système quantique seraient prises en compte, et alourdiraient le temps de calcul associé à la résolution de l'hamiltonien. Afin de vérifier que la troncature de cette partie n'altère pas les charges sur la partie réactive du cofacteur (cycle dihydropyrimidique et groupement ose, cf Figure 34), les charges entre le cofacteur complet et la molécule tronquée ont été comparées au moyen de calculs de chimie quantique (niveau B3LYP//6-31++G**). Les résultats indiquent que les charges sont identiques entre le modèle tronqué et la molécule complète, ce qui permet de valider la pertinence de la coupure pour le cofacteur, ainsi qu'une bonne représentation du comportement de sa partie réactive.

La partie quantique de notre système sera donc constituée de notre molécule tronquée de NADPH, du substrat considéré en totalité et de parties de chaînes latérales de trois résidus de l'enzyme : tyrosine 163, lysine 167 et sérine 128. Plusieurs tests préliminaires ont montré qu'il était nécessaire de considérer la totalité de la triade catalytique de l'enzyme pour obtenir une description correcte de la réaction. Au final, 4 link atoms sont donc retenus pour séparer

la molécule de NADPH et les 3 résidus du site actif du reste de l'enzyme. Celle-ci, ainsi que la seconde moitié du cofacteur et le solvant, font partie intégrante de la partie traitée à l'aide du champ de force. Au total, le nombre d'atomes dans la partie quantique s'élève à 94. (90 atomes plus les 4 "link atoms"). La Figure 34 montre de manière détaillée la partie quantique.

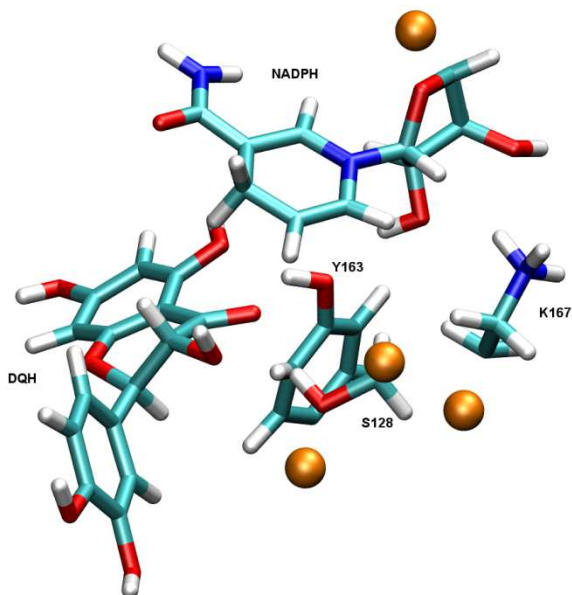


Figure 34 : Représentation de la partie QM. En *licorice* sont représentés les différents atomes considérés dans la partie quantique. Les atomes oranges désignent les quatre "link atom" pour les résidus S128, Y163, K167 et le cofacteur NADPH.

II.2) Apport du protocole d'Umbrella Sampling et choix des contraintes

Un protocole de dynamique moléculaire, couplé à cette représentation par méthode hybride QM/MM, permet de prendre en compte les phénomènes dynamiques de l'environnement et leurs influences sur la partie réactive. Cependant, la dynamique moléculaire obéit à la statistique de Boltzmann, qui impose que les états échantillonnés sont associés à une énergie potentielle fortement négative, gage de stabilité du système, et donc de probabilité élevée. L'échantillonnage d'un état de faible probabilité, comme par exemple un état de transition, ne pourra être que très difficilement, voire quasiment impossible à observer au cours d'une simulation non biaisée. Une solution pour étudier la réaction est de réaliser un couplage des simulations de DM QM/MM avec un protocole d'Umbrella Sampling. L'introduction d'un biais va nous permettre de suivre la variation d'énergie libre par rapport à cette contrainte. Ainsi, une fois le biais défini de façon adéquate, il sera possible

d'échantillonner les états de haute énergie associés à la coordonnée de réaction (si notre contrainte la représente pertinemment), et donc d'étudier une réaction chimique. Une fois les simulations réalisées, un retraitement des trajectoires par le logiciel WHAM[148] permettra d'extraire les variations d'énergie en fonction de la contrainte. Un potentiel de force moyenne, pouvant être assimilé à une variation d'enthalpie libre le long de la coordonnée de réaction, est alors obtenu. D'un point de vue pratique, afin d'observer ces transferts d'atomes hydrogènes (proton ou hydrure), il est nécessaire de ne pas appliquer l'algorithme SHAKE pour la partie quantique. En effet, cet algorithme permet de « geler » les liaisons de type X-H. Le corolaire à cette obligation est qu'il est désormais nécessaire de définir un pas d'intégration plus petit permettant de simuler les vibrations le long des liaisons X-H. Le pas est ainsi fixé à 0,5 ps dans ce type de calcul.

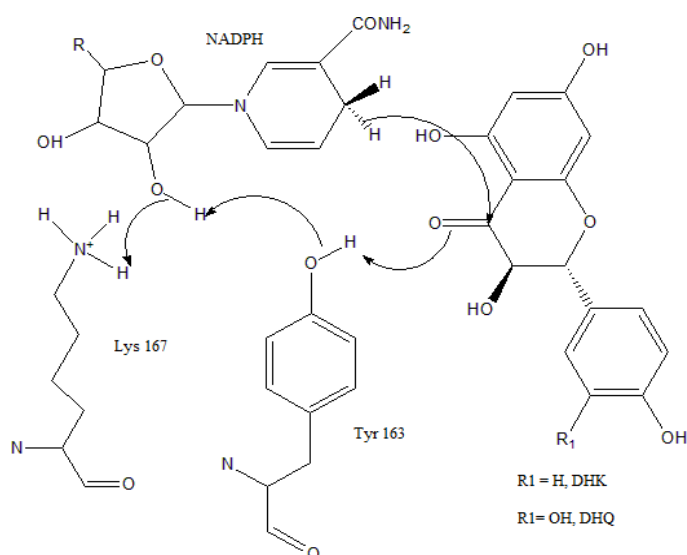


Figure 35 : Mécanisme enzymatique de la réduction du substrat par l'enzyme DFR.

L'hydrure est fourni par le cofacteur NADPH. Les résidus Y163 et K167 participent à la cascade de transferts de protons. Pour rappel, DHQ : Dihydroquercétine, DHK : Dihydrokaempferol.

Notre réaction implique la réduction du groupement cétone du substrat par un transfert d'hydrure provenant du cofacteur. Afin d'accéder aux produits de la réaction, une cascade de transfert de protons impliquant les résidus Tyr 163 et Lys 167 de l'enzyme, ainsi qu'une nouvelle fois le NADPH doit avoir lieu. La lysine doit ensuite être reprotonée par le solvant pour permettre à l'enzyme d'être régénérée et ainsi terminer le cycle catalytique. Le mécanisme proposé est rappelé sur la Figure 35.

Nous avons donc au total quatre différents partenaires biologiques qui sont impliqués dans la totalité de la réaction. Il est donc plus difficile d'identifier de manière précise la coordonnée de réaction. Notre choix s'est porté sur la représentation de ces différents transferts par une seule contrainte, correspondant au transfert d'hydrure provenant du cofacteur dans un premier temps, arguant du fait que ce transfert entrainerait les transferts de protons. Nous avons également testé une contrainte représentant le transfert de proton provenant de la tyrosine afin de vérifier si le transfert d'hydrure pouvait intervenir dans un second temps lors de la réaction. Ces tests nous ont montré qu'une seule contrainte « simple » ne peut représenter pertinemment la coordonnée de réaction du mécanisme. Des simulations impliquant deux coordonnées de réactions ont été effectuées, basées sur deux contraintes associées indépendamment aux transferts d'hydrure et du premier proton.

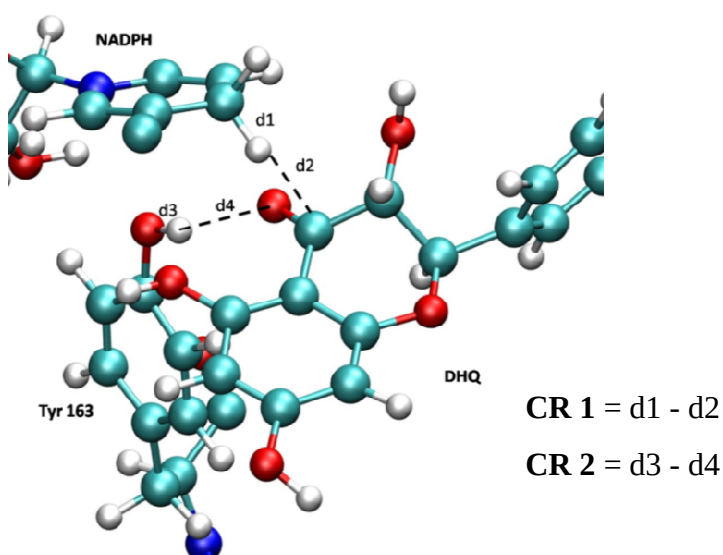


Figure 36 : Schéma des distances impliquées dans la définition des coordonnées de réaction.

Le transfert d'hydrure est modélisé par la contrainte CR1 (égale à d1-d2), tandis que le transfert de proton venant de la tyrosine 163 est représentée par le biais CR 2 (égale à d3 - d4).

Chaque contrainte a été choisie, après plusieurs tests, comme étant une combinaison linéaire de deux distances. Cela permet une meilleure représentation des transferts d'atomes. Par exemple, pour simuler le transfert d'hydrure, la coordonnée de réaction est la combinaison linéaire de la distance entre l'hydrure et l'atome de carbone du NADPH qui le porte (nommée d1) et celle entre l'hydrure et l'atome de carbone du substrat qui va être réduit (nommée d2). Pour la contrainte sur le transfert de proton, la coordonnée est constituée de la combinaison linéaire entre la distance tyrosine-proton transférable et la distance proton-atome d'oxygène

du groupement cétone réduit par la réaction (respectivement appelée d3 et d4). La Figure 36 représente les contraintes choisies.

Les paramètres des potentiels harmoniques associés aux biais sont choisis selon des études déjà publiées[153]. Pour chaque contrainte, les paramètres utilisés sont listés ci-après :

- Transfert d'hydrure : constante de force choisie à 20 kcal/mol; espacement entre fenêtre de 0,15 Å; 35 ps d'échantillonnage dans chaque fenêtre et la contrainte varie de -1,80 à 1,35 Å.
- Transfert de proton : constante de force choisie à 20 kcal/mol; espacement entre fenêtre de 0,15 Å; 35 ps d'échantillonnage pour chaque fenêtre et la contrainte varie de -0,90 à 1,20 Å.

III) Interactions entre les différents substrats et l'enzyme - calculs US-DM QM/MM à une dimension

III.1) Simulation sans contraintes DM classique et hybrides QM/MM

Pour chacun des deux systèmes, complexes enzyme-NADPH-DHQ et enzyme-NADPH-DHK, 5 simulations de DM classique, d'une durée de 20 ns, ont été réalisées afin de vérifier leur stabilité. Les RMSd calculées sur les évolutions structurales de l'enzyme le long des trajectoires montrent que dans chaque cas, la structure générale reste cohérente lors des différentes simulations. De manière générale, le RMSd de l'enzyme reste inférieur à 2,5 Å. En cohérence avec les analyses effectuées sur la structure cristallographique, les simulations du complexe ternaire montrent que les résidus Asn 133 (deux interactions dans le cas du DHQ, une pour DHK) et Glu 227 interagissent de manière forte avec les groupements hydroxyles du cycle B du substrat et peuvent, de ce fait, être considérés comme des résidus de reconnaissance. D'autres acides-aminés, faisant partie de la triade catalytique, sont impliqués dans des interactions par liaison hydrogène : la tyrosine 163 et la sérine 128. A un moindre niveau, car faisant intervenir une molécule d'eau pontante, la thréonine 191, ainsi qu'un des atomes d'oxygène du groupement diphosphate du cofacteur, interagissent avec un groupement hydroxyle du cycle C. Un grand nombre de résidus de la cavité est impliqué dans des interactions de type hydrophobes avec le substrat. Nous pouvons particulièrement noter

l'isoleucine 222 et la valine 193, qui sont en contact avec les cycles aromatiques par interactions de type σ - π .

Afin d'équilibrer et préparer nos systèmes dans le cadre de l'étude de la réactivité, plusieurs simulations de type DM-QM/MM ont été réalisées d'une durée de 1 à 2 ns : 2 simulations de 1 ns et une de 2 ns pour le complexe ternaire avec DHQ, 2 simulations de 1 ns pour le complexe DHK-enzyme, soit un total respectif de 4 et 2 ns simulées. Les interactions relevées le long des simulations classiques sont retrouvées dans les calculs hybrides, avec cependant des valeurs légèrement différentes (par exemple, la thréonine 191 interagit directement avec le substrat). Les interactions de type liaison hydrogène sont notamment plus courtes, de l'ordre de 1,8-1,6 Å. Nous pouvons noter que la seule différence notable entre l'interaction du DHQ et celle du DHK est l'absence d'une liaison hydrogène avec l'asparagine 133 provenant de l'absence d'un groupement hydroxyle sur le cycle B du DHK. La Figure 37 représente de manière détaillée les interactions stabilisantes de la dihydroquercétine avec la DFR. Cet arrangement spatial est compatible avec la réaction telle qu'elle est anticipée. La Figure 37 représente ainsi la structure de départ choisie pour commencer l'étude de réactivité.

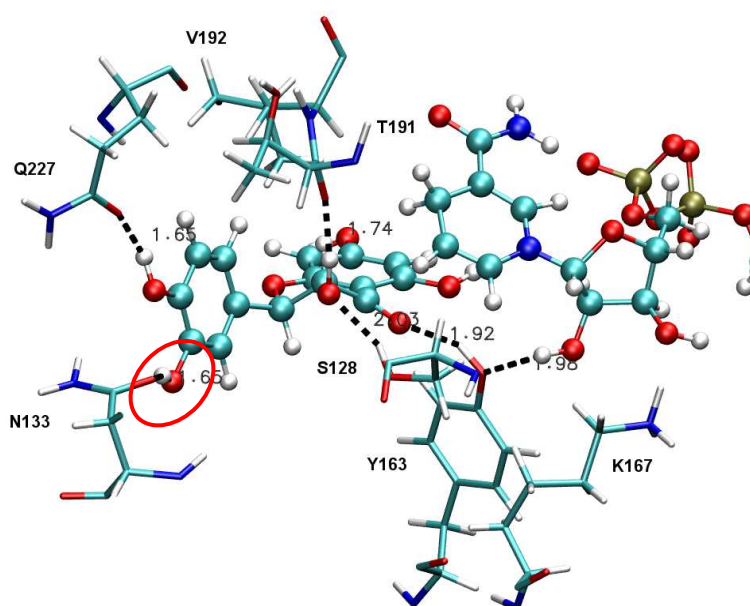


Figure 37 : Représentations des différentes interactions du substrat dans le site actif.
L'interaction entre l'asparagine et la dihydroquercétine, qui n'existe pas dans le cas du dihydrokaempferol, est entourée en rouge.

III.2) Calculs DM-QM/MM US à une dimension

Dans un premier temps, des premiers calculs de DM-QM/MM couplés avec le protocole d'Umbrella Sampling ont été effectués avec une seule coordonnée de réaction représentant uniquement le transfert d'hydrure. La simulation, d'une durée totale de 735 ps (21 fenêtres de 35ps), a permis l'obtention d'un potentiel de force moyenne associé à la réaction de transfert. Ce profil est représenté sur la Figure 38.

Ce potentiel présente trois phases distinctes :

- de -1,80 à 0,5 Å : l'énergie varie de 0 à 6 kcal/mol. Cette phase correspond exclusivement à un rapprochement du NADPH et du substrat. En effet, durant cette phase, et malgré l'application de la contrainte, la distance $C_{\text{NADPH-H}}$ n'évolue pas. L'état obtenu à la fin de ce rapprochement est donc un état intermédiaire idéal pour le transfert d'hydrure.

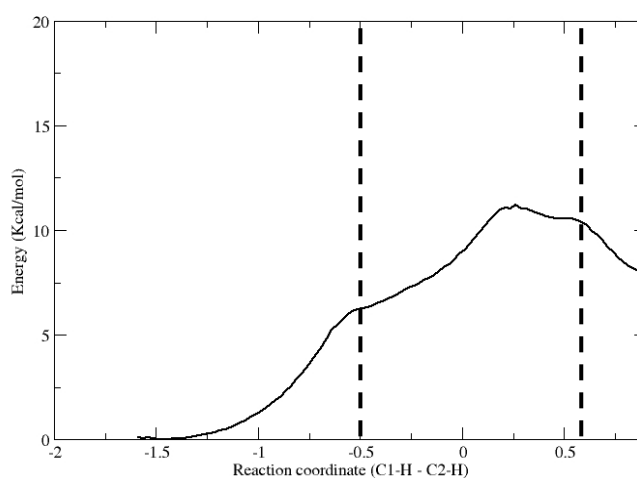


Figure 38 : Potentiel de force moyenne obtenu pour le transfert d'hydrure.

- de -0,5 à 0,2 Å : l'énergie varie de 6 à 12 kcal/mol. Cette phase correspond au transfert d'hydrure. L'état de transition du transfert d'hydrure a été ainsi identifié pour une valeur de 0,2 Å de la coordonnée, ce qui correspond à des distances $C_{\text{NADPH-H}} = 1,43$ Å et $H-C_{\text{DHQ}} = 1,22$ Å.
- de 0,2 à 0,8 Å : l'énergie varie de 12 à 7,5 kcal/mol. Cette phase correspond à la relaxation provoquée par le franchissement de l'état de transition. La cascade de transferts de protons ne s'effectue qu'après le transfert d'hydrure. Nous pouvons

remarquer que les produits de réactions possèdent une énergie plus élevée que les réactifs. Nous pouvons supposer que cela provient du fait que la lysine167 n'est pas reprotonée en fin de réaction lors de notre simulation.

En effectuant un calcul QM au niveau B3LYP/6-31++G** de la reprotonation par un ion oxonium d'une lysine déprotonnée, l'enthalpie libre de Gibbs associée est de l'ordre de -12 kcal/mol. Cela permet d'estimer que l'état final de notre réaction pourrait atteindre une énergie relative de - 6kcal/mol par rapport aux réactifs. Dans le cas du DHK, un comportement similaire a été observé, avec cependant une barrière d'activation plus élevée (état de transition à 16 kcal/mol).

Comme nous venons de le voir, cette campagne de simulations biaisées du transfert d'hydrure entraîne la cascade de transferts de protons dans son intégralité. Afin de vérifier si le transfert du premier proton (biais CR2) entraînerait le transfert d'hydrure, une campagne de calculs simulant le transfert de proton de la tyrosine vers le substrat a été réalisée suivant le même protocole. Une barrière très élevée (18 kcal/mol) est relevée sur le profil énergétique. La simulation conduit à un intermédiaire très faiblement stabilisé par rapport à l'état de transition (17 kcal/mol). Dans cette structure, le transfert de proton est effectif mais n'induit ni le transfert des autres protons, ni le transfert d'hydrure. Il apparaît donc que la première simulation devrait nous permettre de caractériser l'état de transition de la réaction biochimique.

Afin de le vérifier, nous avons réalisé une nouvelle simulation de DM-QM/MM biaisée, pour laquelle est appliqué le biais correspondant à la fenêtre encadrant l'état de transition. Cette valeur correspond à un biais de $CR1 = 0,2 \text{ \AA}$. L'idée étant que si nous sommes exactement sur cet état de transition, chaque système soumis à des simulations de dynamique moléculaire à partir de cette position devrait se relaxer de manière aléatoire vers les produits ou les réactifs de la réaction. Une simulation longue (2 ns), sous contrainte, a été effectuée et certaines structures ont été choisies le long de cette trajectoire comme point de départ pour des simulations sans contrainte. 100% de ces simulations (environ 30 calculs différents) présentent un retour vers les réactifs uniquement. De plus, lors de l'analyse de la trajectoire sous contrainte, il apparaît que la tyrosine présente un comportement particulier. En effet, le proton se transfère de manière naturelle vers le substrat mais revient également auprès de l'atome d'oxygène de la tyrosine, et ceci de manière spontanée. De ces analyses, il apparaît qu'il existe un couplage entre la

coordonnée représentant le transfert d'hydrure et la coordonnée de réaction associée au transfert de proton. Par conséquent, les calculs basés sur l'utilisation d'une contrainte ne peuvent pas être suffisant pour décrire correctement le mécanisme. Nous devons ainsi réaliser des simulations plus élaborées basées sur la définition de deux biais.

IV) Calculs DM-QM/MM US 2D

D'une façon générale, la réalisation de ce type de protocole permet d'obtenir des surfaces d'énergie potentielle en fonction de deux coordonnées. Pratiquement, deux biais seront appliqués selon les deux transferts, celui de l'hydrure et celui du proton provenant de la tyrosine. Ces biais seront appliqués comme précédemment, à l'aide d'un potentiel harmonique appliqué sur une combinaison linéaire des distances impliquées dans les transferts. Ainsi, dans chaque « fenêtre », le système sera échantillonné autour de ces deux contraintes. La totalité de nos simulations sera ensuite débaisée une nouvelle fois à l'aide du protocole WHAM et nous permettra d'obtenir la surface d'énergie en fonction des contraintes appliquées. Ici encore, il faut nous assurer que les contraintes représentent au mieux la coordonnée de réaction attendue. Nous avons considéré les contraintes CR1 et CR2, qui représentent respectivement les transferts d'hydrure et du proton de la tyrosine, et qui, comme nous l'avons vu précédemment, sont les deux transferts d'atomes qui représentent la réaction complète.

Afin d'obtenir la carte $E(\text{CR1}, \text{CR2})$, nous devons définir les limites des variations de ces deux contraintes. Dans ce cadre, il est nécessaire de déborder quelque peu des valeurs associées aux états stables des réactifs et des produits, afin de s'assurer que nous avons bien échantillonné ces états stationnaires. Pour cela, les simulations 1D présentées précédemment nous sont d'une grande aide. Ainsi, pour la contrainte CR1, d'après la Figure 4, les minima d'énergie sont obtenus, coté réactif pour $\text{CR1} \approx -1,5 \text{ \AA}$ et coté produit pour $\text{CR1} \approx +1 \text{ \AA}$. Au final, pour cette contrainte CR1, nous considérerons 22 valeurs allant de $-1,8 \text{ \AA}$ (distances hydrure-DHQ et hydrure-cofacteur respectivement égales à $2,8 \text{ \AA}$ et $1 \text{ \AA}$) à $+1,35 \text{ \AA}$ ($1 \text{ \AA}$ et $2,35 \text{ \AA}$), ce qui nous permet de définir 22 fenêtres de simulations. Concernant la deuxième contrainte, les variations sont beaucoup moins importantes car il n'est pas nécessaire ici que le substrat et la tyrosine se rapprochent pour que le transfert puisse s'effectuer. En effet, la liaison hydrogène entre les deux réactifs est présente de façon constante le long des simulations. Ainsi, pour CR2, nous considérerons 15 valeurs, allant de $-0,90 \text{ \AA}$ (distances proton-DHQ et proton-tyrosine respectivement égales à $1,9 \text{ \AA}$ et $1 \text{ \AA}$) à $1,20 \text{ \AA}$ ($1 \text{ \AA}$ et $2,20 \text{ \AA}$), ce qui nous permet, le long de cette dimension de définir 15 fenêtres de simulations. Au

final, en tenant compte des évolutions des deux contraintes, 330 fenêtres ont été considérées, correspondant à un temps total de simulation de 11,55ns.

IV.1) Substrat n°1 : DHQ (dihydroquercétine)

IV.1.a) Analyse énergétique du profil

La surface d'énergie potentielle décrivant l'énergie du système en fonction des deux contraintes est représentée sur la Figure 39.

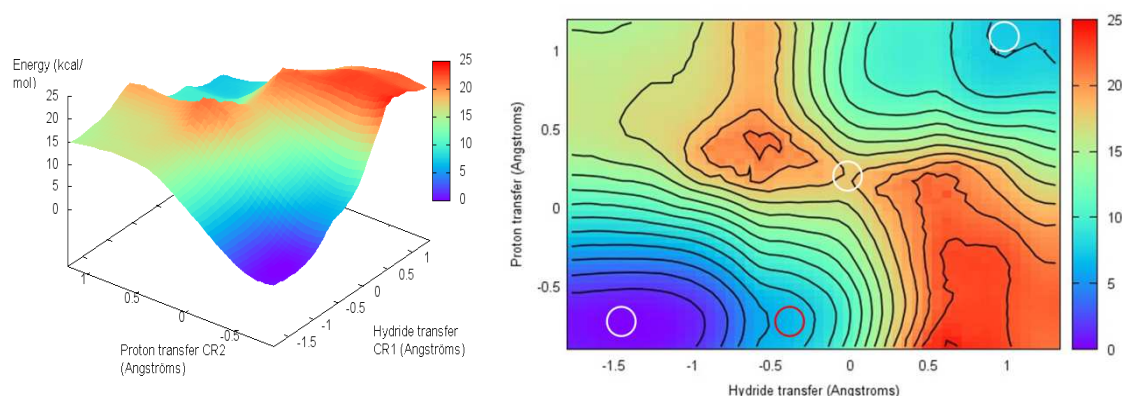


Figure 39: Surface d'énergie potentielle obtenue pour le système DFR-NADPH-DHQ. L'axe x représente la contrainte CR1 et l'axe y CR2. A droite, sur la projection sur le plan horizontal, les cercles blancs désignent les réactifs (en bas à droite), l'état de transition (au centre) et les produits de la réaction (en haut à droite). Le cercle rouge représente un état intermédiaire lié au rapprochement NADPH-DHQ.

Le profil présente un minimum global d'énergie pour des valeurs de CR1 et CR2 négatives, ce qui correspond bien à un état où les atomes ne sont pas transférés, c'est à dire les réactifs. Cet état est en totale cohérence avec les structures échantillonnées lors de simulations sans contraintes. Nous retrouvons en effet les caractéristiques d'un système pré-réactif, stable pendant un temps de simulation relativement important. Ce point qui est celui de plus basse énergie sur toute la surface de potentiel est considéré comme la référence des énergies.

Un point selle, qui peut être considéré en première approximation comme un état de transition, est également caractérisé sur cette surface. L'énergie nécessaire pour atteindre ce point est de 17 kcal/mol. Il est situé pour des valeurs des contraintes représentant les transferts d'hydruure et de proton valant $CR1 = 0,0 \text{ Å}$ et $CR2 = 0,30 \text{ Å}$.

Enfin, un état stable, correspondant à la fin des transferts d'hydrure et de proton, est également caractérisé sur cette surface d'énergie, aux coordonnées $CR1 = 0,90 \text{ \AA}$ et $CR2 = 1,20 \text{ \AA}$. Cette structure correspond à un état où les transferts de protons non contraints, i.e. du fragment ose vers la tyrosine et de la lysine vers le fragment ose, sont également effectifs. Une nouvelle fois, les caractéristiques structurales de cet état sont cohérentes avec celles que nous avons pu échantillonner lors de simulations sans contraintes sur le complexe enzyme-produits (voir chapitre 3). Ce point est associé à une énergie de $8,11 \text{ kcal/mol}$ plus haute que celle des réactifs. Cette valeur ne doit pas être choquante car ici encore, la régénération de la lysine n'est pas effectuée.

Enfin, nous pouvons noter, à nouveau, la présence d'un état intermédiaire métastable aux coordonnées $[-0,45; -0,75]$, à une valeur d'énergie de 6 kcal/mol . Ce point sera discuté plus tard. Le chemin réactionnel associé à la réaction enzymatique peut être représenté en suivant le chemin de moindre pente. Ce profil est représenté sur la Figure 40.

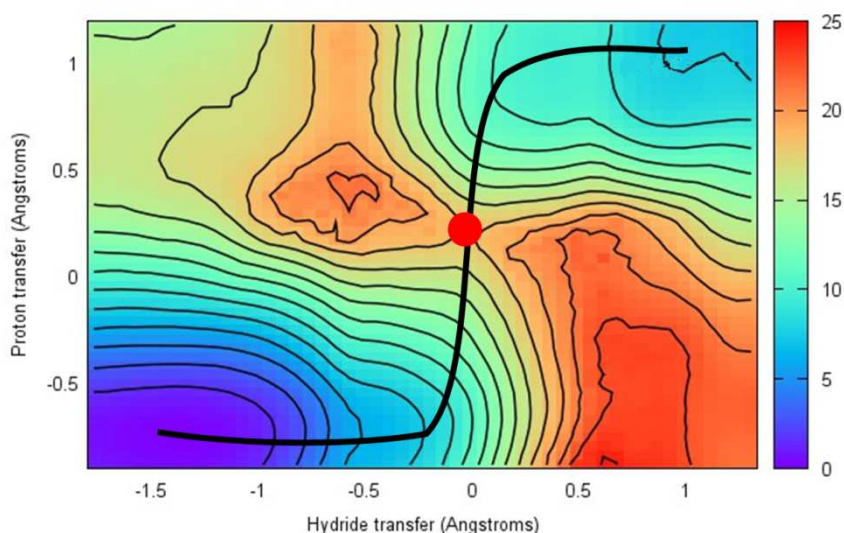


Figure 40 : Chemin réactionnel suivi par le mécanisme enzymatique. Le point rouge correspond à la position de l'état de transition.

Cette vue permet de bien nous rendre compte de l'implication de chaque contrainte en fonction de l'avancée de la réaction. En première approximation, il peut être considéré que le chemin suivi est fonction d'une des coordonnées, indépendamment de l'autre selon trois phases. La première étape, associée au rapprochement des entités réactives, ne fait intervenir que la coordonnée liée au transfert d'hydrure, de $CR1 \approx -1,8 \text{ \AA}$ à $CR1 \approx -0,15 \text{ \AA}$. Durant cette période, la contrainte représentant le transfert de proton n'est pas impliquée dans le chemin de plus basse énergie et reste à une valeur de $-0,60 \text{ \AA}$ environ. A partir d'une valeur de la

contrainte $CR1 \approx 0$ Å, c'est à dire lorsque le transfert d'hydrure est réalisé à moitié, le chemin réactionnel suit alors quasiment exclusivement la coordonnée liée au transfert de proton entre les valeurs $CR2 \approx -0,75$ Å et $CR2 \approx 0,75$ Å. Dans cette partie, le chemin suivi reste à une valeur de $CR1 \approx 0$ Å. Le long de ce transfert de proton, un point selle, que l'on peut considérer comme un état de transition en première approximation, est atteint dans la fenêtre correspondant à $[CR1 = 0,00; CR2 = 0,30]$. Alors que cet état de transition correspond à une structure où l'hydrure est à équidistance entre carbone donneur et carbone accepteur ($C_{NADPH} - H = 1,4$ Å ; $H - C_{DHQ} = 1,4$ Å) il apparaît tardif par rapport à la deuxième contrainte ($O_{TYR} - H = 1,45$ Å ; $H - O_{DHQ} = 1,15$ Å). Une fois ce point selle franchi, le chemin suit encore la coordonnée du transfert de proton, jusqu'à une valeur de $CR2 = 0,75$ Å. A ce point, la descente vers les produits de la réaction (à la position $[0,90;0,90]$) suit plus particulièrement la contrainte représentant le transfert d'hydrure qui sera complet à la fin du chemin.

L'énergie libre de Gibbs associée à la réaction est positive (+ 8 kcal/mol) mais comme dans le cas des calculs impliquant une seule dimension, l'enzyme n'est pas dans son véritable état final (K162 déprotonée). Comme précédemment, on peut ajuster l'énergie des produits de la réaction en soustrayant 12 kcal/mol, ce qui donne un ΔG de réaction de -4 kcal/mol.

La comparaison entre les profils 1D et 2D nous démontre ainsi que la prise en compte d'une seule coordonnée de réaction est réellement insuffisante pour modéliser de manière correcte la réaction enzymatique. En effet, seul le transfert d'hydrure était contraint, le transfert de proton se faisait de manière opportuniste, i.e. lorsqu'étaient échantillonnées des structures pour lesquelles le transfert d'hydrure était au-delà du mi-transfert et pour lesquelles le transfert du proton était possible. Cette partie, modélisée par une contrainte adaptée $CR2$, est alors associée à une énergie qui n'était pas présente lors des simulations à une seule contrainte. Cependant, nous pouvons noter que les portions du chemin correspondant uniquement à une évolution de la coordonnée de l'hydrure sont bien représentées par le profil 1D, notamment au niveau de la barrière de rapprochement des réactifs qui présente la même valeur énergétique entre les deux types de calculs. Ce rapprochement des entités réactives mène à la formation d'un complexe pré-réactif appelé NAC (Near-attack conformer) dont il sera question plus tard.

IV.1.b) Analyse structurale

Les différentes analyses ont été réalisées selon deux protocoles. Tout d'abord, une première analyse sur la totalité de la surface d'énergie sera effectuée. En parallèle sera faite une seconde le long du chemin réactionnel. Dans celles-ci, il s'agira des valeurs moyennes associées à chacune des fenêtres considérées, projetées sur les valeurs exactes des contraintes.

Tout d'abord, l'analyse des distances impliquées dans les contraintes représentant les transferts montre de manière explicite les étapes associées au chemin réactionnel. La Figure 41 représente l'évolution de ces distances dans la définition des coordonnées représentant les transferts d'hydrure et de proton.

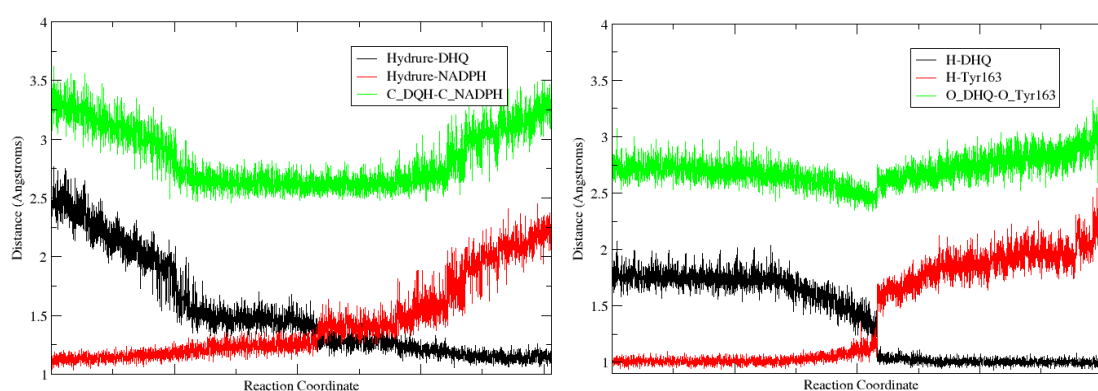


Figure 41 : Evolution des distances impliquées dans la définition des contraintes de calculs DM-QM/MM US 2D. A gauche sont représentées les distances composant la contrainte CR1, à droite celles de CR2. Sont ajoutées aussi les distances entre les atomes lourds concernés.

L'évolution des distances montre que la première étape de la réaction amène à un état dans lequel les entités DHQ et NADPH se sont rapprochées. Cela correspond à la formation du complexe pré-réactif de type NAC, que nous aborderons plus loin dans le chapitre. Nous pouvons voir que pendant cette phase, les distances associées à la contrainte représentant le transfert de proton restent constantes, ce qui confirme que cette partie du mécanisme ne fait intervenir que CR1. Lorsque le rapprochement est fini, c'est à dire que la distance entre les atomes de carbone est stabilisée à une valeur de 2,6 Å, la phase réactive commence. La distance hydrure-NADPH augmente alors pour atteindre une valeur de coordonnée d'environ 0,2 Å. A ce moment, la contribution de CR1 à la coordonnée de réaction est grandement diminuée, et on observe une évolution importante des distances associées à CR2. Suivant la contrainte, le proton se rapproche du substrat. Lorsqu'il est assez proche du DHQ, le transfert est initié et la distance H⁺-tyrosine augmente, mais de manière très rapide et brève dans le cas du proton, tandis que la finalisation du transfert dans le cas de l'hydrure est beaucoup plus

longue. Lorsque l'état de transition est franchi, les deux transferts s'effectuent. S'en suit une étape de relaxation où les partenaires s'éloignent après formation des produits. Les distances hydrure-NADPH et H^+ -tyrosine augmentent grandement tandis que les autres restent constantes, attestant de la fin des transferts vers le DHQ.

L'analyse de la distance $C_{NADPH}-C_{DHQ}$, entre l'atome de carbone du cofacteur donneur d'hydrure et celui du substrat qui va être réduit, montre une zone de stabilité extrêmement large juste avant l'initiation du transfert d'hydrure. Après la phase associée au rapprochement du NADPH et du substrat, la distance entre ces derniers reste remarquablement stable (environ 2,6 Å), afin de permettre le transfert d'hydrure (cf Figure 42). Cette distance reste constante tant que la totalité de la réaction ne s'est pas effectuée. Si nous analysons cette distance, en considérant maintenant toute notre surface explorée, cette zone est beaucoup plus grande, et existe même lorsque l'on en suit pas le chemin de la réaction. En fait, elle existe lors que les valeurs de la contrainte CR1 sont comprises entre -0,45 Å et 0,3 Å.

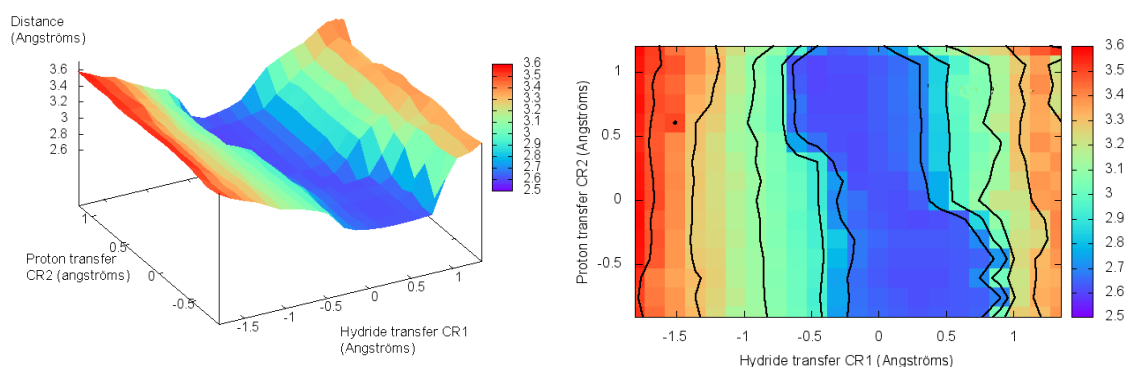


Figure 42 : Evolution de la distance entre les atomes de carbone du cofacteur et du substrat, le long de la surface d'énergie potentielle. A droite est représenté une vue du dessus de la surface présentée.

Cette zone recouvre toutes les fenêtres associées à la contrainte du transfert de proton. Ainsi, cette proximité entre le cofacteur et le substrat est une condition pour que le transfert d'hydrure puisse se réaliser. Ceci signifie que le rapprochement des deux réactifs constituerait une étape préliminaire nécessaire à la réaction enzymatique.

Ce comportement ne se retrouve pas lorsque l'on étudie l'évolution de la distance OTYR-ODHQ. En effet, cette distance suit l'évolution du transfert, les atomes se rapprochant puis s'éloignant en fonction du transfert du proton, mais de manière assez anecdotique. En analysant l'évolution de cette distance sur toute notre surface (voir Figure 43), les variations restent extrêmement faibles même si on observe, pour les dernières valeurs de la contrainte

représentant le transfert d'hydrure, que la distance devient très grande. Si la coordonnée atteint une valeur trop élevée, elle entraîne une modification structurale du site actif, car l'éloignement du NADPH et du DHQ devient important. Du coup, la structure du substrat change dans la cavité et explique cette distance beaucoup plus importante avec la tyrosine.

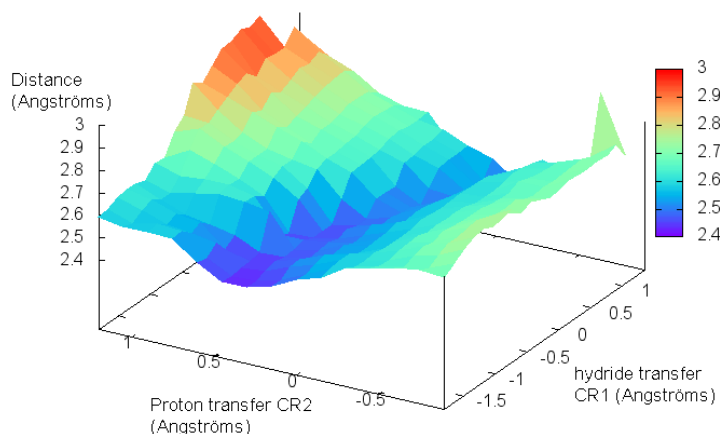


Figure 43 : Evolution de la distance OTyr-ODHQ durant toutes les fenêtres de simulation.

En Figure 44 sont représentés l'ensemble des distances associées aux transferts de proton de la tyrosine, la lysine et le NADPH.

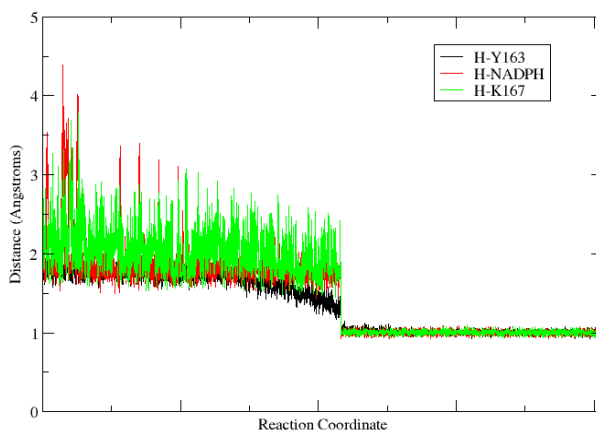


Figure 44 : Evolution des distances représentant les transferts de protons se déroulant lors de la réaction de DHQ dans DFR.

Aucune étape de rapprochement n'est observée pour les transferts du NADPH vers la tyrosine, et celui de la lysine vers le cofacteur. Cependant, on observe une remarquable synchronicité de tous les transferts de la cascade enzymatique.

Les interactions de type liaison hydrogène avec les résidus de reconnaissance G227 et N133 ainsi que les résidus S128 et T191, restent sensiblement constantes la durée de la réaction (respectivement d'une valeur moyenne de 1,8 Å, 2 Å, 1,85 Å et 2 Å). Elle sont par contre moins stables dans les cas extrêmes de la surface (notamment dans le même cas que la distance OTyr-ODHQ), ou l'on observe une perte des interactions à cause d'un changement de conformation de la molécule.

IV.2) Substrat n°2 : DHK (dihydrokaempferol)

IV.2.a) Analyse énergétique du profil

Dans le cas du dihydrokaempferol (DHK), le nombre de calculs est identique à celui du système avec le DHQ. Ainsi le profil a été obtenu dans le même espace et est représenté sur la Figure 45.

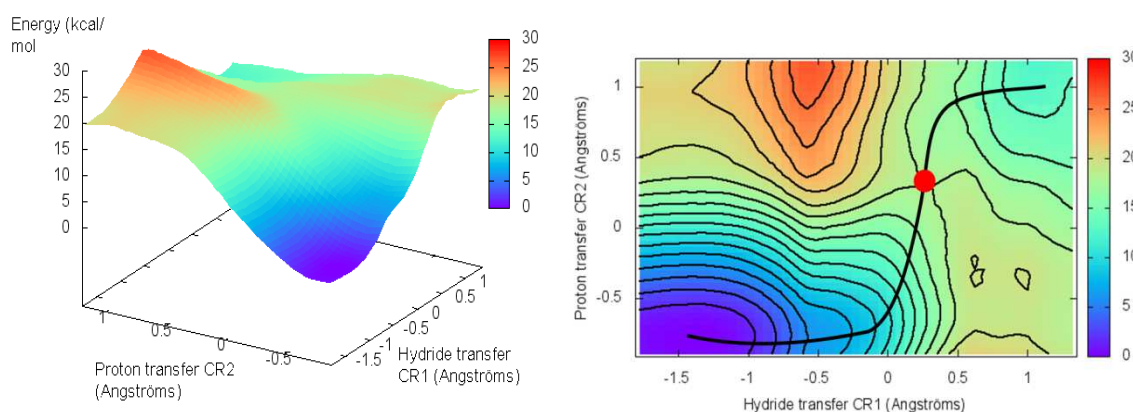


Figure 45 : Profil réactionnel obtenu pour le complexe ternaire enzyme/NADPH/DHK. La courbe noire désigne le chemin de plus basse énergie suivie par le mécanisme réactionnel. Le point rouge désigne l'état de transition.

Dans son ensemble, le profil est grandement similaire à celui obtenu dans le cas du substrat DHQ. Le profil présente un minimum global pour des valeurs de CR1 et CR2 négatives [-1,50;-0,75], correspondant aux réactifs de la réaction. Un point selle, correspondant à l'état de transition de la réaction, est de nouveau caractérisé. L'énergie nécessaire pour l'atteindre est de 17,9 kcal/mol, ce qui est très similaire au cas du DHQ. Les valeurs des contraintes CR1 et CR2 qui permettent d'identifier l'état de transition sont respectivement 0,30 Å et 0,30 Å. Nous observons à nouveau un état intermédiaire, qui ne concerne que le rapprochement des réactifs, à une énergie de 7 kcal/mol, dont nous parlerons par la suite. Enfin, l'état stable correspondant aux produits de la réaction est localisé au point

[1,05;1,20], avec une énergie de 12,5 kcal/mol. Comme dans le cas du DHQ, cet état correspond à une structure dans laquelle la lysine catalytique de l'enzyme n'est pas régénérée.

Les différentes observations apportées par le profil réactionnel impliquant DHQ sont applicables également sur cette surface d'énergie. Cependant, il apparaît que les deux contraintes que nous avons suivi sont moins décorrélées que dans le cas du DHQ. La première étape, correspondant au rapprochement des entités, ne fait intervenir à nouveau que la contrainte liée au transfert d'hydrure, de $CR1 = -1,50 \text{ \AA}$ à $CR1 = -0,30 \text{ \AA}$. A partir de cette dernière, le chemin fait intervenir principalement la contrainte représentant le transfert de proton, mais également, de manière plus importante que dans le cas du DHQ, la contrainte du transfert d'hydrure. Durant cette portion, $CR1$ varie de $-0,30$ à $0,45 \text{ \AA}$, et la contrainte $CR2$ de $-0,75$ à $0,60 \text{ \AA}$. L'état de transition est atteint pour la valeur $[0,30;0,30]$. De ce fait, il est considéré comme étant plus "tardif" sur la contrainte $CR1$ que dans le cas du DHQ d'après le postulat de Hammond, alors que la valeur de $CR2$ reste identique à celle du DHQ. Le fait le plus intéressant est que l'énergie d'activation associée à la réaction avec DHK est pratiquement identique à celle du DHQ. Enfin, après réaction, la dernière partie menant à la formation des produits fait intervenir de nouveau majoritairement $CR1$, avec un couplage de $CR2$ encore présent.

La différence entre les deux substrats est donc essentiellement thermodynamique, avec une énergie libre de Gibbs corrigée (après la reprotonation de la lysine) d'environ 0 kcal/mol pour le cas du DHK. Nous avons donc un retour des produits vers les réactifs plus favorables que dans le cas du DHQ, impliquant que la conversion de la molécule de DHK sera moins efficace. Ces résultats sont directement en accord, d'un point de vue qualitatif, avec les données expérimentales obtenues par Petit *et al.* [68]. Ainsi, notre modèle se montre prédictif sur la conversion de substrats par l'enzyme.

IV.2.b) Analyse structurale du mécanisme

Les analyses ont été réalisées le long du chemin réactionnel et également sur la surface complète. D'une manière générale, la plupart des observations notées dans le cas du DHQ se retrouvent ici (résidus impliqués dans des liaisons hydrogènes toujours présents le long du chemin). On retrouve notamment le même comportement des distances DHQ-hydrure et NADPH-hydrure, avec une zone correspondant uniquement au rapprochement des réactifs. Concernant la distance $C_{\text{NADPH}}-C_{\text{DHK}}$, on observe, comme dans le cas du DHQ, une vaste zone

dans laquelle la distance entre les deux atomes reste constante. Elle est cependant légèrement décalée vis-à-vis de ce qui est observé avec le DHQ car l'ET est plus tardif ici. Cette évolution est représentée sur la Figure 46.

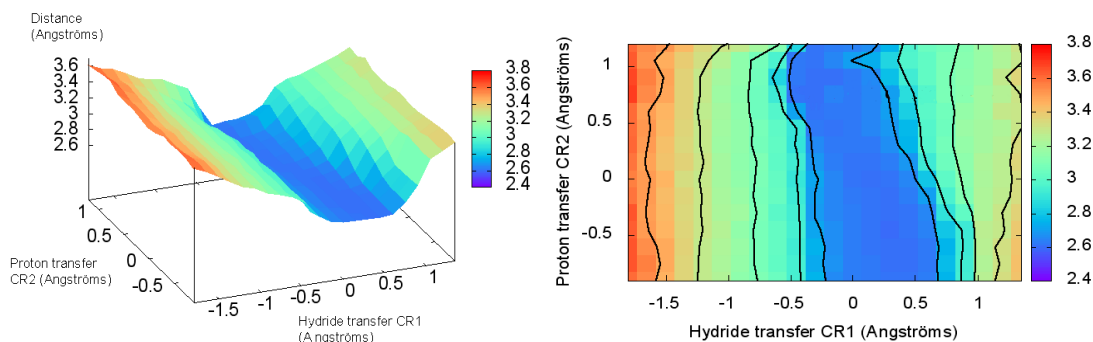


Figure 46 : Evolution de la distance C-C lors des différentes fenêtres calculées pour le cas du DHK. A droite est représentée la vue de dessus de cette surface.

La différence majeure provient du comportement des partenaires impliqués dans la cascade de transfert de protons. En effet, contrairement au cas précédent où il se déroulait de manière spontanée à la fin des transferts d'hydrure et du proton, nous voyons que le transfert n'intervient qu'en fin de réaction. Il n'est obtenu que dans les dernières fenêtres après le franchissement de l'état de transition tardif. L'évolution des distances associées à ces transferts est représentée sur la Figure 47.

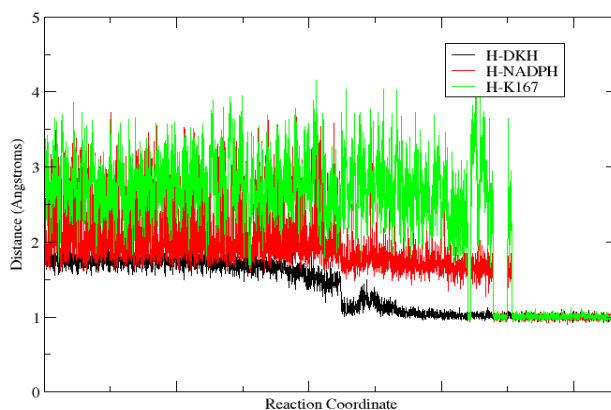


Figure 47 : Evolution des distances impliquant la cascade de transferts de protons.

Cette observation peut être expliquée par le fait que l'état de transition est plus tardif que dans le cas du DHQ. En effet, la contrainte représentant le transfert d'hydrure doit être plus avancée pour franchir l'état de transition, impliquant que la préparation du système à la réaction est, elle aussi, décalée. Ainsi, le cofacteur est préparé moins "rapidement" à la réaction, ce qui retarde le transfert de proton provenant de son groupement ose. Cette

préparation tardive est attribuée au fait que le DHK est plus mobile que le DHQ, phénomène imputé à l'absence d'une interaction du DHK avec l'enzyme (Une interaction avec le résidu de reconnaissance N133), à cause de l'absence de groupement hydroxyle en position 4'. Les analyses de RMSd des deux substrats (obtenues sur les simulations de DM-QM/MM sans contrainte), qui sont représentées sur la Figure 48, montrent cette mobilité accrue du dihydrokaempferol.

Rapidement, lors des simulations, on observe que le DHK subit une déviation plus importante que celle du DHQ. L'absence de liaison hydrogène supplémentaire verrouille de manière moindre le cycle B de la molécule. Par ailleurs, le décrochage observé du DHQ, après 600 ps de simulation de dynamique moléculaire QM/MM, correspond à une perte ponctuelle de l'interaction avec un résidu de reconnaissance (G227). Le substrat passe alors sur un mode d'interaction à une seule liaison hydrogène. Le cycle B est ainsi moins contraint, et on observe une valeur du RMSd environ égale à celle du DHK, représentant une mobilité équivalente. Nous observons que le mode d'interaction est de nouveau de deux liaisons hydrogènes après 820 ps de simulation, et ainsi la valeur du RMSd diminue car le cycle est de nouveau contraint. Cette mobilité accrue du DHK peut expliquer le caractère tardif de l'état de transition. Il faut en effet que l'hydrure soit transféré de manière plus avancée pour réaliser la réaction

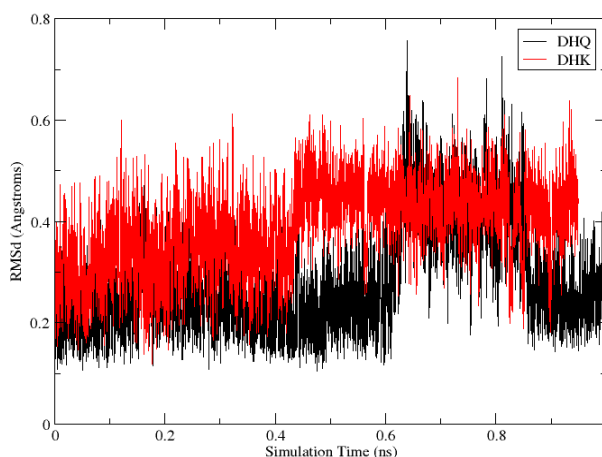


Figure 48 : Evolution du RMSd des substrats par rapport à la pose cristallographique.

Le reste des analyses structurales est très similaire entre le DHQ et le DHK, et c'est justement cette grande constance entre les deux substrats qui peut expliquer cette surprenante barrière d'activation identique pour les deux molécules.

V) Complexe pré-réactif (Near-attack conformation)

Les sites actifs des enzymes possèdent des caractéristiques particulières qui facilitent les réactions, par exemple en stabilisant les états de transition ou en permettant à trois réactifs d'être présents dans un même espace. L'hydrophobicité du site actif, les contraintes stériques ou d'autres propriétés découlant des types d'acides aminés constituant la cavité jouant souvent un rôle majeur dans la réactivité, notamment en abaissant la barrière énergétique à franchir. Depuis de nombreuses années, une hypothèse met en avant que la barrière d'activation de la réaction enzymatique peut être encore plus abaissée que prévu. La théorie du Near Attack Conformer (NAC), ou complexe pré-réactif, a été développée en 1976 par Thomas C. Bruice [154]. Selon cette hypothèse, l'enzyme, de par ses mouvements et sa dynamique, pourrait forcer son substrat à prendre des conformations particulières grâce aux différentes contraintes stériques associées à l'environnement enzymatique[155].

D'un point de vue structural, la conformation du substrat induite par la dynamique de l'enzyme se rapprocherait de la structure de l'état de transition, ce qui diminue de manière significative l'énergie nécessaire pour l'atteindre. Cette hypothèse pourrait venir en complément du concept généralement admis que l'on attribue une stabilisation de l'état de transition à des effets électrostatiques de l'environnement enzymatique[156]. Cette théorie a été grandement débattue au fil des décennies[82, 157, 158], mais elle est reconnue maintenant au même titre que celle de la stabilisation de l'état de transition. Ces deux effets coexistent afin que la barrière énergétique à franchir soit plus faible et ainsi que la réaction soit plus aisée.

V.1) DHQ (Dihydroquercétine)

Sur le profil réactif de nos deux substrats, nous observons une première montée en énergie amenant à une structure d'énergie environ égale à 6-7 kcal/mol par rapport aux réactifs (voir Figure 39, cercle rouge). Ce point apparaît lors de la première phase du chemin réactionnel, qui n'est dépendant que de la contrainte liée au transfert d'hydrure. Toutefois, un point attire l'attention. En effet, si nous décomposons, dans le cas du DHQ, la contrainte selon les deux distances qu'elle implique (distance H-DHQ et H-NADPH), il apparaît que seule la distance hydrure-DHQ voit sa valeur se modifier dans cette partie. En revanche, la distance H-NADPH n'est que très peu modifiée (voir Figure 49).

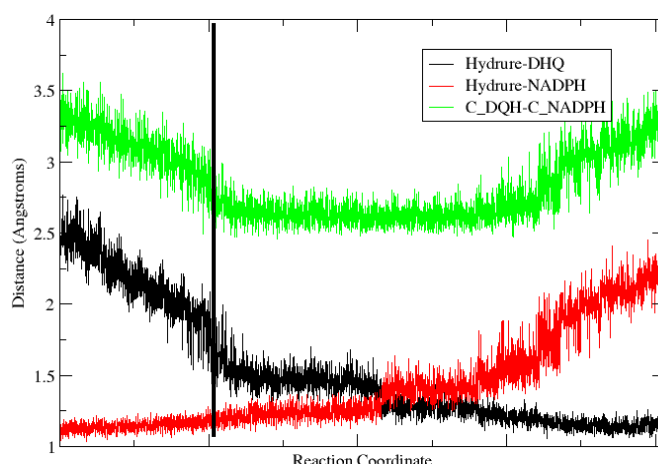


Figure 49 : Représentation de l'évolution des distances composant la contrainte du transfert d'hydrure durant la réaction. Le trait noir marque la fin de la première montée en énergie observée.

Cette étape caractérise donc un rapprochement des réactifs, (cofacteur NADPH et substrat DHQ) et il ne peut être considéré que le transfert d'hydrure ai déjà commencé. Les états échantillonnés lors de cette phase du mécanisme sont donc des conformations qui seront plus proches de la structure de l'état de transition. Ceci correspond à la définition des structures classées en tant que NAC. La distance hydrure-DHQ peut être considérée comme un critère structural principal pour la caractérisation de conformation de type NAC. La plupart du temps, les distances et angles d'attaque sont retenus pour la caractérisation des NAC [159]. Ici, l'angle de pyramidalisation de l'atome de carbone du substrat impliqué dans la réaction peut être également choisi comme critère, car l'hybridation de l'atome passe d'un sp^2 à un sp^3 . La variation de l'angle de pyramidalisation de l'atome de carbone est représenté sur la Figure 50.

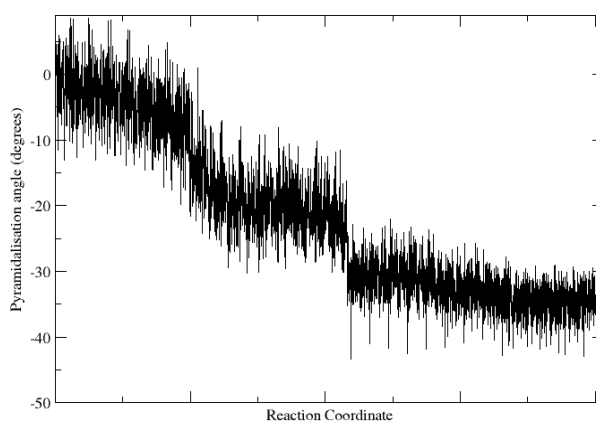


Figure 50 : Evolution de l'angle de pyramidalisation du carbonyle le long du chemin réactionnel

L'analyse du profil réactionnel concernant DHQ montre que l'angle de pyramidalisation de l'atome de carbone impliqué dans la réaction varie très peu lors de la phase de rapprochement des entités réactives. Nous pouvons remarquer toutefois que le carbone n'est pas totalement plan et présente une pyramidalisation légère. Lorsque les réactifs sont en contact, et ce même avant que la réaction en elle-même débute, la pyramidalisation de l'atome devient plus importante. Cela montre que cette première phase de rapprochement permet au final d'obtenir un complexe "préparé" à la réaction, avec un atome de carbone partiellement hybridé sp^3 (valeur de l'angle à environ -20°).

Concernant les valeurs de la distance hydrure-carbonyle et de l'angle de pyramidalisation pour les réactifs, leur valeur moyenne est respectivement de 2,48 Å et de $-1,44^\circ$ dans la fenêtre de calcul $[-1.35;-0.75]$. Nous pouvons comparer ainsi ces valeurs avec ce que nous observons lors de simulations de dynamique moléculaire "brute-force" de type QM/MM. L'analyse de ces simulations non biaisées montre que :

- Concernant la distance hydrure-carbonyle, la distance moyenne obtenue sur des simulations longues de plus de 1 ns est de l'ordre de 2,67 Å, ce qui est légèrement au dessus de la distance moyenne de l'état initial. Nous pouvons par contre noter que nous échantillonnons des états dans lesquels la distance est en deçà de la valeur moyenne (valeur minimale 1,91 Å). D'après les analyses, 28% des structures échantillonnées durant la simulation sont en dessous de cette valeur.
- Pour la pyramidalisation de l'atome de carbone subissant la réduction, la valeur moyenne est de $-2,76^\circ$. Cela correspond à une valeur déjà inférieure (l'atome de carbone est plus pyramidalisé) à celle rencontrée lors de notre simulation des réactifs de la réaction. La valeur maximale obtenue pour cet angle est de -16° . La Figure 51 représente la distribution des valeurs des critères de définition d'un NAC échantillonnées durant la simulation non biaisée.

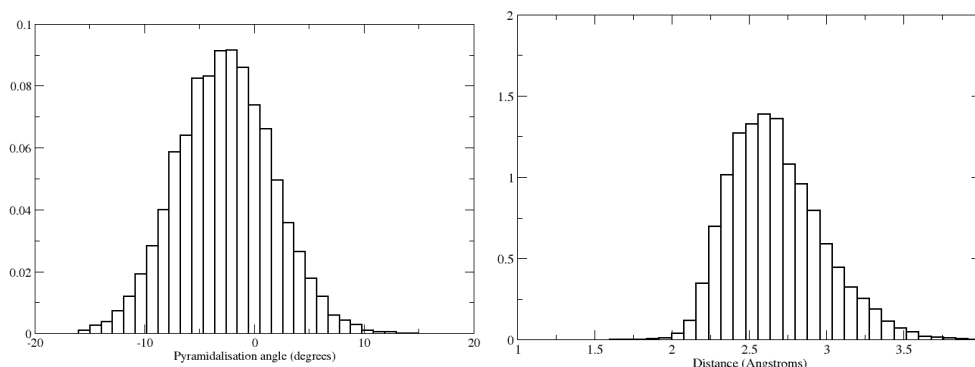


Figure 51 : Distribution des valeurs échantillonnées durant la simulation DM-QM/MM pour l'angle de pyramidalisation (à gauche) et la distance hydrure-DHQ (à droite).

L'échantillonnage de valeur de la pyramidalisation de l'atome de carbone impliqué dans la réaction laisse supposer que malgré l'implication de cet atome de carbone dans un cycle conjugué, les différents mouvements de l'enzyme sont susceptibles de « préparer » cet atome à une modification de son hybridation. De plus, notons que lors de l'étape de rapprochement, la pyramidalisation de l'atome reste très faible (valeur maximale -10°) alors que la distance entre le cofacteur et le substrat diminue fortement. Afin d'estimer ce que coûte ce processus de début de pyramidalisation, nous avons réalisé un scan de l'énergie en fonction de l'angle dièdre au niveau QM B3LYP/6-31+G*. Il s'avère que le changement d'hybridation (de 0° à -15°) ne coûte que très peu d'énergie, environ 1,9 kcal/mol, ce qui rend cette transition très facilement réalisable à température ambiante. Ainsi, le processus de pyramidalisation ne devrait pas être un frein à l'obtention de structure de type NAC. Pour la suite, seule la distance hydrure-DHQ est retenue pour la recherche de conformations que l'on peut considérer comme pré-réactives, i.e. pour lesquelles le rapprochement est effectif, ou tout du moins bien avancé. Nous considérerons également en contrainte supplémentaire (mais pas comme un critère structural nécessaire) que la contrainte représentant le transfert de proton se situe dans l'intervalle que nous avons défini sur notre surface d'énergie potentielle (Valeur max : $-0.90 \text{ \AA}$)

Sur toute la simulation (générant 10000 structures différentes), les points pouvant être associés à des structures de type NAC ont été estimés à un nombre de 1000, soit 10 %. Nous pouvons donc projeter les points échantillonnés en simulation de DM QM/MM les plus pertinents sur notre profil réactionnel afin d'estimer l'énergie gagnée sur la barrière totale. Les points sont représentés sur la Figure 52.

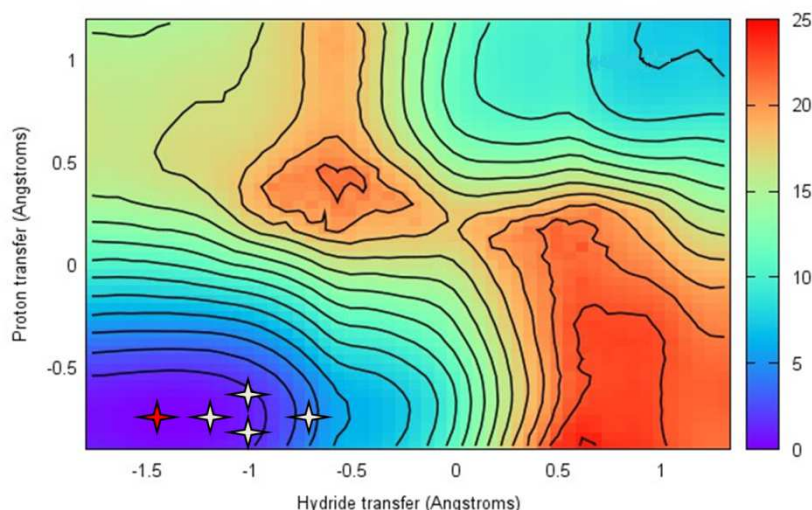


Figure 52 : projection des points échantillonnés en simulation de DM-QM/MM sur le profil réactionnel de DHQ. Ne sont représentés que les points représentatifs de tous les états observés. L'étoile rouge représente le point décrivant l'état initial de la réaction.

Le point le plus élevé est identifié à une position $[-0.80; -0.75]$, ce qui correspond à une énergie d'environ 3,5 kcal/mol (la valeur de l'angle de pyramidalisation est de $-6,26^\circ$). Les mouvements de l'enzyme sont donc capables de mener le substrat à une structure pré-réactive qui permet de réduire la barrière d'activation totale de 3,5 kcal/mol. Même si nous ne l'avons pas échantillonné dans nos simulations de dynamique moléculaire QM/MM, nous pouvons assumer, au vu des résultats obtenus, que l'enzyme doit être capable de produire un complexe pré-réactif correspondant à l'état intermédiaire dans lequel le rapprochement entre les deux réactifs est maximal (position $[-0.30; -0.75]$; énergie de 6-7 kcal/mol), permettant une diminution non négligeable de la barrière réactionnelle totale. Nous pouvons donc dire que l'enzyme catalyse la réaction et l'optimise encore plus par un effet de type NAC.

V.2) DHK (Dihydrokaempferol)

Dans le cas de la surface d'énergie associée à la réduction du DHK, nous observons également un état intermédiaire, correspondant uniquement au rapprochement du cofacteur et du substrat. De la même manière que pour la réaction avec la DHQ, cette première étape n'implique que la diminution de la distance hydrure-substrat mais la distance cofacteur-substrat reste stable. Ainsi, nous avons réalisé des simulations de dynamique moléculaire couplées avec les méthodes hybrides QM/MM, pour essayer de voir si nous échantillonnons de manière aussi importante les états pouvant être caractérisés en tant que NAC. La moyenne sur toute la simulation de la valeur de la distance hydrure-substrat est de 2,72 Å, ce qui est

très proche de la valeur obtenue pour le cas de la dihydroquercétine (valeur moyenne obtenue lors de l'échantillonnage de la fenêtre associée aux réactifs : 2,48 Å). De même, la distribution des valeurs échantillonnées est similaire de celle obtenue avec le DHQ, comme le montre la Figure 53.

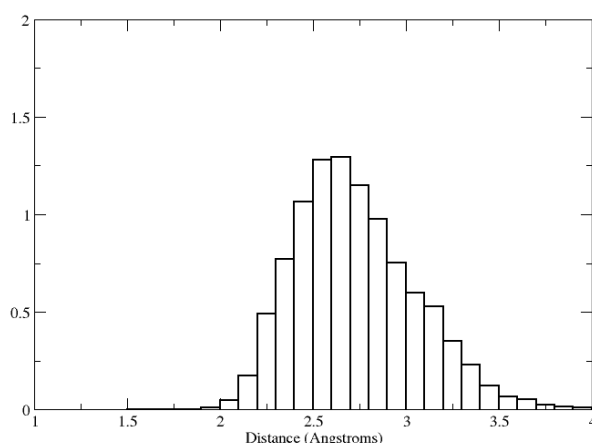


Figure 53 : Distribution des valeurs échantillonnées lors de la simulation de type DM-QM/MM de la distance hydrure-DHK.

En ce qui concerne l'angle de pyramidalisation, le même type de comportement est observé que dans le cas du DHQ. Ce critère peut être également négligé pour la caractérisation d'état comme étant de type NAC. Ainsi, 520 états sur les 10000 obtenus dans la simulation peuvent être considérés comme des Near-Attack Conformer (contrainte représentant le transfert d'hydrure supérieure à -1,35 Å). Sur le même temps de simulation, cela représente une population deux fois moins importante de NAC échantillonnées comparé à la simulation avec la DHQ. Statistiquement parlant, la formation de ce type de structure pré-réactive serait donc moins favorisée pour ce substrat. Comme dans le cas du DHQ, nous pouvons projeter les points échantillonnés les plus intéressants sur la surface d'énergie potentielle associée au mécanisme réactionnel impliquant le DHK, afin d'estimer l'énergie gagnée sur la barrière. Ces points sont représentés sur la Figure 54.

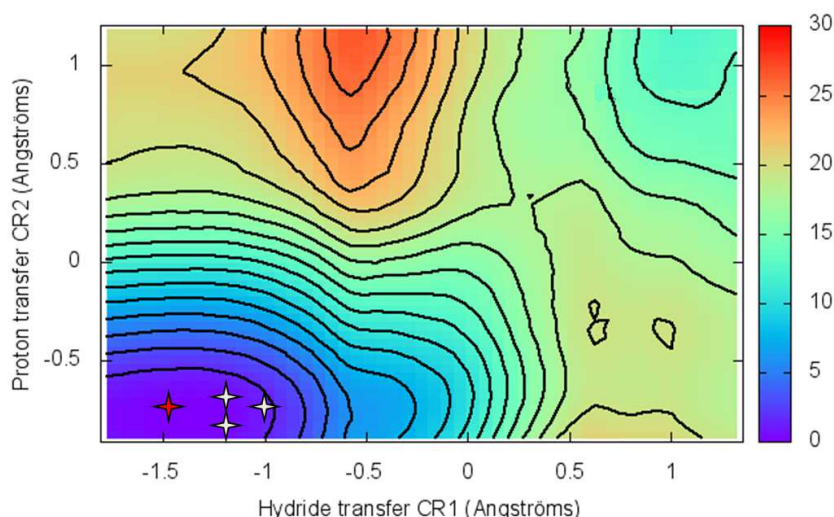


Figure 54: Points échantillonnés dans les simulations DM-QM/MM avec le DHK, et projetés sur le profil .

De manière remarquable, nous voyons que les points échantillonnés caractérisent des états énergétiquement moins élevés que dans le cas du DHQ. Nous observons ainsi une diminution du nombre de points pouvant être caractérisés en tant que NAC, et un gain d'énergie sur la barrière moins important. La raison pour laquelle cet échantillonnage est moins important réside dans l'absence d'une deuxième interaction avec le résidu de reconnaissance R133. En effet, cette interaction fixe de manière plus importante le cycle B dans le cas du DHQ que dans celui du DHK. Si le cycle est moins bien figé, il est de ce fait plus mobile (voir la Figure 48) dans le site actif de l'enzyme, ce qui implique que le positionnement est moins favorable envers la réaction chimique, rendant plus difficile l'échantillonnage d'états pré-réactifs. Ces observations peuvent constituer, en plus de la différence de stabilisation des produits de la réaction défavorable dans le cas du DHK, un autre effet qui pourrait expliquer la différence des taux de conversion entre DHQ et DHK reportée dans l'article de Petit *et al.* [68].

VI) Conclusion

Les enzymes possèdent un rôle très important dans le vivant, de par leur propriétés réactives permettant la transformation de composés. Dans le cas particulier de la biosynthèse des flavonoïdes, la réaction catalysée par l'enzyme DFR constitue une étape majeure de la biosynthèse, puisqu'il s'agit de la dernière étape commune menant à la production des anthocyanes et des proanthocyanidines, deux familles d'intérêts biologiques. Dans le cadre de notre métabolon, cette protéine a été choisie comme pivot de notre étude, nous avons donc

décidé d'en étudier la réactivité, par l'application de méthodes hybrides de type QM/MM MD couplées au protocole d'Umbrella Sampling. Cette combinaison de méthodes constitue l'état de l'art pour l'étude des mécanismes enzymatiques. Ici, deux contraintes ont été nécessaires pour représenter la coordonnée de réaction, une représentant le transfert d'un hydrure et une autre modélisant le transfert du proton provenant de la tyrosine.

La réactivité de l'enzyme a été étudiée pour deux substrats : la dihydroquercétine et le dihydrokaempferol, pour lesquels des données expérimentales existent (le taux de conversion de DHQ est plus important que celui de DHK). Les surfaces d'énergie potentielle obtenues pour les deux substrats montrent que la barrière d'énergie est identique dans les deux cas de figure. Seule la différence de stabilisation énergétique des produits de la réaction est notable, avec une enthalpie libre de Gibbs négative pour la conversion du DHQ, et nulle dans le cas du DHK. Nous observons également la présence d'un état intermédiaire dans les deux cas, qui correspond à une étape de rapprochement du cofacteur et du substrat. Cette portion d'énergie n'enclenche aucun phénomène réactif. Nous avons identifié cet état comme étant un NAC (Near Attack Conformer), c'est à dire un état pré-réactif qui s'approche de l'état de transition de la réaction et permet la diminution de la barrière d'activation de la réaction. De plus, les simulations de dynamique moléculaire QM/MM nous permettent échantillonner des points assimilés à des complexes pré-réactifs. Cela implique que l'enzyme peut atteindre ces états de manière naturelle, diminuant la barrière d'activation pour la réaction enzymatique. Nous observons ainsi que la probabilité d'échantillonner ces états de type NAC est 2 fois plus faible dans le cas du DHK que dans celui du DHQ.

Ainsi, la différence de conversion entre le DHQ et le DHK est due à l'effet conjugué du nombre d'états pré-réactif échantillonnés, ainsi qu'à une stabilisation moins bonne des produits de la réaction. L'absence d'une liaison hydrogène importante avec le résidu N133 est à l'origine de la déstabilisation importante du produit. Quand à l'échantillonnage des structures de type NAC moins probable pour le DHK, il est associé à une plus grande mobilité de ce dernier, de par l'absence d'une interaction supplémentaire (celle du résidu N133), qui rendrait le composé plus mobile.

Dans le cadre de l'étude de complexes multienzymatiques, la compréhension des mécanismes enzymatiques est nécessaire pour comprendre toutes les propriétés intrinsèques au métabolisme. Afin de parfaire cette étude, la prochaine étape serait d'étudier le mécanisme enzymatique de DFR lorsque celle-ci est en interaction avec une ou plusieurs protéines

partenaires. L'idée est de mesurer un possible effet allostérique des interactions protéine-protéine sur la réaction (par exemple, un changement structural du site actif, modifiant la barrière d'activation).

Chapitre 3 : Modèles DFR-LAR et "Substrate channeling"

Les voies biosynthétiques sont constituées d'une série de réactions chimiques ou biochimiques reliées par leurs intermédiaires. Au sein d'une cellule, les enzymes catalysent la biosynthèse de métabolites, et le produit d'une réaction devient le substrat de la réaction ultérieure. Il a été suggéré que les enzymes impliquées dans une voie métabolique donnée peuvent s'associer sous forme de "super-complexe" nommé métabolon[35, 48, 160, 161]. Ce complexe d'enzymes améliore l'efficacité des voies métaboliques par un phénomène appelé "substrate channeling": la proximité des sites actifs des enzymes autorise la diffusion du produit de la première enzyme vers le site actif de l'enzyme suivante dans la cascade biochimique et évite ainsi une re-solvatation du composé[162]. Les métabolons maintiennent et accroissent le flux métabolique en concentrant les intermédiaires réactionnels dans l'environnement proche des enzymes[35, 48, 61]. Ces "super-complexes" sont généralement assemblés par associations transitoires de complexes protéine-protéine dans lesquels certaines enzymes sont ancrées à la membrane pour améliorer la stabilité et la durée de vie du métabolon[57, 58, 101]. Les preuves expérimentales de l'existence de métabolon sont difficiles à obtenir[49]. La modélisation moléculaire peut apporter des premiers éléments d'étude du phénomène de "substrate channeling" survenant dans ces complexes protéine-protéine transitoires, et ainsi proposer un modèle au niveau atomique des différentes phases de ce processus de transfert.

Plusieurs approches de modélisation moléculaire seront utilisées pour caractériser les interactions protéines-protéine et détailler les phénomènes dynamiques impliqués dans la diffusion du substrat d'une enzyme à une autre:

- la proximité des sites actifs des enzymes est nécessaire pour faciliter la diffusion rapide d'un métabolite. Il est donc nécessaire de connaître les portes d'entrée/sortie des sites actifs de chaque enzyme. Dans un premier temps, des simulations de dynamique moléculaire permettant de visualiser les phénomènes d'association/dissociation du complexe enzyme/substrat seront donc mises en œuvre.
- Dans un deuxième temps, des approches de docking protéine-protéine permettront de générer les structures potentielles des complexes multienzymatiques. Pour éliminer les faux positifs dans les solutions obtenues, nous utiliserons les contraintes structurales basées sur la proximité des sites actifs des enzymes.
- Enfin, des simulations de dynamique moléculaire d'un système enzyme-enzyme-métabolite-cofacteurs seront mises en œuvre pour tenter de visualiser la diffusion du métabolite entre deux enzymes et de détailler les forces mises en jeu.

Dans la voie métabolique des flavonoïdes, la dihydroflavonol-4-réductase (DFR) est considérée comme une enzyme centrale pour la biosynthèse des anthocyanes et des procyanidines[42, 68, 71]. La leucoanthocyanidine réductase 1 (LAR) est l'enzyme suivante dans la voie menant aux proanthocyanidines[30, 74]. Il a été récemment prouvé que DFR est physiquement reliée à d'autres enzymes impliquées dans la biosynthèse des flavonoïdes, comme la chalcone synthase (CHS) et la flavanol synthase (FLS1)[49, 98, 160]. Dans ce chapitre, nous chercherons à proposer un modèle tridimensionnel pertinent d'un complexe multienzymatique faisant intervenir DFR et LAR. Nous apporterons pour la première fois un point de vue à l'échelle moléculaire sur le phénomène de "*substrate channeling*" entre ces deux enzymes (LAR et DFR).

L'association de deux enzymes pour créer un complexe moléculaire à deux corps a fait l'objet de nombreuses études. Toutefois, ici, une contrainte supplémentaire vient s'ajouter à la prédiction des zones d'interactions. En effet, il est nécessaire que le complexe formé soit compatible avec un processus de "*substrate channeling*" depuis le site actif de la première enzyme vers celui de la suivante. Afin que ces chemins d'accès ne soient pas obstrués lors de la définition du complexe, il est donc nécessaire de connaître ces voies. C'est l'objet des calculs décrits ci-dessous qui ont été appliqués sur des enzymes individuelles sans interactions.

I) Identification des chemins d'accès aux sites actifs des enzymes

I.1) Protocoles de simulation

Le protocole de simulation de dynamique moléculaire détaillé dans le chapitre méthodologie (page 93), est utilisé ici pour relaxer la structure du complexe enzyme-substrat-cofacteur. Par la suite, chacun des complexes enzyme/cofacteur/substrat a été soumis à des simulations de type RAMD pour identifier les chemins d'entrée/sortie possibles du site actif enzymatique.

Bien que l'on ne puisse pas considérer ces simulations comme pouvant refléter un chemin réversible d'entrée/sortie, il nous est apparu qu'appliquer ce type de protocole à un substrat dans une cavité enzymatique pouvait nous apporter de précieuses informations sur les voies d'accès entre phase solvant et cavité. L'utilisation de méthodes de type brute force nous

permettront de valider les résultats obtenus par les simulations RAMD. Concernant les paramètres, une force permettant d'accélérer le substrat dans une direction aléatoire est appliquée tous les 40 pas de simulation. La valeur de cette accélération est de 0,1 kcal/mol/Å/g. 30 simulations différentes ont été réalisées dans le but d'améliorer la statistique des résultats obtenus. La graine permettant de générer cette force supplémentaire a elle même été fixée aléatoirement pour s'assurer du caractère unique des simulations. Le calcul s'arrête automatiquement dès que la distance entre le centre de masse du substrat et celui de l'enzyme dépasse 40 Å, ce qui nous assure d'un positionnement du substrat hors d'une cavité enzymatique.

I.2) Résultats pour l'enzyme DFR

I.2.a) Simulations de type RAMD du complexe enzyme-cofacteur-substrat

Les simulations RAMD ont mis à jour 3 voies d'accès possibles pour la sortie du site actif de DFR. Ces sorties sont indiquées sur la Figure 55. Les sorties ont été échantillonnées respectivement 12, 10 et 8 fois. La statistique d'échantillonnage de chacune des voies d'accès ne montre pas de différence significative qui permettraient de privilégier un chemin plutôt qu'un autre. Chacune des voies d'accès implique différents résidus :

- Sortie 1 : la molécule de DHQ passe à proximité des résidus M88 et P204. La distance moyenne entre ces résidus est de 6,7 Å, ce qui montre que ce chemin est la plupart du temps ouvert lors des simulations de type RAMD.
- Sortie 2 : le substrat franchit les résidus V290, S289, L139 et P140 avant d'atteindre la phase solvant. Les interactions hydrophobes existante entre ces résidus sont perdues lors de la sortie du DHQ.
- Sortie 3 : les résidus impliqués sont T159, M158, S221 et I222. De même, la sortie de la molécule provoque une ouverture de la porte, impliquant que les résidus cités n'interagissent plus les uns les autres.

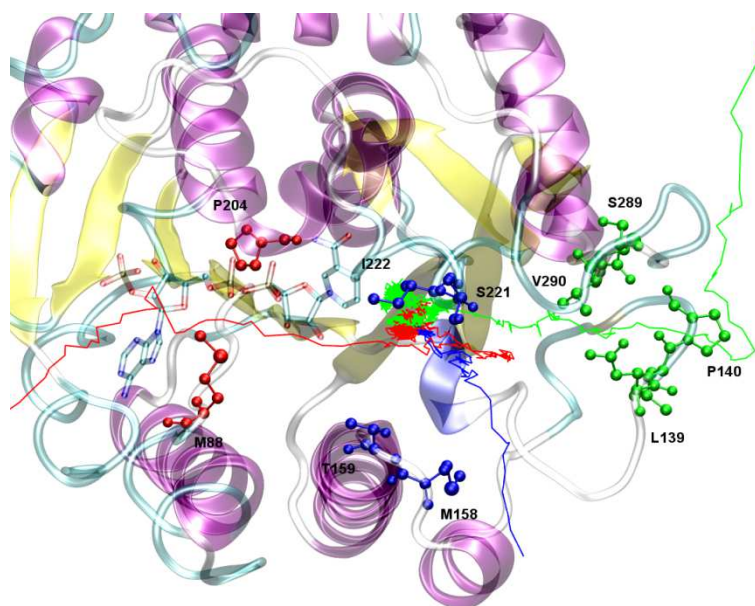


Figure 55 : Détail des 3 voies d'accès détectées par les simulations de type RAMD. La sortie 1 est représentée en rouge, la sortie 2 en vert et la sortie 3 en bleu. Les résidus impliqués dans chacun des chemins sont également représentés de la même couleur.

1.2.b) Simulations de DM du complexe enzyme-cofacteur-substrat

Pour appuyer les résultats des calculs de type RAMD, des simulations de dynamique moléculaire brute force ont été réalisées afin d'observer le phénomène d'association et d'entrée du substrat dans le site actif de l'enzyme. Ces simulations ont été réalisées sur un système DFR-NADPH entouré de plusieurs ligands positionnés dans la phase solvant. Les temps de simulation sont de l'ordre de quelques centaines de nanosecondes (100-600 ns). Une contrainte entre le centre de masse de chaque ligand et le centre de masse de la protéine a été appliquée pour contraindre les molécules à être à une distance inférieure à 35 Å.

Un premier test avec 10 ligands autour de l'enzyme a été effectué. Les simulations montrent que les 10 ligands ont tendance à interagir les uns avec les autres (par interaction de type π - π), et ainsi former des agrégats de molécules ne pouvant plus accéder au site actif. Nous observons néanmoins les entrées d'un (ou parfois deux) ligand(s) au sein de la cavité, par la voie d'accès 1 déjà identifiée en RAMD. Ainsi, pour éviter ces phénomènes d'agrégation, les simulations ont été relancées avec 5 ligands autour de l'enzyme, chacune avec des conditions initiales différentes. Les observations de cette simulation de plus de 700 ns montrent que le substrat pénètre dans le site actif par une voie d'accès pouvant être assimilée à la voie d'accès 3 (cf Figure 55). L'organisation du site actif en fin de simulation montre que le ligand est dans une position très proche de la pose cristallographique observée

dans le fichier PDB 2C29 (cf Figure 56). Bien que le RMSd du ligand, calculé par rapport à sa position cristallographique, soit relativement important (4 Å), on retrouve la plupart des résidus de la position cristallographique en interaction avec la molécule (T191, N133, I222, V192...). Nous pensons qu'une simulation plus longue aurait permis au cycle A de la molécule de se repositionner correctement pour interagir avec la tyrosine 163 via une interaction de type liaison hydrogène avec le groupement carbonyle.

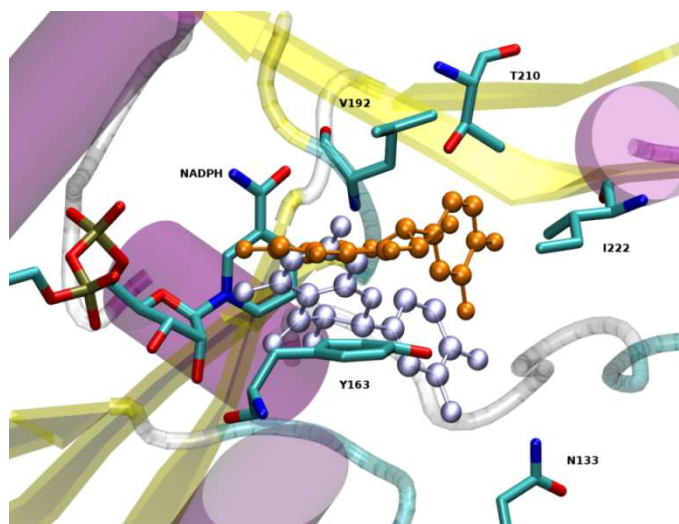


Figure 56 : Position finale échantillonnée lors des simulations de dynamique moléculaire longues avec 5 ligands. La pose obtenue par les calculs est représentée en orange. En bleu clair est représentée la position cristallographique.

1.2.c) Simulations DM sur le complexe enzyme-cofacteur-produit

Pour comparer le comportement du substrat et du produit de l'enzyme, et éventuellement identifier d'autres voies d'accès au site actif, nous avons réalisé des simulations de dynamique moléculaire brute force sur un système enzyme-cofacteur(NADP⁺)-produit (Leucocyanidine, noté LCC). Le produit est placé dans le site actif, calqué sur la position du substrat dans la structure cristallographique. Trois simulations de dynamique moléculaire, avec des conditions initiales différentes, ont été réalisées sur une échelle de temps allant de 200 ns à 1,2 μ s. Nous observons la sortie, presque totale, du produit de la réaction par un chemin pouvant être identifié comme étant la voie d'accès 1 (cf Figure 55), dans un temps d'environ 750 ns. La libération complète de la molécule est difficile à être échantillonnée du fait de la création d'interactions de type liaison hydrogène entre le groupement phosphate du cofacteur et les deux groupements hydroxyles du cycle B du produit. Le chemin suivi par le substrat pour sortir de la cavité est représenté sur la Figure 57.

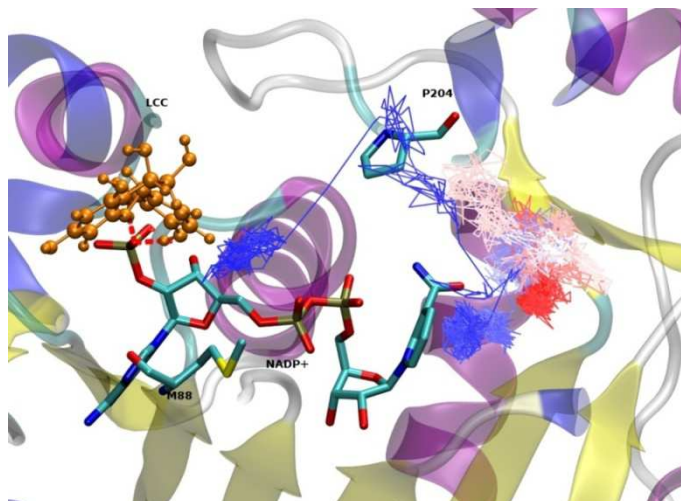


Figure 57 : Représentation du chemin suivi lors de la libération du produit de l'enzyme DFR. Les atomes du cofacteur sont représentés ainsi que le LCC en orange. Les résidus définissant une porte menant à la cavité (M88 et P204) sont également mis en avant.

En conclusion de cette partie, les simulations de dynamique moléculaire brute force confirment les calculs RAMD. Nous observons l'association enzyme/substrat par deux entrées possibles, sans forcer le ligand à pénétrer dans l'enzyme ainsi que la libération du produit par l'une des voies déjà identifiée. Ces chemins d'entrée/sortie serviront de contraintes structurales dans le cadre de l'interaction protéine-protéine entre DFR et LAR.

I.3) Résultats pour l'enzyme LAR

I.3.a) Simulations de type RAMD

Comme pour DFR, l'enzyme LAR a été préparée pour la réalisation de simulations de type RAMD. 30 simulations ont été réalisées avec les mêmes paramètres que ceux définis précédemment. 25 simulations ont abouti, aucune sortie n'ayant été échantillonnée durant les 5 autres simulations. Cette fois, une seule sortie de l'enzyme a été détectée. Cet accès est représenté sur la Figure 58. L'entrée de la cavité est définie par les résidus M136, I263, I171 et H172. On observe lors du phénomène de dissociation du complexe la perte d'interaction entre I263 et I171, résidus qui fermaient la cavité du complexe ternaire. La rupture de cette interaction permet au ligand de sortir librement. Nous pouvons également noter la perte d'interaction entre le ligand et les résidus M136 et H172.

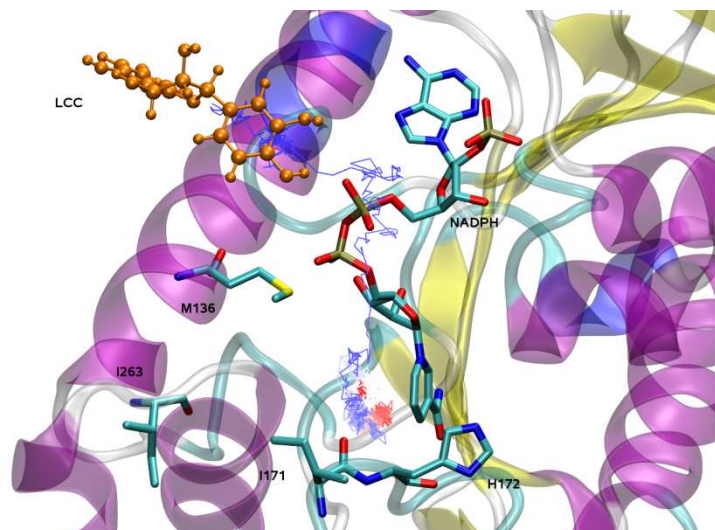


Figure 58 : Sortie du site actif identifiée par les simulations de type RAMD. En orange est représenté le ligand, et en *licorice* sont représentés le cofacteur, et les résidus définissant l'entrée de la cavité. Le chemin de diffusion est également montré.

I.3.b) Simulation DM sur le complexe LAR-cofacteur-substrat

Afin de confirmer ce qui a été observé dans les simulations RAMD, des simulations de dynamique moléculaire de type "brute-force" ont été réalisées avec 5 ligands positionnés autour de l'enzyme dans la boîte de solvant. 4 simulations de DM, avec des conditions initiales différentes, ont ainsi été effectuées afin d'observer le phénomène d'association enzyme/ligand. La durée totale de ces simulations est de 850 ns. Nous n'avons cependant jamais observé le phénomène d'entrée d'un ligand dans le site actif de LAR. En effet, la plupart des molécules interagissent rapidement à la surface de l'enzyme par liaisons hydrogène et interactions hydrophobes. Les temps de résidence des ligands à la surface de l'enzyme sont beaucoup plus importants que pour DFR. Ceci peut s'expliquer par une présence plus marquée de résidus polaires à la surface de LAR. Une analyse de la composition des acides aminés accessibles au solvant de LAR et DFR indique une fraction de résidus polaires légèrement plus importante pour LAR (40% vs 35%). Pour optimiser la voie de biosynthèse des flavonoïdes, la formation d'un métabolite impliquant une proximité des sites actifs des deux enzymes LAR et DFR a du sens. Il permettrait une diffusion plus rapide des métabolites d'une enzyme à une autre en diminuant les temps de résidence à leur surface. La seule contrainte structurale utilisée sur LAR pour le protocole de docking protéine-protéine sera la voie d'accès définie par les simulations de type RAMD.

II) Prédiction de la structure du complexe protéine-protéine DFR-LAR

II.1) Protocole de simulation

Le logiciel de docking protéine-protéine ATTRACT[104, 106] est utilisé pour la prédiction des complexes DFR-LAR. Une vue générale du protocole de docking est détaillée dans la partie méthodologique (page 74). Dans le cas présent, l'enzyme DFR a été choisie comme étant la protéine "réceptrice", LAR est donc la protéine "ligand". Une grille de points de départ (permettant de positionner le centre de masse de la protéine dite "ligand") a été construite autour de DFR avec un espace entre chaque points de 10 Å. Un total de 436 points de départ a été généré et pour chaque position sur cette grille, 258 rotamères de LAR seront considérés. Ainsi plus de 110000 positions initiales du complexe protéine-protéine sont minimisées sur la base du champ de force développé spécifiquement pour ATTRACT. La résolution gros grain utilisée ici permet d'évaluer rapidement toutes les solutions afin d'échantillonner au mieux les interfaces protéine-protéine. Les 2000 structures de plus basse énergie ont été analysées et les solutions similaires ont été regroupées sous forme de cluster. Chaque cluster est composé de structures du complexe avec un RMSd de moins de 1 Å sur la totalité du système.

II.2) Résultats du docking DFR-LAR

Les 10 clusters de plus basse énergie (allant de -16,2 à -14 RT) ont été analysés d'un point de vue structural. Pour trier ces solutions, nous avons choisi d'appliquer les critères structuraux définis précédemment. La distance entre les résidus H172 de LAR et T159 de DFR a été calculée pour vérifier la proximité des voies d'accès des sites actifs des deux enzymes. Seul le résidu T159 a été choisi pour DFR car il est commun aux deux chemins d'entrée/sortie de l'enzyme échantillonnés dans les simulations brute force précédentes. Les valeurs de distance entre les carbones C α des deux résidus s'échelonnent entre 16 et 65 Å.

Les structures choisies doivent ainsi avoir une distance entre les deux résidus faibles (gage d'une proximité des sites actifs) et une énergie d'interaction entre protéine la plus forte possible. Dans notre cas, le cluster 7 se démarque largement des autres, avec une énergie d'interaction forte ($E=-14,2$ RT) et la plus faible distance mesurée (16 Å). Nous pouvons également sélectionner le cluster 4, qui présente la deuxième plus petite distance ($d=30$ Å) et

une énergie plus forte que le cluster 7 ($E = -14,5$ RT). Sur le reste des 10 meilleurs clusters, aucune autre structure ne peut être retenue car le critère structural n'est pas respecté (distance supérieure à 30 Å, ce qui représente la limite de connexion entre les sites actifs des enzymes).

Si nous étendons l'analyse aux 100 meilleurs clusters, les conclusions sont similaires. Nous retrouvons deux complexes assez proche du cluster 4 d'un point de vue structural et deux autres similaires au cluster 7. Aucune autre solution de docking ne satisfait le critère structural et ne sera retenu. La structure de chacun des complexes choisis est représentée sur la Figure 59 et nous déciderons d'appeler par la suite le cluster 4, S1 et le cluster 7, S2 pour plus de clarté.

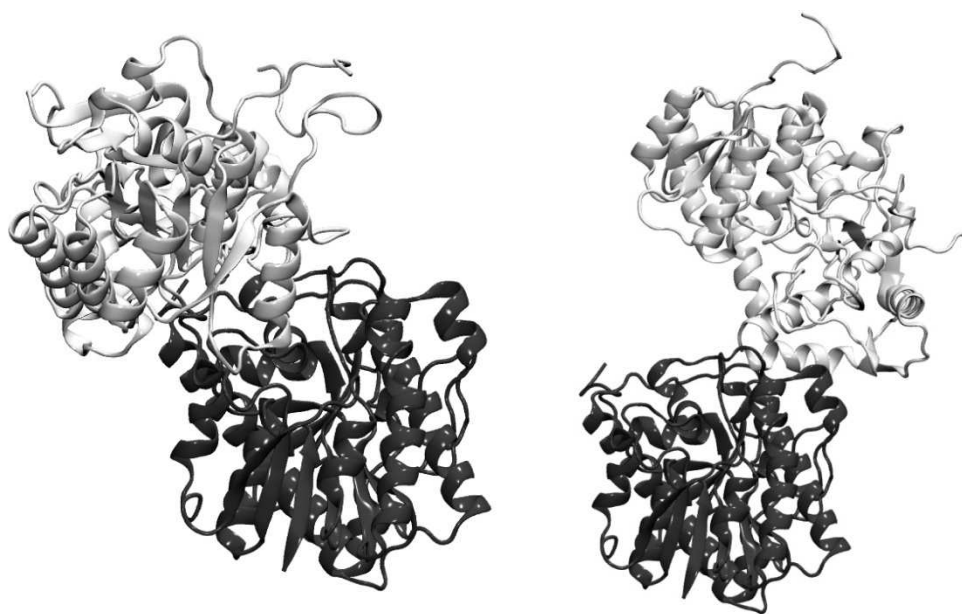


Figure 59 : Structures protéine-protéine obtenues par le protocole ATTRACT. LAR est représentée en blanc tandis que DFR est en noir. A gauche est représenté le cluster 4 et à droite le cluster 7 (appelé respectivement S1 et S2 par la suite).

II.3) Analyse structurale des complexes DFR-LAR retenus

Une interface protéine-protéine se décompose en deux zones distinctes, que sont le cœur de l'interface et sa périphérie[163]. Dans le cas des interactions transitoires protéine-protéine, la première zone est relativement compacte, majoritairement composée de résidus hydrophobes et ne contient que quelques molécules d'eau jouant un rôle structurant. La seconde zone est partiellement accessible au solvant et la fraction de résidus polaires y est plus importante. Les deux zones sont représentées sur la Figure 60.

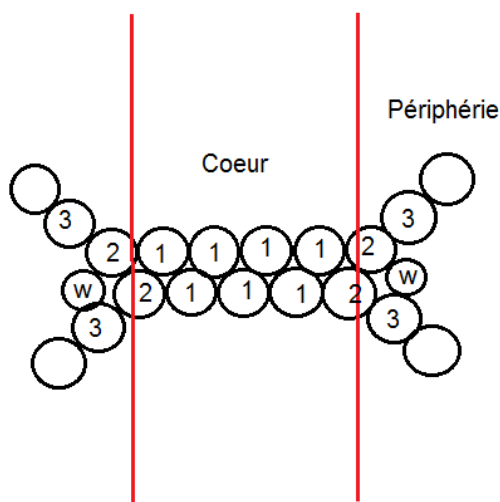


Figure 60 : Détail d'une interface protéine-protéine transitoire. Le cœur de l'interaction est caractérisé par une compacité élevée des résidus, avec des interactions majoritairement hydrophobes. Les résidus sont notés 1 sur le schéma. La périphérie de l'interface est définie par des interactions plus hydrophiles. Elle est composée des résidus de type 2 (interagissant avec le cœur mais partiellement solvatés) et de type 3 (interaction avec l'autre protéine uniquement grâce aux molécules d'eau pontantes, identifiée par la lettre w).

La surface de l'interface protéine-protéine se situe aux alentours de $2000 \text{ \AA}^2$, située dans la moyenne des complexes hétérodimériques transitoires [164]. Une analyse réalisée avec le logiciel Ligplot+[165] montre que le cœur de l'interface protéine-protéine est plus compact pour le cluster S2 que pour le cluster S1. Il implique moins de résidus (23 vs 30) pour une surface légèrement plus importante ($2070 \text{ vs } 1900 \text{ \AA}^2$). Par ailleurs, la fraction des résidus non polaires impliqués dans cette zone est nettement en faveur du cluster S2 (48% vs 33%). En revanche, le nombre d'interaction de type liaison hydrogène à l'interface est plus important pour le cluster S1 (12 vs 7 interactions). L'analyse de la totalité de l'interface protéine-protéine (cœur et périphérie) montre que le nombre de résidus impliqués est toujours en faveur du cluster S1 (63 vs 42 résidus). L'analyse de l'interface protéine-protéine des deux complexes est résumée dans le Tableau 2.

	Cluster S1	Cluster S2
Résidus impliqués (interface totale)	63	42
Nb liaisons H	12	7
Résidus impliqués (cœur interface)	23	30
Fraction résidus non polaire (%) (cœur interface)	33	48
SASA (Å²)	1900	2070
Interaction ionique	0	1

Tableau 2 : Résumé des analyses structurales des deux complexes sélectionnés.

Cette analyse montre que ces deux clusters correspondent parfaitement à la définition d'une interface protéine-protéine transitoire[164, 166, 167]. C'est à dire une fraction de résidus non polaires d'environ 50-60%, un nombre de liaisons hydrogène d'environ 10 et un nombre de résidu à l'interface aux alentours de 50-60. Cependant, le cluster S1 présente une compacité et une fraction non polaire moins compatible avec cette définition qui provient du fait que l'interface protéine-protéine est divisée en deux zones distinctes, séparées par un espace partiellement solvaté (cf Figure 61).

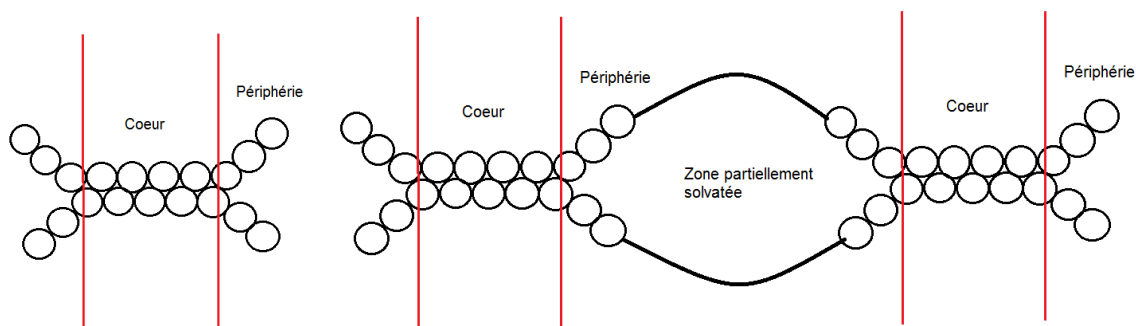


Figure 61 : Schéma représentant la topologie de l'interface protéine-protéine pour les deux clusters sélectionnés. A gauche, la topologie du cluster S2. A droite le cluster S1, avec une zone partiellement solvatée entre deux zones d'interactions.

Il est également à noter que la formation d'un complexe protéine-protéine demande une réorganisation de l'interface pour permettre une bonne complémentarité de surface. Le protocole de docking est basé sur une représentation gros-grain des résidus et sur une association d'objets (ici 2 protéines) rigides. Il est donc nécessaire d'utiliser une approche de

simulation de dynamique moléculaire pour raffiner les interfaces protéine-protéine et observer les phénomènes dynamiques

III) Simulations de dynamique moléculaire des complexes DFR-LAR

Les deux structures des complexes protéine-protéine S1 et S2 seront soumis à des protocoles de simulations de dynamique moléculaire afin de raffiner l'interface protéine-protéine et d'observer les phénomènes dynamiques comme la diffusion de métabolites d'une enzyme à l'autre.

III.1) Protocoles de simulation

Pour comprendre la dynamique des complexes protéine-protéine, nous avons réalisé des simulations de dynamique moléculaire brute force de longue durée (atteignant la micro-seconde). Chaque simulation possède des conditions initiales différentes. Au début des simulations de chaque complexe, la molécule LCC (le produit de DFR) est placée dans le site actif de DFR, en contact avec la molécule de NADP⁺. En tout, 4 simulations de DM d'une durée de 0,15 μ s, 0,17 μ s, 0,44 μ s et 0,71 μ s et 4 simulations de 0,57 μ s, 0,58 μ s, 0,94 μ s et 1,2 μ s ont été effectuées respectivement pour les complexes S1 et S2, pour une durée totale de simulation de 5,5 μ s. Les productions ont été réalisées dans l'ensemble isotherme-isobare (NTP). Les analyses structurales ont été effectuées avec le module ptraj de la suite AMBER12[115].

III.2) Résultats des simulations pour le complexe S1.

III.2.a) Analyse structurale

L'analyse de la structure des deux enzymes est résumée dans le Tableau 3. Tenant compte de toutes les simulations, les structure des enzymes prises de façon indépendante sont stables. Le RMSd de la structure de LAR (à l'exception des parties flexibles N et C-terminale, qui n'interviennent pas dans l'interaction protéine-protéine) calculé sur l'ensemble de la trajectoire est faible et n'excède jamais 2 Å. La libération du produit (LCC) nécessite une ouverture de la cavité de l'enzyme, mais la déformation de la protéine est toutefois limitée (RMSd inférieur à 4 Å).

MD	Temps de simulation (ns)	RMSd DFR (en Å)	RMSd LAR (en Å)	RMSd complexe (en Å)	Dissociation du complexe	Comportement du substrat
1	714	3,6 (4)	1,8 (2,0)	5,5 (7,0)	Non	400 ns : sortie de cavité, mais mauvaise direction
2	440	1,8 (2,3)	1,0 (1,1)	4,0 (7,0)	Non	400 ns : sortie du substrat dans bulk
3	150	2,5 (3,8)	1,8 (2,0)	5,5 (5,0)	Oui (partielle)	RAS
4	120	2,8 (4,0)	1,5 (1,6)	6,0 (9,0)	Oui (partielle)	RAS

Tableau 3 : Résumé des événements importants échantillonnés durant les différentes simulations. Pour les RMSd, la valeur moyenne est donnée et entre parenthèse est affichée la valeur maximum. Les structures de références sont les structures cristallographiques.

L'analyse du système pris dans son ensemble montre en revanche que le complexe protéine-protéine n'est pas très stable. Nous observons une réorganisation modérée de l'interface dans l'ensemble des simulations. La valeur maximale du RMSd du complexe, calculée sur la totalité de la simulation, est comprise entre 5 et 9 Å. Une dissociation partielle du complexe est observée dans deux des simulations (notée MD3 et MD4 dans le Tableau 3). L'une des deux régions à l'interface protéine-protéine (cf Figure 61) est perdue au cours du calcul, comme montré sur la Figure 62.

La réorganisation et/ou la dissociation partielle de l'interface protéine-protéine peut s'expliquer par le protocole de simulation de dynamique moléculaire employé. Premièrement, la surface d'énergie potentielle diffère entre une représentation gros grains de la protéine lors du docking et des simulations de dynamique moléculaire réalisées à la résolution atomique. Deuxièmement, les protéines sont considérées comme des objets rigides lors du calcul de docking, et les simulations de dynamique moléculaire apportent la flexibilité nécessaire à l'association protéine-protéine. Cela explique que les valeurs de RMSd se situent entre 5 et 10 Å lorsque l'on compare les protéines isolées et dans le complexe.

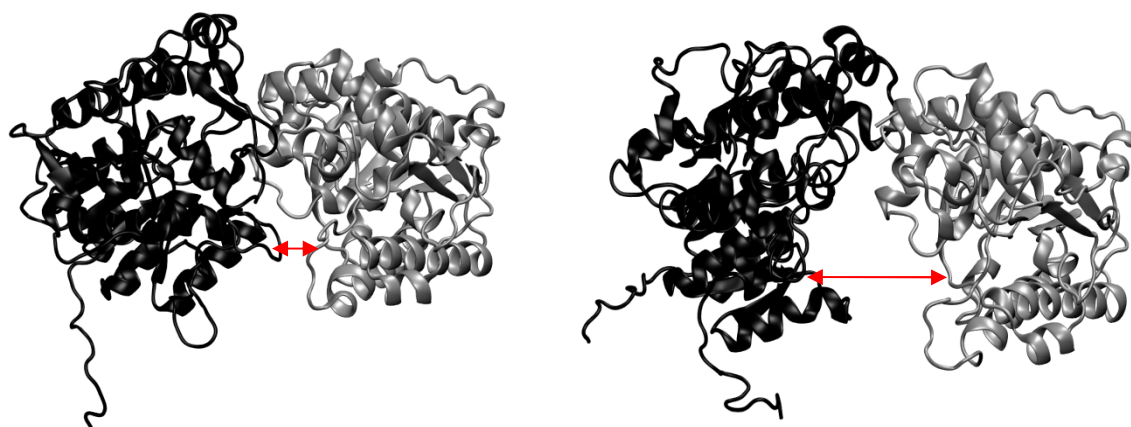


Figure 62 : Evolution de la structure du complexe DFR (en gris)-LAR (en noir). A gauche, structure sortie du protocole ATTRACT; à droite, après simulation de dynamique moléculaire, dans laquelle on observe une dissociation partielle.

III.2.b) Diffusion du substrat

La libération du produit (qui n'est due à aucune contrainte appliquée lors du protocole de simulation) est observée dans deux des simulations de dynamique moléculaire (notée MD1 et MD2 dans le Tableau 3). Lorsque nous observons une dissociation partielle du complexe, le produit reste dans la cavité enzymatique au cours de la simulation. Il est à noter que cette diffusion est plus rapide dans le cas présent que lorsque l'enzyme DFR est considérée sans la deuxième enzyme (cf partie I.2.c). Pour rappel, la sortie partielle a été observée au bout de 750 ns sur un système DFR-NADP⁺-LCC, alors qu'ici on observe une diffusion vers la phase solvant en ~400 ns de simulation. La formation du complexe protéine-protéine induit un léger changement conformationnel de DFR, qui pourrait expliquer cette différence de comportement. Enfin, le chemin emprunté par la molécule pour sortir de l'enzyme n'est pas celui attendu (cf Figure 63). La leucocyanidine se dirige vers la phase solvant au lieu de s'orienter vers le site actif de LAR.

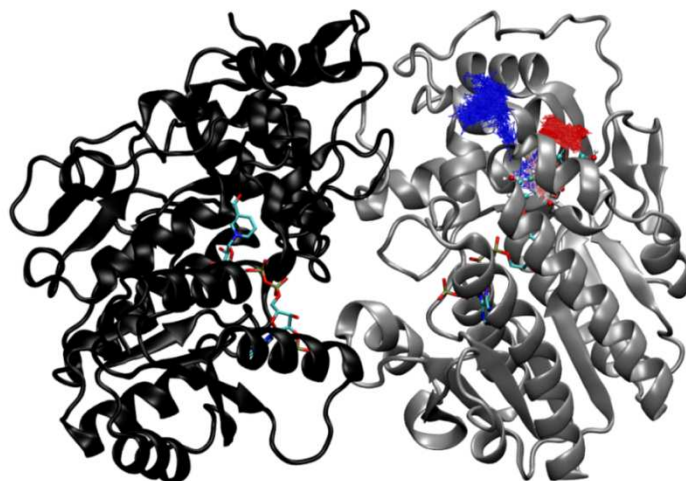


Figure 63 : Chemin de diffusion caractérisé par simulation de dynamique moléculaire dans le cas du complexe S1, observé durant la simulation MD 2. En rouge est représentée la position initiale, et en bleu l'état final obtenu pendant le calcul. DFR est représentée en gris et LAR en noir.

III.3) Résultats des simulations pour le complexe S2

III.3.a) Analyse structurale

L'analyse de la structure des deux enzymes est résumée dans le Tableau 4. Comme pour les simulations sur le système S1, la structure des enzymes prise de façon indépendante est stable. Le RMSd de la structure de LAR (à l'exception des parties flexibles N et C-terminale, qui n'interviennent pas dans l'interaction protéine-protéine) et DFR calculé sur l'ensemble de la trajectoire n'excède respectivement jamais 2,5 Å et 4 Å, pour les mêmes raisons qu'énoncées précédemment.

Dans la simulation MD1, le complexe protéine-protéine est relativement stable pendant environ 1 μ s et nous observons une dissociation complète du complexe au bout de 1,1 μ s. La dissociation intervient après un temps relativement long, compatible avec un comportement de complexe protéine-protéine transitoire.

MD	Temps de simulation (μs)	RMSd DFR (en Å)	RMSd LAR (en Å)	RMSd complexe (en Å)	Dissociation	Comportement du substrat
S2_1	1,2	3,0 (3,5)	2,0 (2,1)	7,5 (16,5)	Oui, complète à 1,1 μs.	200 ns : début de diffusion, mais partielle.
S2_2	0,935	2,2 (2,5)	1,9 (2,1)	6,2 (9,0)	Non	Diffusion totale de LCC vers LAR / début : 140 ns
S2_3	0,57	2,6 (3,0)	2,0 (2,2)	8,0 (10,0)	Non	Oui / 140 ns reste à l'interface DFR-LAR
S2_4	0,58	2,4 (3,0)	1,8 (2,1)	7,5 (9,0)	Non	Oui / 140 ns reste à l'interface DFR-LAR

Tableau 4 : Résumé des événements se déroulant lors des simulations de dynamique moléculaire impliquant le complexe S2. Les valeurs de RMSd dans le tableau correspondent à une valeur moyenne, tandis qu'entre parenthèse est indiquée la valeur maximale.

Pour la simulation MD2, une réorganisation modérée de l'interface est observée. Dans ce cas, la valeur maximale du RMSd du complexe est comprise entre 5 et 9 Å. Cette réorganisation n'est pas très importante d'un point de vue structural, mais l'est beaucoup plus d'un point de vue fonctionnel. C'est ce que nous détaillerons dans la partie III.4. Après 75 ns de simulation, une légère modification de la topologie de l'interface protéine-protéine est observée (cf Figure 64); Sa surface diminue à environ 1400 Å². Cette perte d'interaction reste toutefois compatible avec les critères structuraux d'une interface protéine-protéine transitoire (cf partie II.3. [164, 166, 167]). Elle est également compensée par la formation d'un point salin entre les résidus K93 de DFR et E175 de LAR. La distance entre les atomes NZ du résidu K93 et OD1 du résidu E175 est de 3,2 Å. La Figure 64 illustre le réarrangement structural observé durant la simulation. Pour augmenter l'échantillonnage de ce type d'interaction protéine-protéine observé durant la simulation MD2, nous avons décidé de lancer 2 simulations (MD3 et MD4) en prenant comme point de départ la structure après 75 ns de simulation. Ce réarrangement se montre extrêmement stable et ne montre aucun changement majeur sur les trois simulations (2, 3 et 4).

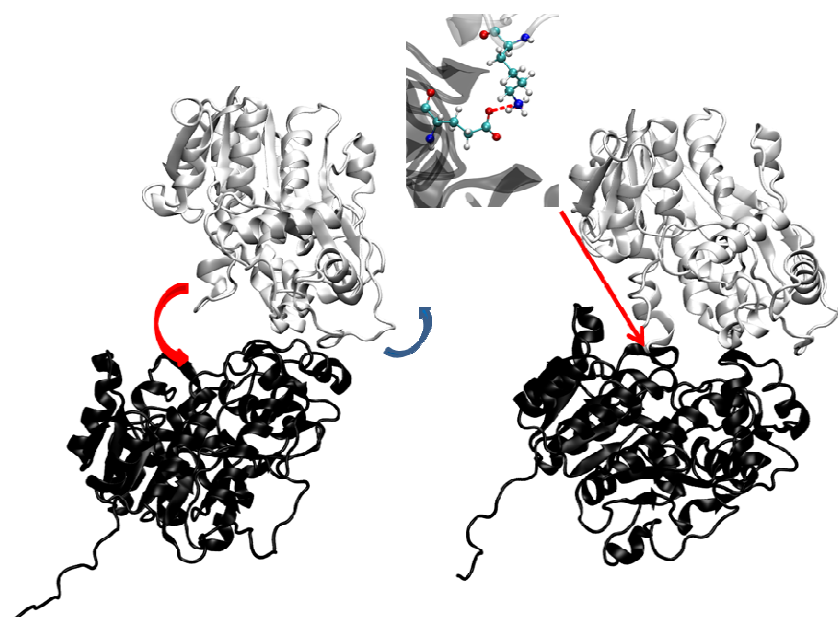


Figure 64 : Réarrangement du complexe S2 observé en simulation de dynamique moléculaire. DFR est en gris, LAR en noir. La flèche rouge montre la formation du pont salin lors du réarrangement, et la flèche bleue l'autre zone d'interaction impliquée.

III.3.b) Libération du produit et diffusion vers la seconde enzyme

La libération de la leucocyanidine est observée dans l'ensemble des simulations de dynamique moléculaire. La sortie du produit ne nécessite pas une modification structurale importante du complexe. Ce phénomène est observé dans les 75 premières nanosecondes de simulation pour MD2, 3 et 4, et est un peu plus longue pour MD1 (200 ns). Cette différence de comportement peut être due au réarrangement structural qui se produit dans MD2, et qui induit un départ plus facilité. Dans les simulations MD3 et 4, le produit migre jusqu'à l'interface protéine-protéine, puis sort de cette interface et diffuse dans la phase solvant. Dans le cas de la MD1, le ligand a été transféré jusqu'à l'interface protéine-protéine avant la dissociation, mais cette dernière empêche la finalisation de la migration.

La diffusion du produit de l'enzyme DFR jusqu'à la cavité de LAR est observée dans la simulation MD2. La Figure 65 représente le chemin suivi ainsi que l'évolution du RMSd du ligand par rapport à sa position cristallographique dans le fichier PDB 3I52, correspondant au complexe ternaire LAR-NADPH-LCC. La valeur de RMSd est naturellement très élevée au début de la simulation pour se retrouver à une valeur d'environ 5 Å à la fin, indiquant un positionnement proche de la structure cristallographique.

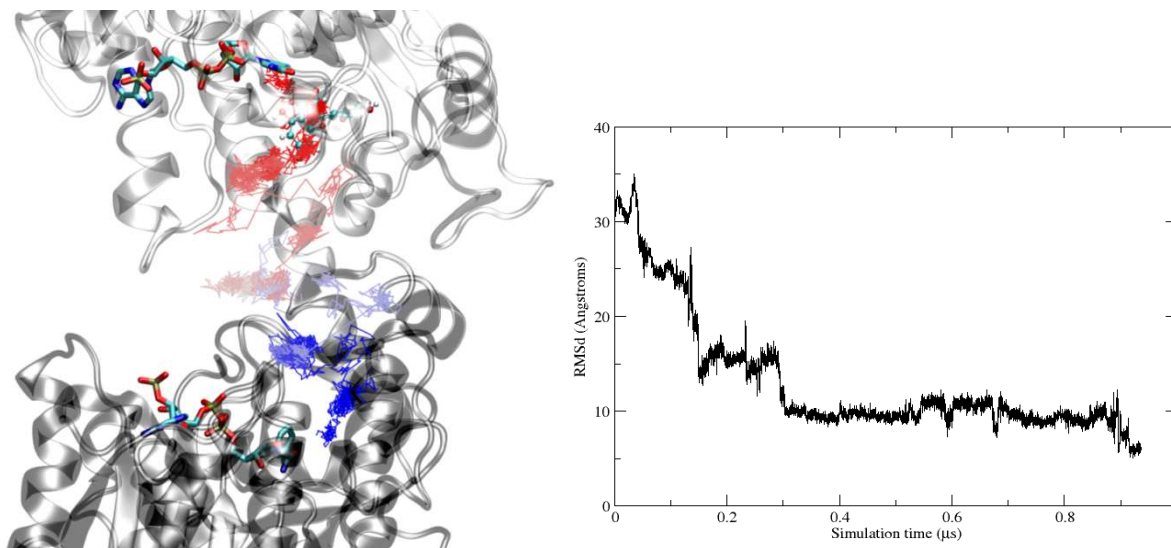


Figure 65 : Chemin de diffusion suivi par le substrat observé lors de la simulation 2 du complexe S2 et évolution du RMSd durant le transfert. La position initiale du substrat est représentée en rouge, tandis que la position finale en bleu.

Le chemin de diffusion peut être scindé en trois grandes étapes indiquées sur la Figure 65 : la première étape intervient lors des 300 premières nanosecondes. La molécule sort de la cavité de DFR et atteint l'interface protéine-protéine en 150 ns environ. Elle continue alors jusqu'à 300 ns, dans lesquelles la molécule explore l'interface protéine-protéine et migre jusqu'à la surface de LAR. La deuxième étape dure 500 ns, pendant laquelle le substrat explore la surface de l'enzyme. La dernière étape consiste en l'entrée de LCC dans la cavité enzymatique de LAR. Chacune des phases de ce processus de "substrate channeling" a été analysée d'un point de vue structural et énergétique. L'énergie d'interaction ligand/récepteur a été évaluée au moyen du protocole MM-GBSA (voir partie méthodologique, page 75) afin de caractériser l'évolution de la stabilisation du métabolite et les différents états du complexe. La simulation a été découpée en 157 sous-ensembles de 600 ps, et pour chaque ensemble, une valeur moyenne et un écart type de l'énergie d'interaction ont été calculés (cf Figure 66 à gauche). Les distances entre les centres de masse du métabolite et de chaque enzyme ont été mesurées le long de la simulation. Cette analyse a permis de représenter, sous forme d'une carte en deux dimensions, l'évolution de l'énergie d'interaction en fonction de ces deux critères structuraux (cf Figure 66 à droite). Les trois étapes du processus de "substrate channeling" sont clairement visibles sur ces deux figures.

Remarque : La contribution entropique à l'énergie d'interaction n'a pas pu être évaluée sur l'ensemble de la trajectoire. La méthode employée est basée sur un protocole d'analyse des modes normaux nécessitant un état d'énergie minimale (qu'il soit global ou local). Nous avons

estimé cette contribution sur deux états : l'état initial et final. Le terme $T\Delta S$ est équivalent dans les deux cas, avec une valeur de -23 kcal/mol.

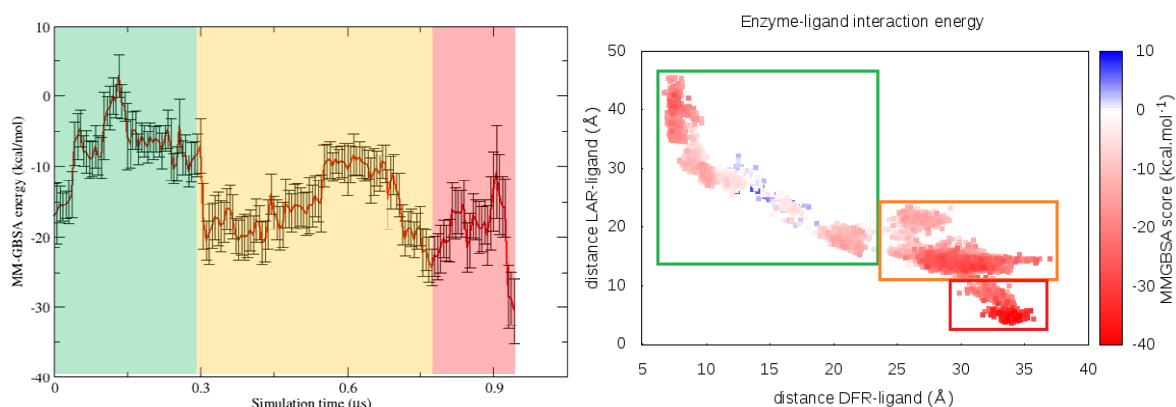


Figure 66 : A gauche, évolution de l'énergie d'interaction LCC/complexe le long du chemin de diffusion. A droite, projection de cette énergie en fonction de la distance entre les centres de masse de DFR et du ligand, et celle entre les centre de masse de LAR et de LCC. Les trois couleurs représentent les trois étapes précédemment citées.

- Etape 1 du chemin de diffusion

La première étape du chemin de diffusion est caractérisée par une diminution de l'énergie d'interaction assez importante, de l'ordre de 23 kcal/mol, entre 0 et 150 ns de simulation, puis d'un renforcement de cette interaction jusqu'à atteindre une interaction d'environ -23,5 kcal/mol. Le $\Delta\Delta E$ de cette étape est d'environ -3,5 kcal/mol.

De nombreux phénomènes se déroulent durant cette étape. Nous observons ainsi une diminution très forte du RMSd (cf Figure 65), dans laquelle nous pouvons identifier trois cas distincts :

i) une diminution d'une valeur de 30 à 25 Å, qui correspond à une réorientation du substrat dans la cavité de DFR. Ceci est visible sur la première portion de la partie encadrée en vert (en haut à gauche) sur la Figure 66, dans laquelle nous voyons une augmentation de la distance entre les centres de masse de LAR et LCC, mais très peu de changement pour la distance entre le centre de masse de DFR et de LCC.

ii) une seconde sous-étape de 40 à 150 ns de simulation, dans laquelle le RMSd chute d'une valeur de 25 à 15 Å. Cette étape correspond à la sortie du substrat de la cavité de DFR, ce qui se traduit par une grande perte d'interaction. Ce phénomène est donc associé à cette importante déstabilisation du substrat observée dans le profil énergétique de la Figure 66. Elle

correspond à la seconde portion de la partie encadrée en vert (la plupart des points bleus). Nous voyons que dans cette étape la distance entre le centre de masse de DFR et LCC augmente tandis que celle entre LAR et LCC diminue, indiquant la diffusion du site actif de DFR vers le suivant. Durant la sortie du site actif, le LCC interagit avec M88 et P204 (définissant la porte de la voie d'accès 1 de DFR) mais également avec la THR159 (impliquée dans la porte de la voie d'accès 3). Nous pouvons en déduire que la molécule suit un chemin hybride entre ces deux sorties.

iii) La dernière sous-étape se déroule de 150 à 300 ns de simulation. Le RMSd reste constant sur cette valeur de 15 Å, puis diminue en fin d'étape à une valeur de 10 Å. Il s'agit ici d'une exploration de l'interface protéine-protéine par le métabolite. Cela correspond au palier observé sur la courbe de l'énergie d'interaction. La leucocyanidine interagit avec les résidus M88, F85, T159 de DFR, ainsi qu'avec les résidus H172, E94, I171, et E265 de LAR. La distance entre les centres de masse des deux enzymes et le ligand continue à évoluer (nous nous situons toujours dans la partie encadrée en vert, en bas à droite). A la fin de cette étape, le complexe tombe dans un minimum d'énergie, correspondant à l'arrivée du LCC sur la surface de LAR.

- Etape 2 du chemin de diffusion

Après l'exploration de l'interface protéine-protéine, la molécule atteint un minimum d'énergie, avec un gain total par rapport au maximum de la courbe d'environ 22 kcal/mol. Commence ainsi une longue phase, jusqu'à 560 ns, durant laquelle l'énergie d'interaction reste constante. Après 560 ns de simulation, nous observons une seconde perte d'interaction, d'une valeur de 12 kcal/mol. Enfin aux alentours de 800 ns de simulation, un deuxième minimum est atteint. Le gain d'énergie associé est de 16 kcal/mol. Le $\Delta\Delta G$ total de cette étape de la diffusion est ainsi de -4 kcal/mol.

L'analyse du RMSd montre que sa valeur (environ 10 Å) reste constante tout le long de cette étape. Pour autant, de nombreux phénomènes dynamiques sont observés durant cette seconde étape. Le premier minimum atteint est associé à une position du substrat particulière, qui lui permet d'interagir de manière forte avec le cofacteur situé dans la cavité de LAR. Deux liaisons hydrogène sont créées entre le groupement phosphate et le cycle B du LCC. Le substrat interagit également avec l'enzyme, en créant deux interactions de type liaison hydrogène avec les résidus E94 et N135. Enfin, le métabolite s'insère entre les hélices α 4

(résidus 94-110) et 6 (résidus 128-150), créant ainsi de nombreuses interactions hydrophobes. La position décrite est représentée sur la Figure 67.

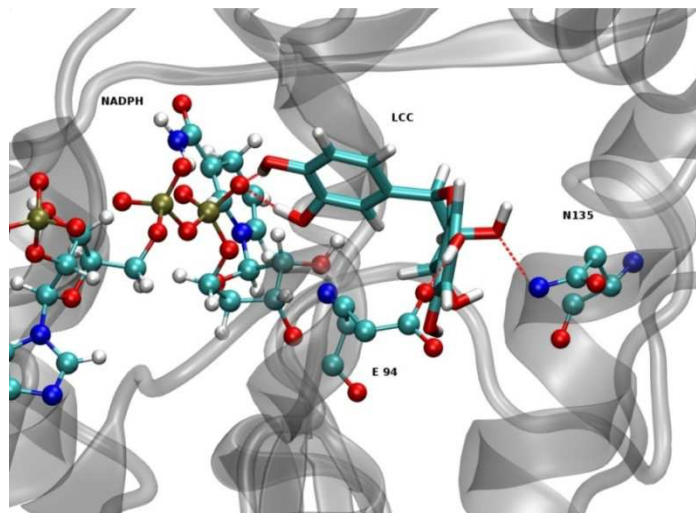


Figure 67 : Position intermédiaire rencontrée lors du chemin de diffusion du LCC. Ne sont affichées ici que les interactions de type liaison hydrogène à la surface de LAR.

Au bout de 560 ns de simulation, la seconde déstabilisation est atteinte, et elle correspond à une perte ponctuelle des interactions du LCC avec les résidus E94 et N135, ainsi qu'à la perte des interactions hydrophobes avec les deux hélices de l'enzyme. Seules les interactions avec le NADPH sont conservées. La position spatiale change peu, ce qui explique la valeur constante du RMSd sur cette période. La fin de cette étape est caractérisée au bout d'environ 800 ns de simulation, par une reformation des interactions avec l'hélice 4 et 6. Sur l'image b) de la Figure 66, cette seconde étape est identifiée par la zone entourée en orange. Nous échantillonnons ainsi un bassin d'attraction énergétique de manière importante durant toute cette étape.

- Etape 3 du chemin de diffusion

La simulation échantillonne de manière importante la conformation associée à ce minimum d'énergie. La simulation a été continuée à ce jour à 1,47 μ s, et ne nous a pas permis de franchir la barrière d'énergie nécessaire à une entrée du LCC dans la cavité de LAR. Nous avons décidé d'appliquer un autre protocole, pour donner un "coup de pouce" au système afin d'échantillonner cette entrée. Il s'agit du protocole d'aMD (accelerated Molecular Dynamics[168]). Le principe est d'introduire un biais dans la simulation sur l'énergie potentielle du système, ou sur l'énergie associée aux angles dièdres (ou sur les deux). La valeur de l'énergie est ainsi augmentée, permettant un "lissage" de la surface d'énergie

potentielle (et/ou des angles dièdres) pour faciliter ainsi les transitions entre différents états. Nous avons appliqué ce protocole d'aMD afin d'observer la transition du LCC vers le site actif de LAR. Les paramètres de "boost" ont été appliqués sur l'énergie potentielle et l'énergie des angles dièdres (3,5 kcal/mol/résidu de boost sur les angles dièdres, et 20 % de boost sur l'énergie potentielle).

Ainsi, avec le protocole d'aMD, nous avons pu échantillonner l'entrée du substrat dans la cavité. L'évolution de l'énergie d'interaction montre une nouvelle perte d'interaction, légèrement plus élevée que la précédente (14 kcal/mol). Après l'avoir franchie, le minimum global de la courbe est atteint. Le gain d'énergie est ici de 20 kcal/mol. Le $\Delta\Delta G$ total de cette étape de la diffusion est ainsi de -8 kcal/mol. Globalement, le $\Delta\Delta G$ total de la diffusion du substrat du site actif de DFR vers LAR est de -13,5 kcal/mol.

La valeur du RMSd diminue jusqu'à environ 5 Å, et représente l'entrée du substrat dans la cavité de la seconde enzyme. Ce mouvement est caractérisé par un réarrangement important. Les interactions entre le substrat et le cofacteur sont encore conservées, mais celles impliquant le résidu N135 et l'hélice 4 sont perdues. La molécule atteint ainsi la porte de la cavité de LAR, caractérisée par des interactions entre le cofacteur et deux résidus : M136 et K140 (il s'agit de la lysine catalytique). Ces interactions sont représentées sur la Figure 68. Ces dernières sont rompues durant la simulation de type aMD, pour finalement donner la possibilité au substrat d'entrer dans la cavité enzymatique. Dans la cavité, la leucocyanidine interagit principalement avec les résidus Y137, K140, H122 et Q285.

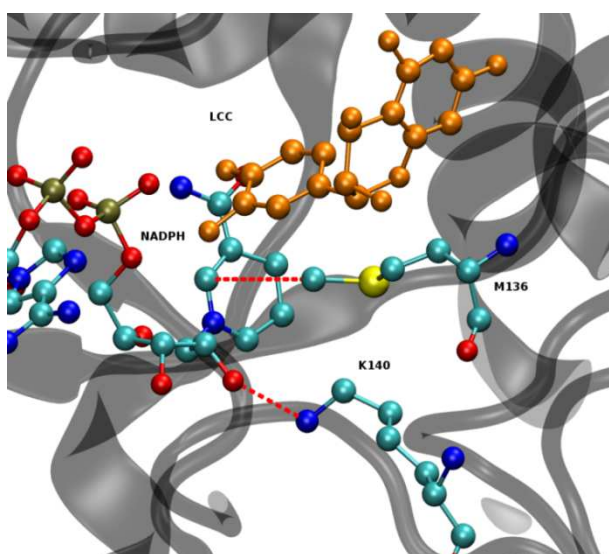


Figure 68 : Interaction jouant le rôle de "porte" de la cavité de LAR, impliquant le cofacteur, la lysine catalytique K140 et la méthionine M136.

III.4) Rôle de la solvation dans le phénomène de "substrate channeling"

Les métabolons sont des assemblages transitoires d'enzymes, ce qui implique une proximité des différentes cavités. Ils maintiennent et accroissent le flux métabolique en concentrant les intermédiaires réactionnels dans l'environnement proche des enzymes. Nous avons observé le transfert du substrat d'un site actif à l'autre sans passer par une phase de diffusion dans une phase purement solvant. Ce processus peut être considéré comme un phénomène de "substrate channeling". Le rôle de la solvation est ici crucial. Une analyse de la micro-solvatation du substrat au cours de la simulation réactive a été réalisée.

Le nombre de molécules d'eau dans la première sphère de solvation (jusqu'à une distance de 3,4 Å du ligand) a donc été mesuré. La simulation a été découpée en 157 sous-ensembles de 600 ps, et pour chaque ensemble, une valeur moyenne et un écart type du nombre de molécules d'eau ont été calculés. Une simulation de DM de la LCC dans une phase purement solvant a été réalisée pour estimer le nombre moyen de molécules d'eau présentes dans la première sphère de solvation. La solvation maximale a été évaluée à environ 36 molécules d'eau.

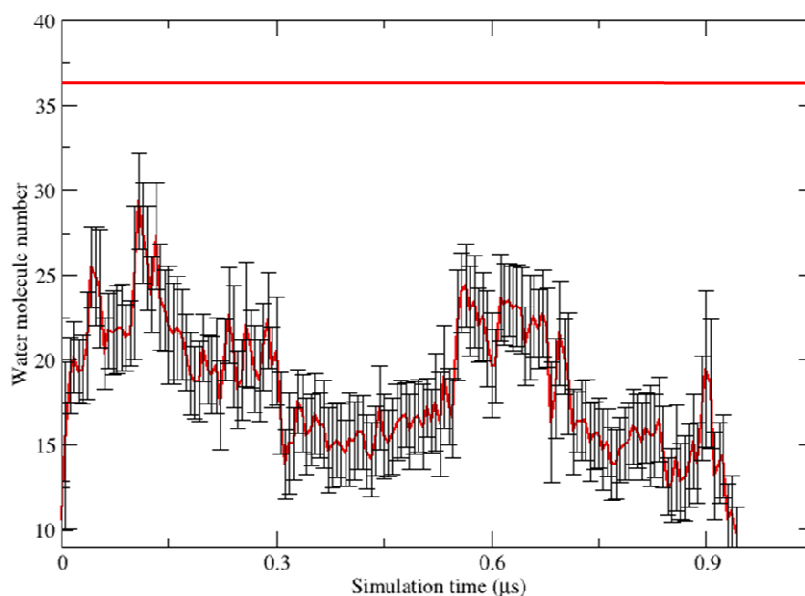


Figure 69 : Evolution du nombre de molécules d'eau autour du substrat lors de sa diffusion du site actif de DFR vers celui de LAR (sphère de solvation prise entre 1 et 3,4 Å autour du solvant). En rouge est représentée l'évolution, et en noire les barres d'erreurs associées. La ligne rouge montre le nombre moyen de molécules d'eau pour un substrat complètement solvato.

Nous remarquons que cette courbe présente de nombreuses similarités avec la courbe de l'évolution de l'énergie d'interaction LCC/complexe calculée par le protocole MM-GBSA. Le profil peut être découpé en trois étapes, de la même manière que précédemment. Ainsi, chaque barrière d'énergie peut être associée à une solvation partielle du substrat, plus ou moins importante selon la valeur de la barrière. Les minimums de micro-solvation sont ainsi associés à des interactions plus importantes avec l'enzyme. Enfin, nous pouvons remarquer que nous n'obtenons jamais une solvation totale de la molécule lors du transfert vers la seconde enzyme. Le nombre de molécules d'eau entourant le métabolite (valeur max égale à 28 molécules d'eau) est toujours inférieur à la valeur estimée dans la simulation de LCC dans une phase solvant pure (estimée à 36 molécules d'eau).

Les effets de solvant jouent un rôle majeur dans le processus d'interaction ligand/récepteurs. Ici, lors du transfert de la molécule d'une enzyme à une autre, un phénomène de solvation/désolvation a lieu. La libération du substrat entraîne une solvation partielle de celui-ci grâce à l'existence du complexe DFR-LAR. Le processus de désolvation lié à l'association LAR-substrat est moins pénalisant d'un point de vue énergétique lors de la formation d'un tel édifice. Le transfert de LCC de DFR vers LAR a lieu lors de la première étape (0-300 ns) et le nombre de molécules d'eau autour du substrat est ici de 28 au maximum. Le gain sur la pénalité de désolvation est donc d'environ 20% (28/36). La valeur de l'énergie de solvation calculée par le protocole MM-GBSA est de 59,39 kcal/mol. Le gain d'énergie sur la pénalité de désolvation est donc de l'ordre de 12 kcal/mol.

La formation de ces complexes multienzymatiques permet donc un transfert de molécules d'un site actif à l'autre, et permet à celle-ci de ne pas être resolvater de manière totale, entraînant un gain d'énergie non négligeable. Ce phénomène de solvation partielle est ainsi une des forces motrices du phénomène de "substrate channeling", qui est la base même du fonctionnement d'un métabolon.

IV) Conclusion

Un métabolon est un assemblage complexe d'enzymes, ce qui rend ce type de structure difficile à étudier expérimentalement. Ainsi, les outils de modélisation moléculaire sont précieux et parfaitement adaptés pour une analyse moléculaire de ce type de complexe. Les différents protocoles (RAMD, DM, Docking protéine-protéine, aMD) utilisés ont permis d'étudier de manière pertinente ce type de structure. Ils nous ont permis de répondre à différentes problématiques associées : i) l'identification des voies d'accès au site actif des deux

enzymes pour les métabolites ii) un modèle d'interaction pertinent entre les deux enzymes, dans lequel une connexion existe entre les deux sites actifs et iii) l'étude des phénomènes dynamiques associés à ce type de structure à une résolution atomique.

Dans le cadre de la biosynthèse des flavonoïdes, la voie métabolique fait intervenir l'une à la suite de l'autre les enzymes DFR et LAR. La formation d'un complexe entre ces deux enzymes est donc envisageable. Dans ce chapitre, nous avons étudié la possible association de ces deux enzymes et fourni des informations au niveau moléculaire et énergétique sur les phénomènes associés à ce complexe multienzymatique. La détection des entrées/sorties des deux enzymes ont permis de faciliter l'identification des complexes pouvant présenter une connexion entre les sites actifs enzymatiques. Les calculs de docking permettent de dégager deux structures, dont l'une présente une interaction classique de type protéine-protéine transitoire. Les simulations de dynamique moléculaire, appliquées sur ces deux structures, montre généralement une libération de la leucyanidine du site actif de DFR, dans des temps plus courts que lorsque le complexe DFR-NADP⁺-LCC est considéré seul. Nous avons réussi à échantillonner dans le cas du complexe S2, un transfert du LCC du site actif de DFR vers celui de LAR, sans appliquer aucune contrainte. Cette observation peut être assimilée à un phénomène de "substrate channeling".

Ce phénomène montre un $\Delta\Delta E$ total de transfert de -8 kcal/mol. Plusieurs étapes ont pu être identifiées avec tout d'abord, une libération de LCC à partir de la cavité de DFR pour une migration vers l'interface protéine-protéine, ensuite une exploration de la surface de l'enzyme LAR par le métabolite. Enfin, l'étape finale consiste en l'entrée du ligand dans le site actif de la seconde enzyme. Le long de ce phénomène, la solvation joue un rôle majeur. En effet, le substrat n'est jamais resolvaté de manière complète, on observe ainsi un gain non négligeable sur les pénalités de désolvatation, ce qui facilite ainsi le transfert de la molécule, et de limiter la déstabilisation de l'interaction complexe/ligand, comme par exemple celles associées à une entrée ou sortie de cavité. Ainsi, la solvation partielle du substrat joue un rôle crucial dans le phénomène de "substrate channeling" et dans le fonctionnement d'un métabolisme.

Les différents résultats reliés au phénomène de "substrate channeling" ont été observés sur la base d'une seule simulation, ce qui ne nous permet pas de dégager, en premier abord, une statistique suffisante. L'idéal aurait été d'effectuer plus de calculs sur ces complexes afin d'échantillonner une seconde fois le processus. Nous ne devons cependant pas oublier que ce

type de phénomène est associé à des temps de simulation de l'ordre de la microseconde, ce qui demande de grandes ressources de calcul. De plus, l'échantillonnage de conformations hautes en énergie sur ces systèmes très hydrophiles est plus difficile que sur des systèmes totalement hydrophobes (par exemple, les RCPG), car il est nécessaire de rompre des interactions plus fortes, comme des interactions de type liaison hydrogène.

Chapitre 4 : Interactions F3'H/membrane et premiers modèles de métabolon

Il est possible que les enzymes impliquées dans la cascade menant à la biosynthèse de molécules naturelles soient en interaction forte avec une membrane, que ce soit la membrane cellulaire ou une autre impliquée dans la compartimentation à l'intérieur de la cellule. Il semblerait, dans notre cas qui traite de la voie de production menant aux flavonoïdes, qu'au moins une enzyme soit dans ce cas. Il s'agit de l'enzyme F3'H (Flavonoïde-3'-hydroxylase). Les membranes sont constituées de bicouches lipidiques qui ont pour rôle de protéger et de compartimenter la cellule. Nous pouvons par exemple citer la membrane du réticulum endoplasmique, qui est une zone de la cellule permettant la création et modification de protéines[169]. Les protéines qui interagissent avec les membranes peuvent le faire de différentes façons. Elles peuvent être en interaction transitoire, relativement faible, en étant simplement "posées" sur la partie hydrophile des bicouches lipidiques. D'un autre côté, elles peuvent être totalement intégrées dans la membrane, permettant ainsi un flux de molécules ou d'ions, ou même le transfert d'informations entre le milieu extérieur et l'intérieur de la cellule. Il s'agit de protéines dites transmembranaires (GPCR, RO...)[170]. Enfin, et c'est le cas qui nous intéresse ici, seule une partie de l'enzyme interagit fortement avec les lipides tandis que la majeure partie reste en interaction forte avec le solvant. F3'H appartient à cette dernière famille. Comme présenté lors du chapitre de préparation des systèmes, F3'H est une enzyme de type cytochrome P450 [59, 60, 75]. La plupart des enzymes de cette famille présentent une partie transmembranaire, mais plusieurs types d'interactions sont envisageables entre la surface de la membrane et l'enzyme (liaison hydrogène, hydrophobe, ionique). Il a été récemment montré que des interactions de type cation- π pouvaient jouer un rôle majeur dans la stabilisation de ces interactions. Elles possèdent un caractère stabilisant plus fort qu'une liaison hydrogène [171] et sont impliquées de manière cruciale dans la stabilisation de grands nombres de systèmes membrane/protéine [172, 173].

Dans le cadre de la biosynthèse des flavonoïdes, le métabolon se formerait à la surface de la membrane du réticulum endoplasmique [58, 160]. Certaines enzymes joueraient ainsi un rôle d'ancre à la membrane, autour de laquelle les autres protéines peuvent s'agréger pour former un métabolon [53, 57]. Ce phénomène d'ancrage à la membrane pourrait avoir une importance dans la statistique de rencontre des enzymes, augmentant la probabilité de la création de ces interactions. De plus, la stabilité de l'édifice multienzymatique pourrait être augmenté grâce à des interactions supplémentaires entre enzymes constituant le métabolon et la bicouche lipidique. La prise en compte de ces interactions est ainsi d'une grande importance.

Ici, comme présenté en page 92, le serveur OPM[102] prédit que l'hélice N-terminale de F3'H doit être transmembranaire. Cette prédiction a également mis en évidence une interaction de surface, de nature hydrophobe, impliquant les résidus 35-37, 69, 222-227. La prédiction sans hélice montre à nouveau que ces résidus sont impliqués dans une interaction de surface, laissant penser que notre modèle d'interaction F3'H/membrane est cohérent. Nous présenterons dans ce chapitre les premières études sur F3'H, son comportement dynamique en présence d'un modèle de membrane explicite, ainsi que les premiers modèles de métabolon obtenu en regroupant tous les résultats obtenus jusqu'alors. Comme précédemment, des simulations de type RAMD ont été effectuées afin de localiser les différentes voies reliant la cavité enzymatique de F3'H et le solvant. Le système a ensuite été soumis à des simulations de dynamique moléculaire en tenant compte d'une membrane décrite explicitement. Enfin des calculs de docking protéine-protéine ont été réalisés afin de proposer un premier modèle d'un métabolon impliquant les enzymes F3'H-DFR-LAR en interaction avec un modèle explicite de membrane.

I) Simulations préparatoires de dynamique moléculaire de systèmes F3'H/membrane.

I.1) Identification des chemins d'accès au site actif de F3'H

I.1.a) Canaux déjà identifiés dans les cytochromes P450

Les cytochromes P450 possèdent un arrangement structural très similaire les uns des autres et surtout très compact [174]. Le site actif de l'enzyme est très enfoui, et la question de l'identification de chemins d'accès au site actif des molécules s'est rapidement posé. De par leur structure tridimensionnelle très similaire, ces chemins d'accès pourraient être commun à la famille entière des cytochromes P450. Ainsi, de nombreuses études ont déjà été publiées sur l'étude d'accès de composés au sein de ces enzymes [175-178]. De manière générale, la structure tertiaire des cytochromes P450 est composée de deux domaines de feuillets β , ainsi que de 10 hélices alpha (sans compter le domaine N-terminal transmembranaire) nommée de A à K. Les différentes études théoriques, utilisant des méthodes telles que des protocoles de steered molecular dynamics (SMD), d'umbrella sampling (US) et de RAMD ont permis l'identification de plusieurs chemins d'accès connectant le site actif au milieu extérieur. La

nomenclature de la structure secondaire des cytochromes P450 et les différents chemins d'accès identifiés sont représentés sur la Figure 70 .

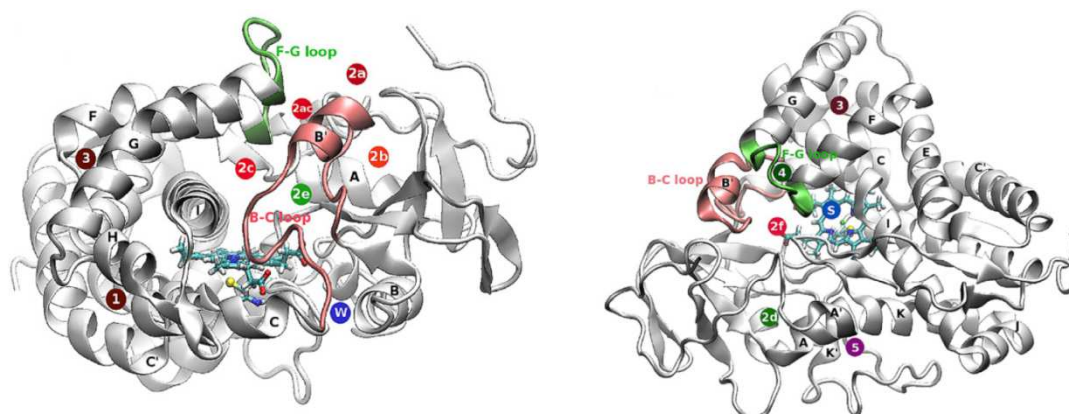


Figure 70 : Nomenclature des différentes structures secondaires des cytochromes P450. Les cercles désignent les différents accès au site actifs identifiés. A gauche, vue de côté; à droite, vue de dessus. Les cercles désignent la totalité des canaux identifiés comme étant des voies d'accès au site actif. Tiré de la ref [175].

Les sept chemins d'accès dérivant du canal 2, notés de 2a à 2f, correspondent aux chemins les plus échantillonnés et les plus favorables énergétiquement pour la sortie et l'entrée des molécules [175, 179]. Les différences entre ces sept chemins sont mineures, et proviennent uniquement de la conformation à un temps donné de la structure des boucles F-G et B-C. Nous pouvons les considérer comme toutes équivalentes. L'ouverture et la fermeture de ces voies dépendent uniquement de la position de la boucle F-G, reliant les hélices F et G, ainsi que celle de l'hélice B' avec la boucle B-C. Les publications montrent que plusieurs conformations existent dans lesquelles l'hélice B' peut se dénaturer afin de permettre la sortie de composés. Un déplacement de cette dernière est également possible et permet l'ouverture et la fermeture de canaux. Les différents chemins peuvent servir soit pour l'entrée des molécules dans le site actif, soit pour la sortie des produits de la réaction, mais aucun rôle n'a été clairement identifié pour chacun. Seuls les canaux W et S ont été caractérisés comme les chemins d'accès du solvant au site actif (la présence d'un atome d'oxygène est nécessaire pour la réaction enzymatique), mais peuvent aussi jouer le rôle de régulateur du degré de solvation du site actif. Il a été également montré que le canal S peut permettre la sortie de composés dans certains cytochromes P450.

Cette nomenclature étant commune à toutes les enzymes de type cytochrome P450, nous pouvons identifier les hélices et les boucles dans la structure de F3'H reconstruite par homologie (voir Figure 71).

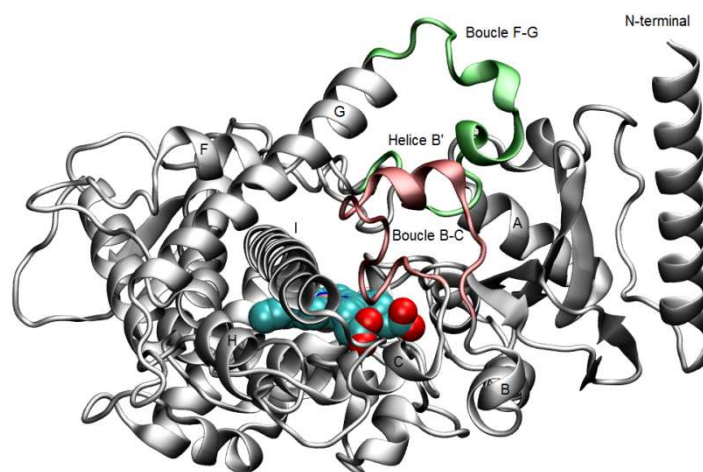


Figure 71 : Identification des hélices de la structure reconstruite de F3'H. La boucle B-C et la boucle F-G sont respectivement représentées en rouge et vert.

Les résidus impliqués dans les boucles B-C et F-G de l'enzyme sont respectivement les résidus 95-110 et 200-215. Les simulations de dynamique moléculaire effectuées sur le complexe F3'H-DHQ-Hème (sans modèle de membrane) ne permettent pas de visualiser une sortie de la dihydroquercétine de manière naturelle (au total 700 ns simulées lors de 4 simulations). En effet, la structure est et reste très compacte durant les simulations. A ce jour, aucune étude sur une sortie naturelle de composé depuis le site actif d'un cytochrome P450 par simulation de dynamique moléculaire brute-force n'a été publiée, il faut nécessairement biaiser la simulation. La structure a ainsi été soumise à des simulations de type RAMD [180], afin de vérifier si nous détectons les mêmes chemins déjà identifiés dans les précédentes études.

I.1.b) Protocole de simulation

Une force permettant d'accélérer le substrat dans une direction aléatoire est appliquée tous les 40 pas de simulation. Trois valeurs d'accélération ont été utilisées afin de détecter un maximum de sorties possibles (0,15, 0,2 et 0,25 kcal/mol/Å/g). La graine permettant de générer aléatoirement cette force supplémentaire a elle même été fixée aléatoirement pour s'assurer du caractère unique des simulations. Pour chaque valeur de ce paramètre, 30 simulations ont été effectuées, pour un total de 90 simulations réalisées en RAMD. Le calcul s'arrête automatiquement dès que la distance entre le centre de masse du substrat et celui de l'enzyme dépasse 40 Å.

I.1.c) Résultat pour l'enzyme F3'H

Le Tableau 5 résume les sorties échantillonnées par DHQ lors de chaque trajectoire. De par leur proximité spatiale et des légères différences structurales entre eux, les résultats concernant les 7 canaux dérivant du chemin 2 et le canal S ont été rassemblés.

Sortie	A = 0,25	A = 0,20	A = 0,15	Total
1	4	10	11	25
2(a, b, c, ac, d, e, f et S)	10	7	3	20
3	2	1	2	5
W	0	0	0	0
4	0	0	0	0
5	3	2	0	5
Autres	11	10	14	35

Tableau 5 : Résumé des sorties échantillonnées dans toutes les simulations de type RAMD.

La ligne "Autres" désigne les sorties non détectées auparavant, dénaturant une partie de la structure et donc inexploitable, et celles qui ne sont pas compatibles avec l'interaction de F3'H avec la membrane.

Environ un tiers des simulations présente soit une sortie proposée dans la zone censée interagir avec la membrane (impliquant notamment les résidus en interaction tels que W35), soit une dénaturation d'une partie de la protéine lors de la sortie du substrat. Ce dernier phénomène peut être dû à l'exploration par le ligand d'une portion moins structurée et plus fragile, donc plus propice à la dénaturation (par exemple, les résidus 315-325). Si nous mettons de côté ces simulations, la sortie par une voie pouvant correspondre au canal 1 est celle qui est la plus échantillonnée (impliquant les résidus S133, G134, K125, L272, E288, V289 et I291). Cependant, au niveau structural, la sortie de composé dans ce canal entraîne la plupart du temps de grandes modifications structurales du site actif. Notamment, les interactions entre l'hème et les différents résidus du site actif sont souvent rompues. Le résidu R124, par exemple, identifié dans notre structure pour faire une interaction ionique forte avec le groupement hème, perd souvent son interaction avec le cofacteur lors de la sortie de DHQ. Ainsi, même si ce chemin est le plus échantillonné, les modifications structurales

qu'implique son parcours ne sont pas cohérentes avec les observations et profils énergétiques présentés précédemment. D'autre part, dans le cas de $A=0,15$, l'accélération n'est peut être pas assez forte pour permettre au substrat d'échantillonner les autres accès, sachant que notre structure est très compacte. Le substrat prend alors le chemin le plus court pour sortir de l'enzyme, qui est le canal 1, ou une sortie "autre" qui perturbe encore plus l'interaction avec l'hème (rupture de l'interaction avec la cystéine coordonnant le métal).

Le canal 2 fait intervenir la boucle B-C (résidus 95-110) et la boucle F-G (résidus 200-215) (voir Figure 71), et selon leur position dans la structure, une variante du canal sera plus ouverte que l'autre. Nous avons ainsi choisi de garder les différentes variantes du canal 2 (a, b, c, d, e, f), le chemin S ainsi que les canaux 3 et 5, (impliquant respectivement K197, E200, R242, F246 et P365, L478, H49, L488), même si ces derniers sont moins échantillonnés, comme voies de sortie principale du DHQ. Des simulations de dynamique moléculaire sur le système F3'H/membrane avec DHQ dans le site actif ont été réalisées afin d'analyser le comportement dynamique des différents canaux avec la présence de la membrane.

I.2) Evolution de l'interaction F3'H/membrane lors des simulations de DM

I.2.a) Protocole de simulation

Le mode d'interaction F3'H/membrane prédit par les résultats du serveur OPM a servi de structure de départ à ces simulations. La membrane a été décrite de façon explicite avant d'appliquer les protocoles de DM, décrits dans le chapitre I (page 91). Deux simulations de dynamique moléculaire (durée respective de 640 et 400 ns) ont été réalisées sur ce système.

I.2.b) Evolution du mode d'interaction F3'H/membrane

La prise en compte de la membrane dans les calculs a une incidence importante tant sur le type d'interaction enzyme/membrane que sur la structure même de l'enzyme. L'hélice transmembranaire voit son angle de tilt changer durant la simulation. La force motrice de ce changement provient majoritairement des interactions ioniques. Nous avons représenté les parties N- et C-terminales sous leur forme zwitterionique. La position du résidu (M1) situé en N-terminale évolue pour créer en milieu de simulation une interaction ionique très forte avec plusieurs groupements phosphates des têtes polaires des lipides, situées sur la face opposée de la couche lipidique (cf Figure 72). Durant la simulation, jusqu'à 4 interactions ioniques

peuvent exister de manière concomitante. L'interaction est constante durant toute la durée de la simulation, mais les groupements phosphates présentent une dynamique d'échange importante. Ce changement est cohérent avec une étude de Cojocar *et al.* [77] dans laquelle cette partie N-terminale traverse intégralement la membrane et interagit avec la partie hydrophile.

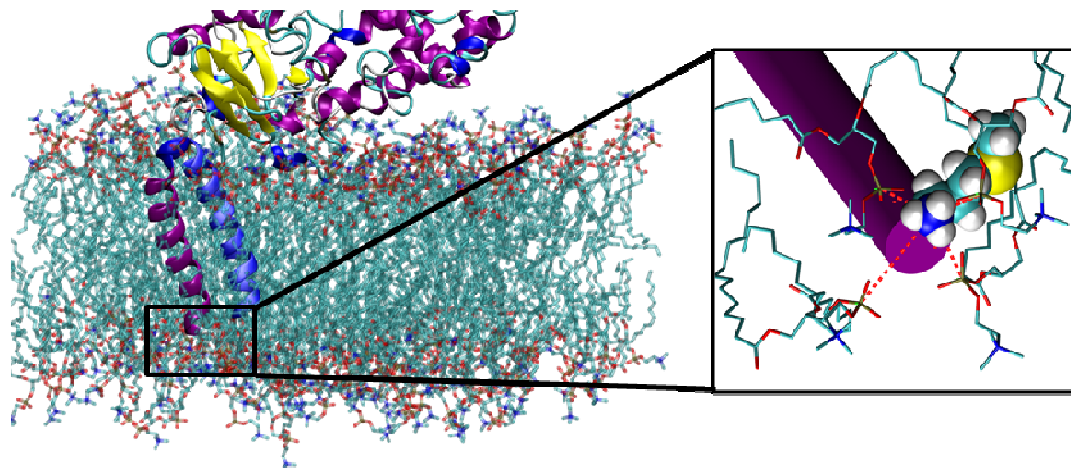


Figure 72 : Représentation de l'évolution de la position de l'hélice transmembranaire au sein de la membrane. A gauche, la position initiale de l'hélice au sein de la membrane est représentée en bleu et la position finale en violet. A droite, détail de l'interaction entre le résidu en N-terminal et les groupements phosphates des lipides

Les interactions entre les acides aminés de l'enzyme et la surface de la membrane évoluent également au cours des simulations, bien que les résidus identifiés lors de la construction du complexe jouent constamment un rôle majeur dans cette interaction. Le changement de l'interaction de l'hélice transmembranaire entraîne des modifications structurales de l'interaction de surface. Les résidus vont notamment interagir avec les parties lipophiles des lipides, mettant en avant un renforcement général de l'interaction enzyme/membrane. La participation de nouveaux résidus dans l'interaction de surface est également notée. Le Tableau 6 résume les résidus impliqués dans l'interaction enzyme/membrane.

	MD1	MD2
Temps de simulation (ns)	640	400
Résidus en interaction type hydrophobe	W35, V36, I38, V70, F69, V221, A222, P223, L224, W226, L227,	V34, W35, V36, I37, I38, F69, V221, A222, P223, L224, W226, L227
Résidus en interaction type liaison H	V36, I37	D228, W35
Résidus en interaction de type cation-π	W35, W226, F69	F69 (x2), W35, W226
Résidus en interaction ionique	D228, R66, R23	R66, R23, K102

Tableau 6 : Résumé des différents résidus impliqués dans l'interaction protéine/membrane, ainsi que la nature de leur interaction propre. Sont répertoriés les résidus dont l'interaction dépasse les 50 % d'existence durant la simulation concernée.

On observe que la chaîne latérale du résidu 35 pénètre de toute sa longueur dans la membrane. Ce mouvement induit ainsi la formation de nouvelles interactions hydrophobes, non prédites par le serveur OPM. La formation d'interaction hydrophiles (ionique et liaison hydrogène) est également observée. La surface chargée de la membrane entraîne la formation de ponts ioniques avec des résidus chargés (K102, D228, R66) proches de l'interface.

Enfin, on observe très rapidement dans nos simulations la formation d'interactions cation- π , qui sont maintenant reconnues pour avoir un rôle structurant important dans la formation de ces interactions protéine/membrane[181]. Notons la formation naturelle d'interaction entre les groupements méthylcholines des lipides et des centres aromatiques proches. Deux groupements tryptophanes (35 et 226) ainsi que la phénylalanine 69, interagissant à la surface, sont pour la plupart du temps impliqués dans ce type d'interaction. Ces acides aminés, de type tryptophane et phénylalanine, ont été identifiés comme étant les résidus principaux impliqués dans la formation de ces interactions cations- π , particulièrement avec les lipides de types phosphocholines (PC)[182]. La Figure 73 représente les interactions cations- π identifiées durant nos simulations.

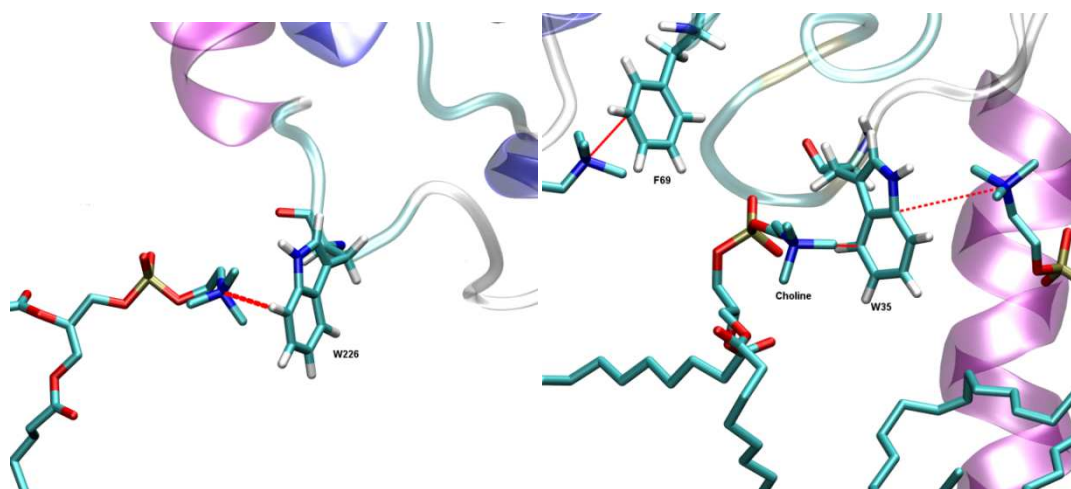


Figure 73 : Représentation des interactions cation- π observées durant les simulations de dynamique moléculaire.

Du fait de la dynamique des lipides constituant la membrane, de nombreux échanges lipide/lipide impliqués dans les interactions avec l'enzyme sont notés. Dans le cas du résidu F69 par exemple dans la simulation MD2, trois lipides interagissent seuls ou de manière concomitante avec la partie aromatique. Nous considérons que l'interaction cation- π existe lorsque la distance méthylcholine-cycle aromatique inférieure à 7 Å [172, 183]. Les analyses montrent qu'au moins une interaction existe durant 85% du temps de simulation, et montrent même l'existence de deux interactions de ce type en même temps sur le même résidu (cf Figure 74). Des résultats similaires sont obtenus pour les résidus W35 et W226.

Ainsi, les simulations de dynamique moléculaire valident les prédictions d'interactions proposées par le serveur OPM. Le réarrangement structural de l'hélice transmembranaire induit de nouvelles interactions hydrophobes mais également la formation d'interactions ioniques et surtout cation- π à la surface de la membrane. Les résidus aromatiques situés à l'interface sont ainsi impliqués dans la formation d'une interaction avec un groupement méthylcholine de la membrane de type POPC.

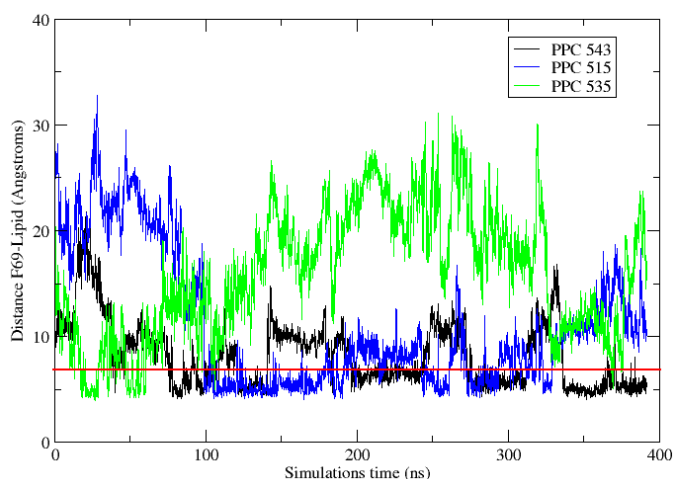


Figure 74 : Evolution des distances entre le résidu F69 et les trois lipides impliqués dans les interactions cation- π . Le trait rouge définit la limite au delà de laquelle on ne considère plus d'interaction.

I.3) Changements structuraux induits par l'interaction avec la membrane

Comme nous l'avons mentionné précédemment, la boucle F-G interagit de manière importante avec la membrane. Cette portion de l'enzyme est liée à l'ouverture des canaux 2 (et ses différentes variantes) et S, pouvant jouer un rôle dans la sortie de composés. La formation de ces interactions pourrait donc provoquer des effets non négligeables sur la structure tertiaire de l'enzyme, comme par exemple des effets d'ouverture et de fermeture des cavités identifiées au sein des cytochromes P450. Dans nos calculs, le RMSD de l'enzyme entière (par rapport à la structure reconstruite par homologie) est respectivement de 6 Å (simulation de 640 ns) et de 4,5 Å (simulation de 400 ns). La valeur élevée observée pour la première simulation est associée à des changements structuraux induits par effet allostérique de l'interaction protéine/membrane. Ainsi, durant nos simulations du complexe F3'H/membrane, même si nous n'avons pas été en mesure d'observer une sortie de ligand, nous avons identifié des phases d'ouverture et de fermeture des différents canaux de l'enzyme.

Au niveau du canal S, le résidu valine 205 joue le rôle de porte pour la sortie des molécules du site actif. Au cours de nos simulations, nous observons une ouverture partielle du canal S, par rotation de la valine 205. Nous n'observons cependant pas la totalité de l'ouverture du canal, les résidus 480 et 484 restant en interaction. Ce mouvement du résidu valine 205, que nous n'observons pas en simulation sans membrane, peut être attribué à l'interaction de la boucle F-G (résidu 218-229) avec la membrane. En effet, elle peut produire

un effet allostérique sur la structure proche, et notamment les résidus proches comme la valine 205. Cette ouverture partielle est montrée sur la Figure 75.

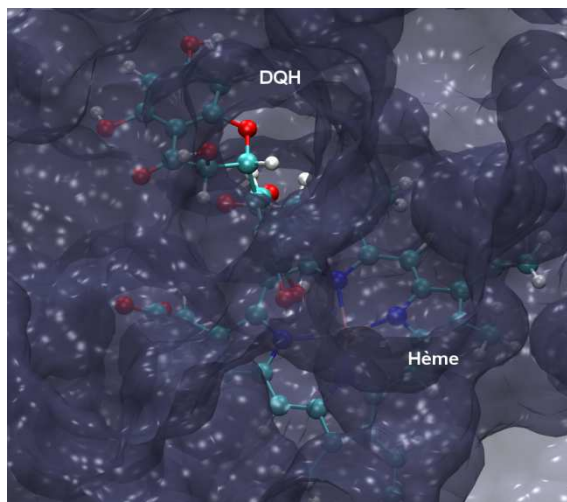


Figure 75 : Ouverture de la porte du canal S après 300 ns, par rotation de la valine 205. Le DQH et le groupement Hème sont représentés en *balls and sticks*

En ce qui concerne le canal 2, la conformation de l'hélice B-C est le facteur dominant définissant préférentiellement l'ouverture d'une des variantes de ce canal. Notamment, les résidus tyrosine 108, lysine 102 et phénylalanine 113 définissent plus précisément la cavité associée au canal 2e. Nous observons dans le cas de ce canal une ouverture très tôt dans notre trajectoire durant la simulation MD1. Les interactions hydrophobes entre ces 3 acides aminés disparaissent et on observe une ouverture totale du canal 2e. Cette ouverture est représentée sur la Figure 76.

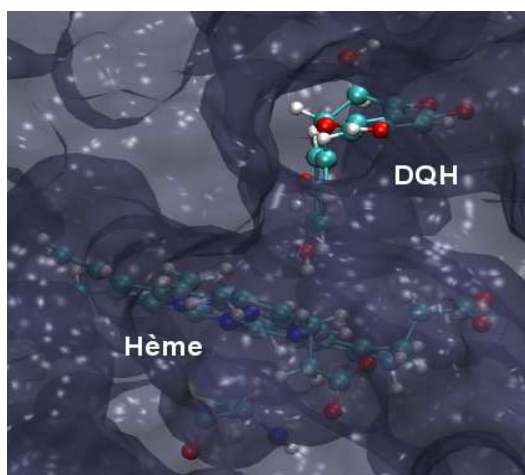


Figure 76: Ouverture du canal 2e après 200 ns de simulation. Le phénomène implique la rotation des résidus Y108 et F113. La molécule de DHQ et le cofacteur Hème sont représentés en *balls and sticks*.

Malgré ce phénomène d'ouverture, nous n'avons pas réussi à échantillonner le long de cette simulation une sortie naturelle de DHQ. Néanmoins, ces résultats présentent un type de changement structural qui n'a pas été observé durant les simulations de dynamique moléculaire sans membrane explicite. Une simulation de plus longue durée pourrait permettre l'observation de la sortie du DHQ par les canaux 2e et S. En ce qui concerne les autres canaux, nous n'observons pas d'ouvertures de leur porte (conformation de la boucle F-G en faveur du canal 2e) qui pourrait permettre la sortie du DHQ du site actif.

La prise en compte des interactions avec la membrane s'avère donc nécessaire si nous voulons prédire de manière convenable le comportement dynamique de l'enzyme F3'H. L'effet allostérique induit par ces interactions implique l'ouverture des canaux 2e et S. Ainsi, nous avons considéré ces deux voies comme sorties principales du DHQ, et donc comme connexions possibles avec les autres enzymes dans le cadre de la formation d'un métabolon à trois enzymes.

II) Premier modèle de métabolon

Avec une compréhension un peu plus avancée sur les interactions entre F3'H et la membrane, nous disposons maintenant de tous les éléments nécessaires pour proposer un premier modèle de notre métabolon impliquant les enzymes F3'H, DFR et LAR. En effet, nous avons construit tout d'abord un premier modèle d'un complexe DFR-LAR, qui a permis de mettre en lumière le phénomène de "substrate channeling", propriété principale et majeure des complexes multienzymatiques. D'autre part, nous avons également caractérisé les interactions F3'H/membrane et démontré que la prise en compte d'une membrane explicite est incontournable.

Afin de proposer un premier modèle de métabolon, nous avons mis en place des protocoles de docking protéine-protéine qui avaient été appliqués pour la construction des complexes bi-enzymatiques DFR-LAR. F3'H est considéré comme étant notre enzyme "récepteur" et la grille a été choisie afin de prendre en compte les facteurs stériques induits par la présence de l'interaction enzyme/membrane. Cette grille a été tronquée afin d'éviter l'échantillonnage d'une interaction mettant en jeu la partie de F3'H interagissant avec la couche bilipidique (cf Figure 77).

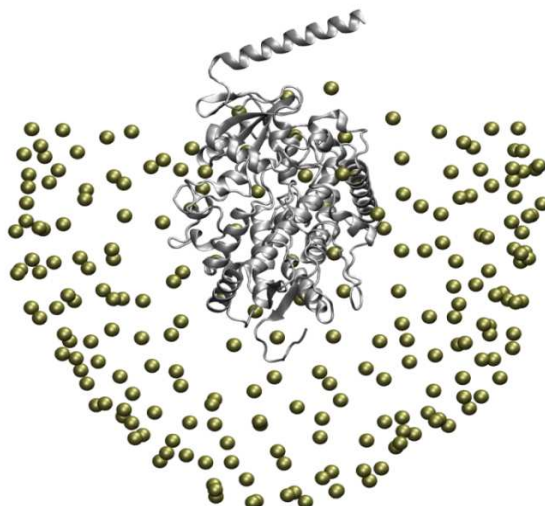


Figure 77 : Grille définie autour de F3'H pour les calculs de docking protéine-protéine du complexe DFR-LAR sur F3'H. La délimitation a été réalisée de sorte que la position de la membrane soit prise en compte.

Le complexe DFR-LAR S2 a été choisi comme protéine "ligand". Cette structure correspond au complexe sélectionné dans le chapitre 3, qui présentait la meilleure interaction protéine-protéine et pour lequel nous avons d'ailleurs observé le phénomène de "substrate channeling". Comme les simulations de dynamique moléculaire n'ont pas permis pour le moment de trouver un chemin préférentiel entre le canal 2e et S pour la sortie des molécules du site actif de F3'H, il a été décidé de proposer un modèle pour chaque cas. De même, aucune entrée préférentielle n'a été définie pour l'accès au site de DFR. Nous avons ainsi décidé de conserver les deux possibilités que sont la voie d'accès 1 et 2.

Au final, quatre modèles sont envisagés pour l'étude du complexe multienzymatique F3'H-DFR-LAR. Pour maximiser nos chances d'obtenir une structure satisfaisant tous nos critères, nous avons décidé de n'exclure aucune solution fournie par le protocole ATTRACT. Ainsi, les 10000 structures générées ont été compilées en 1500 clusters sur la base du RMSd ($\text{RMSd} < 1 \text{ \AA}$) de chacune d'elles. Devant le manque de données objectives, les 1500 clusters obtenus ont été conservés pour analyse.

Le nombre de contraintes structurales pour la sélection de structures pertinentes est ici plus élevé que dans le cas des complexes DFR-LAR. En plus du critère énergétique, deux critères structuraux doivent être considérés :

- De la même manière que précédemment, les cavités des enzymes F3'H et DFR doivent présenter une proximité spatiale. De manière générale, et après plusieurs tests, la limite maximale pour laquelle on observe une connexion entre les cavités est de 30 Å.
- La structure du complexe doit être compatible avec la présence de la membrane. DFR et LAR ne possédant pas de partie transmembranaire d'après le serveur OPM, le complexe DFR-LAR ne doit pas interférer (ou alors le moins possible) avec la surface de la membrane.

Si les trois critères sont satisfaits, le modèle pourra alors être sélectionné.

II.1) Modèle avec sortie par le canal S

Nous avons ici considéré une sortie par le canal S. Ainsi, le critère structural pour la sélection des modèles a été la distance entre un résidu définissant la porte du canal S (V205) et un résidu de la sortie 1 de DFR (F147). Toutes les structures présentant une distance inférieure à 30 Å sont conservées et analysées. La distance obtenue pour chaque pose est représentée sur la Figure 78. Les énergies d'interactions calculées varient ici de -32,16 à -16,35 RT.

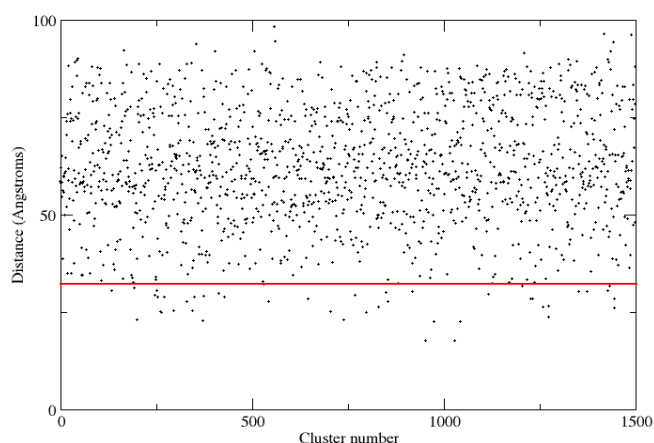


Figure 78 : Mesure de la distance entre la sortie 1 de DFR et la sortie du canal 2e de F3'H pour chaque cluster obtenu. La ligne rouge délimite le critère de sélection et d'analyse des modèles.

Parmi ces structures, les clusters 236, 324, 740 et 1028 sont très similaires (différence inférieure à 2 Å). Les autres clusters présentent soit une interaction d'une des deux enzymes avec la partie transmembranaire de F3'H, soit une non adéquation du complexe DFR-LAR avec notre modèle de membrane. Dans les deux cas, nous avons décidé d'exclure ces modèles.

Ainsi, nous avons choisi de conserver le modèle 236, qui satisfait le mieux les trois critères que nous nous sommes imposés : il s'agit de la meilleure pose sélectionnée en terme d'énergie ($E = -19,93$ RT), il présente à priori une bonne connexion entre les sites actifs, et peu d'interactions entre la membrane et les deux autres enzymes (DFR et LAR). La Figure 79 présente le modèle.

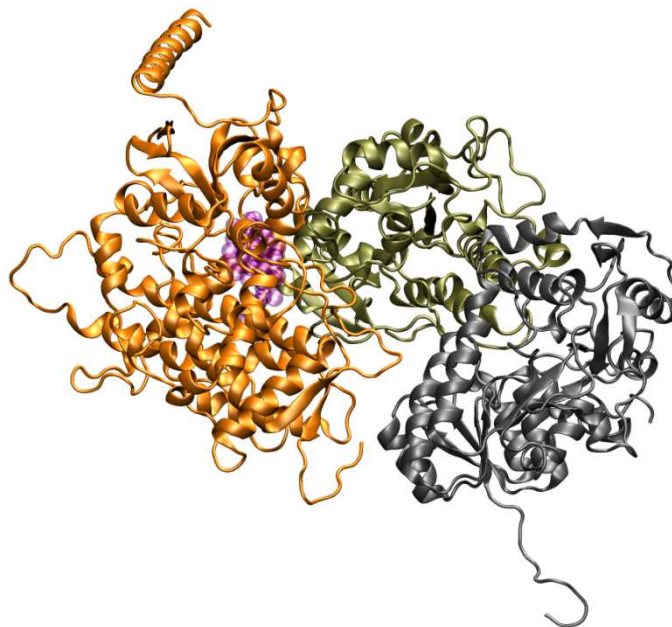


Figure 79 : Modèle du complexe F3'H-DFR-LAR obtenu par les calculs de docking protéine/protéine, dans le cas d'une sortie par le canal S. En violet est représentée la sortie du canal S de l'enzyme F3'H. En orange est représentée F3'H, en gris LAR et en marron-vert, DFR.

Ce cluster présente une interface similaire à celle obtenue pour le complexe S2 de DFR-LAR. 53 résidus participent à l'interface protéine-protéine (27 de type hydrophobe et 26 de type hydrophile). Nous pouvons noter que la majorité de l'interaction concerne les enzymes DFR et F3'H, même si une petite surface d'interaction existe entre F3'H et LAR (environ 10 % des résidus impliqués). Elle est majoritairement apolaire avec des interactions de type liaison hydrogène et deux ponts ioniques (D157 de F3'H-R909 de LAR et D199 de F3'H-N552 de DFR). La surface de l'interface protéine-protéine est de l'ordre de $1380 \text{ \AA}^2$. Ces résultats sont cohérents avec la définition d'une interface protéine-protéine transitoire telle que nous l'avons défini dans le chapitre 3. Les détails structuraux de l'interface sont répertoriés dans le Tableau 7.

Si maintenant nous considérons la voie d'accès 2 de DFR (distance entre V205 de F3'H et F147 de DFR), aucun cluster n'a pu être sélectionné. En effet, aucun des 1500 clusters ne satisfaisait les trois critères que nous avons défini précédemment. Ainsi, seul un modèle de métabolon avec une entrée dans le site actif de DFR par la voie d'accès 1 a pu être établi dans le cas d'une sortie de composé par le canal S de F3'H.

II.2) Modèle avec sortie par canal 2e

Ici, le critère structural a été choisi avec un résidu qui définit la porte du canal 2e, c'est à dire la tyrosine 108. Les valeurs des énergies d'interactions entre les enzymes dans ce cas varient de -28,54 à -15,63 RT. Les valeurs de toutes les distances pour les différents clusters entre le canal 2e et la voie d'accès 1 de DFR sont représentées sur la Figure 80.

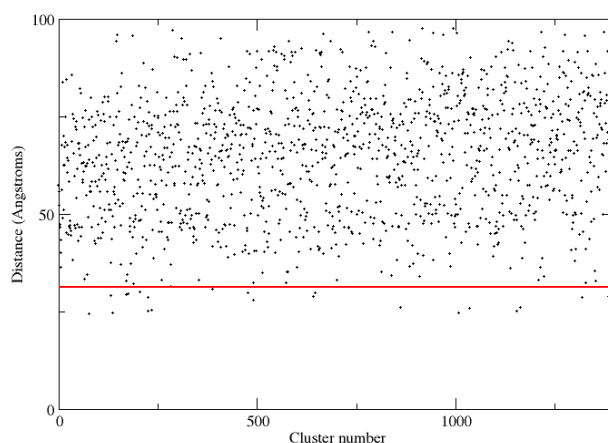


Figure 80 : Valeurs des distances entre la sortie du canal 2e de F3'H et la voie d'accès 1 de DFR pour tous les clusters. La ligne rouge désigne la limite de sélection des modèles, à une distance de 30 Å.

Beaucoup de modèles peuvent, de la même manière que dans le cas du canal S, être écartés de la sélection à cause de la non comptabilité entre la position du complexe DFR-LAR et la membrane. Deux clusters peuvent cependant être considérés : le cluster 136 (similaire au 226, 235, 834 et 1023) et le modèle 176. Le critère énergétique étant meilleur pour la structure du cluster 136 (-19,62 RT contre -19,23 pour la structure 176), c'est ce modèle que nous avons conservé. Ce dernier est représenté sur la Figure 81.

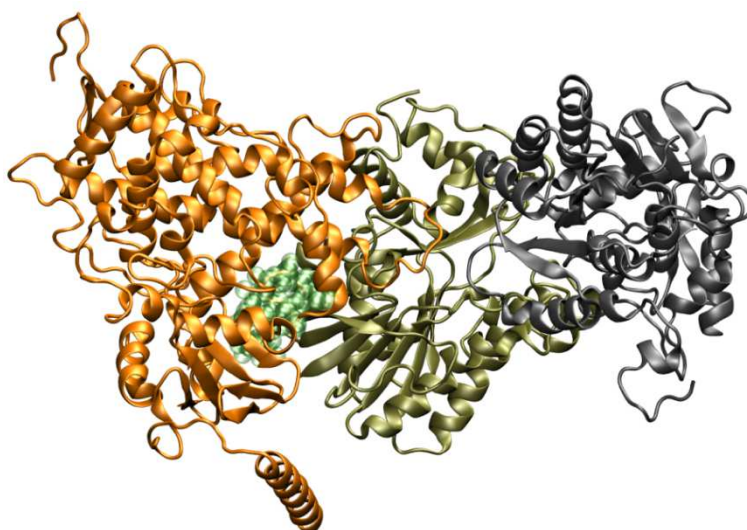


Figure 81 : Modèle du complexe F3'H-DFR-LAR, avec une sortie des composés par le canal 2e. En vert sont représentés les résidus définissant la porte du canal 2e.

L'interface de ce modèle est très similaire à celle obtenue dans le cas d'une sortie par le canal S : 47 résidus sont ainsi impliqués, avec une interface majoritairement hydrophobe (26 résidus hydrophobes vs 21 résidus hydrophiles). Un pont ionique a été identifié, ainsi que 5 liaisons hydrogène. La surface de l'interface est égale à 1314 Å². Elle ne fait cependant pas intervenir l'enzyme LAR. Seules des interactions entre F3'H et DFR sont identifiées. Les détails structuraux sont présentés dans le Tableau 7.

	Modèle avec canal S	Modèle avec canal 2e
Résidus impliqués (interface totale)	50	47
Nb liaisons H	4	3
Résidus impliqués (cœur interface)	27	22
Fraction non polaire (%)	47	40
SASA (Å²)	1370	1315
Interaction ionique	2	1

Tableau 7 : Détails structuraux de l'interface protéine-protéine pour les deux modèles de métabolon obtenus par les calculs de docking.

De la même manière que pour le modèle avec le canal S, nous n'obtenons pas de modèles respectant toutes nos contraintes structurales si nous considérons une entrée de molécules par la voie d'accès 2 de DFR. Ainsi, le modèle présenté par la Figure 81 est celui que nous proposons pour la formation d'un métabolon avec une sortie par le canal 2e.

II.3) Simulation de dynamique moléculaire

Le système à 3 enzymes, avec le canal S comme voie d'accès principale, a été le premier à être choisi pour la réalisation de simulation de dynamique moléculaire. Comme précisé dans le chapitre 1, le placement de la membrane a été réalisé avec le logiciel Maestro de la suite Schrödinger[144]. Dans ce cas-ci, 521 lipides de type POPC ont été rajoutés pour former la membrane. Le complexe multienzymatique DFR-LAR-F3'H et la membrane ont été placés dans une boîte de solvant de dimensions 160x130x170 Å³, pour un total de 310 000 atomes. 20 contre-ions ont été ajoutés pour assurer la neutralité du système (cf Figure 82)

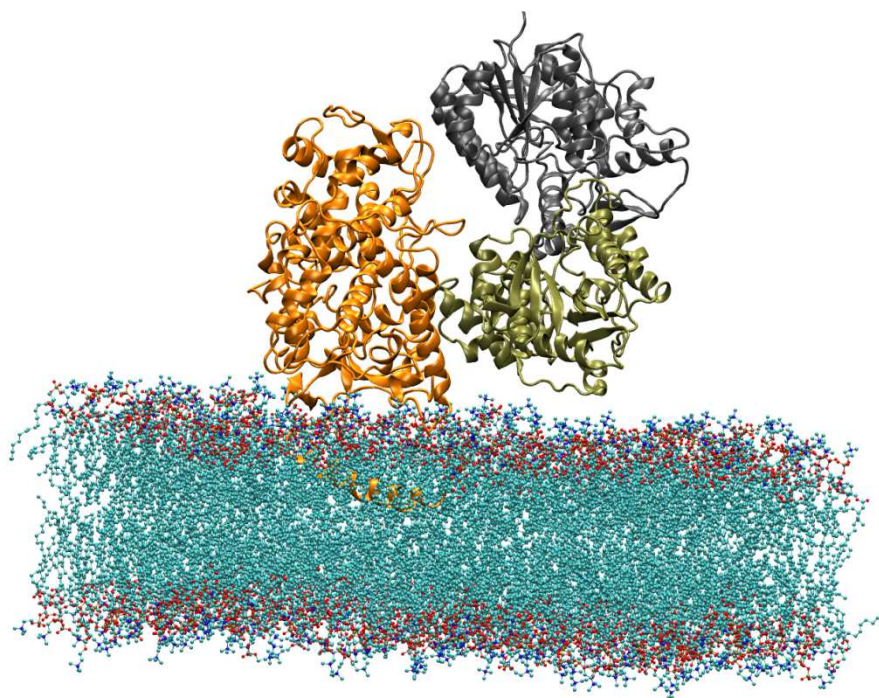


Figure 82 : Représentation du modèle de métabolon avec sortie par le canal S, ainsi que l'interaction avec la membrane. F3'H est représentée en orange, DFR en marron-vert et LAR en gris.

Ce système a été soumis à un protocole de simulation de dynamique moléculaire. Une première approche sur le comportement dynamique d'un tel édifice moléculaire peut être apportée. A ce jour, la simulation de dynamique moléculaire est toujours en cours. Sur les 100 premières ns simulées, nous pouvons noter que les interactions hydrophobes et de type liaison

hydrogène, relevées dans les modèles générés par ATTRACT entre les différentes protéines, restent stables. Une légère réorganisation de l'interface est cependant observée, ainsi que la création de nouvelles interactions ioniques entre les enzymes. Nous pouvons citer l'interaction entre le résidu D238 de F3'H et le résidu N556 de DFR. En ce qui concerne l'interaction D157-R909, elle reste stable durant la simulation (distance moyenne de 3,3 Å). Enfin, une autre interaction ionique est créée (mais avec un temps d'existence plus faible durant la simulation, environ 15%) entre le résidu D189 de F3'H et N552 de DFR, tandis que l'interaction D199-N552 identifiée dans le modèle initial est perdue en contrepartie.

III) Conclusion

Au delà de son rôle protecteur, la membrane cellulaire est impliquée dans de nombreux mécanismes biologiques de transformations/modifications de protéines. Elle permettrait également l'ancrage de certaines enzymes, permettant à d'autres de s'agréger menant ainsi à la formation de complexes multienzymatiques. Dans la biosynthèse des flavonoïdes, et plus particulièrement dans notre métabolon à 3 enzymes, l'enzyme F3'H est celle qui joue ce rôle. L'objectif du travail présenté était de prédire les interactions entre F3'H et la membrane cellulaire et de détecter les voies d'accès des molécules dans le site actif de F3'H. Nous avons par la suite raffiné ces résultats par simulation de dynamique moléculaire.

Les simulations de dynamique moléculaire du système F3'H/membrane valident les observations faites par le serveur OPM. Néanmoins, le nombre de résidus impliqués dans l'interaction augmente lors des simulations, et on observe notamment la formation d'interactions de type cation- π . Les trois résidus aromatiques (W35, W226 et F69) identifiés à l'interface sont tous engagés dans des interactions cation- π avec les groupements méthylcholines des lipides de la membrane. Par ailleurs, des ponts ioniques se forment également entre des résidus chargés et les groupements phosphates des lipides. La formation de ces interactions provoque des modifications structurales assez fortes (valeur de RMSD sur la structure de l'enzyme lors des simulations supérieures à 3,5 Å), et notamment l'ouverture des deux canaux 2e et S identifiés dans les simulations de type RAMD. Nous avons ainsi choisi de conserver ces deux canaux pour la prédiction des complexes multienzymatiques à trois enzymes.

Les calculs de docking nous permettent d'obtenir deux modèles différents avec la sortie 1 de DFR et les deux sorties de F3'H que nous avons retenues. Chacun présente une interaction de surface majoritairement hydrophobe, similaire à une interface protéine-protéine transitoire,

mais impliquant également des interactions de type liaison hydrogène et ioniques. Ainsi, nous pouvons proposer deux modèles de métabolon F3'H-DFR-LAR selon le mode de sortie du DHQ de l'enzyme F3'H. La membrane peut être prise en compte dans les deux cas, car nous n'observons pas de gênes stériques associées à la présence de DFR et LAR avec la couche bilipidique.

Ces résultats prouvent que la présence de la membrane est nécessaire pour une description correcte du comportement dynamique de F3'H. Ainsi, l'analyse d'un complexe multienzymatique nécessite la prise en compte de la membrane si nous la voulons la plus correcte possible. Nous avons ainsi pu proposer un modèle de métabolon à trois enzymes en compilant toutes les informations obtenues lors de ce travail. Les simulations de dynamique moléculaire sur ce complexe à 3 enzymes ont donné de premiers résultats satisfaisants, mais doivent néanmoins être continuées pour affiner les observations.

Conclusion Générale

Les flavonoïdes sont des métabolites secondaires très répandus dans le milieu végétal. Leurs propriétés antioxydantes et antiradicalaires sont bénéfiques pour l'Homme et certains d'entre eux jouent un rôle important dans la pigmentation des plantes et fruits ainsi que dans la sensation d'astringence perçue lors de l'ingestion de certaines boissons. Leur voie de biosynthèse est complexe et fait intervenir un grand nombre d'enzymes dont les différentes réactions aboutissent à la formation d'une grande diversité chimique (quelques milliers de composés identifiés). De récentes études proposent que l'efficacité de la biosynthèse reposerait sur la formation de complexes multienzymatiques appelés métabolons. Leur existence a été montrée par des expériences de biologie moléculaire sur différentes voies de biosynthèse, différentes de celle des flavonoïdes. Cette superstructure se formerait à proximité de la membrane du réticulum endoplasmique grâce à la présence d'enzymes jouant le rôle d'ancre à la membrane. Ce type de complexe autoriserait une connexion entre les sites actifs des différentes enzymes, permettant de faciliter le transfert des métabolites d'une enzyme à l'autre. Entre deux étapes catalytiques, les métabolites ne seraient donc pas amenés à diffuser dans le milieu intracellulaire, évitant ainsi les pénalités de désolvatation et entraînant un gain de temps au niveau de la diffusion. Ce phénomène est appelé "substrate channeling".

Afin de comprendre et de décrire d'un point de vue moléculaire le comportement d'un tel édifice, nous devons étudier les quatre points principaux qui caractérisent un métabolon : i) les interactions protéine-protéine ii) la réactivité propre à chaque enzyme iii) l'interface protéine-membrane et iv) les interactions entre les enzymes et les métabolites. L'objectif de ce projet de thèse était de proposer le premier modèle d'un métabolon impliqué dans la biosynthèse des flavonoïdes. Au terme de ce travail, nous sommes en mesure de présenter une structure multienzymatique impliquant trois enzymes en interaction avec une membrane. Ces trois enzymes, intervenant dans la formation des proanthocyanidines, sont la flavonoïde-3'-hydroxylase (F3'H), la dihydroflavonol-4-réductase (DFR) et la leucoanthocyanidine réductase (LAR). Ce choix est motivé par l'importance que revêt l'enzyme DFR : elle constitue en effet la dernière étape commune entre les proanthocyanidines et les anthocyanes, deux sous-familles des flavonoïdes.

L'étude théorique d'une telle structure ne peut passer que par l'utilisation de nombreuses méthodes de modélisation moléculaire. L'application d'une approche multi-échelle, au delà du défi qu'elle relève, permet de traiter à tous les niveaux les différentes

propriétés de ces complexes. La préparation des systèmes a nécessité une utilisation de divers outils de bioinformatique structurale, notamment la reconstruction par homologie. Les méthodes hybrides de type QM/MM nous ont permis de déchiffrer la réactivité de l'enzyme DFR, qui est le point central de notre étude. Concernant la prédiction de complexes protéine-protéine, un protocole de docking (protocole ATTRACT) a été utilisé en résolution gross-grain, pour permettre un échantillonnage rapide d'un grand nombre de structures possibles. Ces structures ont été ensuite raffinées par simulation de dynamique moléculaire en résolution atomique, nous permettant également de comprendre le comportement dynamique de tels complexes.

La première étape de notre étude fut de préparer les différents systèmes biologiques. Dans le cas de l'enzyme DFR, nous avons pris la structure cristallographique du complexe ternaire DFR-NADP⁺-DHQ représentée par le fichier PDB 2C29. Concernant LAR, une structure cristallographique d'un complexe ternaire existe également (fichier PDB 3I52). Cependant, certaines parties de la structure étaient manquantes et ont dû être reconstruites par diverses méthodes de bioinformatique structurale. En ce qui concerne F3'H, sa structure tridimensionnelle est inconnue. Or, la plupart des cytochromes P450 ont pour caractéristique de posséder une bonne conservation de leur structure tertiaire. Ainsi, un protocole de reconstruction par homologie nous a permis d'obtenir un modèle représentant l'enzyme.

Les études menées par la suite sur ces structures théoriques ont montré la pertinence de nos choix. En effet, outre la stabilité des édifices reconstruits, LAR et F3'H notamment, nos simulations de dynamique moléculaire ont montré que la structure de F3'H est en parfait accord avec les études menées sur des enzymes similaires, qu'elles soient expérimentales ou théoriques. Il est également intéressant de constater que le protocole ATTRACT, destiné à la définition d'interaction protéine/protéine, nous a permis de proposer une structure complète, stable et cohérente de l'enzyme LAR. Seule la position de certains acides aminés dans la structure cristallographique était manquante. Les résultats obtenus lors de cette partie, relativement technique, nous a permis d'enclencher avec confiance la suite de l'étude.

Une des forces des méthodes de modélisation moléculaire est d'être en capacité de décrire les différentes étapes d'un mécanisme enzymatique. Ici nous nous sommes attachés à décrire la réaction associée à la catalyse par l'enzyme DFR. Lors de cette réaction, l'enzyme fait intervenir un cofacteur (NADPH). Celui-ci fournit l'hydrure permettant la réduction du

groupement carbonyle du substrat, alors que deux résidus, Y163 et K167, sont impliqués dans une cascade de transferts de protons, menant à la formation du groupement hydroxyle. Pour analyser ce mécanisme, des méthodes hybrides de type QM/MM-MD, couplées avec un protocole de type Umbrella Sampling ont été appliquées. Ainsi, les différents effets de l'environnement enzymatique sur la réaction ont été pris en compte. Une approche avec deux contraintes, représentant respectivement le transfert d'hydrure et le transfert de proton provenant de la tyrosine, a été réalisée pour modéliser correctement la réaction. Ce choix a été fait après une série de calculs tests qui nous ont montré que une seule contrainte sur le transfert d'hydrure ne serait pas suffisante et surtout que les deux transferts de proton les plus éloignés du substrat n'étaient déclenchés qu'en fin de réaction. Les résultats montrent que, bien que la barrière d'énergie soit similaire pour deux substrats, DHQ et DHK, l'état de transition apparaît comme plus tardif dans le cas du DHK. Une différence notable des énergies libres de réaction peut être mise en avant : -8 kcal/mol pour DHQ, 0 kcal/mol pour DHK. Expérimentalement, il a été observé que le taux de conversion par l'enzyme de DHQ était plus important que celui de DHK. Nos résultats sont en accord avec ces observations dans le cas du DHK car, d'un point de vue thermodynamique, le substrat et le produit existent avec la même probabilité. Dans le cas du DHQ, la formation du produit est par contre clairement favorisée.

Toutes les études menées sur ce mécanisme ont montré que notre choix de deux contraintes pour représenter une coordonnée de réaction nécessairement plus complexe, impliquant notamment des angles de pyramidalisation, était valide. Ainsi, un des résultats le plus intéressant a été de montrer que ce mécanisme ne pouvait être considéré ni comme concerté, ni comme par étape, car il existait un certain découplage entre les transferts d'hydrure et de protons.

Un état pré-réactif, appelé NAC (Near-Attack Conformer), a été identifié le long du profil réactionnel. Ce type d'état correspond généralement à des structures des réactifs plus proches de la structure de l'état de transition, mais n'ayant pas commencé à réagir. Ici, il s'agit plus particulièrement de l'échantillonnage d'une structure s'éloignant de celle des réactifs dans la direction de l'état de transition sans qu'il y ait de début de transferts d'atomes. Ces structures ont été échantillonnées lors de simulations de dynamique moléculaire QM/MM sans contrainte, validant ainsi le choix de la représentation de la coordonnée de réaction de

réduction du substrat. En favorisant la formation de ces états, on peut considérer que l'enzyme DFR diminue l'énergie d'activation de la réaction. Ces calculs ont mis en évidence une réorganisation du complexe enzyme-cofacteur-ligand, de sorte que les réactifs soient dans une conformation facilitant la réaction.

Cette analyse est totalement compatible avec les temps classiquement associés aux différents phénomènes liés à une transformation enzymatique. En effet, outre les phénomènes de diffusion, de l'ordre de quelques nanosecondes à quelques microsecondes, qui permettent l'arrivée du substrat dans la cavité, il est considéré que le placement adéquat enzyme/substrat, qui prend quelques nanosecondes, déclenche la réaction enzymatique, qui est de l'ordre de quelques picosecondes. L'existence de structures de type NAC est pleinement en accord avec cette vision.

Ce modèle simplifié ne prend pas en compte les possibles effets allostériques induits par la formation d'interactions de type protéine-protéine présentes dans un métabolon. Ces complexes peuvent moduler la réactivité intrinsèque de chaque enzyme. Ces résultats pourront être raffinés lors d'une étude de réactivité de DFR prenant en compte les enzymes partenaires.

Une fois le produit de DFR, la leucocyanidine (LCC), formé, la seconde étape fut d'étudier son transfert vers la seconde enzyme. Pour cela, nous avons dû prédire des structures de complexes protéine-protéine présentant une connexion entre sites actifs pour autoriser le phénomène de "substrate channeling". Dans un premier temps, il a donc été nécessaire de déterminer les différentes voies d'accès des molécules au sein des sites actifs de DFR et de LAR. Pour ce faire, des simulations de type brute force et de type RAMD ont été appliquées sur les systèmes tertiaires enzyme-cofacteur-substrat. Trois voies d'accès différentes ont ainsi été mises en évidence pour l'enzyme DFR, et seulement une pour l'enzyme LAR. A noter que ces mêmes protocoles ont été mis en place pour l'enzyme F3'H et sont en complète adéquation avec des études précédemment réalisées, basées sur des simulations de type SMD.

Bien que ces protocoles, qu'ils soient hors équilibre ou non, ne nous aient pas permis dans un premier temps d'établir un profil énergétique lié à l'entrée ou à la sortie des substrats/produits, les conclusions de ces analyses sont très claires et permettent une recherche ciblée des interactions entre protéines. Une fois ces voies identifiées, nous avons utilisé le

protocole ATTRACT pour la réalisation des calculs de docking protéine-protéine. Deux critères ont été considérés pour le choix des structures : l'énergie d'interaction protéine-protéine et la distance entre les résidus à l'entrée des deux sites actifs. En s'appuyant sur ces deux critères, énergétique et/ou structural, deux modèles ont été sélectionnés parmi les 100000 structures générées par le protocole. Ils présentent des propriétés structurales à l'interface protéine-protéine proche de ce que l'on retrouve classiquement pour les complexes protéine-protéine transitoires.

Toutefois, la définition de complexes protéine/protéine en interaction n'est pas gage de stabilité des édifices. Il est donc nécessaire d'appliquer des protocoles de dynamique moléculaire sur ces modèles (avec au sein de DFR, les produits de la réaction), qui permettront d'estimer les réorganisations au niveau des zones d'interaction et la stabilité au cours du temps de ces complexes. Il apparaît que nos deux complexes sont susceptibles de se dissocier. Mais la répétition de ces simulations a également montré que ces interactions peuvent perdurer suffisamment longtemps pour que les produits puissent diffuser hors de la cavité. Ces simulations sont en cohérence avec l'existence transitoire de ce type de complexe enzymatique mais aussi et surtout avec le fait qu'ils puissent être un excellent point de départ pour étudier les phénomènes de "substrate channeling".

Ainsi, même si nous avons pu échantillonner la libération de produits vers la phase solvant, nous avons surtout observé une diffusion d'un produit depuis le site actif de DFR vers celui de LAR. Le long de ce chemin, nous avons calculé l'énergie d'interaction du produit de DFR, LCC, avec le complexe multienzymatique. Trois phases ont été identifiées lors de ce transfert : i) sortie de la cavité de DFR et migration vers l'interface DFR/LAR ii) exploration de la surface de LAR iii) entrée dans la cavité de LAR. La différence d'énergie d'interaction entre l'état initial et l'état final est de -14 kcal/mol et est en faveur d'un transfert vers l'enzyme LAR. La mesure du nombre de molécules d'eau le long du transfert montre que la leucocyanidine n'est jamais entièrement solvatée, ce qui constitue le cœur du phénomène de "substrate channeling". La diminution des pénalités de désolvatation facilite le transfert du substrat d'une enzyme à l'autre sur un temps plus court.

Il est évident que le phénomène de diffusion, qui est un phénomène naturel, est associé à un chemin qui évite les barrières de transition et suit les vallées énergétiques. Ici, l'existence de phases correspondant à une perte d'interaction (complexe

enzymatique/substrat), associés à une montée en énergie, ne concerne pas l'ensemble du système, mais uniquement l'interaction entre deux entités. Cette interaction est spatialement limitée, du fait de la taille relative de LCC par rapport à celle du complexe multienzymatique. Cela nous amène donc à penser que les mouvements de chaque enzyme, la réorganisation de l'interaction protéine/protéine, mais aussi la dynamique des molécules de solvant jouent un rôle dans ce transfert du LCC. Nous avons confiance en ce modèle que nous proposons car il découle de simulations effectuées sans contraintes, laissant la possibilité à la molécule de "trouver" son chemin. Les résultats des analyses effectuées le long du chemin (nombre de molécules d'eau solvatant LCC, énergie d'interaction) sont totalement compatibles avec les hypothèses liées à un phénomène de "substrate channeling".

L'addition d'une troisième enzyme, ici F3'H, et de l'interaction de l'édifice complet avec la membrane, nous permet, dans une dernière partie, d'anticiper ce que peut être un métabolon. Dans le cas de la biosynthèse des flavonoïdes, et plus particulièrement de la participation de l'enzyme F3'H, il s'avère nécessaire de considérer explicitement un modèle de membrane cellulaire. Ce point est extrêmement intéressant car il peut être lié à une augmentation de la statistique de rencontre des enzymes et surtout à la stabilisation du complexe tri-enzymatique. En effet, l'enzyme F3'H possède une partie de sa séquence structurée en hélice, prédite comme étant transmembranaire. Cette protéine pourrait alors jouer le rôle d'ancre à la membrane et cette immobilisation pourrait permettre une augmentation de la probabilité de rencontre avec DFR notamment. Une fois le complexe enzymatique formé, l'interaction de DFR avec la membrane pourrait alors stabiliser le complexe dans son entier. Ici, est présenté un premier modèle de métabolon et si le dernier point a pu être évoqué partiellement, le premier point concernant la statistique n'a pas encore été étudié.

L'ancrage à la membrane de F3'H est assurée par sa partie N-terminale. Le serveur OPM a permis d'identifier une interaction de surface, mettant en jeu des résidus classiquement trouvés dans ce type d'interaction. Le fait le plus important est la création d'interactions cation- π , entre les têtes polaires des lipides (groupement méthylcholines) de la membrane avec les résidus aromatiques présents à l'interface de type tryptophane ou phénylalanine. Ces interactions sont connues pour jouer un rôle important dans la structuration des interactions de surface entre membrane et protéine. Ces interactions de surface avec la membrane provoquent

des effets allostériques sur la structure de F3'H. L'ouverture de deux canaux permettant l'accès au site actif, qui avaient été échantillonnés par les simulations de type RAMD, est observée. Même si nous n'avons pas réussi à observer une sortie du produit de la réaction, ces canaux peuvent être considérés comme les deux voies d'accès principales pour F3'H, car nous n'observons aucune autre ouverture. Ainsi, ces deux voies ont été considérées pour la définition des complexes à 3 corps F3'H-DFR-LAR. Des calculs de docking protéine-protéine, avec le protocole ATTRACT, ont été réalisés pour proposer un premier modèle de métabolon impliquant les trois enzymes. Les contraintes structurales sont plus nombreuses que précédemment car la position du complexe doit être compatible avec la présence de la bicouche lipidique. Ainsi, les calculs nous ont permis d'extraire deux modèles de métabolon, un pour chaque canal considéré de F3'H. L'interface de l'interaction F3'H-DFR correspond à une interface de type complexe transitoire. Les simulations de dynamique moléculaire ont été lancées sur ce type de structure et sont en cours de réalisation. Dans l'attente d'une analyse plus poussée, nous pouvons mettre en avant la cohérence des modèles proposés avec l'hypothèse d'un métabolon : ancrage à la membrane, complexe transitoire, "substrate channeling"...

Au terme de ce travail, nous avons été en mesure de proposer un premier modèle de métabolon, impliquant 3 enzymes et une membrane. Ce travail a été réalisé en analysant plusieurs de ses fonctions indépendamment : catalyse enzymatique, définition d'enzymes en interaction, caractérisation de voie de diffusion entre sites actifs... Ces résultats ouvrent la porte à de nombreuses études complémentaires. Par exemple, il a été montré expérimentalement que l'association de deux enzymes menait à une meilleure efficacité du processus catalytique global alors que chacune des deux étapes est moins efficace. Cette conclusion permet d'imaginer i) que l'étude plus précise des réactions enzymatiques lors de l'interaction entre protéines pourrait mener à des chemins réactionnels modifiés, ii) que les phénomènes de diffusion des substrats dans le milieu intracellulaire seraient le point limitant de ces transformations en cascade. Ainsi, de nouveaux calculs devront être effectués pour corroborer ou infirmer ces hypothèses, en complément de nouvelles expériences. Des phénomènes de compétition entre enzymes peuvent également être imaginés. En effet, dans notre cas, nous avons choisi de modéliser une des voies passant par LAR après que le substrat ait réagit avec F3'H et DFR, menant à la formation des proanthocyanidines. A la sortie de DFR, il est envisageable que le produit de la transformation enzymatique soit transféré vers

ANS (Anthocyanidine Synthase), qui orienterait la biosynthèse totale vers la production d'anthocyanes. Il serait alors possible, moyennant une solution bien établie de calcul d'énergie d'interaction protéine/protéine, de comparer les deux voies en terme de stabilité des complexes et de diffusion des substrats entre enzymes. Au final, ce travail tend à montrer que si des études sur des enzymes isolées restent incontournables, il sera bientôt nécessaire de considérer des complexes multienzymatiques pour déchiffrer les voies métaboliques définies dans les plantes.

Chapitre Annexe :
Modèle d'inhibition de
DFR par excès de
substrat

Les transformations catalysées par des enzymes couvrent un très large spectre de réactions chimiques. La force de ces entités biologiques provient également du fait que leur organisation spatiale peut permettre l'intervention d'un troisième composé, généralement un cofacteur, afin de mener à bien la réaction. Co-facteur et substrat doivent donc avoir une certaine affinité pour le site actif de l'enzyme par la création de nombreuses interactions. Celles-ci peuvent mettre en jeu des atomes qui ne sont pas directement impliqués dans la réaction, on parlera alors de site de reconnaissance. D'autres, au contraire, auront un rôle actif dans la réaction. Ils doivent alors être idéalement placés afin que la molécule soit stabilisée dans sa configuration réactive et seront également « activés » afin que le rôle de catalyseur de l'enzyme devienne effectif. On peut alors parler de site d'activation ou plus généralement de site actif.

L'identification de ces résidus peut être réalisée en inspectant la structure cristallographique d'un complexe enzyme/substrat, si celle-ci est connue. Des expériences de biologie moléculaire, telles que la mutagenèse dirigée, permettent de mettre en avant le rôle d'un acide aminé en particulier. Si la séquence génomique codant la protéine est connue, sa modification permet de modifier la séquence primaire de l'enzyme, et ainsi de remplacer un acide aminé d'une position précise par un autre. Bien entendu, effectuer la mutation de tous les acides aminés serait trop long et trop coûteux, le choix rationnel d'acides aminés à muter peut se faire à l'aide d'une structure tridimensionnelle de l'enzyme. Ainsi, si après mutation il est observé une diminution de l'activité enzymatique, nous pouvons en déduire que le résidu muté doit jouer un rôle dans la reconnaissance du substrat par l'enzyme. La mutation peut aussi induire d'autres effets, comme par exemple une modification de la sélectivité de l'enzyme par rapport à différents substrats. Malgré tout, modifier le type d'un acide aminé peut également jouer un rôle sur le positionnement des acides aminés voisins ou, plus rarement, modifier la structure globale de la molécule. La méthode de choix serait donc de muter les acides aminés d'intérêts et de construire une nouvelle structure tridimensionnelle. Ceci est rarement effectué du fait des coûts, temporels et financiers, de telles études.

Cependant, la stabilisation d'un composé au sein de la cavité ne signifie pas systématiquement que cette molécule réagira avec l'enzyme. Certains composés s'associent avec l'enzyme et induisent une perte de fonctionnalité de celle-ci. Ces molécules sont appelées

inhibiteurs. Lorsqu'un composé interagit avec l'enzyme et engendre une modification de la structure de celle-ci pour la rendre inopérante, on parle alors de régulation allostérique.

Dans le cas de DFR, la réaction fait intervenir un substrat et un cofacteur. Ce dernier fournit l'hydrure qui va permettre la réduction du groupement carbonyle du DHQ. L'enzyme, quant à elle, va permettre la protonation de l'alcoolate pour ainsi former le produit de la réaction. Le cofacteur NADPH s'oxyde durant la réaction pour former le NADP⁺. La Figure 83 montre l'équation bilan de la réaction enzymatique.

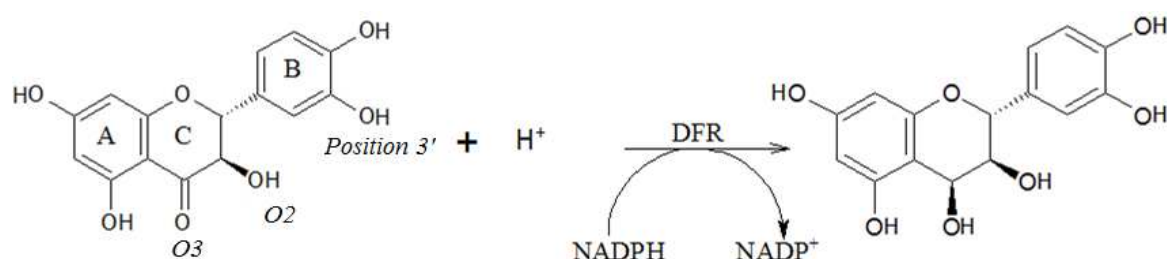


Figure 83 : Equation bilan de la réaction enzymatique se déroulant au sein de l'enzyme DFR.

En italique est précisée la nomenclature officielle des flavonoïdes.

Dans le cas de DFR, l'étude de Trabelsi *et al.* montre qu'un excès de substrat (DHQ) peut inhiber l'enzyme et réguler son activité[71]. Ici, l'hypothèse mise en avant serait que le substrat occuperait le site de reconnaissance du cofacteur, inhibant ainsi l'activité de l'enzyme. Selon leur étude, les auteurs ont ainsi proposé plusieurs modèles cinétiques pouvant expliquer ce phénomène d'inhibition à haute concentration de substrat. Nous avons tenté d'apporter des éléments de réponse à cette hypothèse et dans ce chapitre seront présentés nos résultats concernant ce phénomène.

1) Les différents modèles d'inhibition possibles

L'article de Trabelsi *et al.* présente différents modèles pouvant expliquer le phénomène d'inhibition de l'enzyme DFR par excès de DHQ. Cette étude, basée sur la mesure de la vitesse de conversion de DHQ à différentes concentrations de substrat, de cofacteur ou d'enzymes, propose 7 modèles différents. Parmi ceux là, nous mettons en avant 2 modèles préférentiels : les modèles 1 et 2.

- **Modèle 1 :** Inhibition provenant soit de la formation d'un complexe DFR-DHQ, soit par la création d'un complexe ternaire abortif DFR-NADP⁺-DHQ. Dans le premier

cas, le cofacteur ne peut plus être reconnu dans la cavité soit par gêne de son entrée par le substrat, soit par modulation allostérique induite par ce complexe. Dans le second, on peut imaginer qu'une fois la réaction effectuée, le cofacteur oxydé, NADP⁺, reste en interaction avec l'enzyme alors que le produit de la réaction, la leucocyanidine (LCC) est sortie du site actif. Alors, si les conditions du milieu consistent en un excès de substrat (DHQ), une molécule de DHQ pourrait interagir avec le complexe DFR-NADP⁺, avant la régénération de celui-ci, et ainsi former un complexe ternaire abortif.

- **Modèle 2 : Inhibition par la formation d'un complexe ternaire DFR-DHQ-NADPH abortif ou le complexe DFR-NADP⁺-DHQ.** Dans le premier cas, le DHQ entrerait en premier dans la cavité, suivi du NADPH. Le placement des partenaires biologiques ne correspond pas à celui nécessaire à la faisabilité de la réaction, entraînant une non production de l'enzyme. Le deuxième cas est similaire à celui du modèle 1.

L'article conclut que le mécanisme par inhibition impliquant la formation d'un complexe DFR-DHQ est l'hypothèse la plus probable. Ce type d'inhibition compétitive a déjà été relevée précédemment, mais reste un cas à part et peu étudié[184-186]. Cette régulation de l'activité de l'enzyme en cas d'excès de DHQ permettrait de ne pas accumuler en trop grande quantité la leucocyanidine dans le milieu intracellulaire. En effet, la présence de cette molécule instable et en grande concentration pourrait entraîner des effets néfastes sur la cellule.

Le complexe DFR-NADP⁺-DHQ est abortif. Cette structure a été cristallisée et est représentée par la structure PDB 2C29. Sa formation est donc possible et mène à un ensemble thermodynamiquement stable, qui expliquerait cette inhibition. Ceci dit, l'étude tend à montrer que la formation d'un complexe DFR-DHQ, en cas d'excès de substrat, était plus en accord avec les résultats cinétiques.

Notre travail s'est donc focalisé sur le modèle impliquant la formation d'un complexe binaire abortif DFR-DHQ. Nous avons réalisé des simulations de dynamique moléculaire sur différentes structures de ce complexe, générées initialement par calculs de docking, pour prédire leurs propriétés dynamiques. Dans un second temps, pour estimer l'interaction entre le substrat et l'enzyme, des protocoles de type MM-GBSA ont été appliqués sur les différents systèmes.

II) Etude de l'association DFR-DHQ

II.1) Protocole de simulation

Le protocole de dynamique moléculaire utilisé est présenté dans la méthodologie, pages 91. Afin de vérifier la possibilité d'une entrée de DHQ dans l'apoenzyme, des simulations, sans cofacteur et avec 5 DHQ autour de l'enzyme ont été réalisées. Les temps de simulation sont compris entre 100 et 350 ns. Une contrainte entre le centre de masse de chaque ligand et le centre de masse de la protéine a été appliquée pour empêcher les molécules d'être à une distance supérieure à 35 Å, ceci afin d'augmenter l'échantillonnage d'une possible entrée de DHQ dans le site de DFR.

II.2) Résultats

Les différentes simulations montrent l'entrée du substrat dans la cavité par la voie d'accès 2, mais aucunement du côté de la voie d'accès 1 (cf Chapitre 3). En effet, l'absence de NADPH au sein de la cavité induit la création d'interactions de type hydrophobe entre P204 et M88 (cf Figure 84), qui va provoquer la fermeture de cet accès (distance entre l'atome de soufre du résidu méthionine et l'atome de carbone CE1 du cycle pyridine du résidu proline égale à 4,23 Å). Cette interaction hydrophobe jouerait ainsi le rôle de "porte" et empêcherait l'entrée de substrats de ce côté de la cavité en l'absence de NADPH. Lorsque celui-ci est présent dans la cavité, il interagit avec la proline et la méthionine, ce qui permet une ouverture permanente de la cavité (distance supérieure à 8 Å).

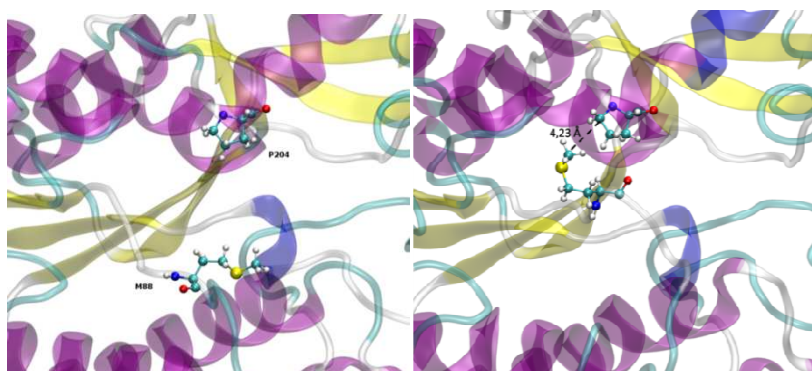


Figure 84 : Identification de la "porte" de la voie d'accès 1. A gauche, au début de la simulation, on observe que la porte est ouverte. A droite, nous voyons au cours du temps la création de cette interaction hydrophobe provoquant la fermeture de la cavité.

Ces simulations montrent que le DHQ peut accéder au site actif même en absence de NADPH. Ainsi la formation d'un complexe binaire de type DFR-DHQ peut être envisagée.

III) Calculs de docking sur l'enzyme seule

III.1) Protocole de simulation

Les simulations de dynamique moléculaire suggèrent que la formation d'un complexe DFR-DHQ est possible. Les calculs de DM brute force nécessitent des moyens de calculs conséquents. Nous avons choisi d'échantillonner différentes conformations du complexe DFR-DHQ à l'aide de calculs de docking. Ceux-ci seront raffinés par la suite par de courtes simulations de DM. Ainsi, une campagne de docking a été réalisée afin de déterminer les différentes positions que peut prendre le DHQ dans l'enzyme vide. Ils ont été réalisés à l'aide du logiciel Autodock Vina. La grille définie pour les calculs est de dimension 26x26x28, permettant ainsi d'englober tout le site actif de DFR. Après quelques tests, le paramètre "Exhaustiveness" a été réglé sur une valeur de 12.

III.2) Résultats

De ces calculs se dégagent trois structures de complexes enzymes-substrats. Les trois clusters regroupent au total 20 positions différentes :

- Un ensemble de positions regroupées autour de la pose cristallographique du complexe ternaire. En absence de cofacteur, différentes orientations du substrat sont mises en avant. Il correspond au site B sur la Figure 85.
- Un ensemble localisé sur le site de liaison du cofacteur, correspondant à la position A.
- Un troisième groupe de position au niveau de la voie d'accès 2 du site actif (Voir chapitre 3), associé au site C.

Ces trois clusters sont représentés sur la Figure 85. Deux positions pour le groupe associé à la position cristallographique sont représentées.

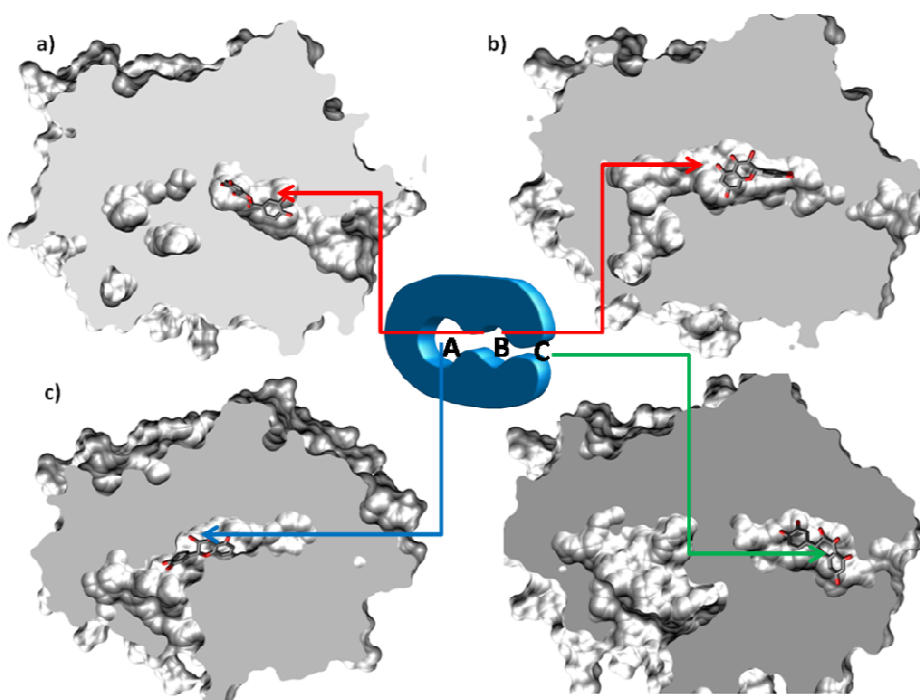


Figure 85 : Coupe transversale de l'enzyme montrant ainsi les 3 clusters obtenus par calcul de docking. Les sites A, B et C représentent respectivement le site du cofacteur, la position cristallographique du DHQ dans le complexe ternaire et la pose associée à l'entrée de la cavité par la voie d'accès 2. Pour le site B, deux positions sont représentées.

La formation d'un complexe abortif DFR-DHQ, ce dernier tant situé dans le site de fixation du NADPH, est un élément de réponse au phénomène d'inhibition de l'enzyme. Par ailleurs, nous échantillonnons plusieurs positions autour de la position cristallographique comme le montre la Figure 85. Certaines révèlent une orientation différente de celle qui permet la réactivité.

Pour le premier cluster, représenté par les images a) et b) de la Figure 85, les résidus impliqués dans l'interaction avec le substrat sont pratiquement tous les mêmes que dans le complexe ternaire. On retrouve des liaisons hydrogènes fortes avec les résidus de reconnaissance (N133 et Q227) ainsi qu'un des résidus de la triade catalytique, S128. Les résidus A129, M88, L192, V193 et V194 sont impliqués dans des interactions de type hydrophobe avec les cycles C et B de la dihydroquercétine.

Dans le cas du second cluster, représenté par l'image c), le DHQ est situé dans le site de liaison du NADPH. Les résidus G15, A85, M88, T86, P190, et K167 interagissent de

manière hydrophobe avec le substrat (notamment le cycle C), tandis que des liaisons hydrogène sont créées avec les résidus S127, I17 et S14.

Enfin, un troisième cluster a été identifié à l'interface entre la phase solvant et l'enzyme, au niveau de la voie d'accès 2. Ce site peut être assimilé à une position transitoire lors de l'entrée de la molécule dans la cavité vide. Le substrat interagit avec A129, I134, L192, I222, R224 et Q225 avec des interactions de type Van der Waals, tandis que les résidus , N133, I222, R224 et Q225 forment de faibles interactions hydrophiles. Le détail des résidus impliqués dans les trois clusters sont représentés sur la Figure 86.

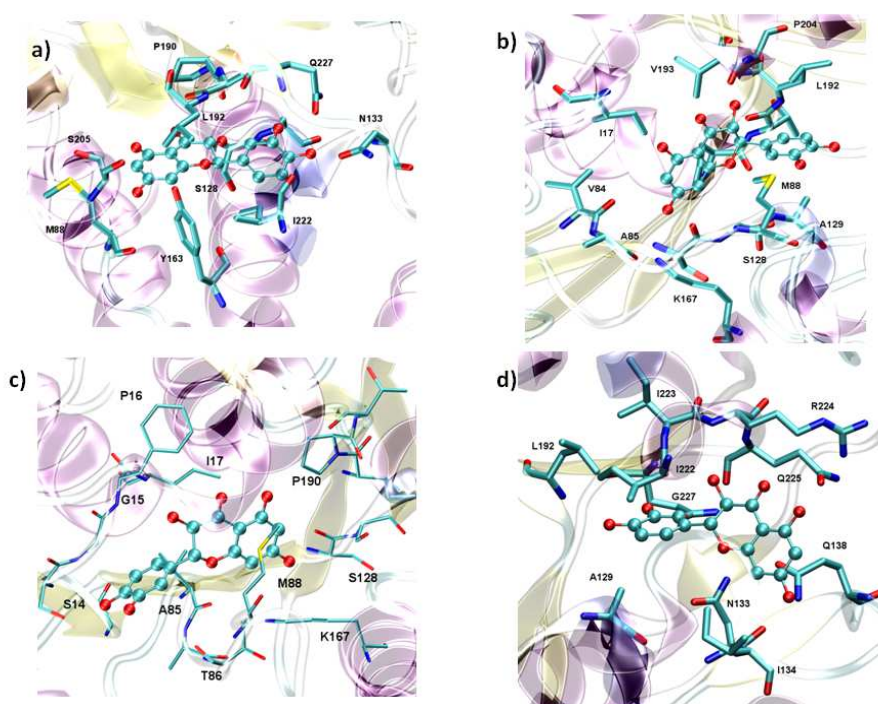


Figure 86 : Détails des interactions dans le cas des trois clusters identifiés durant les calculs de docking. a) et b) montrent les deux meilleures poses du cluster de la pose cristallographique, de la même manière que la Figure 85, c) représente la meilleure pose du cluster associé au site de liaison du cofacteur et d) présente la pose associée à la sortie 2.

Une analyse des propriétés physico-chimique des différents sites de la cavité montre que le site A présente un caractère amphiphile (autant hydrophile qu'hydrophobe), tandis que les deux autre sites ont un caractère plutôt hydrophobe (ratio hydrophobe/hydrophile respectif des sites B et C : 60/40 et 61,5/38,5). Ces observations apportent des éléments de réponses quant au site préférentiel d'interaction avec le DHQ. En effet, les flavonoïdes présentent des aspects autant hydrophobes (caractère aromatique) qu'hydrophiles (liaisons hydrogène avec

groupements hydroxyles), mais sont solubles dans de nombreux solvants protiques polaires (qui permettent par ailleurs leur extraction)[187]. Ainsi, le DHQ interagirait de manière préférentielle avec le site présentant le plus de caractère hydrophile. Un calcul local du potentiel électrostatique montre que le site du NADPH présente également des propriétés électrostatiques plus prononcées que les autres sites, ce qui appuie notre précédente observation sur la préférence du DHQ pour la liaison avec le site A. L'évolution de ce potentiel est représentée sur la Figure 87.

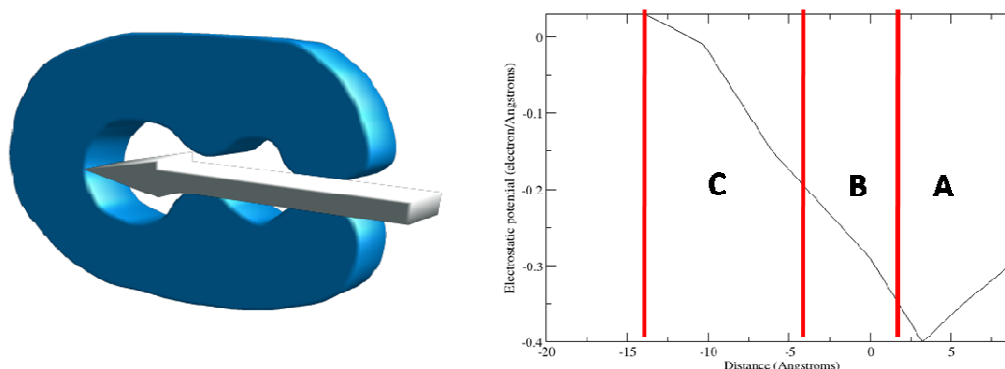


Figure 87 : Evolution du potentiel électrostatique le long de la cavité de DFR. A gauche est représenté le sens de la coordonnée de réaction. A droite est montrée l'évolution du potentiel électrostatique en fonction de la position du substrat dans la cavité. L'origine a été choisie à la position cristallographique de DHQ dans le complexe ternaire enzyme-cofacteur-substrat, dans le site B.

Pour raffiner l'interaction ligand-récepteur et les observations obtenues sur la base des calculs de docking, chaque cluster a été soumis à des simulations de dynamique moléculaire. Dans le cas du cluster représentant la position cristallographique, nous avons décidé d'améliorer l'échantillonnage en prenant un grand nombre d'orientation du substrat dans cette position. Ainsi, 8 conformations (3 obtenues par les calculs de docking et 5 orientées manuellement avec le logiciel Xleap) ont été retenues. Les différentes orientations ont été obtenues selon une approche "head-tail" and "upside-down". Pour les deux autres clusters (représentés en c et d dans les Figures 1 et 2), les trois positions associées aux meilleures énergies d'interaction obtenues par protocoles de docking ont été conservées. Au total 14 systèmes différents ont été soumis à des simulations de DM.

IV) Simulation de dynamique moléculaire sur les différents systèmes

IV.1) Protocole de simulation

Pour chacun des 14 systèmes, 6 ns de dynamique moléculaire ont été effectuées pour estimer la stabilité des complexes formés par l'application du protocole MMGBSA. Il apparaît rapidement que la cavité est accessible au solvant. Des molécules d'eau sont souvent présentes autour du substrat. Il est bien connu que l'interaction enzyme-substrat est souvent assurée par la présence de molécules d'eau [188-190]. Nous avons donc choisi de modifier le protocole MM-GBSA en prenant en compte de manière explicite les molécules d'eau pontantes entre le substrat et le récepteur. Celles qui étaient à une distance inférieure à 2,5 Angströms d'un atome donneur ou accepteur de liaison hydrogène du substrat, respectivement par rapport à l'atome d'oxygène ou les atomes d'hydrogène de la molécule d'eau, ont été sélectionnées une première fois. Afin d'être conservée dans le récepteur, cette molécule doit également se trouver à une distance inférieure à 2,5 angströms d'un atome donneur ou accepteur de l'enzyme. Si ces deux conditions sont satisfaites, la molécule d'eau est conservée de manière explicite dans le complexe et prise en compte pour le calcul de type MM-GBSA [191]. L'établissement d'un classement entre les différentes positions peut donc être envisagé par un simple calcul de l'enthalpie libre. Comme le même ligand sera conservé dans chacune des simulations, la contribution entropique sera négligée lors de nos calculs.

IV.2) Résultats des simulations de DM

Les résultats obtenus pour les 14 systèmes sont répertoriés dans le Tableau 8.

	Site B (Position dans placement cristallographique)								Site A (Site NADPH)			Site C (Site Sortie 2)		
Poses	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Energie MM-GBSA avec prise en compte des molécules d'eau (kcal/mol)	-29,5 ± 3,7	-31,7 ± 3,7	-21,5 ± 3,8	-9,95 ± 2,5	-29,8 ± 3,7	-10,6 ± 2,9	-19,7 ± 4,5	-26,6 ± 3,4	-32,9 ± 3,9	-24,9 ± 3,3	-21,4 ± 6,1	-21,1 ± 4,4	-27,0 ± 3,3	-23,0 ± 3,4
Energie MM-GBSA classique (kcal/mol)	-26,6 ± 3,1	-31,1 ± 3,4	-19,1 ± 3,8	-10,81 ± 2,2	-28,6 ± 3,3	-9,78 ± 2,8	-17,3 ± 4,6	-24,3 ± 3,3	-30,6 ± 3,5	-21,1 ± 3,1	-21,1 ± 5,9	-20,86 ± 4,3	-26,3 ± 3,1	-19,9 ± 3,2
Observations	Déplacement vers site NADPH	Déplacement vers site NADPH	RAS	Décalage vers le haut de la cavité	RAS	Décalage vers une boucle hydrophobe	RAS	RAS	RAS	Rotation cycle B	RAS	RAS	RAS	RAS
Résidus avec interaction type hydrophobe	129, 88, 191, 192, 193	17, 192, 193, 190, 191, 88, 204	192, 222, 134, 191	222, 193, 129	129, 222, 164	224, 281, 220, 221, 224	155, 129, 192, 226	192, 193, 17, 128, 129, 205	86, 79, 88, 167	15, 16, 17, 192, 204, 205, 206	16, 17, 193, 204	222, 192, 227, 226, 223, 224, 134	134, 224, 138, 129, 192, 223, 222	134, 222, 129, 133, 191, 192
Résidus avec interaction de type liaison H	133, 227, 128	130, 205, 193	130, 227	225, 227	227, 191, 128, 129, 222	227, 217	133, 227, 128, 191, 222	191, 193, 167, 205	127, 17, 15	12, 18	193, 167	222, 225, 227	227, 225, 133, 224	225, 130
RMSd DFR (Å)	1	1,3	1,2	1,4	1,1	1,3	1,3	0,8	1,2	1,3	1,1	1,7	1,4	1,3
RMSd ligand (Å)	0,3	0,5	0,4	0,6	0,5	0,5	0,4	0,5	0,4	0,5	0,5	0,6	0,3	0,4

Tableau 8 : Résumé des différents résultats obtenus pour les 14 systèmes enzyme-ligand construits par les calculs de docking. Sont présentés les résidus en interaction, ainsi que les énergies d'interaction calculées par le protocole MM-GBSA classique, et celles obtenues par le protocole modifié avec la prise en compte des molécules d'eau explicite au sein du récepteur. Enfin, les valeurs moyennes des RMSd du ligand et de l'enzyme sont présentées.

Toutes les simulations montrent un RMSd très faible calculé sur les atomes du squelette de l'enzyme, ce qui montre que la structure de l'enzyme reste stable ($< 2 \text{ \AA}$). Le placement différent de DHQ ne provoque pas de changements allostériques notables. Toutefois, dans la moitié des simulations concernant un substrat en « position cristallographique », celui-ci migre rapidement vers de nouvelles positions. Les simulations 1 et 2 présentent de manière naturelle un transfert du DHQ vers le site de liaison de la molécule de NADPH, indiquant qu'il n'y a pas de barrière à franchir pour atteindre cette nouvelle position ou que les mouvements de l'enzyme permettent cette diffusion. De nouvelles interactions sont créées, notamment avec les résidus M88 et P204, qui normalement interagissent avec le cofacteur. Ce changement de position montre que le DHQ est en compétition avec le cofacteur pour l'interaction avec le site A. Concernant les autres clusters, les positions dans les autres sites ne montrent que peu de modifications structurales par rapport à la pose initiale obtenue après docking et les interactions initiales se retrouvent ici durant la totalité des simulations.

Concernant l'énergie d'interaction, nous observons tout d'abord que la prise en compte des molécules d'eau pontantes modifie l'énergie calculée par le protocole MM-GBSA et pour la plupart du temps tend à améliorer les interactions enzyme/substrat. D'une manière générale, le positionnement associé à la sortie 2 correspond aux énergies d'interaction les moins favorables. Sur la base de l'énergie d'interaction, quatre positions se détachent clairement. Il s'agit des positions 1, 2 et 5 concernant la localisation initiale autour de la position cristallographique et la position 9 du positionnement en localisation du site NADPH.

Nous avons comparé ces énergies d'interactions obtenues avec celle du complexe ternaire DFR-NADPH-DHQ (pour rappel, 5 simulations de dynamique moléculaire d'une durée de 20ns ont été effectuées, cf chapitre 2, III.1). Les différentes énergies calculées par le protocole MM-GBSA (classique et modifié) ainsi que les détails associés sont résumés dans le Tableau 9.

Simulation	DFR1	DFR2	DFR3	DFR4	DFR5
Energie d'interaction avec protocole modifié (kcal/mol)	-14,75 ± 3,3	-14,28 ± 3,5	-19,84 ± 3,1	-17,22 ± 3,5	-11,87 ± 3,9
Energie d'interaction avec protocole classique (kcal/mol)	-14,06 ± 3,2	-12,92 ± 3,4	-17,13 ± 3,1	-15,55 ± 3,6	-10,6 ± 4,0

Tableau 9 : Energie d'interaction DHQ/récepteur calculée pour chaque simulation de dynamique moléculaire d'un système DFR-NADPH-DHQ.

Il s'avère que l'énergie d'interaction DHQ/récepteur est plus forte sans le cofacteur dans le site actif. En l'absence de NADPH dans la cavité, la position cristallographique n'est plus le minimum global de la surface d'énergie potentielle du système. D'après nos simulations, la barrière associée au transfert d'un site de liaison à l'autre est assez faible puisque nous observons de manière naturelle et très rapidement une diffusion du substrat vers le site de reconnaissance de NADPH. La molécule de dihydroquercétine interagit avec le site de liaison du cofacteur et gêne son entrée ou son positionnement dans la cavité, entraînant une inhibition de l'activité catalytique de l'enzyme. Ces résultats sont en direct accord avec les analyses structurales des trois cavités, basées sur les calculs de docking, qui favorisaient le site A pour l'interaction de DHQ dans l'enzyme vide.

Ces simulations de dynamique moléculaires de courtes durées nous ont permis d'identifier les systèmes les plus stables. Sur les 5 systèmes les plus stables (poses 1, 2, 5, 9 et 12), nous avons vérifié leur stabilité au cours du temps (RMSd enzyme < 3 Å et RMSd DQH < 1,5 Å), en augmentant le temps de simulation à 200 ns. Les positions sont toutes restées stables et nous assistons à des rotations de cycle du DHQ, mais en aucun cas à un déplacement ou à une sortie du substrat. Ceci confirme les interprétations données par le protocole MM-GBSA.

V) Conclusion

Le travail de Trabelsi *et al.* proposait différents modèles pour expliquer le phénomène d'inhibition de l'enzyme par excès de substrat. Leur résultats proposaient la formation d'un complexe binaire DFR-DHQ inopérant ou la formation du complexe ternaire abortif DFR-NADP⁺-DHQ comme hypothèses principales. Sur la base de leur étude, nous proposons un

modèle tridimensionnel pour expliquer au niveau moléculaire cette inhibition. Les résultats des simulations de dynamique moléculaire, ainsi que les estimations des énergies d'interactions ligand/DFR obtenues par deux versions du protocole MM-GBSA montrent que la formation d'un complexe binaire DFR-DHQ entraînerait la fixation du substrat soit sur le site de liaison du cofacteur, soit sur le site déjà identifié de la structure cristallographique du complexe ternaire. Les simulations de dynamique moléculaire montrent également un transfert naturel d'un site à l'autre, indiquant que la barrière associée à ce transfert est faible. L'inhibition par excès de DHQ serait donc compétitive avec le cofacteur pour l'interaction avec le site A lorsque nous sommes en excès de substrat, et empêcherait ainsi l'enzyme de jouer son rôle catalytique.

Bibliographie

1. Berkessel, A., Gröger, H., and MacMillan, D., *Asymmetric Organocatalysis: From Biomimetic Concepts to Applications in Asymmetric Synthesis* **2006**: Wiley.
2. Kennedy, J.F., Kumar, H., Panesar, P.S., Marwaha, S.S., Goyal, R., Parmar, A., and Kaur, S., Enzyme-catalyzed regioselective synthesis of sugar esters and related compounds. *J. Chem. Technol. Biot.*, **2006**. 81(6): p. 866-876.
3. Varma, I.K., Albertsson, A.-C., Rajkhowa, R., and Srivastava, R.K., Enzyme catalyzed synthesis of polyesters. *Prog. Polym. Sci.*, **2005**. 30(10): p. 949-981.
4. Sanchez, C., Arribart, H., and Giraud Guille, M.M., Biomimetism and bioinspiration as tools for the design of innovative materials and systems. *Nat Mater*, **2005**. 4(4): p. 277-288.
5. Baggiolini, M., Les stades repères dans le développement annuel de la vigne et leur utilisation pratique. *Rev.Romande Agric.Vitic.Arboric.*, **1952**. 8: p. 4-6.
6. Kennedy, J., Understanding grape berry development. *Practical Winery and Vineyard journal*, **2002**(July/August): p. 14-23.
7. Coombe, B.G., Distribution of Solutes within the Developing Grape Berry in Relation to Its Morphology. *Am. J. Enol. Viticult.*, **1987**. 38(2): p. 120-127.
8. Lacampagne, S., Localisation et caractérisation des tannins dans la pellicule du raisin : Etude de l'impact de l'organisation physico-chimique des parois cellulaires sur la composante tannique, la qualité du fruit et la typicité des raisins de Bordeaux. Thèse de doctorat, **2010**.
9. Winkel-Shirley, B., Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant. Biol.*, **2002**. 5(3): p. 218-223.
10. Forkmann, G. and Heller, W., 1.26 - *Biosynthesis of Flavonoids*, in *Comprehensive Natural Products Chemistry*, S.D. Barton, K. Nakanishi, and O. Meth-Cohn, Editors. **1999**, Pergamon: Oxford. p. 713-748.
11. Dixon, R.A. and Steele, C.L., Flavonoids and isoflavonoids – a gold mine for metabolic engineering. *Trends in Plant Science*, **1999**. 4(10): p. 394-400.
12. Manthey, J. and Buslig, B., *Flavonoids in the Living System*, in *Flavonoids in the Living System*, J. Manthey and B. Buslig, Editors. **1998**, Springer US. p. 1-7.
13. Lewandowska, U., Szewczyk, K., Hrabec, E., Janecka, A., and Gorlach, S., Overview of Metabolism and Bioavailability Enhancement of Polyphenols. *J. Agr. Food Chem.*, **2013**. 61(50): p. 12183-12199.
14. D'Archivio, M., Filesì, C., Vari, R., Sczzocchio, B., and Masella, R., Bioavailability of the Polyphenols: Status and Controversies. *International Journal of Molecular Sciences*, **2010**. 11(4): p. 1321-1342.
15. Chen, D., Daniel, K.G., Kuhn, D.J., Kazi, A., Bhuiyan, M., Li, L., Wang, Z., Wan, S.B., Lam, W.H., Chan, T.H., and Dou, Q.P., Green tea and tea polyphenols in cancer prevention. *Frontiers in bioscience : a journal and virtual library*, **2004**. 9: p. 2618-2631.
16. Frankel, E.N., German, J.B., Kinsella, J.E., Parks, E., and Kanner, J., Inhibition of oxidation of human low-density lipoprotein by phenolic substances in red wine. *The Lancet*, **1993**. 341(8843): p. 454-457.

17. de Freitas, V. and Mateus, N., Protein/Polyphenol Interactions: Past and Present Contributions. Mechanisms of Astringency Perception. *Curr. Org. Chem.*, **2012**. 16(6): p. 724-746.
18. Golebiowski, J., Fiorucci, S., Adrian-Scotto, M., Fernandez-Carmona, J., and Antonczak, S., Molecular Features Underlying the Perception of Astringency as Probed by Molecular Modeling. *Molecular Informatics*, **2011**. 30(5): p. 410-414.
19. Cushnie, T.P.T. and Lamb, A.J., Antimicrobial activity of flavonoids. *Int. J. Microbiol. Ag.*, **2005**. 26(5): p. 343-356.
20. Friedman, M., Overview of antibacterial, antitoxin, antiviral, and antifungal activities of tea flavonoids and teas. *Molecular Nutrition & Food Research*, **2007**. 51(1): p. 116-134.
21. Ververidis, F., Trantas, E., Douglas, C., Vollmer, G., Kretzschmar, G., and Panopoulos, N., Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health. *Biotech. J.*, **2007**. 2(10): p. 1214-1234.
22. Sarni-Manchado, P. and Cheynier, V., *Les polyphénols en agroalimentaire*. Tec & Doc ed. Collection sciences & techniques alimentaires **2006**: Lavoisier.
23. Simmonds, M.S.J., Importance of flavonoids in insect-plant interactions: feeding and oviposition. *Phytochemistry*, **2001**. 56(3): p. 245-252.
24. Hemingway, R.W., Foo, L.Y., and Porter, L.J., LINKAGE ISOMERISM IN TRIMERIC AND POLYMERIC 2,3-CIS-PROCYANIDINS. *Journal of the Chemical Society-Perkin Transactions 1*, **1982**(5): p. 1209-1216.
25. Porter, L., *Flavans and proanthocyanidins*, in *The Flavonoids*, J.B. Harborne, Editor **1988**, Springer US. p. 21-62.
26. Rodríguez Montealegre, R., Romero Peces, R., Chacón Vozmediano, J.L., Martínez Gascueña, J., and García Romero, E., Phenolic compounds in skins and seeds of ten grape *Vitis vinifera* varieties grown in a warm climate. *Journal of Food Composition and Analysis*, **2006**. 19(6-7): p. 687-693.
27. Mattivi, F., Guzzon, R., Vrhovsek, U., Stefanini, M., and Velasco, R., Metabolite Profiling of Grape: Flavonols and Anthocyanins. *J. Agr. Food Chem.*, **2006**. 54(20): p. 7692-7702.
28. Forkmann, G., Flavonoids as Flower Pigments: The Formation of the Natural Spectrum and its Extension by Genetic Engineering. *Plant Breeding*, **1991**. 106(1): p. 1-26.
29. Ganguly, A.K., Seshadri, T.R., and Subramanian, P., A study of leucoanthocyanidins of plants—I : Isomers of leucodelphinidin from karada bark and eucalyptus gum. *Tetrahedron*, **1958**. 3(3): p. 225-229.
30. Stafford, H.A. and Lester, H.H., Enzymic and Nonenzymic Reduction of (+)-Dihydroquercetin to Its 3,4,-Diol. *Plant Physiology*, **1982**. 70(3): p. 695-698.
31. Haslam, E., In vino veritas: Oligomeric procyanidins and the ageing of red wines. *Phytochemistry*, **1980**. 19(12): p. 2577-2582.
32. Haslam, E., Symmetry and promiscuity in procyanidin biochemistry. *Phytochemistry*, **1977**. 16(11): p. 1625-1640.

33. Andersen, O.M. and Markham, K.R., *Flavonoids: Chemistry, Biochemistry and Applications* **2005**: Taylor & Francis.
34. <http://www.genome.jp/kegg/kegg2.html>.
35. Winkel-Shirley, B., Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiology*, **2001**. 126(2): p. 485-493.
36. Ververidis, F., Trantas, E., Douglas, C., Vollmer, G., Kretzschmar, G., and Panopoulos, N., Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part II: Reconstruction of multienzyme pathways in plants and microbes. *Biotech. J.*, **2007**. 2(10): p. 1235-1249.
37. Ferrer, J.-L., Jez, J.M., Bowman, M.E., Dixon, R.A., and Noel, J.P., Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat Struct Mol Biol*, **1999**. 6(8): p. 775-784.
38. Dao, T.T.H., Linthorst, H.J.M., and Verpoorte, R., Chalcone synthase and its functions in plant resistance. *Phytochemistry Reviews*, **2011**. 10(3): p. 397-412.
39. Bednar, R.A. and Hadcock, J.R., Purification and characterization of chalcone isomerase from soybeans. *J. Biol. Chem.*, **1988**. 263(20): p. 9582-8.
40. Cain, C., Saslowsky, D., Walker, R., and Shirley*, B., Expression of chalcone synthase and chalcone isomerase proteins in Arabidopsis seedlings. *Plant Mol. Biol.*, **1997**. 35(3): p. 377-381.
41. Owens, D.K., Crosby, K.C., Runac, J., Howard, B.A., and Winkel, B.S.J., Biochemical and genetic characterization of Arabidopsis flavanone 3 β -hydroxylase. *Plant Physiol. Biochem.*, **2008**. 46(10): p. 833-843.
42. Wang, H., Fan, W., Li, H., Yang, J., Huang, J., and Zhang, P., Functional Characterization of Dihydroflavonol-4-Reductase in Anthocyanin Biosynthesis of Purple Sweet Potato Underlies the Direct Evidence of Anthocyanins Function against Abiotic Stresses. *PLoS ONE*, **2013**. 8(11): p. e78484.
43. Xie, D.-Y. and Dixon, R.A., Proanthocyanidin biosynthesis – still more questions than answers? *Phytochemistry*, **2005**. 66(18): p. 2127-2144.
44. Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., and Caboche, M., GENETICS AND BIOCHEMISTRY OF SEED FLAVONOIDS. *Annu. Rev. Plant Biol.*, **2006**. 57(1): p. 405-430.
45. Zhao, J., Pang, Y., and Dixon, R.A., The Mysteries of Proanthocyanidin Transport and Polymerization. *Plant Physiology*, **2010**. 153(2): p. 437-443.
46. Verdier, J., Zhao, J., Torres-Jerez, I., Ge, S., Liu, C., He, X., Mysore, K.S., Dixon, R.A., and Udvardi, M.K., MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula*. *Proc Natl Acad Sci U S A*, **2012**.
47. Stafford, H.A., Changes in phenolic compounds and related enzymes in young plants of sorghum. *Phytochemistry*, **1969**. 8(4): p. 743-752.
48. Winkel, B.S.J., METABOLIC CHANNELING IN PLANTS. *Annu. Rev. Plant Biol.*, **2004**. 55(1): p. 85-107.

49. Crosby, K.C., Pietraszewska-Bogiel, A., Gadella Jr, T.W.J., and Winkel, B.S.J., Förster resonance energy transfer demonstrates a flavonoid metabolon in living plant cells that displays competitive interactions between enzymes. *FEBS Letters*, **2011**. 585(14): p. 2193-2198.
50. Spivey, H.O. and Ovádi, J., Substrate Channeling. *Methods*, **1999**. 19(2): p. 306-321.
51. Sharma, H., Landau, M.J., Vargo, M.A., Spasov, K.A., and Anderson, K.S., First Three-Dimensional Structure of *Toxoplasma gondii* Thymidylate Synthase–Dihydrofolate Reductase: Insights for Catalysis, Interdomain Interactions, and Substrate Channeling. *Biochemistry*, **2013**. 52(41): p. 7305-7317.
52. Ellis, R.J., Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci*, **2001**. 26(10): p. 597-604.
53. Nielsen, K.A., Tattersall, D.B., Jones, P.R., and Møller, B.L., Metabolon formation in dhurrin biosynthesis. *Phytochemistry*, **2008**. 69(1): p. 88-98.
54. Maier, T., Leibundgut, M., and Ban, N., The crystal structure of a mammalian fatty acid synthase. *Science*, **2008**. 321(5894): p. 1315-22.
55. Lomakin, I.B., Xiong, Y., and Steitz, T.A., The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell*, **2007**. 129(2): p. 319-32.
56. Anselmi, C., Grininger, M., Gipson, P., and Faraldo-Gómez, J.D., Mechanism of Substrate Shuttling by the Acyl-Carrier Protein within the Fatty Acid Mega-Synthase. *J. Am. Chem. Soc.*, **2010**. 132(35): p. 12357-12364.
57. Jørgensen, K., Rasmussen, A.V., Morant, M., Nielsen, A.H., Bjarnholt, N., Zagrobelny, M., Bak, S., and Møller, B.L., Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant. Biol.*, **2005**. 8(3): p. 280-291.
58. Ralston, L. and Yu, O., Metabolons involving plant cytochrome P450s. *Phytochemistry Reviews*, **2006**. 5(2-3): p. 459-472.
59. Denisov, I.G., Makris, T.M., Sligar, S.G., and Schlichting, I., Structure and chemistry of cytochrome P450. *Chem Rev*, **2005**. 105(6): p. 2253-77.
60. Scott, E.E., He, Y.A., Wester, M.R., White, M.A., Chin, C.C., Halpert, J.R., Johnson, E.F., and Stout, C.D., An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution. *Proc Natl Acad Sci U S A*, **2003**. 100(23): p. 13196-201.
61. Møller, B.L., Dynamic Metabolons. *Science*, **2010**. 330(6009): p. 1328-1329.
62. Joernvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., and Ghosh, D., Short-chain dehydrogenases/reductases (SDR). *Biochemistry*, **1995**. 34(18): p. 6003-6013.
63. Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B., and Jörnvall, H., Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chemico-Biological Interactions*, **2003**. 143–144(0): p. 247-253.
64. Oppermann, U.T., Persson, B., Filling, C., and Jörnvall, H., *Structure-Function Relationships of SDR Hydroxysteroid Dehydrogenases*, in *Enzymology and Molecular Biology of Carbonyl Metabolism 6*, H. Weiner, et al., Editors. **1997**, Springer US. p. 403-415.
65. Rao, S.T. and Rossmann, M.G., Comparison of super-secondary structures in proteins. *J. Mol. Biol.*, **1973**. 76(2): p. 241-256.

66. King, J.D., Harmer, N.J., Preston, A., Palmer, C.M., Rejzek, M., Field, R.A., Blundell, T.L., and Maskell, D.J., Predicting Protein Function from Structure—The Roles of Short-chain Dehydrogenase/Reductase Enzymes in *Bordetella* O-antigen Biosynthesis. *J. Mol. Biol.*, **2007**. 374(3): p. 749-763.
67. Gani, O.A., Adekoya, O.A., Giurato, L., Spyraakis, F., Cozzini, P., Guccione, S., Winberg, J.O., and Sylte, I., Theoretical calculations of the catalytic triad in short-chain alcohol dehydrogenases/reductases. *Biophys J*, **2008**. 94(4): p. 1412-27.
68. Petit, P., Granier, T., d'Estaintot, B.L., Manigand, C., Bathany, K., Schmitter, J.-M., Lauvergeat, V., Hamdi, S., and Gallois, B., Crystal Structure of Grape Dihydroflavonol 4-Reductase, a Key Enzyme in Flavonoid Biosynthesis. *J. Mol. Biol.*, **2007**. 368(5): p. 1345-1357.
69. Johnson, E.T., Ryu, S., Yi, H., Shin, B., Cheong, H., and Choi, G., Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. *The Plant Journal*, **2001**. 25(3): p. 325-333.
70. Trabelsi, N., Petit, P., Manigand, C., Langlois d'Estaintot, B., Granier, T., Chaudiere, J., and Gallois, B., Structural evidence for the inhibition of grape dihydroflavonol 4-reductase by flavonols. *Acta Crystallogr D*, **2008**. 64(8): p. 883-891.
71. Trabelsi, N., d'Estaintot, B., Sigaud, G., Gallois, B. and Chaudière, J., Kinetic and binding equilibrium studies of dihydroflavonol 4-reductase from *Vitis vinifera* and its unusually strong substrate inhibition. *J. Biophys Chem.*, **2011**. 2: p. 332-344.
72. Gang, D.R., Kasahara, H., Xia, Z.-Q., Vander Mijnsbrugge, K., Bauw, G., Boerjan, W., Van Montagu, M., Davin, L.B., and Lewis, N.G., Evolution of Plant Defense Mechanisms: RELATIONSHIPS OF PHENYLCOUMARAN BENZYLIC ETHER REDUCTASES TO PINORESINOLARICRESINOL AND ISOFLAVONE REDUCTASES. *J Biol Chem*, **1999**. 274(11): p. 7516-7527.
73. Wang, X., He, X., Lin, J., Shao, H., Chang, Z., and Dixon, R.A., Crystal Structure of Isoflavone Reductase from Alfalfa (*Medicago sativa* L.). *J. Mol. Biol.*, **2006**. 358(5): p. 1341-1352.
74. Maugé, C., Granier, T., d'Estaintot, B.L., Gargouri, M., Manigand, C., Schmitter, J.-M., Chaudière, J., and Gallois, B., Crystal Structure and Catalytic Mechanism of Leucoanthocyanidin Reductase from *Vitis vinifera*. *J. Mol. Biol.*, **2010**. 397(4): p. 1079-1091.
75. Guengerich, F.P., Cytochrome P450 and Chemical Toxicology. *Chem. Res. Toxicol.*, **2007**. 21(1): p. 70-83.
76. Lamb, D.C., Lei, L., Warrilow, A.G., Lepesheva, G.I., Mullins, J.G., Waterman, M.R., and Kelly, S.L., The first virally encoded cytochrome p450. *J Virol*, **2009**. 83(16): p. 8266-9.
77. Cojocar, V., Balali-Mood, K., Sansom, M.S.P., and Wade, R.C., Structure and Dynamics of the Membrane-Bound Cytochrome P450 2C9. *PLoS Comput Biol*, **2011**. 7(8): p. e1002152.
78. Sansen, S., Yano, J.K., Reynald, R.L., Schoch, G.A., Griffin, K.J., Stout, C.D., and Johnson, E.F., Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *J Biol Chem*, **2007**. 282(19): p. 14348-55.
79. Schoch, G.A., Yano, J.K., Wester, M.R., Griffin, K.J., Stout, C.D., and Johnson, E.F., Structure of Human Microsomal Cytochrome P450 2C8: Evidence for a peripheral fatty acid binding site. *J. Biol. Chem.*, **2004**. 279(10): p. 9497-9503.

80. Jez, J.M., Bowman, M.E., Dixon, R.A., and Noel, J.P., Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat Struct Biol*, **2000**. 7(9): p. 786-91.
81. Hur, S. and Bruice, T.C., Enzymes do what is expected (chalcone isomerase versus chorismate mutase). *J Am Chem Soc*, **2003**. 125(6): p. 1472-3.
82. Hur, S. and Bruice, T.C., The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc Natl Acad Sci U S A*, **2003**. 100(21): p. 12015-12020.
83. Ruiz-Pernia, J.J., Silla, E., and Tunon, I., Enzymatic effects on reactant and transition states. The case of chalcone isomerase. *J Am Chem Soc*, **2007**. 129(29): p. 9117-24.
84. Ruiz-Pernía, J.J., Martí, S., Moliner, V., and Tuñón, I., A Novel Strategy to Study Electrostatic Effects in Chemical Reactions: Differences between the Role of Solvent and the Active Site of Chalcone Isomerase in a Michael Addition. *J. Chem. Theory. Comput.*, **2012**. 8(5): p. 1532-1535.
85. Turnbull, J.J., Nakajima, J., Welford, R.W., Yamazaki, M., Saito, K., and Schofield, C.J., Mechanistic studies on three 2-oxoglutarate-dependent oxygenases of flavonoid biosynthesis: anthocyanidin synthase, flavonol synthase, and flavanone 3 β -hydroxylase. *J Biol Chem*, **2004**. 279(2): p. 1206-16.
86. Saito, K., Kobayashi, M., Gong, Z., Tanaka, Y., and Yamazaki, M., Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red form of *Perilla frutescens*. *Plant J*, **1999**. 17(2): p. 181-9.
87. Nakajima, J., Sato, Y., Hoshino, T., Yamazaki, M., and Saito, K., Mechanistic study on the oxidation of anthocyanidin synthase by quantum mechanical calculation. *J Biol Chem*, **2006**. 281(30): p. 21387-98.
88. Chebil, L., Chipot, C., Archambault, F., Humeau, C., Engasser, J.M., Ghoul, M., and Dehez, F., Solubilities inferred from the combination of experiment and simulation. Case study of quercetin in a variety of solvents. *J Phys Chem B*, **2010**. 114(38): p. 12308-13.
89. Steiner, R.A., Kooter, I.M., and Dijkstra, B.W., Functional analysis of the copper-dependent quercetin 2,3-dioxygenase. 1. Ligand-induced coordination changes probed by X-ray crystallography: inhibition, ordering effect, and mechanistic insights. *Biochemistry*, **2002**. 41(25): p. 7955-62.
90. Steiner, R.A., Meyer-Klaucke, W., and Dijkstra, B.W., Functional analysis of the copper-dependent quercetin 2,3-dioxygenase. 2. X-ray absorption studies of native enzyme and anaerobic complexes with the substrates quercetin and myricetin. *Biochemistry*, **2002**. 41(25): p. 7963-8.
91. Fiorucci, S., Golebiowski, J., Cabrol-Bass, D., and Antonczak, S., Oxygenolysis of flavonoid compounds: DFT description of the mechanism for quercetin. *Chemphyschem*, **2004**. 5(11): p. 1726-33.
92. Siegbahn, P.E., Hybrid DFT study of the mechanism of quercetin 2,3-dioxygenase. *Inorg Chem*, **2004**. 43(19): p. 5944-53.

93. Fiorucci, S., Golebiowski, J., Cabrol-Bass, D., and Antonczak, S., DFT study of quercetin activated forms involved in antiradical, antioxidant, and prooxidant biological processes. *J. Agr. Food Chem.*, **2007**. 55(3): p. 903-911.
94. Chakraborty, S. and Biswas, P.K., Elucidation of the mechanistic pathways of the hydroxyl radical scavenging reaction by daidzein using hybrid QM/MM dynamics. *J Phys Chem A*, **2012**. 116(34): p. 8775-85.
95. Deepha, V., Praveena, R., Sivakumar, R., and Sadasivam, K., Experimental and theoretical investigations on the antioxidant activity of isoorientin from *Crotalaria globosa*. *Spectrochim. Acta A*, **2014**. 121: p. 737-745.
96. Dana, C.D., Bevan, D.R., and Winkel, B.S., Molecular modeling of the effects of mutant alleles on chalcone synthase protein structure. *J Mol Model*, **2006**. 12(6): p. 905-14.
97. Fujimoto, K., Hara, M., Yamada, H., Sakurai, M., Inaba, A., Tomomura, A., and Katoh, S., Role of the conserved Ser-Tyr-Lys triad of the SDR family in sepiapterin reductase. *Chem Biol Interact*, **2001**. 130-132(1-3): p. 825-32.
98. Owens, D.K., Alerding, A.B., Crosby, K.C., Bandara, A.B., Westwood, J.H., and Winkel, B.S.J., Functional Analysis of a Predicted Flavonol Synthase Gene Family in *Arabidopsis*. *Plant Physiology*, **2008**. 147(3): p. 1046-1061.
99. Busk, P.K. and Møller, B.L., Dhuririn Synthesis in *Sorghum* Is Regulated at the Transcriptional Level and Induced by Nitrogen Fertilization in Older Plants. *Plant Physiology*, **2002**. 129(3): p. 1222-1231.
100. Kristensen, C., Morant, M., Olsen, C.E., Ekstrom, C.T., Galbraith, D.W., Moller, B.L., and Bak, S., Metabolic engineering of dhuririn in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome. *Proc Natl Acad Sci U S A*, **2005**. 102(5): p. 1779-84.
101. Jensen, K., Osmani, S.A., Hamann, T., Naur, P., and Moller, B.L., Homology modeling of the three membrane proteins of the dhuririn metabolon: catalytic sites, membrane surface association and protein-protein interactions. *Phytochemistry*, **2011**. 72(17): p. 2113-23.
102. Lomize, M.A., Lomize, A.L., Pogozheva, I.D., and Mosberg, H.I., OPM: orientations of proteins in membranes database. *Bioinformatics*, **2006**. 22(5): p. 623-5.
103. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A., Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics*, **2006**. Chapter 5: p. Unit 5 6.
104. Zacharias, M., ATTRACT: protein-protein docking in CAPRI using a reduced protein model. *Proteins*, **2005**. 60(2): p. 252-6.
105. Fiorucci, S. and Zacharias, M., Prediction of protein-protein interaction sites using electrostatic desolvation profiles. *Biophys J*, **2010**. 98(9): p. 1921-30.
106. Schneider, S., Saladin, A., Fiorucci, S., Prévost, C., and Zacharias, M., ATTRACT and PTOOLS: Open Source Programs for Protein-Protein Docking #, in *T Computational Drug Discovery and Design* **2012**. p. 221-232.
107. Ludemann, S.K., Lounnas, V., and Wade, R.C., How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J Mol Biol*, **2000**. 303(5): p. 797-811.

108. Senn, H.M. and Thiel, W., QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Edit.*, **2009**, 48(7): p. 1198-1229.
109. Leach, A.R., *Molecular Modelling: Principles and Applications* **2001**: Prentice Hall.
110. Allen, P. and Tildesley, D.J., *Computer Simulation of Liquids* **1989**: Clarendon Press.
111. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **2003**, 24, 1999–2012.
112. Lee, M.C.; Duan, Y. Distinguish protein decoys by using a scoring function based on a new Amber force field, short molecular dynamics simulations, and the generalized Born solvent model. *Proteins*, **2004**, 55, 620–634.
113. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham III, T.E., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, **1995**, 91(1–3): p. 1-41.
114. Toukmaji, A.Y. and Board Jr, J.A., Ewald summation techniques in perspective: a survey. *Computer Physics Communications*, **1996**, 95(2–3): p. 73-92.
115. Case, D.A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A.W., Kolossváry, I., Wong, K.F., Paesani, F., Vanicek, J., Wolf, R.M., Liu, J., Wu, X., Brozell, S.R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Hsieh, M.J., Cui, G., Roe, D.R., Mathews, D.H., Seetin, M.G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P.A., *AMBER 12*, 2012, University of California, San Francisco.
116. Salomon-Ferrer, R., Götz, A.W., Poole, D., Le Grand, S., and Walker, R.C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory. Comput.*, **2013**, 9(9): p. 3878-3888.
117. Kirkwood, J.G., Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.*, **1935**, 3(5): p. 300-313.
118. Torrie GM, V.J., Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.*, **1974**, 28: p. 578-581.
119. Torrie GM, V.J., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling *J. Comput. Phys.*, **1977**, 23: p. 187-199.
120. Roux, B., The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, **1995**, 91(1-3): p. 275-82.
121. Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., and Kollman, P.A., Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, **1995**, 16(11): p. 1339-50.
122. Souaille, M. and Roux, B., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications*, **2001**, 135(1): p. 40-57.
123. Frenkel D, S.B., *Understanding Molecular Simulation: From Algorithms to Applications* ed. C. Academic Press San Diego **2002**

124. Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A., and Cheatham, T.E., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts Chem. Res.*, **2000**. 33(12): p. 889-897.
125. Hartree, D.R., The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, **1928**. 24(01): p. 89-110.
126. Fock, V., Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik*, **1930**. 61(1-2): p. 126-148.
127. Hohenberg, P. and Kohn, W., Inhomogeneous Electron Gas. *Physical Review*, **1964**. 136(3B): p. B864-B871.
128. Kohn, W. and Sham, L.J., Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, **1965**. 140(4A): p. A1133-A1138.
129. Zhao, Y. and Truhlar, D., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, **2008**. 120(1-3): p. 215-241.
130. McGuffin, L.J., Bryson, K., and Jones, D.T., The PSIPRED protein structure prediction server. *Bioinformatics*, **2000**. 16(4): p. 404-5.
131. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J., Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **2009**. 25(9): p. 1189-1191.
132. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M., PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **1993**. 26(2): p. 283-291.
133. Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S., and Onufriev, A., H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res*, **2005**. 33(Web Server issue): p. W368-71.
134. Ryde, U., Molecular dynamics simulations of alcohol dehydrogenase with a four- or five-coordinate catalytic zinc ion. *Proteins*, **1995**. 21(1): p. 40-56.
135. Holmberg, N., Ryde, U., and Bulow, L., Redesign of the coenzyme specificity in L-lactate dehydrogenase from bacillus stearothermophilus using site-directed mutagenesis and media engineering. *Protein Eng*, **1999**. 12(10): p. 851-6.
136. Autenrieth, F., Tajkhorshid, E., Baudry, J., and Luthey-Schulten, Z., Classical force field parameters for the heme prosthetic group of cytochrome c. *J Comput Chem*, **2004**. 25(13): p. 1613-22.
137. Oda, A., Yamaotsu, N., and Hirono, S., New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J. Comput. Chem.*, **2005**. 26(8): p. 818-826.
138. Shahrokh, K., Orendt, A., Yost, G.S., and Cheatham, T.E., 3rd, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J Comput Chem*, **2012**. 33(2): p. 119-33.

139. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A., Development and testing of a general amber force field. *J Comput Chem*, **2004**. 25(9): p. 1157-74.
140. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J.A., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Laham, A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C., and Pople, J.A., *Gaussian 03, Revision C.02*, **2003**.
141. Trott, O. and Olson, A.J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, **2010**. 31(2): p. 455-61.
142. Guengerich, F.P. and Isin, E.M., Mechanisms of cytochrome P450 reactions. *Acta Chimica Slovenica*, **2008**. 55(1): p. 7.
143. Meunier, B., de Visser, S.P., and Shaik, S., Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. *Chem Rev*, **2004**. 104(9): p. 3947-80.
144. Schrödinger Release 2014-2: Maestro, version 9.8. *Schrödinger, LLC, New York, NY*, **2014**.
145. Martí, S., Andrés, J., Moliner, V., Silla, E., Tuñón, I., and Bertrán, J., Conformational equilibrium of chorismate. A QM/MM theoretical study combining statistical simulations and geometry optimisations in gas phase and in aqueous solution. *Journal of Molecular Structure: THEOCHEM*, **2003**. 632(1-3): p. 197-206.
146. Marti, S., Roca, M., Andres, J., Moliner, V., Silla, E., Tunon, I., and Bertran, J., Theoretical insights in enzyme catalysis. *Chem Soc Rev*, **2004**. 33(2): p. 98-107.
147. Rajamani, R., Naidoo, K.J., and Gao, J., Implementation of an adaptive umbrella sampling method for the calculation of multidimensional potential of mean force of chemical reactions in solution. *J. Comput. Chem.*, **2003**. 24(14): p. 1775-1781.
148. Hub, J.S., de Groot, B.L., and van der Spoel, D., g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory. Comput.*, **2010**. 6(12): p. 3713-3720.
149. Lin, H. and Truhlar, D., QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.*, **2007**. 117(2): p. 185-199.
150. Mulholland, A.J., Computational enzymology: modelling the mechanisms of biological catalysts. *Biochem Soc Trans*, **2008**. 36(Pt 1): p. 22-6.
151. Seabra, G.d.M., Walker, R.C., Elstner, M., Case, D.A., and Roitberg, A.E., Implementation of the SCC-DFTB Method for Hybrid QM/MM Simulations within the Amber Molecular Dynamics Package†. *J. Phys. Chem. A*, **2007**. 111(26): p. 5655-5664.

152. Han, W.-G., Elstner, M., Jalkanen, K.J., Frauenheim, T., and Suhai, S., Hybrid SCC-DFTB/Molecular Mechanical Studies of H-Bonded Systems and of N-Acetyl-(L-Ala)_n N'-Methylamide Helices in Water Solution. *Int. J. Quantum Chem.*, **2000**. 78(6): p. 459-479.
153. Garcia-Viloca, M., Alhambra, C., Truhlar, D.G., and Gao, J., Hydride transfer catalyzed by xylose isomerase: mechanism and quantum effects. *J Comput Chem*, **2003**. 24(2): p. 177-90.
154. Bruice, T.C., Some Pertinent Aspects of Mechanism as Determined with Small Molecules. *Annu. Rev. Biochem.*, **1976**. 45(1): p. 331-374.
155. Schwartz, S.D. and Schramm, V., Enzymatic transition states and dynamic motion in barrier crossing. *Nature Chem. Biol.*, **2009**. 5: p. 551.
156. Claeysens, F., Ranaghan, K.E., Manby, F.R., Harvey, J.N., and Mulholland, A.J., Multiple high-level QM/MM reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization. *Chem Commun (Camb)*, **2005**(40): p. 5068-70.
157. Griffin, J.L., Bowler, M.W., Baxter, N.J., Leigh, K.N., Dannatt, H.R.W., Hounslow, A.M., Blackburn, G.M., Webster, C.E., Cliff, M.J., and Waltho, J.P., Near attack conformers dominate β -phosphoglucomutase complexes where geometry and charge distribution reflect those of substrate. *Proc Natl Acad Sci U S A*, **2012**. 109(18): p. 6910-6915.
158. Ranaghan, K.E., Ridder, L., Szefczyk, B., Sokalski, W.A., Hermann, J.C., and Mulholland, A.J., Transition state stabilization and substrate strain in enzyme catalysis: ab initio QM/MM modelling of the chorismate mutase reaction. *Org Biomol Chem*, **2004**. 2(7): p. 968-80.
159. Hur, S. and Bruice, T.C., Comparison of formation of reactive conformers (NACs) for the Claisen rearrangement of chorismate to prephenate in water and in the E. coli mutase: the efficiency of the enzyme catalysis. *J Am Chem Soc*, **2003**. 125(19): p. 5964-72.
160. Burbulis, I.E. and Winkel-Shirley, B., Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. *Proc Natl Acad Sci U S A*, **1999**. 96(22): p. 12929-12934.
161. Winkel-Shirley, B., It Takes a Garden. How Work on Diverse Plant Species Has Contributed to an Understanding of Flavonoid Metabolism. *Plant Physiology*, **2001**. 127(4): p. 1399-1404.
162. Smith, N.E., Tie, W.J., Flematti, G.R., Stubbs, K.A., Corry, B., Attwood, P.V., and Vrielink, A., Mechanism of the dehydrogenase reaction of DmpFG and analysis of inter-subunit channeling efficiency and thermodynamic parameters in the overall reaction. *Int J Biochem Cell Biol*, **2013**. 45(8): p. 1878-85.
163. Lo Conte, L., Chothia, C., and Janin, J., The atomic structure of protein-protein recognition sites. *J Mol Biol*, **1999**. 285(5): p. 2177-98.
164. Bahadur, R.P. and Zacharias, M., The interface of protein-protein complexes: Analysis of contacts and prediction of interactions. *Cell. Mol. Life Sci.*, **2008**. 65(7-8): p. 1059-1072.
165. Laskowski, R.A. and Swindells, M.B., LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model*, **2011**. 51(10): p. 2778-86.
166. Nooren, I.M. and Thornton, J.M., Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol*, **2003**. 325(5): p. 991-1018.
167. Bahadur, R.P., Chakrabarti, P., Rodier, F., and Janin, J., A dissection of specific and non-specific protein-protein interfaces. *J Mol Biol*, **2004**. 336(4): p. 943-55.

168. Hamelberg, D., Mongan, J., and McCammon, J.A., Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, **2004**. 120(24): p. 11919-11929.
169. Soltys, B.J., Falah, M., and Gupta, R.S., Identification of endoplasmic reticulum in the primitive eukaryote *Giardia lamblia* using cryoelectron microscopy and antibody to Bip. *J Cell Sci*, **1996**. 109 (Pt 7): p. 1909-17.
170. Rasmussen, S.G., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., Kobilka, T.S., Thian, F.S., Chae, P.S., Pardon, E., Calinski, D., Mathiesen, J.M., Shah, S.T., Lyons, J.A., Caffrey, M., Gellman, S.H., Steyaert, J., Skinotitis, G., Weis, W.I., Sunahara, R.K., and Kobilka, B.K., Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature*, **2011**. 477(7366): p. 549-55.
171. Mahadevi, A.S. and Sastry, G.N., Cation- π Interaction: Its Role and Relevance in Chemistry, Biology, and Material Science. *Chem Rev*, **2012**. 113(3): p. 2100-2138.
172. Grauffel, C., Yang, B., He, T., Roberts, M.F., Gershenson, A., and Reuter, N., Cation- π Interactions As Lipid-Specific Anchors for Phosphatidylinositol-Specific Phospholipase C. *J. Am. Chem. Soc.*, **2013**. 135(15): p. 5740-5750.
173. Broemstrup, T. and Reuter, N., How does Proteinase 3 interact with lipid bilayers? *Phys.Chem.Chem.Phys.*, **2010**. 12(27): p. 7487-7496.
174. Poulos, T. and Johnson, E., *Structures of Cytochrome P450 Enzymes*, in *Cytochrome P450*, P. Ortiz de Montellano, Editor 2005, Springer US. p. 87-114.
175. Cojocaru, V., Winn, P.J., and Wade, R.C., The ins and outs of cytochrome P450s. *Biochim Biophys Acta*, **2007**. 1770(3): p. 390-401.
176. Cojocaru, V., Winn, P.J., and Wade, R.C., Multiple, ligand-dependent routes from the active site of cytochrome P450 2C9. *Curr Drug Metab*, **2012**. 13(2): p. 143-54.
177. Berka, K., Hendrychova, T., Anzenbacher, P., and Otyepka, M., Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site. *J Phys Chem A*, **2011**. 115(41): p. 11248-55.
178. Li, W., Shen, J., Liu, G., Tang, Y., and Hoshino, T., Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations. *Proteins*, **2011**. 79(1): p. 271-81.
179. Fishelovitch, D., Shaik, S., Wolfson, H.J., and Nussinov, R., Theoretical Characterization of Substrate Access/Exit Channels in the Human Cytochrome P450 3A4 Enzyme: Involvement of Phenylalanine Residues in the Gating Mechanism. *J. Phys. Chem. B*, **2009**. 113(39): p. 13018-13025.
180. Lüdemann, S.K., Lounnas, V., and Wade, R.C., How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.*, **2000**. 303(5): p. 797-811.
181. Gromiha, M.M., Influence of cation- π interactions in different folding types of membrane proteins. *Biophys Chem*, **2003**. 103(3): p. 251-8.
182. Sanderson, J.M., Refined models for the preferential interactions of tryptophan with phosphocholines. *Org Biomol Chem*, **2007**. 5(20): p. 3276-86.

183. Petersen, F.N., Jensen, M.O., and Nielsen, C.H., Interfacial tryptophan residues: a role for the cation- π effect? *Biophys J*, **2005**. 89(6): p. 3985-96.
184. Kuhl, P.W., Excess-substrate inhibition in enzymology and high-dose inhibition in pharmacology: a reinterpretation [corrected]. *Biochem J*, **1994**. 298 (Pt 1): p. 171-80.
185. Reed, M.C., Lieb, A., and Nijhout, H.F., The biological significance of substrate inhibition: a mechanism with diverse functions. *Bioessays*, **2010**. 32(5): p. 422-9.
186. He, S. and Lai, L., Molecular docking and competitive binding study discovered different binding modes of microsomal prostaglandin E synthase-1 inhibitors. *J Chem Inf Model*, **2011**. 51(12): p. 3254-61.
187. Laville, R., Castel, C., Fattarsi, K., Roy, C., Legendre, L., Delbecq, C., Garry, P.-P., Audran, A., and Fernandez, X., Low sclareol by-product of clary sage concrete: chemical analysis of a waste product of the perfume industry. *Flavour Frag. J.*, **2013**. 28(2): p. 93-101.
188. Baron, R., Setny, P., and Andrew McCammon, J., Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.*, **2010**. 132(34): p. 12091-12097.
189. Huggins, D.J. and Tidor, B., Systematic placement of structural water molecules for improved scoring of protein-ligand interactions. *Protein Eng Des Sel*, **2011**. 24(10): p. 777-89.
190. Hayes, J.M., Skamnaki, V.T., Archontis, G., Lamprakis, C., Sarrou, J., Bischler, N., Skaltsounis, A.L., Zographos, S.E., and Oikonomakos, N.G., Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: the role of water molecules examined. *Proteins*, **2011**. 79(3): p. 703-19.
191. Mikulskis, P., Genheden, S., and Ryde, U., Effect of explicit water molecules on ligand-binding affinities calculated with the MM/GBSA approach. *J Mol Model*, **2014**. 20(6): p. 2273.

Résumé

Les flavonoïdes, molécules naturelles possédant des propriétés antiradicalaires et antioxydantes, sont produits au cours de cascades enzymatiques impliquant plusieurs enzymes. Il a récemment été proposé que certaines de ces enzymes s'assembleraient pour former un complexe multienzymatique transitoire, appelé métabolon, ancré à la membrane cellulaire. Cette structure rendrait possible des phénomènes de transfert direct de métabolites d'un site actif à l'autre (substrate channeling), réduisant ainsi les temps de diffusion et minimisant les effets de solvation/désolvation du substrat. L'objectif de ce travail est de proposer un premier modèle de ce type de complexe, impliquant trois enzymes de la biosynthèse des flavonoïdes : la dihydroflavonol-4-réductase (DFR), la flavonoïd-3'-hydroxylase (F3'H) et la leucoanthocyanidine réductase (LAR). L'étude de ces complexes moléculaires requiert la mise en œuvre d'une méthodologie multi-échelles basée sur l'utilisation i) de méthodes hybrides QM/MM pour l'étude de la réactivité enzymatique, ii) de simulations de dynamique moléculaire à résolution atomique se déroulant sur des échelles de temps de l'ordre de la microseconde pour estimer des propriétés thermodynamiques et cinétiques, iii) de calculs de docking protéine-protéine en résolution gros grain. L'application des différents niveaux de théorie nous a permis d'établir un premier modèle de métabolon à trois enzymes en interaction avec une membrane cellulaire.

Mots-clés : modélisation moléculaire, flavonoïde, enzyme, métabolon, approche multi-échelle, QM/MM.

Abstract

Flavonoids, some natural compounds exhibiting antioxidant properties, are synthesized along enzymatic cascades involving several enzymes. It has been recently proposed that some of these enzymes are involved in the formation of large transient macromolecular edifices, called metabolon, interacting with cellular membrane. Such molecular complexes should allow direct metabolites transfer from one active site to the other (substrate channeling phenomenon), minimizing diffusion time and solvation effects. The present work aims to establish a first model of a metabolon involving 3 enzymes of the flavonoid biosynthesis: the dihydroflavonol 4-reductase (DFR), the flavonoid 3'-hydroxylase (F3'H) and the leucoanthocyanidin reductase (LAR). The study of such large molecular system requires a multiscale approach based on i) hybrid QM/MM methods to decipher enzymatic mechanism, ii) molecular dynamic simulations in microsecond timescale to estimate thermodynamic and kinetic properties and iii) protein-protein docking at coarse-grained resolution. These different levels of theory allow us to establish a first model of a three-enzymes-metabolon in interaction with a model of cellular membrane.

Keywords: molecular modeling, flavonoid, enzyme, metabolon, multiscale approach, QM/MM