



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Institut National Polytechnique de Lorraine
École Nationale Supérieure d'Agronomie et des Industries Alimentaires
Laboratoire d'Ingénierie des Biomolécules

THÈSE

présentée à l'INPL par

Eduardo Basilio DE OLIVEIRA

Ingénieur des Industries Alimentaires - Grade de Master

pour obtenir le grade de

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

spécialité

Procédés Biotechnologiques et Alimentaires

Sujet

**Simulations moléculaires appliquées à l'acétylation de flavonoïdes
catalysées par des lipases :
*influence des structures de la lipase et du flavonoïde sur la
régiosélectivité de la bioconversion***

soutenue publiquement le 7 décembre 2009

MEMBRES DU JURY

Président : Mr. Bernard MAIGRET, D.R. (LORIA-CNRS, Nancy, France)

Rapporteurs : Mr. Serge ANTONCZAK, Prof. (Université de Nice Sophia-Antipolis, France)
Mr. Alain MARTY, Prof. (INSA, Toulouse, France)

Examineurs : Mme. Aude FALCIMAIGNE, M.d.C. (Un. Technologique de Compiègne, France)
Mme. Catherine HUMEAU, M.d.C. (ENSAIA, Nancy, France)
Mr. Jean-Marc ENGASSER, Prof. (ENSAIA, Nancy, France)

Invités : Mr. Mohamed GHOU, Prof. (ENSAIA, Nancy, France)
Mme. Elaine Rose MAIA, Prof. (Un. de Brasilia, Brésil)

*« L'homme n'est pas en possession de la Sagesse :
il ne fait qu'y tendre et peut seulement avoir de l'amour pour elle,
ce qui est déjà assez méritoire ».*

Emmanuel KANT

Remerciements

Que ce soit d'un point de vue scientifique ou humain, la réalisation de cette thèse fut pour moi une expérience d'une valeur immense. Scientifiquement, ces trois années ont représenté une opportunité précieuse d'apprendre un peu davantage sur les procédés enzymatiques et de découvrir, avec une grande fascination, la modélisation moléculaire. Humainement, cette période a été marquée par d'énormes leçons d'amitié, de résilience et d'humilité scientifique. Ainsi, je ne peux que remercier tous ceux qui étaient à mes côtés au cours de cette expérience.



Avant tout, merci à mes directeurs de thèse, M. Jean-Marc Engasser, M. Mohamed Ghoul et Mme. Catherine Humeau, pour m'avoir accueilli dans leur équipe, en me faisant confiance pour relever le défi de mener cette première thèse entièrement axée sur la modélisation moléculaire au sein de leur laboratoire.

Merci, M. Engasser, pour vos nombreux conseils et pour votre aide intensive pendant la rédaction et la structuration de ce manuscrit. Votre esprit critique et votre sens de l'organisation sont pour le moins remarquables.

Merci, M. Ghoul, d'avoir toujours tout fait pour assurer que ce travail se déroule dans les meilleures conditions. Votre implication constante et votre soutien ont été essentiels pour qu'on en arrive là aujourd'hui, et ne sont pas du tout passés inaperçus.

Merci, Catherine... pour tout ! Pour m'avoir suivi et soutenu patiemment au cours de ces trois années. Pour les discussions scientifiques fréquentes, au cours desquelles tu me poussais à la réflexion sans perdre ta douceur et ton humour !. Ce talent si rare que tu possèdes –faire coexister en harmonie professionnalisme, esprit scientifique et charisme – n'est qu'une des qualités qui te rendent si appréciée de tous ceux qui t'entourent. Merci aussi pour m'avoir permis de connaître ta famille si adorable et chaleureuse.

Bien entendu, un grand merci aux nombreux collaborateurs externes, experts en modélisation moléculaire, pour leur soutien incontournable.

Mme. Elaine Maia, professeur à l'Université de Brasilia qui, au cours de ses séjours à Nancy ou via e-mail, nous a appris les bases de la modélisation moléculaire et nous a guidé tout au long de la réalisation des simulations. Merci, Elaine, d'avoir été toujours disponible et d'avoir partagé ton énorme connaissance avec nous. La distance n'a pas été une barrière à ta bonne volonté. Ta contribution à ce travail a été inestimable !

M. Bernard Maigret, directeur de recherche au LORIA-CNRS (Université Henri Poincaré, Nancy) : merci pour vos conseils concernant les simulations de docking et de dynamique moléculaire, et aussi pour les outils de calcul que vous avez récemment mis à disposition de notre équipe. Merci également pour avoir accepté de présider le jury de soutenance de cette thèse.

M. Manuel Ruiz-Lopez, directeur de recherche, et M. Gérard Monard, professeur au Laboratoire de Chimie et Biochimie Théoriques (Université Henri Poincaré, Nancy) : merci pour votre gentillesse et votre immense patience pour nous conseiller sur les simulations de chimie quantique.

Je voudrais également remercier les membres du jury : M. Serge Antonczak, professeur à l'Université de Nice Sophia-Antipolis, M. Alain Marty, professeur à l'INSA de Toulouse (rapporteurs) et Mme. Aude Falcimagne, maître de conférence à l'Université Technologique de Compiègne (examinatrice). Mes sincères remerciements à vous, pour nous avoir fait l'honneur d'accepter de juger ce travail, pour le temps consacré à la lecture et à l'évaluation du manuscrit et, finalement, pour la discussion très riche d'enseignements pour nous lors de la soutenance.

Enfin, je ne saurais me passer de laisser un petit mot à ceux qui, d'une façon ou d'une autre, ont contribué à rendre mes journées plus agréables.

Mes remerciements les plus chaleureux et sincères à tout l'ensemble de l'équipe du LIBio : ce fut un honneur et un plaisir de vivre cette période de ma vie avec vous. Plus particulièrement, merci aux collègues doctorants Nidal, Ghada, Julie, Nizar, Guillaume, Leïla, Éric, Jordane et la nouvelle arrivée Christelle, pour votre bienveillance et tous les moments partagés (les drôles et les moins drôles...). À Latifa, Celine, Jennifer, Loanne, Fred, Cédric, Olivier, Adrien, Anne, Sylvia, Delphine, Sandrine et Sabrina pour leur gentillesse de tous les jours, leur attention et leur professionnalisme. Mention spéciale à Evelyne, qui était toujours là avec un calme et une fiabilité inébranlables pour assurer le bon fonctionnement du système informatique. Sans ton aide, tout aurait été beaucoup plus difficile ! À l'équipe de Chimie et Biochimie Alimentaire – M. Michel Girardin, Bernard, Bernadette, Lionel, Marie Noëlle, Jocelyne, Annelore – qui m'avaient déjà accueilli pour le master et n'ont pas hésité à m'ouvrir à nouveau les portes pour faire des vacations en TP, en me permettant ainsi de m'éclater un peu avec le métier d'enseignant qui me passionne depuis toujours. Et pour finir, je présente d'avance mes excuses à ceux dont j'aurais oublié de citer le nom : ma mémoire peut me jouer de tours, mais ceci ne change rien à la considération que j'ai et que j'aurai toujours pour chacun de vous.

Cette thèse est dédiée à mon père et à ma mère, qui ont fait tant de sacrifices pour offrir à moi et à mes sœurs l'opportunité de faire des études

TABLE DE MATIÈRES

LISTE DES FIGURES	11
LISTE DES TABLEAUX	15
LISTE DES SIGLES ET ABBREVIATIONS	17
INTRODUCTION	19

CHAPITRE I ETUDE BIBLIOGRAPHIQUE

1. LES FLAVONOÏDES	27
1.1. Généralités : origine et structure	27
1.2. Activités biologiques	29
1.2.1. Activité photoprotectrice	29
1.2.2. Activité anti-oxydante	29
1.2.3. Relations entre structure et activités biologiques	31
1.3. Acylation des flavonoïdes	32
1.3.1. Acylation par voie chimique	32
1.3.2. Acylation par voie enzymatique	33
1.4. La quercétine, de l'isoquercitrine et de la rutine	34
1.4.1. Distribution	34
1.4.2. Propriétés	35
1.4.3. Acylation enzymatique	36
2. LES PRINCIPALES APPROCHES DE MODELISATION MOLECULAIRE	39
2.1. Introduction	39
2.2. La mécanique moléculaire	40
2.2.1. Principes théoriques	40
2.2.2. Formalisme : les champs de force	41
2.2.3. Le champ de forces CHARMM	43
2.2.4. Minimisation d'énergie (optimisation de géométrie)	44
2.3. Les simulations de docking	45
2.3.1. Principes théoriques	46
2.3.2. Algorithmes de docking	46
2.3.3. Fonctions de score	47
2.4. Les simulations de dynamique moléculaire	49
2.4.1. Principes théoriques	49
2.4.2. Paramètres de mise en œuvre	49

2.5. Les simulations de chimie quantique	51
2.5.1. Principes théoriques	51
2.5.2. Méthodes classiques de chimie quantique	53
2.5.3. Méthode de la Théorie Fonctionnelle de la Densité (DFT)	53
2.5.3.1. Formalisme : l'équation de Kohn-Sham	54
2.5.3.2. Les fonctionnelles d'échange-corrélation	55
2.5.3.3. Les fonctions de base (orbitales atomiques)	55
2.6. Les méthodes hybrides de simulation (QM/MM)	57
2.7. Applications dans l'étude de systèmes du type protéine-ligand	57
2.7.1. Simulations de mécanique et dynamique moléculaires	57
2.7.2. Simulations de mécanique quantique	58
2.7.3. Simulations hybrides (QM/MM)	59
3. APPLICATIONS DE LA MODELISATION MOLECULAIRE POUR LA COMPREHENSION DES INTERACTIONS PROTEINES-FLAVONOIDES	59
3.1. Interactions entre les flavonoïdes et les acides aminés	60
3.2. Bioconversion enzymatique de flavonoïdes	60
3.2.1. La O-méthylation de flavonoïdes	61
3.2.2. L'oxygénolyse de flavonoïdes	62
3.2.3. L'inhibition de la lipoxigénase-3 du soja par la quercétine	65
3.3. Inhibition de cibles biologiques par des flavonoïdes	65
4. APPLICATIONS DE LA MODELISATION MOLECULAIRE POUR LA COMPREHENSION DES INTERACTIONS LIPASES-SUBSTRATS	69
4.1. Introduction	69
4.2. INVITED REVIEW : Contribution of molecular modelling to the understanding of lipase selectivities	70
4.3. Conclusions de l'article de revue	82
5. RÉFÉRENCES	84

CHAPITRE II

MATERIELS ET METHODES DE SIMULATION

1. RESSOURCES INFORMATIQUES	99
1.1. Ordinateurs	99
1.2. Logiciels de simulation moléculaire	99
2. STRUCTURES DE DÉPART	99
2.1. Structures des flavonoïdes	99
2.2. Structures des lipases	100

2.2.1. Structures cristallographiques	100
2.2.1.1. La lipase B de <i>Candida antarctica</i> (CALB)	100
2.2.1.2. La lipase de <i>Pseudomonas cepacia</i> (PCL)	100
2.2.2. Relaxation des structures cristallographiques	101
2.2.3. Construction des structures des acétyl-lipases	103
3. SIMULATIONS DE DOCKING	103
3.1. La définition du site de liaison	103
3.1.1. Principe	103
3.1.2. Mode opératoire	104
3.2. L'obtention des complexes	106
3.2.1. Principe	106
3.2.2. Mode opératoire	107
3.3. Le classement des complexes obtenus	109
3.4. L'optimisation des complexes retenus après le classement	109
3.4.1. Principe	109
3.4.2. Mode opératoire	110
4. SIMULATIONS DE DYNAMIQUE MOLÉCULAIRE	111
4.1. Principe	111
4.2. Mode opératoire	111
4.3. Analyse des résultats	111
5. SIMULATIONS DE CHIMIE QUANTIQUE	112
5.1. Principe	112
5.2. Mode opératoire	112
5.3. Analyse des résultats	113
6. RÉFÉRENCES	114

CHAPITRE III

ÉTUDE DE LA RÉGIOSÉLECTIVITÉ DANS L'ACÉTYLATION DES FLAVONOÏDES GLYCOSYLÉS ISOQUERCITRINE ET RUTINE CATALYSÉE PAR LA LIPASE B DE *CANDIDA ANTARCTICA*

1. INTRODUCTION	119
2. ARTICLE : A molecular modelling study to rationalize the regioselectivity in acylation of flavonoid glycosides catalyzed by <i>Candida antarctica</i> lipase B	120
3. CONTRIBUTION DE L'ARTICLE	138
4. ÉVALUATION DES COMPLEXES PRODUCTIFS PAR DYNAMIQUE MOLÉCULAIRE (DM)	138

4.1. DM du complexe productif pour l'isoquercitrine (I1)	141
4.2. DM du complexe productif pour la rutine (R1)	142
4.3. Sommaire et conclusions de l'étude complémentaire par DM	142

CHAPITRE IV

ETUDE DES SPÉCIFICITÉS DES LIPASES DE *CANDIDA ANTARCTICA* ET *PSEUDOMONAS CEPACIA* DANS L'ACÉTYLATION DU FLAVONOÏDE AGLYCONE QUERCÉTINE

1. INTRODUCTION	147
2. ARTICLE : Combined docking and molecular dynamics simulations to enlighten the specificities of <i>Pseudomonas cepacia</i> and <i>Candida antarctica</i> lipases in quercetin acetylation	148
3. CONTRIBUTION DE L'ARTICLE	167
4. ANALYSE COMPLÉMENTAIRE DE LA MOLÉCULE DE QUERCÉTINE	168

CHAPITRE V

ETUDE DE LA REACTIVITE DE LA QUERCÉTINE, L'ISOQUERCITRINE ET LA RUTINE AVEC LA LIPASE B DE *CANDIDA ANTARCTICA* COMME BIOCATALYSEUR POUR LEUR ACÉTYLATION

1. INTRODUCTION	171
2. ARTICLE : Density Functional Theory (DFT) modelling of the <i>Candida antarctica</i> lipase B selectivity in quercetin, isoquercitrin and rutin acetylation	172
3. CONTRIBUTION DE L'ARTICLE	184
CONCLUSION ET PERSPECTIVES	187

ANNEXES

A1. LES FONCTIONS DE SCORE UTILISEES POUR LE CONSENSUS SCORE	197
A2. L'ALGORITHME INTEGRATEUR DE VERLET-LEAPFROG	198
A3. LA FONCTIONNELLE D'ÉCHANGE-CORRELATION PW91 ET LES FONCTIONS DE BASE NUMERIQUES DND	199

LISTE DES FIGURES

CHAPITRE I ETUDE BIBLIOGRAPHIQUE

Figure I1 : Motif de base des flavonoïdes, avec la numérotation systématique des atomes de carbone et la nomenclature des cycles

Figure I2 : Les différentes classes de flavonoïdes

Figure I3 : Quatre exemples de flavonoïdes glycosylés abondants dans la nature : (a) naringine, (b) hespéridine, (c) rutine et (d) isoquercitrine

Figure I4 : Complexation de cations métalliques (Me^{n+}) par un flavonoïde

Figure I5 : Déterminants structuraux des bandes d'absorption des UV des flavonoïdes

Figure I6 : Représentation schématique de différence de régiosélectivité des procédés d'acylation par voie (a) chimique et (b) enzymatique d'un composé polyhydroxylé

Figure I7 : Schéma et forme fonctionnelle des termes liés d'un champ de forces classique

Figure I8 : Schéma et forme fonctionnelle des termes non liés d'un champ de forces classique

Figure I9 : Représentation schématique du principe d'une minimisation d'énergie statique : à partir d'un ensemble de coordonnées atomiques (conformation) de départ, l'algorithme retrouve le minimum local de la fonction d'énergie le plus proche, ainsi que la nouvelle conformation correspondant à ce minimum

Figure I10 : Les méthodes automatisées de docking les plus connues, avec les exemples plus fréquents dans la littérature, d'après les références

Figure I11 : Les fonctions de score les plus connues, avec les exemples les plus fréquents dans la littérature, d'après les références

Figure I12 : (a) Une molécule de soluté (sphère bleue) entourée par une cellule cubique de solvant répliquée en trois dimensions. (b) Vue schématique du système, montrant que chaque molécule interagit avec, au maximum, une image de chaque molécule « réelle » (B1 et B5 étant la même molécule, A1 ne doit interagir qu'avec une des deux)

Figure I13 : Une molécule d'éthane représentée en bâtons et sphères (à gauche) et en orbitales (à droite). Dans la représentation orbitale, les points foncés représentent les noyaux atomiques et les zones translucides les zones autour des noyaux où les électrons sont (dé)localisés.

Figure I14 : Représentation d'une fonction Gaussienne et d'une fonction de Slater.

Figure I15 : Schéma d'un complexe protéine-ligand traité par une du type QM/MM. la partie du système traitée par mécanique quantique est désignée par QM, incluant tout atome de substrat et des résidus catalytiques intervenant directement dans le mécanisme réactionnel. Le reste de la protéine et le solvant, désignés par MM, sont traités par un champ de force standard. La zone d'interaction entre les deux parties est représentée par une ligne rouge.

Figure I16 : Modèles obtenus par docking de la quercétine dans le site actif de la SOMT-9 : (a) l'enzyme native et (b) le mutant Y207G.

Figure I17 : Modèles obtenus par docking de (a) la lutéoline et (b,c) la 3'-méthoxy-lutéoline dans le site actif de la POMT-7.

Figure I18 : iinteractions stabilisant la quercétine dans le site actif de la 2,3-QD : en (a), les interactions hydrophobiques avec les résidus Phe75, Met51, Phe136, Pro164 et val63 ; en (b) et (c), les liaisons hydrogène établies grâce à une molécule d'eau soit avec la Thr53, soit avec l'Asn74.

Figure I19 : Succession des intermédiaires (I1, I2 et I3) et états de transition (TS1, TS2, TS3 et TS4) au cours de l'oxygénolyse de la quercétine. RI représente les réactifs et Pdt les produits. Les atomes d'hydrogène sont colorés en blanc, ceux d'oxygène en noir et ceux de carbone en gris.

Figure I20 : Snapshots des complexes, montrant les semi quinones (a) 3-O, (b) 7-O et (c) 4'-O pointées vers le cation Fe^{2+} du site actif de la LOX-3.

ARTICLE

Figure 1 : Classical mechanism of lipases : if $R_2 = H$ and $R_3 = \text{carbon chain}$, this is an esterification; if $R_2 = \text{carbon chain}$ and $R_3 = H$, this is an ester hydrolysis; if $R_2 = \text{carbon chain}$ and $R = \text{carbon chain}$, this is a transesterification.

CHAPITRE II

MATERIELS ET METHODES DE SIMULATION

Figure II1 : Structures des flavonoïdes étudiés. Dans l'ordre : la quercétine, l'isoquercitrine (quercétine-O-3-Glucose) et la rutine (quercétine-O-3-Glucose-Rhamnose).

Figure II2 : Les cellules contenant la (a) CALB et la (b) PCL, utilisées dans la procédure de préparation des cibles de docking.

Figure II3 : Représentation schématisée bidimensionnelle de la grille autour et au travers la protéine (traits noirs), et de ses points d'intersection. (a) Les points sont classés en occupés (cercles noirs) et libres (cercles gris). (b) Des points libres sont éliminés (cercles blancs) de manière à obtenir des cavités dont l'ouverture maximale est définie par l'utilisateur. Les points libres restants (cercles gris) définissent des cavités.

Figure II4 : À gauche, surface de Connolly, mettant en évidence la cavité catalytique de l'acétyl-CALB et les résidus qui l'entourent. À droite, en bleu, le nuage de points couvrant l'intérieur de cette cavité, correspondant à l'espace où seront dockés les flavonoïdes (le site actif).

Figure II5 : À gauche, surface de Connolly, mettant en évidence la cavité catalytique de l'acétyl-PCL et les résidus qui l'entourent. À droite, en bleu, le nuage de points couvrant l'intérieur de cette cavité, correspondant à l'espace où seront dockés les flavonoïdes (le site actif).

Figure II6 : Quatre orientations relatives de chaque conformère du ligand (poses) sont dockées dans le site actif.

Figure II7 : Représentation schématisée du fonctionnement de l'algorithme de docking LigandFit.

Figure II8 : Exemple de complexe solvaté avec une couche de molécules d'eau de 5 Å.

CHAPITRE III

ÉTUDE DE LA RÉGIOSÉLECTIVITÉ DANS L'ACÉTYLATION DES FLAVONOÏDES GLYCOSYLÉS ISOQUERCITRINE ET RUTINE CATALYSÉE PAR LA LIPASE B DE *CANDIDA ANTARCTICA*

ARTICLE

Figure 1 : Chemical structures of (a) isoquercitrin and (b) rutin, with the systematic numbering of carbon atoms and nomenclature of aglycon rings.

Figure 2 : The reaction mechanism of the CALB-catalyzed acylation of a flavonoid, evidencing the two tetrahedral intermediates and the acyl-enzyme complex. In this study, the acyl donor substrate is vinyl acetate, F-OH is the flavonoid to be acetylated (docked ligand) and the docking target is the acetyl-CALB.

Figure 3 : Scheme of structural requirements that flavonoid hydroxyl groups must satisfy in the models in order to be considered as reactive.

Figure 4 : Ribbons showing CALB crystal structure superimposed to its structure after the modelling procedure complexed with (a) isoquercitrin and (b) rutin. Docked flavonoids are represented in sticks. The crystal structure is colored in gray in both cases; that after the modelling procedure is colored in green in (a) and in violet in (b). Main chain differences are found in the sequences: (A) Cys311-Gly313, corresponding to the loop between the sheets $\beta 8$ and $\beta 9$; (B) Val190-Pro192, contained in the loop that connects the sheets $\beta 6$ and $\beta 7$; (C) Asn259-Pro262, contained in the loop that connects the helices $\alpha 9$ and $\alpha 10$; (D) Asn85-Ile87, contained in the helix $\alpha 3$; (E) Thr158-Ala162, contained in the loop between the helices $\alpha 6$ and $\alpha 7$.

Figure 5 : Orientations and hydrogen bonds interactions (dashed lines) of the substrates in the three best scored acetyl-CALB isoquercitrin complexes (I1, I2 and I3). For clarity, only the residues useful for discussion are shown and non-interacting hydrogen atoms are hidden.

Figure 6 : Orientations and hydrogen bonds interactions (dashed lines) of the substrates in the three best scored acetyl-CALB rutin complexes (R1, R2 and R3). For clarity, only the residues useful for discussion are shown and non-interacting hydrogen atoms are hidden.

Figure 7 : Connolly accessible surface of CALB binding pocket evidencing the hydrophobic interactions of isoquercitrin in I1 and rutin in R1. Hydrophobic regions of the enzyme pocket are colored in red and hydrophilic ones in white/blue.

ETUDE COMPLÉMENTAIRE

Figure III1 : Suivi des distances entre l'atome d'oxygène de l'acétate (Ace:O) et les atomes Gln106:NH, Thr40:NH et Thr40:OH au cours de la dynamique moléculaire des complexe I1 (à gauche) et R1 (à droite). Les flèches indiquent les frames échantillonnés pour l'analyse.

Figure III2 : Superposition des frames échantillonnés dans la trajectoire de I1 à 0 ps (noir), 400 ps (bleu), 800 ps (vert), 1200 ps (orange), 1600 ps (rose) et 2000 ps (violet). Pour une meilleure clarté, seuls les résidus et les atomes d'hydrogène importants pour la discussion sont montrés. En (a), une vue frontale, montrant la cavité, l'isoquercitrine et les résidus avec lesquels elle interagit directement à l'entrée ; en (b), une vue latérale, mettant en évidence la disposition de l'isoquercitrine par rapport aux résidus catalytiques et l'acétate par rapport au trou oxyanionique.

Figure III3 : Superposition des frames échantillonnés dans la trajectoire de R1 à 0 ps (noir), 100 ps (bleu), 250 ps (vert), 750 ps (orange), 1100 ps (rose) et 2000 ps (violet). Pour une meilleure clarté, seuls les résidus et les atomes d'hydrogène importants pour la discussion sont montrés. En (a), une vue frontale, montrant la cavité, la rutine et les résidus avec lesquels elle interagit directement à l'entrée ; en (b), une vue latérale, mettant en évidence la disposition de la rutine par rapport aux résidus catalytiques et l'acétate par rapport au trou oxyanionique.

Figure III4 : Représentation schématique des complexes (a) I1 et (b) R1. Dans les deux cas, l'acétate demeure correctement orienté vers les résidus Thr40 et Gln106 et les parties aglycones des flavonoïdes sont stabilisées à l'entrée de la cavité. L'hydroxyle accessible aux résidus catalytiques Ser105 et His224 sont le 6''-OH de l'isoquercitrine et le 4'''-OH de la rutine.

CHAPITRE IV

ETUDE DES SPÉCIFICITÉS DES LIPASES DE *CANDIDA ANTARCTICA* ET *PSEUDOMONAS CEPACIA* DANS L'ACÉTYLATION DU FLAVONOÏDE AGLYCONE QUERCÉTINE

Figure 1 : Chemical structure of quercetin ($R = H$), showing the systematic numbering of the carbon atoms, the nomenclature of the rings and the dihedral angle τ , defined by the atoms O1-C2-C1'-C2'. The monoglycosylated isoquercitrin corresponds to $R = \text{glucose}$ and the diglycosylated rutin to $R = \text{rutinose}$.

Figure 2 : The reaction mechanism of a lipase-catalyzed transesterification reaction, between the substrates $R_1\text{COOR}_2$ (acyl donor) and XOH (acyl acceptor) to give the products $R_2\text{OH}$ and $R_1\text{COOX}$. The involved first tetrahedral intermediate, the acyl-enzyme complex and the second tetrahedral intermediate are indicated.

Figure 3 : Fluctuations of the distances between the carbonyl oxygen of the serine-bound acetate and the oxyanion hole residues throughout the 2 ns molecular dynamics trajectories, for each quercetin binding mode within the catalytic cavities of PCL and CALB. The times corresponding to frames sampled for more detailed analysis are indicated by black arrows.

Figure 4 : Superposition of trajectory frames representative of the conformational changes observed during the molecular dynamics simulations for each quercetin binding mode within the catalytic cavities of PCL and CALB. In the first column, frontal views of the binding pocket surfaces are shown, focusing on the docked quercetin and the residues directly interacting with it. In the second column, lateral views of the active site region are presented, showing the docked quercetin, the serine-bound acetate and the residues constituting the catalytic triad and the oxyanion hole. The acetate carbonyl oxygen (Ace:O), the oxygen of the quercetin hydroxyl group which is the closest to the catalytic residues and the H of the oxyanion hole residues are represented by spheres. For clarity other hydrogen atoms and residues not referred in the discussion were hidden. For each case, the colours corresponding to the selected frames are: in (a) and (b), 0 ps (black), 1000 ps (blue), 2000 ps (orange); in (c) and (d), 0 ps (black), 1000 ps (blue), 2000 ps (orange); in (e) and (f), 0 ps (black), 100 ps (blue), 500 ps (green), 1000 ps (orange), 1500 ps (rose), 2000 ps (violet); in (g) and (h) 0 ps (black), 200 ps (blue), 600 ps (green), 900 ps (orange), 1010 ps (rose), 2000 ps (violet).

Figure 5 : Schematic representation of the final configuration of the investigated quercetin binding modes within the catalytic cavities of PCL and CALB, after a 2 ns trajectory.

CHAPITRE V

ETUDE DE LA REACTIVITÉ DE LA QUERCÉTINE, L'ISOQUERCITRINE ET LA RUTINE AVEC LA LIPASE B DE *CANDIDA ANTARCTICA* COMME BIOCATALYSEUR POUR LEUR ACÉTYLATION

Figure 1 : Chemical structures of the flavonoids (a) quercetin and its glycosylated analogues (b) isoquercitrin and (c) rutin. The hydroxyl groups accessible to the lipase catalytic residues, according to the complexes obtained by docking simulations, are indicated by arrows.

Figure 2 : Representation of the optimized mini systems used for DFT calculations. In the first column, the reactant states, which correspond to the Michaelis complexes. The atoms referred in the text are represented by spheres and identified by arrows. In the second column, the obtained intermediates. The regions of the systems that underwent modifications are bordered with dashed circles. All hydrogen atoms are coloured in white, oxygen in red and nitrogen in blue. Carbon atoms are coloured in grey for the amino-acid fragments, dark green for the serine-bound acetate, orange for the quercetin, purple for isoquercitrin and greenish for rutin.

Figure 3 : Scheme showing possible hydrogen bonds in the reactant states (RS) and in intermediates (Int). This latter is formed in the cases of isoquercitrin and rutin, but not in that of quercetin.

LISTE DES TABLEAUX

CHAPITRE I ETUDE BIBLIOGRAPHIQUE

Tableau I1 : Effet des éléments de structure sur l'activité antioxydante des flavonoïdes

Tableau I2 : Quelques activités biologiques de la quercétine, l'isoquercitrine et la rutine.

Tableau I3 : Solubilité (mM) des flavonoïdes quercétine, isoquercitrine et rutine dans l'eau, le tert-amyl alcool, l'acétone et l'acétonitrile

Tableau I4 : Sommaire des données de régiosélectivité dans l'acylation par voie enzymatique des flavonoïdes quercétine, isoquercitrine et rutine.

Tableau I5 : Sommaire de quelques études employant des outils de modélisation moléculaire pour identifier les interactions intermoléculaires qui sont à la base de l'inhibition des protéines par des flavonoïdes.

ARTICLE

Table 1 : Some review papers dealing with biotechnological applications of lipases

Table 2 : Some examples of the use of molecular modeling aiming to understand the enantioselectivity based on the recognition of substrates by lipases binding pocket.

Table 3 : Some examples of the use of molecular modeling aiming to understand the regioselectivity based on the recognition of substrates by lipases binding pocket.

CHAPITRE II MATERIELS ET METHODES DE SIMULATION

Tableau II1 : Paramètres de simulation employés dans les étapes de minimisation d'énergie du système contenant les protéines solvatées

Tableau II2 : Paramètres de simulation employés dans l'optimisation des complexes enzyme-substrat solvatés

CHAPITRE III ÉTUDE DE LA RÉGIOSÉLECTIVITÉ DANS L'ACÉTYLATION DES FLAVONOÏDES GLYCOSYLÉS ISOQUERCITRINE ET RUTINE CATALYSÉE PAR LA LIPASE B DE *CANDIDA ANTARCTICA*

Table 1 : Minimal and maximal transacetylating distances a during the MD production phase carried out on the productive complexes

Table 2 : Hydrogen bond distances and angles involving the substrates in the three best-scored complexes after optimization

Table 3 : Distances a of buried hydroxyl groups of the docked flavonoids from the His224:Ne and the acetate carbonyl carbon (Ace:C) atoms in the three best scored complexes after optimization

ETUDE COMPLÉMENTAIRE

Tableau III1 : Distances d'intérêt pour le processus catalytique au cours de la trajectoire des complexes I1

Tableau III2 : Distances d'intérêt pour le processus catalytique au cours de la trajectoire du complexe R1

CHAPITRE IV

ETUDE DES SPÉCIFICITÉS DES LIPASES DE *CANDIDA ANTARCTICA* ET *PSEUDOMONAS CEPACIA* DANS L'ACÉTYLATION DU FLAVONOÏDE AGLYCONE QUERCÉTINE

Table 1 : Maximal *rmsd* values (Å) for the best scored complex of each orientation family, with regard to the crystal structures of the enzymes.

Table 2 : Catalytically relevant distances (Å) for the substrates and catalytic residues in the A-ring binding mode for PCL

Table 3 : Catalytically relevant distances (Å) for the substrates and catalytic residues in the B-ring binding mode for PCL

Table 4 : Catalytically relevant distances (Å) for the substrates and catalytic residues in the A-ring binding mode for CALB

Table 5 : Catalytically relevant distances (Å) for the substrates and catalytic residues in the A-ring binding mode for CALB

CHAPITRE V

ETUDE DE LA REACTIVITE DE LA QUERCÉTINE, L'ISOQUERCITRINE ET LA RUTINE AVEC LA LIPASE B DE *CANDIDA ANTARCTICA* COMME BIOCATALYSEUR POUR LEUR ACÉTYLATION

Table 1 : Relevant interatomic distances and bond lengths in the reactant state (RS) and the intermediates (Int) for the optimized mini-systems

Table 2 : Relevant Mulliken atomic charges in the reactant states (RS) and the intermediates (Int) for the optimized mini-systems

LISTE DES SIGLES ET ABBREVIATIONS

[Bmin(BF) ₄]	1-Butyl-3-méthylimidazolium-tétrafluoroborate
[Bmin(PF) ₆]	1-Butyl-3-méthylimidazolium-hexafluorophosphate
2,3-QD	Quercétine-2,3-dioxygénase
Å	Angström (= 10 ⁻¹⁰ m)
ABNR	<i>Adopted-Basis Newton-Raphson</i>
AM1	<i>Austin Model 1</i>
AMBER	<i>Assisted Model Building and Energy Refinement</i>
BCL	<i>Burkholderia cepacia</i> Lipase (= <i>Pseudomonas cepacia</i> Lipase)
BO	Born-Oppenheimer
BSLA	<i>Bacillus subtilis</i> Lipase A
CALA	<i>Candida antarctica</i> Lipase A
CALB	<i>Candida antarctica</i> Lipase B
CFF	<i>Consistent Force Field</i>
CG	<i>Conjugate Gradient</i>
CHARMm	<i>Chemistry at HARvard for Macromolecular mechanics</i>
CRL	<i>Candida rugosa</i> Lipase
CVFF	<i>Consistent Valence Force Field</i>
DFT	<i>Density Functional Theory</i>
DNP	<i>Double Numerical with Polarization</i>
E	Rapport énantiomérique
ES	Enzyme-Substrat
GROMACS	<i>GROningen MACHine for Chemical Simulation</i>
GROMOS	<i>GROningen MOlecular Simulation</i>
HF	Hartree-Fock
HLL	<i>Humicola lanuginosa</i> Lipase
INDO	<i>Intermediate Neglect of Differential Overlap</i>
KS	Kohn-Sham
LJ	Lennard-Jones
LOX-3	Lipoxygénase 3
MD	<i>Molecular Dynamics</i>
MM	<i>Molecular Mechanics</i>
MM-PBSA	<i>Molecular Mechanics - Poission-Boltzman Surface Area</i>
MMFF	<i>Merck Molecular Force Field</i>
MNDO	<i>Modified Neglect of Differential Overlap</i>
NAMD	<i>NAnoscale Molecular Dynamics</i>
NPT	<i>Number-Pressure-Temperature</i>
NVE	<i>Number-Volume-Energy</i>

NVT	<i>Number-Volume-Temperature</i>
OPLS	<i>Optimized Potentials for Liquid Simulations</i>
PAL	<i><u>Pseudomonas aeruginosa</u> Lipase</i>
PBC	<i>Periodic Boundary Conditions</i>
PCL	<i><u>Pseudomonas cepacia</u> Lipase (= <u>Burkholderia cepacia</u> Lipase)</i>
PDB	<i>Protein Data Bank</i>
PGa	<i>Penicillin G acylase</i>
PLP	<i>Pairwise Linear Potential</i>
PM3	<i>Parameterized Model 3</i>
PMF	<i>Potential of Mean Force</i>
PML	<i><u>Pseudomonas mendocina</u> Lipase</i>
POMT-7	<i>Poppler Methyltransferase 7</i>
PPL	<i>Porcine Pancreatic Lipase</i>
QM	<i>Quantum Mechanics</i>
QM/MM	<i>Quantum Mechanics/Molecular Mechanics</i>
RDL	<i><u>Rhizopus delamar</u> Lipase</i>
RML	<i><u>Rhizomucor miehei</u> Lipase</i>
RMN	<i>Résonance Magnétique Nucléaire</i>
ROL	<i><u>Rhizopus orizae</u> Lipase</i>
<i>rmsd</i>	<i>Root Mean Square Deviation</i>
SC	<i>Subtilisine Carlsberg</i>
SCF	<i>Self-Consistent Field</i>
SD	<i>Steepest Descent</i>
SOMT-9	<i>Soybean O-Methyltransferase 9</i>
supCO ₂	<i>Supercritical carbon dioxide</i>
vdW	<i>van der Waals</i>
UV	<i>Ultra-Violet</i>
ZINDO	<i>Zerner's Intermediate Neglect of Differential Overlap</i>

Introduction

L'utilisation d'enzymes en synthèse organique connaît un essor depuis les années 1990, et ceci pour deux raisons principales : premièrement, les biocatalyseurs présentent souvent des propriétés uniques de sélectivité vis-à-vis d'un substrat ou d'une classe de substrat donné. Ainsi, des conversions hautement énantiosélectives ou régiosélectives deviennent plus aisées, permettant la synthèse de nouvelles molécules à structure contrôlée, très difficiles voire impossibles à obtenir par les voies « classiques » de synthèse organique. Deuxièmement, ces procédés enzymatiques nécessitent, de manière générale, des conditions opérationnelles plus douces que leurs analogues chimiques : températures moins élevées, pression ambiante, absence de catalyseurs inorganiques toxiques. Ils répondent donc aux besoins actuels de développement de technologies et de procédés moins agressifs pour l'environnement.

Depuis quelques années notre équipe développe et applique, avec succès, des procédés enzymatiques pour l'obtention de dérivés régiosélectivement acylés de flavonoïdes. Les biocatalyseurs utilisés sont des lipases, qui catalysent en milieu organique la formation de liaisons ester. Les flavonoïdes, abondants dans la plupart des aliments d'origine végétale, possèdent un éventail assez vaste de propriétés biologiques potentiellement bénéfiques : prévention de la propagation d'espèces radicalaires néfastes aux cellules, pouvoir anti-inflammatoire, renforcement et protection des vaisseaux sanguins, action photoprotectrice pour la peau, pour ne citer que les plus connues. Ces activités ont été clairement corrélées à des éléments structuraux de ces molécules, notamment le nombre et la position des hydroxyles libres. Toutefois, ces flavonoïdes sont peu miscibles que ce soit dans des milieux hydrophiles ou hydrophobes, ce qui constitue un obstacle pour la valorisation technologique de leurs propriétés. Ainsi, leur acylation constitue une solution pour produire des dérivés plus apolaires, donc plus facilement solubles dans les formulations cosmétiques et alimentaires, qui ont souvent un caractère hydrophobe. Les procédés enzymatiques apparaissent comme une bonne solution pour établir un compromis entre amélioration de la solubilité, de la stabilité et maintien des propriétés des flavonoïdes, puisque l'utilisation d'enzymes rend possible l'obtention de dérivés acylés de manière régiosélective. Ainsi, les groupements hydroxyle à l'origine des effets biologiques recherchés peuvent être conservés intacts.

Le nombre d'articles relatant l'acylation enzymatique de ces trois flavonoïdes, en utilisant des lipases comme biocatalyseurs, s'est multiplié dernièrement. De nouveaux dérivés plus solubles et plus stables, ayant des propriétés biologiques maintenues voire accrues par rapport aux molécules naturelles, ont pu être synthétisés. Les paramètres réactionnels ont été améliorés, pour rendre les procédés économiquement et écologiquement plus performants. Mais ces travaux expérimentaux ne fournissent pas d'information permettant de répondre à des questions plus fondamentales. Pourquoi tel hydroxyle précis de tel flavonoïde est acylé par une lipase donnée ? Pourquoi une lipase donnée catalyse la réaction pour tel flavonoïde, et non pour un autre ? Pour répondre à des questions de ce

type, une analyse des interactions entre les enzymes et les substrats est nécessaire. Il existe des approches expérimentales permettant d'évaluer directement de telles interactions. Par exemple, des méthodes basées sur la diffraction de rayons X ou la RMN peuvent mener à l'élucidation des modes d'ancrage des substrats dans le site actif des enzymes ou l'évaluation de la densité électronique autour des différents groupements hydroxyle, ce qui pourrait être ensuite corrélé à leur réactivité. Cependant, la mise en œuvre de ces techniques requiert la cristallisation préalable du complexe enzyme-substrat ou des intermédiaires réactionnels. Ceci est en pratique très difficile car la durée d'existence de ces espèces est très courte. Par conséquent, il est indispensable de trouver d'autres alternatives afin de comprendre les bases moléculaires dictant la sélectivité des réactions enzymatiques d'acylation des flavonoïdes.

Une alternative prometteuse est l'utilisation de simulations moléculaires : il s'agit de l'application de lois plus fondamentales de la physique et de la chimie théorique aux biomolécules, à l'aide de ressources informatiques permettant d'exécuter les calculs nécessaires et d'analyser les informations qui en sont issues. En effet, nombre de travaux récents, publiés dans des périodiques ou présentés dans des congrès scientifiques, illustrent l'application de ces approches pour comprendre, expliquer ou même prédire les modes d'interaction entre enzymes et substrats. Alliés aux connaissances expérimentales, les résultats de simulation fournissent une compréhension de la sélectivité d'enzymes diverses, dans des contextes variés. En particulier, les structures tridimensionnelles à bonne résolution de nombreuses lipases, ont été déterminées par diffraction de rayons X et sont disponibles dans le domaine public (www.pdb.org). Depuis, de nombreuses études de modélisation moléculaire sur les propriétés de sélectivité de ces enzymes sont à disposition dans la littérature. Toutefois, aucune d'entre elles ne traite spécifiquement des flavonoïdes.

C'est pourquoi dans ce travail de thèse nous avons l'objectif d'introduire l'application d'outils de modélisation moléculaire pour comprendre et expliquer, au niveau atomique, l'influence des structures de la lipase et des flavonoïdes sur la régiosélectivité de leur acylation. Comme réactions modèles, nous avons choisi l'acétylation de la quercétine et de deux de ses dérivés glycosylés, l'isoquercitrine (quercétine-3-O-glucose) et rutine (quercétine-3-O-rhamnose), en utilisant les lipases de *Candida antarctica* (CALB) ou de *Pseudomonas cepacia* (PCL). Ce choix se justifie par plusieurs raisons : d'abord, ces trois flavonoïdes figurent parmi les plus abondants dans le règne végétal et présentent toutes les propriétés bénéfiques précédemment mentionnées, ou la plupart d'entre elles. En conséquence, la plupart des données expérimentales disponibles sur l'acylation de flavonoïdes concernent ces trois composés. Les deux lipases mentionnées, CALB et PCL, ont été utilisées dans ces études. Par ailleurs, la réaction d'acétylation a été choisie car il s'agit de la seule pour laquelle des résultats expérimentaux de régiosélectivité sont disponibles pour tous les trois flavonoïdes. En effet, la confrontation des résultats de simulation à des données empiriques est essentielle pour l'élaboration de modèles fiables. En outre, l'acétate a l'avantage d'avoir une chaîne alkyle très courte, à bas degré de liberté, permettant de focaliser l'analyse, pour une enzyme donnée, sur la seule structure du flavonoïde.

Plus précisément, nous tenterons d'apporter des réponses plausibles à trois questions :

- (1) *Comment les modes d'interaction enzyme-substrats influencent la régiosélectivité observée dans l'acétylation des flavonoïdes glycosylés isoquercitrine et rutine catalysée par la lipase B de Candida antarctica (CALB) ?*
- (2) *Comment les modes d'interaction enzyme-substrats peuvent expliquer le fait que la lipase de Pseudomonas cepacia (PCL) catalyse l'acétylation du flavonoïde quercétine tandis que la CALB ne catalyse pas cette réaction ?*
- (3) *Dans le cas de la CALB, le type de groupement OH de ces trois flavonoïdes (quercétine, isoquercitrine et rutine) qui atteignent les résidus catalytiques peut-elle aussi contribuer à la sélectivité ?*

Dans le **premier chapitre** nous commençons par récapituler l'importance des flavonoïdes et l'intérêt de leur acylation enzymatique, ainsi que les principaux résultats expérimentaux publiés sur ce thème. Ensuite, nous proposons une description des approches de simulation moléculaire les plus utilisées pour l'étude des interactions enzyme-substrat. Puis, nous présentons un état de l'art des travaux récents portant sur l'utilisation d'outils de modélisation moléculaire pour une meilleure compréhension des modes d'interaction entre flavonoïdes et protéines. Enfin, dans la dernière partie de ce chapitre, rédigée sous la forme d'un article de revue (suite à l'invitation de Mr. Joseph Boudrand, éditeur du journal *Process Biochemistry*), nous présentons et commentons plusieurs travaux portant sur l'application de ces outils pour la rationalisation des propriétés de sélectivité (principalement l'énantiosélectivité et la régiosélectivité) de divers bioprocédés catalysés par des lipases. Ces deux dernières parties nous ont été particulièrement utiles, en nous fournissant des informations générales sur les interactions protéine-ligand qui se sont avérées précieuses par la suite, lors de l'analyse de nos modèles.

Le **deuxième chapitre** contient une description minutieuse du mode opératoire suivi pour les simulations. Nous y décrivons le protocole utilisé dans chaque étape, de l'obtention des structures de départ à l'analyse des modèles finaux. Dans chaque étape, nous avons pris le soin de présenter le principe et le but de la simulation, de spécifier et justifier les paramètres utilisés. Les éléments pris en compte pour l'analyse des résultats y sont aussi présentés et justifiés.

Dans le **troisième chapitre**, nous étudions l'acétylation régiosélective des flavonoïdes glycosylés isoquercitrine et rutine catalysée par la CALB. Toute la procédure de docking pour l'obtention de modèles fiables des modes d'interaction CALB-substrats (flavonoïdes et acétate) y est décrite, ainsi que l'analyse structurale des complexes obtenus. La corrélation entre la structure de ces complexes et les données expérimentales de régiosélectivité de la réaction est aussi discutée. Cette partie a fait l'objet d'une publication parue dans le *Journal of Molecular Catalysis B : Enzymatic* (v. 59, p. 96-105, 2009). En outre, sont aussi présentés et discutés les résultats d'une évaluation complémentaire de la stabilité des complexes par dynamique moléculaire.

Dans le **quatrième chapitre**, nous nous focalisons sur l'acétylation du flavonoïde aglycone quercétine, en utilisant comme biocatalyseurs les lipases CALB et PCL. La combinaison de simulations de docking et dynamique moléculaire, qui s'était avéré performant lors de l'étude précédente, a été reprise comme base méthodologique pour étudier les orientations et les différences de stabilisation des substrats (quercétine et acétate) dans la cavité des deux enzymes. Nous discutons ici de la corrélation entre les résultats de simulation et les comportements catalytiques de ces deux enzymes observées expérimentalement lors de l'acétylation de la quercétine. Ce chapitre a fait l'objet d'un article scientifique, à soumettre prochainement.

Le **cinquième et dernier chapitre** de notre étude porte sur l'application de la Théorie Fonctionnelle de la Densité (DFT) comme outil pour compléter les informations relatives au docking des trois flavonoïdes mentionnés (quercétine, isoquercitrine et rutine) dans le site actif de la CALB. Ici, il s'agit d'une première évaluation de la réactivité des flavonoïdes au sein des complexes enzyme-substrats et de confronter les informations recueillies aux cours des simulations quantiques aux observations expérimentales. Ce chapitre a été aussi rédigé sous la forme d'un projet de publication.

CHAPITRE I :

Étude
Bibliographique

1.

LES FLAVONOÏDES

1.1. Généralités : origine et structure

Les flavonoïdes sont des polyphénols produits comme métabolites secondaires de la plupart des plantes. Ces composés sont abondamment présents dans la plupart des aliments d'origine végétale, en particulier dans les légumes feuilles (salade, chou, épinard etc.), dans l'oignon, dans les fruits rouges et dans les téguments externes des fruits en général [1].

Ces molécules dérivent de la voie biosynthétique de l'acide shikimique. Leur précurseur est le 4-hydroxycinnamate-coenzyme A, synthétisé à partir de l'acide aminé phénylalanine. A l'exception des chalcones et aurones, les flavonoïdes présentent le même motif structural de base formé par 15 atomes de carbone : 2 cycles en C6 (cycles A et B) reliés entre eux par une chaîne en C3 (cycle C, un noyau 2-phényl-1-benzopyrane) en position 2 ou 3 (Figure I1).

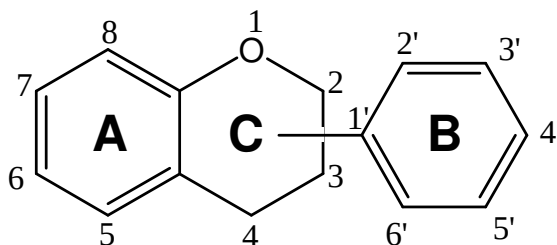


Figure I1 - Motif structural de base des flavonoïdes, avec la numérotation systématique des atomes de carbone et la nomenclature des cycles [1].

En fonction du degré d'oxydation et d'insaturation du cycle C, les flavonoïdes sont subdivisés en différentes classes. Les composés appartenant à une même classe se distinguent entre eux par le nombre et les positions des hydroxyles, le degré de polymérisation et, parfois, la présence d'un substituant méthoxyle, méthyle ou prényle [2]. Les classes des flavonoïdes sont schématisées dans la Figure I2, avec des exemples représentatifs de chacune. Les flavonoïdes sont fréquemment présents dans la nature sous forme glycosylée. La présence d'un sucre modifie les propriétés de la molécule par rapport à son analogue aglycone, en particulier l'hydrophobicité et la solubilité. La partie aglycone est liée à la partie glycosydique généralement par l'un des hydroxyles phénoliques, en particulier ceux en position 3 (pour les flavonols) ou 7 (pour les flavones) et parfois, en 6 ou 8 dans les cas où ces positions comportent un groupement hydroxyle. Les principaux monosaccharides entrant dans leur constitution structurale sont : l'acide D-glucuronique, l'acide D-galacturonique, le L-arabinose, L-mannose, le D-xylose, le D-allose, le D-galactose, le D-glucose et le D-rhamnose. Ces deux derniers sont les plus couramment rencontrés [2, 3]. Quelques exemples de flavonoïdes glycosylés abondants dans la nature sont donnés dans la Figure I3.

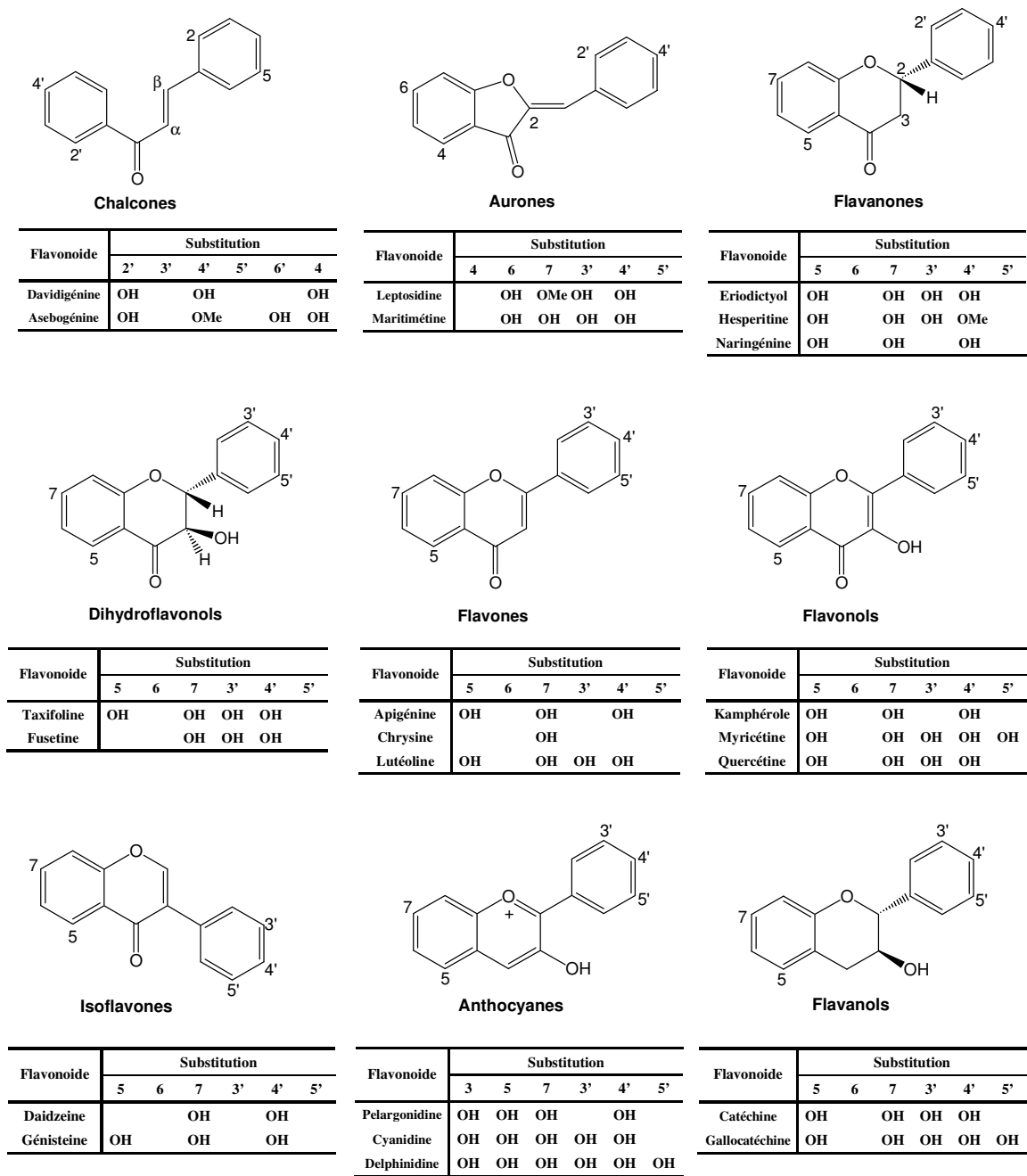


Figure I2 - Les différentes classes de flavonoïdes [4, 5].

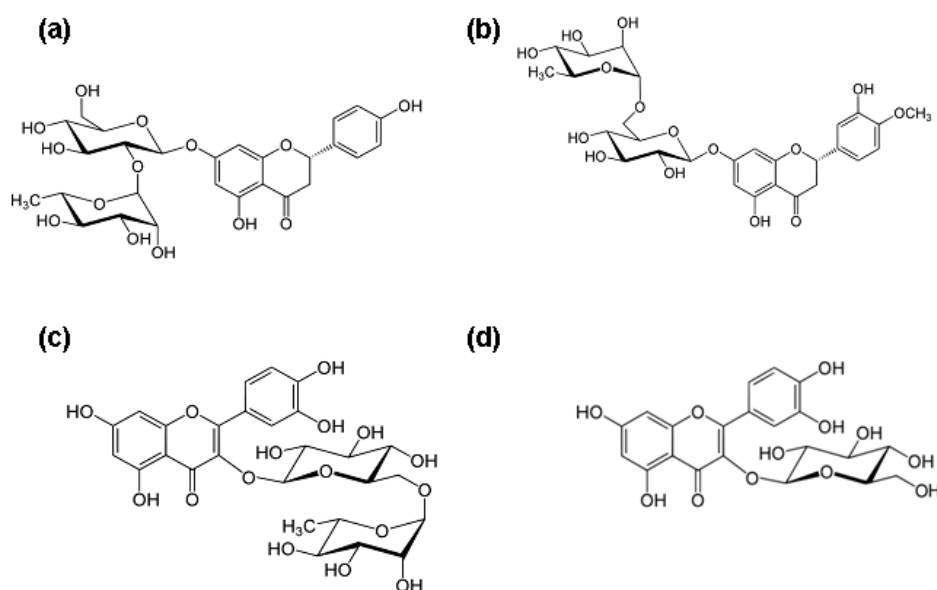


Figure 13 - Quatre exemples de flavonoïdes glycosylés abondants dans la nature : (a) naringine, (b) hespéridine, (c) rutine et (d) isoquercitrine

1.2. Activités biologiques

Au cours des dernières années, il a été démontré que les flavonoïdes présentent d'importantes activités biologiques, essentiellement de type anti-oxydantes et photoprotectrices. Grâce à ces propriétés, un grand intérêt est actuellement porté à ces molécules aussi bien par les milieux académiques que les industriels, s'intéressant aux domaines des cosmétiques, des produits agro-alimentaires et pharmaceutiques.

1.2.1. Activité photoprotectrice

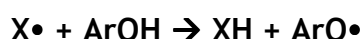
Les flavonoïdes exercent une activité photoprotectrice au sein des plantes, résultant en grande partie de leur forte absorption dans le domaine de l'UV [6]. Suivant leur structure, leurs maxima d'absorption se situent dans la zone 270-350 nm. Ceci permet d'envisager leur utilisation pour prévenir ou diminuer les dommages induits par le stress oxydant au niveau de la peau, en particulier ceux dus aux radiations UV-B (290-320 nm), connues pour leurs effets délétères. En effet, plusieurs brevets ont revendiqué l'utilisation de flavonoïdes dans la formulation de filtres solaires et agents anti-age [7-10].

1.2.2. Activité anti-oxydante

Un anti-oxydant peut être défini comme étant une substance qui, lorsqu'elle est présente dans un milieu à des concentrations plus faibles que celle d'un substrat oxydable (protéines, acides gras, acides nucléiques etc.), évite ou ralentit considérablement l'oxydation de ce substrat [11]. Les

flavonoïdes peuvent présenter une action anti-oxydante (ou parfois pro-oxydante), selon la concentration et les caractéristiques du milieu. Au vu de ces propriétés et de l'absence de toxicité aux concentrations habituellement utilisées [12], ils sont souvent ajoutés comme anti-oxydant dans des formulations alimentaires (en tant qu'additif) et cosmétiques (en tant qu'ingrédients actifs prévenant l'action néfaste d'espèces radicalaires sur la peau) [13, 14]. L'activité anti-oxydante de ces composés peut se produire selon trois mécanismes distincts : piégeage de radicaux libres, inhibition d'enzymes et complexation de cations métalliques [15].

- *Piégeage de radicaux libres* : les flavonoïdes sont capables de piéger des radicaux libres oxygénés ($X\bullet$) par transfert d'un atome d'hydrogène ou d'un électron:



Le radical aryloxy formé ($\text{ArO}\bullet$ ou $\text{ArOH}^{\bullet+}$) est stabilisé par résonance électronique. L'électron non apparié peut se délocaliser sur l'ensemble du système aromatique. Celui-ci peut continuer à évoluer selon plusieurs processus (dimérisation, dismutation, recombinaison avec d'autres radicaux, réduction en molécule parent ou oxydation en quinone), ou réagir avec d'autres radicaux ou biomolécules (protéines, acides nucléiques).

- *Inhibition d'enzymes* : les flavonoïdes ont une activité inhibitrice vis-à-vis de certaines enzymes notamment des oxydoréductases qui font intervenir des espèces radicalaires au cours de leur cycle catalytique (peroxydases, lipoxygénases, cyclo-oxygénases, monooxygénases, xanthine-oxydases, protéine-kinases et bien d'autres).
- *Chélation de cations métalliques* : le pouvoir anti-oxydant des flavonoïdes peut s'exercer indirectement par la complexation de cations de métaux de transition comme le Fe^{3+} , Al^{3+} , Zn^{2+} , Cu^{2+} et d'autres (Figure I4). En effet, ces ions accélèrent la formation d'espèces oxygénées pro oxydantes. La complexation de flavonoïdes par des métaux de transition peut aussi améliorer leur pouvoir antioxydant en diminuant leur potentiel d'oxydation.

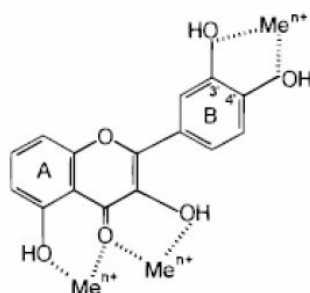


Figure I4 - Complexation de cations métalliques (Me^{n+}) par un flavonoïde [16].

1.2.3. Relations entre structure et activités biologiques

Les propriétés photoprotectrices des flavonoïdes sont une conséquence de l'absorption des rayonnements UV par les cycles aromatiques. Typiquement, les spectres UV des flavonoïdes exhibent deux bandes : la bande I (300-380 nm) est associée à l'absorption de la partie cinnamoyle (noyau B) ; la bande II (240-280 nm) à celle de la partie benzoyle (Figure I5).

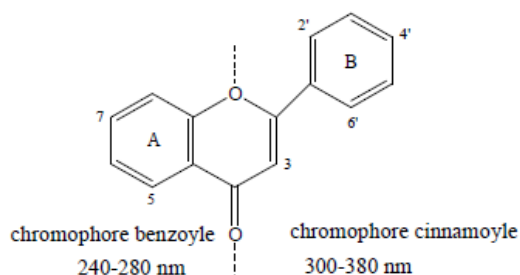


Figure I5 - Déterminants structuraux des bandes d'absorption des UV des flavonoïdes [5].

L'étude de l'effet de la structure des flavonoïdes sur leur activité antioxydante a fait l'objet de nombreux travaux. La présence de certains éléments structuraux a été corrélée à cette activité. En fonction du mécanisme d'action antioxydante, quelques différences dans ces relations structure-activité ont été observées. Le tableau I1 résume l'effet des éléments de structure des flavonoïdes sur les activités anti-radicalaires, de complexation de cations métalliques, d'inhibition de peroxydation de lipides et d'inhibition de la xanthine-oxydase.

Tableau I1 - Effet des éléments de la structure des flavonoïdes sur leur activité antioxydante, d'après les références [17-23]

Activité	Facteurs accroissant le pouvoir antioxydant
<i>Antiradicalaire</i>	▪ Proportionnelle au nombre de groupements -OH libres ▪ Présence d'un groupement -C=O en C4 ▪ Double liaison entre les atomes de carbone C2 et C3 ▪ Présence de groupements -OH en C3' et en C4' ▪ La méthylation a des effets variables et difficiles à quantifier
<i>Complexation de cations métalliques</i>	▪ Présence d'un -C=O en C4 + un -OH en C5 et/ou C3 ▪ Présence de groupements -OH en C3' et en C4' ▪ La présence d'un sucre n'a guère d'effet
<i>Inhibition de la peroxydation lipidique</i>	▪ Présence de groupements -OH en C3' et en C4' ▪ Présence d'un groupement -C=O ou -OH en C4 ▪ Présence de groupements -OH en C3, C5, C7, C3' et C4'
<i>Inhibition de la xanthine-oxydase</i>	▪ Double liaison entre les atomes de carbone C2 et C3 ▪ Présence d'un groupement -C=O en C4 ▪ Présence d'un groupement -OH en C5 ou C7 ▪ Angle de torsion entre les cycles B et C d'environ 27°

1.3. Acylation des flavonoïdes

Les flavonoïdes ont des propriétés physico-chimiques qui peuvent constituer une entrave à la valorisation de leurs bienfaits biologiques. En effet, ces composés sont plutôt instables à la lumière, à l'oxygène et à des températures élevées et généralement peu solubles, que ce soit en milieux aqueux ou hydrophobes [1]. Ces caractéristiques engendrent des difficultés pour les incorporer et les stabiliser dans des formulations alimentaires et cosmétiques. Afin de contourner ces inconvénients, des modifications de la structure des flavonoïdes peuvent être envisagées. Ainsi, il a été montré que l'hydroxylation [24], l'oligomérisation [25] et l'acylation [26-28] de ces molécules peuvent conduire à des dérivés dont la solubilité et la stabilité dans des milieux complexes sont améliorées. Dans cette étude, nous nous focalisons plus spécifiquement sur l'acylation des flavonoïdes, qui consiste en un greffage de chaînes acyle.

Il a été montré que les dérivés acylés des flavonoïdes sont plus hydrophobes (ou lipophiles) que la molécule d'origine et par conséquent, plus solubles et plus stables dans des formulations à caractère hydrophobe [27, 28]. Ainsi, le terme « acylation » (qui fait référence à la transformation chimique exécutée) est couramment substitué par « lipophilisation » (un terme qui évoque une caractéristique renforcée du produit de la réaction). L'acylation des flavonoïdes peut être accomplie par voie chimique ou par voie enzymatique.

1.3.1. Acylation par voie chimique

Des procédés d'acylation par voie chimique des flavonoïdes ont fait objet de quelques brevets [29-31]. Toutefois, ces procédés sont problématiques en raison de la sensibilité des flavonoïdes aux températures élevées et aux milieux alcalins requis pour la réaction d'esterification. De plus, ces procédés chimiques ne sont pas sélectifs, menant à l'estérification de plusieurs (voire tous) les groupements hydroxyles présents dans la molécule du flavonoïde. Deux inconvénients en découlent alors :

- un mélange de produits présentant différents degrés d'esterification est obtenu, augmentant le nombre d'opérations de séparation et de purification et donc le coût du procédé.
- les produits multi estérifiés ne gardent souvent pas les propriétés biologiques du flavonoïde d'origine. En effet, comme discuté dans la section 1.2.3, les activités biologiques des flavonoïdes sont directement corrélées à certains motifs structuraux, dont la présence de groupements hydroxyles libres sur des positions spécifiques de la molécule [18-23].

Ainsi, il est impératif que ces hydroxyles ne soient pas acylés, afin de conserver les propriétés biologiques d'intérêt. C'est pourquoi l'acylation des flavonoïdes doit être régiosélective, nécessitant ainsi des alternatives aux procédés chimiques.

1.3.2. Acylation par voie enzymatique

Les procédés d'acylation enzymatique présentent deux grands avantages qui les rendent préférables aux procédés chimiques précédemment évoqués [32]:

- Le premier est le fait que ces procédés sont menés sous des conditions opérationnelles plus douces, s'insérant parfaitement dans le cadre actuel de recherche de technologies plus économiques et moins agressives vis-à-vis de l'environnement.
- Le deuxième avantage est la spécificité et la sélectivité (régio-, énantio- ou chimio-) que les enzymes présentent vis-à-vis des substrats, permettant des modifications ciblées et contrôlées de ceux-ci (Figure I6). Ceci est essentiel dans le cas de l'acylation des flavonoïdes, où la greffe régiosélective d'un groupement acyle établit un compromis entre l'amélioration de la solubilité et de la stabilité de la molécule et le maintien voire le renforcement de ses activités biologiques.

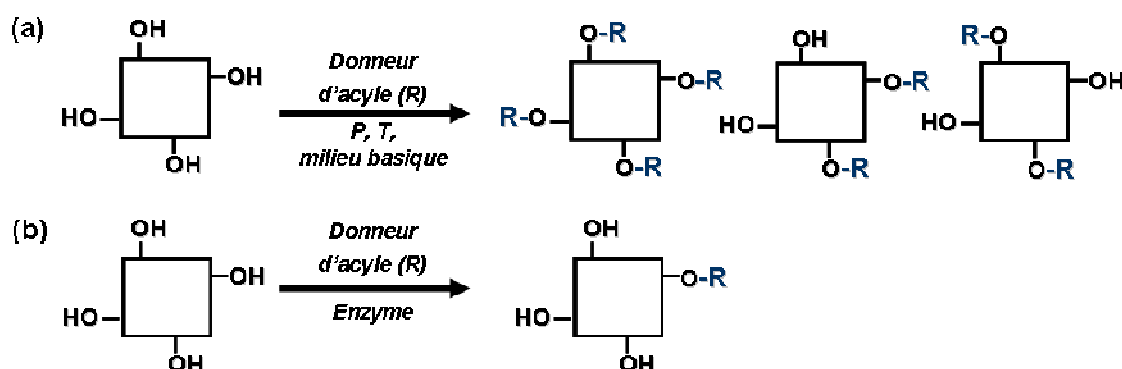


Figure I6 - Représentation schématique de la différence de régiosélectivité des procédés d'acylation par voie (a) chimique et (b) enzymatique d'un composé polyhydroxylé.

De nombreux travaux portant sur l'acylation enzymatique de composés polyphénoliques, notamment des flavonoïdes, ont été menés au cours des dix dernières années, montrant que des dérivés acylés de ces composés peuvent être obtenus de manière régiosélective. Des revues de littérature sur le sujet ont été récemment publiées [26-28]. Le résultat d'une analyse objective des informations rapportées dans l'ensemble de ces travaux peut être résumée comme suit :

- L'acylation enzymatique est très souvent efficace avec les flavonoïdes glycosylés. Les rendements de conversion sont généralement bons (70% à ≈100% dans les premières 24 h) et les procédés sont remarquablement régiosélectifs. La plupart des fonctions hydroxyle du flavonoïde ne sont pas touchées ; ces sont principalement des hydroxyles de la partie glycosydique qui sont acylés. Dans la plupart des cas, seuls des mono-esters ont été détectés. Parfois, la formation de diesters a été relatée, mais dans des proportions moins importantes que les mono-esters. Ces synthèses ont pu être accomplies en utilisant divers donneurs d'acyle comme des acides gras saturés à courte, moyenne ou longue chaîne (C2 à C18), des acides phénoliques (notamment les acides caféique, cinnamique et p-coumarique) et leurs esters de vinyle ou éthyle.

- L'acylation enzymatique est plus difficile avec les flavonoïdes aglycones. Dans les quelques cas rapportés, des produits ont pu être obtenus à condition d'utiliser des donneurs d'acyle activés de type esters vinyliques. Malgré tout, les rendements restent plutôt faibles (environ 30% à 60% dans les premières 24 h) et les procédés sont peu régiosélectifs (formation de mono-, di- et parfois tri-esters).
- Les enzymes qui catalysent le plus efficacement ces réactions sont les lipases (EC 3.1.1.3). Parmi les plus citées, les lipases de *Candida antarctica* (CALB) et *Pseudomonas cepacia* (PCL) se révèlent particulièrement efficaces pour ces procédés d'acylation. L'utilisation d'autres lipases, comme celles de *Candida rugosa* (CRL), *Humicola lanuginosa* (HLL), *Rhizomucor miehei* (RML) et *Rhizopus oryzae* (ROL), d'estérases et de subtilisines a également été rapportée par certains auteurs, mais avec des performances plus limitées.
- Puisque les lipases catalysent naturellement l'hydrolyse d'esters dans des milieux aqueux, les milieux utilisés pour les bioprocédés d'estérification sont soit le donneur d'acyle fondu soit des solvants organiques, purs ou en mélanges. Parmi eux, les solvants aux meilleures performances de réaction sont le *tert*-butanol, l'acétonitrile, l'acétone et le tétrahydrofurane. L'extraction de l'eau du milieu réactionnel favorise la synthèse des dérivés acylés en déplaçant l'équilibre thermodynamique de la réaction. La faisabilité de ces synthèses dans des milieux réactionnels de type liquides ioniques [Bmin(PF₆)] ou [Bmin(BF₄)] a été plus récemment démontrée.
- Les esters de flavonoïdes produits au cours de ces bioprocédés ont été caractérisés comme étant plus hydrophobes, plus résistants à la lumière et plus thermostables. De plus, leurs activités anti-oxydantes *in vitro* et/ou anti-microbiennes sont conservées ou accrues par rapport au flavonoïde d'origine.

1.4. La quercétine, l'isoquercitrine et la rutine

1.4.1. Distribution

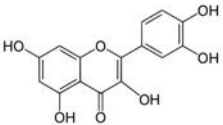
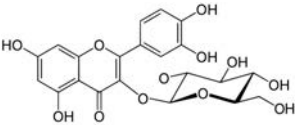
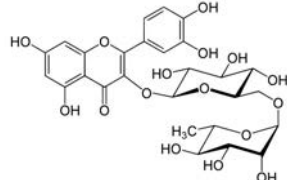
Parmi les flavonoïdes dont l'acylation par voie enzymatique a été étudiée figurent principalement le flavonol quercétine et deux de ses analogues glycosylés, l'isoquercitrine (quercétine-3-O-glucose) et la rutine (quercétine-3-O-glucose-rhamnose). Leurs formules structurales peuvent être trouvées dans le Tableau I2 (page suivante). C'est sur ces molécules que porte également notre étude. Ces trois flavonoïdes sont en effet très abondants dans le règne végétal, étant donc présents dans de nombreux aliments de grande consommation.

La quercétine peut atteindre jusqu'à 160 mg/kg dans les fruits rouges et jusqu'à 430 mg/Kg dans les oignons, le thé vert et la cannelle [33-35]. L'isoquercitrine est présente à des concentrations importantes dans la rhubarbe [36]. La rutine, quant à elle, est présente notamment dans les feuilles de persil, dans les oignons, dans la papaye et dans d'autres fruits comme les mûres et les myrtilles [37, 38].

1.4.2. Propriétés

Des investigations menées *in vitro* et *in vivo* ont mis en évidence un nombre considérable d'activités biologiques de la quercétine, l'isoquercitrine et la rutine, comme illustré dans le Tableau I2. Les données réunies dans ce tableau montrent leur remarquable activité anti-oxydante *in vitro* et leur potentiel thérapeutique varié.

Tableau I2 - Quelques activités biologiques de la quercétine, l'isoquercitrine et la rutine.

Flavonoïde	Activités biologiques	Ref.
QUERCÉTINE 	- Un des plus puissants capteurs d'espèces réactives oxygénées <i>in vitro</i> parmi tous les flavonoïdes - Puissant inhibiteur de la cytokine TNFα (protéine des macrophages impliquée dans des processus inflammatoires)	[39]
	- Régulation de l'hypertension artérielle - Contrôle de la prostatite chronique - Contrôle de la sarcoidose (type d'inflammation des voies respiratoires)	[40-42]
	Prévention d'épisodes de dépression et d'anxiété	[42, 43]
ISOQUERCITRINE 	Complexation de cations Al³⁺	[44]
	Résistance à l'exposition aux rayonnements UV chez les plantes des genres <i>Hyptis</i> et <i>Rheum</i> (rhubarbe)	[36, 45]
RUTINE 	Action anti-oxydante <i>in vitro</i> comparable à celle de la quercétine	[39, 46]
	Ralentissement du développement de certains cancers	[47]
	Amélioration de la mémoire spatiale chez la souris	[48]
	Amélioration d'oedèmes associés aux varices et à l'arthrite	[49]
	Action antimicrobienne contre les bactéries pathogènes <i>Pseudomonas aeruginosa</i> et <i>Klebsiella pneumoniae</i>	[50]

La solubilité de ces trois molécules dans différents types de milieux est généralement faible. Pour l'isoquercitrine la solubilité dans l'eau, même à ébullition, est négligeable. Dans des solvants organiques d'usage courant en bioprocédés, leurs solubilités sont significativement accrues par rapport à l'eau, mais restent encore très limitées. À titre d'exemple, les valeurs de solubilité obtenues expérimentalement pour ces flavonoïdes dans l'eau, le *tert*-amylalcool, l'acétone et l'acétonitrile sont données dans le Tableau I3.

Tableau 13 - Solubilité (mM) des flavonoïdes quercétine, isoquercitrine et rutine dans l'eau, le tert-amyl alcool, l'acétone et l'acétonitrile, d'après les références [51, 52].

Flavonoïde	Solvant			
	eau	tert-amyl alcool (50°)	acétone (50°C)	acétonitrile (50°C)
Quercétine	0,2 ^a	67	80	5,4
Isoquercitrine	≈ 0 ^b	66	30	3,9
Rutine	0,2 ^a	60	13	0,5

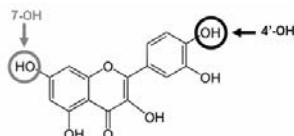
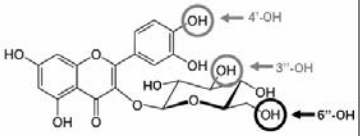
^a à 27°C ; ^b ≈ 100°C.

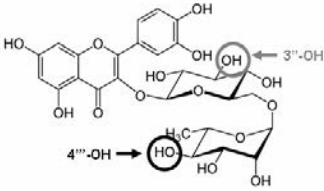
1.4.3. Acylation enzymatique

Au vu du nombre et de la variété d'activités biologiques de la quercétine, de l'isoquercitrine et de la rutine d'une part, et de leurs faibles solubilités d'autre part, il est intéressant de procéder à des modifications structurales minimales et ciblées sur ces molécules. De telles modifications permettraient d'obtenir des dérivés plus solubles et stables dans des formulations hydrophobes, tout en gardant les propriétés biologiques des flavonoïdes d'origine. Comme mentionné, ce compromis peut être atteint par des procédés régiosélectifs d'acylation, ce qui explique l'intérêt porté à l'acylation par voie enzymatique. Les conditions réactionnelles et les résultats les plus marquants des travaux portant sur l'acylation enzymatique de la quercétine, l'isoquercitrine et la rutine sont présentés de façon résumée dans le Tableau 14 (les deux pages suivantes).

L'analyse de ces données indique que l'enzyme la plus utilisée pour la biosynthèse de dérivés acylés de l'isoquercitrine et de la rutine est la lipase CALB. L'hydroxyle 6''-OH de l'isoquercitrine est le site préférentiellement acylé, bien que la formation de diesters ait été rapportée par certains auteurs ; Ainsi, les hydroxyles 4'-OH (avec PCL) et 3''-OH (avec la CALB) du monoester produit dans la première période de la réaction peuvent être acylés après des temps réactionnels élevés. La CALB ne catalyse pas l'acylation de la quercétine. Celle-ci a pu être acylée en utilisant la lipase PCL comme catalyseur, conduisant à la formation d'un monoester en position 4'-OH, lequel est ensuite converti en diester, en raison de l'acylation de l'hydroxyle 7-OH. Ces résultats indiquent que, dans les bioprocédés d'acylation de flavonoïdes catalysés par des lipases, la spécificité et la régiosélectivité sont dépendants des structures du flavonoïde et de l'enzyme. Néanmoins, aucun de ces travaux ne fournit d'explication, au niveau moléculaire, des propriétés de sélectivité et de spécificité expérimentalement observées.

Tableau I4 - Sommaire des données de régiosélectivité dans l'acylation par voie enzymatique des flavonoïdes quercétine, isoquercitrine et rutine.

Flavonoïde (F)	Donneur(s) d'acyle (D)	Conditions réactionnelles ^a	Produits (% conversion)	Ref.
QUERCÉTINE 	- Acide hexanoïque (C6:0) - Acide nanoïque (C9:0) - Acide laurique (C12:0) - Acide myristique (C14:0) - Acide palmitique (C16:0) - Acide stéarique (C18:0) - Acide oléique (C18:1)	Milieu : <i>tert</i>-amyl alcool ; enzyme : CALB ; ratio D/F = 1 ; T = 60 °C ; 150 h.	Aucun produit	[53]
	acétate de vinyle	Milieu : <i>tert</i>-amyl alcool et acetonitrile ; enzymes : CALB et PCL ; ratio D/F = 40 ; T = 50 °C ; 96 h.	- CALB : aucun produit - PCL : 4'-O-acétate et 4',7-O-diacétate (96%) ^b	[54]
ISOQUERCITRINE 	acétate de vinyle	Milieu : acétone ; enzyme : CALB ; ratio D/F = 10 ; T = 45 °C ; 60 h.	6''-O-acétate et 3'',6''-O-diacétate (77%) ^b	[55]
	cinnamate de vinyle	Milieu : acétone ; enzyme : CALB ; ratio D/F = 10 ; T = 45 °C ; 60 h.	6''-O-cinnamate (68%)	[56]
	- Acide palmitique - Acide 2-OH-phenylpropionique - Acide 4-OH-phenylpropionique	Milieu : <i>tert</i>-amyl alcool ; enzymes : CALB et PCL ; ratio D/F = 2 ; T = 50 °C ; 96 h.	6''-O-palmitate (81%) 6''-O-(2-OH-phenylpropionate) (89%) 6''-O-(4-OH-phenylpropionate) (88%)	[57]
	acétate de vinyle	Milieu : acétone ; enzymes : CALB et PCL ; ratio D/F = 40 ; T = 50 °C ; 96 h.	- CALB : 6''-O-acétate et 3'',6''-O-diacétate (≈100%) ^b - PCL : 6''-O-acétate et 4',6''-O-diacétate (≈100%) ^b	[54]

Flavonoïde (F)	Donneur(s) d'acyle (D)	Conditions réactionnelles ^a	Produits (% conversion)	Ref.
RUTINE 	acétate de vinyle	Milieu : pyridine + acétone ; enzyme : CALB ; ratio D/F = 10 ; T = 45 °C ; 45 h.	4'''-O-acétate et 3'',4'''-O-diacétate (79%) ^b	[55]
	- Acide octanoïque (C8:0) - Acide décanoïque (C10:0) - Acide laurique (C12:0)	Milieux : <i>tert</i>-amyl alcool ; enzyme : CALB ; ratio D/F = 20 mmol ; T = 45 °C, 240h.	4'''-O-octanoate (≈50%) 4'''-O-decanoate (≈50%) 4'''-O-dodecanoate (≈50%)	[58, 59]
	- Acide hexanoïque (C6:0) - Acide nanoïque (C9:0) - Acide laurique (C12:0) - Acide myristique (C14:0) - Acide palmitique (C16:0) - Acide stéarique (C18:0) - Acide oléïque (C18:1)	Milieu : <i>tert</i>-amyl alcool ; enzyme : CALB ; ratio D/F = 1 ; T = 60 °C ; 150 h.	4'''-O-hexanoate (42%) 4'''-O-nanoate (65%) 4'''-O-laureate (76%) 4'''-O-myristate (48%) 4'''-O-palmitate (72%) 4'''-O-stéarate (74%) 4'''-O-oléate (70%)	[53, 60]
	- Acide palmitique (C16:0) - Acide hydrocinnamique	Milieu : <i>tert</i>-amyl alcool ; enzyme : CALB ; ratio D/F = 1 ; T = 60 °C ; 150 h.	4'''-O-hexadecanoate (82%) 4'''-O-hydroxycinnamate (76%)	[61]
	- Acide oléïque (C18:1) - Acide linoléïque (C18:2) - Acide linolénique (C18:3) - linoate de vinyle	Milieux : acétone ; enzyme : CALB ; ratio D/F = 40 ; T = 50 °C ; 96 h.	4'''-O-oleate (70%) 4'''-O-linoleate (80%) 4'''-O-linolenate (68%) 4'''-O-linoleate (50%)	[62]
	butyrate de vinyle	Milieux : liquides ioniques (bmin)PF₆ et (bmin)BF₄ ; enzyme : CALB ; ratio D/F = 40 ; T = 60 °C ; 96 h.	4'''-O-butyrates (≈65%)	[63]

^a Dans les cas où plusieurs donneurs d'acyle et plusieurs conditions réactionnelles ont été testés, seules les conditions donnant les rendements les plus élevés sont présentées, car la régiosélectivité est la même dans les autres cas.

^b Rendement de conversion total du substrat ; les proportions entre les différents produits obtenus ne sont pas mentionnées dans les articles. Toutefois, ces travaux ont montré que la formation du diester n'arrive qu'après de longues périodes de réaction, dû à l'acylation du monoester préalablement formé.

2. LES PRINCIPALES APPROCHES DE MODÉLISATION MOLÉCULAIRE

2.1. Introduction

La modélisation moléculaire est un domaine scientifique multidisciplinaire : elle exploite des lois de la chimie, de la physique et de la biologie dans des programmes informatiques spécifiques afin de calculer structures et propriétés d'entités chimiques et biochimiques (protéines, acides nucléiques, complexes moléculaires, solides, cristaux etc.). Le but est la compréhension ou la prédiction des phénomènes auxquels s'intéressent ces disciplines [64].

Depuis les années 80, la modélisation moléculaire connaît un essor continu, non seulement comme outil scientifique précieux pour des domaines fondamentaux (tels que la chimie et la biologie) ou appliqués (tels que la pharmacie et les ingénieries), mais aussi comme une branche scientifique à part entière, consacrée à l'amélioration de la robustesse des logiciels et de l'efficacité des processeurs. Ce développement peut être attribué à deux facteurs principaux : le premier est l'évolution de l'informatique au cours des dernières décennies. Ceci a permis l'intégration de concepts de chimie théorique dans les algorithmes et le développement de machines de plus en plus puissantes, capables d'exécuter les lourds calculs numériques requis et de stocker les données générées. Le deuxième facteur est l'élucidation des structures tridimensionnelles de nombreuses protéines, grâce aux progrès de la génomique, et des techniques de cristallisation et résolution structurale par diffraction de rayons X. En effet, les structures tridimensionnelles de ces biomolécules sont le point de départ pour des simulations visant à comprendre leurs interactions, entre elles ou avec des petites molécules, ce qui est à la base d'un grand nombre de phénomènes chimiques, biologiques et biochimiques [65]. Ainsi, la modélisation moléculaire trouve de nos jours d'importantes applications, parmi lesquelles trois exemples classiques sont :

- *L'Étude des propriétés des matériaux* : à partir de simulations moléculaires, un lien peut être établi entre les résultats de calculs menés à l'échelle microscopique (sur un échantillon représentatif d'atomes en interaction) et des propriétés mesurables à l'échelle macroscopique (sur les systèmes matériels réels que l'on souhaite maîtriser). Ainsi, peuvent être étudiées les propriétés mécaniques des polymères, les propriétés électroniques de solides cristallins ou encore les propriétés thermodynamiques et spectroscopiques d'une large gamme de composés d'intérêt scientifique et technologique [66].
- *Le développement de nouveaux médicaments* : le mécanisme d'action de nombreux médicaments consiste à agir comme inhibiteur (ligand) d'une enzyme (récepteur) impliquée dans le développement de la maladie, que ce soit une protéine d'un microorganisme pathogène ou du propre organisme humain. L'élucidation de la structure tridimensionnelle de protéines impliquées dans plusieurs pathologies a permis, *via* des simulations informatiques, la découverte d'inhibiteurs puissants pour ces protéines, en réduisant considérablement le nombre d'essais de *screening* nécessaires pour aboutir à un nouveau médicament (*rational drug design*) [67-72].

- *La rationalisation de l'ingénierie enzymatique* : le fondement est le même que celui du développement de médicaments : en permettant d'étudier les interactions entre les substrats (ligand) et les enzymes (récepteurs), la modélisation moléculaire apporte une meilleure compréhension de la sélectivité enzymatique, au niveau moléculaire. Ceci peut rendre possible, par la suite, le développement d'approches prédictives pour la sélectivité des enzymes natives vis-à-vis d'un substrat ou d'une classe de substrats donnée (*rational process design*) [32, 73, 74], ou encore, d'identifier des acides aminés dont la mutation ponctuelle permettrait de modifier la sélectivité ou la spécificité enzymatiques (*site-directed mutagenesis*) [75, 76].

Dans un contexte de catalyse enzymatique, on s'intéresse dans un premier temps aux simulations de docking, qui visent à étudier les modes d'interaction entre les protéines et les ligands (inhibiteurs ou substrats). Les simulations de docking peuvent être complétées, en amont et en aval, par des simulations impliquant des calculs basés sur la mécanique moléculaire (MM), la dynamique moléculaire (MD), la mécanique quantique (QM) ou encore sur des méthodes hybrides (QM/MM). Le type de simulation à exécuter dans une situation donnée dépend directement de trois facteurs : la problématique scientifique à étudier, les ressources informatiques disponibles et le niveau de précision requis. Dans les sections qui suivent, chacune de ces approches est brièvement présentée.

2.2. La mécanique moléculaire

2.2.1. Principes théoriques

Dans les approches de mécanique moléculaire, quatre simplifications sont considérées de manière à permettre la description du système modélisé en appliquant des approximations empiriques, à partir des concepts de la mécanique classique :

- Chaque atome est considéré comme une particule sphérique, dotée d'un rayon (typiquement le rayon de van der Waals) et d'une charge électrique constante (généralement dérivés de calculs quantiques ou de déterminations expérimentales).
- Chaque liaison covalente est considérée comme un ressort, avec une distance d'équilibre soit calculée par des méthodes quantiques, soit déterminée par des méthodes expérimentales (rayons X, RMN etc.)
- Les atomes non liés peuvent interagir *via* des interactions électrostatiques et de van der Waals.
- Les propriétés des atomes dans les molécules sont transférables à d'autres molécules structuralement corrélées.

Ces simplifications permettent d'exprimer l'énergie potentielle d'un système moléculaire en fonction des paramètres géométriques de celui-ci. L'expression mathématique correspondante est connue sous le nom de « fonction de potentiel » ou « champ de forces », qui est la dénomination la plus couramment utilisée.

Parmi les champs de forces le plus connus, figurent : MMFF, CFF et CVFF (utilisés dans le traitement de petites molécules organiques et de biomolécules), et GROMOS, AMBER et CHARMM (spécialisés dans le traitement de molécules biologiques, telles que les acides nucléiques et les protéines) [77].

2.2.2. Formalisme : les champs de force

Dans l'expression fonctionnelle des champ de forces classiques, l'énergie potentielle totale du système (V) est exprimée comme étant la somme des contributions énergétiques des termes liés et de celles des termes non liés [78] :

$$V_{\text{totale}} = V_{\text{liés}} + V_{\text{non liés}} \quad (\text{Eq. 11})$$

Les termes liés correspondent aux liaisons covalentes (b), angles de valence (θ), angles de torsion (φ) et dièdres impropres (ω), aussi connus comme « torsions oop » (de l'anglais, *out of plane*), soit :

$$V_{\text{liés}} = V_{\text{liaisons covalentes}} + V_{\text{angles de valence}} + V_{\text{torsions}} + V_{\text{impropres}} \quad (\text{Eq. 12})$$

Une représentation de la signification physique de chacune de ces composantes, ainsi que leurs formes fonctionnelles, est donnée dans la Figure 17.

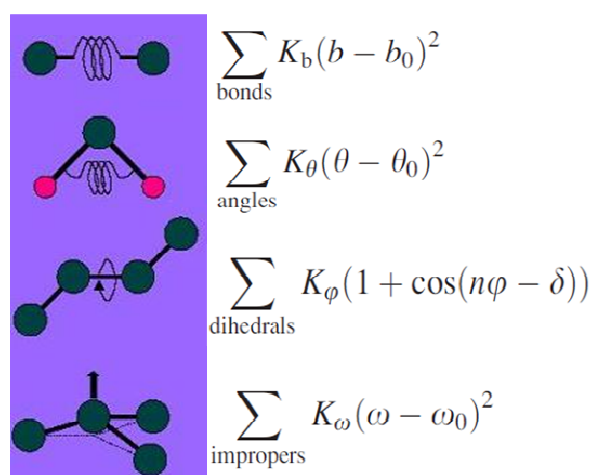


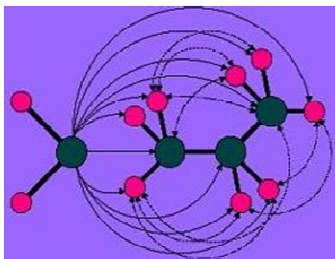
Figure 17 - Schéma et forme fonctionnelle des termes liés d'un champ de forces classique.

Les termes des liaisons covalentes (b), des angles de valence (θ) et des dièdres impropres (ω) sont généralement des potentiels harmoniques centrés sur les valeurs d'équilibre des longueurs de liaison (indexées 0 dans les équations). Les termes des angle de torsion (φ) ont souvent plusieurs minima et ne peuvent donc pas être modélisés comme des oscillateurs harmoniques simples, étant fréquemment représentés comme la somme des premiers termes d'une série de Fourier. Les valeurs d'équilibre, ainsi que les constantes de force (K_b , K_θ , K_φ et K_ω) sont ajustées pour une classe de molécules donnée, *via* des calculs *ab initio* ou des données empiriques (rayons X, FT-IR ou RMN le plus souvent) [78].

Les termes non liés, quant à eux, correspondent aux interactions de van der Waals, représentées par un terme de Lennard-Jones, et aux interactions électrostatiques, représentées par un terme de Coulomb, soit :

$$V_{\text{non liées}} = V_{\text{électrostatique}} + V_{\text{van der Waals}} \quad (\text{Eq. 13})$$

Ces contributions sont illustrées dans la Figure I8 :



$$\sum_{\text{non-bonded pairs}} \left\{ \varepsilon_{ij}^{\min} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}} \right\}$$

Figure I8 - Schéma et forme fonctionnelle des termes non liés d'un champ de forces classique.

- Le premier terme de la sommation est un potentiel de Lennard-Jones 12-6 (LJ 12-6), qui représente la contribution des interactions de van der Waals. La distance entre deux atomes i et j est représentée par r_{ij} ; $\varepsilon_{ij}^{\min}$ et $R_{ij}^{\min}$ sont respectivement les valeurs qui minimisent ce potentiel. Une variante de ce potentiel, le Lennard-Jones 9-6 (LJ 9-6, Eq. 14) est utilisée dans certains cas ; les résultats ne diffèrent pratiquement pas pour des très courtes distances, de l'ordre des distances interatomiques [79] :

$$E_{\text{vdW}} = \sum_{i,j} \varepsilon_{ij} \left[2 \left(\frac{r_{ij}^*}{r_{ij}} \right)^9 - 3 \left(\frac{r_{ij}^*}{r_{ij}} \right)^6 \right] \quad (\text{Eq. 14})$$

(ε_{ij} et r_{ij}^* ont la même signification que $\varepsilon_{ij}^{\min}$ et $R_{ij}^{\min}$, respectivement, dans l'équation du potentiel LJ 12-6, dans la Figure I8).

- Le deuxième terme est un potentiel de Coulomb, qui représente la contribution des interactions électrostatiques. Les charges atomiques partielles des atomes i et j sont représentées respectivement par q_i et q_j ; ϵ_0 est la constante diélectrique du vide et ϵ est la constante diélectrique relative du milieu (par défaut, $\epsilon = 1$ dans les simulations avec des molécules de solvant explicites).

Les termes non liés sont beaucoup plus coûteux en termes de calcul car, même si un atome n'est lié qu'à quelques atomes voisins, il interagit avec chacun des autres atomes non liés, dans la même molécule ou dans des molécules environnantes. Le plus souvent, une distance de troncature (*cut-off*

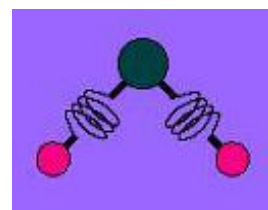
distance, en anglais) est utilisée pour alléger et accélérer ces calculs : si la distance entre deux atomes non liés est supérieure à cette distance de troncature, les contributions énergétiques de van der Waals et électrostatiques dans le champ de forces sont négligées. Certains champs de force plus « généralistes », comme le CFF par exemple, comportent des « termes croisés » qui sont des combinaisons entre les termes décrits précédemment, permettant de prendre en compte le couplage des déformations moléculaires (distance-torsion, distance-angle, angle-torsion etc.) [80].

2.2.3. Le champ de forces CHARMM

Dans ce travail, le champ de forces **CHARMM** (Chemistry at **HAR**vard **M**olecular **m**echanics) a été utilisé [81]. La performance de CHARMM a été démontrée pour la simulation de biomolécules, notamment de protéines et d'acides nucléiques, mais aussi de peptides, de lipides, de carbohydrates et de petites molécules pouvant s'associer à ces biomolécules, dans des environnements diversifiés comme des solutions, des cristaux et des membranes [82]. La forme générale de l'équation de ce champ de forces est :

$$\begin{aligned}
 V = & \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\
 & + \sum_{\text{dihedrals}} K_\phi(1 + \cos(n\phi - \delta)) \\
 & + \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} K_{UB}(S - S_0)^2 \\
 & + \sum_{\text{non-bonded pairs}} \left\{ \epsilon_{ij}^{\min} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}} \right\}
 \end{aligned} \tag{Eq. 15}$$

En plus des termes liés et non liés précédemment définis, CHARMM comporte un cinquième terme lié, appelé terme de Urey-Bradley (S). Celui-ci (schéma à droite) correspond à la contribution énergétique en fonction de la distance entre les atomes 1 et 3 (en rose) d'une séquence d'atomes liés. Son but est de mieux reproduire la contribution énergétique due aux déformations non planaires et aux étirements asymétriques observés dans les spectres infrarouge de certains acides aminés [81].



Un grand nombre de travaux employant deux autres champs de force, GROMOS [83] et AMBER [84], sont disponibles dans la littérature. Des études comparatives récentes indiquent que dans le cadre de la modélisation des protéines en général, les performances de ces deux champs de forces sont comparables à celle de CHARMM [85, 86].

2.2.4. Minimisation d'énergie (optimisation de géométrie)

Une simulation typique de mécanique moléculaire est la minimisation d'énergie, le but étant de trouver une conformation de faible énergie d'une molécule ou d'un système moléculaire. La fonction d'énergie potentielle V , possède $3N$ variables (N = nombre d'atomes du système étudié). C'est cette fonction objective $V(r_i)$ qu'il faut minimiser, où $r_i = (x_i, y_i, z_i)$ représente les coordonnées de chaque atome i du système. Formellement, la minimisation de l'énergie potentielle d'un système moléculaire consiste donc à résoudre un problème d'optimisation à plusieurs variables indépendantes : à partir des coordonnées cartésiennes des atomes dans la structure initiale (x_i, y_i, z_i) , on recherche un nouveau jeu de coordonnées optimisées (x_i^*, y_i^*, z_i^*) qui réduit à son minimum l'énergie du système, $V(x_i^*, y_i^*, z_i^*)$. Pour cette raison, la minimisation d'énergie est aussi appelée optimisation de géométrie, les deux phénomènes étant indissociables. Le principe consiste à prendre la dérivée première de l'énergie V par rapport à chacun des degrés de liberté du système et à trouver l'endroit sur son hypersurface énergétique où, pour chaque coordonnée :

$$\partial V / \partial r_i = 0 \quad (\text{Eq. 16})$$

Les algorithmes numériques qui exécutent cette tâche sont itératifs et ont, au moins, un point en commun : on commence en un endroit donné de l'hypersurface et on descend vers le minimum le plus proche, sans savoir si ce minimum est local ou absolu (Figure 19).

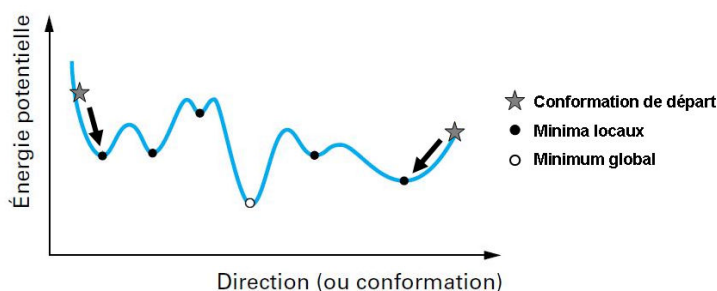


Figure 19 - Représentation schématique du principe d'une minimisation d'énergie « statique » : à partir d'un ensemble de coordonnées atomiques (conformation) de départ, l'algorithme retrouve le minimum local de la fonction d'énergie le plus proche, ainsi que la nouvelle conformation correspondant à ce minimum.

Les procédures sont de deux types : les unes utilisent uniquement la pente de la surface (dérivée première), comme les algorithmes Steepest Descent (SD) et les Conjugate Gradients (CG) ; les autres, utilisent à la fois cette pente et la courbure de la surface (les dérivées première et seconde), comme l'algorithme Adopted-basis Newton-Raphson (ABNR) [87, 88]. Il est important de noter que, étant donné le nombre important de variables de la fonction d'énergie $V(x_i, y_i, z_i)$, celle-ci présente, généralement, plusieurs minima et maxima. Ces algorithmes de minimisation ne permettent de trouver que le minimum local le plus proche de la configuration de départ, sans affranchir les barrières énergétiques.

2.3. Les simulations de docking

2.3.1. Principes théoriques

Docking (ancrage ou amarrage, en français) est le nom donné aux simulations moléculaires dans lesquelles différentes approches sont combinées pour étudier les modes d'interaction entre deux molécules. Dans la plupart des cas, il s'agit d'un récepteur macromoléculaire (cible de docking) dont la structure tridimensionnelle est connue et d'une petite molécule (ligand). Le récepteur macromoléculaire étant le plus souvent une protéine, le terme *docking* seul est couramment employé pour désigner un « docking protéine-ligand » [89]. Une simulation de docking comprend essentiellement deux étapes : le *docking* proprement dit et le *scoring*.

- La première (le *docking*) est l'étape de sélection, consistant à placer le ligand dans le site actif de la protéine et à échantillonner les conformations, positions et orientations (poses) possibles, en ne retenant que celles qui représentent les modes d'interaction les plus favorables. Bien que pouvant être faite « manuellement », cette étape est le plus souvent exécutée de manière automatisée à l'aide d'algorithmes de docking, ce qui améliore la vitesse et la précision des simulations.
- La deuxième (le *scoring*) est l'étape de classement, qui consiste à évaluer l'affinité entre le ligand et la protéine dans chacun des complexes sélectionnés lors de l'étape précédente. Cette tâche est accomplie par des outils informatiques nommées fonctions de score. Celles-ci évaluent d'abord les différentes contributions énergétiques (et parfois entropiques) pour la stabilisation du complexe protéine-ligand, et ordonnent ensuite ces complexes, permettant ainsi l'identification des modes d'interaction les plus probables [89, 90].

2.3.2. Algorithmes de docking

En principe, un docking peut être fait de façon « manuelle », directement par le modélisateur, en plaçant le ligand dans le site actif de la protéine à l'aide d'une interface graphique. Ensuite, la géométrie de l'ensemble est optimisée de manière à corriger les problèmes stériques et obtenir un complexe énergétiquement stable. Cette approche est appliquée quand on a une idée précise du mode d'interaction réel du ligand [91]. Néanmoins, le plus souvent, le mode d'interaction réel n'est pas connu. Dans ce cas, tester manuellement toutes les conformations et orientations des ligands s'avère impossible d'un point de vue pratique, même en considérant la protéine comme un corps rigide. Pour contourner cette difficulté, les algorithmes de docking ont été conçus pour rechercher de façon objective, rapide et efficace les modes d'association protéine-ligand les plus favorables [92].

Un système biologique comprend généralement au moins un ligand, un récepteur macromoléculaire et des molécules de solvant (de l'eau dans la plupart des cas). Ainsi, un algorithme idéal devrait pouvoir prendre en compte la présence du solvant et la flexibilité totale des systèmes moléculaires. Toutefois, le nombre de degrés de liberté à explorer dans une telle situation rendrait le problème intraitable par les ressources informatiques disponibles à ce jour. Des simplifications sont donc nécessaires pour rendre le problème abordable d'un point de vue pratique.

Une première simplification consiste à enlever les molécules de solvant pendant le docking. Le duo protéine-ligand possédant un nombre encore très élevé de degrés de liberté, dû à la flexibilité de la protéine, une deuxième simplification est adoptée dans une grande partie des algorithmes de docking : la non prise en compte de la flexibilité de la protéine [93-95]. Ainsi, les algorithmes de docking (Figure I10, page suivante) peuvent être séparés en deux grandes classes : ceux qui ne tiennent pas compte de la flexibilité de la protéine, en traitant celle-ci comme un corps rigide, et ceux qui sont capables de prendre en compte, partiellement ou totalement, la flexibilité du récepteur.

- Dans les procédures de docking considérant la protéine comme un corps rigide, la flexibilité du ligand seule est prise en compte pour l'obtention de complexes. Selon la méthode utilisée pour générer les conformères du ligand et les placer dans la cavité catalytique du récepteur, les algorithmes peuvent être sous-divisés en algorithmes de simulation de MM et DM, de forme, systématiques et stochastiques. Plusieurs articles de revue décrivant le principe, les points forts et les limitations de ces méthodes sont disponibles [89-92, 96-99]. Toutefois, cette classification doit être considérée avec précaution, puisque un bon nombre d'algorithmes combinent plus d'une méthode pour la génération et l'échantillonnage de conformères du ligand. Dans la plupart des cas, l'utilisation d'algorithmes considérant la protéine comme un corps rigide mène à de bons résultats, principalement quand la protéine a une flexibilité limitée. En effet, dans de tels cas, la structure cristallographique peut être considérée comme plus représentative de l'état de la protéine dans son environnement naturel, ce qui augmente les chances de simuler correctement la complexation des ligands [93-95, 100-102].
- Certaines protéines présentent naturellement des régions de grande flexibilité, subissant des réarrangements considérables en présence d'un ligand. Dans ce cas, négliger la flexibilité de la protéine peut mettre en péril la fiabilité des résultats de docking, et rend nécessaire l'utilisation d'approches capables de tenir compte de la flexibilité du système entier [100, 101]. Des méthodes indirectes ou directes, où la flexibilité de la protéine est partiellement ou totalement prise en compte, sont décrites dans la littérature. Néanmoins, ces méthodes ne sont pas souvent utilisées car le gain en précision par rapport aux algorithmes traditionnels est généralement trop petit par rapport à l'augmentation du temps de simulation [103-106].

2.3.3. Fonctions de score

La procédure de docking permet de générer une liste de complexes représentant les modes d'association favorables entre le ligand et le récepteur macromoléculaire. L'étape suivante consiste à évaluer ces complexes, afin de trouver celui ou ceux les plus susceptibles de reproduire au mieux le mode d'association réel.

L'association entre protéines et ligands est gouvernée par plusieurs paramètres thermodynamiques : les interactions hydrophobes, les interactions électrostatiques, les liaisons hydrogène, les effets de solvation et les effets d'entropie. Théoriquement, le complexe est favorable si la variation d'énergie libre globale de complexation est négative ($\Delta G_{\text{complexation}} < 0$). En

pratique, l'évaluation de l'énergie libre des complexes est une tâche souvent coûteuse d'un point de vue informatique, ce qui limite son utilisation en routine [107]. De ce fait, des méthodes approximatives ont été développées pour distinguer (évaluer et classer) les meilleurs complexes parmi ceux générés par une procédure de docking : les fonctions de score [108]. Selon les principes utilisés dans leur conception, les fonctions de score sont classées en : fonctions de score basées sur des champs de force, fonctions de score *knowledge-based* et fonctions de score empiriques [109, 110]. Les fonctions de score le plus couramment utilisées sont présentées schématiquement dans la Figure I11.

Une manière alternative de classer les complexes issus du docking est l'utilisation d'un « consensus score ». Il s'agit d'une technique de classement de complexes qui combine les résultats individuels de plusieurs fonctions de score afin de compenser les faiblesses des unes avec les forces des autres, en augmentant ainsi la fiabilité du résultat final. Le principe consiste à vérifier la fréquence à laquelle chaque complexe apparaît parmi les X complexes les mieux classés selon différentes fonctions de score (X = 5 ou 10 le plus souvent). D'un point de vue statistique, plus cette fréquence est élevée pour un certain complexe, plus la probabilité que ce complexe représente le réel mode d'interaction protéine-ligand est élevée, vu qu'il est reconnu comme tel par plusieurs fonctions de score distinctes [111, 112]. En effet, des études comparatives récentes confirment que l'application d'un consensus score augmente les chances, par rapport aux scores individuels, de retrouver par simulation un complexe identifié expérimentalement par des techniques de diffraction de rayons X [113-116].

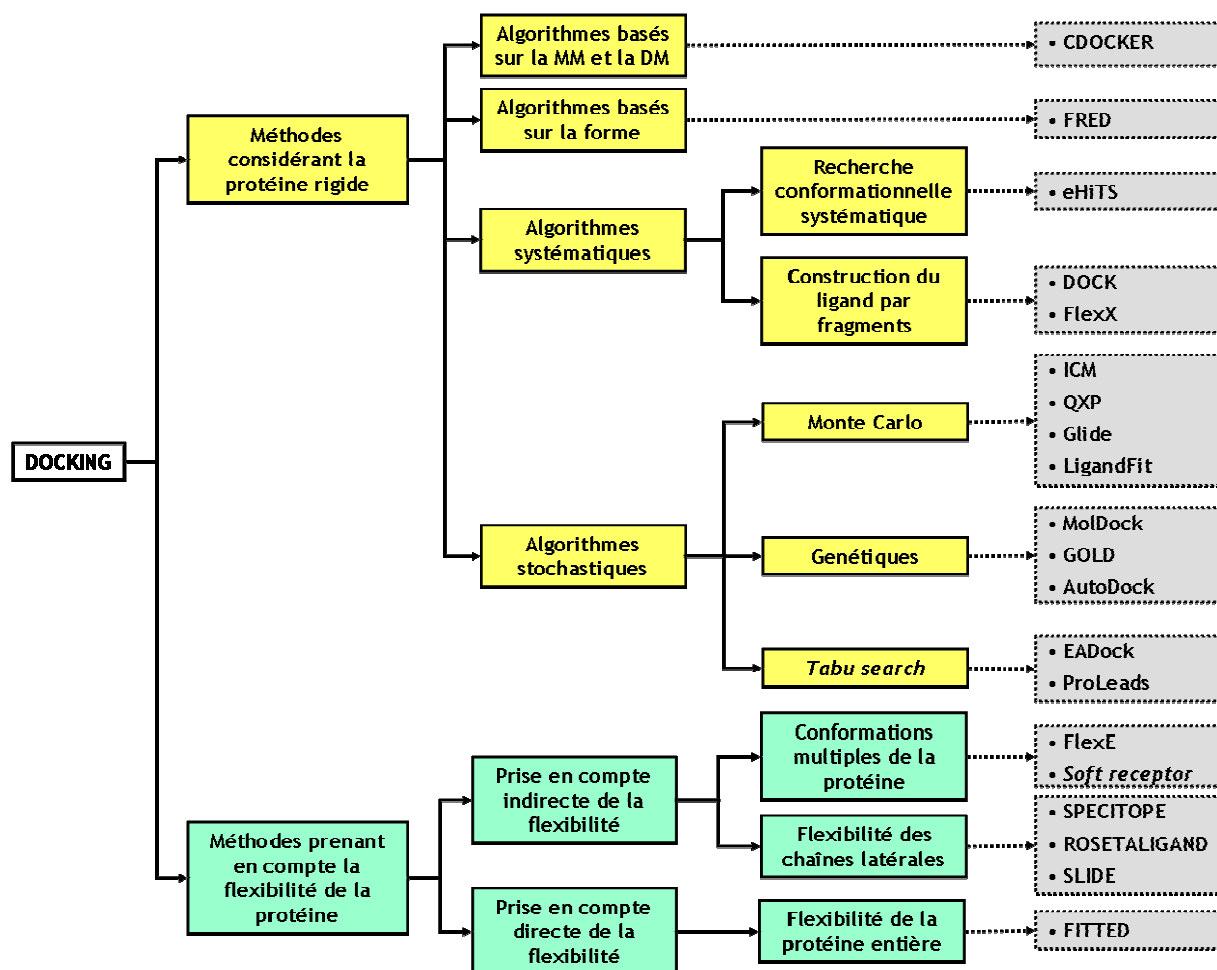


Figure I10 - Méthodes automatisées de docking les plus connues et exemples d'algorithmes (en gris) fréquemment cités dans la littérature (d'après les références [89-92, 96-99]).

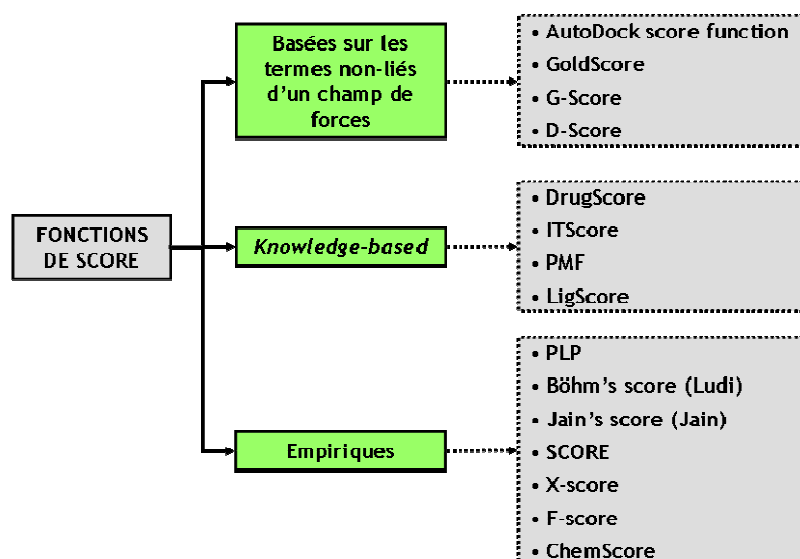


Figure I11 - Différents types de fonctions de score les plus connues et les exemples (en gris) les plus fréquemment cités dans la littérature (d'après les références [108-110]).

2.4. Les simulations de dynamique moléculaire

2.4.1. Principes théoriques

La dynamique moléculaire est une approche de simulation qui prend en compte les interactions entre atomes et molécules pendant une période de temps, à une certaine température, permettant ainsi une évaluation des mouvements atomiques [117].

La base théorique de la dynamique moléculaire est l'équation de mouvement de Newton ($F = m.a$), qui s'écrit sous la forme différentielle suivante :

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r}{dt^2} \quad (\text{Eq. 17})$$

où V est la fonction d'énergie potentielle (selon un champ de forces), m_i est la masse et r_i représente la position spatiale (x_i, y_i, z_i) de chaque atome i du système. Cette équation montre que les changements de l'énergie potentielle du système sont corrélés avec les changements de position des atomes au cours du temps.

Dans un premier temps, étant donné une configuration de départ du système moléculaire, une vitesse initiale « aléatoire » est estimée pour chaque atome, selon une fonction de distribution de probabilités du type Boltzmann (Eq. 18), qui permet de calculer la vitesse la plus probable pour chaque atome à une température T .

$$f(v) = 4\pi \left(\frac{m}{2\pi kT} \right)^{3/2} v^2 \exp \left[\frac{-mv^2}{2kT} \right] \quad (\text{Eq. 18})$$

k est la constante de Boltzmann, m est la masse de l'atome et la vitesse v est la résultante des composantes dans les trois directions spatiales x , y et z :

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2} \quad (\text{Eq. 19})$$

À partir des coordonnées spatiales initiales des atomes, de leurs vitesses ainsi estimées et de la discrétisation dt ($dt \equiv \Delta t$), un algorithme d'intégration numérique est appliqué sur l'équation de mouvement : les nouvelles coordonnées spatiales de chaque atome du système sont calculées au temps $(t + \Delta t)$. En répétant itérativement la procédure, on obtient une séquence de configurations qui donnent un aperçu (qui peut être visualisé sous la forme d'animation) de l'évolution temporelle du système. Cet ensemble de configurations est appelé « trajectoire ». Parmi les algorithmes intégrateurs les plus souvent implémentés dans les logiciels se trouvent les algorithmes de Verlet et de Verlet-Leapfrog [118].

2.4.2. Paramètres de mise en oeuvre

La mise en oeuvre des simulations de dynamique moléculaire nécessite aussi la définition d'autres paramètres : la température, le pas d'intégration (Δt), le temps de trajectoire, et l'ensemble thermodynamique utilisés [117].

- La température est un paramètre fondamental, car c'est en fonction de celle-ci que les vitesses initiales (v) des atomes seront établies, le plus souvent déterminées de manière « aléatoire » selon une distribution de probabilités de Maxwell-Boltzmann (Eq. I8, ci-dessus).
- Le pas d'intégration (Δt) doit être suffisamment petit pour éviter les erreurs dues à la discrétisation de l'équation de mouvement. Typiquement, on adopte $\Delta t = 1$ femtoseconde (10^{-15} s). Cette valeur correspond à la fréquence d'élongation la plus rapide des liaisons covalentes dans les systèmes moléculaires courants : celles impliquant un atome d'hydrogène.
- Le temps de trajectoire doit être compatible avec les ressources informatiques et le système modélisé : il doit être suffisamment court pour que les calculs puissent se faire dans un délai de temps raisonnable, mais suffisamment long pour mettre en évidence de manière réaliste le phénomène à étudier. Par exemple, des trajectoires de l'ordre de la nanoseconde sont généralement suffisantes pour étudier le comportement d'un ligand dans le site actif d'une protéine, tandis que des changements conformationnels d'une chaîne protéique ne peuvent être observés qu'après des temps de l'ordre de la milliseconde.
- L'utilisation d'ensembles thermodynamiques dans les simulations permet d'imposer des conditions aux variables macroscopiques du système (température, pression et volume) de manière à ce que les conditions de simulation reproduisent au mieux les conditions expérimentales correspondantes. Les trois ensembles thermodynamiques disponibles sont :
 - *NVT (nombre d'atomes, volume et température constants)* : aussi appelé « ensemble canonique ». Dans ce cas, l'énergie des processus endothermiques ou exothermiques est échangée avec un thermostat de manière à ce que la température soit maintenue constante.
 - *NVE (nombre d'atomes, volume et énergie constants)* : aussi appelé « ensemble micro canonique ». La pression et la température ne sont pas contrôlées. Ceci correspondant à un processus adiabatique sans échange de chaleur.
 - *NPT (nombre d'atomes, pression et température constants)* : il s'agit de la condition de simulation la plus proche des conditions expérimentales, correspondant à un flacon de laboratoire ouvert, soumis à pression atmosphérique et température ambiante. L'ensemble NPT ne peut être utilisé que pour des systèmes solvatés dans une boîte sous conditions périodiques.

Lorsqu'un système est soumis à des trajectoires de dynamique moléculaire dans le vide, les molécules de solvant peuvent s'évaporer, en s'écartant ainsi les unes des autres. Afin d'éviter ce phénomène, des « conditions périodiques », ou « des conditions de frontière » (PBC, de l'anglais *Periodic Boundary Conditions*) peuvent être utilisées. Le principe des PBC consiste à reproduire dans l'espace tridimensionnel la maille unitaire de simulation, qui contient une ou plusieurs macromolécules entourées de molécules de solvant. Quand une molécule de solvant traverse la face d'une cellule, elle réapparaît dans la face opposée avec la même vitesse et n'est donc pas « perdue » (Figure I12) [119]. Chaque copie de la cellule est appelée simplement « image » et les

molécules qu'elle contient, « molécules fantômes ». Théoriquement on peut reproduire la maille centrale indéfiniment dans les trois directions de l'espace. Néanmoins, sa reproduction en triplicata dans chacune des trois directions représente le cas le plus fréquent.

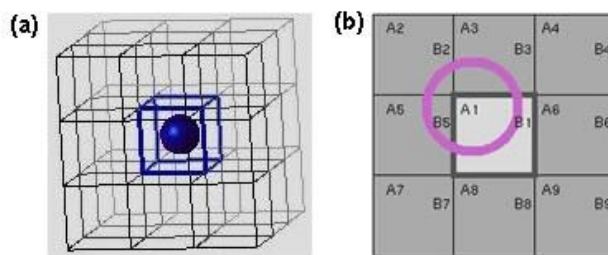


Figure I12 - (a) Une molécule de soluté (sphère bleue) entourée par une cellule cubique de solvant répliquée en trois dimensions. (b) Vue schématique du système, montrant que chaque molécule interagit avec, au maximum, une image de chaque molécule « réelle » (B1 et B5 étant la même molécule, A1 ne doit interagir qu'avec l'une des deux) [119].

2.5. Les simulations de chimie quantique

2.5.1. Principes théoriques

La chimie quantique est l'application des fondements de la mécanique quantique aux systèmes atomiques et moléculaires. L'idée de base de cette branche de la chimie est la description d'un atome comme étant constitué par un noyau chargé positivement, autour duquel les électrons se déplacent comme des nuages (Figure I13).

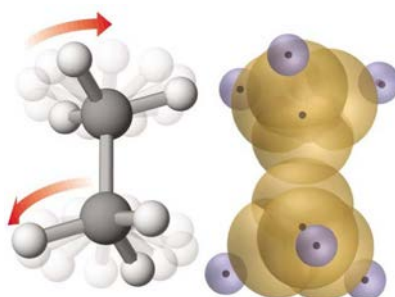


Figure I13 - Une molécule d'éthane représentée en bâtons et sphères (à gauche) et en orbitales (à droite). Dans la représentation orbitale, les points foncés représentent les noyaux atomiques et les zones translucides l'espace autour des noyaux où les électrons sont (dé)localisés.

Le formalisme théorique repose sur deux axiomes :

- Le premier axiome de la théorie quantique stipule que tout état d'un système de N particules peut être décrit par une fonction mathématique ψ (fonction d'onde), qui est dépendante des coordonnées de toutes les particules. Bien que la fonction ψ ne possède pas de signification physique concrète, la probabilité de la présence des particules dans un élément de volume (l'orbitale) est corrélée à la valeur de $|\psi|^2$.

- Le deuxième axiome établit que l'application d'un opérateur mathématique sur la fonction d'onde ψ permet d'atteindre la grandeur physique observable correspondante. L'opérateur associé à l'énergie E du système est l'opérateur Hamiltonien H . Ces trois éléments sont reliés entre eux selon l'équation proposée par Erwin Schrödinger [120] :

$$H\psi = E\psi \quad (\text{Eq. I10})$$

En théorie, tout système chimique peut être décrit en se basant sur la mécanique quantique pure. Néanmoins, la complexité de l'équation de Schrödinger augmente exponentiellement avec le nombre de particules du système. Ainsi, en pratique, seuls quelques systèmes modèles très simples, ne contenant que quelques atomes, peuvent être investigués analytiquement. Avec les ressources informatiques existantes à ce jour, des simplifications sont nécessaires pour traiter des systèmes moléculaires de taille plus importante, plus réalistes [121, 122].

La première de ces simplifications consiste à considérer l'équation de Schrödinger comme non relativiste, indépendante du temps. Dans ce cas, l'Hamiltonien H peut être écrit sous la forme :

$$H = -\sum_i^N \frac{\hbar^2}{2m} \nabla_i^2 - \sum_I^A \frac{\hbar^2}{2M} \nabla_I^2 - \sum_{i,I} \frac{Z_I e^2}{|\vec{r}_i - \vec{R}_I|} + \sum_{i<j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} + \sum_{I<J} \frac{Z_I Z_J e^2}{|\vec{R}_I - \vec{R}_J|} \quad (\text{Eq. I11})$$

Les deux premiers termes de l'Hamiltonien sont respectivement les opérateurs d'énergie cinétique des N électrons (indexés i) et des A noyaux atomiques (indexés I). Les trois autres termes représentent les différents potentiels d'interaction électron-noyau, électron-électron et noyau-noyau.

La deuxième simplification est l'approximation de Born-Oppenheimer (BO). Le principe repose sur le fait que, la masse des électrons m_e étant négligeable par rapport à celle des noyaux atomiques m_N ($m_N/m_e \approx 1836$), les positions spatiales de ces derniers sont considérées comme constantes. Par conséquent, l'énergie cinétique des noyaux est négligeable ; seul le mouvement des électrons est pris en compte. Le terme d'interaction noyau-noyau peut être approximé par une constante (E_{II}). L'opérateur Hamiltonien devient alors :

$$H = -\sum_i^N \frac{\hbar^2}{2m} \nabla_i^2 - \sum_{i,I} \frac{Z_I e^2}{|\vec{r}_i - \vec{R}_I|} + \sum_{i<j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} + E_{II} \quad (\text{Eq. I12})$$

Grâce à ces simplifications, des méthodes numériques plus facilement applicables ont pu être développées. Ces méthodes permettent de trouver des solutions approximées de l'équation de Schrödinger pour un grand nombre de systèmes moléculaires, en élargissant ainsi les applications de la chimie quantique en modélisation moléculaire.

2.5.2. Méthodes classiques de chimie quantique

L'utilisation d'approximations numériques a permis le développement et l'implémentation de méthodes numériques qui ont longtemps servi à la détermination de solutions approximées de l'équation de Schrödinger pour des systèmes contenant jusqu'à quelques dizaines d'atomes.

Dans une grande partie de ces cas, les solutions obtenues avec ces méthodes se sont avérées suffisantes pour décrire le comportement électronique de tels systèmes avec une précision satisfaisante [122, 123]. Les méthodes traditionnellement utilisées peuvent être divisées en deux catégories : les méthodes *ab initio* et les méthodes semi empiriques.

- L'expression latine *ab initio* signifie littéralement « depuis le début », « depuis le principe » ; c'est-à-dire, que les méthodes *ab initio* utilisent des approximations des fondements théoriques primordiaux de la mécanique quantique, mais sans inclure des paramètres empiriques ou semi empiriques dans ses équations. La méthode *ab initio* la plus connue est celle de Hartree-Fock (HF) [124]. Leur lourdeur en termes de calculs est telle que l'application de ces méthodes reste limitée à des systèmes de taille restreinte, nécessitant des processeurs puissants et des délais parfois déraisonnables pour l'exécution des calculs. Ces inconvénients les rendent peu viables pour les applications courantes.
- Les méthodes semi empiriques se basent sur le formalisme de la méthode HF mais, contrairement aux méthodes *ab initio*, utilisent des paramètres ajustés sur des données obtenues empiriquement (d'où leur dénomination). Les méthodes semi empiriques sont classées en deux catégories : celles qui ne traitent explicitement que les électrons de valence, dont les plus couramment utilisées sont MNDO, AM1 et PM3, et celles qui ne traitent explicitement que les électrons π , comme la méthode de Pariser-Parr-Pople (PPP), utilisée dans des calculs impliquant des états excités de polyènes. L'intégration de paramètres empiriques dans ces algorithmes a pour conséquence un allègement des calculs par rapport aux méthodes *ab initio* permettant ainsi le traitement de systèmes contenant un nombre plus élevé d'atomes (parfois plus d'une centaine), dans des délais plus raisonnables. Néanmoins, si les molécules traitées ne sont pas similaires à celles utilisées dans la paramétrisation empirique de la méthode, les solutions obtenues peuvent différer significativement de la solution exacte [125].

2.5.3. Méthode de la théorie fonctionnelle de la densité (DFT)

De nos jours, la Théorie Fonctionnelle de la Densité (Density Functionnal Theory, DFT) constitue l'une des méthodes largement utilisée pour les calculs quantiques de la structure électronique de la matière, aussi bien en physique de la matière condensée (propriétés des matériaux) qu'en chimie quantique (réactivité chimique). Les méthodes quantiques classiques (HF) se fondent sur une fonction d'onde multiélectronique (ψ). Le principal atout de la DFT est de remplacer la fonction d'onde ψ par la densité électronique ρ comme grandeur de base pour la description du système. La densité électronique $\rho(r)$, pour un état électronique donné d'un système à N électrons, est définie comme étant le nombre d'électrons par unité de volume au point de l'espace de coordonnées $r = (x,y,z)$ [126]. Alors que la fonction d'onde ψ dépend de $3N$ variables, la

densité électronique n'est fonction que de trois variables (les coordonnées x,y,z) ; elle est donc plus facile à appréhender conceptuellement et à traiter mathématiquement. Ainsi, la littérature est assez riche d'exemples de systèmes comprenant jusqu'à à environ 200 atomes et plusieurs milliers d'électrons de valence, étudiés de manière efficace par l'approche DFT, moyennant des processeurs relativement modestes. Les applications incluent la prévision de spectres infra-rouge de composés, de paramètres géométriques et électriques de petites molécules des cristaux et l'estimation des variations d'énergie au cours d'une réaction chimique [126, 127].

2.5.3.1. Formalisme : l'équation de Kohn-Sham

Le formalisme mathématique de cette théorie repose sur deux théorèmes établis par de Pierre Hohenberg et Walter Kohn [126]. Ce dernier a été honoré par le prix Nobel de Chimie en 1998, pour ses contributions décisives au développement de la DFT.

- *Premier théorème de Hohenberg-Kohn* : ce théorème établit que, à l'état fondamental, le potentiel externe V_{ext} est complètement déterminé par la densité électronique (ρ). Dans le cas d'une molécule, le potentiel externe est le potentiel d'attraction électrons-noyaux, qui définit de façon complète la nature du système polyélectronique. Une conséquence importante de ce théorème est que toute observable moléculaire découle de la connaissance de la densité électronique, car elle détermine l'Hamiltonien du système. En particulier, l'énergie électronique est une fonctionnelle de la densité.
- *Deuxième théorème de Hohenberg-Kohn* : celui-ci stipule que, pour tout système approché de l'état fondamental, la valeur de l'énergie E est supérieure à l'énergie exacte E du système. En d'autres termes, si le système est à l'état fondamental, la fonctionnelle énergie est minimale pour la distribution spatiale de la densité électronique

Plus tard, à partir de ces théorèmes, Walter Kohn et Lu Sham [126] ont montré que cette énergie électronique E de l'état fondamental d'un système réel est la somme de différentes contributions, prenant la forme suivante :

$$E = -\frac{1}{2} \sum_k \int \phi_k(\vec{r}) \nabla^2 \phi_k(\vec{r}) d\vec{r} - \sum_i \int \frac{Z_i}{|\vec{R}_i - \vec{r}|} \rho(\vec{r}) d\vec{r} + \frac{1}{2} \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + E_{xc}[\rho] \quad (\text{Eq. I13})$$

où ϕ_k représente les fonctions d'espace (qui déterminent les orbitales atomiques). Le premier terme de l'équation est l'énergie cinétique des électrons, le deuxième est l'énergie d'interaction électrons-noyaux, le troisième est l'énergie de répulsion interélectronique, et le dernier terme, $E_{xc}(\rho)$, représente l'énergie d'échange et corrélation, qui regroupe toutes les interactions électroniques non classiques. L'application du principe variationnel par rapport aux fonctions ϕ_k mène aux équations de Kohn-Sham (KS) :

$$\left(-\frac{1}{2} \Delta + \sum_i \frac{Z_i}{|\vec{R}_i - \vec{r}|} + \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' + v_{xc}[\rho] \right) \phi_k(\vec{r}) = \varepsilon_k \phi_k(\vec{r}) \quad (\text{Eq. I14})$$

où $V_{xc}(\rho)$ représente le potentiel d'échange-corrélation, dérivée de la fonctionnelle $E_{xc}(\rho)$, dont l'expression analytique pour les systèmes polyélectroniques est impossible à déterminer. Ainsi, afin de rendre la théorie applicable, il est nécessaire (2) d'adopter une fonctionnelle adéquate pour l'approximation du terme d'échange-corrélation et (1) de choisir les fonctions de base ϕ_k centrées sur les différents noyaux, dont les combinaisons linéaires permettront de construire les orbitales atomiques.

2.5.3.2. Les fonctionnelles d'échange-corrélation

Trois générations de fonctionnelles existent pour l'approximation de ce terme : les Approximations de la Densité Locale, les Approximations du Gradient Généralisé et les fonctionnelles hybrides [127].

- *Approximations de la Densité Locale (LDA, de l'anglais Local Density Approximation)* : ces fonctionnelles ne dépendent que de la densité en chaque point du système, en négligeant toute influence de l'inhomogénéité de la densité électronique de celui-ci et ne sont plus guère utilisées de nos jours. Elles donnent des résultats corrects dans les cas où la densité varie lentement au long du système étudié. Un des exemples les plus connus est la fonctionnelle VWN.
- *Approximations du Gradient Généralisé (GGA, de l'anglais Generalized Gradient Approximation)* : ces fonctionnelles introduisent dans leur expression un terme prenant en compte le gradient de la densité électronique, permettant de tenir compte de la non homogénéité de la distribution électronique. Les fonctionnelles les plus courantes appartenant à cette classe sont P86, B88, PW91 et LYP.
- *Fonctionnelles hybrides* : celles-ci incluent un terme pour l'énergie de corrélation basée sur la DFT et un terme pour l'énergie d'échange basé sur les formalismes HF et DFT (d'où leur dénomination). La fonctionnelle B3LYP, l'une des plus citées dans la littérature, est un exemple de fonctionnelle hybride.

2.5.3.3. Les fonctions de base (orbitales atomiques)

Mathématiquement, les fonctions de base représentant les orbitales peuvent être classées en deux catégories : les fonctions de Slater et les fonctions Gaussiennes [122, 123].

- *Fonctions de Slater (STO, Slater Type Orbital)* : ces fonctions sont bien adaptées pour des calculs atomiques car elles ont un bon comportement à très courtes et très longues distances du noyau. Néanmoins, elles restent peu utilisées dans l'étude de systèmes moléculaires, dû à la complexité du calcul des intégrales contenant un terme du type $\exp(-\zeta r)$, où ζ est l'exposant de Slater (nombre positif qui dépend du numéro atomique Z et qui peut être déterminé à l'aide des règles empiriques de Slater ou par optimisation).

- Fonctions gaussiennes (GTO, *Gaussian Type Orbital*) : pour alléger les calculs, une option est de remplacer les fonctions de Slater par une combinaison adéquate de fonctions Gaussiennes, qui engendrent un terme du type $\exp(-\alpha r^2)$, où α est un nombre positif. D'une part, ces fonctions rendent les calculs plus faciles et donc plus rapides. D'autre part, elles possèdent un comportement physique peu satisfaisant très près du noyau, car leur dérivée est nulle en ce point, et à longue distance, elles décroissent trop abruptement (Figure I14).

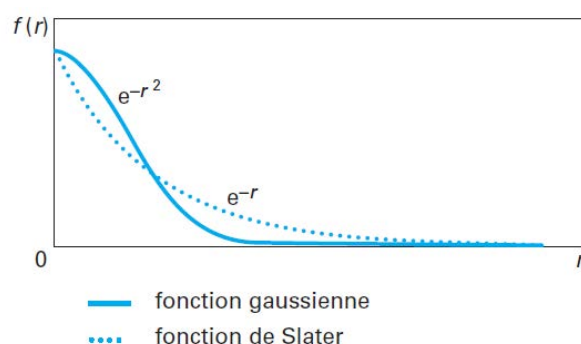


Figure I14 - Représentation graphique d'une fonction Gaussienne et d'une fonction de Slater, illustrant comment varient les termes exponentiels en fonction de la distance (r) au noyau [122].

Ces bases atomiques ne tiennent pas compte du fait que dans la molécule les atomes présentent une dissymétrie du nuage électronique par rapport aux noyaux et subissent des distorsions dues à l'environnement. Ces limitations peuvent être minimisées par l'inclusion des fonctions diffuses (représentées par +) et de polarisation (représentées par *). Par exemple, utiliser une base **6-31+G**** signifie que les **électrons de coeur** sont décrits à l'aide d'une combinaison linéaire de 6 Gaussiennes, les **électrons de valence** à l'aide de deux fonctions (une combinaison linéaire de 3 Gaussiennes plus une autre constituée d'une seule Gaussienne). De plus, chaque atome porte une **fonction de polarisation** (+), les atomes lourds portant également deux **fonctions diffuses** (**).

Les simulations basées sur la DFT sont considérées actuellement comme des méthodes *ab initio* par la plupart des scientifiques. En effet, les théorèmes de Hohenberg et Kohn ainsi que le développement amenant aux équations de Kohn-Sham sont fondamentaux, obtenus sans avoir recours à des approximations. Cependant, la fonctionnelle d'échange-corrélation qui apparaît dans les équations rend impossible toute résolution exacte car sa forme analytique est inconnue. Ainsi, il est nécessaire d'approximer cette fonctionnelle, soit en formulant une forme mathématique approchée, ou en y injectant un certain nombre de données expérimentales. Cette approche étant typique des méthodes semi-empiriques, la méthode DFT pourrait donc tout aussi bien être classée dans cette catégorie. Mais, vu que l'énergie totale en DFT est exprimée en termes dépendant de la densité électronique plutôt qu'en termes de fonctions d'onde, il semble plus pertinent de considérer la méthode DFT comme une classe à part dans les méthodes numériques de chimie quantique.

2.6. Les méthodes hybrides de simulation (QM/MM)

On entend par méthodes hybrides (qui ne doivent pas être confondues avec les *fonctionnelles hybrides* en DFT) des approches combinant un traitement classique par champ de forces et un traitement quantique au sein d'un même système. Le but général d'une telle approche est le traitement de modèles atomistiques de grande taille en trouvant le meilleur compromis possible entre précision et effort de calcul. Le traitement quantique est réservé à un « coeur », tandis que les parties restantes sont traitées par un champ de force classique. On utilise souvent à cet égard l'abréviation QM/MM (*Quantum Mechanics/Molecular Mechanics*) [128-130]. L'énergie totale d'un système traité par une méthode QM/MM est donnée par :

$$E_{totale} = E_{QM} + E_{MM} + E_{QM/MM} \quad (\text{Eq. 115})$$

où E_{QM} et E_{MM} sont les énergies des sous-ensembles MM et QM, respectivement, et $E_{QM/MM}$ est le terme d'interaction entre les deux sous-ensembles QM et MM. Ce terme doit prendre en compte la polarisation des atomes du sous-ensemble QM par les atomes de la région MM. C'est ce phénomène qui rend délicate la paramétrisation des méthodes QM/MM, nécessitant des méthodes spécifiques pour le traitement de la frontière entre les régions QM et MM [131].

Les méthodes hybrides du type QM/MM sont un thème de recherche ouvert et d'actualité car elles offrent l'avantage d'accroître sensiblement la taille des modèles moléculaires, tout en conservant toute la précision des méthodes quantiques (classiques ou DFT) dans les zones critiques des systèmes étudiés.

2.7. Applications dans l'étude des systèmes du type protéine-ligand

Bien qu'utiles pour la compréhension de phénomènes d'interaction protéine-ligand dont l'investigation expérimentale est très difficile (voire impossible), les simulations de docking présentent encore quelques limitations. Les principales sont les difficultés à prendre en compte : la flexibilité des protéines, la présence des molécules de solvant dans le milieu et la contribution entropique pour la stabilisation des complexes [99]. Dans certains cas, ces limitations pourraient nuire à la précision des modèles, menant à des mauvaises interprétations des phénomènes étudiés. Afin de pallier ces limitations, une stratégie souvent adoptée consiste à associer à la procédure de docking-scoring des simulations de basées sur des champs de forces (mécanique moléculaire/dynamique moléculaire), sur la mécanique quantique ou des simulations hybrides (QM/MM).

2.7.1. Simulations de mécanique et dynamique moléculaires

Au vu de la forme fonctionnelle mathématiquement simple des champs de force, des simulations de mécanique moléculaire (minimisations d'énergie) et de dynamique moléculaire sont utilisées pour modéliser des systèmes macromoléculaires (protéine, ADN, polymères) et des milieux avec des molécules de solvant explicites, contenant jusqu'à plusieurs milliers d'atomes au total [117]. Si les paramètres du champ de force sont correctement adaptés pour la classe de molécules

étudiée et les paramètres utilisés dans simulations sont plausibles, les résultats obtenus sont fiables. Sans être exhaustif, deux situations typiques impliquant des minimisations d'énergie et la dynamique moléculaire dans les études de catalyse enzymatique sont :

- *Pour la relaxation de la structure d'une protéine* : d'abord, en partant d'une structure cristalline de la protéine, souvent obtenue par diffraction de rayons X, des minimisations d'énergie du système sont effectuées afin trouver une configuration protéique optimale en milieu solvant. Ensuite, à partir de cette structure ainsi optimisée, des simulations de dynamique moléculaire sont appliquées pour étudier l'évolution de la structure de la protéine dans ce nouvel environnement sur une période de temps, identifier les séquences plus flexibles et plus rigides, ou encore identifier les régions interagissant plus ou moins favorablement avec le solvant [132].
- *Pour l'évaluation de l'évolution temporelle d'un complexe enzyme-substrat* : ces simulations permettent d'évaluer les changements structuraux éventuels sur la structure de la protéine, dûs à la présence du ligand (*induced fit effect*) ou du solvant [133]. En effet, ces phénomènes ne sont pas pris en compte par les algorithmes de docking les plus courants, qui traitent la protéine-cible comme une entité rigide et en absence de molécules de solvant explicites.

2.7.2. Simulations de mécanique quantique

Dû à la lourdeur des calculs numériques nécessaires à la mise en œuvre des simulations quantiques, seules des atomes directement impliqués dans la réaction, des substrats et des résidus du site actif de l'enzyme, peuvent être pris en compte dans l'étude des réactions enzymatiques par ces méthodes. Ainsi, la mise en œuvre de simulations basées sur une approche de chimie quantique se justifie lorsqu'on souhaite [134, 135] :

- *Étudier, au niveau électronique, la rupture et la formation de liaisons chimiques au cours du processus catalytique* : ceci permet une appréciation de la réactivité d'un substrat donné, avec une enzyme donnée. À partir d'un système représentatif des résidus catalytiques et des substrats dans le complexe, on cherche à déterminer des états stationnaires (minima locaux dans le profil énergétique de la réaction), correspondant aux réactifs, aux intermédiaires et aux produits.
- *Rechercher des états de transition (maximum d'énergie entre deux états stationnaires)* : la détermination d'un état de transition nécessite la connaissance préalable des structures optimisées des intermédiaires qui l'encadrent. Si l'on connaît les états stationnaires et les états de transition, la proposition de mécanismes réactionnels plausibles devient possible. La recherche d'états de transition est une procédure souvent laborieuse, comprenant la recherche, l'affinement et la vérification ; chaque étape impliquant des algorithmes spécifiques. Alors qu'un état stationnaire est caractérisé par $3N-6$ modes normaux de fréquence positive, un état de transition se caractérise par $3N-5$ modes normaux de fréquence positive + 1 fréquence imaginaire, qui n'est pas expérimentalement observable.

2.7.3. Simulations hybrides (QM/MM)

En catalyse enzymatique, les méthodes QM/MM permettent d'étudier, au cours d'une même simulation, les interactions intermoléculaires stabilisant les substrats dans cavité catalytique de l'enzyme et le processus catalytique en soi (la rupture et de la formation de liaisons dû aux interactions correspondant à la réaction chimique). Dans de tels cas, le substrat et le site réactif de la protéine sont traités par une méthode quantique, tandis que les autres parties de la protéine et les molécules de solvant sont modélisées *via* un champ de forces [74, 130]. Cette situation est illustrée dans la Figure I15.

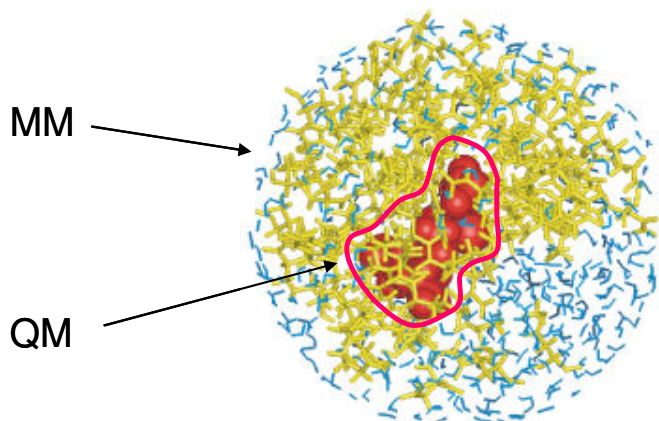


Figure I15 : Schéma d'un complexe protéine-ligand traité par une du type QM/MM. la partie du système traitée par mécanique quantique est désignée par QM, incluant tout atome de substrat et des résidus catalytiques intervenant directement dans le mécanisme réactionnel. Le reste de la protéine et le solvant, désignés par MM, sont traités par un champ de force standard. La zone d'interaction entre les deux parties est représentée par une ligne rouge [74].

3. APPLICATIONS DE LA MODELISATION MOLECULAIRE POUR LA COMPREHENSION DES INTERACTIONS PROTEINES-FLAVONOIDES

Les flavonoïdes sont des molécules hautement fonctionnalisées, possédant dans leur structure des cycles aromatiques et des fonctions hydroxyle (-OH) et carbonyle (-C=O). Grâce à ces éléments structuraux, ces molécules peuvent établir différents types d'interactions, avec un large éventail d'autres molécules. En particulier, les interactions que peuvent établir les flavonoïdes avec certaines protéines expliquent, sur un plan moléculaire, une grande partie des activités biologiques de ces polyphénols. Pour la compréhension de ces activités, un élément-clé est la connaissance des interactions susceptibles de se former entre les acides aminés du site actif des protéines et les flavonoïdes, que ceux-ci soient inhibiteurs ou substrats.

Dans la section suivante, les interactions intermoléculaires susceptibles de s'établir entre les flavonoïdes et les acides aminés sont présentées. Ensuite, à l'aide de quelques exemples récents, sont discutés les apports de la modélisation moléculaire pour l'étude de la bioconversion des flavonoïdes et pour la compréhension de leur pouvoir inhibiteur d'enzyme.

3.1. Interactions entre les flavonoïdes et les acides aminés

Cordoniu-Hernandez *et al.* [136] ont appliqué des calculs quantiques semi empiriques (AM1 et PM3) pour caractériser individuellement les interactions entre deux flavonoïdes modèles (la catéchine et la robinétinédine) et chacun des 20 acides aminés naturels connus. En analysant les valeurs d'énergie, de distances et d'angles d'interaction par la méthode de l'hypersurface des minima multiples (MMH), ces auteurs ont proposé un ordre théorique d'interaction (TAO, de l'anglais *theoretical affinity order*):

Lys, Glu, Asp > His, Arg > Trp, Leu, Gln, Asn, Phe > les autres acides aminés

Dans une étude postérieure [137], ce TAO a été confronté à des données expérimentales directes (modes d'interaction identifiés dans plusieurs complexes protéines-flavonoïdes de la PDB) et indirectes (mesure du pouvoir inhibiteur d'enzymes par différents flavonoïdes). Ainsi, la présence d'acides aminés hydrophyles et aromatiques dans le site actif des enzymes a pu être corrélée à la capacité des flavonoïdes à former des complexes avec celles-ci.

En effet, ces résultats sont en accord avec les informations plus générales relatives aux interactions intermoléculaires disponibles dans la littérature: des liaisons hydrogène ou d'autres interactions attractives de nature électrostatique sont susceptibles d'être formées entre les -OH et les -C=O des flavonoïdes et les acides aminés hydrophiles Lys, Glu, Asp, His, Arg, Gln, Asn [138]. Des interactions aromatiques de type σ - π , π - π , OH- π , ou NH- π peuvent aussi s'établir entre les cycles aromatiques présents dans les molécules des flavonoïdes et les acides aminés à chaîne latérale aromatique (Trp, Phe) [139, 140]. Par ailleurs, les acides aminés aliphatiques hydrophobes peuvent éventuellement contribuer à la stabilisation des flavonoïdes, en formant des interactions plus faibles du type CH- π avec les centres aromatiques [138, 141].

3.2. Bioconversion enzymatique de flavonoïdes

Des outils de modélisation moléculaire ont été employés pour étudier les événements moléculaires impliqués dans la bioconversion des flavonoïdes catalysée par des enzymes. Dans de tels cas, l'identification des interactions stabilisant le complexe enzyme-substrat peuvent être identifiées par des simulations de docking, souvent combinées avec des simulations de dynamique moléculaire. Dans quelques cas, des investigations plus poussées sur le mécanisme réactionnel sont réalisées à l'aide de calculs quantiques. Des exemples de travaux portant sur la O-méthylation et l'oxygénolyse enzymatique de flavonoïdes sont décrits ci-après.

3.2.1. La O-méthylation de flavonoïdes

Quelques études appliquant des simulations moléculaires à l'étude de la O-méthylation de flavonoïdes catalysée par des O-méthyltransférases (OMT) sont disponibles. Cette réaction repose sur le transfert régiosélectif du groupement méthyle du co-facteur S-adenosyl-L-méthionine (SAM) vers un groupement hydroxyle du flavonoïde. Son intérêt réside dans le fait que les dérivés méthylés sont plus stables et ont une activité microbienne accrue par rapport aux flavonoïdes d'origine [142, 143].

- Park *et al.* [142] ont étudié la méthylation de la quercétine par une O-méthyltransférase du soja (SOMT-9), qui catalyse la méthylation sur le 3'-OH. La mutation du résidu Tyr207 par une glycine (Y207G) change la régiosélectivité, menant à la méthylation des hydroxyles 3'-OH et 4'-OH. Les simulations de docking ont permis l'identification des bases moléculaires à l'origine de ce changement : pour la SOMT-9 native, la quercétine est stabilisée dans le site actif par des interactions hydrophobes formées avec les résidus Met60 et Tyr207. Le groupement 3'-OH est placé entre le co-facteur SAM et l'ion Mg^{2+} . Pour le mutant Y207G, le remplacement de la tyrosine par la glycine, un acide aminé moins volumineux, mène à une augmentation de la taille de la cavité, permettant une rotation de la quercétine. Dans ce cas, les deux hydroxyles 3'-OH et 4'-OH peuvent accéder également au centre catalytique de l'enzyme (Figure I16, page suivante).
- Kim *et al.* [143] se sont intéressés à la méthylation de la lutéoline par une O-méthyltransférase produite par le peuplier (POMT-7). Expérimentalement, ces auteurs ont constaté que cette enzyme catalyse la méthylation de l'hydroxyle 7-OH de la lutéoline. Quand la 3'-méthoxy-lutéoline (préalablement obtenue par l'action d'une autre OMT) a été utilisée comme substrat pour la POMT-7, les hydroxyles 7-OH ou 4'-OH ont été méthylés. Pour expliquer ce changement de régiosélectivité, les auteurs ont appliqué des simulations de docking. Les résultats (Figure I17, page suivante) des simulations ont montré que, dans la cavité catalytique de l'enzyme POMT-7, la lutéoline est orientée avec le 7-OH proche du cofacteur SAM, stabilisée par une interaction du type π - π avec le Trp126 et des liaisons H avec les résidus Arg253, Asp256 et Gln117. Avec la 3-O-méthoxy-lutéoline, deux orientations possibles ont été obtenues : la première, semblable à celle obtenue pour la lutéoline, présente le groupement 3'-méthoxy stabilisé par des interactions hydrophobes avec les résidus Phe122, Leu123 et Phe163. Le 7-OH est alors accessible au centre catalytique. Dans la deuxième orientation, opposée à la première, le 3'-méthoxy est stabilisé par des résidus hydrophobes localisés plus au fond de la poche (Leu148, Trp149 et Phe162). Ainsi, c'est le 4'-OH qui est accessible au centre catalytique de l'enzyme.

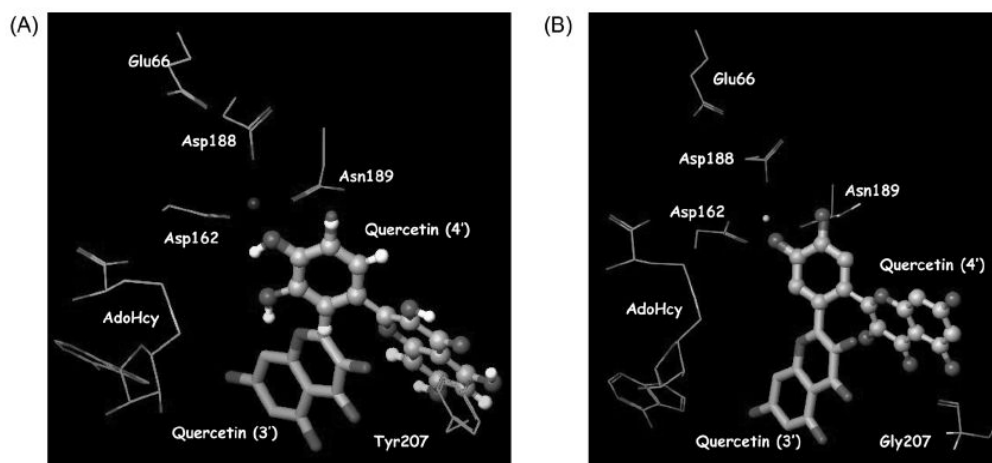


Figure 116 - Modèles obtenus par docking de la quercétine dans le site actif de la SOMT-9 : (a) l'enzyme native et (b) le mutant Y207G [142].

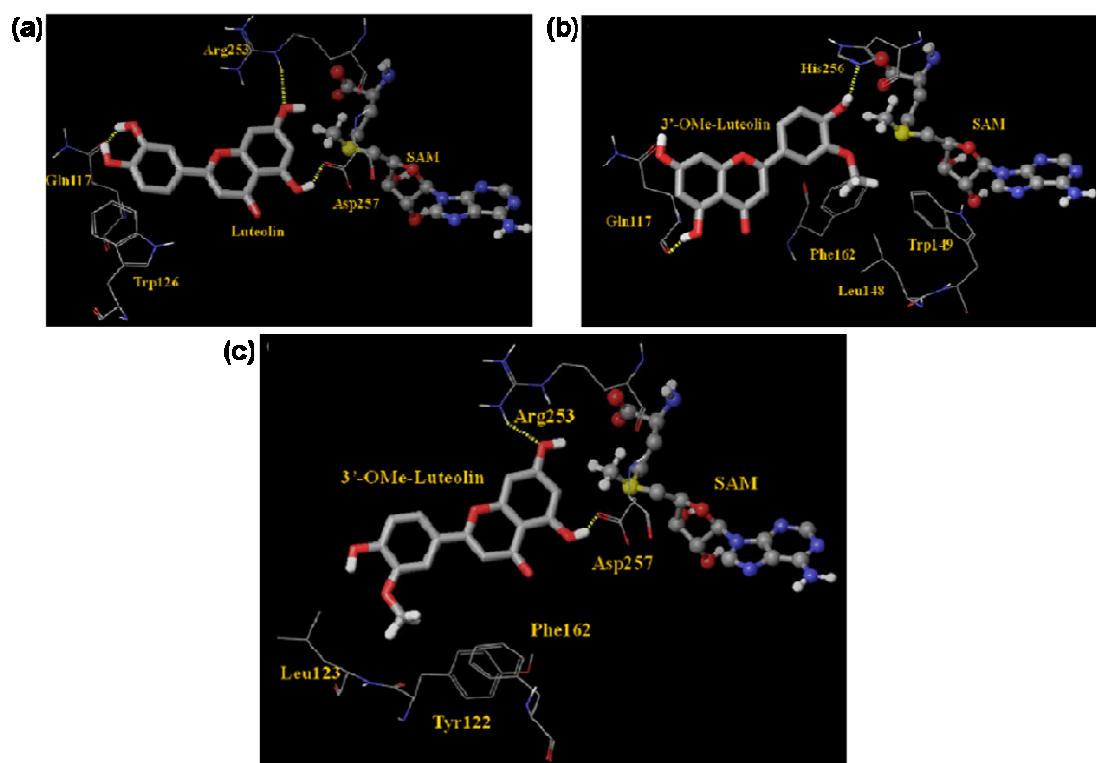
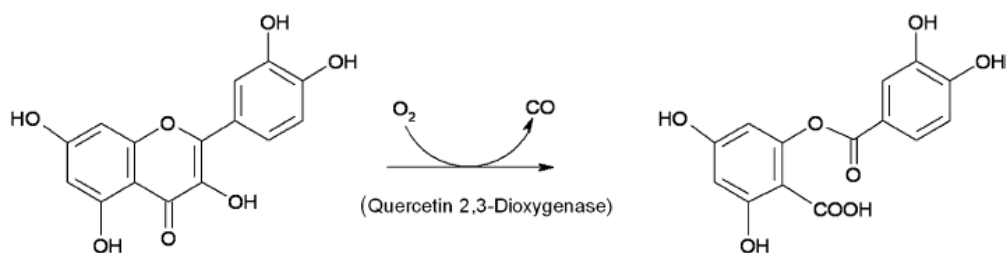


Figure 117 - Modèles obtenus par docking de (a) la lutéoline et (b,c) de la 3'-méthoxy-lutéoline dans le site actif de la POMT-7 [143].

3.2.2. L'oxygénolyse de flavonoïdes

Fiorucci et *al.* ont utilisé des simulations de dynamique moléculaire pour étudier l'oxygénolyse de la quercétine catalysée par l'enzyme quercétine-2,3-dioxygénase (2,3-QD) en présence d'oxygène moléculaire (O₂), une réaction-clé dans le métabolisme de ce flavonoïde :



- Leurs premières investigations, basées sur des dynamiques moléculaires et des analyses énergétiques, ont abouti à l'identification d'un canal étroit qui permet à l'O₂ d'accéder au site actif de l'enzyme [144]. Ce canal, formé par les résidus hydrophobes Phe132, Leu135, Phe136 et Phe175, connecte directement la surface de la protéine aux atomes du site actif et de la quercétine qui sont directement impliqués dans la réaction.
- Dans une autre article, ces auteurs relatent l'utilisation de simulations de dynamique moléculaire pour étudier la stabilisation de la quercétine dans la cavité du site actif de l'enzyme 2,3-QD [145]. Cette étude a permis d'identifier les régions de l'enzyme qui subissent un réarrangement conformationnel significatif dû à la présence de la quercétine (sub-unités 154-159 et 200-204) et les interactions formées par le flavonoïde avec les résidus de la cavité du site actif (Figure I18). Il a été montré que la quercétine forme des interactions électrostatiques avec le cation Cu²⁺ du site actif et des interactions hydrophobes avec les résidus tapissant la cavité catalytique. En outre, une molécule d'eau présente dans la structure cristalline de la protéine crée un réseau de liaisons H entre la quercétine et les résidus Thr53 et Asn74, tour à tour. Une autre contribution importante de cette étude a été la mise en évidence des déformations structurales subies par la quercétine, notamment la rotation du cycle B et la pyramidalisation du carbone C2, dues aux interactions de celui-ci avec les acides aminés.
- Plus récemment, ces auteurs ont publié une étude portant sur le mécanisme de l'oxygénolyse de la quercétine et de quelques flavonoïdes de structures similaires (galangine, kaempférol, myricétine, morine et fisétine), catalysée par l'enzyme 2,3-QD [146]. Dans cette étude, des méthodes quantiques (DFT) et hybrides (QM/MM) ont été appliquées pour caractériser structuralement et énergétiquement les points stationnaires (réactifs, intermédiaires et produits) et les états de transition du mécanisme réactionnel : un atome d'oxygène (O_a) de la molécule d'O₂ se lie sur le carbone C2 du flavonoïde, le deuxième atome d'oxygène (O_b) se lie ensuite au carbone C4, puis les liaisons C2-C3, C3-C4 et O_a-O_b sont rompues, libérant les produits monoxyde de carbone (CO) et le flavonoïde oxydé (Figure I19). En outre, les changements structuraux subis par les flavonoïdes ont été corrélés non seulement à la stabilisation du complexe ES, mais aussi à l'abaissement des barrières énergétiques d'activation tout au long de la réaction.

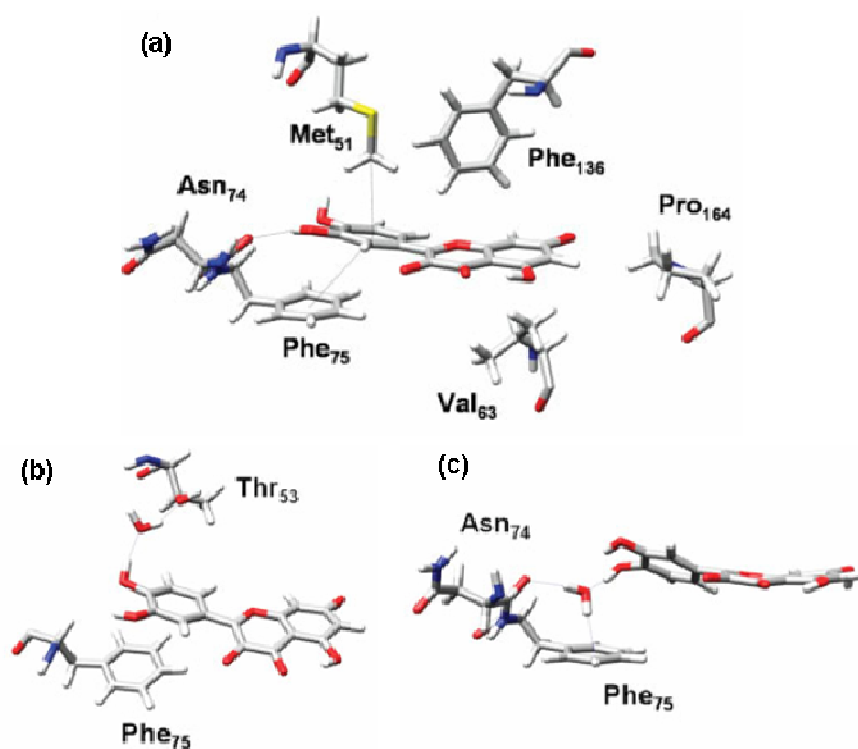


Figure I18 - interactions stabilisant la quercétine dans le site actif de la 2,3-QD : (a), les interactions hydrophobes avec les résidus Phe75, Met51, Phe136, Pro164 et val63 ; (b) et (c), les liaisons hydrogène établies, via une molécule d'eau, soit avec la Thr53, soit avec l'Asn74 [145].

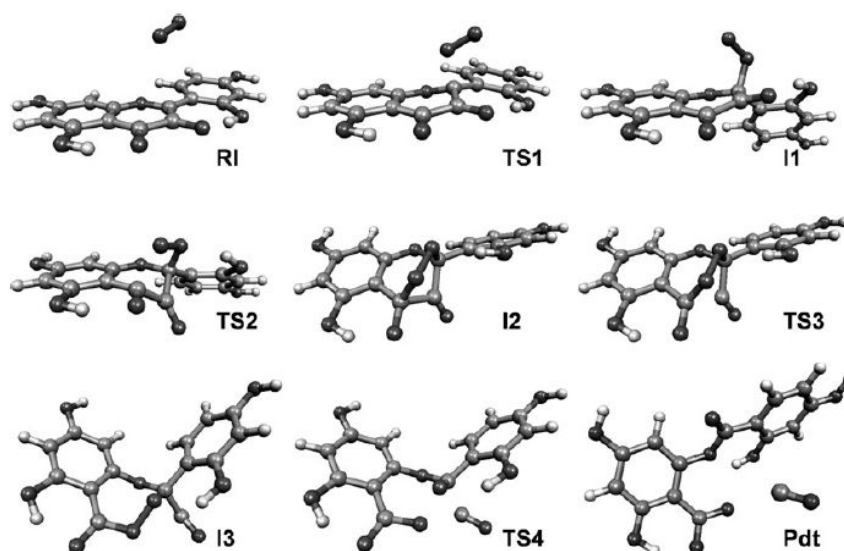


Figure I19 - Succession des intermédiaires (I1, I2 et I3) et états de transition (TS1, TS2, TS3 et TS4) au cours de l'oxygénolyse de la quercétine. RI représente les réactifs et Pdt les produits. Les atomes d'hydrogène sont colorés en blanc, ceux d'oxygène en noir et ceux de carbone en gris [146].

3.2.3. L'inhibition de la lipoxygénase par la quercétine

Fiorucci *et al.* se sont aussi intéressés à l'inhibition de la lipoxygénase-3 de soja (LOX-3) par la quercétine [147]. Des complexes différant les uns des autres par l'orientation de la quercétine sous forme semi quinone (3-O ou 7-O ou 4'-O orienté vers l'ion Fe^{2+} du site catalytique) ont été modélisés par dynamique moléculaire. Dans les trois cas, il a été montré que les complexes sont stabilisés par des interactions hydrophobes avec les résidus de la poche et des interactions électrostatiques entre le flavonoïde et le cation métallique et des molécules d'eau du site actif (Figure I19). Il a été observé que le complexe avec la semi quinone 3-O présente les distances entre l'atome d'oxygène et le Fe^{2+} les plus faibles et les plus stables au cours de la trajectoire, ainsi que la valeur d'énergie libre de complexation (évaluée par la méthode MM-PBSA) la plus faible. Ces résultats suggèrent que la quercétine se lie au cation métallique de cette enzyme via le 3-OH.

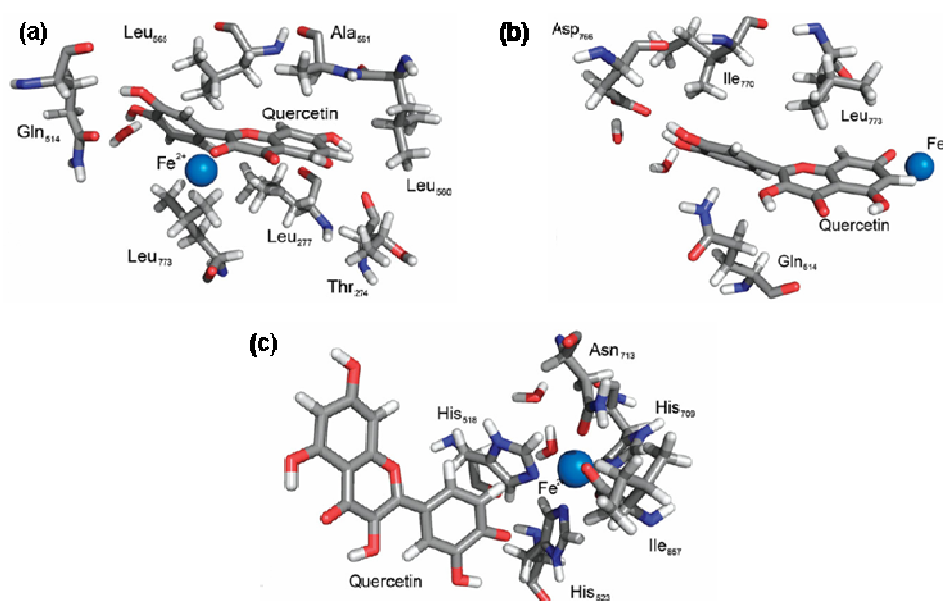


Figure 120 - Snapshots des complexes, montrant les semi quinones (a) 3-O, (b) 7-O et (c) 4'-O pointées vers le cation Fe^{2+} du site actif de la LOX-3 [147].

3.3. Inhibition de cibles biologiques par des flavonoïdes

En plus des études vouées à l'étude de réactions de bioconversion enzymatique, la littérature est riche de travaux relatant l'utilisation d'outils de modélisation moléculaire pour mieux comprendre l'action inhibitrice des flavonoïdes vis-à-vis de protéines impliquées dans certaines pathologies. Cette démarche a pour but principal d'identifier les interactions intermoléculaires stabilisant les complexes protéine-flavonoïde. L'importance de ces interactions est souvent corrélée aux propriétés inhibitrices des flavonoïdes, déterminées *in vitro*. Sans être exhaustif, le Tableau I5 (les trois pages suivantes) présente un résumé de quelques études récentes portant sur ce thème.

Tableau I5 - Sommaire de quelques études employant des outils de modélisation moléculaire pour identifier les interactions intermoléculaires qui sont à la base de l'inhibition des protéines par des flavonoïdes.

Protéines / flavonoïdes	Activité biologique Données expérimentales	Outils de simulation moléculaire Interactions protéine-flavonoïde identifiées	Ref.
Transport de la quercétine par la serum albumine humaine (HSA)	- Biodisponibilité du flavonoïde - Complexation mise en évidence par spectroscopie UV-visible et dichroïsme circulaire	- Simulations : docking (AutoDock) - Interactions identifiées : liaisons hydrogène 3-OH...Lys195 et 4'-OH...Asp451 ; π-π stacking Trp241-cycle C	[40]
Inhibition de la protéase 3CLPro du coronavirus (CoV) par la quercétine-3-β-galactoside	- Traitement des troubles respiratoires provoqués par le CoV - IC₅₀ : enzyme native = 43 μM mutant Q189A = 128 μM	- Simulations : docking (DOCK) - Interactions identifiées dans l'enzyme native: liaisons hydrogène 3-O...Glu166, Gln166...4-C=O, 3'-OH...Gln189, 4'-OH...Gln189 et 3''-OH (sucre)...Gln189. - Les 3 liaisons hydrogène avec le résidu 189 ne sont pas possibles dans le mutant Q189A, diminuant ainsi l'affinité du ligand.	[148]
Régulation de l'activité amidolytique de la thrombine humaine par la quercétine	- Prévention de troubles de coagulation sanguine et de certaines tumeurs - Constante de dissociation k_d : 0,34 μM	- Simulations : docking (MOE) combiné avec des optimisations de géométrie (champ de forces MMFF94) - Interactions identifiées : liaison hydrogène 3-OH...His57 ; attractions électrostatiques avec Lys60, Glu192 et Ser195 ; interactions σ-π Trp60-bicyclo A,C	[149]
Inhibition sélective des mono-amine oxydases A et B (MAO-A et MAO-B) par la quercétine	- Amélioration des symptômes de la dépression et de l'anxiété chronique - IC₅₀ : MAO-A = 0,01 μM MAO-B = 20 μM	- Simulations : docking (MOLINA) combiné avec des optimisations de géométrie (champ de forces AMBER) - Interactions identifiées dans la MAO-A : liaisons hydrogène 3-OH...Asn181, 5-OH...Asn181, 5-OH...Tyr197 et 3'-OH...Glu216 ; π-π stacking Phe352-cycle B, Tyr407-cycle A et Tyr444-cycle A - Interactions identifiées dans la MAO-B : liaisons hydrogène 3-OH...Gln206, 5-OH...Ser433, 3'-OH...Ile198 ; interaction σ-π Tyr398-cycle C	[43]

Protéines / flavonoïdes	Activité biologique Données expérimentales	Outils de simulation moléculaire Interactions protéine-flavonoïde identifiées	Ref.
Inhibition de la P-glycoprotéine humaine par la chrysine, la 6-prényl-chrysine et la galangine	- Augmentation de la biodisponibilité de certaines molécules utilisées dans le traitement du cancer - Constantes de dissociation k_d : chrysine = 8,9 μM galangine = 5,3 μM 6-prényl-chrysine = 0,3 μM	- Simulations : docking (AutoDock) combiné avec des optimisations de géométrie (champ de forces GROMOS) - Interaction identifiée pour la chrysine : liaison hydrogène 5-OH...Lys1076 - Interactions identifiées pour la galangine : liaisons hydrogène Ser1077...1-O, 3-OH...Ser1071, 3-OH...Cys1074, 3-OH...Gly1075 - Interactions identifiées pour la 6-prenyl-chrysine : liaisons hydrogène 4-C=O...Thr1078, 5-OH...Lys1076, 5-OH...Ser1077, 7-OH...Ser1072 ; interaction σ-π Tyr1044-cycle B	[150]
Inhibition de la peroxydase de raifort (HRP) par la quercétine, la myricétine, le kaempférol et la galangine	- Activité anti-oxydante IC_{50} : quercétine = 3,0 μM myricétine = 3,1 μM kaempférol = 10,2 μM galangine = 16,3 μM	- Simulations : docking (GOLD) combiné avec des optimisations de géométrie (champ de forces MMFF) - Interactions identifiées pour les 4 flavonoïdes : liaisons hydrogène entre les hydroxyles sur le cycle B et les résidus catalytiques His42 et Asp38 au fond de la cavité; interactions aromatiques entre le cycle A et les résidus Phe68 et Phe142 à l'entrée.	[151]
Inhibition de de la phospholipase humaine A2-IIA par la quercétine et la naringénine	- Activité anti-inflammatoire et modulation de l'agrégation plaquettaire IC_{50} : quercétine = 6,2 μM naringénine = 50 μM	- Simulations : docking (FlexX) combiné avec des simulations de dynamique moléculaire (10 ns) sur les complexes - Interactions identifiées pour la quercétine : liaisons hydrogène 7-OH...His47 et 7-OH...Cys44 ; π-π stacking cycle A-Phe5 et Phe98-cycle A ; interaction NH-π Lys62-cycle B. - Interactions identifiées pour la naringénine : liaisons hydrogène 7-OH...His47 et 7-OH...Cys44 ; interactions hydrophobiques CH-π entre le bicyclo AC et les résidus Ala17 et Ala18	[152]
Inhibition de la caspase-3 par la myricétine	- Prévention de l'apoptose neuronale, qui est à l'origine de plusieurs maladies neurodégénératives IC_{50} : 10,2 μM	- Simulations : docking (LigandFit) combiné avec des optimisations de géométrie (champ de forces CHARMM) - Interactions identifiées : liaisons hydrogène Arg207...1-O, 3'-OH...Glu123, 4'-OH...Glu123 ; blocage stérique du résidu catalytique Cys163	[153]

Protéines / flavonoïdes	Activité biologique Données expérimentales	Outils de simulation moléculaire Interactions protéine-flavonoïde identifiées	Ref.
Inhibition de la β-secretase (BACE-1) par la myricétine, la quercétine, le kaempférol, la morine et l'apigénine	- Modulation de la production du peptide β-amyloïde, impliqué dans la maladie d'Alzheimer - IC₅₀ : myricétine = 2,8 μM quercétine = 5,4 μM kaempférol = 14,7 μM morine = 21,7 μM apigénine = 38,5 μM	- Simulations : docking (LigandFit) combiné avec des optimisations de géométrie (champ de forces CHARMM) - Interactions identifiées pour la myricétine : liaisons hydrogène Trp198...4'-O, Trp198...5'-O, 3-OH...Asp32 (2 liaisons H), 7-OH...Gln73 - Interactions identifiées pour la quercétine : liaisons hydrogène Trp198...4'-O, Trp198...5'-O, 3-OH...Asp32 (1 liaison H), 7-OH...Gln73 - Interactions identifiées pour le kaempférol : liaisons hydrogène Trp198...4'-O et 3-OH...Asp32 (1 liaison H) - Interaction identifiée pour la morine : liaison hydrogène 2'-OH...Asp228 - Interaction identifiée pour l'apigénine : liaison hydrogène 7-OH...Gln73	[154]
Inhibition de la carbonyl-réductase humaine par la rutine, la quercétine et la quercitrine	- Diminution des effets secondaires des anthracyclines, utilisées dans le traitement de certains cancers - IC₅₀ : rutine = 2,1 μM quercitrine = 6,5 μM quercétine = 6,2 μM	- Simulations : docking (ArgusDock) combiné avec des optimisations de géométrie (champ de forces MMFF) - Interactions identifiées pour la rutine et pour la quercitrine : liaisons hydrogène 7-OH...Met234, Trp229...3-O, 3'-OH...Cys226 - Interactions identifiées pour la quercétine : liaisons hydrogène Ser139...4-C=O, 3-OH...Tyr193, 3'-OH...Met234	[155]
Inhibition d'une trihydroxynaphthalène réductase fongique (3HNR) par l'apigénine et par biochanine A	- Traitement d'allergies d'origine fongique et de mycoses - IC₅₀ : apigénine = 3,1 μM biochanine A = 6,5 μM	- Simulations : docking (AutoDock) - Interactions identifiées pour l'apigénine : liaisons hydrogène 7-OH...Ser149 et 4'-OH...Ser228 ; π-π stacking cycle A-Tyr208 et cycle A-NADPH (coenzyme) - Interactions identifiées pour la biochanine A : liaisons hydrogène 5-OH...Ser149 et 7-OH...Tyr163 ; π-π stacking cycle A-Tyr208 et cycle A-NADPH (coenzyme) ; interactions hydrophobes entre le groupe 3'-OCH₃ (méthoxy) et les résidus Trp227 et Ala224	[156]

L'analyse de ces études, qu'elles portent sur la transformation des flavonoïdes catalysée par des enzymes ou sur l'action inhibitrice de ces composés sur des protéines diverses, permet de vérifier le potentiel que la modélisation moléculaire peut avoir dans ces contextes : elle permet une évaluation des modes d'interaction entre les ligands et les protéines, ce qui est très difficilement accessible par voie expérimentale. L'ordre d'affinité entre flavonoïdes et acides aminés, proposé par Codorniu-Hernandez et al. [136, 137], corrobore les travaux analysés : les liaisons hydrogène entre les groupements -C=O et -OH des flavonoïdes avec des acides aminés polaires, ainsi que les interactions aromatiques entre leurs cycles aromatiques et des acides aminés de ce type, sont souvent identifiés comme éléments stabilisateurs majeurs des complexes protéine-flavonoïde. Associée aux résultats expérimentaux d'inhibition et de biosynthèse, la connaissance de ces interactions est précieuse : d'un point de vue fondamental, ceci contribue à une meilleure compréhension des phénomènes biochimiques étudiés. D'un point de vue plus applicatif, ces connaissances sont indispensables pour le design de nouveaux inhibiteurs plus puissants et plus sélectifs (médicaments plus efficaces) ou de nouvelles enzymes à sélectivité et stabilité contrôlées (biocatalyseurs plus performants).

4. APPLICATIONS DE LA MODELISATION MOLECULAIRE POUR LA COMPREHENSION DES INTERACTIONS LIPASES-SUBSTRATS

4.1. Introduction

Les lipases (EC 3.1.1.3) sont les parmi les enzymes les plus utilisées dans des procédés de bioconversion industriels. Elles catalysent une grande diversité de réactions d'hydrolyse et de synthèse et sont souvent utilisées pour leurs propriétés d'énantiosélectivité et de régiosélectivité. Comme mentionné précédemment dans la section 1.3, ce sont les enzymes les plus performantes, donc les plus utilisées, pour l'obtention de dérivés de flavonoïdes régiosélectivement acylés. C'est pourquoi nous nous focalisons maintenant sur ces biocatalyseurs. Dans cette dernière partie de l'étude bibliographique, nous résumons les contributions de nombreux travaux antérieurs sur la sélectivité de biotransformations catalysées par des lipases. Nous examinerons plus particulièrement les publications décrivant l'utilisation de techniques de modélisation moléculaire pour expliquer des résultats expérimentaux démontrant l'énantiosélectivité ou la régiosélectivité de l'action de lipases. La revue est structurée en trois parties: (1) un résumé des connaissances sur la structure et les mécanismes d'action des lipases, (2) une analyse des travaux expérimentaux et de modélisation montrant l'influence de l'étape initiale de fixation des substrats sur la sélectivité de la lipase et (3) une analyse des travaux modélisant également la réactivité des complexes enzyme-substrat formés pour expliquer la sélectivité de bioconversions.

Cette partie est rédigée en anglais, sous la forme d'un projet de *Review*, préparée en réponse à une invitation de l'éditeur du journal *Process Biochemistry*.

4.2. INVITED REVIEW

Contribution of molecular modelling to the understanding of lipase selectivities

Eduardo B. De Oliveira ^a, Catherine Humeau ^{a*}, Elaine R. Maia ^b,
Mohamed Ghoul ^a, Jean-Marc Engasser ^a

^a *Laboratoire Biocatalyse Bioprocédés (LBB), ENSAIA-INPL, Nancy Université, 2 av. de la Forêt d'Haye, 54500, Vandoeuvre-lès-Nancy, France.*

^b *Laboratorio de Estudos Estruturais Moleculares (LEEM), Instituto de Química, Universidade de Brasília, CP 4478, 70904-970, Brasília-DF, Brazil.*

* Corresponding author - phone: (+33) 03.83.59.57.84 / fax: (+33) 03.83.59.57.78

Abstract:

Scientific research has progressively increased its computational efforts to the treatment and the analysis of data derived from 3D-structures and dynamics of proteins, with the goal of both scientific and commercial breakthrough, most evidently in drug discovery and in bioprocesses conception, development and optimization. This review endeavors to dress a panel of the applications of molecular modelling methods to lipase-catalyzed bioprocesses in the last years. These enzymes constitute the most versatile and important biocatalysts studied and utilized worldwide in a huge array of biotechnological processes, due to their impressive substrate spectra, selectivity properties and stability in various solvents and in a broad temperature range. Studies grouped here deal with the molecular basis for the enantio- and regioselectivity displayed by these enzymes in a broad spectrum of bioprocesses are discussed. They exemplify how molecular modelling tools can aid to understand substrate recognition modes of lipases and correlate it to the enzyme selectivity.

Keywords:

Molecular modeling, Bioprocesses, Lipase, Enantioselectivity, Regioselectivity.

1. Introduction

Lipases (triacylglycerol hydrolases, EC 3.1.1.3) are among the most widely used enzymes in biotechnological processes. Formally, they are defined as enzymes that catalyze hydrolysis of long chain triglycerides, monoacylglycerols, diacylglycerols and other water-insoluble esters of fatty acids [157]. These biocatalysts can also catalyze hydrolysis reverse reactions in non-aqueous media, consisting in acylating alcohol or amine functions and esterifying carboxylic acids [158-160]. Moreover, it has been reported the lipase catalytic ability in some atypical reactions, involving the formation of C-C, C-N and C-S bonds [161]. Their wide activity spectrum makes them one of the most important and versatile group of biocatalysts in a wide range of reactions, for both industrial and academic purposes. [28, 162, 163]. During the last years, the vast biotechnological importance of lipases has been extensively underlined and numerous papers reviewing some of their specific applications have been published. Without being exhaustive, some of these review papers are listed in Table 1.

Table 1 – Some review papers dealing with biotechnological applications of lipases

Application field	Ref.
<i>Chemicals :</i>	
▪ Production of bio-diesel	[164-166]
▪ Synthesis of polyesters	[167-169]
▪ Fabrication of paper	[170-172]
<i>Food technology :</i>	
▪ Production of ingredients/additives	[173-176]
▪ Wastewater treatment	[177-179]
<i>Pharmaceutical, cosmetrical and agricultural domains :</i>	
▪ Enantiomeric resolution of chiral active compounds	[180-184]
▪ Regioselective acylation of poly-hydroxylated active compounds	[27, 185-187]

Many lipases are characterized by a high degree of enantioselectivity in hydrolysis and synthesis reactions. This makes them particularly attractive catalysts for asymmetric transformations. They are frequently used for kinetic resolutions of enantiomers mixtures, for instance of racemic carboxylic acids or alcohols, and for the syntheses of enantiomerically pure products. In addition lipases can also display interesting regioselective properties, characterized by their aptitude to catalyze a reaction onto a specific functional group or region of their substrates, while conserving unmodified the remaining parts of the molecule.

When designing a lipase-catalyzed bioconversion processes, a key step are the selection of an enzyme that yields a high catalytic activity with the desired selectivity. The biocatalyst may be selected among existing lipases, generally produced in high yields by microbial organisms (mainly bacteria or fungi). Presently the main used industrial lipases are *Candida antarctica* lipase B (CALB), *Burkholderia cepacia* (formerly *Pseudomonas cepacia*, PCL), *Rhizomucor miehei* lipase (RML) and *Candida rugosa* lipase (CRL). Alternatively, if improved enzymatic activities or selectivities are

necessary for process competitiveness, a novel lipase may be engineered by mutation or evolution of existing enzymes [32, 188]. For an accelerated and rational selection or engineering of a lipase catalyst, a prerequisite is the understanding the molecular basis of lipase action. One should preferably have a minimal knowledge of how the lipase amino acid residues interact with the substrates functional groups to determine the biotransformation activity and selectivity [189, 190].

During the last decade, an increasing number of publications addressed the question of the basic mechanism of lipase catalysis. Many of them entailed molecular modeling techniques to investigate the processes responsible for the lipase enantio- or regioselectivities. Indeed, these modelling-based studies established that lipase selectivity can be rationalized on the basis of interactions of lipase amino acids residues with the reaction substrates, some at the initial step of the substrate binding into the catalytic cavity, others at the subsequent step of substrates transformation. The aim of the present review is to analyze recent results on the use of molecular modelling to explain the experimentally observed selectivities of lipase catalyzed bioprocesses. It contains: (1) a description of the common lipases structure and catalytic mechanisms, (2) a brief summary of the underlying principals of molecular modeling techniques, and their expected contributions to enzyme catalysis; (3) a review of studies attributing lipase selectivity to the initial substrate binding step, (4) a review of the publications explaining lipase selectivity on the basis of reactivities of the bound substrates.

2. Lipases structures and catalytic mechanism

2.1. General structure of lipases

On a structural point of view, lipases are 20-60 kDa proteins belonging to the fold family of the α/β -hydrolases. Their architecture consists of a central hydrophobic β -sheet, enfolded by amphiphilic α -helices. Most of these enzymes present a lid consisting of one or two short α -helices covering the catalytic pocket and linked to the enzyme body. As the enzyme draws near a hydrophobic surface, the movement of this lid leads to the exposition of the catalytic pocket to the enviroing solvent and substrate molecules. This phenomenon is often referred as interfacial activation and is recognized as an important part of the activation of the enzyme activity [65, 191, 192]. However, some lipases do not present this structural feature, as occurs with *Candida antarctica* lipase B (CALB), one of the most widely used biocatalysts [193, 194]. In such cases, only minor structural changes correlated to catalytic activity are normally observed.

The catalytic machinery of these enzymes is constituted by a triad of residues complemented by the so-called oxyanion hole. The catalytic triad consists of the residues the Ser...His...Asp/Glu, placed in the bottom of the catalytic pocket, at a highly conserved geometry from one lipase to another. The oxyanion hole is made of two residues that stabilize the substrate in the catalytic intermediates through H bond interactions [189]. The catalytic pocket has a hydrophobic cleft that lodges the acyl moiety of the substrate and another more hydrophilic region that accommodates the alcohol moiety, in either esterification or ester hydrolysis reactions. The size and shape of this pocket deeply differ from one lipase to another and this is one of the factors that greatly influence their different specificity and selectivity properties [190].

2.2. Mechanism of lipase catalysis

The reaction mechanism classically accepted for lipases is known as “ping-pong bi-bi mechanism” [195], represented in Figure 1.

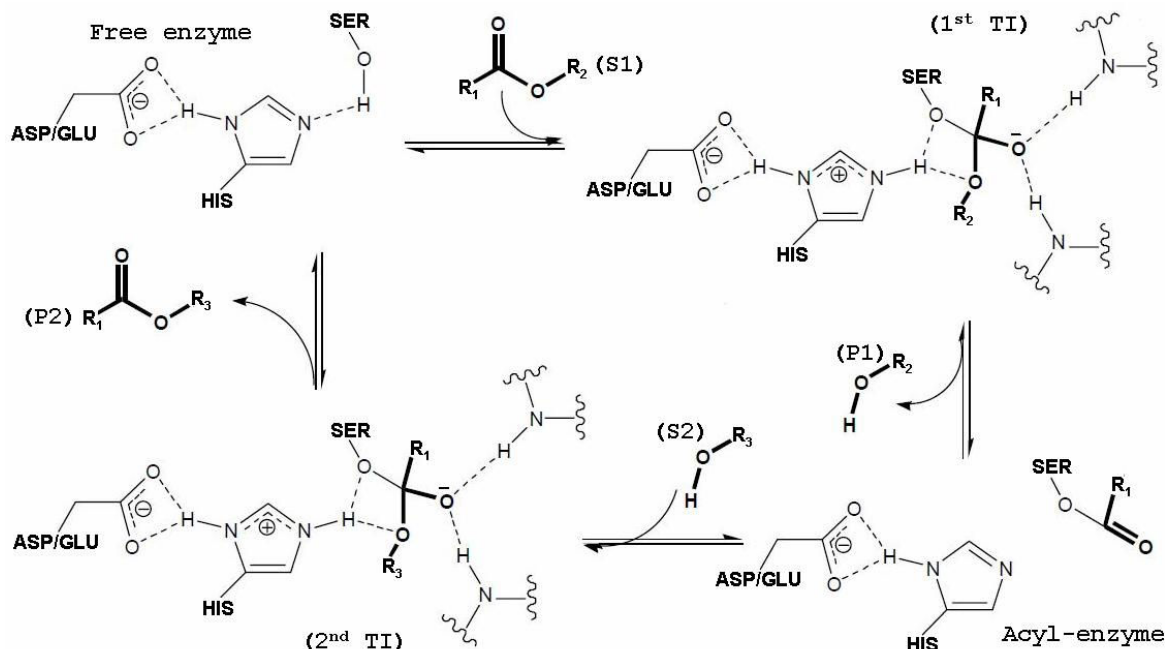


Figure 1 – Classical mechanism of lipases : if $R_2 = H$ and $R_3 = \text{carbon chain}$, this is an esterification; if $R_2 = \text{carbon chain}$ and $R_3 = H$, this is an ester hydrolysis; if $R_2 = \text{carbon chain}$ and $R = \text{carbon chain}$, this is a transesterification.

According to this later, the acyl-donor substrate (S1) enters the catalytic cavity and the hydroxyl proton of the serine residue is transferred to the histidine $N\epsilon$ atom. Then, the serine executes a nucleophilic attack on the acyl donor substrate, forming the first tetrahedral intermediate (1st TI). This leads to the formation of the acyl-enzyme complex, with the release of the first product (P1). Next, the acyl acceptor substrate (S2) enters the active site and transfers the proton of the function to be acylated to the $N\epsilon$ atom of the histidine, improving its nucleophilicity. Then, this “activated” function executes a nucleophilic attack on the acyl-enzyme carbonyl carbon, leading to the formation of the second tetrahedral intermediate (2nd TI) with the subsequent release of the second product (P2), which leaves the active site. The histidine $N\epsilon$ proton is then transferred to the serine, making the enzyme ready for another catalytic cycle. The role of the aspartate (or glutamate) residue has been identified as to enhance the nucleophilicity of the serine to provide its electrostatic stabilization, due to the electric field of the charged aspartate side chain. The backbone amides of the oxyanion hole residues stabilize the transition states and tetrahedral intermediates through hydrogen bond interactions.

3. Lipase selectivity determined at the substrates binding level

Most of the molecular modeling studies aimed at investigating the selectivity of lipases featured simulations of substrates docking in the enzyme catalytic site. Docking protocols were

generally used to select the most favorable binding modes. The simulation results are successively reviewed for studies addressing the issue of lipase enantioselectivity and regioselectivity.

3.1. Lipase enantioselectivity explained by the substrates binding modes

First studies aiming to describe and model the origin of enantioselectivity of an enzyme lead to the elaboration of some empirical models, which did not attempt to quantify the degree of enantioselectivity, but simply predict which enantiomer is preferred by the enzyme. The first of such models was the Prelog's rule [196] related the enantioselectivity of alcohol dehydrogenases in the reduction of ketones to the size of the two substituents on the carbonyl group. In the case of lipases, the most known model is the empirical rule of Kazlauskas [197] for chiral recognition of secondary alcohols. It establishes that enantioselectivity is directly correlated to the difference in size between the large (L) and the medium (M) substituents of the stereocenter. It was originally exemplified for the lipases PCL and CRL, and further extended to the active site model for all lipases that have two pockets of different sizes, a small one (M) and a large one (L). More recent modeling studies aimed at rationalizing experimentally observed lipases enantioselectivities are reported in Table 2. They cover several lipases and families of bioconversion reactions, and use different docking and energy minimization protocols. Most of them evaluate the protein-ligand interaction energies characterizing the binding mode. According to these studies, several phenomena can explain why binding modes may be more favorable for one enantiomeric substrate than for the other: differences in the protein-ligand interaction energies for the two enantiomeric substrates, the shape adaptation between substrates and the lipase binding pocket, the possible sterical clashes provoked by the non-preferred enantiomer, the kind and number of interactions formed between the lipase and the docked substrates (hydrogen bond interactions, electrostatic interactions, hydrophobic contacts and π -stacking) and, occasionally, the distortion of the whole complex or in a part of it, evaluated by *rmsd* calculations.

Table 2 – Some examples of the use of molecular modeling aiming to understand the enantioselectivity based on the recognition of substrates by lipases binding pocket.

Lipase	Experimental data		Molecular modelling data		Ref.
	Reaction / Substrates	Experimental results	Modelling methods	Molecular explanations for experimental results	
<i>CALB</i>	Acetylation of 9 chiral secondary alcohols with short chains and 8 ones with short chains.	(R)-enantiomers react faster - E value not given	- Manual covalent docking (SYBIL version 6.1) - Energy minimizations and MD simulations (AMBER force field) - Quantum chemical semi-empirical optimizations (MNDO)	Sterical hindrance in the cavity observed for (S)-enantiomers, due to residues Trp104, Pro38, Gly39, Thr40, Thr42 and Ser27.	[198]
<i>CALB</i>	Hydrolysis of chiral esters of aliphatic secondary alcohols	(R)-enantiomers react faster - E values in the range 8-260	- Manual covalent docking (SYBIL version 6.6) ΔG estimations based on MD simulations (Kollman All Atoms force field)	ΔG values for formation of the reaction intermediates lower for (R)-enantiomers than for (S)-ones	[199]
<i>CALB</i>	Acetylation of N-substituted (3R,4R)- and (3S,4S)-dihydropiperidines	(3R,4R)-enantiomers react faster - E value $\approx$ 200	- Manual covalent docking (Insight II version 2000.1) - Systematic conformational search and energy minimizations (AMBER force field)	Sterical hindrance in the cavity observed for (3S,4S)-enantiomers, due to the residues Trp104, Ala281, Gly39, Ile285, Glu188 and Leu278.	[200]
<i>CALB</i>	Esterification of (R,S)-pentan-2-ol with propanoic acid	(R)-enantiomers react faster - E value $\approx$ 100 (for reaction medium a_w close to 0) - E value $\approx$ 320 (for reaction medium a_w about 0.2)	- Manual covalent docking of the substrates, with and without a water molecule (SYBIL version 6.6). - Energy minimizations and MD simulations (Kollman All Atoms force field)	- Sterical hindrance in the cavity observed for (S)-enantiomers, due to the residue Trp104. 3 H bonds mediated by the water molecule. - No sterical hindrance for (R)-enantiomers (better fitting). 5 H bonds mediated by the water molecule confer additional stabilization.	[201]

Lipase	Experimental data		Molecular modelling data		Ref.
	Reaction / Substrates	Experimental results	Modelling methods	Molecular explanations for experimental results	
<i>CALB</i>	Acetylation of the (R,S)-hexan-2-ol and its δ-substituted –OH, –C=O and –COOH derivatives	(R)-enantiomers react faster - Hexan-2-ol: E value > 300 δ-OH derivative: E value > 200 δ-C=O derivative: E value < 100 δ-COOH derivative: E value $\approx$ 10	- Manual covalent docking (Moloc package) - MD simulations (MAB force field)	- Shorter H bonds with the catalytic His224 and lower deviations from the oxyanion hole observed for all (R)-enantiomers. - Better stabilization for δ-OH derivatives due to electrostatic attractions with Thr40. - Weaker stabilization for δ-C=O and δ-COOH due to repulsive interactions and sterical hindrance with Ala281 and Ala282	[202]
<i>CALB</i> <i>CRL</i>	Acetylation of (cis9,trans11)- and (trans10,cis12)-linolenic acid	- With CALB, the (trans10,cis12)-isomer reacts about 2-fold faster - With CRL, the (cisp,trans11)-isomer reacts about 3-4-fold faster	- Mulliken atomic charges for tetrahedral intermediates adjusted by DFT (B3LYP/6-31+G) - Manual covalent docking (GROMACS version 3.3.3) - Energy minimizations and MD simulations (GROMOS force field)	- More stable H bonds observed for the fast-reacting substrates: substrate O...His:Nϵ and His:Nϵ...Ser:Oγ - No correlation found between substrate preference and binding energy or linolenic acid conformation.	[203]
<i>CRL</i>	Hydrolysis of methyl esters of (R,S)-atliprofen, (R,S)-naproxen and (R,S)-ketoprofen	(S)-enantiomers react faster - Atliprofen: E value = 206 - Naproxen: E value = 100 - Ketoprofen: E value = 58	- Manual covalent docking (SYBYL version 6.9) - Energy minimizations (Tripos force field) ΔH estimations with a quantum semi-empirical method (AM1)	ΔH values for formation of the reaction intermediates lower for (S)-enantiomers than for (R)-ones - Lower enantiopreference for ketoprofen correlated with a repulsive interaction with the Leu297, formed by both enantiomers.	[204]

Lipase	Experimental data		Molecular modelling data		Ref.
	Reaction / Substrates	Experimental results	Modelling methods	Molecular explanations for experimental results	
<i>PCL</i>	Ring opening of (R,S)-δ-lactones and (R,S)-γ-lactones	(R)-enantiomers react faster δ-lactones: E values in the range 2-100 γ-lactones: E values in the range 4-11	- Atomic charges for tetrahedral intermediates and catalytic residues adjusted by a quantum semi-empirical method (MNDO94) - Manual covalent docking (SYBYL version 6.4) - Energy minimizations and MD simulations (Tripos force field)	(S)-lactones alkyl chains push the Tyr29, which push the catalytic His224, provoking the rupture of the H bond with the substrate. - Larger displacements in these residues observed for longer chain length lactones (mainly in the case of δ-lactones)	[205]
<i>PCL</i>	Hydrolysis of chiral esters of aromatic secondary alcohols	(R)-enantiomers react faster - E values in the range 22 to >1000	- Manual covalent docking (MacroModel/BatchMin version 4.0) - Monte Carlo conformational search of the docked substrates - Energy minimizations (AMBER force field)	Better stabilization of (R)-enantiomers: stronger H bonds with the oxyanion hole residues Leu17 and Gln88 and stronger aromatic interactions between the substrates phenyl rings and the Tyr23 and Tyr29 residues.	[206]
<i>PCL</i>	Transesterification between (R,S)-2-bromophenyl acetic acid ethyl ester and octan-1-ol	(R)-enantiomers react faster - E value = 57	- Manual docking (Insight II) - Energy minimizations (CFF91 force field) - Two distinct protocols	- First protocol (failed) : no significant difference observed between H bond networks stabilizing the intermediates of the (R)- and (S)-enantiomers - Second protocol (successful): lower interaction energies observed for (R)-enantiomers in passing through the active site channel.	[207]

3.2 Lipase regioselectivity explained at the substrate binding level

Several combined experimental and modeling studies also addressed the question of lipases regioselectivity: when acting on a polyfunctional substrate, which is the functional group that is preferentially transformed and what are the structural elements of the lipase that determine this bioconversion selectivity? The published studies, summarized in Table 3, deal with the esterification, by several lipases, of polyhydroxylated compounds such as sugars, glycosides and prostaglandins. In all studies, docking protocols based on a combination of structural or energetic considerations were found capable of differentiating between the OH groups being either accessible or inaccessible to the lipase. For the studied substrates, calculations generally identified a single favorable orientation of the bound alcohol displaying a specific hydroxyl group in close proximity to the catalytic triad. Other substrate orientations were found more unlikely due to sterical repulsion with lipase residues. Observed differences in substrate acylation regioselectivities, when comparing different enzymes, could also be explained by accounting for the interaction energy of the ligands and for the vibrational entropy of ligands in their binding pockets.

Docking simulations also demonstrated the importance of a strong stabilization of the bound substrates in the enzyme cavity. For the preferential binding modes there is, always, hydrogen bond interactions that stabilizes both the acyl donor in the oxyanion hole and the acyl acceptor in proximity to the catalytic triad. Substrates stabilization may also be strengthened by hydrophobic interactions with the wall cavity. This is particularly important for substrates containing aromatic nuclei which can be stabilized by π -interactions with aromatic residues present in the cavity wall.

Table 3 – Some examples of the use of molecular modeling aiming to understand the regioselectivity based on the recognition of substrates by lipases binding pocket.

Lipase	Experimental data		Molecular modelling data		Ref.
	Reaction / Substrates	Experimental results	Modelling methods	Molecular explanations for experimental results	
CALB	▪ Acylation of the glucose 6-OH with different arylaliphatic acids	▪ Higher yields for acyl donors with a -CH₂- between the -COOH and the aromatic part ▪ Lower yields for acyl donors with a polar group in the aromatic part ▪ No product detected with α,β-unsaturated acyl donors	▪ Manual covalent docking (SYBYL 6.4) ▪ Energy minimizations (Tripos force field)	▪ Better stabilization of acyl donors with a -CH₂- between the -COOH and the aromatic part, due to favorable hydrophobic interactions in enzyme cavity. ▪ Worse stabilization of acyl donors with a polar group in the aromatic part, due to electrostatic disturbing of Thr40 ▪ Bad fitting observed between the enzyme cavity and α,β-unsaturated acyl donors, due to their higher molecular rigidity.	[208]
CALB	▪ Acylation of five stereoisomeric F-type prostaglandins (polyhydroxylated compounds which carry 3 free OH groups each)	Obtained products: ▪ For 3 of the substrates, only the monoester acetylated in position 11-OH ▪ For the 2 others, the monoesters acetylated in positions 9-OH et 11-OH + the diester simultaneously acetylated in the 9-OH and 11-OH positions	▪ Manual covalent docking (SYBIL version 6.7), mimicking the acetylation of all OH groups of all substrates ▪ Energy minimizations and MD simulations (Kollman All Atoms force field)	In the modeled structures that actually leads to the formation of an acetylated product, the following structural characteristics were simultaneously observed: ▪ Global protein rmsd ≤ 3 Å ▪ At least 3 of the 6 possible H bonds involving the substrates and the catalytic residues ▪ Maximal subset energy of -95 KJ.mol⁻¹	[209]

Lipase	Experimental data		Molecular modelling data		Ref.
	Reaction / Substrates	Experimental results	Modelling methods	Molecular explanations for experimental results	
CALB HLL SC	Acylation of sucrose (which carry 9 free OH groups) with palmitic acid	Obtained products: - With CALB : acylation of the 6'-OH, in the glucose moiety or the 6-OH, in the fructose moiety - With HLL: acylation of the 6-OH, in the fructose moiety - With SC: acylation of the 1'-OH, in the fructose moiety	- Manual covalent docking (software non specified), mimicking the acylation of the 1'-OH, the 6'-OH and the 6-OH for the 3 enzymes - Energy minimizations (CHARMm force field) - Estimations of enthalpy (ΔH) entropy (ΔS) and free energy (ΔG) (methods not specified) for the intermediate formation	- Correct prediction of regioselectivity for lipases, but not for subtilisin (6-OH predicted instead to 1'-OH), when evaluating enthalpy contributions (ΔH) only - Correct prediction of regioselectivity for all the three enzymes, when entropic contributions (ΔS) were included in the analysis	[210, 211]
CRL	Acetylation of 6-trityl-glycopyranoside (carrying 3 free OH groups) in hydrophobic medium (CCl_4)	Obtained products: - Monoester acetylated in the 3-OH position (about 80%) - Monoester acetylated in the 2-OH position (about 20%)	- Monte Carlo conformational search of the substrate - Manual, non covalent docking of the minimum energy structure within the acetyl-CRL cavity - Energy minimizations (AMBER force field)	- Serine-bound acetate correctly orientated towards the oxyanion hole residues - Closest substrate OH from the catalytic His449:N and Ser109:O atoms = 3-OH (4 Å) 2-OH forms H bonds with Glu2008 and Gly122 and 4-OH forms intramolecular H bonds	[212]
RML	Acetylation of a synthetic c-alkyl-resorcin-4-arene (which carries four free OH groups)	Obtained product: Monoester acetylated in the 2-OH position	- Monte Carlo conformational search of the substrate - Manual, non covalent docking of the minimum energy structures within the acetyl-RML cavity - Energy minimizations (AMBER United Atoms force field)	- Acetate correctly stabilized through 2 H bonds in the oxyanion hole (Ser82 and Leu145) - Only 2-OH comprised between the catalytic His257 and Ser144. - Additional stabilizing interaction between the substrate phenyl ring and the Trp88 residue	[213]

4. Lipase selectivity determined by the reactivity of the substrates binding modes

Whereas most of the modeling studies featured docking simulations of substrates in the lipase active site to determine the most favorable binding modes, some publications also applied molecular modeling to further evaluate the reactivity of these binding modes. They used different quantum computational techniques, such as *ab initio*, semi empirical and DFT. They all contributed to analyze the energetic profile of the reaction pathways and to characterize the involved transition states and reaction intermediates. Some also performed quantum calculations to rationalize experimentally observed selectivities of lipase catalyzed bioconversions.

4.1. Analysis of the reaction pathway energetics

Quantum simulations of lipase catalysis were generally aimed at studying the possibility of formation and transformation of the two tetrahedral intermediates of the hydrolysis or esterification pathway. They also evaluated the involved activation energies that determine the rates of the bioconversion reactions. The most in depth study is that of Hu *et al.* [214] that applied *ab initio* and density functional theory (DFT) modeling to methyl formate hydrolysis by CALB. It considered a very simplified model of the lipase active site consisting of the catalytic triad, represented by formate anion, imidazole, and methanol, and the oxyanion hole, represented by two water molecules. It calculated the system energy and the main bond lengths all along the reaction pathway, going from the initial Michaelis-Menten complex to the final liberated products. The study reported activation energies of 13.4 kcal.mol⁻¹ and 9.6 kcal.mol⁻¹ for the formation of the first and second tetrahedral intermediates, respectively, suggesting that acylation represents the rate limiting step of the hydrolysis reaction. It also demonstrated the essential contribution of the oxyanion hydrogen bonds and the Asp group in lowering the activation barrier. Each hydrogen bond in the oxyanion hole was calculated to contribute up to about 5 kcal.mol⁻¹ to stabilization of the reaction intermediate, whereas the Asp-Hist contributed up to about 6 kcal.mol⁻¹. Activations energies were determined for other lipase catalyzed reactions and found in the same range of values.

4.2. Comparative evaluation of the reactivity of lipase-enantiomers complexes

A few studies applied reaction pathway analysis to rationalize the experimentally observed enantioselectivities. They compared the energy barriers involved in the hydrolysis or esterification reaction starting from the binding modes of enantiomeric substrates, and evaluated the stabilization of the respective formed intermediates or transition states.

For instance, semi empirical calculations were used to study the selectivity of secondary alcohols esterification and their esters hydrolysis by *Burkholderia cepacia* lipase (BCL) [215]. The model system entailed the substrates and the twenty two amino acids residues that enclose the BCL binding site. Starting from optimized enzyme-substrate complexes for the two enantiomers, the model investigated the hydrogen transfer from histidine to serine oxygen during ester formation, and from histidine to the alcohol oxygen during hydrolysis. Results indicated that for the fast substrate enantiomers all the identified binding modes lead to the formation of products, whereas for the slow reacting substrate enantiomers not all of the binding modes were found productive. Interestingly, the

complexes with the fast reacting enantiomer had lower energy barriers than the ones with the slow reacting enantiomer. The reactivity of binding modes could also be correlated to the number of hydrogen bonds stabilizing the intermediates, namely five for the productive binding modes and only four for the non-productive modes.

Other studies [216, 217] combined semi empirical quantum calculations and molecular dynamics to determine its enantiomeric preference of CALB in hydrolysis and esterification reactions. An illustrated example is the reaction of (R)-naproxen ethyl ester and (S)-naproxen ethyl ester. A simplified semi empirical method was first used to identify the formed transition states and intermediates along the reaction pathway. The hydrogen bonds within the obtained CALB-enantiomer complexes were then analyzed by molecular dynamics. According to the simulations, hydrogen bonds were found more stable for the fast reacting enantiomeric substrates than for the slow reacting enantiomers.

5. Final considerations

In this literature study, we have a significant sample of how much molecular modelling can aid to rationalize the molecular features inherent to lipases. The correct utilization of molecular modelling tools can provide a better understanding, at the atomic level, of the positions, orientations and interactions of substrates in the lipase binding pocket. Although here we focused on lipases, due to its relevance in biotechnology, the modelling methods and analysis criteria can be applied to situations involving other biocatalysts. In general, the microscopic features provided by molecular modelling can be correlated to a previous experimental knowledge on the enantioselectivity or the regioselectivity of the bioconversions. In such cases, the modelling procedure can improve experimental results, by providing a rational understanding of the microscopic events that govern the selectivity properties. When the modelling procedure is employed before experiments, this can aid to predict the enantio-, region- or substrate selectivity in a given lipase-catalyzed reaction. However, some limitations persists nowadays, including mainly difficulties in considering entropy and solvation effects in the modeled systems [99]. Such limitations still represent a barrier to make more quantitative predictions in bioprocesses based on molecular simulations. Despite these shortcomings, trends indicate that molecular modelling is likely to be progressively extended to a wider array of enzymatic reactions, or even become a routine tool in design of biocatalysis experiments, like occurs already in drug design [67, 133].

4.3. Conclusions de l'article de revue

Les nombreuses études publiées sur les réactions d'hydrolyse et de synthèse catalysées par des lipases ont montré que ces enzymes peuvent agir sur une grande diversité de structures de substrats. En présence de mélanges racémiques, les lipases ont souvent une préférence pour l'un des énantiomères. Par ailleurs elles catalysent généralement l'esterification des substrats polyhydroxylés avec une forte régiosélectivité. Pour toutes ces réactions la sélectivité des lipases peut dépendre à la fois de la nature de l'enzyme et des structures des substrats mis en jeu.

Au cours de la dernière décennie un nombre croissant de travaux a décrit l'utilisation de techniques de modélisation moléculaire pour rationaliser l'énantiosélectivité ou la régiosélectivité des réactions catalysées par les lipases. La plupart ont utilisé des techniques de docking de substrats au site actif de l'enzyme. Dans certains d'entre eux, des techniques complémentaires de dynamique moléculaire et de chimie quantique ont également été mises en œuvre. Ces études ont d'abord montré que la modélisation moléculaire pouvait fournir des filtres intéressants pour identifier les molécules ou les groupements fonctionnels sur une même molécule pouvant être transformés par une lipase. Les protocoles de docking basés sur des considérations géométriques et énergétiques sont particulièrement efficaces pour sélectionner les complexes favorables de fixation des substrats dans la cavité de l'enzyme, des complexes dans lesquels les substrats sont positionnés à proximité de la triade catalytique de la lipase et l'énergie d'interaction enzyme-substrat est minimale. La réactivité de ces complexes peut ensuite être évaluée par des différentes approches de calcul quantique (ab initio, semi empiriques ou DFT), qui permettent d'évaluer les processus électroniques menant à la formation ou la rupture de liaisons ainsi que les barrières énergétiques au cours du chemin réactionnel.

Les études de modélisation moléculaire ont également contribué à mieux comprendre les interactions enzyme-substrats qui gouvernent la sélectivité des réactions de bioconversions catalysées par les lipases. Elles mettent clairement en évidence le rôle essentiel de la stabilisation des complexes enzyme-substrat dans l'action catalytique des lipases. Cette stabilisation est en grande partie assurée par un ensemble de liaisons hydrogène stabilisant le substrat accepteur d'acyle à proximité de la triade catalytique et le substrat donneur d'acyle au niveau du trou oxyanionique. Elle peut aussi être renforcée par des interactions hydrophobes, en particulier interactions aromatiques entre les substrats et les résidus présents dans les parois de la cavité catalytique de l'enzyme. Par ailleurs les modélisations quantiques ont également démontré le rôle clé des interactions électrostatiques sur la réactivité des complexes de fixation des substrats. Elles indiquent qu'un réseau suffisant de liaisons hydrogène doit être établi au niveau de la triade pour stabiliser les intermédiaires tétraédriques et ainsi abaisser les énergies d'activation des processus réactionnels mis en jeu.

Si dans le passé les études de modélisation moléculaire ont couvert un ensemble très large de substrats et de réactions, jusqu'à présent aucune n'a porté sur l'acylation enzymatique de flavonoïdes. C'est pourquoi dans cette étude nous avons entrepris une première analyse du potentiel de trois techniques de modélisation, le docking, la dynamique moléculaire et la Théorie Fonctionnelle de la Densité (DFT), pour rationaliser les régiosélectivités de bioconversions de flavonoïdes récemment reportées par notre laboratoire. Plus précisément pour trois flavonoïdes choisis, il s'agit d'établir l'apport de ces techniques de modélisation d'une part à sélectionner les positions des groupements hydroxyle pouvant être acétylées par deux lipases différentes, d'autre part à mieux comprendre les processus d'interactions et de stabilisation intervenant dans le mécanisme d'action de ces enzymes.

5. REFERENCES

- [1] P. Sarni-Manchado, V. Cheynier. *Les polyphénols en agroalimentaire*, Tec & Doc, Paris, 2006.
- [2] B.H. Havsteen. *The biochemistry and the medical significance of flavonoids*. Pharmacology & Therapeutics 96 (2002) 67-202.
- [3] A.M. Boudet. *Evolution and current status of research in phenolic compounds*. Phytochemistry 68 (2007) 2722-2735.
- [4] L. Chebil. *Acylation des flavonoïdes par les lipases de Candida antarctica et de Pseudomonas cepacia : études cinétique, structurale et conformationnelle*. Institut National Polytechnique de Lorraine (thèse), Nancy, 2006.
- [5] J. Anthoni. *Synthèse enzymatique, modélisation moléculaire et caractérisation d'oligomères de flavonoïdes*. Institut National Polytechnique de Lorraine (thèse), Nancy, 2007.
- [6] W.J. Zhang, L.O. Björn. *The effect of ultraviolet radiation on the accumulation of medicinal compounds in plants*. Fitoterapia 80 (2009) 207-218.
- [7] H. Gorou. *Cosmetic for fair-complexioned face containing fatty acid ester of quercetin*. JP 58131911, Japon, 1983.
- [8] J. Vercauteren, J.F. Weber, J.L. Bisson, J. Bignon. *Polyphenol derivatives compositions and preparation thereof*. US 5844061, Etats Unis, 1993.
- [9] M.C. Levy, M.C. Andry. *Microcapsule à paroi de flavonoïde reticule et composition en contenant*. FR 2715582, France, 1994.
- [10] P. Amalia. *Flavonoid compounds and their pharmaceutical uses*. US 20044266699, Etats Unis, 2004.
- [11] B. Halliwell. *How to characterize a biological antioxydant*. Free Radical Research Communications 9 (1990) 1-32.
- [12] H. Schröder, E. Navarro. *Effects and safety of antioxidant supplementation*. Current Topics in Nutraceutical Research 4 (2006) 23-32.
- [13] E.N. Frankel, A.S. Meyer. *The problems of using one-dimensional methods to evaluate multifunctional food and biological antioxidants*. Journal of Science in Food and Agriculture 80 (2000) 1925-1941.
- [14] F. Shahidi. *Natural antioxidants: an overview*, in: F. Shahidi (Ed.). Natural Antioxidants: Chemistry, Health Effects and Applications, AOC Press, Champaing, 1997, 414pp.
- [15] B. Halliwell, R. Aeschbach, J. Löliger, O.I. Aruoma. *The characterization of antioxidants*. Food Chemistry and Toxicology 33 (1995) 601-617.
- [16] P.G. Pietta. *Flavonoids as antioxidants*. Journal of Natural Products 63 (2000) 1035-1042.
- [17] Z.Y. Chen, P.T. Chan, K.Y. Ho, K.P. Fung, J. Wang. *Antioxidant activity of natural flavonoids is governed by number and location of their aromatic hydroxyl groups*. Chemistry and Physics of Lipids 79 (1996) 157-163.
- [18] S. Van Acker, D.J. Van Den Berg, M.N.J.L. Tromp, D.H. Griffioen, W.P.V. Van Bannekom, W.J.F. Van Der Vijgh, A. Bast. *Strucural aspects of antioxydant activity of flavonoids*. Free Radical Biology and Medicine 20 (1996) 331-342.
- [19] C.A. Rice-Evans, N.J. Miller, G. Paganga. *Structure-antioxydant activity relationships of flavonoids and phenolic acids*. Free Radical Biology and Medicine 20 (1996) 933-956.
- [20] G. Cao, E. Sofic, R.L. Prior. *Antioxidant and prooxidant behaviour of flavonoids: structure-activity relationships*. Free Radical Biology and Medicine 22 (1997) 749-760.

- [21] S. Burda, W. Oleszek. *Antioxidant and antiradical activities of flavonoids*. Journal of Agricultural and Food Chemistry 49 (2001) 2774-2779.
- [22] C.M. Lin, C.S. Chen, C.T. Chen, Y.C. Liang, J.K. Lin. *Molecular modelling of flavonoids that inhibits xanthine oxidase*. Biochemistry and Biophysics Research Communications 294 (2002) 167-172.
- [23] S.L. Da Silva, A. Da Silva, K.M. Honorio, S. Marangoni, M.H. Toyama, A.B.S. Da Silva. *The influence of electronic, steric and hydrophobic properties of flavonoid compounds in the inhibition of xanthine oxidase*. Journal of Molecular Structure: THEOCHEM 684 (2004) 1-7.
- [24] I. Limem, E. Guedon, A. Hehn, F. Bourgaud, L. Chekir Ghedira, J.M. Engasser, M. Ghoul. *Production of phenylpropanoid compounds by recombinant microorganisms expressing plant-specific biosynthesis genes*. Process Biochemistry 43 (2008) 463-479.
- [25] H. Uyama, S. Kobayashi. *Enzymatic synthesis and properties of polymers from polyphenols*. Advances in Polymer Science 194 (2006) 51-67.
- [26] M.C. Figueroa-Espinoza, P. Villeneuve. *Phenolic acids enzymatic lipophilization*. Journal of Agricultural and Food Chemistry 53 (2005) 2779-2787.
- [27] L. Chebil, C. Humeau, A. Falcimaigne, J.M. Engasser, M. Ghoul. *Enzymatic acylation of flavonoids* Process Biochemistry 41 (2006) 2237-2251.
- [28] P. Villeneuve. *Lipases in lipophilization reactions*. Biotechnology Advances 25 (2007) 515-536.
- [29] H. Tanagushi, E. Nomura, T. Chikuno, H. Minami. *Antioxidant, cosmetic and new ferulic acid esters*. JP 9040613, Japon, 1997.
- [30] E. Perrier, A.M. Mariotte, A. Boumendjel, D. Bresson-Rival. *Nouveaux esters de flavonoids, leur utilisation en cosmetique, dermopharmacie, pharmacie et agro-alimentaire*. FR 2778663A1, France, 1998.
- [31] S.H. Bok, T.S. Jeong, S.K. Lee, J.R. Kim, S.S. Moon, M.S. Choi. *Flavanone derivatives and composition for preventing or treating blood lipid level-related diseases comprising same*. US 20010006978A1, Etats Unis, 2001.
- [32] R. Wohlgemuth. *The locks and keys to industrial biotechnology*. New Biotechnology (2009).
- [33] S.H. Häkkinen, S.O. Kärenlampi, I. Marina Heinonen, H.M. Mykkänen, A.R. Riitta Törrönen. *Content of the flavonols quercetin, myricetin, and kaempferol in 25 edible berries*. Journal of Agricultural and Food Chemistry 47 (1999) 2274-2279.
- [34] C. Smith, K.A. Lombard, E.B. Peffley, W. Liu. *Genetic Analysis of Quercetin in Onion (Allium cepa L.) 'Lady Raider'*. The Texas Journal of Agriculture and Natural Resource 16 (2003) 24-28.
- [35] K.N. Prasad, B. Yang, X. Dong, G. Jiang, H. Zhang, H. Xie, Y. Jiang. *Flavonoid contents and antioxidant activities from Cinnamomum species*. Innovative Food Science and Emerging Technologies (2009) sous presse.
- [36] T. Iwashina, Y. Omori, J. Kitajima, S. Akiyama, T. Suzuki, H. Ohba. *Flavonoids in translucent bracts of the Himalayan Rheum nobile (Polygonaceae) as ultraviolet shields*. Journal of Plant Research 117 (2004) 101-107.
- [37] I. Orhan, M. Kartal, M. Abu-Asaker, F. Sezer Senol, G. Yilmaz, B. Sener. *Free radical scavenging properties and phenolic characterization of some edible plants*. Food Chemistry 114 (2009) 276-281.
- [38] M.J. Simirgiotis, P.D.S. Caligari, G. Schmeda-Hirschmann. *Identification of phenolic compounds from the fruits of the mountain papaya Vasconcellea pubescens A. DC. grown in Chile by liquid chromatography-UV detection-mass spectrometry*. Food Chemistry 115 (2009) 775-784.

- [39] A. Kampkötter, C. Timpel, R.F. Zurawski, S. Ruhl, Y. Chovolou, P. Proksch, W. Wätjen. *Increase of stress resistance and lifespan of Caenorhabditis elegans by quercetin*. Comparative Biochemistry and Physiology (2008) 314-323.
- [40] F. Zsila, Z. Bikadi, M. Simonyi. *Probing the binding of the flavonoid quercetin to human serum albumin by circular dichroism, electronic adsorption spectroscopy and molecular modelling methods*. Biochemical Pharmacology 65 (2003) 447-456.
- [41] M. Harwood, B. Danielewska-Nikiel, J.F. Borzelleca, G.W. Flamm, G.M. Williams, T.C. Lines. *A critical review of the data related to the safety of quercetin and lack of evidence of in vivo toxicity, including lack of genotoxic/carcinogenic properties* Food and Chemical Toxicology 45 (2007) 2179-2205.
- [42] A.W. Boots, G.R.M.M. Haenen, A. Bast. *Health effects of quercetin: from antioxidant to nutraceutical*. European Journal of Pharmacology 585 (2008) 325-337.
- [43] F. Chimenti, F. Cottiglia, L. Bonsignore, L. Casu, M. Casu, C. Floris, D. Secci, A. Bolasco, P. Chimenti, A. Granese, O. Befani, P. Turini, S. Alcaro, F. Ortuso, G. Trombetta, A. Loizzo, I. Guarino. *Quercetin as the active principle of Hypericum hircinum exerts a selective inhibitory activity against MAO-A: extraction, biological analysis and computational study*. Journal of Natural Products 69 (2006) 945-945.
- [44] J.P. Cornard, J.C. Merlin. *Complexes of aluminium(III) with isoquercitrin: Spectroscopic characterization and quantum chemical calculations* Polyhedron 21 (2002) 2801-2810
- [45] C.G. Silva, R.J. Raulino, D.M. Cerqueira, S.C. Mannarino, M.D. Pereira, A.D. Panek, J.F.M. Silva, F.S. Menezes, E.C.A. Eleutherio. *In vitro and in vivo determination of antioxidant activity and mode of action of isoquercitrin in Hyptis fasciculata*. Phytomedicine 16 (2009) 761-767.
- [46] J. Tabart, C. Kevers, J. Pincemail, J.O. Defraigne, J. Dommes. *Comparative antioxidant capacities of phenolic compounds measured by various tests*. Food Chemistry 113 (2009) 1226-1233.
- [47] T. Sawa, M. Nakao, T. Akaike, K. Ono, H. Maeda. *Alkylperoxyl radical scavenging activity of various flavonoids and other phenolics compounds : implication for the antitumor-promoter effects of vegetables*. Journal of Agricultural and Food Chemistry 47 (1999) 397-402.
- [48] T. Koda, Y. Kuroda, H. Imai. *Protective effect of rutin against spatial memory impairment induced by trimethyltin in rats*. Nutrition Research 28 (2008) 629-634.
- [49] Y. Han. *Rutin has therapeutic effect on septic arthritis caused by Candida albicans*. International Immunopharmacology 9 (2009) 207-211.
- [50] M. Singh, R. Govindarajan, A.K.S. Rawat, P.B. Khare. *Antimicrobial flavonoid rutin from Pteris vittata L. against pathogenic gastrointestinal microflora* American Fern Journal 98 (2008) 98-103.
- [51] L. Chebil, C. Humeau, J. Anthoni, F. Dehez, J.M. Engasser, M. Ghoul. *Solubility of flavonoids in organic solvents*. Journal of Chemical Engineering Data 52 (2007) 1552-1556.
- [52] Advanced Nutri Solutions (ANS). *Certificate of analysis of Isoquercitrin 85% IsoQmax™*. Vista, Canada, 2005.
- [53] M. Ardhaoui, A. Falcimaigne, J.M. Engasser, P. Moussou, G. Pauly, M. Ghoul. *Acylation of natural flavonoids using lipase of Candida antarctica as biocatalyst*. Journal of Molecular Catalysis B: Enzymatic 29 (2004) 63-67.
- [54] L. Chebil, J. Anthoni, C. Humeau, C. Gerardin, J.M. Engasser, M. Ghoul. *Enzymatic acylation of flavonoids : effect of the nature of the substrate, origin of lipase and operating conditions on conversion yield and regioselectivity*. Journal of Agricultural and Food Chemistry 53 (2007) 9496-9502.

- [55] B. Danieli, M. Luisetti, G. Sampognaro, G. Carrea, S. Riva. *Regioselective acylation of polyhydroxylated natural compounds catalyzed by Candida antarctica lipase B (Novozyme 435) in organic solvents*. Journal of Molecular Catalysis B: Enzymatic 3 (1997) 193-201.
- [56] N. Nakajima, K. Ishihara, T. Itoh, T. Furuya, H. Hamada. *Lipase-catalyzed direct and regioselective acylation of flavonoid glucoside for mechanistic investigation of stable plant pigments*. Journal of Bioscience and Bioengineering 87 (1999) 105-107.
- [57] D.E. Stevenson, R. Wibisono, D.J. Jensen, R.A. Stanley, J.M. Cooney. *Direct acylation of flavonoid glycosides with phenolic acids catalyzed by Candida antarctica lipase B (Novozyme 435)*. Enzyme and Microbial Technology 39 (2006) 1236-1241.
- [58] A. Kontogianni, V. Skouridou, V. Sereti, H. Stamatis, F.N. Kolisis. *Regioselective acylation of flavonoids catalyzed by lipase in low toxicity media*. European Journal of Lipid Science and Technology 103 (2001) 655-660.
- [59] A. Kontogianni, V. Skouridou, V. Sereti, H. Stamatis, F.N. Kolisis. *Lipase-catalyzed esterification of rutin and naringin with fatty acid of medium carbon chain*. Journal of Molecular Catalysis B: Enzymatic 21 (2003) 59-62.
- [60] M. Ardhaoui, A. Falcimaigne, S. Ognier, J.M. Engasser, P. Moussou, G. Pauly, M. Ghoul. *Effect of acyl donor chain length and substitutions pattern on the enzymatic acylation of flavonoids*. Journal of Biotechnology 110 (2004) 265-271.
- [61] M. Ardhaoui, A. Falcimaigne, J.M. Engasser, P. Moussou, G. Pauly, M. Ghoul. *Enzymatic synthesis of new aromatic and aliphatic esters of flavonoids using Candida antarctica lipase as biocatalyst*. Biocatalysis and Biotransformation 22 (2004) 253.
- [62] F. Mellou, H. Loutrari, H. Stamatis, C. Roussos, F.N. Kolisis. *Enzymatic esterification of flavonoids with unsaturated fatty acids : effect of the nouvel esters on vascular endothelial growth factor release from K562 cells*. Process Biochemistry 41 (2006) 2029-2034.
- [63] M.H. Katsoura, A.C. Polydera, L. Tsironis, A.D. Tselepis, H. Stamatis. *Use of ionic liquids as media for the biocatalytic preparation of flavonoid derivatives with antioxidant potency*. Journal of Biotechnology 123 (2006) 491-503.
- [64] A.R. Leach. *Molecular Modelling: Principles and Applications*, Prentice Hall, New Jersey, 2001.
- [65] J.G.S. Mala, S. Tekeuchi. *Understanding structural features of microbial lipases - an overview*. Analytical Chemistry Insights 2008 (2008) 9-19.
- [66] H. Toulhoat. *Modélisation moléculaire - Bases théoriques : des propriétés microscopiques aux propriétés macroscopiques*. Techniques de l'Ingénieur J 1-013 (2007).
- [67] D. Joseph-McCarthy. *Computational approaches to structure-based ligand design*. Pharmacology & Therapeutics 84 (1999) 179-191.
- [68] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney. *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Advanced Drug Delivery Reviews 46 (2001) 3-26.
- [69] C.A. Lipinski. *Lead- and drug-like compounds : the rule-of-five revolution*. Drug Discovery Today 1 (2004) 337-341.
- [70] T. Solmajer, J. Zupan. *Optimization algorithms and natural computing in drug discovery*. Today Technologies 1 (2004) 247-252.
- [71] P. Karande, A. Jain, K. Ergun, V. Kisperky, S. Mitragotri. *Design principles of chemical penetration enhancers for transdermal drug delivery*. PNAS 102 (2005) 4688-4693.
- [72] W.J. Geldenhuys, K.E. Gaasch, M. Watson, D.D. Allen, C.J. Van Der Schyf. *Optimizing the use of open-source software applications in drug discovery*. Drug Discovery Today 11 (2006) 127-132.

- [73] P. Braiuca, C. Ebert, A. Basso, P. Linda, L. Gardossi. *Computational methods to rationalize experimental strategies in biocatalysis*. Trends in Biotechnology 24 (2006) 419-425.
- [74] A.J. Mulholland. *Modelling enzyme reaction mechanisms, specificity and catalysis*. Drug Discovery Today 10 (2005) 1393-1402.
- [75] Z. Qian, C.J. Fields, Y. Yu, S. Lutz. *Recent progress in engineering α/β hydrolase-fold family members*. Biotechnology Journal 2 (2007).
- [76] J. Damborsky, J. Brezovsky. *Computational tools for designing and engineering biocatalysts*. Current Opinion in Chemical Biology 13 (2009) 26-34.
- [77] R.R.S. Pissurlenkar, M.S. Shaikh, R.P. Iyer, E.C. Coutinho. *Molecular mechanics force fields and their applications in drug design*. Anti-Infective Agents in Medicinal Chemistry 8 (2009) 128-150.
- [78] A. Soldera. *Modélisation moléculaire des polymères linéaires*. Techniques de l'Ingénieur AF 6-042 (2007).
- [79] M.J. McQuaid, H. Sun, D. Rigby. *Development of COMPASS force field for molécules with aliphatic azide chains*. Journal of Computational Chemistry 25 (2004) 61-71.
- [80] J.R. Maple, M.J. Hwang, T.P. Stockfisch, U. Dinur, M. Waldman, C.S. Ewig, A.T. Hagler. *Derivation of class II force fields: V. Quantum force field for amides, peptides, and related compounds*. Journal of Computational Chemistry 15 (1994) 162-182.
- [81] A.D. MacKerell Jr., D. Bashford, M. Bellot, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus. *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. Journal of Physical Chemistry 102 (1998) 3586-3616.
- [82] B.R. Brooks, C.L. Brooks III, A.D. Mackerell Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus. *CHARMM: The biomolecular simulation program*. Journal of Computational Chemistry 30 (2009) 1545-1614.
- [83] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen. *GROMACS: fast, flexible and free*. Journal of Computational Chemistry 26 (2005) 1701-1718.
- [84] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman. *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 91 (1995) 1-41.
- [85] M. Rueda, C. Ferrer-Costa, T. Meyer, A. Pérez, J. Camps, A. Hospital, J.L. Gelpi, M. Orozco. *A consensus view of protein dynamics*. PNAS 104 (2007) 796-801.
- [86] O. Guvench, A.D. MacKerell Jr. *Comparison of protein force fields for molecular dynamics simulations*. Methods in Molecular Biology 443 (2008) 63-88.
- [87] M.R. Hestenes, E. Stiefel. *Methods of conjugate gradients for solving linear systems*. Journal of Research of the National Bureau of Standards 49 (1952) 409-438.
- [88] J. Nocedal, S.J. Wright. *Numerical Optimization* 2006.

- [89] G.L. Warren, C. Webster Andrews, A.M. Capelli, B. Clark, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head. *A critical assessment of docking programs and scoring functions*. Journal of Medicinal Chemistry 49 (2006) 5912-5931.
- [90] I. Helperin, B. Ma, H. Wolfson, R. Nussinov. *Principles of dockings : an overview of search algorithms and a guide to scoring functions*. PROTEINS: Structure, Function and Bioinformatics 47 (2002) 409-443.
- [91] B. Coupez, R.A. Lewis. *Docking and Scoring - theoretically easy, practically impossible?* Current Medicinal Chemistry 13 (2006) 2995-3003.
- [92] A.R. Leach, B.K. Shoichet, C.E. Peishoff. *Prediction of protein-ligand interactions. Docking and scoring : successes and gaps*. Journal of Medicinal Chemistry 49 (2006) 5851-5855.
- [93] I. Helperin, B. Ma, H. Wolfson, R. Nussinov. *Principles od dockings : an overview of search algorithms and a guide to scoring functions*. PROTEINS : Structure, Function and Bioinformatics 47 (2002) 409-443.
- [94] O. Sperandio, M.A. Miteva, F. Delfaud, B.O. Villoutreix. *Receptor-based computational screening of compounds databases : the main docking-scoring engines*. Current Protein and Peptide Science 7 (2006) 369-393.
- [95] S.F. Sousa, P.A. Fernandes, M.J. Ramos. *Protein-ligand docking : current status and future challenges*. PROTEINS : Structure, Function and Bioinformatics 65 (2006) 15-26.
- [96] B.D. Bursulaya, M. Trotoev, R. Abagyan, C.L. Brooks III. *Comparative study of several algorithms for flexible ligand docking*. Journal of Computer-Aided Molecular Design 17 (2004) 755-763.
- [97] M. Kontoyianni, L.M. McClellan, G.S. Sokol. *Evaluation of docking performance : comparative data on docking algorithms*. Journal of Medicinal Chemistry 47 (2004) 558-565.
- [98] S.F. Sousa, P.A. Fernandes, M.J. Ramos. *Protein-ligand docking : current status and future challenges*. PROTEINS: Structure, Function and Bioinformatics 65 (2006) 15-26.
- [99] N. Moitessier, P. Eglebienne, D. Lee, J. Lawandi, C.R. Corbeil. *Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go*. British journal of Pharmacology 153 (2008).
- [100] J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth. *Lessons in molecular recognition : the effects of ligand and protein flexibility in molecular docking accuracy*. Journal of Medicinal Chemistry 47 (2004) 45-55.
- [101] S.J. Teague. *Implications of protein flexibility for drug discover*. Nature Reviews Drug Discovery 2 (2003) 527-541.
- [102] O. Sperandio, M.A. Miteva, F. Delfaud, B.O. Villoutreix. *Receptor-based computational screening of compounds databases : the main docking-scoring engines*. Current Protein and Peptide Science 7 (2006) 369-393.
- [103] V. Schnecke, C.A. Swanson, E.D. Getzoff, J.A. Tainer, L.A. Kuhn. *Screening a peptidyl database for potential ligands to proteins with side-chains flexibility*. PROTEINS : Structure, Function and Bioinformatics 33 (1998) 74-87.
- [104] M.I. Zavodszky, L.A. Kuhn. *Side-chain flexibility in protein-ligand binding: The minimal rotation hypothesis*. Protein Science 14 (2005) 1104-1114.
- [105] J. Meiler, D. Baker. *ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility*. PROTEINS: Structure, Function, and Bioinformatics 65 (2006) 538-548.
- [106] C.R. Corbeil, P. Englebienne, N. Moitessier. *Docking ligands into flexible and solvated macromolecules -1. Development and validation of FITTED 1.0*. Journal of Chemical Informatic Modelling 47 (2007) 435-449.

- [107] N. Foloppe, R. Hubbard. *Towards predictive ligand design with free-energy based computational methods?* Current Medicinal Chemistry 13 (2006) 3583-3608.
- [108] A. Jain. *Scoring functions for protein-ligand docking.* Current Protein and Peptide Science 7 (2006) 407-420.
- [109] G. Schneider, H.J. Böhm. *Virtual screening and fast automated docking methods.* Combinatorial Chemistry 7 (2002) 64-70.
- [110] T. Schulz-Gasch, M. Stahl. *Scoring functions for protein-ligand interactions : a critical perspective.* Today Technologies 1 (2004) 231-239.
- [111] R.D. Clark, A. Strizhev, J.M. Leonard, J.F. Blake, J.B. Matthew. *Consensus scoring for ligand/protein interactions.* Journal of Molecular Graphics and Modelling 20 (2002) 281-295.
- [112] M. Feher. *Consensus scoring for protein-ligand interactions.* Drug Discovery Today 11 (2006) 421-428.
- [113] X.D. Li, T.J. Hou, X.J. Xu. *Comparative studies of 14 binding free energies scoring functions.* Acta Physico - Chimica Sinica 21 (2005) 504-507.
- [114] M. Stahl, M. Rarey. *Detailed analysis of scoring functions for virtual screening.* Journal of Medicinal Chemistry 44 (2001) 1035-1042.
- [115] R. Teramoto, H. Fukunishi. *Surpevised concensus scoring for docking and virtual screening.* Journal of Chemical Informatics Modelling 47 (2007) 526-534.
- [116] R. Wang, Y. Lu, S. Wang. *Comparative evaluation of 11 scoring functions for molecular docking.* Journal of Medicinal Chemistry 46 (2003) 2287-2303.
- [117] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten. *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry 26 (2005) 1781-1802.
- [118] P. Rabinowitz, P.J. Davis. *Methods of Numerical Integration* 2007.
- [119] M.P. Allen, D. Tildesley. *Computer Simulations of Liquids* 1987.
- [120] M.F. Manning. *Exact solutions of the Schrödinger equation.* Physical Review 48 (1935) 161-164.
- [121] H. Toulhoat. *Modélisation moléculaire - Bases théoriques : de la mécanique moléculaire à la mécanique quantique.* Techniques de l'Ingénieur J 1-011 (2007).
- [122] H. Toulhoat. *Modélisation moléculaire - Bases théoriques : la chimie quantique de Hartree-Fock à Kohn-Sham.* Techniques de l'Ingénieur J 1-012 (2007).
- [123] G. Boucekkine. *Méthodes de la chimie quantique.* Techniques de l'Ingénieur AF 6-050 (2006).
- [124] P. Echenique, J.L. Alonso. *A mathematical and computational review of Hartree-Fock SCF methods in quantum chemistry.* Molecular Physics 105 (2007) 3057-3098.
- [125] A.N. Mayeno, J.L. Robinson, R.S.H. Yang, B. Reisfeld. *Predicting activation enthalpies of cytochrome-p450-mediated hydrogen abstractions. Comparison of semiempirical PM3, SAM1, and AM1 with a density functional theory method* Journal of Chemical Information and Modeling 49 (2009) 1692-1703.
- [126] W. Kohn, A.D. Becke, R.G. Parr. *Density functional theory of electronic structure.* Journal of Physical Chemistry 100 (1996) 12974-12980.
- [127] F. Neese. *Prediction of molecular properties and molecular spectroscopy with density functional theory: from fundamental theory to exchange coupling.* Coordination Chemistry Reviews 253 (2008) 526-563.

- [128] H. Lin, D.G. Truhlar. *QM/MM: What have we learned, where are we, and where do we go from here?* Theoretical Chemistry Accounts 117 (2007) 185-199.
- [129] R.A. Friesner. *Combined quantum and molecular mechanics (QM/MM)*. Drug Discovery Today 1 (2004) 253-260.
- [130] T.C. Bruice. *Computational approaches: reaction trajectories, structures and atomic motions. Enzyme reactions and proficiency*. Chemical Reviews 106 (2006) 3119-3139.
- [131] N. Reuter, A. Dejaegere, B. Maigret, M. Karplus. *Frontier bonds in QM/MM methods: a comparison of different approaches*. Journal of Physical Chemistry A 104 (2000) 1720-1735.
- [132] D.H.J. Mackay, A.J. Cross, A.T. Hagler. *The role of energy minimization in simulation strategies of biomolecular systems*, in: G.D. Fasman (Ed.). Prediction of Protein Structure and the Principles of Protein Conformation, Plenum Press, New York and London, 1990, pp. 317-358.
- [133] H. Alonso, A.A. Bliznyuk, J.E. Gready. *Combining docking and molecular dynamic simulations in drug design*. Medicinal Research Reviews 26 (2006) 531-568.
- [134] P. Geerlings, F. De Proft. *Chemical reactivity as described by quantum chemical methods*. International Journal of Molecular Sciences 3 (2002) 276-309.
- [135] R.A. Friesner, V. Guallar. *Ab initio quantum mechanics and mixed quantum mechanics / molecular mechanics (QM/MM) methods for studying enzymatic catalysis*. Annual Review of Physical Chemistry 56 (2005) 389-427.
- [136] E. Codorniu-Hernandez, A. Mesa-Ibárrico, R. Hernandez-Santesteban, L. Montero-Cabrera, F. Martínez-Luzardo, J.L. Santana-Romero, T. Borrmann, W.D. Stohrer. *Essential amino acids interacting with flavonoids: a theoretical approach*. International Journal of Quantum Chemistry 103 (2005) 82-104.
- [137] E. Codorniu-Hernandez, A. Rolo-Naranjo, L. Montero-Cabrera. *Theoretical affinity order among flavonoids and amino acid residues: an approach to understand flavonoid-protein interactions*. Journal of Molecular Structure: THEOCHEM 819 (2007) 121-129.
- [138] A.M. Davis, S.J. Teague. *Hydrogen bonding, hydrophobic interactions and the failure of the rigid receptor hypothesis*. Angewandte Chemie - International Edition 38 (1999) 736-749.
- [139] C.A. Hunter, J.K.M. Sanders. *The nature of π - π interactions*. Journal of American Chemical Society 112 (1990) 5525-5534.
- [140] E.A. Meyer, R.K. Castellano, F. Diederich. *Interactions with aromatic rings in chemical and biological recognition*. Angewandte Chemie - International Edition 42 (2003) 1210-1250.
- [141] D.H. Williams, E. Stephens, D.P. O'Brien, M. Zhou. *Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand induced reductions in motion within receptors and enzymes*. Angewandte Chemie - International Edition 43 (2004) 6596-6616.
- [142] S.H. Park, B.G. Kim, S.H. Lee, Y. Lim, Y. Cheong, J.H. Ahn. *Molecular modelling and site directed mutagenesis of the O-methyltransferase SOMT-9 reveal amino acids important for its reaction and regioselectivity*. Bulletin of Korean Chemical Society 28 (2007) 2248-2252.
- [143] B.G. Kim, Y.J. Lee, S. Lee, Y. Lim, Y. Cheong, J.H. Ahn. *Altered regioselectivity of a poplar O-methyltransferase POMT-7*. Journal of Biotechnology 138 (2008) 107-111.
- [144] S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. *Molecular simulations reveal a new entry site in quercetin-2,3-dioxygenase. A pathway for dioxygen?* PROTEINS: Structure, Function & Bioinformatics 64 (2006) 845-850.
- [145] S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. *Molecular Simulations Bring New Insights Into Flavonoid/Quercetinase Interaction Modes*. PROTEINS: Structure, Function & Bioinformatics 67 (2007) 961-970.

- [146] S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass. *Theoretical investigations of the role played by quercetinase enzymes upon the flavonoids oxygenolysis mechanism*. Physical Chemistry Chemical Physics 11 (2009) 1491-1501.
- [147] S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. *Molecular simulations enlighten the binding mode of quercetin to lipoxygenase-3*. PROTEINS: Structure, Fonction & Bioinformatics 73 (2008) 290-298.
- [148] L. Chen, J. Li, C. Luo, H. Liu, W. Xu, G. Chen, O.W. Liew, W. Zhu, C.M. Puah, X. Shen, H. Jiang. *Binding interaction of quercetin-3- β -galactoside and its synthetic derivatives with SARS-CoV 3CL^{pro}: Structure-activity relationship studies reveal salient pharmacophore features*. Bioorganic & Medicinal Chemistry 14 (2006) 8295-8306.
- [149] M. Mozzicafreddo, M. Cuccioloni, A.M. Eleuteri, E. Fioretti, M. Angeletti. *Flavonoids inhibit the amidolytic activity of human thrombin*. Biochimie 88 (2006) 1297-1306.
- [150] R. Badhan, J. Penny. *In silico modelling of the interaction of flavonoids with P-glycoprotein nucleotide-binding domain*. European Journal of Medicinal Chemistry 41 (2006) 285-295.
- [151] L.M. Kabeya, A.A. De Marchi, A. Kanashiro, N.P. Lopes, C.H.T.P. Da Silva, M.T. Pupo, Y.M. Lucisano-Valim. *Inhibition of horseradish peroxidase catalytic activity by new 3-phenylcoumarin derivatives: synthesis and structure-activity relationships*. Bioorganic & Medicinal Chemistry 15 (2007) 1516-1524.
- [152] J. Lättig, M. Böhl, P. Fischer, S. Tischer, C. Tietböhl, M. Menschikowski, H.O. Gutzeit, P. Metz, M. Teresa Pisabarro. *Mechanism of inhibition of human secretory phospholipase A2 by flavonoids: rationale for lead design*. Journal of Computer Aided Molecular Design 21 (2007) 473-483.
- [153] Y. Shimmyo, T. Kihara, A. Akaike, T. Niidome, H. Sugimoto. *Three distinct neuroprotective functions of myricetin against glutamate-induced neuronal cell death: involvement of direct inhibition of caspase-3*. Journal of Neuroscience Research 86 (2008) 1836-1845.
- [154] Y. Shimmyo, T. Kihara, A. Akaike, T. Niidome, H. Sugimoto. *Flavonols and flavones as BACE-1 inhibitors: structure-activity relationship in cell-free, cell-based and in silico studies reveal novel pharmacophore features*. Biochimica et Biophysica Acta 1780 (2008) 819-825.
- [155] M. Carlquist, T. Fredj, M.F. Gorwa-Graslund. *Flavonoids as inhibitors of human carbonyl reductase 1*. Chemico-Biological Interactions 174 (2008) 98-108.
- [156] M. Brunskole, K. Zorko, V. Kerbler, S. Martens, J. Stojan, S. Gobec, T. Lanisnik Rizner. *Trihydroxynaphthalene reductase of *Curvalaria lunata* - a target for flavonoid action?* Chemico-Biological Interactions 178 (2009) 259-267.
- [157] P. Adlercreutz. *Enzyme-catalyzed lipid modification*. Biotechnology and Genetic Engineering Reviews 12 (1994) 231-254.
- [158] K.E. Jaeger, M.T. Reetz. *Microbial lipases form versatile tools for biotechnology*. Trends in Biotechnology 16 (1998) 396-403.
- [159] K.E. Jaeger, T. Eggert. *Lipases for biotechnology*. Current Opinion in Biotechnology 13 (2002) 390-397.
- [160] F. Hasan, A.A. Shah, A. Hameed. *Industrial applications of microbial lipases*. Enzyme and Microbial Technology 39 (2006) 235-251.
- [161] K. Hult, P. Berglund. *Enzyme promiscuity: mechanism and applications*. Trends in Biotechnology 25 (2007) 231-238.
- [162] G.D. Haki, S.K. Rakshit. *Developments in industrially important thermostable enzymes : a review*. Bioresource Technology 89 (2003) 17-34.
- [163] F.W. Paques, G. Alves Macedo. *Plant lipases from latex: properties and industrial applications*. Quimica Nova 29 (2006) 93-99.

- [164] Y.H. Ju, S.R. Vali. *Rice bran oil as a potential resource for biodiesel: a review*. Journal of Scientific and Industrial Research 64 (2005) 866-882.
- [165] J.M. Marchetti, V.U. Miguel, A.F. Errazu. *Possible methods for biodiesel production*. Renewable and Sustainable Energy Reviews 11 (2007) 1300-1311.
- [166] S.V. Ranganathan, S.L. Narasimhan, K. Muthukumar. *An overview of enzymatic production of biodiesel*. Bioresource Technology 99 (2008) 3975-3981.
- [167] I.K. Varma, A.-C. Albertsson, R. Rajkhowa, R.K. Srivastava. *Enzymatic catalyzed synthesis of polyesters*. Progress in Polymer Science 30 (2005) 949-981.
- [168] S. Matsumura. *Enzymatic synthesis of polyesters via ring-opening polymerization*. Enzyme-catalyzed synthesis of polymers vol. 194, Springer Berlin / Heidelberg, 2006, pp. 95-132.
- [169] H. Uyama, S. Kobayashi. *Enzymatic synthesis of polyesters via polycondensation*. Enzyme-catalyzed synthesis of polymers. vol. 194, Springer Berlin / Heidelberg, 2006, pp. 133-158.
- [170] P. Bajpai. *Enzymatic deinking*. Advances in Applied Microbiology 45 (1997) 241-269.
- [171] P. Bajpai. *Application of enzymes in pulp and paper industry*. Biotechnology Progress 15 (1999) 147-157.
- [172] A. Gutiérrez, J.C. Del Rio, M.J. Martinez, A.T. Martinez. *The biotechnological control of pitch in paper pulp manufacturing*. Trends in Biotechnology 19 (2001) 340-348.
- [173] S. Serra, C. Fuganti, E. Brenna. *Biocatalytic preparation of natural flavours and fragrances*. Trends in Biotechnology 23 (2005) 193-198.
- [174] S.R. Couto, M.A. Sanromán. *Application of solid-state fermentation to food industry: a review*. Journal of Food Engineering 76 (2006) 297-302.
- [175] A. Joshi, S.G. Paratkar, B.N. Thorat. *Modification of lecithin by physical, chemical and enzymatic methods*. European journal of Lipid Science and Technology 108 (2006) 363-373.
- [176] R. Aravindan, P. Anbumathi, T. Viruthagiri. *Lipase applications in food industry*. Indian Journal of Biotechnology 6 (2007) 141-158.
- [177] A.A. Mendes, H.F. De Castro, E.B. Pereira, A. Furigo Jr. *Application of lipases for wastewater treatment containing high levels of lipids*. Quimica Nova 28 (2005) 296-305.
- [178] M.C. Cammarota, D.M.G. Freire. *A review on hydrolytic enzymes in the treatment of wastewater with high oil and grease content*. Bioresource Technology 97 (2006) 2195-2210.
- [179] T.P. Rugabber, J.W. Talley. *Enhancing bioremediation with enzymatic processes: a review*. Practice Periodical of Hazardous, Toxic and Radioactive Waste Management 10 (2006) 73-85.
- [180] W. Bonrath, R. Karge, T. Netscher. *Lipase-catalyzed transformations as key-steps in the large-scale preparation of vitamins*. Journal of Molecular Catalysis B : Enzymatic 19-20 (2002) 67-72.
- [181] V. Gotor-Fernandez, R. Brieva, V. Gotor. *Lipases : useful biocatalysts for the preparation of pharmaceuticals*. Journal of Molecular Catalysis B : Enzymatic 40 (2006) 111-120.
- [182] A. Ghanem, H.Y. Aboul-Enein. *Application of lipase in kinetic resolution of racemates*. Chirality 17 (2005) 1-15.
- [183] J. Aleu, A.J. Bustillo, R. Hernandez-Galan, I.G. Collado. *Biocatalysis applied to the synthesis of agrochemicals*. Current Organic Synthesis 10 (2006) 2037-2054.
- [184] R. Kourist, P. Dominguez de Maria, U.T. Bornscheuer. *Enzymatic synthesis of optically active tertiary alcohols: expanding the biocatalysis toolbox*. ChemBioChem 9 (2008) 941-948.

- [185] O. Parve, I. Jarving, I. Martin, A. Metsala, I. Vallikivi, M. Aidnik, T. Pehk, N. Samel. *Lipase-catalyzed acylation of prostaglandins*. Bioorganic and Medicinal Chemistry Letters 9 (1999) 1853-1858.
- [186] F.J. Plou, M. Cruces Angeles, M. Ferrer, G. Fuentes, E. Pastor, M. Bernabé, M. Christensen, F. Comelles, J.L. Parra, A. Ballesteros. *Enzymatic acylation of di and trisaccharides with fatty acids : chosing the appropriate enzyme, support and solvent*. Journal of Biotechnology 96 (2002) 55-66.
- [187] I. Vallikivi, U. Lille, A. Lookene, A. Metsala, P. Sikk, V. Tougu, H. Vija, L. Villo, O. Parve. *Lipase action on some non-triglyceride substrates*. Journal of Molecular Catalysis B : Enzymatic 22 (2003) 279-298.
- [188] K.E. Jaeger, B.W. Dijkstra, M.T. Reetz. *Bacterial Biocatalysts : Molecular Biology, Three-Dimensional Structures, and Biotechnological Applications of Lipases*. Annual Review of Microbiology 53 (1999) 315-351.
- [189] J. Pleiss, M. Fischer, M. Peiker, C. Thiele, R.D. Schmid. *Lipase engineering database: understanding and exploiting sequence - structure - function relationships*. Journal of Molecular Catalysis B: Enzymatic 10 (2000) 491-508.
- [190] J. Pleiss, M. Fischer, R.D. Schmid. *Anatomy of lipase binding sites : the scissile fatty acid binding site*. Chemistry and Physics of Lipids 93 (1998) 67-80.
- [191] R. Verger. *'Interfacial activation' of lipases : facts and artifacts*. Tibtech 15 (1997) 32-38.
- [192] A. Svendsen. *Lipase protein engineering*. Biochimica et Biophysica Acta 1543 (2000) 223-238.
- [193] J. Uppenberg, M.T. Hansen, S. Patkar, T.A. Jones. *The sequence, crystal structure determination and refinement of two forms of lipase B from Candida antarctica*. Structure 2 (1994) 293-308.
- [194] J. Uppenberg, N. Ohrner, M. Norin, K. Hult, G.J. Kleywegt, S. Patkar, V. Waagen, T. Anthonsen, T.A. Jones. *Cristallographic and molecular-modeling studies od lipase B from Candida antarctica reveal a stereospecificity pocket for secondary alcohols*. Biochemistry 34 (1995) 16838-16851.
- [195] A. Magnusson. *Rational redesign of Candida antarctica lipase B*. Department of Biochemistry, Royal Institute of Technology, Stockholm, 2005, 54.
- [196] V. Prelog. *Specifiation of the stereospecificity of some oxidoreductases by diamond lattice sections*. Pure and Applied Chemistry 9 (1964) 119.
- [197] R.J. Kazlauskas, A.N.E. Weissfloch, T. Rappaport, L.A. Cuccia. *A rule to predict which enantiomer of a secondary alcohol reacts faster in reactions catalyzed by cholesterol esterase, lipase from Pseudomonas cepacia, and lipase from Candida rugosa*. Journal of Organic Chemistry 56 (1991) 2656-2665.
- [198] C. Orrenius, F. Haeffner, D. Rotticci, N. Ohrner, T. Norin, K. Hult. *Chiral recognition of alcohol enantiomers in acyl transfer reactions catalyzed by Candida antarctica lipase B*. Biocatalysis and Biotransformation 16 (1998) 1-15.
- [199] S. Raza, L. Fransson, K. Hult. *Enantioselectivity in Candida antarctica lipase B : a molecular dynamics study*. Protein Science 10 (2001) 329-338.
- [200] L.F. Solares, I. Lavandera, V. Gotor-Fernandez, R. Brieva, V. Gotor. *Biocatalytic preparation of enantioenriched 3,4-dihydroxypiperidines and theoretical study of Candida antarctica lipase B enantioselectivity*. Tetraedron 62 (2006) 3284-3291.
- [201] V. Léonard, L. Fransson, S. Lamare, K. Hult, M. Graber. *A water molecule in the stereospecificity pocket of Candida antarctica lipase B enhances enantioselectivity towards pentan-2-ol*. ChemBioChem 8 (2007) 662-667.

- [202] J. Nyhlén, B. Martin-Matute, A.G. Sandström, M. Bocola, J.E. Bäckvall. *Influence of δ -functional groups on the enantioselectivity of secondary alcohols by Candida antarctica lipase B*. ChemBioChem 9 (2008) 1968-1974.
- [203] W. Li, B. Yang, Y. Wang, D. Wei, C. Whiteley, X. Wang. *Molecular modelling of substrate selectivity of Candida antarctica lipase B and Candida rugosa lipase towards c9,t11- and t10,c12-conjugated linoleic acid*. Journal of Molecular Catalysis B: Enzymatic (2009).
- [204] S. Mittal, S. Khanna, A. Roy, P.V. Bharatam, H.P.S. Chawla. *Candida rugosa lipase mediated multigram synthesis of acid part of S(+)-atliprofen, a new NSAID and molecular modelling studies aimed at predicting selectivity of the enzyme*. Enzyme and Microbial Technology 36 (2005) 232-238.
- [205] B.Y. Hwang, H. Scheib, J. Pleiss, B.G. Kim, R.D. Schmid. *Computer aided molecular modelling of the enantioselectivity of Pseudomonas cepacia lipase toward γ and δ -lactones*. Journal of Molecular Catalysis B: Enzymatic 10 (2000) 223-231.
- [206] A. Tafi, A. Van Almsick, F. Corelli, M. Crusco, K.E. Laumen, M.P. Schneider, M. Botta. *Computer simulations of enantioselective ester hydrolysis catalyzed by Pseudomonas cepacia lipase*. Journal of Organic Chemistry 65 (2000) 3659-3665.
- [207] D. Guiesse, C. Salagnad, P. Monsan, M. Remaud-Simeon, V. Tran. *Towards a novel explanation of Pseudomonas cepacia lipase enantioselectivity via molecular modelling of the enantiomer trajectory into the active site*. Tetrahedron: Asymmetry 14 (2003) 1807-1817.
- [208] R.T. Otto, H. Scheib, U.T. Bornscheuer, J. Pleiss, C. Syldatk, R.D. Schmid. *Substrate specificity of lipase B from Candida antarctica in the synthesis of arylaliphatic glycolipids*. Journal of Molecular Catalysis B: Enzymatic 8 (2000) 201-211.
- [209] I. Vallikivi, L. Fransson, K. Hult, I. Järving, T. Pehk, N. Samel, V. Tõugu, L. Villo, O. Parve. *The modelling and kinetic investigation of the lipase-catalysed acetylation of stereoisomeric prostaglandins*. Journal of Molecular Catalysis B : Enzymatic 35 (2005) 62-69.
- [210] G. Fuentes, M. Cruces Angeles, F.J. Plou, A. Ballesteros, C.S. Verma. *Computational studies of subtilisin-catalyzed transesterification of sucrose: importance of entropic effects*. ChemBioChem 9 (2002) 907-910.
- [211] G. Fuentes, A. Ballesteros, C.S. Verma. *Enthalpic and entropic contributions in the transesterification of sucrose: computational study of lipases and subtilisin*. Journal of Biomolecular Structure and Dynamics 25 (2007) 145-155.
- [212] C. Palocci, M. Falconi, S. Alcaro, A. Tafi, R. Puglisi, F. Ortuso, M. Botta, L. Alberghina, E. Cernia. *An approach to address Candida rugosa lipase regioselectivity in the acylation reactions of trytilated glycosides*. Journal of Biotechnology 128 (2007) 908-918.
- [213] B. Botta, G. Zappia, A. Tafi, M. Botta, F. Manetti, E. Cernia, G. Milana, C. Palocci, S. Soro, G. Delle Monache. *Lipase-catalyzed regioselective acylation of resorcin[4]arenes*. Journal of Molecular Catalysis B : Enzymatic 16 (2002) 241-247.
- [214] C.H. Hu, T. Brinck, K. Hult. *Ab initio and density functional theory studies of the catalytic mechanism for ester hydrolysis in serine hydrolases*. International Journal of Quantum Chemistry 69 (1998) 89-103.
- [215] S. Tomic, M. Ramek. *Quantum mechanical study of Burkholderia cepacia lipase enantioselectivity*. Journal of Molecular Catalysis B: Enzymatic 28 (2006) 139-147.
- [216] C.H. Kwon, J.Y. Jeong, J.W. Kang. *Molecular modelling and experimental verification of lipase-catalyzed enantioselective esterification of racemic naproxen in supercritical carbon dioxide*. Korean Journal of Chemical Engineering 26 (2009) 214-219.

- [217] C.H. Kwon, D.Y. Shin, J.H. Lee, S.W. Kim, J.W. Kang. *Molecular modelling and its experimental verification for the catalytic mechanism of Candida antarctica lipase B*. Journal of Microbiology and Biotechnology 17 (2007) 1098-1105.

CHAPITRE II :

Matériels & méthodes de simulation

1.

RESSOURCES INFORMATIQUES

1.1. Ordinateurs

Deux ordinateurs ont été utilisés pour l'exécution et l'analyse des simulations :

- Un processeur Intel Core Duo CPU 6550 (2,33 GHz), équipé avec 2 Go de mémoire RAM.
- Un bi-processeur AMD Dual Core 280 (2,4 GHz), équipé avec 8 Go de mémoire RAM.

1.2. Logiciels de simulation moléculaire

- *Discovery Studio 2.0 (Accelrys, Inc)* : ce logiciel comporte de nombreux modules de simulation moléculaire voués à l'étude structurale des protéines et acides nucléiques, ainsi que de leurs interactions avec d'autres molécules. Dans la présente étude, ont été employés : le champ de forces CHARMM [1] pour toutes les simulations de mécanique moléculaire et dynamique moléculaire ; le module *LigandFit* [2] pour les simulations de docking ; le *Consensus score* [3, 4] entre 6 fonctions de score pour le classement des complexes obtenus (cf. Annexe A1).
- *Materials Studio 4.4 (Accelrys, Inc)* : ce logiciel est interfacé avec un vaste ensemble de modules permettant l'étude des propriétés physiques et de la réactivité chimique de matériaux solides, cristaux et biomolécules. Dans ce travail, les calculs quantiques ont été exécutés avec le module *DMol³* [5], qui utilise la Théorie Fonctionnelle de la Densité (DFT) [6] pour la description électronique des systèmes moléculaires.

2. STRUCTURES DE DEPART

2.1. Structures des flavonoïdes

Les molécules de rutine, isoquercitrine et quercétine (Figure 3) ont été construites en utilisant les outils du module *Builder* de Discovery Studio. Comme l'algorithme de docking de *LigandFit* explore l'espace conformationnel des ligands de manière randomisée (via une approche Monte Carlo) [2], la conformation de départ des ligands peut être prise au hasard. Ainsi, Les structures des flavonoïdes ont été directement utilisées pour les simulations de docking.

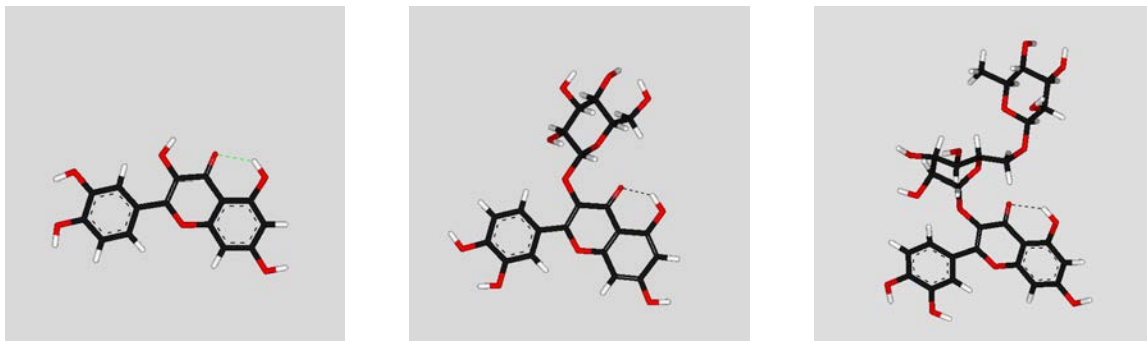


Figure II1 - Structures des flavonoïdes étudiés. Dans l'ordre : la quercétine, l'isoquercitrine (quercétine-O-3-glucose) et la rutine (quercétine-O-3-glucose-rhamnose).

2.2. Structures des lipases

2.2.1. Structures cristallographiques

2.2.1.1. La lipase B de *Candida antarctica* (CALB)

Cinq structures tridimensionnelles pour la lipase B de *Candida antarctica* sont disponibles sur la PDB, identifiées par les codes 1TCA, 1TCB, 1TCC, 1LBS et 1LBT. La structure 1LBS [7] a été choisie pour notre étude, car elle constitue un compromis entre une bonne résolution (2,10 Å) et la présence d'un inhibiteur phosphonate co-cristallisé, lié de manière covalente à la serine catalytique. En effet, ceci permet d'avoir des informations d'origine expérimentale quant au mode d'ancrage d'un ligand, notamment son orientation vis-à-vis du trou oxyanionique. Bien que la structure 1TCA soit déterminée avec une meilleure résolution (1,55 Å), elle n'a pas de ligand co-cristallisé dans la cavité du site actif. En outre, le *rmsd* résultant de la superposition de ces deux structures (1TCA et 1LBS) est proche de zéro, indiquant qu'elles ne présentent pas de différence conformationnelle significative.

Le cristal 1LBS est constitué de six chaînes par cellule, dont les dimensions sont $a = 229,5 \text{ Å}$, $b = 95,6 \text{ Å}$, $c = 86,8 \text{ Å}$; $\alpha = \beta = \gamma = 90^\circ$. Chaque chaîne est composée de 317 acides aminés, 92 molécules d'eau absorbées, un inhibiteur éthyl-hexyl-phosphonate (HEE900) lié de manière covalente à la serine catalytique (Ser105) et un dimère de N-acétyl-glucosamine (NAG400 et NAG401) lié au résidu Asn74. Cette structure a été modifiée pour les simulations comme suit :

- (1) Seule la première des six chaînes a été conservée ; les autres ont été effacées.
- (2) L'inhibiteur éthyl-hexyl-phosphonate (HEE900) et le dimère de N-acétyl-glucosamine (NAG400 et NAG401) ont été effacés de la chaîne restante.
- (3) les molécules d'eau adsorbées ont été conservées.
- (4) Les résidus catalytiques Asp187 et His224 ont été assignés comme déprotonés.
- (5) Les atomes d'hydrogène manquants ont été ajoutés dans les positions théoriques.
- (6) L'ensemble a été centré dans une cellule de dimensions pré optimisées ($a = 64 \text{ Å}$, $b = 59 \text{ Å}$, $c = 63 \text{ Å}$; $\alpha = \beta = \gamma = 90^\circ$) et solvaté avec 6145 molécules d'eau. Le système ainsi obtenu (Figure II2a) a été le point de départ pour l'étape suivante.

2.2.1.2. La lipase de *Pseudomonas cepacia* (PCL)

Pour la lipase de *Pseudomonas cepacia* (aussi appelée *Burkholderia cepacia*), neuf structures tridimensionnelles sont disponibles sur la PDB, identifiées par les codes 1HQD, 1OIL, 2NW6, 1YS1, 1YS2, 2LIP, 3LIP, 4LIP, 5LIP. La PCL est équipée d'un clapet recouvrant le site actif, pouvant donc se présenter sous les formes « ouverte » et « fermée ». Le but étant d'étudier l'ancrage de ligands, la forme ouverte obtenue avec la meilleure résolution (2.00 Å) a été choisie pour notre étude, correspondant à la structure de code 3LIP [8].

La cellule du cristal 3LIP est constituée d'une seule chaîne et possède les dimensions $a = 91,3 \text{ Å}$, $b = 47,3 \text{ Å}$, $c = 85,4 \text{ Å}$; $\alpha = 90,0^\circ$, $\beta = 121,4^\circ$, $\gamma = 90^\circ$. Cette chaîne est composée de 320 acides aminés, 193 molécules d'eau adsorbées et un cation de calcium (Ca^{2+}) complexé par le carboxylates des chaînes latérales des résidus Asp242 et Asp288. Cette structure a été modifiée comme suit :

- (1) Le cation Ca^{2+} a été effacé.
- (2) Les molécules d'eau adsorbées ont été conservées.
- (3) Les résidus catalytiques Asp264 et His286 ont été assignés comme déprotonés.
- (4) Les atomes d'hydrogène manquants ont été ajoutés dans les positions théoriques.
- (5) L'ensemble a été centré dans une cellule de dimensions pré optimisées ($a = 60 \text{ \AA}$, $b = 56 \text{ \AA}$, $c = 60 \text{ \AA}$; $\alpha = \beta = \gamma = 90^\circ$) et solvaté avec 4873 molécules d'eau. Le système ainsi obtenu (Figure II2b) a été le point de départ pour l'étape suivante.

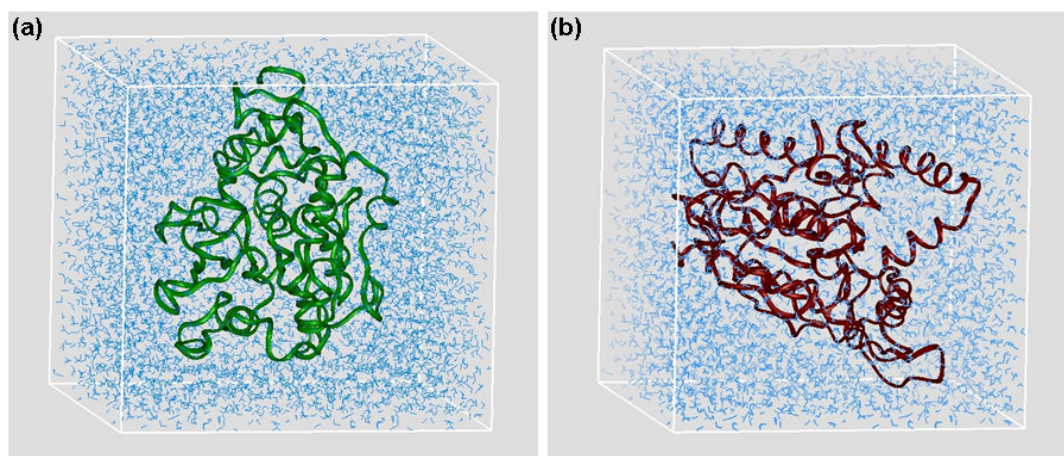


Figure II2 - Les cellules contenant la (a) CALB et la (b) PCL, utilisées dans la procédure de préparation des cibles de docking.

2.2.2. Relaxation des structures cristallographiques

La structure la plus stable d'une protéine cristallisée peut présenter des différences par rapport à sa structure en solution. Pour cette raison, pour que les résultats de simulation reproduisent au mieux le réel, une procédure de relaxation de structure doit être menée au préalable. Le but est de permettre à la protéine de retrouver une configuration structuralement et énergétiquement stable en solution, sans que des distorsions trop importantes soient pour autant infligées à la structure qu'elle présentait initialement au sein du cristal.

Bien que l'acylation de flavonoïdes soit un bioprocédé mené en milieux non aqueux, nous avons utilisé des molécules d'eau comme solvant explicite lors des simulations. Ce choix, à première vue contradictoire, se justifie par la connaissance préalable des particularités des systèmes étudiés. En effet, D'une part, les études expérimentales de l'équipe de Kilbanov ont montré que dans des solvants organiques, la plupart des enzymes dont les lipases conservent une couche de molécules d'eau de solvation. Cette couche d'eau serait un élément stabilisateur de leur structure et serait donc essentielle pour le maintien de l'activité enzymatique hors du milieu naturel de l'enzyme. D'autre part, les travaux théoriques de l'équipe de Pleiss ont suggéré que la CALB ne subit que de faibles variations structurales dans différents solvants. En outre, ces petites

variations concernent des régions éloignées du site actif. Des simulations de dynamique moléculaire en présence de molécules explicites d'eau, de méthanol, de cyclohexane, de chloroforme, de isopentane de toluène ont montré que, dans tous les cas le rmsd maximal de la protéine, après stabilisation, est d'environ 0,9 Å - 1,0 Å. Ces données nous ont motivé pour utiliser dans un premier temps des molécules d'eau pour les simulations, car celles-ci sont correctement prises en compte par le champ de forces.

La relaxation des structures a été faite selon la stratégie proposée par MacKay *et al.* [9], qui comprend une série de minimisations d'énergie suivie d'une équilibration du système par dynamique moléculaire, comme suit :

- (1) Les systèmes contenant les protéines solvatées (Figure 4) ont été soumis à quatre étapes successives de minimisation d'énergie, dans lesquelles des contraintes de mouvement ont été imposées à la protéine et ensuite enlevées de manière progressive. Les paramètres de simulation sont donnés dans le Tableau II1.

Tableau II1- Paramètres de simulation employés dans les étapes de minimisation d'énergie du système contenant les protéines solvatées

Étape	Contraintes	Algorithme	Critère d'arrêt
1	▪ Contrainte <i>fix</i> appliquée à tous les atomes lourds ($\neq$ H) de la protéine ▪ Atomes de H et molécules d'eau libres	Steepest Descent [10]	1000 itérations
2	▪ Contrainte <i>fix</i> appliquée aux atomes constituant le backbone de la protéine ▪ Contrainte <i>harmonic</i> (mouvement restreint) appliquée aux atomes des chaînes latérales des acides aminés de la protéine ▪ Molécules d'eau et atomes de H libres	Conjugate Gradient [11]	Gradient de 0.1 kcal.mol ⁻¹ . Å ⁻¹
3	▪ Contrainte <i>harmonic</i> appliquée aux atomes du backbone de la protéine ▪ Chaînes latérales des acides aminés de la protéine, molécules d'eau et atomes de H libres	Conjugate Gradient	Gradient de 0.01 kcal.mol ⁻¹ . Å ⁻¹
4	▪ Aucune contrainte appliquée au système (tous les atomes libres)	Conjugate Gradient	Gradient de 0.001 kcal.mol ⁻¹ . Å ⁻¹

- (2) Les systèmes issus de la quatrième étape de minimisation ont été soumis à une trajectoire standard de dynamique moléculaire. Ces simulations ont été faites sous conditions périodiques (PBC), en adoptant l'ensemble canonique (NVT), selon les paramètres suivants :
 - Température = 300 K
 - Algorithme d'intégration : algorithme de Verlet-Leapfrog [12] (cf. Annexe A2)
 - *Cut-off* pour les interactions non liées = 12 Å (sphérique)
 - Pas d'intégration (Δt) = 1 fs (10⁻¹⁵s)
 - Fréquence de mise à jour des vitesses atomiques = 50 fs

- Fréquence d'enregistrement de la configuration du système = 1 ps (10^{-12} s)
- Temps de trajectoire = 1 ns (10^{-9} s)

(3) Pour chaque lipase, le dernier *frame* de la trajectoire de dynamique moléculaire a été sélectionnée pour la construction de l'acétyl-enzyme.

2.2.3. Construction des structures des acétyl-lipases

Dans le mécanisme de transestérification accepté pour les lipases [13], le premier substrat (donneur d'acyle) se lie d'abord à la sérine catalytique de l'enzyme, formant l'acyle-lipase. Ensuite, le deuxième substrat (accepteur d'acyle) réagit avec l'acyle-lipase formé. Ainsi, au cours de la réaction d'acétylation, le flavonoïde s'associe à la lipase préalablement acétylée (acétyl-lipase). Pour cette raison, les structures de acétyl-CALB et acétyl-PCL ont été utilisées comme cible de docking des flavonoïdes. Ces structures ont été préparées comme suit :

- (1) Les structures des acétyl-lipases ont été construites en remplaçant le groupement hydroxyle de la chaîne latérale de la sérine catalytique (Ser105 de la CALB et Ser87 de la PCL) par un groupement acétyl.
- (2) Pour optimiser la position de l'acétate, l'acétyl-CALB et l'acétyl-PCL ont été soumis à 1000 étapes de minimisation supplémentaires (Conjugate Gradient), avec une contrainte *fix* appliquée sur les *backbones* des protéines.
- (3) Les molécules d'eau situées dans la cavité catalytique des lipases ont été enlevées et les deux structures d'acétyl-lipases optimisées ont été utilisées comme cible pour le docking des flavonoïdes dans l'étape suivante.

3. SIMULATIONS DE DOCKING

La procédure de docking comprend successivement les étapes de définition du site actif, de génération et sélection des complexes (le docking proprement dit), de classement des complexes sélectionnés et d'optimisation des complexes retenus après le classement.

3.1. La définition du site actif

3.1.1. Principe

Le module LigandFit comprend un outil pour la détection de cavités dans les protéines. Cet outil se base sur un système de grilles (Figure II3) et fonctionne en deux étapes successives :

- (1) D'abord, une grille tridimensionnelle d'espacement fixé à 0,5 Å, ayant les mêmes dimensions maximales de la protéine selon les axes x, y et z, est construite au travers de la structure de la molécule. Pour chaque point de la grille, si la distance à l'atome lourd ($\neq$ hydrogène) de la protéine le plus proche est inférieure ou égale à 2,5 Å, le point est classé comme « occupé », sinon le point est classé comme « libre » (Figure II3a).

- (2) Ensuite, les points libres sont effacés, en partant de l'extérieur de la grille vers la surface de la protéine, jusqu'à ce que les nuages de points restants aient une largeur maximale pré-définie par l'utilisateur. Chaque nuage de points non effacés représente une cavité (Figure II3b). Pour reproduire au mieux les dimensions du site actif dans les simulations, la largeur maximale de celui-ci doit être réglée selon les dimensions déterminées expérimentalement, quand cela est possible.

Les enzymes ayant très souvent des topologies de surface assez irrégulières, plusieurs cavités sont détectées par cette procédure automatique. Il appartient au modélisateur, en se basant sur ses connaissances de l'enzyme étudiée, de sélectionner parmi les cavités détectées celle qui constitue le vrai site de liaison pour y faire le docking des ligands dans l'étape suivante.

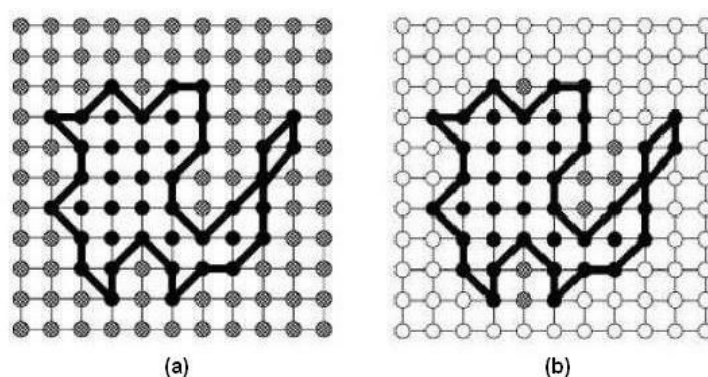


Figure II3 - Représentation schématique bidimensionnelle de la grille autour et au travers de la protéine (traits noirs), et de ses points d'intersection. (a) Les points sont classés en occupés (cercles noirs) et libres (cercles gris). (b) Des points libres sont éliminés (cercles blancs) de manière à obtenir des cavités dont l'ouverture maximale est définie par l'utilisateur. Les points libres restants (cercles gris) définissent des cavités [2].

3.1.2. Mode Opérateur

Les sites actifs des lipases pour le docking des flavonoïdes ont été définis en se basant sur la description de leurs cavités, à partir de données cristallographiques trouvées dans la littérature.

- Pour la CALB [7], la cavité catalytique est étroite et profonde, avec des dimensions approximatives de 10 Å x 4 Å de largeur et 12 Å de profondeur (Figure II4). La triade catalytique, constituée par les résidus Ser105, His224 et Asp187, est localisée au fond de la cavité. Les résidus du trou oxyanionique sont Thr40 et Gln106. La cavité de la CALB est essentiellement hydrophobe, tapissée par les résidus aliphatiques Leu140, Ala141, Leu144, Val149, Val154, Ile189, Leu278, Ala281, Ala282, Ile285 et Ala286. Seule une petite région autour des résidus catalytiques possède un caractère hydrophile, dû à la présence des chaînes latérales des résidus Thr40, Asp134 et Gln157.

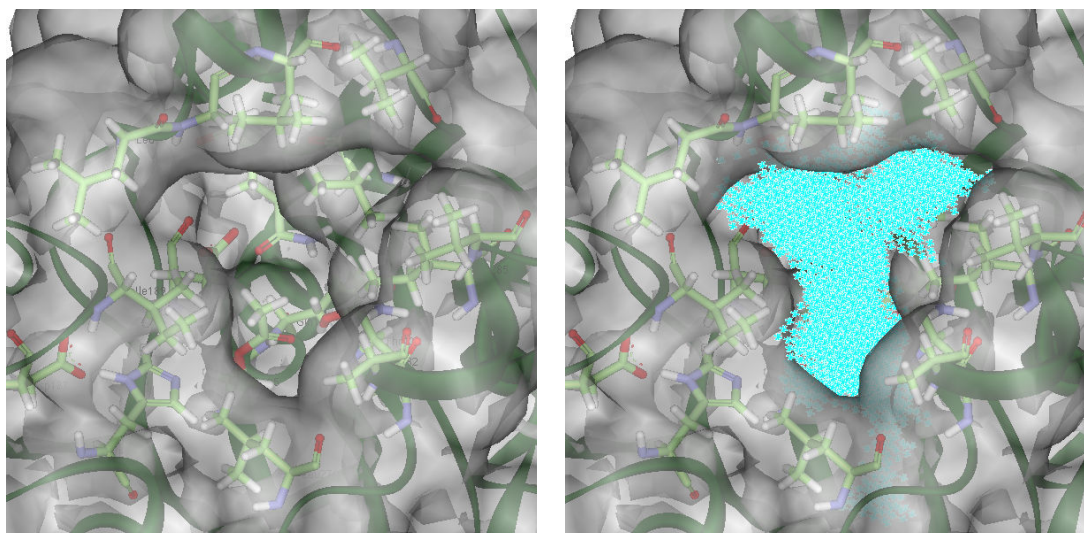


Figure II4 - À gauche, surface de Connolly, mettant en évidence la cavité catalytique de l'acétylcholinesterase (AChE) et les résidus qui l'entourent. À droite, en bleu, le nuage de points couvrant l'intérieur de cette cavité, correspondant à l'espace où seront dockés les flavonoïdes (le site actif).

- Pour la PCL [8], la cavité du site actif est plutôt large et allongée, avec des dimensions approximatives de 17 Å x 7 Å de largeur et 8 Å de profondeur (Figure II5). La triade catalytique, localisée dans la partie centrale du fond de la cavité, est constituée par les résidus Ser87, His286 et Asp264. A côté, se trouvent les résidus du trou oxyanionique, Leu17 et Gln88. La partie interne de la cavité possède deux régions : une région hydrophobe, constituée par les résidus Pro113, Ser117, Phe119, Val123, Leu164, Leu167 et Val267, et une région faiblement hydrophile, délimitée par les résidus Thr18, Val26, Leu27, Tyr29, Phe146, Ile290, Gln292 et Leu293. Vers l'entrée, se trouvent les résidus Leu248, Thr251, Val266 et Leu287.

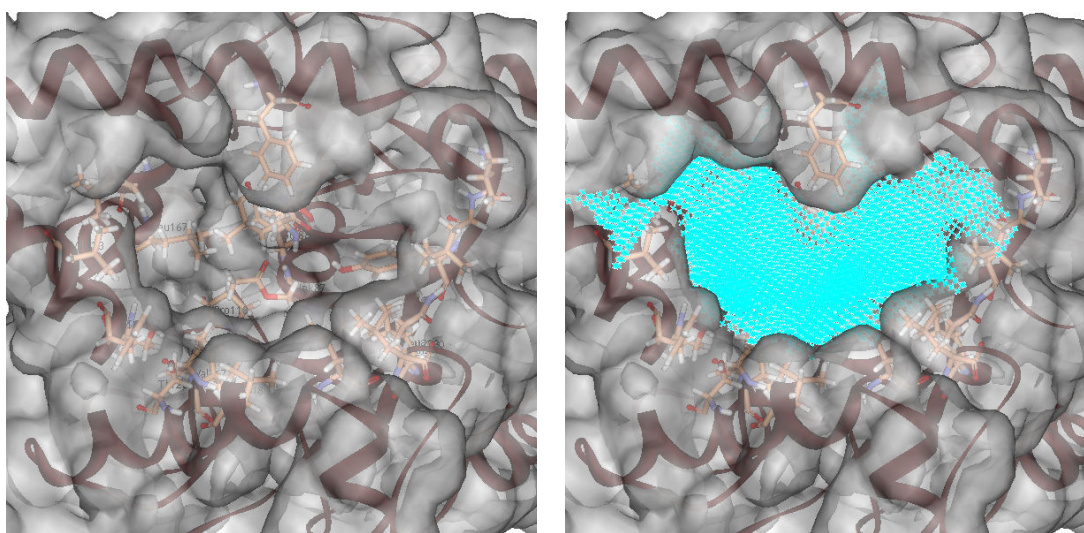


Figure II5 - À gauche, surface de Connolly, mettant en évidence la cavité catalytique de l'acétylcholinesterase (AChE) et les résidus qui l'entourent. À droite, en bleu, le nuage de points couvrant l'intérieur de cette cavité, correspondant à l'espace où seront dockés les flavonoïdes (le site actif).

3.2. L'obtention des complexes

3.2.1. Principe

Avec LigandFit [2], l'exploration de l'espace conformationnel du ligand dans le site actif de la protéine est exécutée selon une méthode stochastique (Monte Carlo), en réalisant des incréments randomisés (multiples de 30°) sur les torsions du ligand [14, 15]. Les méthodes stochastiques de recherche conformationnelle sont généralement considérées comme les plus performantes pour les simulations de docking. En effet, la génération randomisée de conformères permet une meilleure exploration de l'espace du ligand sans risque de *trapping* (blocage du complexe dans un minimum d'énergie). En outre, les méthodes stochastiques permettent un traitement plus exhaustif et dans des délais raisonnables des ligands comportant un grand nombre de torsions, ce qui constitue un avantage par rapport aux algorithmes de recherche systématique [16-18].

La génération et la sélection de conformères du ligand se font de la façon suivante :

- Le nombre maximal de conformères du ligand à générer (N_{trials}) et le nombre maximal de ces conformères à sauvegarder (N_{save}) sont définis par le modélisateur.
- En partant d'une conformation initiale arbitraire n du ligand, chaque angle de torsion de la molécule est soumis à une incrémentation randomisée [15], ce qui donne origine à un nouveau conformère $n+1$. Tous les conformères ainsi générés sont placés dans le site actif selon 4 orientations différentes (Figure II6), conduisant à 4 poses du ligand (une pose = une orientation donnée d'un conformère donné). Les complexes résultants sont conservés dans la liste de sauvegarde, jusqu'à ce que le nombre maximal de conformères dans cette liste (N_{save}), prédéfini par le modélisateur, soit atteint. Quand la liste est pleine, la prochaine pose générée est comparée avec chacune de celles de la liste, selon deux critères : l'adéquation géométrique avec la cavité du site actif et l'énergie potentielle d'interaction dans le complexe formé.

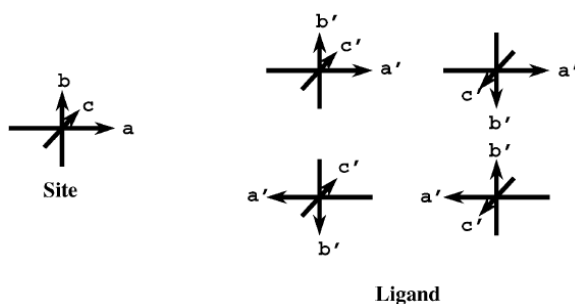


Figure II6 - Quatre orientations relatives de chaque conformère du ligand (poses) sont dockées dans le site actif [2].

- Pour l'évaluation de l'adéquation de forme, le principe mathématique de la comparaison entre les poses du ligand et le site actif se base sur leur représentation par un ensemble de points dans l'espace cartésien tridimensionnel. A partir des sommations des coordonnées (x,y,z) de ces points, une matrice de forme M est définie pour la pose du ligand et une autre pour le site :

(Eq. II1)

$$M = \begin{bmatrix} \sum x^2 & \sum xy & \sum xz \\ \sum xy & \sum y^2 & \sum yz \\ \sum xz & \sum yz & \sum z^2 \end{bmatrix}$$

En utilisant les valeurs propres de la matrice de forme du ligand (L_1 , L_2 , et L_3 ; avec $L_1 > L_2 > L_3$) et de celle du site actif (S_1 , S_2 et S_3 ; avec $S_1 > S_2 > S_3$), l'écart de forme ρ (Eq. 2) est calculé. Plus la valeur de ρ s'approche de zéro, plus les deux formes se ressemblent.

(Eq. II2)

$$\rho = \sqrt{\left(\frac{S_1}{S_2} - \frac{L_1}{L_2}\right)^2 + \left(\frac{S_2}{S_3} - \frac{L_2}{L_3}\right)^2 + \left(\frac{S_1}{S_3} - \frac{L_1}{L_3}\right)^2}$$

Si la valeur de ρ est supérieure à celle de la pire des structures dans la liste (soit celle avec la valeur de ρ la plus élevée), la pose en test est rejetée ; sinon, elle doit encore passer par un deuxième « filtre », le filtre énergétique, pour être acceptée dans la liste de sauvegarde.

- Pour les poses ayant passé le filtre géométrique, l'énergie d'interaction E (Eq. 3) est calculée comme la somme des contributions de van der Waals (représentées par un terme LJ 9-6) et électrostatiques (représentées par un terme de Coulomb).

(Eq. II3)

$$E = \sum_{i,j} \varepsilon_{ij} \left[2 \left(\frac{r_{ij}^*}{r_{ij}} \right)^9 - 3 \left(\frac{r_{ij}^*}{r_{ij}} \right)^6 \right] + \frac{332.0716}{\varepsilon} \sum_{i,j} \frac{q_i q_j}{r_{ij}}$$

$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ (ε = paramètre d'énergie de l'atome i du ligand et j de la protéine) ; $r_{ij}^* = \sqrt{r_i^* r_j^*}$ = distance entre les atomes i de la protéine et j du ligand. (r^* = rayon de van der Waals de l'atome) ; q = charge électrique en unités atomiques. Si la valeur d'énergie d'interaction est supérieure à celle de la pire des structures dans la liste (soit celle avec la valeur de E la plus élevée), la pose en test est rejetée ; sinon, elle remplace la pire des structures.

Cette procédure est répétée pour chaque nouveau conformère, jusqu'à ce que le nombre maximal de conformères générés (N_{trials}), défini par le modélisateur, soit atteint. En fonctionnant ainsi, la liste de sauvegarde est continuellement mise à jour, en éliminant les poses défavorables et les poses redondantes. Seules les meilleures structures, d'un point de vue géométrique et énergétique, sont donc accumulées au cours du processus. L'ensemble de la procédure est représenté de façon schématique dans la Figure II7 (page suivante).

3.2.2. Mode Opératoire

La procédure de docking des flavonoïdes a été effectuée avec les paramètres suivants :

- **Nombre de conformères générés (N_{trials}) = 15000** → les flavonoïdes étudiés sont des molécules plutôt rigides (trois angles dièdres pour la rutine, deux pour l'isoquercitrine et un seul pour la quercétine). Un nombre élevé de conformères a été généré, afin de s'assurer que l'espace conformationnel des ligands a été couvert.

- *Nombre maximal de poses retenues (N_{save}) = 50* → plutôt que de garder uniquement le meilleur complexe issu de la procédure de docking, la rétention des 50 meilleurs permet d'identifier des familles d'orientations possibles, au cas où il y en aurait plusieurs.
- *Écart de forme maximal (rmsd) entre le site et le ligand (ρ) = 4 Å* → il a été montré que cette limite représente un bon compromis [2] : des valeurs supérieures engendrent la rétention de plusieurs poses ayant une adéquation de forme cavité/ligand insatisfaisante ; des valeurs inférieures mènent à l'exclusion excessive de poses qui formeraient, par la suite, des complexes énergétiquement favorables avec la protéine.
- *Écart minimal pour discrimination des poses du ligand = 1,5 Å* → cette valeur évite que des poses trop semblables soient retenues dans la liste de sauvegarde, ce qui mènerait à la rétention de complexes redondants.

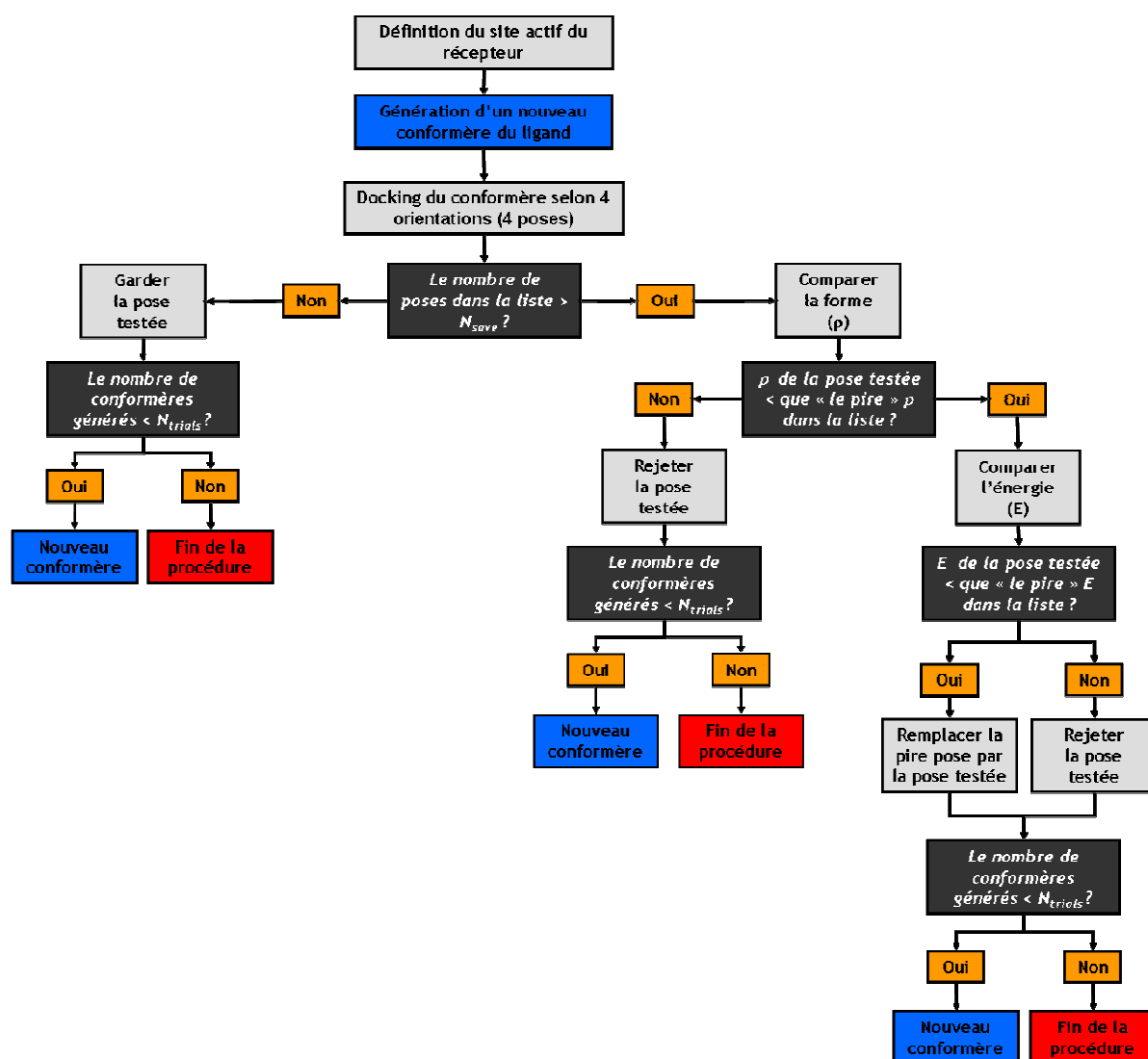


Figure II7 - Représentation schématique du fonctionnement de l'algorithme de docking LigandFit.

3.3. Le classement des complexes obtenus

Dans cet étude, le classement des complexes issus de la procédure de docking a été fait en appliquant le *Consensus score* sur les fonctions de score implémentées dans Discovery Studio 2.0 : LigScore1, LigScore2 [19], PMF [20, 21], PLP1, PLP2 [22, 23] et Jain [24] (cf. Annexe A1). La démarche a été la suivante :

- (1) D'abord, les 50 poses retenues dans la procédure de docking ont été classées en utilisant individuellement les 6 fonctions de score disponibles. Il a été observé que l'ordre de classement n'est pas le même d'une fonction de score à l'autre, mais change peu.
- (2) Ensuite, plutôt que de choisir arbitrairement le classement fourni par l'une de ces fonctions, le *Consensus score* a été appliqué, avec un pourcentage de 10 %. Autrement dit, dans le cas où l'on retient 50 poses, le consensus vérifie la fréquence à laquelle chaque pose apparaît parmi les 5 poses les mieux classées (top 5).
- (3) Lorsque les 50 poses retenues ont présenté le flavonoïde avec des orientations très similaires, les trois complexes les mieux classées (selon le consensus score) ont été sélectionnés pour la suite. Lorsque des familles distinctes d'orientations pour le flavonoïde ont été repérées parmi les 50 poses retenues, le complexe le mieux classé pour chaque famille a été sélectionné.

3.4. L'optimisation des complexes retenus après le classement

3.4.1. Principe

Comme LigandFit ne considère pas la flexibilité de la protéine lors de la procédure de docking, les complexes ont été optimisés, pour permettre le réajustement éventuel des acides aminés de la région de la poche catalytique dû à la présence du ligand. Ce phénomène est connu sous la dénomination de *induced fit effect* [25].

Pour l'exécution de la procédure de docking des flavonoïdes, les structures des acétyl-lipases avaient dû être enlevées de la boîte d'eau (restriction liée à l'algorithme de docking). Or, l'optimisation des complexes doit être faite dans le même environnement que celui dans lequel la protéine a été préalablement relaxée (molécules d'eau explicites). Par ailleurs, avec les processeurs disponibles, l'utilisation de cellules PBC contenant plusieurs milliers de molécules d'eau, comme celles utilisées pour la relaxation initiale des structures, requiert des temps de calcul conséquents. C'est pourquoi, afin de réduire les temps de calcul tout en gardant la cohérence des modèles, nous avons décidé d'utiliser une couche de molécules d'eau explicites pour solvater les complexes à étudier.

3.4.2. Mode Opérateur

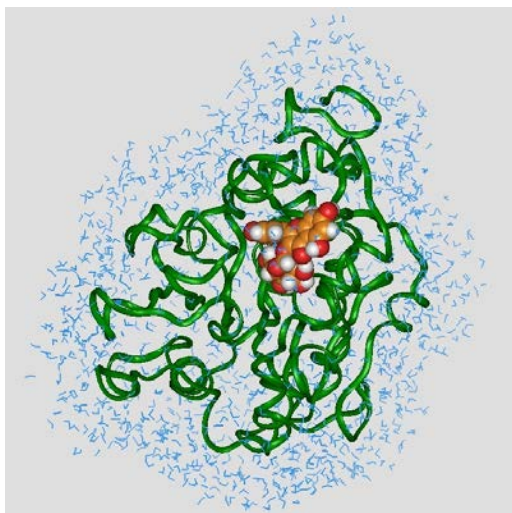


Figure II8 - Exemple de complexe solvato avec une couche de molécules d'eau.

- (1) Les complexes retenus dans l'étape précédente ont été solvatés avec une couche d'eau de 5 Å (1442 molécules), afin d'établir un bon compromis entre précision et efficacité (Figure II8, ci-contre).
- (2) L'optimisation des nouveaux systèmes solvatés a été faite en trois étapes, selon la méthodologie résumée dans le Tableau II2. Comme dans le cas de la relaxation initiale de la structure cristallographique des protéines, l'optimisation progressive a pour objectif d'éviter que les complexes subissent, au départ, des déformations trop brusques et non réalistes.

Tableau II2 - Paramètres de simulation employés dans l'optimisation des complexes enzyme-substrat solvatés.

Étape	Contraintes	Algorithme	Critère d'arrêt
1	▪ Contrainte <i>fix</i> appliquée aux atomes constituant le backbone de la protéine ▪ Contrainte <i>harmonic</i> (mouvement restreint) appliquée aux atomes des chaînes latérales des acides aminés de l'acétyl-lipase et au flavonoïde docké ▪ Atomes de H et molécules d'eau libres	Conjugate Gradient	Gradient de 0.1 kcal.mol ⁻¹ . Å ⁻¹
2	▪ Contrainte <i>harmonic</i> appliquée aux atomes du backbone de la protéine ▪ Chaînes latérales des acides aminés de la protéine, molécules d'eau et atomes de H libres	Conjugate Gradient	Gradient de 0.01 kcal.mol ⁻¹ . Å ⁻¹
3	▪ Aucune contrainte appliquée au système (tous les atomes libres)	Conjugate Gradient	Gradient de 0.001 kcal.mol ⁻¹ . Å ⁻¹

À l'issue de cette étape, les complexes solvatés et optimisés ont été soumis à d'autres simulations. Les complexes impliquant la PCL ont été étudiés par dynamique moléculaire. Les complexes avec la CALB ont été étudiés également par dynamique moléculaire et aussi par la méthode quantique de la Théorie Fonctionnelle de la Densité (DFT).

4. SIMULATIONS DE DYNAMIQUE MOLECULAIRE

4.1. Principe

Dans une revue récente, Alonso et al. [26] ont cité et analysé un grand nombre d'exemples où l'utilisation de simulations de dynamique moléculaire post-docking a permis d'affiner les complexes obtenus lors du docking. En effet, la dynamique moléculaire prend en compte à la fois la flexibilité du ligand et celle de la protéine et des molécules explicites de solvant (si présentes dans le système modélisé). De ce fait, d'éventuelles modifications structurales, résultant des interactions entre toutes ces entités, peuvent être modélisées. Dans notre étude l'objectif des simulations de dynamique moléculaire a été d'examiner la stabilité structurale des complexes lipase-substrats. Une attention particulière a été portée aux interactions entre les résidus enzymatiques et les substrats (flavonoïde et acétate) et à la stabilité de ces derniers dans la cavité catalytique des lipases. Ces facteurs ont été corrélés à la productivité (prédisposition à former un produit) des complexes.

4.2. Mode Opérateur

- (1) Les structures de départ ont été : les différents complexes optimisés obtenus avec la quercétine, l'isoquercitrine et la rutine dans la cavité de l'acétyle-CALB et ceux obtenus avec la quercétine dockée dans la cavité de l'acétyle-PCL.
- (2) Paramètres de simulation :
 - Température = 300 K
 - Algorithme d'intégration : algorithme de Verlet-Leapfrog (cf. Annexe A2)
 - *Cut-off* pour les interactions non liées = 12 Å (sphérique)
 - Pas d'intégration (Δt) = 1 fs (10^{-15} s)
 - Fréquence de mise à jour des vitesses atomiques = 50 fs
 - Fréquence d'enregistrement de la configuration du système = 1 ps (10^{-12} s)
 - Temps de la trajectoire = 2 ns (10^{-9} s)

4.3. Analyse des résultats

Pour l'analyse des trajectoires, les distances entre l'oxygène de l'acétate (Ace:O) et les hydrogènes des résidus du trou oxyanionique ont été suivies. Ces distances indiquent si l'acétate présente ou non une orientation catalytiquement productive, et donc si le complexe est susceptible de mener à la formation d'un ester de flavonoïde. La démarche d'analyse suivie est la suivante :

- (1) Pour les trajectoires au cours desquelles ces distances sont stables, trois *frames* ont été échantillonnés : un au début de la trajectoire, un au milieu et un autre à la fin, à partir du graphe *distances = f(temps)*. Pour les trajectoires instables par rapport à ces distances, un *frame* représentant chaque famille conformationnelle a été sélectionné. Pour chaque complexe, les frames sélectionnés ont été superposés.

- (2) À partir de la superposition des frames et de la visualisation de l'animation de la trajectoire :
- Les interactions (attractives ou répulsives) entre les résidus de la cavité et les substrats ont été identifiées.
 - Toutes les distances interatomiques d'intérêt pour le processus catalytique ont été mesurées, d'après la méthodologie décrite dans les travaux de l'équipe de Hult [27-29].

5. SIMULATIONS DE MÉCANIQUE QUANTIQUE (DFT)

5.1. Principe

La méthode de la Théorie Fonctionnelle de la Densité (DFT) [6] constitue un outil de plus en plus utilisé pour étudier la catalyse enzymatique, comme discuté par Friesner et Guallar [30]. L'application de cette approche quantique permet d'évaluer les énergies et les changements électroniques du système moléculaire au cours du processus catalytique (la rupture et la formation de liaisons chimiques). Dans ce travail, des simulations basées sur la DFT ont été réalisées afin d'étudier la réactivité des flavonoïdes quercétine, isoquercitrine et rutine au sein des complexes formés avec la CALB. Nous nous sommes plus précisément focalisés sur la prédisposition des hydroxyles des flavonoïdes situés à proximité des résidus catalytiques à former une liaison ester avec l'acétyle enzyme.

5.2. Mode Opérateur

- (1) Les structures de départ ont été : les complexes obtenus avec la quercétine, l'isoquercitrine et la rutine dockées dans la cavité de l'acétyle-CALB
- (2) À partir de ces complexes, la molécule de flavonoïde, l'acétate lié à la Ser105 et des fragments des résidus directement impliqués dans la réaction ont été sélectionnés :
 - Le carboxylate de l'Asp187
 - Le cycle imidazole de l'His224
 - Le -NH du backbone de la Gln106
 - Le C α , -NH du backbone et la chaîne latérale de la Thr40
- (3) Ces parties ont été conservées ; le reste de la protéine et les molécules d'eau ont été effacées.
- (4) Les valences libres ont été complétées par l'ajout de groupements méthyle (-CH₃). Les mini-systèmes ainsi obtenus ont constitué les structures d'entrée pour les simulations.
- (5) Paramètres de simulation :
 - Les coordonnées spatiales des atomes appartenant au backbone de la protéine ont été fixées, afin de maintenir la disposition d'origine de ceux-ci.
 - Charge totale des systèmes = -1 (dû à l'Asp187 déprotoné)
 - *Cut-off* global des orbitales = 3,7 Å
 - Fonctionnelle d'échange-corrélation : GGA-PW91 [31] (cf. Annexe A3)

- Bases des fonctions atomiques : DNP (*Double Numerical with Polarization*) [32] (cf. Annexe A3) - fonctions de base numériques équivalentes en performance à la base Gaussienne 6-31G** [33].

- Procédure SCF, avec les critères de convergence : variation maximale d'énergie = 1.0×10^{-5} Ha ; déplacement atomique maximal = 3.0×10^{-3} Å ; force maximale = 4.0×10^{-3} Ha.Å⁻¹.

(6) À partir des structures optimisées, les intermédiaires ont été modélisés comme suit :

- Le proton de l'hydroxyle du flavonoïde a été effacé et l'atome Nε de l'His224 a été protoné, afin de mimer le transfert de proton du mécanisme réactionnel.
- L'atome de carbone Ace:C (carbonyle de l'acétate) a été assigné comme sp³.
- Après ces deux modifications, les structures obtenues ont été ré-optimisées, en utilisant les mêmes paramètres décrits en (5).

5.3. Analyse des résultats

L'analyse des résultats a été faite en se basant sur les variations de trois paramètres entre les structures optimisées des « réactifs » (5) et des « intermédiaires » correspondants (6) :

- (1) Variations des charges Mulliken des atomes impliquées dans la réaction [34, 35].
- (2) Variations des longueurs de liaison et des distances interatomiques [36].
- (3) Variation de l'énergie électronique (1 Hartree (Ha) = 627,51 kcal.mol⁻¹) [36].

6. REFERENCES

- [1] A.D. MacKerel Jr., D. Bashford, M. Bellot, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus. *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. Journal of Physical Chemistry 102 (1998) 3586-3616.
- [2] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman. *LigandFit : a novel method for the shape-directed rapid docking of ligands to protein active sites*. Journal of Molecular Graphics and Modelling 21 (2003) 289-307.
- [3] R.D. Clark, A. Strizhev, J.M. Leonard, J.F. Blake, J.B. Matthew. *Consensus scoring for ligand/protein interactions*. Journal of Molecular Graphics and Modelling 20 (2002) 281-295.
- [4] M. Feher. *Consensus scoring for protein-ligand interactions*. Drug Discovery Today 11 (2006) 421-428.
- [5] B. Delley. *DMol³ DFT studies: from molecules and molecular environments to surfaces and solids*. Computational Materials Science 17 (2000) 122-126.
- [6] W. Kohn, A.D. Becke, R.G. Parr. *Density functional theory of electronic structure*. Journal of Physical Chemistry 100 (1996) 12974-12980.
- [7] J. Uppenberg, M.T. Hansen, S. Patkar, T.A. Jones. *The sequence, crystal structure determination and refinement of two forms of lipase B from Candida antarctica*. Structure 2 (1994) 293-308.
- [8] K.K. Kim, H.K. Song, D.H. Shin, K.Y. Hwang, S.W. Suh. *The crystal structure of a triacyl glycerol lipase from Pseudomonas cepacia reveals a highly open conformation in the absence of a bound inhibitor*. Structure 5 (1997) 173-185.
- [9] D.H.J. Mackay, A.J. Cross, A.T. Hagler. *The role of energy minimization in simulation strategies of biomolecular systems*, in: G.D. Fasman (Ed.). Prediction of Protein Structure and the Principles of Protein Conformation, Plenum Press, New York and London, 1990, pp. 317-358.
- [10] M. Levitt, S. Lifson. *Refinement of protein conformations using a macromolecular energy minimization procedure*. Journal of Molecular Biology 46 (1969) 269-279.
- [11] R. Fletcher, C.M. Reeves. *Function minimization by Conjugate Gradients*. Journal of Computational Chemistry 7 (1964) 149-154.
- [12] A.R.S. Janzen, J.W. Leech. *Periodic multistep methods in molecular dynamics*. Computer Physics Communications 32 (1984) 349-359.
- [13] A. Svendsen. *Lipase protein engineering*. Biochimica et Biophysica Acta 1543 (2000) 223-238.
- [14] J.Y. Trosset, H.A. Scheraga. *Flexible docking simulations: scaled collective variable Monte Carlo minimization approach using Bezier splines, and comparison with a standard Monte Carlo algorithm*. Journal of Computational Chemistry 20 (1999) 244-252.
- [15] T.J. Oldfield. *A number of real-space torsion angle refinement techniques for proteins, nucleic acids, ligands and solvent*. Acta Crystallographica D 57 (2001) 82-94.
- [16] I. Helperin, B. Ma, H. Wolfson, R. Nussinov. *Principles of dockings : an overview of search algorithms and a guide to scoring functions*. PROTEINS: Structure, Function and Bioinformatics 47 (2002) 409-443.
- [17] S.F. Sousa, P.A. Fernandes, M.J. Ramos. *Protein-ligand docking : current status and future challenges*. PROTEINS: Structure, Function and Bioinformatics 65 (2006) 15-26.

- [18] N. Moitessier, P. Eglebienne, D. Lee, J. Lawandi, C.R. Corbeil. *Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go*. British Journal of Pharmacology 153 (2008).
- [19] A. Krammer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman. *LigScore : a novel scoring function for predicting binding affinities*. Journal of Molecular Graphics and Modelling 23 (2005) 395-407.
- [20] I. Muegge. *PMF revisited*. Journal of Medicinal Chemistry 49 (2006) 5895-5902.
- [21] I. Muegge, Y.C. Martin. *A general and fast scoring function for protein-ligand interactions : a simplified potential approach*. Journal of Medicinal Chemistry 42 (1999) 791-804.
- [22] D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel, S.T. Freer. *Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming*. (1995).
- [23] D.K. Gehlhaar, D. Bouzida, P.A. Rejto. *Rational drug design: novel methodology and practical applications* in: L. Parrill, M. Rami Reddy (Eds.). American Chemical Society, Washington, DC., 1999, pp. 292-311.
- [24] A.N. Jain. *Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities*. Journal of Computer Aided Molecular Design 10 (1996) 427-440.
- [25] D.H. Williams, E. Stephens, D.P. O'Brien, M. Zhou. *Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand induced reductions in motion within receptors and enzymes*. Angewandte Chemie - International Edition 43 (2004) 6596-6616.
- [26] H. Alonso, A.A. Bliznyuk, J.E. Gready. *Combining docking and molecular dynamic simulations in drug design*. Medicinal Research Reviews 26 (2006) 531-568.
- [27] C. Branneby, P. Carlqvist, K. Hult, T. Brinck, P. Berglund. *Aldol additions with mutant lipase : analysis by experiments and theoretical calculations*. Journal of Molecular Catalysis B: Enzymatic 31 (2004) 123-128.
- [28] M. Cammenberg, K. Hult, S. Park. *Molecular basis for the enhanced lipase-catalyzed N-acetylation of 1-phenylethanamine with methoxyacetate*. ChemBioChem 7 (2006) 1745-1749.
- [29] M. Svedendahl, P. Carlqvist, C. Branneby, O. Allnér, A. Frise, K. Hult, P. Berglund, T. Brinck. *Direct epoxidation in Candida antarctica lipase B studied by experiment and theory*. ChemBioChem 9 (2008) 2443-2451.
- [30] R.A. Friesner, V. Guallar. *Ab initio quantum mechanics and mixed quantum mechanics / molecular mechanics (QM/MM) methods for studying enzymatic catalysis*. Annual Review of Physical Chemistry 56 (2005) 389-427.
- [31] J.P. Perdew, Y. Wang. *Accurate and simple analytic representation of the electron-gas correlation energy*. Physical Review B 45 (1992) 13244-13249.
- [32] B. Delley. *An all-electron numerical method for solving the local density functional for polyatomic molecules*. Journal of Chemical Physics 92 (1990) 508-517.
- [33] Y. Inada, H. Orita. *Efficiency of numerical basis sets for predicting the binding energies of hydrogen bonded complexes: evidence of small basis set superposition error compared to Gaussian basis sets*. Journal of Computational Chemistry 29 (2007) 225-232.
- [34] R.S. Mulliken. *Electronic population analysis on LCAO-MO molecular wave functions I*. The Journal of Chemical Physics 23 (1955) 1833-1840.

- [35] R.S. Mulliken. *Electronic population analysis on LCAO-MO molecular wave functions II. Overlap populations, bond orders and covalent bond energies.* The Journal of Chemical Physics 23 (1995) 1841-1846.
- [36] C.H. Hu, T. Brinck, K. Hult. *Ab initio and density functional theory studies of the catalytic mechanism for ester hydrolysis in serine hydrolases.* International Journal of Quantum Chemistry 69 (1998) 89-103.

**Étude de la régiosélectivité dans
l'acetylation des flavonoïdes glycosylés
isoquercitrine et rutine catalysée par la
lipase B de *Candida antarctica***

INTRODUCTION

Plusieurs études expérimentales publiées ces dernières années, aussi bien par notre équipe que par d'autres laboratoires, ont démontré l'efficacité de la lipase B de *Candida antarctica* (CALB) comme biocatalyseur, en milieu organique, pour l'acylation des flavonoïdes glycosylés isoquercitrine et rutine. Ces bioprocédés d'acylation présentent une remarquable régiosélectivité : dans les deux cas, les produits majoritaires obtenus sont des mono-esters, résultant de l'acylation du groupement 6''-OH présent sur la partie glucose de l'isoquercitrine et du 4'''-OH de la partie rhamnose la rutine. En outre, de manière intéressante, cette régiosélectivité est indépendante du solvant et du donneur d'acyle utilisés. Néanmoins, aucune explication de ces résultats, au niveau moléculaire, n'a été proposée. Au vu de ces résultats expérimentaux et du mécanisme réactionnel des lipases, nous avons fait l'hypothèse que la régiosélectivité de ces procédés d'acylation devrait être aux modes d'interaction entre les flavonoïdes et le site actif de l'enzyme.

Pour vérifier cette hypothèse :

- Un protocole de docking a été établi pour étudier les positions, les orientations et les interactions de l'isoquercitrine et de la rutine dans la cavité du site actif de la lipase CALB. La réaction d'acétylation (acétate comme donneur d'acyle) a été choisie comme réaction-modèle. Ce choix a été motivé par la disponibilité de données expérimentales dans la littérature pour la réaction de ces deux flavonoïdes avec ce donneur d'acyle, permettant ainsi la validation des modèles sur la base d'arguments empiriques.
- Les complexes favorables obtenus par la procédure de modélisation ont été analysés selon des critères basés sur le mécanisme réactionnel des lipases, dans le but de déterminer la propension des différents groupements -OH des flavonoïdes, accessibles aux résidus catalytiques, de mener à la formation d'un ester (dits « complexes productifs »).

Les résultats de cette partie ont fait l'objet d'une publication, parue dans le *Journal of Molecular Catalysis B : Enzymatic* (volume 59, pages 96-105, 2009).

1. ARTICLE :

A molecular modelling study to rationalize the regioselectivity in acylation of flavonoid glycosides catalyzed by *Candida antarctica* lipase B

Eduardo B. De Oliveira ^a, Catherine Humeau ^{a*}, Latifa Chebil ^a, Elaine R. Maia ^b,
François Dehez ^c, Bernard Maignet ^d, Mohamed Ghouil ^a, Jean-Marc Engasser ^a

^a Laboratoire Biocatalyse Bioprocédés (LBB), ENSAIA-INPL, Nancy Université, 2 av. de la Forêt d'Haye, 54500, Vandoeuvre-lès-Nancy, France.

^b Laboratorio de Estudos Estruturais Moleculares (LEEM), Instituto de Química, Universidade de Brasília, CP 4478, 70904-970, Brasília-DF, Brazil.

^c Equipe de Dynamique des Assemblages Membranaires (EDAM), CNRS, Université Henri Poincaré, 7565, 54500, Vandoeuvre-lès-Nancy, France.

^d Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), CNRS, Université Henri Poincaré, BP 239, 54506, Vandoeuvre-lès-Nancy, France.

* Corresponding author - phone: (+33) 03.83.59.57.84 / fax: (+33) 03.83.59.57.78

Abstract:

The regioselective behaviour of the *Candida antarctica* lipase B (CALB) towards two flavonoid glycosides, rutin and isoquercitrin, in the acetylation reaction was investigated through molecular modelling. A protocol constituted by a Monte-Carlo based docking procedure and classical force fields calculations was applied to find probable binding modes of the substrates inside the catalytic cavity and optimize the corresponding complexes. The analysis of these complexes allowed identifying productive ones (that means, those able to lead to the formation of the ester product) according to three parameters: (1) protein distortion; (2) stability of hydrogen bond interactions with the oxyanion hole residues; (3) localization of hydroxyl groups with regard to the region comprised between the catalytic histidine and serine residues. Results showed that the aglycon part of both rutin and isoquercitrin was localized at the entrance of the binding pocket, stabilized by hydrogen bond and hydrophobic interactions. The sugar part of the flavonoids was placed close to the pocket bottom. In particular, only the primary 6''-OH of the isoquercitrin glucose and the secondary 4'''-OH of the rutin rhamnose were expected to be acetylated, as they were the only ones to stabilize simultaneously near to the catalytic histidine and the acetate bound to the catalytic serine. These findings are in accordance with experimental data and give a suitable explanation, at an atomic level, of the regioselectivity of CALB in the flavonoid glycosides acetylation.

Keywords:

Candida antarctica lipase B, Docking, Flavonoid acylation, Molecular modelling, Regioselectivity.

1. Introduction

Flavonoids are a broad class of low molecular weight polyphenols, produced as secondary metabolites by the greater part of vegetal species. Chemically, they are benzo- γ -pyrone derivatives, with a phenyl-type substituent on the C2 or C3 position. Frequently, but not obligatory, they are found in nature covalently attached to a glycoside part. These compounds exhibit a wide range of biological, pharmacological and medicinal properties, such as anti-oxidant, anti-allergenic, anti-viral and cancer preventive effects, making them very attractive for use as functional ingredient in cosmetic, pharmaceutical and food products [1, 2]. However, most of them present a poor solubility in both polar and non-polar media [3], which strongly restricts their incorporation in many formulations. A solution to improve the hydrophobic nature (lipophilization) of flavonoids consists in their acylation [4], which can be accomplished by chemical or enzymatic processes. With the chemical acylation, many of the hydroxyl groups present on the flavan skeleton and on the sugar moieties can be esterified [5], which can lead to the acylation of some phenol groups that are directly implicated in the beneficial properties of these molecules [6].

On the contrary, using enzymes, such as lipases, only a single or a few hydroxyl groups were reported to be esterified [7]. In this case the number and position of the reactive hydroxyls depend on the structures of both the flavonoid and the lipase. With *Candida antarctica* lipase B (CALB, EC 3.1.1.3), the most frequently used enzyme to perform esterification reactions in non-aqueous media [4], the diglycosylated flavonoid rutin (quercetin-3-O-glucose-rhamnose, Figure 1), which carries ten hydroxyl groups, was found to be acylated only on the secondary 4'''-OH of rhamnose. This strict regioselectivity was reported for a wide range of acyl donors, such as fatty acids with different chain length and their vinyl or ethyl derivatives [8-11]. When CALB was used for the acylation of the monoglycosylated flavonoid isoquercitrin (quercetin-3-O-glucose, Figure 1), which exhibits eight hydroxyl groups, it is the primary 6''-OH of the glucose moiety that was esterified. With vinyl acetate as the acyl donor, the obtained isoquercitrin-6''-acetate could be further esterified to produce isoquercitrin-3'',6''-diacetate [12]. With other acyl donors, the formation of isoquercitrin-6''-monoester as the sole product has been reported [13-15]. In these studies no explanation on a molecular level was provided to the observed regioselectivity of the CALB-catalyzed acylation of either rutin or isoquercitrin.

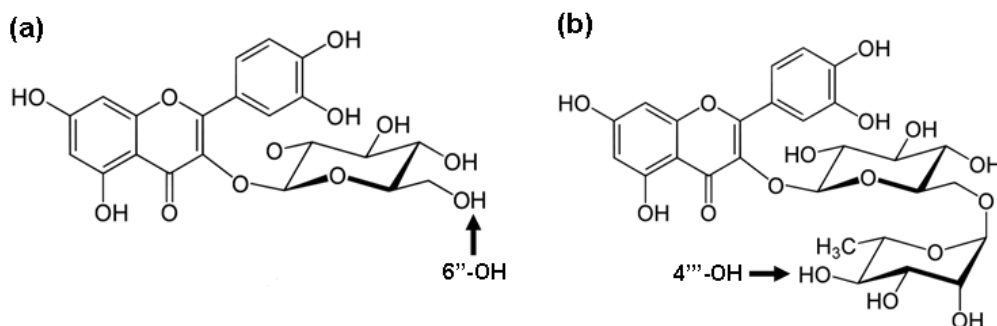


Figure 1 – Chemical structures of (a) isoquercitrin and (b) rutin. The hydroxyl groups regioselectively acetylated with CALB as biocatalyst are indicated by arrows.

Recently, computer-based modelling methodology has been increasingly used to understand the mechanism and selectivity of enzyme-catalyzed reactions and to suggest how to change the selectivity of these reactions by modification of the enzyme or the substrate structure [16, 17]. With lipases, many studies address the use of molecular modelling tools to furnish a rational explanation for their enantioselectivity towards a large set of chiral substrates, in both hydrolysis and esterification reactions [18-23]. Also, some works deal with the regioselectivity of lipase-catalyzed acylation of poly-hydroxylated compounds, such as resorcinarenes [24], glycosides [25-27] and prostaglandins [28]. To our knowledge, there is no published work on the application of computer-based modelling to describe the molecular interactions involved in flavonoids recognition by lipases. Thus, the present study aims to provide a molecular-level explanation for the observed regioselectivity of CALB towards the 6''-OH and 4'''-OH groups of isoquercitrin and rutin, respectively. A combined docking, molecular mechanics and molecular dynamics approach was used to study the positions, the orientations, the interactions of these two compounds in the active site of CALB and the accessibility of their hydroxyl groups to the catalytic residues. All these factors are important to determine if a given substrate binding mode is productive or not, thus allowing to elucidate the regioselectivity in the reaction.

2. Methodology

All simulations were carried out on a bi-processor AMD Dual Core 280 2.4 GHz. Molecular mechanics and molecular dynamics calculations and docking simulations were performed with CHARMM force field [29], adopting a 12 Å non-bound spherical cut-off, using the program-package Discovery Studio version 1.7 (Accelrys, Inc.). In molecular dynamics calculations, the isothermal-isochoric ensemble (NVT) was used, under periodic boundary conditions (PBC). A time step of 1 fs and the Verlet-Leapfrog algorithm were systematically applied for the integration of the equations of motion. Connolly surfaces [30] were calculated using Insight II LS (Accelrys, Inc.). The study comprises of the following steps: preparation of starting structures; modelling of the acetyl-enzyme (docking target); docking of the flavonoid substrate and subsequent scoring of generated poses; optimization and structural analysis of the best-scored poses; verification of the reliability of the final models.

2.1. Target preparation

The CALB crystal structure was taken from the Protein Data Bank (PDB entry: 1LBS; resolution: 2.60 Å) [31]. The complex has six chains per unit cell of dimensions $a = 229.5$ Å, $b = 95.6$ Å, $c = 86.8$ Å; $\alpha = \beta = \gamma = 90.0^\circ$. Each chain system is composed of 317 amino acids, one ethyl-hexylphosphonate inhibitor (HEE) covalently attached to the Ser105 residue, a dimer of N-acetyl-glucosamine covalently bound to the Asn74 residue and 92 adsorbed water molecules. Only one protein chain and its water molecules were kept and used as initial atomic coordinates set for the simulations; the others were erased. The inhibitor and the two N-acetyl-glucosamines were also removed. All hydrogen atoms were added in their theoretical positions. The catalytic residues Asp187 and His224 were assigned as deprotonated. Next, this new system was solvated with 6145 water

molecules in a box with previous optimized dimensions of $a = 64 \text{ \AA}$, $b = 59 \text{ \AA}$, $c = 63 \text{ \AA}$; $\alpha = \beta = \gamma = 90.0^\circ$.

A structure relaxation procedure was executed. In this procedure, constraints and restraints were added then progressively removed [32], in order to preserve the crystalline organization of protein atoms: (a) only protein hydrogen atoms and water molecules were allowed to move, keeping all other heavy atoms fixed, in order to eliminate initial strains; (b) the protein backbone was kept fixed and harmonic restraints were imposed on the protein side chains; (c) harmonic restraints were applied only on the protein backbone, allowing the side chains free to move; (d) finally, the whole system was minimized without any constraints. The first step was done by performing 1000 Steepest Descent iterations. The steps (b), (c) and (d) were performed using the Conjugate Gradient method until the root mean square deviation (*rmsd*) gradient fell below 0.01, 0.001 and 0.0001 kcal.mol⁻¹.Å⁻¹, respectively. In order to equilibrate the optimized system, a standard dynamic trajectory was carried out at 300K, for 1ns. The protein structure in the last frame was taken for subsequent modelling steps.

The binding pocket of CALB consists of an elliptical and steep funnel whose walls predominantly constitute by hydrophobic aliphatic amino acid residues [42]. The catalytic machinery is located at the bottom of the pocket, a more hydrophilic region, and constitutes by the triad Ser105...His224...Asp187 [31]. The well-known reaction mechanism is based on the acylation and deacylation of the Ser105 residue and involves two tetrahedral intermediates (Figure 2).

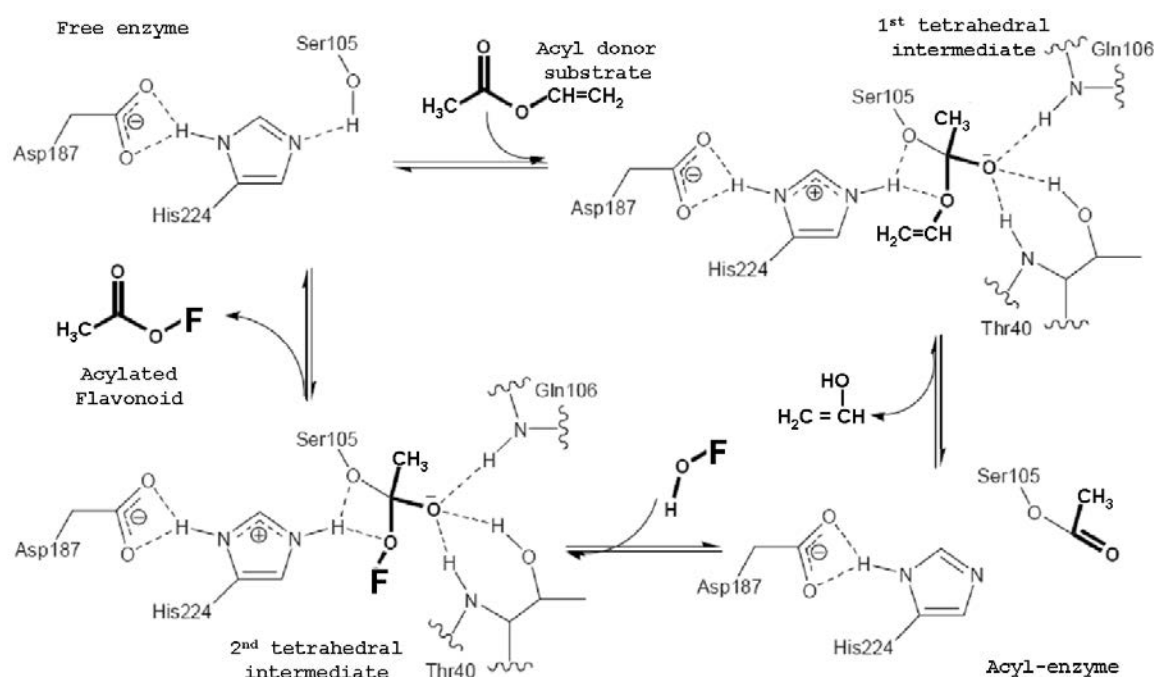


Figure 2 – The reaction mechanism of the CALB-catalyzed acylation of a flavonoid, evidencing the two tetrahedral intermediates and the acyl-enzyme complex. In this study, the acyl donor substrate is vinyl acetate, F–OH is the flavonoid to be acetylated (docked ligand) and the docking target is the acetyl-CALB.

The first tetrahedral intermediate results from the nucleophilic attack of the catalytic serine on the acylating substrate and originates the acyl-enzyme complex. The second one derives from the nucleophilic attack of the acyl acceptor substrate on the acyl-enzyme and leads to the release of the ester product [31, 42]. Therefore, the Ser105 side chain hydroxyl was replaced by acetate in order to mimic the acyl-enzyme complex [24, 27]. To allow the adjustment of the acetate, further 1,000 steps of Conjugate Gradient energy minimization were performed with a fix constraint applied to the protein backbone. This optimized acetyl-CALB complex was taken as target in the docking procedure.

2.2. Docking and scoring

Isoquercitrin and rutin were non-covalently docked against the acetyl-CALB. This approach allows exploring all the different orientations of one ligand within the catalytic pocket in one single simulation, contrarily to another approach which consists in docking the ligands covalently attached to the acyl-enzyme, mimicking the second tetrahedral intermediate. In this case, all the potentially reacting functions must be separately modelled. Docking simulations were performed with LigandFit [33], which combines a shape comparison filter with a Monte Carlo conformational search algorithm. Moreover, LigandFit owns an internal scoring function (DOCKSCORE) that estimates iteratively the interaction energy (E) between each docked pose and the receptor, taken as the sum of van der Waals and electrostatic energies, represented in the equation below:

$$E = \sum_{i,j} \varepsilon_{ij} \left[2 \left(\frac{r_{ij}^*}{r_{ij}} \right)^9 - 3 \left(\frac{r_{ij}^*}{r_{ij}} \right)^6 \right] + \frac{332.0716}{\varepsilon} \sum_{i,j} \frac{q_i q_j}{r_{ij}}$$

wherein ε is the dielectric constant, $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ and $r_{ij}^* = \sqrt{r_i^* r_j^*}$ (r_i^* is the van der Waals radius and ε_i is the energy parameter of the i -th ligand atom, while r_j^* and ε_j are similarly parameters for the j -th protein atom). The distance between the i -th ligand atom and the j -th protein atom is represented by r_{ij} and their respective charges (in atomic units) are q_i and q_j . Thus, in addition to the geometric fit, the docking procedure assures to keep only the ligand poses having favourable energy of interaction with the protein. For each substrate 15,000 Monte Carlo trials were performed and only the 50 most favourable poses were retained. Docked poses were scored by six different scoring functions: LigScore 1 and 2 [34], PLP 1 and 2 [35], PMF [36] and Jain [37], followed by a consensus scoring. In fact, the consensus combines data obtained from different scores to compensate for inaccuracy from individual scoring functions and has been demonstrated to improve the probability of finding correct solutions [38]. The three best scored poses according to the consensus scoring were selected and the corresponding complexes were optimized and further analyzed.

2.4. Post-docking optimization

The docking/scoring procedure of LigandFit is based on both shape fitting and energetic stability of the protein-ligand system, guaranteeing that only viable complexes are saved [33]. As this docking algorithm treats the protein as a rigid body, enzyme-substrate complexes were submitted to a careful post-docking optimization, in order to take into account a potential induced fit effect; that means, little displacements in the protein structure due to the presence of the ligand [39]. In the

selected complexes, the acetate oxygen was orientated to the oxyanion hole, so that it can form hydrogen bonds interactions with the oxyanion hole residues (Thr40 and Gln106), mimicking the oxyanion orientation in the second tetrahedral intermediate [19]. Then three consecutive energy minimizations were performed: (a) a fix constraint was applied on the protein backbone and all the remaining parts of the systems were tethered with a harmonic restraint; (b) the protein backbone was kept fix, the side chains were kept tethered and the docked ligand was free to move; (c) only the backbone was kept fix, with side chains and the ligands free to move. The Conjugate Gradient algorithm was used in the three steps, with root mean square deviation (*rmsd*) gradients of 0.01, 0.001 and 0.0001 kcal.mol⁻¹.Å⁻¹, respectively.

2.5. Analysis of the optimized complexes

In this step, in order to determine if optimized complexes corresponded to productive binding modes, they were analyzed by combining three criteria (Figure 3) used in previous studies dealing with molecular modelling of lipases: The first criterion is the *rmsd* between the protein structure in the complex and its initial crystal structure. Even though some enzymes may undergo relatively large conformational changes upon the binding of substrates (induced fit effect), characterized by *rmsd* values up to 6 Å [39], we chose a *rmsd* value around 3 Å for CALB as the limit between reactive and unreactive complexes, as found by Vallikivi *et al.* [28] for prostaglandins acetylation. The second one is that complexes must conserve at least two of the three hydrogen bonds between the acetate oxygen (namely Ace:O) and the oxyanion hole residues Thr40 and Gln106, postulated to stabilize the oxyanion in the second tetrahedral intermediate [31]. The third one is the distance between the flavonoid hydroxyl group and the acetate sp² carbon (namely Ace:C) and the N ϵ atom of His224 (namely His224:N ϵ). We assumed that, in order to be considered as reactive, a hydroxyl group must be simultaneously placed at 4 Å maximum from the Ace:C and His224:N ϵ atoms, as stated by Palloci *et al.* [27] for acetylation of trytilated glycosides catalyzed by *Candida rugosa* lipase. The breach of any of these conditions would be indicative that the complex is not able to lead to the formation of the second tetrahedral intermediate; that means, the binding mode is non-productive.

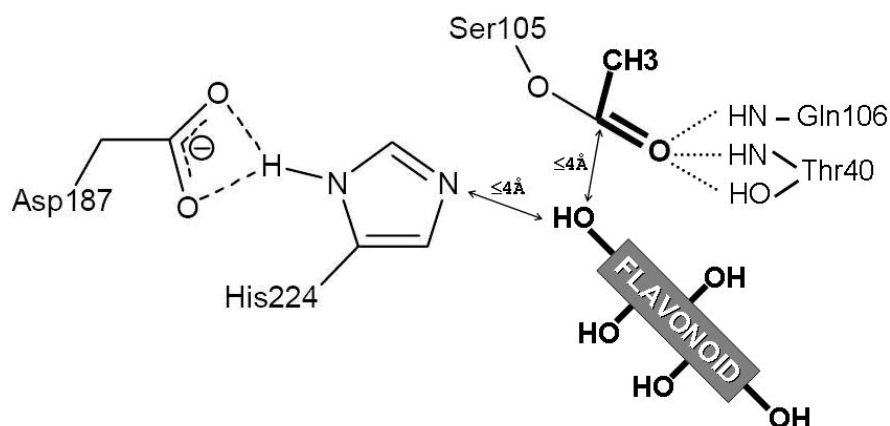


Figure 3 – Scheme of structural requirements that flavonoid hydroxyl groups must satisfy in the models in order to be considered as reactive.

2.6. Molecular dynamics simulations on the productive complexes

After the identification of the productive binding modes, further 100 ps MD trajectory was executed on the corresponding complexes, maintaining the backbone fix, in order to evaluate the dynamics of the docked substrates. In particular, the stability of the essential hydrogen bond interactions and the critical distances for reactivity were verified. The trajectories were divided in 5 ps of heating from 50K to 300K, 5 ps of equilibration at 300K and 90 ps of production at 300 K. Frames were registered every 500 fs. Only production phases were considered for analysis.

3. Results

3.1. Target preparation and analysis

Prior to docking simulations, the CALB structure was submitted to a careful structure relaxation protocol [32] to optimize the system without harming the disposition of the atoms from the crystal structure. In order to monitor eventual distortions in the lipase structure in the procedure, the root-mean-square deviation (*rmsd*) was computed all along the equilibration step (MD trajectory of 1 ns at 300K), considering the superimposition of all the heavy atoms. The relaxation procedure was performed in presence of explicit water molecules, rather than to choose a particular solvent. This decision was based on two arguments: firstly, even though an enzymatic reaction is carried out in organic media, the presence of a certain amount of water molecules in the reaction medium contributes to the structural integrity of the enzyme and thus to its optimal activity [40]. Secondly, Trodler and Pleiss [41] showed, using molecular dynamics simulations with explicit solvent molecules, that CALB is a little more flexible in water than in organic solvents. In that way, if CALB structure is well conserved after the simulations in water, it is reasonable to assume that it should be also preserved in organic solvents. The *rmsd* values (all heavy atoms) were stabilized at about 1.1 Å, after approximately 600 ps of equilibration. This result indicates a high structural conservation, evidencing a successful relaxation procedure. Then, starting from the relaxed CALB structure, the acetyl-enzyme intermediate was manually built and optimized. When superimposing it on the relaxed CALB structure, no structural changes were observed, as reflected by the negligible *rmsd* value (all protein heavy atoms). Moreover, the hydrogen bond interaction between the side chains of Asp187 and His224 was maintained and the distance between the Ser105:O γ and His224:N ϵ atoms remained at 3.2 Å. This indicates that the initial disposition of the catalytic triad residues was not damaged. Therefore, this acetyl-CALB structure was taken as target for the docking of the two flavonoid substrates.

3.2. Docking results

In the case of isoquercitrin the 50 complexes presented the substrate in two orientations: the sugar part or the aglycon part (ring A) orientated towards the enzyme core. In the case of rutin, all the retained complexes presented the substrate with its sugar part orientated towards the interior of the cavity, with only small conformational differences among them. For both flavonoids, the three best-scored complexes, according to the consensus scoring, were representative of these trends and thus were optimized and analyzed in order to determine if they corresponded to productive binding modes or not, according to three criteria: protein distortion; content of essential interactions between the

acetate and the oxyanion hole residues; proximity of the hydroxyl groups to the His224:N ϵ and Ace:C atoms.

Although some protein conformational rearrangements may be needed for a better ligand affinity, if the protein distortion is too high the orientation of some essential residues may be disturbed, leading to a loss of activity. Vallikivi *et al.* [28] showed that, in complexes between CALB and prostaglandins (also polyhydroxylated compounds), *rmsd* values higher than 3.0 Å were experimentally well correlated with unreactive enzyme-substrate complexes. Therefore, we also adopted the maximal value of 3.0 Å as a first criterion for a productive complex. In our case, for all the retained CALB-flavonoid complexes, the superimposition of the CALB backbone to the protein crystal structure showed only minor displacements (*rmsd* values about 0.9 Å). Minimal chain distortions were found mainly in loop regions, which are usually the more mobile regions in proteins [43]. Sequences presenting some distortion due to the simulations are specified in Figure 4. When all heavy atoms were superimposed, *rmsd* values went up to about 3.0 Å, as side chains were free to move in the last step of the optimization procedure. In spite of this, the cavity shape and the general visual appearance of the protein structure were not substantially altered.

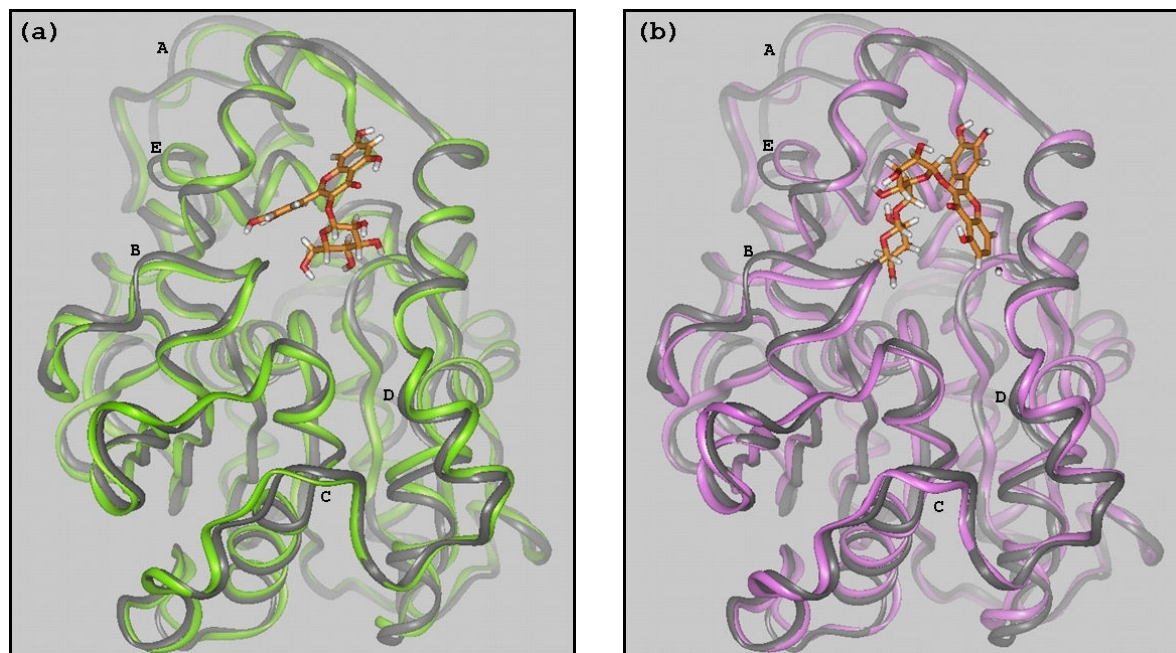


Figure 4 – Ribbons showing CALB crystal structure superimposed to that after the modelling procedure complexed with (a) isoquercitrin and (b) rutin. The crystal structure is colored in gray; that after the modelling procedure is colored in green in (a) and in violet in (b). Main chain differences are found in the sequences: (A) Cys311-Gly313, which is the loop between the sheets $\beta 8$ and $\beta 9$; (B) Val190-Pro192, contained in the loop connecting the sheets $\beta 6$ and $\beta 7$; (C) Asn259-Pro262, contained in the loop connecting the helices $\alpha 9$ and $\alpha 10$; (D) Asn85-Ile87, contained in the helix $\alpha 3$; (E) Thr158-Ala162, contained in the loop between the helices $\alpha 6$ and $\alpha 7$.

To achieve the reaction, the Ace:O atom must form hydrogen bonds interactions in the oxyanion hole, with the backbone –NH of the residues Thr40 and Gln106 and the side chain –OH of Thr40. In fact, these hydrogen bonds enhance the partial positive charge on the Ace:C, enabling the

nucleophilic attack of the acyl acceptor substrate. Therefore, as previously mentioned, another criterion to consider a complex as productive was that at least two of these three hydrogen bonds must be conserved after the post-docking optimization. Hydrogen bond distances and angles values are given in Table 1.

Table 1 - Hydrogen bond distances and angles involving the substrates in the three best-scored complexes after optimization

Ligand	Complex	Donor	Acceptor	Distance (Å)	Angle (°)
<i>Isoquercitrin</i>	I1	Isoq:HO7	Val149:O	2.1	137.2
		Isoq:HO5	Isoq:O4	1.8	149.7
		Isoq:HO4'	Ile189:O	1.8	164.2
		Thr40:HN	Ace:O	1.9	125.5
		Thr40:H _{G1}	Ace:O	1.4	123.4
		Gln106:HN	Ace:O	2.2	8.5
	I2	Thr40:HN	Ace:O	2.1	169.7
	I3	Thr40:H _{G1}	Ace:O	1.5	136.5
		Gln106:HN	Ace:O	2.1	97.0
<i>Rutin</i>	R1	Rut:HO7	Leu278:O	2.1	146.3
		Rut:HO4'	Val149:O	1.9	149.6
		Rut:HO5	Rut:O4	2.0	159.3
		Rut:HO3'''	Gln157:O _{E1}	1.9	126.2
		Thr40:HN	Ace:O	1.9	121.6
		Thr40:H _{G1}	Ace:O	1.7	152.6
		Gln106:HN	Ace:O	2.2	15.7
	R2	Rut:HO4'	Val149:O	2.0	148.7
		Rut:HO7	Leu278:O	2.1	137.6
		Thr40:HN	Ace:O	2.4	110.2
	R3	Rut:HO4''	Rut:O6''	2.1	143.7
		Rut:HO4'''	Gln157:O _{E1}	2.0	150.4
		Thr40:HN	Ace:O	2.4	119.9
		Thr40:H _{G1}	Ace:O	1.8	132.3
		Gln106:HN	Ace:O	1.9	14.7

Another imperative criterion that a complex must obey in order to be considered as productive is the proximity of at least one flavonoid hydroxyl group from catalytic residues. More specifically, it must be positioned simultaneously close to His224:Nε to transfer its proton and enhance its

nucleophilicity and close to Ace:C to execute the nucleophilic attack itself. We decided to adopt for these distances a maximal value of 4 Å, as did by Palocci *et al.* [27]. The distance values obtained for the most buried substrate hydroxyl groups in each complex are noted in Table 2.

Table 2 - Distances *a* of buried hydroxyl groups of the docked flavonoids from the His224:Ne and the acetate carbonyl carbon (Ace:C) atoms in the three best scored complexes after optimization.

Ligand	Complex	Buried –OH groups	Distance from Ace:C (Å)	Distance from His224:Ne (Å)
<i>Isoquercitrin</i>	I1	3''-OH	7.4	8.7
		4''-OH	4.6	6.2
		6''-OH	3.5	4.0
	I2	5-OH	5.6	8.3
		7-OH	5.2	7.1
	I3	3''-OH	8.6	8.7
		4''-OH	5.9	6.5
		6''-OH	4.2	5.9
<i>Rutin</i>	R1	2'''-OH	6.3	8.2
		3'''-OH	3.9	7.0
		4'''-OH	2.9	3.9
	R2	2'''-OH	6.7	7.8
		3'''-OH	5.9	8.5
		4'''-OH	4.1	7.8
	R3	2'''-OH	7.1	12.5
		3'''-OH	4.7	10.1
		4'''-OH	3.7	9.4
^a Distance measured considering the hydroxyl oxygen atoms.				

3.2.1. Docking of isoquercitrin

Among the three retained complexes for isoquercitrin (namely I1, I2 and I3; Figure 5), two possible orientations were found: the sugar part or the aglycon part (ring A) orientated towards the enzyme core. In I2, the isoquercitrin sugar moiety is orientated towards the exterior of the cavity and the aglycon moiety is orientated inwards. The hydroxyl groups closer to the active site are the 5-OH and 7-OH phenolic groups, with respective distances of 5.6 Å and 5.2 Å from Ace:C and 8.3 Å and 7.1 Å from the His224:Ne. In addition, in this complex the acetate forms only one hydrogen bond with the oxyanion hole residues (Thr40:NH...Ace:O). This may be attributed to a hydrophobic contact between the acetate methyl chain and the C6 of the ring A of the substrate (distance of 3.4 Å), turning the acetate oxygen away from the Thr40 and Gln106 residues. For these reasons, I2 cannot represent a

productive binding mode and was rejected. In I1 and I3, the aglycon moiety lodges at the entrance of the binding pocket, whereas the sugar is oriented inwards. In I3 the acetate interacts with the oxyanion hole residues through two hydrogen bonds (Thr40:OH...Ace:O and Gln106:NH...AceO). However, the closest substrate hydroxyl group to the catalytic residues (6''-OH) is placed at 5.8 Å from His224:N ϵ and 4.2 Å from Ace:C, which exceeds the maximal values admitted as to allow the proton transfer and the nucleophilic attack. Therefore, I3 was also discarded. In the complex I1, the acetate conserved the three possible hydrogen bonds interactions with the Thr40 and Gln106 residues. In addition, the nearest hydroxyl group to the catalytic residues (6''-OH) is positioned simultaneously at 4.0 Å from His224:N ϵ and 3.5 Å from Ace:C, fulfilling the distance criteria to react. Thus, I1 may be considered as a productive binding mode.

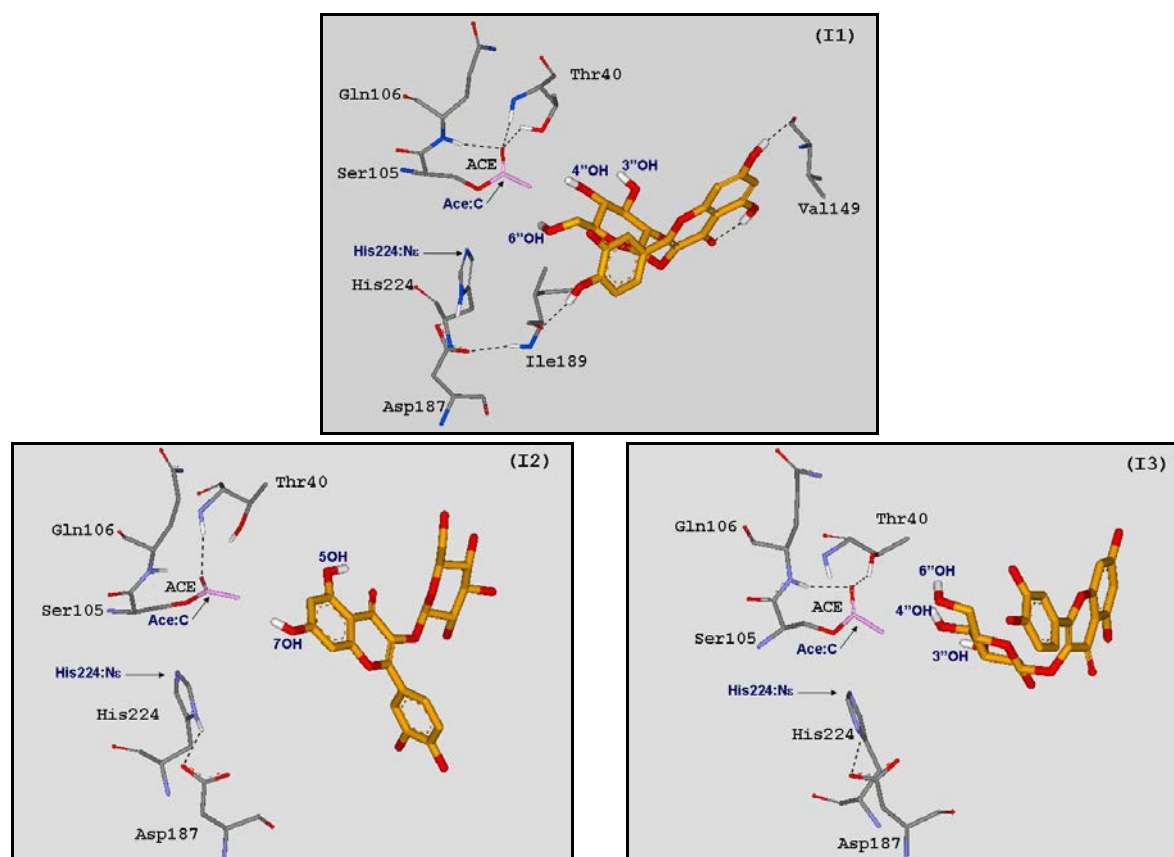


Figure 5 – Orientations and hydrogen bonds interactions (dashed lines) of the substrates in the three best scored acetyl-CALB isoquercitrin (I1, I2 and I3) complexes. For clarity, only the residues useful for discussion are shown and non-interacting hydrogen atoms are hidden.

More specifically, in I1 isoquercitrin docked within the target with the aglycon moiety anchored at the entrance of the binding pocket, where it is stabilized by two types of interactions: hydrogen bond interactions involving the phenolic groups (7-OH...Val149:O and 4'-OH...Ile189:O) and nonpolar interactions between the aromatic cycles and hydrophobic aliphatic residues, which abound in this region of the enzyme (Leu140, Ala141, Leu144, Val 149, Val154, Ile189 and Val 286). The glucose moiety is placed within the cavity. The three most buried hydroxyl groups are the 3''-OH, 4''-OH and 6''-OH, but only this last is at a distance inferior to 4.0 Å from the Ace:C and His224:N ϵ atoms (Table 1). In fact, in this primary alcohol function the -CH₂- alkyl group proceeds as an “extension piece” that

allows the –OH to draw near the catalytic residues. Therefore, the molecular model hints that only the 6''-OH group of isoquercitrin is expected to react. These findings agree with experimental studies showing that isoquercitrin-6''-monoester was the sole product directly obtained in the CALB-catalyzed acetylation of isoquercitrin [12]. The additional formation of isoquercitrin-3'',6''-diacetate was already reported [44], but further investigations indicated that it is likely to result from the acetylation of the initially synthesized isoquercitrin-6''-acetate, in a sequential way [12]. Several other authors also showed that CALB catalyzes the mono-acylation of isoquercitrin with phenolic acids as cinnamic acid [13], p-coumaric acid [14] and phenyl propionic acid [15] as acyl donors. In all cases the preference of this enzyme for the 6''-OH group stood unaffected, which further supports our model for the binding of isoquercitrin within the CALB active site.

3.2.2. Docking of rutin

In the case of rutin, the three retained complexes (namely R1, R2 and R3; Figure 6) presented the flavonoid docked in similar orientations. In all of them, the aglycon moiety stabilizes at the entrance of the cavity, while the sugar part of the substrate points towards the bottom. Notably, the 4'''-OH group of the rhamnose unit is the closest one to the catalytic residues. However, the three complexes present rutin in different conformations, in which the dihedral angle between the two sugars is more or less bent (162° for R1, 149° for R2 and 102° for R3). In R2, only one hydrogen bond between the residues of the oxyanion hole and the acetate is present (Thr40:NH...Ace:O). Indeed, the orientation of the acetate may be influenced by the 6'''-CH₃ of the rutin rhamnose through a hydrophobic attraction (distance of 3.6 Å to the 6'''-C atom). Furthermore, in this complex the distance from the 4'''-OH to the His224:Nε is 7.7 Å, exceeding the value admitted as to permit the proton transfer. Consequently, R2 was rejected. In R3, the acetate presents all the three possible hydrogen bond interactions with the oxyanion hole residues (Thr40:OH...Ace:O, Thr40:NH...Ace:O and Gln106:NH...Ace:O). The 4'''-OH is positioned at 3.6 Å from Ace:C, but at 9.4 Å from the His224:Nε. Therefore, rutin in R3 is also unable to react and the complex was also discarded. In R1 all the three possible hydrogen bond interactions with the oxyanion hole are also maintained. The rutin conformation is quite different from that in R3: the sugar moiety is somewhat benter, so that the 4'''-OH is placed at 3.9 Å from His224:Nε and 2.9 Å from Ace:C. So, this complex may be considered as a productive binding mode.

In this complex, the rutin aglycon part stabilized at the entrance of the cavity thanks to hydrogen bonds interactions involving phenolic groups (4'-OH...Val149:O and 7-OH...Leu278:O) and nonpolar interactions between its aromatic cycles and the side chains of residues Leu144, Val149, Val154, Ala281, Ala282, Ile285 and Val286. The rhamnose part, that has three hydroxyl groups (2'''-OH, 3'''-OH and 4'''-OH), is located near the little hydrophilic region at the bottom of the cavity, neighbouring the catalytic residues. Among these three hydroxyl groups, only the 4'''-OH satisfies the criteria of distance from His224:Nε and Ace:C atoms to be considered as reactive. The 3'''-OH group forms a hydrogen bond interaction with the Gln157 residue (3'''-OH...Gln157:OE1), which confers additional stability to the rhamnose cycle in this orientation. All these observations hint that the acylation reaction can only occur on the 4'''-OH. This model is well corroborated by experimental

studies showing that the CALB-catalyzed acylation of rutin with different acyl donors and under diverse reaction conditions led to the sole formation of the 4'''-monoester [8-11].

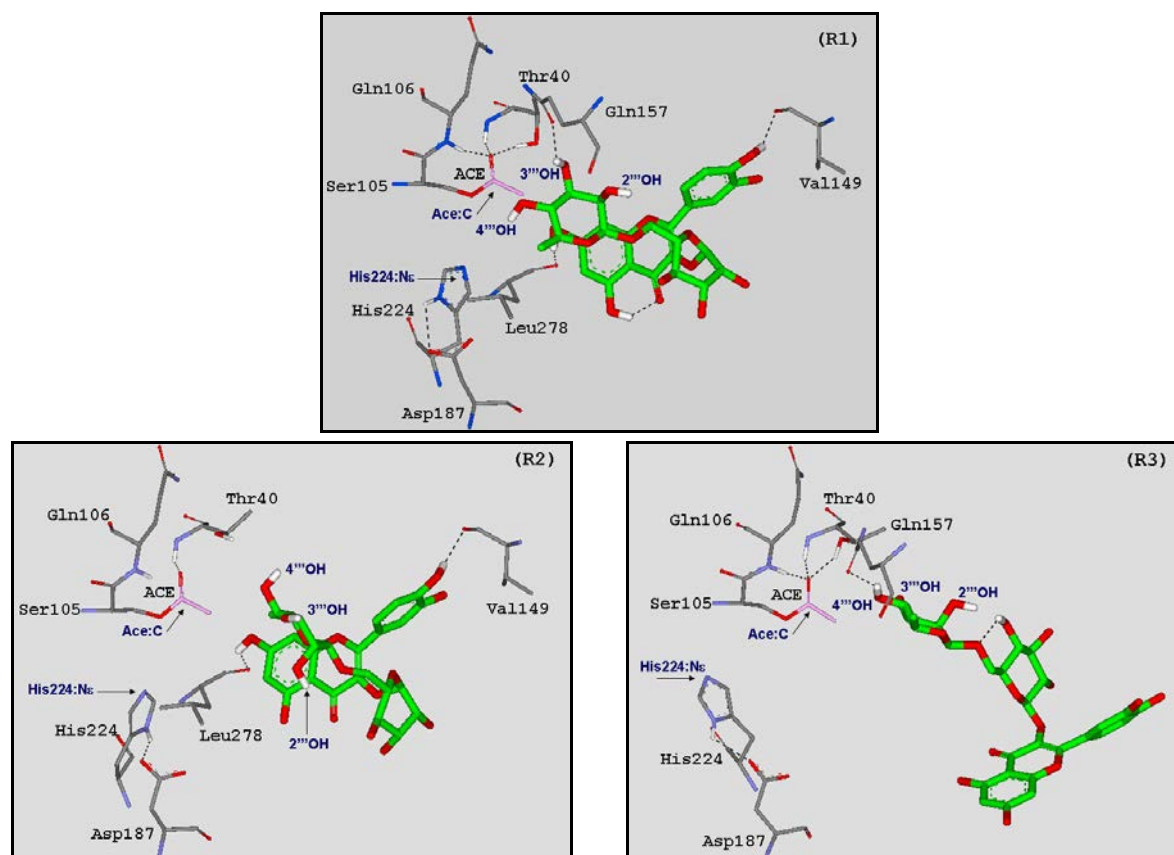


Figure 6 – Orientations and hydrogen bonds interactions (dashed lines) of the substrates in the three best scored acetyl-CALB rutin complexes (R1, R2 and R3). For clarity, only the residues useful for discussion are shown and non-interacting hydrogen atoms are hidden.

4. Discussion

In order to understand the orientations of rutin and isoquercitrin within the CALB binding pocket, the observed interactions between these two substrates and the enzyme residues were compared with already published data about protein-flavonoid interactions. A theoretical affinity order between flavonoids and twenty amino acids was proposed by Cordoniu-Hernandez *et al.* [45, 46], based on quantum mechanics semi-empirical calculations and the analysis of crystal structures of protein-flavonoid complexes from PDB. These authors stated that flavonoids display a more pronounced affinity towards hydrophilic and aromatic residues rather than hydrophobic ones. Indeed, the abundance of hydroxyl groups in flavonoid structure enables interactions of hydrogen bond type with polar residues, whereas their aromatic rings can easily form π - π interactions with aromatic residues side chains. In the case of CALB, major part of the binding pocket area is nonpolar, enfolded mainly by hydrophobic aliphatic residues of the helices $\alpha 5$ and $\alpha 10$ and a loop region that projects the Ile189 into the channel [31]. This can justify the fact that the aglycon moiety of both studied substrates is stabilized at the cavity entrance, despite the absence of aromatic amino acid residues (Figure 7).

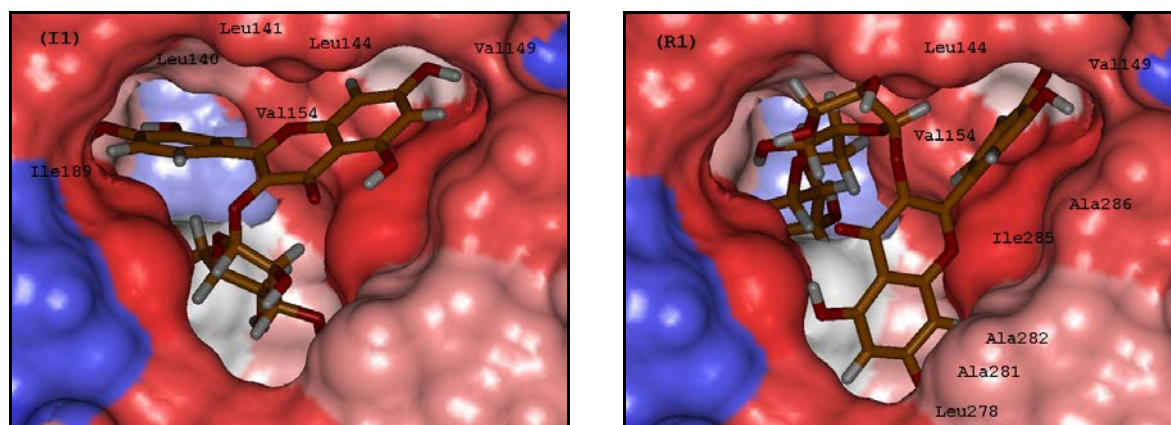


Figure 7 – Connolly accessible surface of CALB binding pocket evidencing the hydrophobic interactions of isoquercitrin in I1 and rutin in R1. Hydrophobic regions of the enzyme pocket are colored in red and hydrophilic ones in white/blue.

Favourable interactions between phenolic substrates and the nonpolar residues Leu144, Ile189 and Leu278 were already evidenced by Otto *et al.* [25], in a computational study of the CALB-catalyzed synthesis of arylaliphatic glycolipids. Only a little region at the bottom of the cavity, around the catalytic Ser105, presents a character somewhat polar, due to the side chains of the polar residues Thr40, Asp134 and Gln157. This justifies the orientation of the glycoside moiety of the two substrates towards the active site.

In the productive complexes, the aglycon parts of the two substrates are located at the cavity's doorway, but surprisingly, with distinct orientations, one perpendicular to the other (Figure 7). In order to explain such behaviour, we docked isoquercitrin and rutin employing another approach that is frequently used when docking structurally similar ligands against one given target protein: one ligand is initially docked and the others are superimposed over the analogue parts of the first one. Then, the resulting systems are sequentially optimized, analyzed and compared [47]. In our case, the rutin aglycon part was superimposed over its counterpart in the previously docked isoquercitrin and vice versa. In the first case, the rutin disaccharide had not enough space to lodge within the cavity, engendering an important sterical clash with the acetate and the Ser105. Conversely, in the second case, the isoquercitrin glucose is not long enough to reach the catalytic residues, resulting in a non-productive binding mode (see Supplementary Data). Therefore, it can be concluded that, although these two substrates have similar structures, they effectively present distinct binding modes within the CALB active site, as determined by LigandFit.

We complemented our study by additional molecular dynamics simulations on the productive complexes (100 ps at 300K), to verify the stability of the hydrogen bond interactions and to evaluate the dynamic behaviour of the docked substrates. In fact, applying molecular dynamics on complexes obtained by docking was shown to be efficient to refine the models, accounting for the flexibility of both receptor and ligand, and to check the complex temporal stability [48]. Results showed that, in any case, the proteins did not undergo significant distortions, as revealed by the maximal *rmsd* values (about 0.2 – 0.5 Å, considering all heavy atoms). The presence of the essential hydrogen bond interactions between the acetate carbonyl oxygen (Ace:O) and Thr40 and Gln106 was monitored. The

visual analysis of trajectories shows that at least two of them are present in most of trajectory frames during the production phase. In order to evaluate the stability of the hydroxyl groups positions, their transacetylating distances (distance between the hydroxyl oxygen atoms and the carbonyl carbon of acetate, Ace:C) were monitored all along the trajectory. Among them, only the 6''-OH of isoquercitrin and the 4'''-OH of rutin reached distances inferior to 4 Å from Ace:C. Furthermore, maximal fluctuations in these distances were weak (about 0.4 – 0.6 Å), reflecting the good stability of the docked flavonoids (Table 3). These results from molecular dynamics simulations on the productive complexes reinforce our previous conclusions about regioselectivity in CALB-catalyzed acylation of isoquercitrin and rutin.

Table 3 - Minimal and maximal transacetylating distances ^a during the MD production phase carried out on the productive complexes

Ligand	Buried –OH groups	Minimal distance (Å)	Maximal distance (Å)	Maximal variation (Å)
<i>Isoquercitrin</i>	3''-OH	8.0	8.4	0.4
	4''-OH	4.6	5.8	1.2
	6''-OH	4.0	4.6	0.6
<i>Rutin</i>	2'''-OH	7.6	8.2	0.6
	3'''-OH	5.6	6.5	0.9
	4'''-OH	3.9	4.3	0.4

^a Distance between the oxygen atoms of the flavonoid hydroxyl groups and the carbonyl carbon atom of the acetate (Ace:C) bound to Ser105

5. Summary and conclusions

Docking simulations combined with molecular mechanics and molecular dynamics calculations were used to study CALB-catalyzed acetylation of the two glycosylated flavonoids isoquercitrin and rutin. Both flavonoids were found to dock with their aglycon part at the cavity entrance, stabilized by hydrogen bond interactions between their phenolic groups and the backbone carbonyl of hydrophobic residues and hydrophobic interactions with the side chains of several Ala, Ile, Leu and Val residues enfolding the pocket. In spite of the aromatic character of the substrates, no π - π interactions were possible due to the lack of aromatic residues in the binding pocket walls. Their sugar part was stabilized in the neighbourhood of the catalytic residues, a little polar region at the bottom of the cavity. Further analyses of the complexes showed that the 4'''-OH of rutin and the 6''-OH of isoquercitrin, both localized at the extremity of the substrates glycosidic part, were the only hydroxyl groups satisfying the proximity criteria to the catalytic residues. This strict regioselectivity suggested by the models corroborated previous published experimental results on the CALB-catalyzed acetylation of isoquercitrin and rutin.

Combined theoretical and experimental studies are now pursued to investigate the capacity of our molecular modelling procedure to predict the regioselectivity of CALB for the acetylation of other glycosylated and aglycon flavonoids. In fact, as flavonoids show a very large structural diversity

(number and positions of glycoside units and hydroxyl groups, degree of oxidation of the C-ring...), they have different sizes, flexibilities, charge distributions, hydrophilic-hydrophobic balance properties, and are likely to yield different interaction modes with the lipase. Thus the entire modelling protocol outlined here must be applied to a broader range of flavonoid substrates, in order to verify if the geometrical cut-off between reactive and unreactive complexes (protein *rmsd* and proximity of flavonoid hydroxyl groups to the lipase catalytic residues) can be generalized to predict the regioselectivity of CALB-catalyzed acetylation of other members of this class of compounds.

Aknowledgments

We are thankful to Evelyne Ronat for her precious and essential technical support all along this work.

REFERENCES:

- [1] B.H. Havsteen, *Pharmacol. Ther.* 96 (2002) 67-202.
- [2] A.M. Boudet, *Phytochem.* 68 (2007) 2722-2735.
- [3] L. Chebil, C. Humeau, J. Anthoni, F. Dehez, J.M. Engasser, M. Ghoul, *J. Chem. Eng. Data* 52 (2007) 1552-1556.
- [4] P. Villeneuve, *Biotechnol. Adv.* 25 (2007) 515-536.
- [5] A.M. Mariotte, E. Perrier, A. Boumendjel, D. Bresson-Rival, Nouveaux esters de flavonoïdes et leur utilisation en cosmétique, en dermopharmacie, en pharmacie et en agro-alimentaire, in FR 2778663, France, 1999.
- [6] Z.Y. Chen, P.T. Chan, K.Y. Ho, K.P. Fung, J. Wang, *Chem. Phys. Lipids* 79 (1996) 157-163.
- [7] L. Chebil, C. Humeau, A. Falcimaigne, J.M. Engasser, M. Ghoul, *Process Biochem.* 41 (2006) 2237-2251.
- [8] F. Mellou, H. Loutrari, H. Stamatis, C. Roussos, F.N. Kolisis, *Process Biochem.* 41 (2006) 2029-2034.
- [9] M.H. Katsoura, A.C. Polydera, L. Tsironis, A.D. Tselepis, H. Stamatis, *J. Biotechnol.* 123 (2006) 491-503.
- [10] M. Ardhaoui, A. Falcimaigne, J.M. Engasser, P. Moussou, G. Pauly, M. Ghoul, *J. Mol. Catal. B: Enzym.* 29 (2004) 63-67.
- [11] M. Ardhaoui, A. Falcimaigne, S. Ognier, J.M. Engasser, P. Moussou, G. Pauly, M. Ghoul, *J. Biotechnol.* 110 (2004) 265-271.
- [12] L. Chebil, J. Anthoni, C. Humeau, C. Gerardin, J.M. Engasser, M. Ghoul, *J. Agric. Food Chem.* 53 (2007) 9496-9502.
- [13] N. Nakajima, K. Ishihara, T. Itoh, T. Furuya, H. Hamada, *J. Biosci. Bioeng.* 87 (1999) 105-107.
- [14] K. Ishihara, N. Nakajima, *J. Mol. Catal. B: Enzym.* 23 (2003) 411-417.
- [15] D.E. Stevenson, R. Wibisono, D.J. Jensen, R.A. Stanley, J.M. Cooney, *Enzyme Microb. Technol.* 39 (2006) 1236-1241.
- [16] R.J. Kazlauskas, *Curr. Opin. Chem. Biol.* 4 (2000) 81-88.
- [17] P. Braiuca, C. Ebert, A. Basso, P. Linda, L. Gardossi, *Trends Biotechnol.* 24 (2006) 419-425.
- [18] C. Orrenius, F. Haefner, D. Rotticci, N. Ohrner, T. Norin, K. Hult, *Biocatal. Biotransf.* 16 (1998) 1-15.

- [19] T. Schultz, J. Pleiss, R.D. Schmid, *Prot. Sci.* 9 (2000) 1053-1062.
- [20] A. Tafi, A. Van Almsick, F. Corelli, M. Crusco, K.E. Laumen, M.P. Schneider, M. Botta, *J. Org. Chem.* 65 (2000) 3659-3665.
- [21] S. Raza, L. Fransson, K. Hult, *Prot. Sci.* 10 (2001) 329-338.
- [22] J. Ottosson, L. Fransson, K. Hult, *Prot. Sci.* 11 (2002) 1462-1471.
- [23] V. Léonard, L. Fransson, S. Lamare, K. Hult, M. Graber, *ChemBioChem* 8 (2007) 662-667.
- [24] B. Botta, G. Zappia, A. Tafi, M. Botta, F. Manetti, E. Cernia, G. Milana, C. Palocci, S. Soro, G. Delle Monache, *J. Mol. Catal. B: Enzym.* 16 (2002) 241-247.
- [25] R.T. Otto, H. Scheib, U.T. Bornscheuer, J. Pleiss, C. Syldatk, R.D. Schmid, *J. Mol. Catal. B: Enzym.* 8 (2000) 201-211.
- [26] L. Cipolla, M. Lotti, L. De Gioia, F. Nicotra, *J. Carbohydr. Chem.* 22 (2003) 631-644.
- [27] C. Palocci, M. Falconi, S. Alcaro, A. Tafi, R. Puglisi, F. Ortuso, M. Botta, L. Alberghina, E. Cernia, *J. Biotechnol.* 128 (2007) 908-918.
- [28] I. Vallikivi, L. Fransson, K. Hult, I. Järving, T. Pehk, N. Samel, V. Tõugu, L. Villo, O. Parve, O. J. *Mol. Catal. B: Enzym.* 35 (2005) 62-69.
- [29] A.D. MacKerrel Jr., D. Bashford, M. Bellot, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem.* 102 (1998) 3586-3616.
- [30] M.L. Connolly, *Science* 221 (1983) 709-713.
- [31] J. Uppenberg, M.T. Hansen, S. Patkar, T.A. Jones, *Structure* 2 (1994) 293-308.
- [32] D.H.J. Mackay, A.J. Cross, A.T. Hagler, in: G.D. Fasman (Ed.), *Prediction of protein structure and the principles of protein conformation*, Plenum Press. New York & London, 1990, pp. 317-358.
- [33] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, *J. Mol. Graphics Modell.* 21 (2003) 289-307.
- [34] A. Krammer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman, *J. Mol. Graphics Modell.* 23 (2005) 395-407.
- [35] X.D. Li, T.J. Hou, X.J. Xu, *Acta Phys. Chim. Sin.* 21 (2005) 504-507.
- [36] I. Muegge, *J. Med. Chem.* 49 (2006) 5895-5902.
- [37] A. Jain, *Curr. Prot. Pept. Sci.* 7 (2006) 407-420.
- [38] M. Feher, *Drug Discovery Today* 11 (2006) 421-428.
- [39] A.M. Davis, S.J. Teague, *Angew. Chem. Int. Ed.* 38 (1999) 736-749.
- [40] A.M. Kilbanov, *Tibtech* 15 (1997) 97-100.
- [41] P. Trodler, J. Pleiss, *BMC Struct. Biol.* 8 (2008) art. no. 9.
- [42] J. Pleiss, M. Fischer, R.D. Schmid, *Chem. Phys. Lipids* 93 (1998) 67-80.
- [43] A. May, F. Sieker, M. Zacharias, *Curr. Comput. Aided Drug Des.* 4 (2008) 143-153.
- [44] B. Danieli, M. Luisetti, G. Sampognaro, G. Carrea, S. Riva, *J. Mol. Catal. B: Enzym.* 3 (1997) 193-201.

- [45] E. Codorniu-Hernandez, A. Mesa-Ibirico, R. Hernandez-Santiesteban, L.A. Montero-Cabrera, F. Martinez-Luzardo, J.L. Santana-Romero, T. Borrmann, W.D. Stohrer, *Int. J. Quantum Chem.* 103 (2005) 82-104.
- [46] E. Codorniu-Hernandez, A. Rolo-Naranjo, L.A. Montero-Cabrera, *THEOCHEM* 819 (2007) 121-129.
- [47] C.K.Z. Andrade, W.A. Silva, E.R. Maia, J. *Biomol. Struct. Dyn.* 25 (2007) 35-48.
- [48] H. Alonso, A.A. Bliznyuk, J.E. Gready, *Med. Res. Rev.* 26 (2006) 531-568.

2. CONTRIBUTION DE L'ARTICLE

Dans cette partie, nous avons étudié par docking les modes d'ancrage des flavonoïdes glycosylés isoquercitrine et rutine dans la cavité catalytique de la lipase CALB. Le but était de comprendre, au niveau moléculaire, la régiosélectivité expérimentalement observée dans l'acétylation de ces deux composés en utilisant cette lipase comme biocatalyseur. À notre connaissance, cet article a été le premier à relater l'utilisation d'outils de modélisation moléculaire pour étudier les modes d'association de flavonoïdes dans la cavité catalytique d'une lipase.

Une définition de complexe enzyme-substrat « productif » (susceptible de mener à la formation d'un produit) a été proposée, en se basant sur le mécanisme catalytique et des études de modélisation de lipases publiées par d'autres auteurs : l'acétate doit être orienté vers les résidus formant le trou oxyanionique de la CALB (Thr40 et Gln106) et au moins un groupement -OH du flavonoïde doit être à la proximité de deux des résidus de la triade catalytique (Ser105 et His224). L'analyse des résultats des simulations ont montré que les deux flavonoïdes sont stabilisés dans la cavité catalytique de la CALB avec leur partie aglycone ancrée à l'entrée de la cavité. Des liaisons hydrogène formées entre les fonctions phénol des flavonoïdes et les carbonyles peptidiques des résidus Val149, Ile189 et Leu278 contribuent à stabiliser les complexes. Les parties glycosidiques des flavonoïdes s'orientent vers le fond de la cavité. En particulier, lorsque l'acétate est correctement orienté vers le trou oxyanionique (Thr40 et Gln106), seuls les hydroxyles 6''-OH de l'isoquercitrine et 4'''-OH de la rutine sont accessibles aux résidus catalytiques Ser105 et His224. Des trajectoires courtes de dynamique moléculaire (100 ps à 300 K), exécutées sur les complexes productifs identifiés n'ont montré aucune modification conformationnelle importante des substrats dockés, ni dans le cas de l'isoquercitrine ni dans celui de la rutine. Ainsi, l'application de la définition proposée de complexe productif sur les meilleurs complexes obtenus par la procédure de docking/scoring a apporté une explication rationnelle aux résultats expérimentaux de régiosélectivité de la réaction d'acétylation. En outre, l'étude a confirmé que la façon dont les substrats (flavonoïde et donneur d'acyle) se positionnent et s'orientent dans la cavité du site actif de l'enzyme est directement corrélée à la régiosélectivité de la réaction.

3. ÉTUDE COMPLÉMENTAIRE : EVALUATION DES COMPLEXES PRODUCTIFS PAR DYNAMIQUE MOLÉCULAIRE (DM)

Bien que les résultats reportés dans l'article aient apporté une explication plausible à la régiosélectivité observée au cours de l'acétylation de ces deux flavonoïdes glycosylés, la flexibilité de la protéine y est peu exploitée. Or, les complexes enzyme-substrats ne sont pas statiques, pouvant subir des modifications structurales, plus ou moins importantes, en fonction de la flexibilité du substrat et de l'enzyme ou encore du type et de l'intensité des interactions intermoléculaires formées entre eux-ci. Ceci laisse place à un autre questionnement : de telles modifications pourraient-elles être mises en évidence avec des simulations de dynamique moléculaire plus

étendues ? Pour répondre à cette question dans le cas des complexes productifs obtenus, nous avons complété cette étude par des simulations de dynamique moléculaire plus longues (2 ns à 300 K). L'objectif a été de vérifier si les configurations de ces complexes, observées dans les modèles statiques, sont stables au cours du temps ou si d'autres phénomènes liés aux mouvements atomiques peuvent être mis en évidence. Les points de départ pour ces simulations étaient les complexes classifiés de productifs obtenus pour l'isoquercitrine (complexe I1) et pour la routine (complexe R1). Ces complexes ont été solvatés avec une couche d'eau de 5 Å et ensuite soumis à une trajectoire de 2 ns (5 ps de thermalisation + 5 ps d'équilibration + 1990 ps de production), à T = 300K. Des *frames* ont été sauvegardés toutes les 2,5 ps, en totalisant 800 frames. Seules les phases de production ont été considérées pour l'analyse des trajectoires. Pour cette analyse, il fallait se baser d'abord sur un élément commun aux deux systèmes et essentiel pour le processus catalytique. C'est pourquoi nous avons choisi de suivre l'orientation de l'acétate vis-à-vis des résidus du trou oxyanionique. Les distances entre l'atome d'oxygène de celui-ci (Ace:O) et les hydrogènes des résidus Thr40 et Gln106 susceptibles de former des liaisons hydrogène avec lui (Gln106:NH, Thr40:NH et Thr40:OH) ont été mesurées au cours de la simulation (Figure III1).

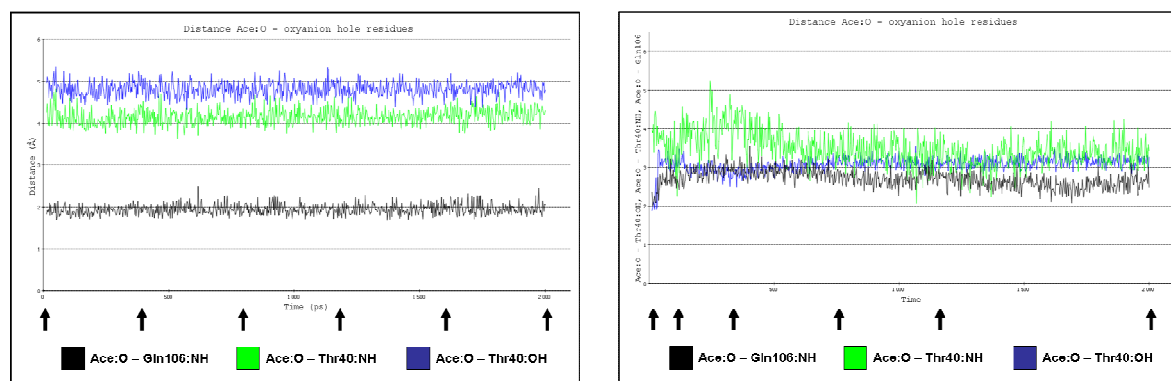


Figure III1 - Suivi des distances entre l'atome d'oxygène de l'acétate (Ace:O) et les atomes Gln106:NH, Thr40:NH et Thr40:OH au cours de la dynamique moléculaire (2 ns) des complexe I1 (à gauche) et R1 (à droite). Les flèches indiquent les frames échantillonnés pour l'analyse.

A partir des graphes correspondant au suivi de ces distances en fonction du temps, des *frames* ont été échantillonnés et superposés pour une analyse plus détaillée des changements conformationnels éventuels des complexes.

Pour le complexe I1 (isoquercitrine), l'orientation de l'acétate vis-à-vis du trou oxyanionique est très stable : elle n'a subi aucun changement significatif au cours de la trajectoire, comme montré dans la Figure III1 et confirmé par la visualisation de l'animation de la trajectoire. Étant donné la stabilité des distances suivies, six *frames* espacés entre eux de 400 ps (0 ps, 400 ps, 800 ps, 1200 ps, 1600 ps et 2000 ps) ont été échantillonnés pour l'analyse (Figure III2 et Tableau III1). Pour le complexe R1 (routine), le seul changement notable, concernant l'acétate, est une augmentation de la distance Ace:O-Thr40:NH entre 200 ps et 400 ps. A part ceci, les trois distances suivies restent plutôt stables à partir de 100 ps. Ainsi, les structures correspondant à 0 ps, 100 ps, 250 ps, 750 ps, 1100 ps et 2000 ps ont été prises pour l'analyse (Figure III3 et Tableau III2).

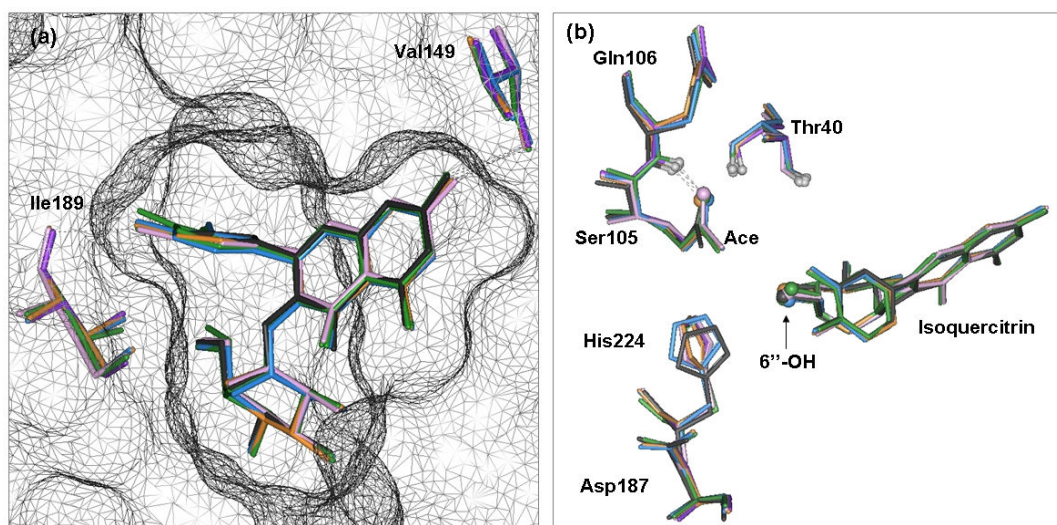


Figure III2 - Superposition des frames échantillonnés dans la trajectoire de I1 à 0 ps (noir), 400 ps (bleu), 800 ps (vert), 1200 ps (orange), 1600 ps (rose) et 2000 ps (violet). Pour une meilleure clarté, seuls les résidus et les atomes d'hydrogène importants pour la discussion sont montrés. En (a), une vue frontale, montrant la cavité, l'isoquercitrine et les résidus avec lesquels elle interagit directement à l'entrée ; en (b), une vue latérale, mettant en évidence la disposition de l'isoquercitrine par rapport aux résidus catalytiques et la disposition de l'acétate par rapport au trou oxyanionique.

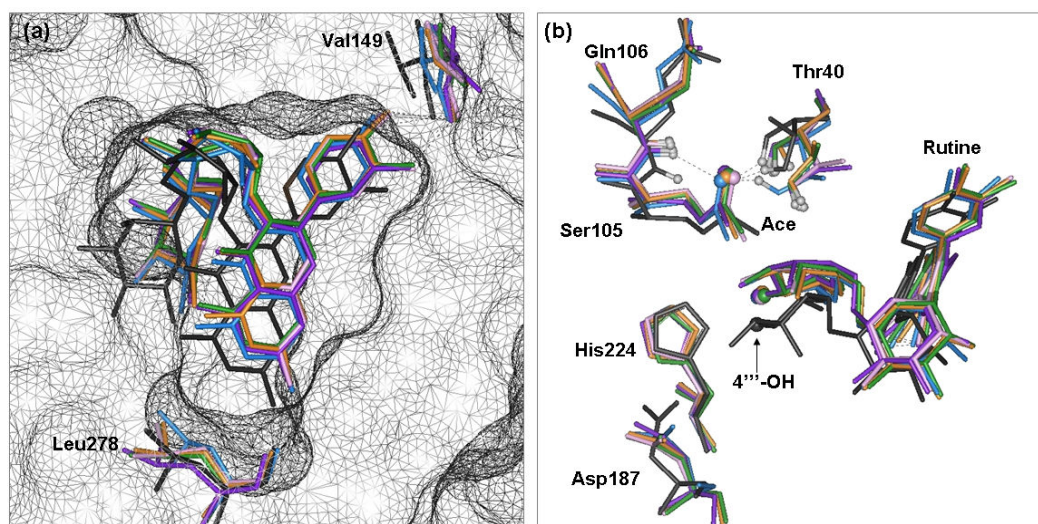


Figure III3 - Superposition des frames échantillonnés dans la trajectoire de R1 à 0 ps (noir), 100 ps (bleu), 250 ps (vert), 750 ps (orange), 1100 ps (rose) et 2000 ps (violet). Pour une meilleure clarté, seuls les résidus et les atomes d'hydrogène importants pour la discussion sont montrés. En (a), une vue frontale, montrant la cavité, la rutine et les résidus avec lesquels elle interagit directement à l'entrée ; en (b), une vue latérale, mettant en évidence la disposition de la rutine par rapport aux résidus catalytiques et l'acétate par rapport au trou oxyanionique.

Tableau III2 - Distances d'intérêt pour le processus catalytique, relevées à différents temps au cours de la trajectoire dans le cas de complexe R1

Conf. (ps)	Asp187:O _D - His224:N _D	His224:N _ε - Ser105:O _γ	Ace:O - trou oxyanionique			4'''-OH	
			Gln106:NH	Thr40:NH	Thr40:OH	His224:N _ε	Ace:C
0	2.50	3.23	1,98	3,94	1,65	4.71	4.34
100	3.89	4.47	2,37	3,29	1,89	6.72	4.24
250	3.84	4.86	2,83	4,31	3,14	6.54	4.31
750	3.87	4.56	2,72	3,93	3,15	6.53	4.52
1100	4.04	4.65	2,78	3,20	3,89	6.74	5.06
2000	3.41	4.85	2,35	3,39	3,11	6.72	4.53

Tableau III1 - Distances d'intérêt pour le processus catalytique, relevées à différents temps de la trajectoire dans le cas du complexe I1

Conf. (ps)	Asp187:O _D - His224:N _D	His224:N _ε - Ser105:O _γ	Ace:O - trou oxyanionique			6''-OH	
			Gln106:NH	Thr40:NH	Thr40:OH	His224:N _ε	Ace:C
0	4.35	3.94	1.83	4.26	4.95	4.63	3.72
400	4.54	3.45	1.97	3.96	4.67	4.13	4.72
800	3.98	4.17	1.96	3.98	4.86	4.58	4.42
1200	3.99	4.04	1.89	4.31	4.75	4.93	4.52
1600	4.09	4.07	1.95	4.42	4.62	4.84	4.44
2000	4.11	3.98	1.99	4.27	4.69	4.73	4.31

3.1. DM du complexe productif pour l'isoquercitrine (I1)

Pour le complexe I1, la valeur maximale de *rmsd* observée au cours de la trajectoire est d'environ 1,0 Å, et se stabilise à partir d'environ 900 ps. Cette faible valeur suggère qu'aucun changement global important n'est survenu dans la structure du complexe. Néanmoins, une analyse plus minutieuse de la trajectoire a été exécutée, visant à évaluer des changements locaux éventuels, principalement dans la région du site actif.

L'analyse conjointe de la Figure III2 et du Tableau III1 montre que la molécule d'isoquercitrine ne subit pas de changement conformationnel ou positionnel important au cours de la simulation. Les liaisons hydrogène 7-OH...Val149:O et 4'-OH...Ile189 :O qui stabilisent la partie aglycone à l'entrée de la cavité catalytique sont stables tout au long de la trajectoire. Le groupement hydroxyle 6''-OH du glucose de l'isoquercitrine demeure le plus proche des résidus Ser105 et His224. Sa distance par rapport à l'atome d'azote ε de l'histidine catalytique (His224:N_ε) varie entre 4,1 Å et 4,9 Å ; celle à l'atome de carbone du carbonyle de l'acétate (Ace:C) oscille entre 3,7 Å et 4,7 Å.

Les résidus catalytiques conservent globalement leur disposition, à l'exception d'une rotation de l'imidazole de l'His224, et d'un léger déplacement de la Thr40. Le retournement de l'hydroxyle de la chaîne latérale vers la molécule de l'isoquercitrine pourrait être attribué à une

attraction de nature électrostatique avec le flavonoïde. En conséquence, les distances de l'atome Ace:O par rapport à la Thr40 sont plutôt élevées (environ 4,0 - 4,4 Å pour Ace:O-Thr40:NH et environ 4,6 - 5,0 Å pour Ace:O-Thr40:OH). Ceci n'engendre cependant pas la déstabilisation de l'acétate, qui garde une orientation productive tout au long de la simulation. Notamment la distance Ace:O-Gln106:NH reste dans l'intervalle 1,8 - 2,2 Å, indiquant qu'une liaison H stable s'est établie entre ces deux atomes.

3.2. DM du complexe productif pour la rutine (R1)

Comme pour I1, le complexe R1 ne subit pas de variation structurale globale importante (*rmsd* maximal de 1,1 Å, stabilisé à environ 900 ps de la trajectoire).

La comparaison des frames superposés dans la Figure III3 montre que la molécule de rutine subit un réarrangement positionnel dans les premiers 200 ps de la trajectoire, se déplaçant vers le résidu Val149. En conséquence, les hydroxyles 3'-OH et 4'-OH peuvent établir une liaison hydrogène avec le carbonyle peptidique de la Val149 (3'-OH...val:O et 4'-OH...Val:O), tout en conservant l'interaction 7-OH...Leu278:O. L'hydroxyle 4'''-OH du rhamnose de la rutine reste le groupement le plus proche des résidues catalytiques Ser105 et His224 au cours de la trajectoire.

En raison du réajustement du flavonoïde dans la cavité, sa distance par rapport à l'atome His224:N_ε augmente d'environ 4,7 Å à environ 6,5-6,7 Å. La distance par rapport à l'atome Ace:C a oscille entre 4,3 et 5,1 Å. Un léger écartement de l'Asp187 par rapport à l'His224 est aussi observé durant les 200 ps initiaux, sans toutefois que la distance Asp187:O_D-His224:N_D ne dépasse 4,0 Å. Le résidu Thr40 subit des oscillations, notamment vers la molécule du flavonoïde, expliquant les fluctuations plus importantes observées pour les distances Ace:O-Thr40:NH (3,2 - 4,3 Å) et Ace:O-Thr40:OH (1,7 - 3,9 Å). La distance Ace:O-Gln106:NH est inférieure et subit des fluctuations plus faibles (2,0 - 2,8 Å). Il en résulte que l'acétate, ne subissant que des faibles oscillations, garde une orientation catalytiquement active vers le trou oxyanionique.

3.3. Conclusions de l'étude complémentaire par DM

Des simulations de DM ont été exécutées sur les complexes identifiés comme productifs d'après l'analyse des résultats de docking de l'isoquercitrine et de la rutine dans la cavité catalytique de la CALB acétylée (complexes I1 et R1, respectivement). Le but de ces simulations était d'évaluer l'évolution temporelle de ces complexes, vérifier leur stabilité et identifier des changements éventuels dans leurs structures. Les faibles variations de *rmsd* obtenues pour la CALB au cours des simulations indiquent que, globalement, cette enzyme est plutôt rigide. Ceci est en accord avec les données de la littérature à ce sujet.

Dans les deux complexes, au cours des trajectoires de 2 ns, il a été observé que le résidu Thr40 subit des oscillations plus marquées par rapport aux autres résidus catalytiques, ce qui pourrait être dû à un effet électrostatique attractif exercé par les flavonoïdes. Néanmoins, ces oscillations de la Thr40 n'ont pas d'effet apparent sur l'orientation catalytiquement active de l'acétate lié à la Ser105 vers les résidus formant le trou oxyanionique (Thr40 et Gln106).

Dans I1, la molécule d'isoquercitrine ne subit pas de changement conformationnel ou positionnel important. La partie aglycone reste ancrée à l'entrée de la cavité, formant des liaisons hydrogène avec les carbonyles peptidiques des résidus Val149 et Ile189. L'hydroxyle 6''-OH du glucose est, tout au long de la simulation, le groupement hydroxyle plus proche de l'acétate (entre 3,7 Å et 4,7 Å) et de l'His224 (entre 4,1 Å et 4,9 Å).

Dans R1, la molécule de rutine ne présente pas de changement conformationnel important, mais subit un réarrangement positionnel, en se déplaçant vers les résidus Val149 et Leu278. Ces deux résidus forment des liaisons hydrogène avec la partie aglycone du flavonoïde tout au long de la trajectoire. En raison de ce déplacement de la molécule, les distances entre ses groupements hydroxyle par rapport à l'His224 augmentent. Néanmoins, le 4'''-OH demeure l'hydroxyle le plus proche de l'acétate (entre 4,3 Å et 5,1 Å) et de l'His224 (entre 6,5 Å et 6,7 Å).

Les tendances observées dans les trajectoires calculées pour les deux complexes sont représentées schématiquement dans la Figure III4, ci-dessous :

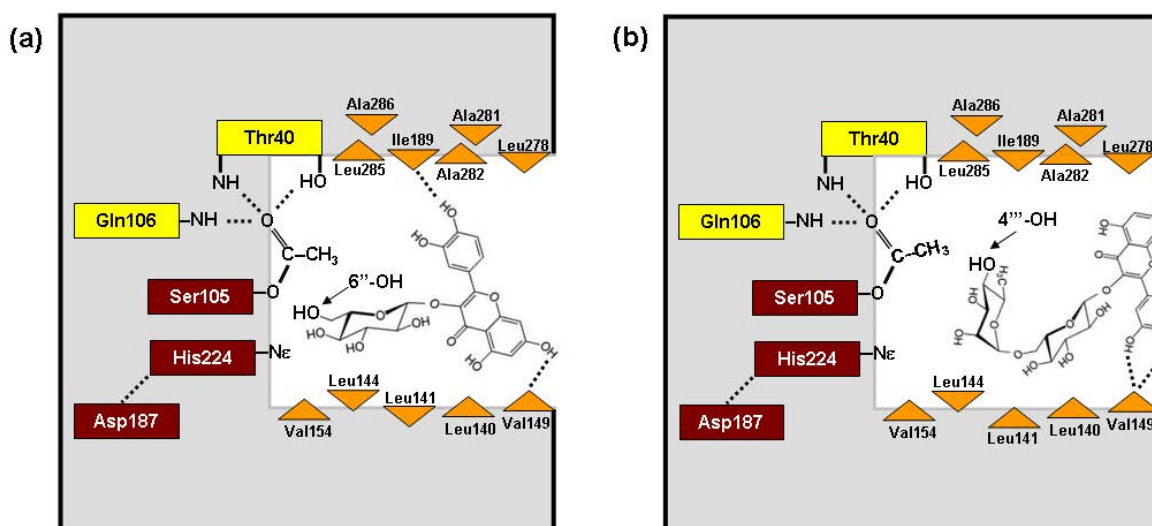


Figure III4 - Représentation schématique des complexes (a) I1 et (b) R1. Dans les deux cas, l'acétate demeure correctement orienté vers les résidus Thr40 et Gln106 et les parties aglycones des flavonoïdes sont stabilisées à l'entrée de la cavité. L'hydroxyle accessible aux résidus catalytiques Ser105 et His224 est le 6''-OH dans le cas de l'isoquercitrine et le 4'''-OH dans celui de la rutine.

Ainsi, ces résultats confirment que le maintien de l'orientation de l'acétate vers le trou oxyanionique est directement corrélé à l'occurrence de la réaction. La régiosélectivité dans l'acylation des deux flavonoïdes glycosylés étudiés, prédite par le docking et observée expérimentalement, a été aussi confirmée par les simulations de dynamique moléculaire : le 6''-OH de l'isoquercitrine et le 4'''-OH de la rutine sont ceux les plus proches de l'histidine catalytique et de l'acétate au cours des trajectoires.

Néanmoins, au vu des distances de stabilisation de ces hydroxyles par rapport aux résidus catalytiques, la limite maximale de 4,0 Å adoptée précédemment pour déterminer si un hydroxyle est susceptible de réagir ou non, doit être revue. En effet, les simulations ont indiqué que

l'hydroxyle acétylé expérimentalement correspond toujours à celui le plus proche des résidus catalytiques dans les modèles, avec des distances à l'histidine catalytique allant de 3,5 Å à 6,5 Å. Afin d'établir une corrélation statistiquement valable entre les valeurs de ces distances et l'occurrence de la réaction, des études expérimentales et des simulations moléculaires combinées, en utilisant un ensemble plus large de flavonoïdes, sont encore nécessaires. L'approche QSAR, par exemple, semblerait être un outil plausible pour entreprendre une telle démarche.

CHAPITRE IV :

**Étude des spécificités des lipases de
Candida antarctica et *Pseudomonas*
cepacia dans l'acétylation du
flavonoïde aglycone quercétine**

1.

INTRODUCTION

Un autre résultat expérimental intéressant obtenu lors de l'acylation des flavonoïdes est l'absence d'activité d'acylation de la lipase B de *Candida antarctica* (CALB) vis-à-vis du flavonoïde aglycone quercétine. Ceci est assez surprenant car cette lipase est connue pour catalyser, en milieu organique, l'estérification d'une gamme très large de substrats, et en particulier l'acétylation régiosélective de l'isoquercitrine et de la rutine (deux analogues glycosylés de la quercétine), comme vu dans le chapitre précédent. Néanmoins, l'acétylation enzymatique de la quercétine a pu être réalisée de manière efficace en utilisant une autre lipase comme biocatalyseur : celle de *Pseudomonas cepacia* (PCL). Dans ce cas, avec l'acétate de vinyle comme donneur d'acyle, le mono-ester quercétine-4'-acétate a été le seul produit détecté dans les premières heures de réaction. À des temps de réaction plus longs, il a été constaté qu'une fraction de ce mono-ester peut subir une deuxième acylation, régiosélective elle aussi, en formant des faibles quantités du diester quercétine-4',7-diacétate. Ces résultats empiriques montrent que la structure des lipases peut jouer un rôle déterminant dans leur spécificité vis-à-vis des flavonoïdes. Plus spécifiquement, ils suggèrent que la reconnaissance des substrats par les sites actifs des lipases CALB et PCL au niveau moléculaire présente des différences significatives, qui se traduisent par la différence de spécificité observée au niveau macroscopique.

À partir de cette hypothèse, nous avons entamé une deuxième étude de simulation, visant à étudier l'influence de la structure des lipase CALB et PCL sur leur spécificité lors de l'acétylation de la quercétine.

- Le même protocole de docking/scoring établi lors de l'étude précédente a été appliqué pour obtenir les positions et les orientations les plus probables de la quercétine dans les cavités catalytiques des lipases CALB et PCL. Les complexes obtenus ont été ensuite étudiés par dynamique moléculaire, dans le but d'obtenir une évaluation temporelle de la configuration globale du complexe et de la stabilité des interactions intermoléculaires au sein de celui-ci.
- Pour évaluer si les complexes modélisés sont susceptibles d'être productifs, les critères utilisés ont été, d'une part, la disposition et l'accessibilité des groupements OH par rapport aux résidus catalytiques, selon la définition proposée dans le chapitre précédent, d'autre part, la nature et la stabilité temporelle des interactions entre les résidus des cavités catalytiques des lipases et les substrats.

Les résultats de cette étude sont reportés ci-après,
sous la forme d'un projet de publication.

2. ARTICLE :

Combined docking and molecular dynamics simulations to enlighten the specificities of *Pseudomonas cepacia* and *Candida antarctica* lipases in quercetin acetylation

Eduardo B. De Oliveira ^a, Catherine Humeau ^{a*}, Latifa Chebil ^a, Evelyne Ronat ^a,
Elaine R. Maia ^b, Bernard Maigret ^c, Mohamed Ghoul ^a and Jean-Marc Engasser ^a

^a Nancy University, Biomolecules Engineering Laboratory, Biocatalysis and Bioprocess Engineering Group, ENSAIA-INPL, 2 av. de la Forêt de Haye, 54500 Vandoeuvre-lès-Nancy, France

^b Laboratorio de Estudos Estruturais Moleculares (LEEM), Instituto de Quimica, Universidade de Brasília, CP 4478, 70904-970, Brasília-DF, Brazil.

^c Laboratoire Lorrain de Recherche en Informatique et ses Applications, CNRS, Nancy Université, BP 239, 54506, Vandoeuvre-lès-Nancy, France

* Corresponding author - phone: (+33) 03.83.59.57.84 / fax: (+33) 03.83.59.57.78

Abstract:

In order to better understand the structural factors determining the specificity and the regioselectivity of lipase-catalyzed acylation of flavonoids, a combined docking and molecular dynamics protocol was applied to investigate the binding of quercetin within the catalytic cavity of two lipases: the *Candida antarctica* lipase B (CALB) and the *Pseudomonas cepacia* lipase (PCL). For both lipases, the initial docking of quercetin in the cavity of the acetyl-enzyme yielded two families of conformers with either the quercetin hydroxylated A or B-ring pointing towards the catalytic residues. Molecular dynamics calculations over a 2 ns time were subsequently performed on the best-scored complex of each family, in order to evaluate the stability of the binding modes and the proximity of the two substrates with the lipase catalytic residues. According to the modelling results, with the PCL lipase the 4'-OH on the quercetin B-ring was found as the only hydroxyl group likely to be acetylated. In this binding mode quercetin is stabilized by hydrophobic aromatic interactions with the cavity walls residues and the serine-bound acetate by hydrogen bond interactions with the oxyanion hole. At the same time the 4'-OH group is located sufficiently close to the catalytic histidine for proton transfer and to acetate for nucleophilic attack. With CALB, on the contrary, due to displacements of both quercetin and serine-bound acetate within the active site, none of the A or B-ring type conformers can be expected to be esterified. These theoretical findings are in agreement with available experimental results.

Keywords:

Bioprocesses, Lipase, Substrate specificity, Quercetin acetylation, Docking, Molecular Dynamics.

1. Introduction

Flavonoids are naturally occurring polyphenolic antioxidants abundantly present in fruits, vegetables, and beverages such as wine and tea. They have recently gained increasing interest because of their beneficial health effects such as anti-inflammatory, cardio-protective and anticancer activities. These compounds have a characteristic fifteen-carbon backbone structure consisting of two aromatic rings (A and B) joined by a three-carbon linked γ -pyrone ring (C). Many of them have glycoside moieties covalently bound to the flavan structure (Havsteen 2002; Boudet 2007). Among the most abundant in human diet one finds quercetin and its glycosylated derivatives isoquercetin and rutin (Figure 1). The antioxidant activity of flavonoids depends on the number and position of functional groups present on their backbone. For instance, the highly reactive hydroxyls on the B-ring have a dominant contribution to their free radical scavenging capacity (Chen et al. 1996; Burda and Oleszek 2001; Da Silva et al. 2004).

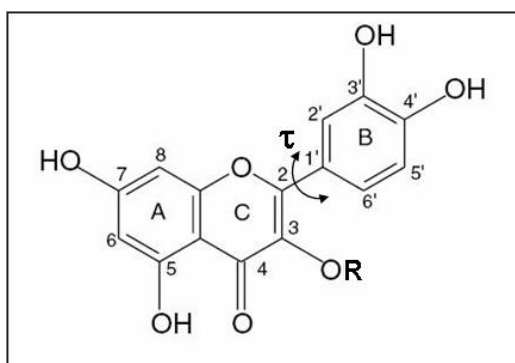


Figure 1 – Chemical structure of quercetin ($R = H$), showing the systematic numbering of the carbon atoms, the nomenclature of the rings and the dihedral angle τ , defined by the atoms $O1-C2-C1'-C2'$. The monoglycosylated isoquercitrin corresponds to $R = \text{glucose}$ and the diglycosylated rutin to $R = \text{rutinose}$.

To be used as nutritional, cosmetic and pharmaceutical active ingredients, acylated derivatives of flavonoids characterized by a higher solubility, stability or bioavailability are often preferred. They can be obtained by grafting one or several fatty acids of variable chain length on the polyhydroxylated flavonoid. It is important to control the regioselectivity of the acylation process in order to preserve the hydroxyl groups responsible for antioxidant activities. Many acylated flavonoids are efficiently synthesized using lipases as biocatalysts in non aqueous media. In such cases, the number and position of the acylated OH groups depends on the nature of the flavonoid and the lipase structure. Most often the sugars hydroxyl groups are preferentially esterified. With some lipases acylation has also been reported to occur on the aromatic rings hydroxyls (Chebil et al. 2006; Villeneuve 2007). At present there is only a limited understanding of structural factors that determine the regioselectivity of the lipase catalyzed acylation of aglycon or glycosylated flavonoids. This constitutes a major bottleneck when considering the use of site-directed mutation to design improved lipases tailored for flavonoids regioselective acylation.

Recently, a first molecular modelling approach was proposed to rationalize the regioselectivity in acetylation of the two flavonoid glycosides isoquercitrin and rutin catalyzed by *Candida antarctica*

lipase B (De Oliveira et al. 2009). Modelling experiments were based on the well established bi-bi ping-pong mechanism of the enzymatic transesterification reaction (Figure 2), which involves two tetrahedral intermediates, the first one resulting from the nucleophilic attack of the catalytic serine on the acylating substrate and the second one coming from the nucleophilic attack of the acyl acceptor substrate on the acyl-enzyme (Pleiss *et al.* 2000).

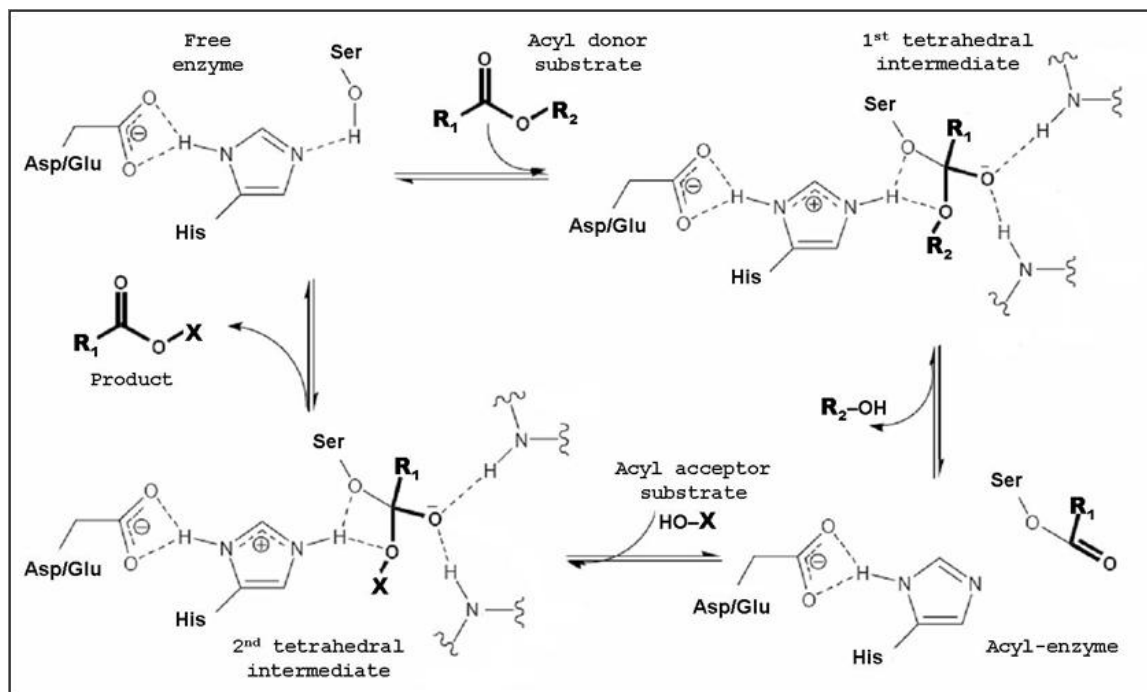


Figure 2 – The reaction mechanism of a lipase-catalyzed transesterification reaction, between the substrates R_1COOR_2 (acyl donor) and XOH (acyl acceptor) to give the products R_2OH and R_1COOX . The involved first tetrahedral intermediate, the acyl-enzyme complex and the second tetrahedral intermediate are indicated.

The protocol considered the binding of the flavonoid to the acetylated lipase to form the second tetrahedral intermediate. It involved a docking procedure to find probable binding configurations of the substrates inside the lipase catalytic cavity combined with classical force fields calculations, in order to assess their positions, orientations and interactions and to evaluate the proximity of the flavonoids hydroxyl groups to the catalytic residues. Computational results showed that for both flavonoids the aglycon part was localized at the entrance of the cavity whereas the sugar moieties pointed towards the bottom of the pocket. In the case of the monoglycosylated isoquercitrin that carries eight hydroxyl groups, only the primary hydroxyl of the glucose part was found close enough to the catalytic histidine and to the serine-bound acetate for the acylation to occur. On the other hand, with the diglycosylated rutin that has ten hydroxyl groups, the models identified a single secondary hydroxyl on rhamnose to be located near the catalytic residues. These findings on the flavonoids docking within the lipase cavity provided an explanation of the experimentally observed regioselectivities of the CALB-catalyzed acetylation of isoquercetin and rutin (Ardhaoui *et al.* 2004; Mellou *et al.* 2005; Katsoura *et al.* 2006; Chebil *et al.* 2007).

In the present study, which is aimed to gain a better understanding of the influence of lipase structure in flavonoid acetylation reactions, the previous molecular modelling protocol was extended by submitting the obtained best-scored complexes to molecular dynamics simulations and analyzing the temporal stability of their configurations and interactions (Alonso *et al.* 2006). The protocol was applied to investigate the binding of the aglycon flavonoid quercetin within the cavity of two lipases: the *Candida antarctica* lipase B (CALB) and the *Pseudomonas cepacia* lipase (PCL).

2. Modelling Methodology

Simulations were carried out on a bi-processor AMD Dual Core 280 2.4 GHz. Docking was performed using the LigandFit module (Venkatachalam *et al.* 2003) of the program-package Discovery Studio version 2.0 (Accelrys, *Inc.*). All molecular mechanics and molecular dynamics calculations were performed with the CHARMM force field (MacKerrel Jr. *et al.* 1998), adopting a 14 Å spherical cut-off for non-bound interactions. In all molecular dynamics simulations, the Verlet-Leapfrog algorithm was employed for the integration of the equations of motion, with a time step of 1 fs.

2.1. Preparation of the systems

2.1.1. Starting structures

The crystal structures of both PCL (Kim *et al.* 1997) and CALB (Uppenberg *et al.* 1994) were downloaded from the Protein Data Bank (PDB entries 3LIP and 1LBS, respectively). In both cases, missing hydrogen atoms were added in their theoretical positions, the catalytic Asp and His residues were assigned as deprotonated. In the case of PCL, the crystal has one single protein chain per unit cell of dimensions $a = 91.3 \text{ Å}$, $b = 47.3 \text{ Å}$, $c = 85.4 \text{ Å}$; $\alpha = 90.0^\circ$, $\beta = 121.4^\circ$, $\gamma = 90^\circ$. The chain system is composed of 320 amino acids, one calcium cation and 193 adsorbed water molecules. The crystal water molecules were kept and the calcium cation was removed. The resulting structure was centred in a rectangular box with optimized dimensions of $a = 60 \text{ Å}$, $b = 56 \text{ Å}$, $c = 60 \text{ Å}$; $\alpha = \beta = \gamma = 90.0^\circ$ and solvated with 4873 water molecules. In the case of CALB, the complex has six chains per unit cell of dimensions $a = 229.5 \text{ Å}$, $b = 95.6 \text{ Å}$, $c = 86.8 \text{ Å}$; $\alpha = \beta = \gamma = 90.0^\circ$. Each chain system is composed of 317 amino acids, one ethylhexylphosphonate inhibitor (HEE) covalently bound to the Ser105, a dimer of N-acetyl-glucosamine covalently bound to the Asn74 and 92 adsorbed water molecules. Only one protein chain and its water molecules were kept; the five others were erased. In the remaining chain, the phosphonate inhibitor and the two N-acetyl-glucosamine dimers were removed. The resulting structure was centred in a rectangular box with optimized dimensions of $a = 64 \text{ Å}$, $b = 59 \text{ Å}$, $c = 63 \text{ Å}$; $\alpha = \beta = \gamma = 90.0^\circ$ and solvated with 6145 water molecules. These new solvated systems were used as starting atomic coordinates set for the subsequent modelling steps.

2.1.2. Relaxation of the systems

The adjusted systems were relaxed following a careful relaxation protocol, in order to preserve the crystalline organization of protein atoms (Da Cunha *et al.* 2004; Andrade *et al.* 2007). In this protocol, constraints and restraints were initially added to the systems and then progressively removed throughout the procedure, totalizing four steps: (a) In order to eliminate initial strains, all protein heavy

atoms were fixed; only protein hydrogen atoms and water molecules were allowed to move; (b) A fix constraint was applied on the protein backbone, harmonic restraints were applied on the side chains and protein hydrogen atoms and water molecules were allowed to move; (c) An harmonic restraint was applied on the protein backbone and all the remaining parts of the system were free to move; (d) Finally, the whole system was minimized without any constraints or restraints. The step (a) was done by performing 1000 Steepest Descent iterations. The steps (b)-(d) were done using the Conjugate Gradient method until the root mean square deviation (rmsd) gradient fell below 0.01, 0.001 and 0.0001 kcal.mol⁻¹.Å⁻¹, respectively. Next, the optimized systems were equilibrated during 1ns of standard dynamic trajectories at 300 K, under periodic boundary conditions, using the NVT thermodynamic ensemble. For both PCL and CALB, the protein structure in the last frame was taken for the construction of the corresponding acetyl-lipase.

2.1.3. Construction of the acetyl-lipase structures

As the mechanism of lipase-catalyzed acylation involves the prior esterification of the catalytic serine by the acyl donor substrate, the side chain hydroxyls of the Ser87 of PCL and Ser105 of CALB were replaced by an acetate, in order to mimic the acetyl-lipase structure (Botta et al. 2002; Palocci et al. 2007; De Oliveira et al. 2009). The acetate carbonyl was oriented towards the oxyanion hole, in order to obtain the catalytic competent orientation of the acyl-enzyme. Then, the energy of these new systems was further minimized by executing 1000 Conjugate Gradient steps, with a fix constraint applied to the protein backbone. The resulting acetyl-PCL and acetyl-CALB structures were taken as targets for the docking of quercetin.

2.2. Docking simulations

2.2.1. Docking and scoring procedures

Quercetin was non-covalently docked against the acetyl-CALB and acetyl-PCL using LigandFit (Venkatachalam et al. 2003), which combines a Monte Carlo conformational search algorithm and a shape comparison filter. In addition to the geometric fit with the binding pocket, LigandFit entails an internal scoring function (DockScore) that assures to keep only the ligand poses having favourable energy of interaction with the protein, taken as the sum of van der Waals energies (represented by a 6-9 Lennard-Jones term) and electrostatic energies (represented by a Coulomb term). For both targets, 15,000 Monte Carlo trials were performed on the quercetin. Only the 50 most favourable poses were retained and scored based on a consensus scoring of six different scoring functions : LigScore1, LigScore2, PLP1, PLP2, PMF and Jain (Feher 2006). Families of poses were defined according to the orientation of quercetin within the binding pocket of the lipases. For each of these families, the best-scored complex was taken for the subsequent modelling steps.

2.2.2. Optimization of the selected complexes

Because LigandFit algorithm treats the protein as a rigid body, the Michaelis-Menten complexes retained in the preceding step were submitted to a careful post-docking optimization, in order to take into account potential structural rearrangements in the proteins due to the presence of

the ligand. As the protein structures were previously relaxed and stabilized, complexes were solvated with a 5 Å layer of water molecules (1446 molecules for PCL complexes and 1321 molecules for CALB ones), in order to speed the calculations up. So, these solvated systems were optimized by three consecutive energy minimizations: (a) A fix constraint was applied on the protein backbone, harmonic restraints were applied on the side chains and protein hydrogen atoms and water molecules were allowed to move; (b) An harmonic restraint was applied on the protein backbone and all the remaining parts of the system were free to move; (c) The whole system was free to move. The Conjugate Gradient algorithm was used in the three steps, with root mean square deviation (rmsd) gradients of 0.01, 0.001 and 0.0001 kcal.mol⁻¹.Å⁻¹, respectively.

2.3. Molecular dynamics simulations

2.3.1. Molecular dynamics procedure

Molecular dynamics (MD) simulations aimed to check the temporal stability of the enzyme-substrate complexes and to evaluate their propensity to lead to the formation of a tetrahedral intermediate and then a quercetin ester. For the optimized complexes, MD trajectories of 2 ns were performed, consisting of 5 ps of heating from 50 K to 300 K, 5 ps of equilibration at 300 K and 1990 ps of production at 300 K. Frames were registered every 1 ps. Only production phases were considered for analysis.

2.3.2. Analysis of the trajectories

The MD trajectories of the complexes were firstly analyzed based on the distances between the acetate carbonyl oxygen and the oxyanion hole residues. Indeed, these distances are indicative of the correct or incorrect orientation of the acetate and thus of the potential reactivity of the enzyme-substrate complexes. When the trajectory was stable with regard to the acetate dynamics, three frames were sampled for superimposition: one in the beginning of the trajectory, one in the middle and one in the end. For unstable trajectories, one frame was sampled to represent each conformational family identified in the graph. The analyses of the whole systems were performed based on the superimposition of the sampled frames and the visualization of the trajectory animations, focusing on the interactions that contribute to stabilize the substrates within the catalytic binding pocket of the enzymes (Alonso *et al.* 2006) and the catalytically important distances (Branneby *et al.* 2004; Svedendahl *et al.* 2008).

3. Results

3.1. Quercetin acetylation catalyzed by Pseudomonas cepacia lipase (PCL)

3.1.1. Quercetin-PCL docking configurations

Docking simulations were first used to identify the possible binding modes of quercetin in the active site of PCL following the formation of the acetyl-lipase intermediate. The PCL catalytic site is approximately 7 Å x 17 Å wide and 8 Å deep. The catalytic triad, formed by the residues Ser87, His286 and Asp264, is located at the center of the cavity bottom. In close vicinity are the oxyanion hole residues Leu17 and Gln88, whose backbone –NH groups can form two hydrogen bonds with the

carbonyl oxygen of the serine-bound acetate. The internal part of the cavity has a hydrophobic region, formed by the residues Pro113, Ser117, Phe119, Val123, Leu164, Leu167 and Val267, and a slightly hydrophilic region, constituted by the residues Thr18, Val26, Leu27, Tyr29, Phe146, Ile290, Gln292 and Leu293. The cavity entrance, consisting of Leu248, Thr251, Val266 and Leu287, is mostly hydrophobic. It entails the $\alpha 5$ helix lid (residues 134-150). The crystal structure used in the simulations corresponds to the opened form of the enzyme.

The fifty most favourable complexes retained after the automated docking procedure can be classified in two families of poses: the complexes with the flavonoid A-ring orientated towards the enzyme core and those with the B-ring orientated towards the enzyme core, pointing towards the catalytic triad. Within each family, the obtained quercetin conformers presented only minor differences in terms of conformation and position inside the enzyme cavity. Therefore, only the top-ranked complex belonging to each family was selected as input structure for the subsequent steps. These complexes were further optimized and then submitted to molecular dynamic simulations. During these simulations, the root mean square deviation (*rmsd*) values relative to the enzyme initial structure were monitored and the maximal value observed for each complex are given in Table 1. The small values observed indicate that the binding of quercetin brought only minor protein rearrangements to the global structure of the lipases. However, in order to verify if local rearrangements in the active site region occurred, a more detailed structural analysis of the trajectories was performed.

Table 1 – Maximal *rmsd* values (Å) for the best scored complex of each orientation family, with regard to the crystal structures of the enzymes.

Protein	Complex	<i>rmsd</i> values (Å)	
		<i>Backbone</i>	<i>Heavy</i>
PCL	A-ring binding mode	0.63	2.91
	B-ring binding mode	0.62	2.91
CALB	A-ring binding mode	0.74	2.96
	B-ring binding mode	0.75	2.96

3.1.2. Dynamics of the A-ring binding mode

The molecular dynamic trajectory of the A-ring binding mode was first analyzed with regard to the distance between the carbonyl oxygen of acetate (Ace:O) and the oxyanion hole residues. The complex presented a relatively stable trajectory over a 2000 ps time period (Figure 3a), characterized by an Ace:O-Gln106:NH distance comprised between 2.5 Å and 3.5 Å and an Ace:O-Leu17:NH distance between 2.0 Å and 3.0 Å. Other molecular dynamics findings are illustrated by superimposing the frames at 0 ps, 1000 ps and 2000 ps (Figures 4a and 4b) and measuring the corresponding relevant interatomic distances (Table 2). During the trajectory, the lipase cavity essentially retained its initial configuration. The catalytic triad residues did not undergo significant displacements as indicated by the weak variations of the distances between Asp264 and His286 and between His286 and Ser87.

Small variations, however, were observed for some of the residues surrounding the wall of the cavity: for instance Leu17 and Leu287 were pushed away from quercetin, while Tyr23 and Phe146 presented the trend to be attracted towards the flavonoid molecule.

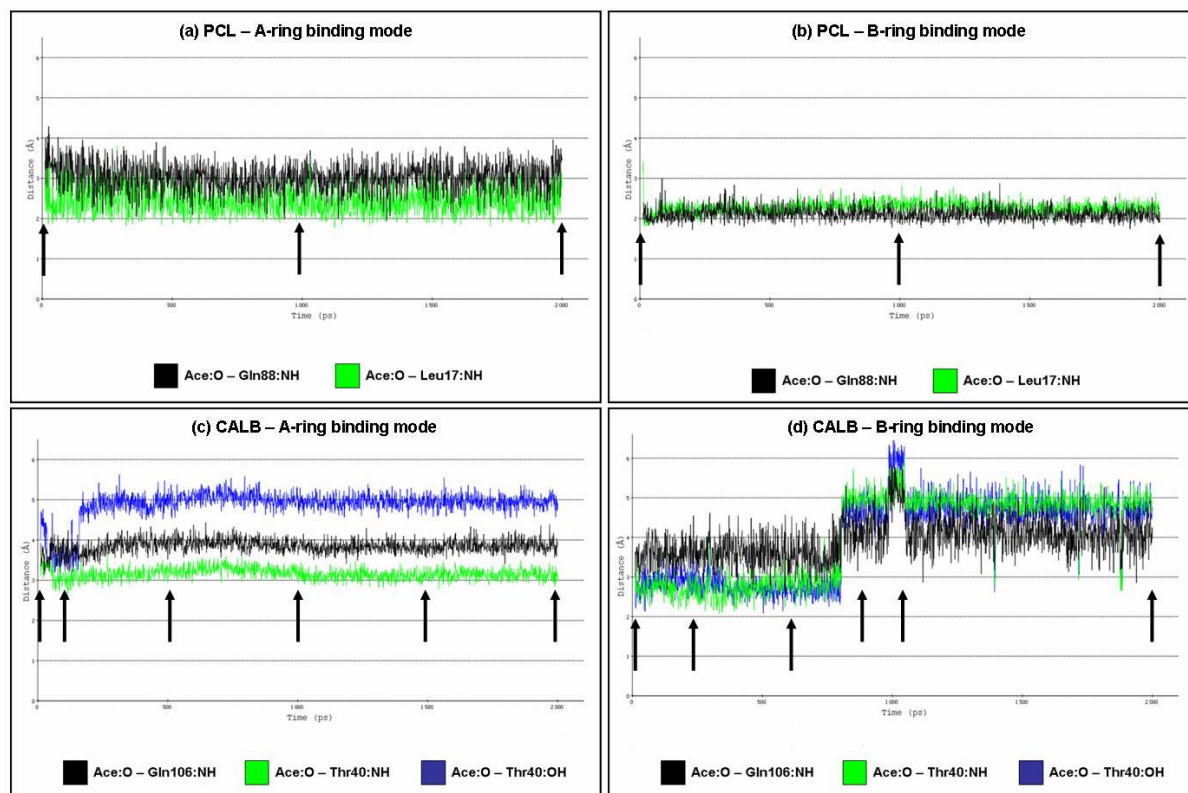


Figure 3 – Fluctuations of the distances between the carbonyl oxygen of the serine-bound acetate and the oxyanion hole residues throughout the 2 ns molecular dynamics trajectories, for each quercetin binding mode within the catalytic cavities of PCL and CALB. The times corresponding to frames sampled for more detailed analysis are indicated by black arrows.

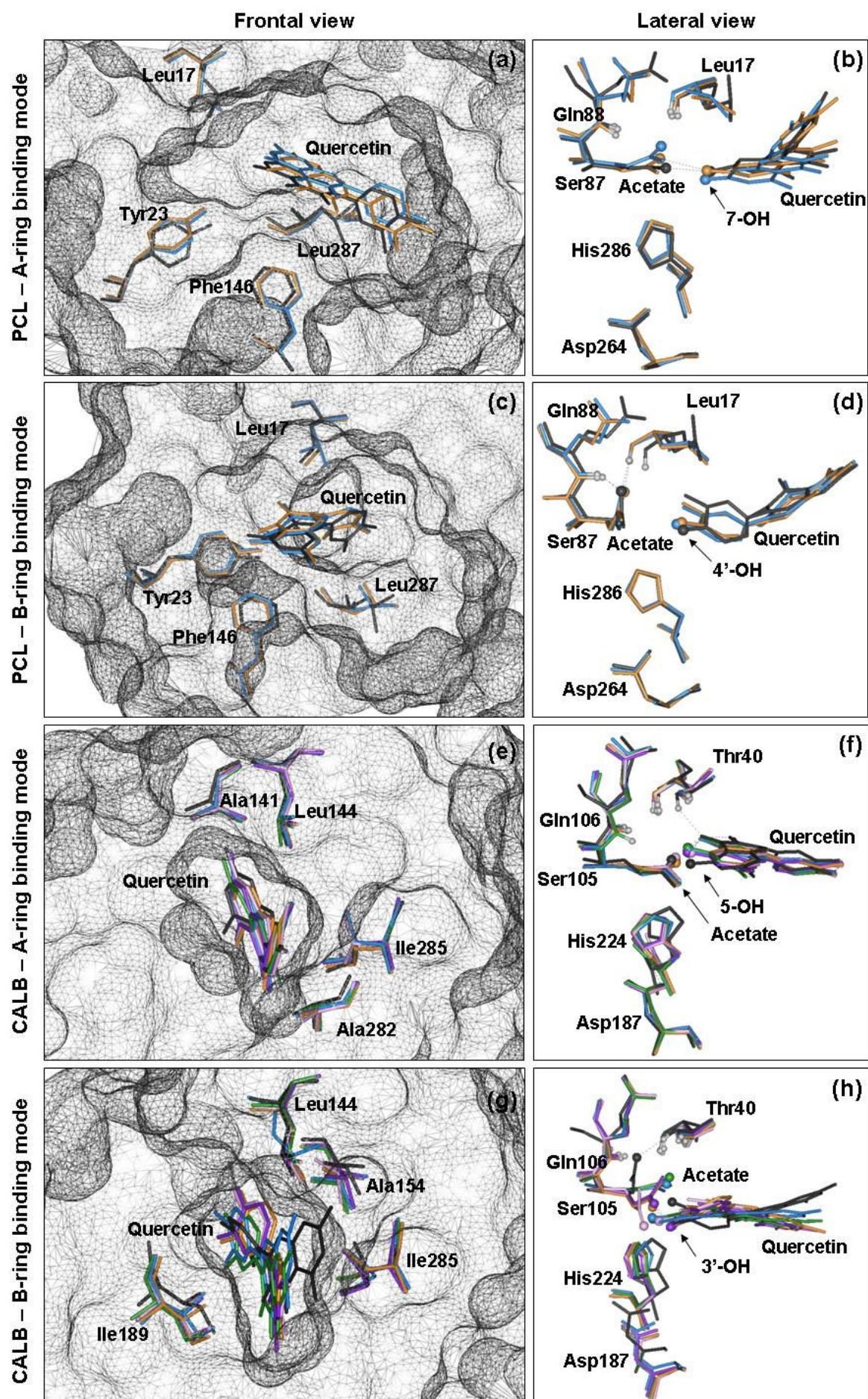


Figure 4 – Superposition of trajectory frames representative of the conformational changes observed during the molecular dynamics simulations for each quercetin binding mode within the catalytic cavities of PCL and CALB. In the first column, frontal views of the binding pocket surfaces are shown, focusing on the docked quercetin and the residues directly interacting with it. In the second column, lateral views of the active site region are presented, showing the docked quercetin, the serine-bound acetate and the residues constituting the catalytic triad and the oxyanion hole. The acetate carbonyl oxygen (Ace:O), the oxygen of the quercetin hydroxyl group which is the closest to the catalytic residues and the H of the oxyanion hole residues are represented by spheres. For clarity other hydrogen atoms and residues not referred in the discussion were hidden. For each case, the colours corresponding to the selected frames are: in (a) and (b), 0 ps (black), 1000 ps (blue), 2000 ps (orange); in (c) and (d), 0 ps (black), 1000 ps (blue), 2000 ps (orange); in (e) and (f), 0 ps (black), 100 ps (blue), 500 ps (green), 1000 ps (orange), 1500 ps (rose), 2000 ps (violet); in (g) and (h) 0 ps (black), 200 ps (blue), 600 ps (green), 900 ps (orange), 1010 ps (rose), 2000 ps (violet).

Table 2 – Catalytically relevant distances (Å) for the substrates and catalytic residues in the A-ring binding mode for PCL

Frame (ps)	Asp264:O _D – His286:N _D	His286:N _ε - Ser87:O _γ	Ace:O – oxyanion hole residues		7-OH	
			Gln88:NH	Leu17:NH	His286:N _ε	Ace:C
0	2.13	3.62	3.42	2.93	4.47	4.07
1000	2.14	3.38	2.79	1.98	3.89	3.97
2000	2.39	3.24	2.89	2.43	4.16	3.45

The quercetin molecule presented a dihedral angle τ of about 54° in the beginning and about 20° in the end, but its position within the cavity was not significantly modified. It is stabilized by hydrophobic interactions between the flavan backbone and the side chains of the neighbouring aliphatic hydrophobic residues Leu17 and Leu287 and aromatic residues Tyr23 and Phe14. In this complex it is the 7-OH group of quercetin that was located closest to the catalytic triad, at a relatively stable distance to His286:N_ε (between 3.9 Å and 4.5 Å) and to the Ace :C (between 3.5 Å and 4.1 Å). The other hydroxyl on the A-ring, the 5-OH, was pointed towards the cavity entrance. As a major finding of the dynamic simulation, the 7-OH functional group is also engaged in a stabilizing hydrogen bond interaction with the acetate carbonyl oxygen. This resulted from the higher proximity of the acetate carbonyl to the quercetin 7-OH (between 1.8 Å and 2.4 Å) than to the backbone -NH of the oxyanion hole residues Gln88 and Leu17 (between 2.0 Å and 3.5 Å).

The results concerning the distances and interactions between PCL and quercetin are schematically summarized in Figure 5a. Quercetin is stabilized within the cavity through interactions with aliphatic and aromatic hydrophobic residues of the cavity walls. The acetate carbonyl is not correctly orientated towards the oxyanion residues and does not form the catalytically important hydrogen bond interactions.

So, even though the quercetin 7-OH is close to the catalytic histidine and serine, the formation of the second tetrahedral intermediate cannot be expected. Therefore, the A-ring binding mode of quercetin with the acetyl-PCL is unlikely to be catalytically favourable.

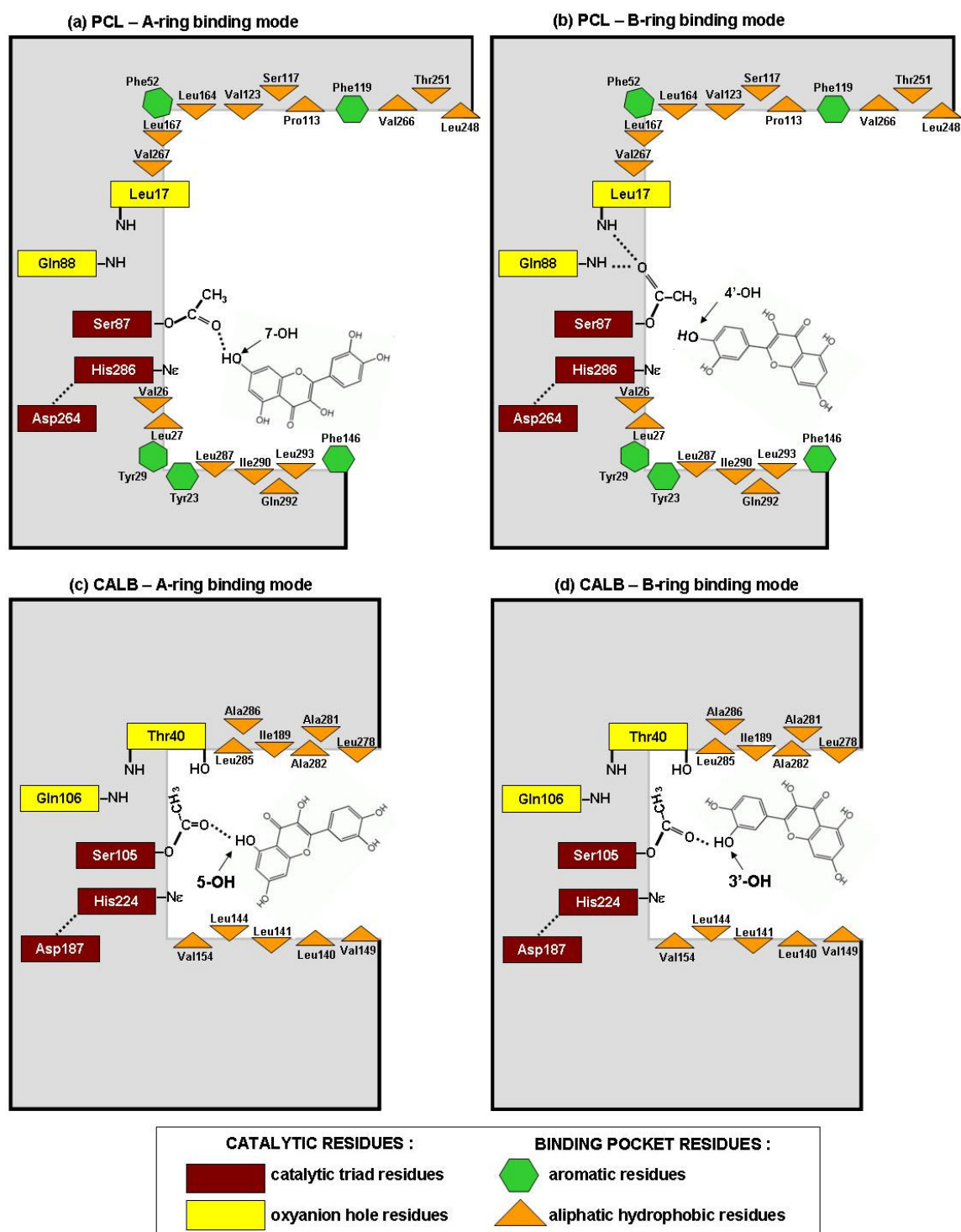


Figure 5 – Schematic representation of the final configuration of the investigated quercetin binding modes within the catalytic cavities of PCL and CALB, after a 2 ns trajectory.

3.1.3. Dynamics of the B-ring binding mode

A similar molecular dynamic analysis was performed on the second binding mode of quercetin, with its B-ring oriented towards the bottom of the cavity. This binding mode displayed a very stable dynamic trajectory, since the Ace:O-Gln106:NH and Ace:O-Leu17:NH distances fluctuated between 1.9 Å and 2.5 Å (Figure 3b), indicating a catalytically competent orientation of the acetate. The positions of the catalytic residues for the 0 ps, 1000 ps and 2000 ps frames are shown in Figures 4c and 4d. The interatomic distances of interest are given in Table 3.

Table 3 -Catalytically relevant distances (Å) for the substrates and catalytic residues in the B-ring binding mode for PCL

Frame (ps)	Asp264:O _D – His286:N _D	His286:N _E - Ser87:O _γ	Ace:O – oxyanion hole residues		4'-OH	
			Gln88:NH	Leu17:NH	His286:N _E	Ace:C
0	2.20	2.60	1.97	3.18	3.45	3.84
1000	2.49	2.72	2.07	2.45	3.55	3.94
2000	2.67	2.83	2.03	2.28	3.86	4.05

As previously observed for the A-ring binding mode, small displacements occurred for the binding pocket wall residues, in close proximity with quercetin: the aromatic Tyr23 and Phe146 were attracted towards the flavonoid molecule, while the aliphatic Leu17 and Leu287 were slightly pushed away. Quercetin showed a rotation of its dihedral angle of about 59° in the beginning decreasing to about 43°, but its position within the cavity did not significantly change, due to the hydrophobic interactions with the aromatic Tyr23 and Phe14 and the aliphatic Leu17 and Leu287. The quercetin 4'-OH was located close to the catalytic triad residues, at stable distances to His286:N_E (between 3.5 Å and 3.9 Å) and to Ace:C (between 3.8 Å and 4.1 Å). The other hydroxyl on the B-ring (3'-OH) pointed towards the cavity entrance. As schematically summarized in Figure 5b, in the B-ring binding mode, quercetin is stabilized in the PCL cavity by its interactions with the wall residues. The serine-bound acetate remains correctly orientated towards the oxyanion residues, with the carbonyl oxygen close enough to form the catalytically efficient hydrogen bonds. Furthermore, in this stable configuration, the quercetin 4'-OH is in close proximity to histidine and acetate, favourable for the formation of the second tetrahedral intermediate.

According to the above results obtained for the two binding modes, the combined docking and molecular dynamic protocol suggests that with the PCL lipase the 4'-OH of quercetin is the only hydroxyl group likely to be acetylated. These modelling results are corroborated by previous published experimental investigations about the PCL-catalyzed quercetin acetylation (Chebil *et al.* 2007): With vinyl-acetate as acyl donor the acetylation of quercetin was found to yield the monoester quercetin-4'-acetate during the first hour of reaction. At longer reaction times, the formation of a small amount of the diester quercetin-4',7-diacetate was also observed, as a result of the consecutive acetylation of the initially formed monoester. So, the modelling predicted regioselectivity of the quercetin acetylation by PCL lipase is in close agreement with experimental results.

3.2. Quercetin acetylation catalyzed by *Candida antarctica* lipase B (CALB)

3.2.1. Quercetin-CALB docking configurations

The previously outlined molecular modelling protocol was applied to the prediction of the regioselectivity of quercetin acetylation catalyzed by a second lipase, the CALB. Its catalytic cavity is a channel with approximately 10 Å x 4 Å of width and 12 Å of length. Differently to most of known lipases, CALB has no lid covering the active site cavity. The triad is composed by the residues Ser105, His224 and Asp187. The oxyanion hole contains the residues Thr40 and Gln106. In this case, in addition to the backbone –NH of these residues, the side chain –OH of Thr40 can form a third hydrogen bond with the tetrahedral intermediate oxyanion. The cavity wall is essentially lined with the aliphatic hydrophobic residues Leu140, Leu141, Leu144, Val149, Val154, Ile189, Leu278, Ala281, Ala282, Ile285 and Ala286. Contrary to PCL, there are no aromatic residues. It also entails a small slightly polar region around the catalytic Ser105, made of the aliphatic polar side chains of Thr40, Asp134 and Gln157.

As in the case of PCL, the docking and scoring procedures of quercetin with the acetyl-lipase target gave fifty complexes, classified in two families: with either the A-ring or the B-ring of quercetin orientated towards the enzyme core. The top-ranked complex of each family were further optimized and then submitted to molecular dynamic simulations. According to their maximal *rmsd* values (Table 1), the binding of quercetin produced only small displacements on the global structure of the protein.

3.2.2. Dynamics of the A-ring binding mode

For this binding mode the monitored variations of the distance between the carbonyl oxygen of acetate (Ace:O) and the oxyanion hole residues are reported in Figure 3c. The trajectory was globally stable. The Ace:O-Gln106:NH distance fluctuated between 3.8 Å and 4.1 Å and the Ace:O-Thr40:NH distance between 2.8 Å and 3.3 Å. For the Ace:O–Thr40:OH distance, a sudden increase from 3.5 Å to 4.8 Å was observed at 200 ps, due to the rotation of the OH group of the Thr40, which is electrostatically attracted by the quercetin 4-C=O. After this, only minor fluctuations occurred. For a more detailed analysis, frames at 0 ps, 100 ps, 500 ps, 1000 ps, 1500 ps and 2000 ps were superimposed (Figures 4e and 4f) and the distances of interest were measured (Table 4).

Quercetin was stabilized by the hydrophobic interactions between its flavan backbone and the side chains of the neighbouring aliphatic Ala141, Leu144, Ala 282 and Ile285 and its position in the cavity remained practically unchanged. No important displacements of these cavity wall residues were observed. The catalytic triad also remained stable, except for the His224, which underwent a small displacement towards the enzyme core, resulting in an increased distance between histidine and quercetin. On the A-ring the 5-OH is the closest hydroxyl group to histidine and acetate. The hydrogen bonds between acetate and the the Thr40 and Gln106 residues, were not established, as reflected by the large distances between the acetate carbonyl oxygen, indicating a non productive orientation of the acetate with regard to the oxyanion hole.

Table 4 – Catalytically relevant distances (Å) for the substrates and catalytic residues in the A-ring binding mode for CALB

Frame (ps)	Asp187:O _D – His224:N _D	His224:N _E - Ser105:O _Y	Ace:O – oxyanion hole residues			5-OH	
			Gln106:NH	Thr40:NH	Thr40:OH	His224:N _E	Ace:C
0	2.92	3.50	3.93	3.73	4.03	5.90	4.02
100	2.61	2.84	3.77	2.98	5.04	6.08	3.83
500	2.66	3.00	3.88	3.25	5.05	6.24	3.90
1000	2.49	3.08	3.95	3.29	4.81	5.93	3.52
1500	2.48	3.14	3.78	3.02	4.92	6.19	3.80
2000	2.24	3.08	3.98	3.25	5.07	6.16	3.88

As schematically summarized in Figure 5c, in its A-ring binding mode quercetin is weakly stabilized within the CALB cavity through interactions with the hydrophobic aliphatic cavity aliphatic residues. This triggers a rotation of the Thr40 OH group towards quercetin. As previously mentioned, in the absence of acetate stabilization in the oxyanion hole, acetylation is unlikely to occur on the 5-OH of quercetin located near the catalytic triad.

3.2.3. Dynamics of the B-ring binding mode

When the B-ring is oriented towards the catalytic triad, the quercetin-lipase complex displayed higher instability, as indicated by the fluctuations of the distances between the carbonyl oxygen of acetate (Ace:O) and the oxyanion hole residues (Figure 3d). In order to analyze these fluctuations more deeply, the frames at 0 ps, 200 ps, 600 ps, 900 ps, 1010 ps and 2000 ps were superimposed (Figures 4g and 4h) and the distances of interest involving the substrates and the surrounding residues were measured (Table 5).

Table 5 – Catalytically relevant distances (Å) for the substrates and catalytic residues in the A-ring binding mode for CALB

Frame (ps)	Asp187:O _D – His224:N _D	His224:N _E - Ser105:O _Y	Ace:O – oxyanion hole residues			3'-OH	
			Gln106:NH	Thr40:NH	Thr40:OH	His224:N _E	Ace:C
0	2.78	3.52	1.27	2.22	2.23	4.90	3.53
200	3.05	3.56	3.69	3.03	3.10	4.52	2.28
600	3.52	3.79	3.97	2.78	2.52	4.45	3.50
900	3.36	2.84	4.38	5.06	4.71	5.74	3.62
1010	3.24	2.64	5.71	5.87	5.97	6.02	3.40
2000	3.16	3.30	4.21	5.03	5.07	6.24	3.75

As shown in Figures 4g and 4h, a first finding is the displacement of quercetin within the cavity. Initially close to the neighbouring Leu144, Ala154 and Ile285 wall residues, the quercetin molecule displaced laterally toward the center of the cavity and glided towards the bottom. At the same time the Ile189 was pushed away, indicating a repulsive interaction with the quercetin. After 600

ps the A-ring part of the molecule glided in the molecule plane, in the perpendicular direction of the precedent movement. A second structural modification occurred at the catalytic triad. It consists in the initial displacement of Asp187 and a progressive movement of His224 towards the enzyme core, and resulted in an increase of the distance between the His224:Ne and the nearest quercetin hydroxyl (3'-OH), from 4.9 Å to 6.2 Å. The most important change observed is the progressive removal of the acetate from the oxyanion hole as induced by the quercetin displacement. From the beginning of the trajectory until about 800 ps, acetate rotated, moving away from the Thr40 and Gln106 oxyanion hole residues. At 800 ps, the acetate carbonyl formed a hydrogen bond interaction with the quercetin 3'-OH, with a simultaneous increase in its distance from the oxyanion hole. During a short period between 1000 ps and 1050 ps, the acetate displayed an opposite orientation with regard to that observed in the beginning of the trajectory. Afterwards it returned to the configuration observed between 800 ps and 1000 ps, until the end of the simulation, as schematized in Figure 5d. Clearly, the B-ring binding mode cannot be expected to lead to the formation of the second tetrahedral intermediate between quercetin and the acetyl-CALB.

According to the modelling results obtained for the two binding modes (A-ring and B-ring), the combined docking and molecular dynamic protocol predicts that the quercetin acetylation is unlikely with CALB as biocatalyst. These modelling predictions are corroborated by previous experimental results, which demonstrated the absence of any acetylation activity of CALB towards quercetin in presence of vinyl acetate as acyl donor (Ardhaoui *et al.* 2004; Chebil *et al.* 2007).

4. Discussion

In this study a combined docking and molecular dynamics protocol was applied to theoretically investigate the lipase catalyzed acetylation of the aglycon flavonoid quercetin. Within the global transesterification reaction mechanism, the modelling only considered the binding of the flavonoid to the acetylated lipase to form the second tetrahedral intermediate. It involved an initial docking procedure which identified two possible orientations of quercetin with either the A or the B-ring pointing towards the catalytic residues, followed by molecular dynamics calculations over a 2 ns time scale, to assess the stability of the binding modes and evaluate the proximity of the two substrates to the lipase catalytic residues. According to the modelling results, with PCL lipase the 4'-OH on the B-ring is the only hydroxyl group on quercetin likely to be acetylated. With CALB, on the contrary, none of the five hydroxyls of quercetin is expected to be esterified. These theoretical predictions are in agreement with experimental results.

The outlined methodology which aims to predict the productivity of quercetin binding modes is based on several stability and proximity criteria. The formation of the second tetrahedral intermediate first requires a sufficient stabilization of the two substrates within the lipase cavity: on the one hand, quercetin has to be stabilized through interactions with the cavity wall residues; on the other hand, the serine-bound acetate must be stabilized through hydrogen bounding with the oxyanion hole residues. Indeed, for the favorable B-ring binding mode within PCL cavity, quercetin was found to be stabilized by its close proximity to two aromatic phenylalanine and tyrosine residues, and the carbonyl of the serine-bound acetate was located at a very close distance (between 2.0 Å and 2.5 Å) to the oxyanion

hole residues. A second criterion for a binding mode to be productive is the proximity of the flavonoid hydroxyl both to the catalytic histidine for proton transfer and to the serine-bound acetate for nucleophilic attack. In the case of the B-ring binding mode, the quercetin 4'-OH group was found at a distance from the His286:N ϵ atom fluctuating between 3.5 Å and 3.9 Å and a distance from the Ace:C atom comprised between 3.8 Å and 4.1 Å. For the other non-productive binding modes, these stability and proximity criteria were not satisfied during the dynamic trajectories.

Substrates stabilization criteria clearly explain the difference of catalytic activities observed with the two lipases PCL and CALB. Whereas they have essentially similar catalytic triad and oxyanion hole residues, the lipases widely differ in the amino acids making up their cavity walls. Within the PCL cavity, composed of aliphatic and aromatic residues, quercetin is strongly stabilized through hydrophobic interactions between its two aromatic A and B-rings and the phenylalanine and tyrosine residues, which is in good agreement with the theoretical affinity order proposed in previous works (Codorniu-Hernandez *et al.* 2005; Codorniu-Hernandez *et al.* 2007). This stabilization contributes to minimize displacements of the catalytic residues and to maintain hydrogen bond interactions between the serine-bound acetate and the oxyanion hole residues. On the contrary, within the CALB cavity, which is only composed of aliphatic amino acids, stabilizing interactions with quercetin are much weaker. During dynamic trajectories, this can result both in important displacements of the flavonoid and the lipase catalytic residues, and in a shift of the serine-bound acetate outside of the oxyanion hole.

An interesting observation during the molecular dynamics simulations of quercetin-PCL and quercetin-CALB complexes is the variation of the dihedral angle of quercetin which reached deviations from the planarity of up to 30°–40° after the stabilization of the trajectories. Similar findings were reported in molecular dynamics investigations of quercetin complexed with other enzymes, such as quercetin-2,3-dioxygenase (Fiorucci *et al.* 2007) and soybean lipoxygenase-3 (Fiorucci *et al.* 2008). According to previous *ab initio* calculations in gas phase, when considered alone the quercetin molecule has a planar geometry (Mendonza-Wilson and Glossman-Mitnik 2004; Mendonza-Wilson and Glossman-Mitnik 2005). The observed conformational change in quercetin structure upon docking to enzymes results from interactions between the flavonoid and amino acid residues surrounding the binding pocket walls and is likely to contribute to the complex stabilization.

The crucial role of the flavonoid stabilization within the lipase cavity can also account for the differences in CALB activities observed for the acetylation of the aglycone quercetin and the glycosylated isoquercetin and rutin (De Oliveira *et al.* 2009). Whereas CALB was shown to unsufficiently stabilize quercetin, according to similar molecular dynamic studies (unpublished results) the lipase efficiently binds the glycan part of both isoquercetin and rutin. Indeed, with isoquercitrin, hydrogen bonds are established between the phenolic hydroxyl groups of the A and B rings and the carbonyl groups of the Val149 and Ile189 residues located at the cavity entrance. Rutin, on the other hand, is also stabilized at the entrance by the formation of hydrogen bonds between its phenolic hydroxyls and the carbonyl groups of Val149 and Leu278. For these stabilized flavonoids, the serine-bound acetate remains in its catalytically competent orientation towards the oxyanion hole residues,

and the hydroxyl group placed at the end of the sugar chain is close to the catalytic histidine and to the acetate and is then expected to be acetylated.

The present docking and molecular dynamics protocol was found to correctly identify the potentially productive flavonoid-lipase binding modes based on criteria of complex stability and substrate-catalytic residues proximity. In order to further assess the reactivity of these favorable complexes, additional quantum mechanics calculations are being pursued. They are expected to yield the activation energies necessary for the formation of the tetrahedral intermediates involving the hydroxyl groups present on either the aromatic rings or on the sugar moieties of the flavonoids and, more generally, to determine the rate limiting steps of flavonoids acylation processes catalyzed by lipases.

REFERENCES:

- Alonso, H., Bliznyuk, A. A. and Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews* 26, 531-568.
- Andrade, C. K. Z., Silva, W. A. and Maia, E. R. (2007). Computational approach for the design of AP1867 analogs: aiming at new synthetic routes for potential immunosuppressant agents. *Journal of Biomolecular Structure and Dynamics* 25, 35-48.
- Ardhaoui, M., Falcimaigne, A., Engasser, J. M., Moussou, P., Pauly, G. and Ghoul, M. (2004). Acylation of natural flavonoids using lipase of *Candida antarctica* as biocatalyst. *Journal of Molecular Catalysis B: Enzymatic* 29, 63-67.
- Botta, B., Zappia, G., Tafi, A., Botta, M., Manetti, F., Cernia, E., Milana, G., Palocci, C., Soro, S. and Delle Monache, G. (2002). Lipase-catalyzed regioselective acylation of resorcin[4]arenes. *Journal of Molecular Catalysis B: Enzymatic* 16, 241-247.
- Boudet, A. M. (2007). Evolution and current status of research in phenolic compounds. *Phytochemistry* 68, 2722-2735.
- Branneby, C., Carlqvist, P., Hult, K., Brinck, T. and Berglund, P. (2004). Aldol additions with mutant lipase : analysis by experiments and theoretical calculations. *Journal of Molecular Catalysis B: Enzymatic* 31, 123-128.
- Burda, S. and Oleszek, W. (2001). Antioxidant and antiradical activities of flavonoids. *Journal of Agricultural and Food Chemistry* 49, 2774-2779.
- Chebil, L., Anthoni, J., Humeau, C., Gerardin, C., Engasser, J. M. and Ghoul, M. (2007). Enzymatic acylation of flavonoids : effect of the nature of the substrate, origin of lipase and operating conditions on conversion yield and regioselectivity. *Journal of Agricultural and Food Chemistry* 53, 9496-9502.
- Chebil, L., Humeau, C., Falcimaigne, A., Engasser, J. M. and Ghoul, M. (2006). Enzymatic acylation of flavonoids *Process Biochemistry* 41, 2237-2251.
- Chen, Z.Y., Chan, P.T., Ho, K.Y., Fung, K.P. and Wang, J. (1996). Antioxidant activity of natural flavonoids is governed by number and location of their aromatic hydroxyl groups. *Chemistry and Physics of Lipids* 79, 157-163.
- Codorniu-Hernandez, E., Mesa-Ibirico, A., Hernandez-Santiesteban, R., Montero-Cabrera, L. A., Martinez-Luzardo, F., Santana-Romero, J. L., Borrmann, T. and Stohrer, W. D. (2005). Essential amino acids interacting with flavonoids: a theoretical approach. *International Journal of Quantum Chemistry* 103, 82-104.
- Codorniu-Hernandez, E., Rolo-Naranjo, A. and Montero-Cabrera, L. A. (2007). Theoretical affinity order among flavonoids and amino acid residues: an approach to understand flavonoid-protein interactions. *Journal of Molecular Structure: THEOCHEM* 819, 121-129.

- Da Cunha, E. F. F., De Castro Ramalho, T. and Maia, E. R. (2004). Interactions of 5-deazapteridine derivatives with Mycobacterium tuberculosis and with human dihydrofolate reductases. *Journal of Biomolecular Structure and Dynamics* 22, 119-130.
- Da Silva, S.L., Da Silva, A., Honorio, K.M., Marangoni, S., Toyama, M;H. and Da Silva, A.B.S. (2004). The influence of electronic, steric and hydrophobic properties of flavonoid compounds in the inhibition of xanthine oxidase. *Journal of Molecular Structure: THEOCHEM* 684, 1-7.
- De Oliveira, E.B., Humeau, C., Chebil, L., Maia, E.R., Dehez, F., Maigret, B., Ghoul, M. and Engasser, J.M. (2009). A molecular modelling study to rationalize the regioselectivity in acylation of flavonoid glycosides catalyzed by Candida antarctica lipase B. *Journal of Molecular Catalysis B: Enzymatic* 59, 96-105.
- Feher, M. (2006). Consensus scoring for protein–ligand interactions. *Drug Discovery Today* 11, 421-428.
- Fiorucci, S., Golebiowski, J., Cabrol-Bass, D. and Antonczak, S. (2007). Molecular Simulations Bring New Insights Into Flavonoid/Quercetinase Interaction Modes. *PROTEINS: Structure, Fonction & Bioinformatics* 67, 961-970.
- Fiorucci, S., Golebiowski, J., Cabrol-Bass, D. and Antonczak, S. (2008). Molecular simulations enlighten the binding mode of quercetin to lipoxygenase-3. *PROTEINS: Structure, Fonction & Bioinformatics* 73, 290-298.
- Havsteen, B. H. (2002). The biochemistry and the medical significance of flavonoids. *Pharmacology & Therapeutics* 96, 67-202.
- Katsoura, M. H., Polydera, A. C., Tsironis, L., Tselepis, A. D. and Stamatis, H. (2006). Use of ionic liquids as media for the biocatalytic preparation of flavonoid derivatives with antioxidant potency. *Journal of Biotechnology* 123, 491-503.
- Kim, K. K., Song, H. K., Shin, D. H., Hwang, K. Y. and Suh, S. W. (1997). The crystal structure of a triacyl glycerol lipase from Pseudomonas cepacia reveals a highly open conformation in the absence of a bound inhibitor. *Structure* 5, 173-185.
- MacKerell Jr., A. D., Bashford, D., Bellot, M., Dunbrack Jr., R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher III, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry* 102, 3586-3616.
- Mellou, F., Lazari, D., Skaltsa, H., Tselepis, A. D., Kolisis, F. N. and Stamatis, H. (2005). Biocatalytic preparation of acylated derivatives of flavonoid glycosides enhances their antioxidant and microbial activity. *Journal of Biotechnology* 116, 295-304.
- Mendonza-Wilson, A.M. and Glossman-Mitnik, D. (2004). CHIH-DFT determination of the molecular structure, infrared and ultraviolet spectra of the flavonoid quercetin. *Journal of Molecular Structure: THEOCHEM* 681, 71-76.
- Mendonza-Wilson, A.M. and Glossman-Mitnik, D. (2005). CHIH-DFT study of the electronic properties and chemical reactivity of quercetin. *Journal of Molecular Structure: THEOCHEM* 716, 67-72.
- Palocci, C., Falconi, M., Alcaro, S., Tafi, A., Puglisi, R., Ortuso, F., Botta, M., Alberghina, L. and Cernia, E. (2007). An approach to address Candida rugosa lipase regioselectivity in the acylation reactions of trytilated glucosides. *Journal of Biotechnology* 128, 908-918.
- Pleiss, J., Fischer, M., Peiker, M., Thiele, C. and Schmid, R.D. (2000). Lipase engineering database: understanding and exploiting sequence - structure - function relationships. *Journal of Molecular Catalysis B: Enzymatic* 10, 491-508.
- Svedendahl, M., Carlqvist, P., Branneby, C., Allnér, O., Frise, A., Hult, K., Berglund, P. and Brinck, T. (2008). Direct epoxidation in Candida antarctica lipase B studied by experiment and theory. *ChemBioChem* 9, 2443-2451.

- Uppenberg, J., Hansen, M. T., Patkar, S. and Jones, T. A. (1994). The sequence, crystal structure determination and refinement of two forms of lipase B from Candida antarctica. *Structure* 2, 293-308.
- Venkatachalam, C. M., Jiang, X., Oldfield, T. and Waldman, M. (2003). LigandFit : a novel method for the shape-directed rapid docking of ligands to protein active sites. *Journal of Molecular Graphics and Modelling* 21, 289-307.
- Villeneuve, P. (2007). Lipases in lipophilization reactions. *Biotechnology Advances* 25, 515-536.

3. CONTRIBUTION DE L'ARTICLE

Dans cet article, nous avons proposé une explication, jusqu'alors manquant dans la littérature, de la différence de spécificité des lipases PCL et CALB dans l'acylation de la quercétine. Pour ce faire, deux approches complémentaires de modélisation moléculaire ont été utilisées. D'abord, des simulations de docking ont été exécutées pour identifier les modes d'association les plus probables - conformations, orientations et interactions - des substrats dans les cavités catalytiques de ces enzymes. Pour les deux lipases, deux orientations possibles ont été trouvées pour la quercétine : avec le cycle A ou avec le cycle B pointé au fond de la cavité, proche des résidus catalytiques. Ensuite, ces quatre complexes obtenus ont été ré-solvatés et soumis à des simulations de dynamique moléculaire, afin d'évaluer leur stabilité temporelle. À partir de l'analyse des trajectoires calculées, les événements moléculaires qui pourraient expliquer la différence de spécificité entre ces deux lipases vis-à-vis de la quercétine ont été identifiés.

Deux conclusions importantes ressortent de cette étude. La première conclusion est que présence de résidus aromatiques dans la cavité catalytique de la lipase contribue significativement à la stabilisation de la quercétine. Ceci est dû au fait que ce type de résidu interagit plus fortement que les résidus aliphatiques hydrophobes avec les noyaux aromatiques des flavonoïdes. En effet, les parois de la cavité de la CALB ne présente pas de résidu aromatique, tandis que celle de la PCL contient ce type de résidu. En particulier, les résidus Tyr23 et Phe146 de cette dernière interagissent directement avec les noyaux aromatiques de la quercétine. Des fluctuations positionnelles plus marquées ont été observée pour les substrats dans la cavité de CALB, malgré le fait que celle-ci soit plus étroite que celle de la PCL. Ceci indique que l'intensité des interactions intramoléculaires dans la cavité influence plus la stabilisation du complexe que le degré de liberté spatiale des substrats. Ou de façon plus formelle, ces résultats suggèrent, au moins qualitativement, que les contributions enthalpiques sont plus déterminantes que les entropiques pour la stabilisation de la quercétine dans les cavités de ces deux lipases. La deuxième conclusion est que la prédiction de la réactivité d'un complexe acétyle-lipase/flavonoïde, obtenu par docking, est plus fiable lorsqu'elle inclut aussi une dimension temporelle. Autrement dit, un complexe « statique » présentant au départ une configuration potentiellement productive (l'oxygène de l'acétate orienté vers le trou oxyanionique et un hydroxyle du flavonoïde accessible simultanément à la sérine et à l'histidine catalytiques) doit garder stable cette configuration sur une période de temps, afin de pouvoir être considéré effectivement comme productif. Étudier les complexes obtenus par docking par dynamique moléculaire s'avère indispensable, car ces simulations permettent d'évaluer la stabilité temporelle des complexes. En effet, la rotation de l'acétate par rapport au trou oxyanionique de la CALB, qui peut expliquer l'absence d'activité d'acylation de cette enzyme vis-à-vis de la quercétine, n'a pu être mise en évidence qu'au cours de les trajectoires dynamiques calculées pour les complexes ES correspondants.

4. REMARQUE COMPLEMENTAIRE : REACTIVITE DE LA QUERCETINE

Au cours de cette étude, la différence de spécificité des lipases PCL et CALB vis-à-vis de la quercétine a été corrélée aux différences structurales des deux lipases : ces différences engendrent des interactions intermoléculaires qui stabilisent mieux le flavonoïde dans la cavité catalytique de la PCL, et perturbent moins l'orientation de l'acétate vers le trou oxyanionique. On remarque cependant que dans les quatre complexes, l'OH de la quercétine le plus proche des résidus histidine et sérine catalytiques n'est pas le même : pour la PCL, ce sont le 7-OH (orientation A) et le 4'-OH (orientation B), tandis que pour la CALB, ce sont le 5-OH (orientation A) et le 3'-OH (orientation B).

Or, les travaux de Fiorucci *et al.* (2007), ont indiqué que les hydroxyles de la quercétine n'ont pas tous la même réactivité : les valeurs d'énergie libre de déprotonation (ΔG) en phase gaz pour la quercétine, calculées au niveau B3LYP/6-31+G*, sont : 5-OH (cycle A), 338,0 kcal.mol⁻¹ ; 7-OH (cycle A) : 322,6 kcal.mol⁻¹ ; 3'-OH (cycle B) : 324,4 kcal.mol⁻¹ ; 4'-OH (cycle B) : 317,5 kcal.mol⁻¹. Plus cette valeur est élevée, plus l'arrachement du proton est difficile. Ainsi, dans le cycle A, le 5-OH est intrinsèquement moins réactif que le 7-OH. Dans le cycle B, le 3'-OH est moins réactif que le 4'-OH.

En conséquence, la non réactivité de la quercétine avec la CALB pourrait être attribuée à deux facteurs : d'une part, la mauvaise stabilisation de la molécule dans la cavité, comme montré par notre article (contribution de l'enzyme) ; d'autre part, le fait que pour cette enzyme seuls les hydroxyles intrinsèquement plus réactifs de la quercétine ne sont pas accessibles aux résidus catalytiques (contribution du substrat).

**Étude de la réactivité de la quercétine,
l'isoquercitrine et la rutine avec la
lipase B de *Candida antarctica* comme
biocatalyseur pour leur acetylation**

1.

INTRODUCTION

Dans les deux chapitres précédents, nous avons relaté les résultats obtenus au cours de simulations de docking et de dynamique moléculaire, pour étudier les modes d'association des flavonoïdes dans les cavités catalytiques des lipases lors de l'acétylation de ces composés. De façon succincte, les résultats de ces simulations ont permis de corréler l'occurrence de la réaction d'acétylation aux trois facteurs suivants : (1) l'orientation de l'acétate lié à la sérine catalytique par rapport aux résidus formant le trou oxyanionique, ce qui détermine l'aptitude de l'acétyl-lipase à recevoir l'attaque nucléophile de l'hydroxyle du flavonoïde ; (2) la position et l'orientation du flavonoïde dans la cavité du site actif de l'enzyme, ce qui influence l'accessibilité des groupements hydroxyles aux résidus catalytiques ; (3) la nature et la stabilité des interactions établies entre les deux substrats et les résidus de la cavité du site actif, ce qui influence la stabilité du complexe et, plus particulièrement, le maintien de celui-ci dans une configuration catalytiquement productive. Les modèles obtenus ont fourni des explications structurales plausibles pour la régiosélectivité stricte observée lors de l'acétylation des flavonoïdes glycosylés isoquercitrine et rutine catalysée par la lipase B de *Candida antarctica* (CALB), ainsi que pour l'absence d'activité de cette lipase pour l'acétylation du flavonoïde aglycone quercétine. Toutefois, ces explications ne tiennent pas compte du fait que pour chacun des flavonoïdes étudiés, les hydroxyles les plus proches des résidus catalytiques sont chimiquement différents : le 3'-OH de la quercétine est une fonction phénol, le 6''-OH de l'isoquercitrine est un alcool primaire et le 4'''-OH de la rutine est un alcool secondaire. Ces hydroxyles peuvent donc avoir des réactivités chimiques différentes.

En conséquence, afin d'examiner la prédisposition de ces trois types d'OH des flavonoïdes à former une liaison ester avec le carbone trigonal de l'acétate, une étude complémentaire basée sur une approche de chimie quantique a été réalisée.

- À partir des complexes obtenus par docking, les substrats (flavonoïde et acétate) et des fragments des résidus catalytiques de la CALB ont été sélectionnés et isolés, en conservant la disposition spatiale de leurs atomes. Les mini-systèmes ainsi obtenus ont été optimisés par la méthode quantique de la Théorie Fonctionnelle de la Densité (DFT).
- La prédisposition des trois hydroxyles (3'-OH de la quercétine, 6''-OH de l'isoquercitrine et 4'''-OH de la rutine) à être acétylés a été évaluée sur la base des modifications de géométrie des mini-systèmes et des charges des atomes directement impliqués dans la formation du deuxième intermédiaire tétraédrique, suite au transfert du proton du flavonoïde à l'His224. Lorsque ces intermédiaires ont été formés, la variation énergétique subie par les mini-systèmes a été évaluée.

Les résultats préliminaires de cette étude sont reportés ci-après,
sous la forme d'un projet de publication.

2. ARTICLE :

Density Functional Theory (DFT) modelling of the *Candida antarctica* lipase B selectivity in quercetin, isoquercitrin and rutin acetylation

Eduardo B. De Oliveira ^a, Catherine Humeau ^{a*}, Elaine R. Maia ^b,
Mohamed Ghoul ^a, Jean-Marc Engasser ^a

^a Nancy University, Biomolecules Engineering Laboratory, Biocatalysis and Bioprocess Engineering Group, ENSAIA-INPL, 2 av. de la Forêt de Haye, 54500 Vandoeuvre-lès-Nancy, France

^b Laboratorio de Estudos Estruturais Moleculares (LEEM), Instituto de Química, Universidade de Brasília, CP 4478, 70904-970, Brasília-DF, Brazil.

* Corresponding author - phone: (+33) 03.83.59.57.84 / fax: (+33) 03.83.59.57.78

Abstract:

Flavonoids are naturally occurring polyphenolic antioxidants of increasing nutritional and pharmaceutical interest. By using lipase biocatalysts, they can be selectively converted into acylated derivatives with improved stabilities and bioavailabilities. In order to better understand the structural factors determining the substrate selectivity of these lipase-catalyzed flavonoid acetylation, in previous studies we described the application of an enzyme-substrate docking protocol to identify the favourable binding modes of the aglycon flavonoid quercetin and its glycosylated derivatives isoquercitrin and rutin within the catalytic cavity of *Candida antarctica* lipase B (CALB). The aim of the present work was to use Density Functional Theory (DFT) to determine the reactivity of the flavonoids in these binding modes. The quantum chemical calculations indicated that the docked quercetin is not expected to react with the serine bound acetate. On the contrary, with isoquercitrin and rutin, an ester bond can be formed between the acetate and the closest sugar hydroxyl (the glucose 6"-OH for isoquercitrin and the rhamnose 4'''-OH for rutin). These theoretically predicted lipase selectivities are in agreement with experimental results.

Keywords:

Candida antarctica lipase B, Docking, DFT, Flavonoid acylation, Substrate selectivity.

1. Introduction

Flavonoids are the most abundant family of plant polyphenols. They have a characteristic fifteen-carbon backbone structure consisting of two aromatic rings (A and B) linked through three carbons that usually form an oxygenated heterocycle (C ring). In nature, many occur as conjugates, with the flavonoid aglycon linked to a variable sugar moiety. In recent years flavonoids have attracted strong interest for their antioxidant properties and their beneficial health effects [1]. To be used as nutritional, cosmetic and pharmaceutical active ingredients, acylated derivatives of flavonoids characterized by a higher solubility, stability or bioavailability are often preferred. Combining flavonoids and fatty acids, specially unsaturated, can also yield novel active ingredients with improved functional properties. The flavonoid analogues are obtained by grafting one or several acyl groups on the polyhydroxylated aglycon or glycosylated molecule [2, 3]. As several biological activities of flavonoids depend on the number and position of their hydroxyl groups, it is important to control the regioselectivity of the acylation process in order to preserve the hydroxyls responsible for these activities [4, 5]. Acylated flavonoids can be efficiently synthesized with lipase biocatalysts in non-aqueous media. For the design of such processes, a key step is the selection of an enzyme able to catalyze the reaction between the flavonoid and the acyl donor with the desired regioselectivity [6]. The biocatalyst can be selected among available microbial lipases. Alternatively it may be a novel engineered lipase obtained by mutation of an existing enzyme. For now, due to the limited understanding of structural factors that determine the selectivity of the flavonoid acylation processes, a rational lipase selection or engineering approach can hardly be envisioned.

During the past decade, molecular modelling techniques have been increasingly used to investigate the mechanisms of enzyme catalysis [7]. Applied to lipase-catalyzed reactions, modelling studies based on classical or quantum mechanics approaches contributed both to identify reaction intermediates and to assess the role of their catalytic amino acids in controlling the bioconversion selectivity [8-13]. In previous studies, we described the application of an enzyme-substrate docking protocol to rationalize the regioselectivity of the lipase-catalyzed acetylation of the flavonoid quercetin and its two glycosylated conjugates isoquercitrin (quercetin-3-O-glucose) and rutin (quercetin-3-O-glucose-rhamnose) [14], whose chemical structures are represented in Figure 1. The modelling was based on the well-known bi-bi ping-pong mechanism of the enzymatic esterification reaction [15], which involves two tetrahedral intermediates: the first one resulting from the nucleophilic attack of the catalytic serine on the acylating substrate and the second one coming from the nucleophilic attack of the acyl acceptor substrate on the acyl-enzyme. When using *Candida antarctica* lipase B (CALB) as biocatalyst, the docking protocol identified a single favourable binding mode of quercetin within the lipase cavity after the initial formation of the acetyl-lipase. In this configuration the 3'-OH group on the B-ring is in close proximity to the catalytic Ser105 and His224 residues, while the Ser105-bound acetate is stabilized in the oxyanion hole (Thr40 and Gln106 residues). The two glycosylated flavonoids isoquercitrin and rutin, on the other hand, were found to bind through their flavan ring at the cavity entrance, while their sugar moieties were pointing towards the bottom of the pocket. For isoquercitrin, only the primary hydroxyl on the glucose part (6''-OH) is close to the catalytic histidine and to the serine attached acetate, whereas with rutin, the modelling identified a single secondary

hydroxyl on rhamnose to be located near the catalytic residues. The aim of the present work was to use additional quantum mechanical calculations to determine if chemical acetylation was possible for these binding modes of quercetin, isoquercitrin and rutin in the CALB cavity.

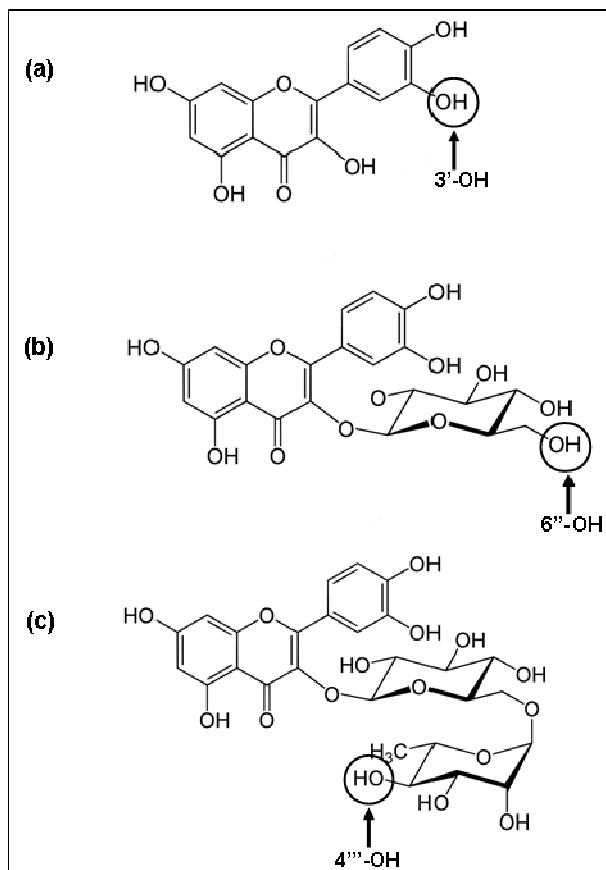


Figure 1 – Chemical structures of the flavonoids (a) quercetin and its glycosylated analogues (b) isoquercitrin and (c) rutin. The hydroxyl groups accessible to the lipase catalytic residues, according to the complexes obtained by docking simulations, are indicated by arrows.

In the past, computational *quantum* chemical approaches, such as *ab initio* and Density Functional Theory (DFT), were applied to investigate the pathway of lipase catalyzed reactions based on simplified model system of the lipase catalytic site [16]. A few quantum chemical modelling studies also aimed at predicting the enantioselectivity of esterification and hydrolysis reactions. For instance, semi empirical and *ab initio* calculations were used to study the selectivity of secondary alcohols esterification and hydrolysis of their esters by *Burkholderia cepacia* lipase (BCL) [17]. Starting from optimized enzyme-substrate complex, the model investigated the hydrogen transfer from histidine to serine oxygen during the formation of the ester and from the catalytic histidine to the alcohol oxygen during hydrolysis. Whereas in the case of the fast reacting substrate enantiomers all the identified binding modes led to the formation of products, which did not happen with the slow reacting ones. In another approach, combined quantum semi empirical and molecular dynamics calculations were performed on CALB to determine its enantiomeric preference in hydrolysis and esterification reactions of naproxen [18]. A simplified semi empirical method was first used to identify the transition states and intermediates along the reaction pathway. The hydrogen bonds within the CALB-enantiomer

complexes were then analyzed by molecular dynamics simulations and found more stable for the fast reacting enantiomeric substrates than for the slow reacting enantiomers.

In the present study, the reactivity of the binding modes of quercetin, isoquercitrin and rutin in the active site of CALB, previously identified by docking, was investigated by the DFT method. The considered model system entails fragments of the lipase catalytic triad and oxyanion hole residues as well as the whole flavonoid and acetate substrates. The changes in interatomic distances were examined, as well as bond orders and charge distributions triggered by a proton transfer from the docked flavonoid hydroxyl to the catalytic His224. For productive binding modes the energy barrier between the formed tetrahedral intermediate and the initial state of the unbound flavonoid was also evaluated.

2. Methodology

All the calculations were performed on a processor Intel Core Duo CPU 6550 2.33 GHz, using the DFT DMol3 module of the software-package Materials Studio version 4.4 (Accelrys, Inc.).

2.1. Definition of the systems

The optimized Michaelis complexes consisting of the flavonoid rutin, isoquercitrin and quercetin docked in the acetyl-CALB, were obtained as described in our previous paper [14]. These complexes were the departure structures to construct the mini-systems for the DFT calculations. Starting from these complexes, the entire flavonoid, the entire Ser105-bound acetate and fragments of the catalytic residues directly implicated in the reaction were conserved. These kept residues fragments are: the side chain carboxylate of Asp187, the imidazole ring of His224, the backbone and side chain of Ser105, the backbone -NH of Gln106, the backbone C α and -NH and the entire side chain of Thr40. the other parts of the systems were rubbed out. The remaining free valences in the systems were then filled with methyl groups (-CH₃). Simple hydrogen atoms could have been used to fill the free valences in order to further reduce the size of the systems [18]. Methyl groups were preferentially chosen to avoid the formation of artificial and undesired hydrogen bonds during the modelling procedure. The resulting systems thus included a total of 135, 115 and 94 atoms in the cases of rutin, isoquercitrin and quercetin, respectively. These systems were taken as input structures of the “reactant states” (RS) for the DFT structure optimizations.

2.2. DFT calculations

For the DFT optimizations of the RS structures, the spatial coordinates of the non-hydrogen atoms that constituted the protein backbone and the carbon atoms of the added methyl groups were kept frozen, in order to conserve the spatial disposition of the residues and substrates in the complexes [11, 12]. The total charge of the systems was assigned as -1, due to the deprotonated Asp187 carboxylate. A global orbital cut-off of 3.7 Å was adopted. The exchange-correlation functional approximation GGA-PW91 [20] and the DNP basis set [20] were used. This basis set includes double-numerical basis functions together with polarization functions (i.e., with angular momentum one higher than that of the highest occupied orbital in free atom). It has been demonstrated that DNP is

comparable in quality to the well-known Gaussian basis set 6-31G** [21]. The convergence tolerances were set at 1.0×10^{-5} Ha (Hartree) for energy, 3.0×10^{-3} Å for the maximum displacement, and 4.0×10^{-3} Ha.Å⁻¹ for the maximum force. The optimized structures were verified to have no imaginary vibrational frequencies. Thereafter, starting from the optimized RS structures, the “intermediates structures” (Int) were constructed according to the reaction mechanism, as follows: the flavonoid hydroxyl closest to the catalytic His224 was deprotonated, the His224:Nε atom was protonated and the acetate carbonyl carbon (Ace:C) was assigned as sp³. After these modifications, the systems were optimized again using the same parameters than those used for the corresponding RS optimizations. The energy differences (ΔE) between the reactant states (RS) and the intermediates (Int) were calculated by simple difference between the energies of the corresponding optimized structures [16], with the conversion factor 1 Ha = 627.51 kcal.mol⁻¹. Atomic charge distribution was determined by the Mulliken population analysis approach [22].

3. Results

In this study the reactivity of the binding modes of the three flavonoids quercetin, isoquercitrin and rutin in the CALB cavity, previously identified by a docking protocol, were analyzed by quantum DFT simulations. The mini-systems constructed starting from the complexes consisting of the flavonoids docked in the catalytic cavity of the acetyl-CALB were optimized, by adopting the PW91 exchange-correlation functional and a double numerical basis set (DNP), which was shown to account accounts adequately for hydrogen bonds in the systems. Next, after the proton transfer from the closest flavonoid hydroxyl group to the lipase catalytic histidine, the mini-systems were again optimized. The obtained initial and final key interatomic distances within the lipase catalytic residues and between the flavonoid and the acetylated lipase are reported in Table 1. The corresponding calculated Mulliken atomic charges on the involved atoms are reported in Table 2. The respective structural modifications between the RS state and the intermediate are illustrated in Figure 2. In the following sections the DFT modelling results are analyzed successively for quercetin, isoquercitrin, and rutin.

3.1. Quercetin binding mode

For the docked quercetin-lipase complex, the initially imposed proton transfer resulted in the formation of an NH bond on histidine, as shown by the reduced His224:Nε - Fla:H distance, from 1.5 to 1.1 Å and by the simultaneous decrease of the negative charge on His224:Nε and of the positive charge on the flavonoid H. But the distance between the quercetin deprotonated oxygen and the serine-bound acetate carbon only slightly decreased from 3.6 to 3.0 Å. This suggests that no covalent bond is established between the acetate and quercetin substrates, thus the second tetrahedral intermediate is not formed. One can also notice the relatively long distances (from 2.1 and 3.4 Å) between the acetate carbonyl and the oxyanion hole residues, suggesting only a weak hydrogen bonding stabilization of acetate.

Table 1 – Relevant interatomic distances (Å) and bond lengths in the reactant states (RS) and the intermediates (Int) for the optimized mini-systems

Interatomic distance or bond length	Rutin		Isoquercitrin		Quercetin	
	RS	Int	RS	Int	RS	Int
Asp187:O _D – His224:H _{ND}	1.544	1.050	1.515	1.063	1.490	1.059
His224:H _{ND} - His224:N _D	1.093	1.580	1.091	1.506	1.116	1.624
His224:N _ε - Ser105:O _γ	3.158	3.019	3.164	2.999	3.397	3.223
His224: N _ε - Fla:H	2.207	1.027	2.392	1.039	1.563	1.095
Fla:H – Fla:O	0.981	3.030	0.980	3.029	1.058	1.520
Fla:O – Ace:C	3.017	1.501	3.409	1.482	3.616	3.020
Ace:C - Ser105:O _γ	1.333	1.509	1.337	1.549	1.346	1.351
Ace:C – Ace:O	1.239	1.300	1.242	1.314	1.232	1.230
Ace:O – Gln106:NH	2.061	2.318	1.876	1.932	1.994	1.947
Ace:O – Thr40:NH	2.132	2.237	2.208	2.369	3.216	3.175
Ace:O – Thr40:OH	1.987	1.722	1.793	1.658	2.625	2.857

Table 2 – Relevant Mulliken atomic charges in the reactant states (RS) and the intermediates (Int) for the optimized mini-systems

Atom	Rutin		Isoquercitrin		Quercetin	
	RS	Int	RS	Int	RS	Int
Asp187:O _D	-0.606	-0.486	-0.620	-0.499	-0.614	-0.482
His224:H _{ND}	+0.295	+0.341	+0.295	+0.342	+0.313	+0.343
His224:N _ε	-0.410	-0.299	-0.398	-0.273	-0.472	-0.359
Ser105:O _γ	-0.391	-0.551	-0.387	-0.530	-0.399	-0.414
Fla:H	+0.311	+0.271	+0.298	+0.286	+0.347	0.324
Fla:O	-0.511	-0.510	-0.526	-0.487	-0.510	-0.639
Ace:C	+0.564	+0.591	+0.544	+0.633	+0.531	+0.549
Ace:O	-0.534	-0.673	-0.561	-0.744	-0.494	-0.488
Gln106:NH	+0.256	+0.269	+0.257	+0.279	+0.248	+0.255
Thr40:NH	+0.224	+0.236	+0.217	+0.231	+0.179	+0.187
Thr40:OH	+0.295	+0.336	+0.369	+0.369	+0.284	+0.285

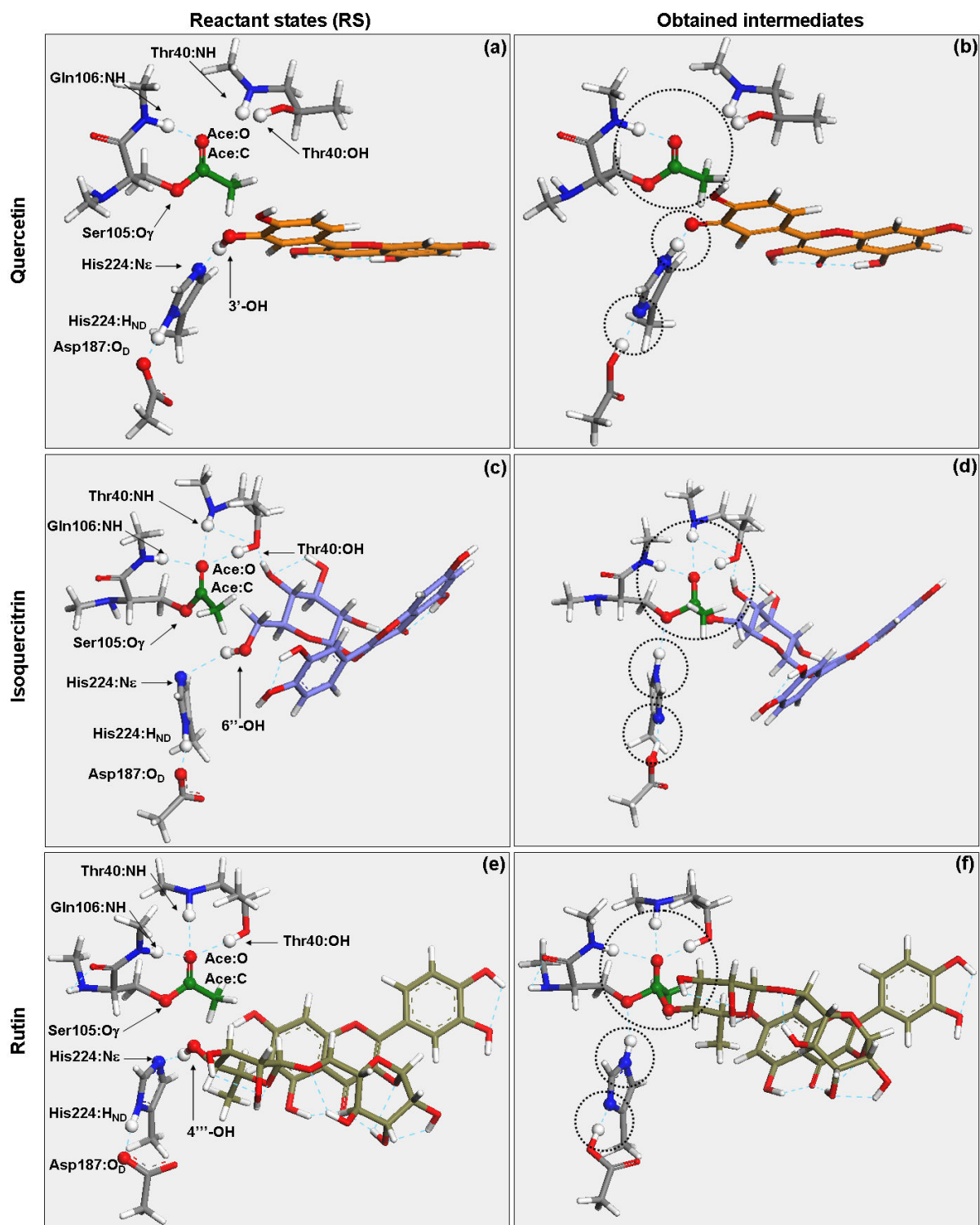


Figure 2 – Representation of the optimized mini systems used for DFT calculations. In the first column, the reactant states, which correspond to the Michaelis complexes. The atoms referred in the text are represented by spheres and identified by arrows. In the second column, the obtained intermediates. The regions of the systems that underwent modifications are bordered with dashed circles. All hydrogen atoms are coloured in white, oxygen in red and nitrogen in blue. Carbon atoms are coloured in grey for the amino-acid fragments, dark green for the serine-bound acetate, orange for the quercetin, purple for isoquercitrin and greenish for rutin.

3.2. Isoquercitrin binding mode

For the docked isoquercitrin-lipase complex, a first geometrical modification is the simultaneous increase in the interatomic Fla:H - Fla:O distance from 1.0 to 3.0 Å and the decrease in the His224:N ϵ - Fla:H distance from 2.4 to 1.0 Å, indicative of the formation of the covalent NH bond on histidine. This is corroborated by the simultaneous decrease of the negative charge on His224:N ϵ and of the positive charge on the flavonoid hydrogen following the transformation of an initial hydrogen bond to a final covalent bond. At the same time an ester bond is established between the flavonoid and the serine bound acetate as shown by the important reduction of the Fla:O - Ace:C distance from 3.0 to 1.5 Å. For the formed tetrahedral intermediate, one observes an increased distance between the Ser105:O γ and the Ace:C atoms, together with an increased negative charge on Ser105:O γ . This suggests a weakening of the existing covalent bond which is favourable for the final product release. A proton transfer from His and Asp also occurred as evidenced by the reduced Asp187:O $_D$ - His224:H $_{ND}$ distance and the simultaneous decrease of the negative charge on Asp:O $_D$ and of the positive charge on the hydrogen atom.

DFT computations also showed a strong stabilization of the tetrahedral intermediate by five hydrogen bonds, as schematized in Figure 3: two between the flavonoid oxygen and the His224:N ϵ or the Ser105:O γ , one between the protonated Asp187 and the His224, and three between the acetate carbonyl and the oxyanion hole residues. The latter interactions are strengthened for the tetrahedral intermediate due to the decreased negative charge on Ace:O and the slightly increased positive charge on the Gln106:NH and Thr40:NH atoms. The calculated energy barrier between the initial binding complex and the tetrahedral intermediate structures is +9.4 kcal.mol⁻¹.

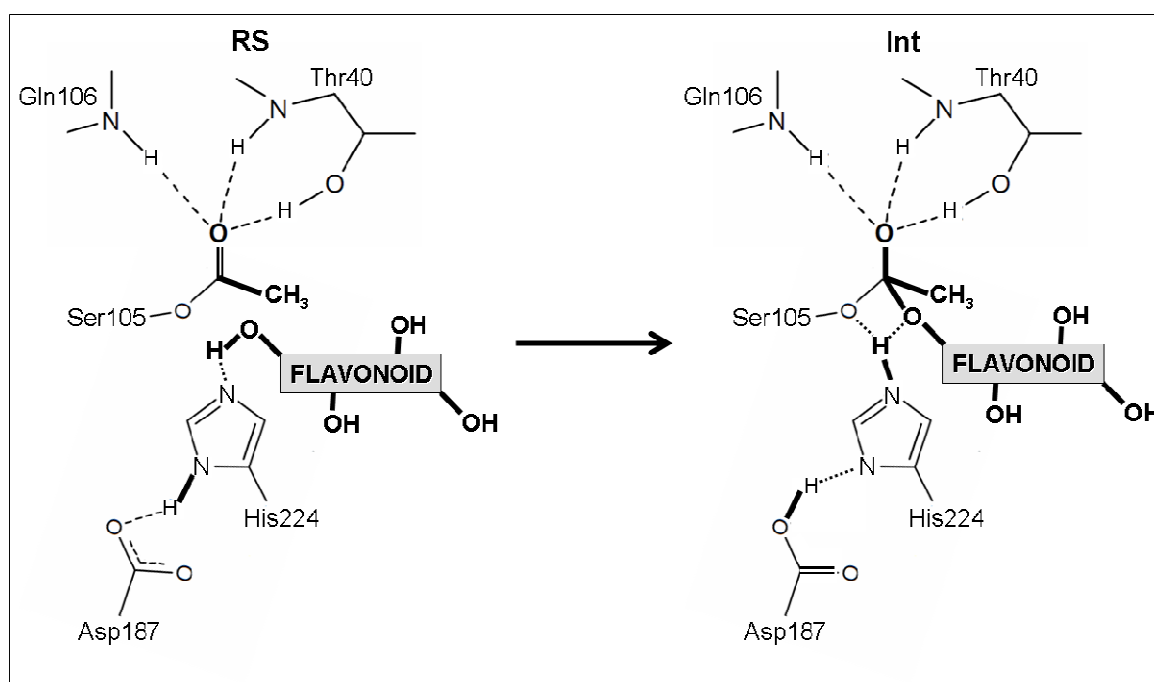


Figure 3 – Scheme showing possible hydrogen bonds in the reactant states (RS) and in intermediates (Int). This latter is formed in the cases of isoquercitrin and rutin, but not in that of quercetin.

3.3. Rutin binding mode

For the docked rutin-lipase complex, the results of DFT modelling in terms of interatomic distances and atomic charges variations are very similar to those obtained for isoquercitrin. They show that the initial proton transfer from the rutin hydroxyl group to histidine results in the formation of a NH bond. There is also a proton transfer from His224 to Asp187. At the same time an ester bond is established between the flavonoid and the serine bound acetate, as demonstrated by the reduced Fla:O - Ace:C distance from 3.0 to 1.5 Å.

The formed tetrahedral intermediate is stabilized by the same six hydrogen bonds. The intermediate is maintained by stronger electrostatic interactions in the oxyanion hole as indicated by the decreased negative charge on the Ace:O and the increased positive charge on the involved hydrogen atoms of the Gln106 and Thr40 residues. The main difference with isoquercitrin mini-system is the higher barrier between the initial binding complex and the tetrahedral intermediate, calculated as +13.8 kcal.mol⁻¹.

4. Discussion

In this study quantum quantum chemical calculations, based on the Density Functional Theory (DFT), were used to evaluate the reactivity of previously identified binding modes of quercetin, isoquercitrin and rutin in the active site of CALB. According to results, no covalent bond is expected to be formed between the aglycon quercetin and the serine bound acetate. On the contrary, with the two glycosylated flavonoids isoquercitrin and rutin, an ester bond can be established between the acetate and the closest sugar hydroxyl (the glucose 6''-OH for isoquercitrin and the rhamnose 4'''-OH for rutin). For these two substrates, the calculated energetic barriers between the formed second tetrahedral intermediate and the initial Michaelis complexes were calculated as +9.4 kcal.mol⁻¹ for isoquercitrin-6''-acetate and +13.8 kcal.mol⁻¹ for rutin-4'''-acetate. These values are similar to those previously reported from Hartree-Fock (HF) and Density Functional Theory (DFT) calculations for lipase-catalyzed transesterification reactions, using a simplified model system [16]. Moreover, the theoretical results predictions are in agreement with previously reported experimental data referring to the acetylation of these the three flavonoids acetylation with CALB as biocatalyst: when using vinyl acetate as acyl donor, no acetylated product was detected for quercetin. Under the same conditions, both the glycosylated flavonoids were found to be regioselectively acetylated (rutin on the secondary 4'''-OH of its rhamnose moiety and isoquercitrin on the primary 6''-OH of its glucose moiety).

Our DFT modelling results also contribute to the better understanding of the molecular factors controlling the regioselectivity of lipase catalyzed esterification reactions. In addition to the recognition and binding of the flavonoid in the enzyme cavity, with one of its hydroxyl group in close proximity to the catalytic triad, a successful acetylation requires the formation of tetrahedral intermediate and its stabilization through hydrogen bond networks. For the productive isoquercitrin and rutin binding modes, six hydrogen bonds involving the catalytic residues and the substrates were found to stabilize the tetrahedral intermediate. Hydrogen bond formation was less efficient for the structure obtained after the deprotonation of the quercetin (Figure 2). This crucial role of the intermediate stabilization through hydrogen bonds was previously suggested to account for the enantiomeric preference of

lipases in esterification and hydrolysis reactions [17, 18]. In these studies, quantum chemical-based calculations indicated the formation of a higher number of hydrogen bonds and molecular dynamics simulations showed that these interactions were more stable for the fast reacting substrate enantiomer. A more detailed DFT investigation of lipase-catalyzed esterification reaction also demonstrated that each hydrogen bond in the oxyanion hole contributes up to about 5 kcal.mol⁻¹ to the stabilization of the transition state, while that formed between the catalytic aspartate and histidine residues contributes up to about 6 kcal.mol⁻¹ [16].

The effects of additional stabilization of flavonoids in the lipase cavity was shown by complementary molecular dynamic simulations of the quercetin, isoquercitrin and rutin the binding modes [15]: the two glycosylated flavonoids are anchored at the entrance of the CALB cavity through hydrogen bond interactions between their flavan ring and the peptidic carbonyl of residues located in these region. The sugar moieties of these molecules display considerable structural stability and proximity with the catalytic histidine and the serine-bound acetate. This later keeps its catalytically competent orientation towards the residues Thr40 and Gln106, constituting the oxyanion hole. On the contrary, the aglycon flavonoid quercetin is more weakly stabilized in the CALB cavity, due less strong hydrophobic interactions between its aromatic structure and the aliphatic amino acids lining up the cavity walls. This results in both positional fluctuations of the flavonoid and in a shift of the serine-bound acetate outside of the oxyanion hole, as observed in molecular dynamics simulations.

To gain a deeper understanding of the influence of the amino acids residues making up the lipase cavity in the reactivity of the flavonoids hydroxyl groups, further hybrid quantum mechanics/molecular mechanics (QM/MM) modelling seems to be a more robust approach. With this methodology, the reactive region of the active site is modelled at the QM level, defining a QM region of sufficient size to encompass the atoms involved in bond break and formation and in the intermediates stabilisation all along the catalytic process. The remainder of the protein, and eventually the solvent are modelled at the MM level (classical force fields), accounting for electrostatic and van der Waals polarisation effects of the binding pocket residues on the substrate molecule, and providing the appropriate structural constraints with the core reactive region. A correctly parameterized QM/MM protocol is expected to constitute a valuable *in silico* screening platform for future rational choice or redesign of biocatalysts for flavonoid acylation.

REFERENCES:

- [1] A.M. Boudet. *Evolution and current status of research in phenolic compounds*. Phytochemistry 68 (2007) 2722-2735.
- [2] F. Mellou, D. Lazari, H. Skaltsa, A.D. Tselepis, F.N. Kolisis, H. Stamatis. *Biocatalytic preparation of acylated derivatives of flavonoid glycosides enhances their antioxidant and microbial activity*. Journal of Biotechnology 116 (2005) 295-304.
- [3] F. Mellou, H. Loutrari, H. Stamatis, C. Roussos, F.N. Kolisis. *Enzymatic esterification of flavonoids with unsaturated fatty acids : effect of the novel esters on vascular endothelial growth factor release from K562 cells*. Process Biochemistry 41 (2006) 2029-2034.
- [4] L.M. Kabeya, A.A. De Marchi, A. Kanashiro, N.P. Lopes, C.H.T.P. Da Silva, M.T. Pupo, Y.M. Lucisano-Valim. *Inhibition of horseradish peroxidase catalytic activity by new 3-phenylcoumarin*

- derivatives: synthesis and structure-activity relationships. *Bioorganic & Medicinal Chemistry* 15 (2007) 1516-1524.
- [5] Y. Shimmyo, T. Kihara, A. Akaike, T. Niidome, H. Sugimoto. *Flavonols and flavones as BACE-1 inhibitors: structure-activity relationship in cell-free, cell-based and in silico studies reveal novel pharmacophore features*. *Biochimica et Biophysica Acta* 1780 (2008) 819-825.
 - [6] L. Chebil, C. Humeau, A. Falcimaigne, J.M. Engasser, M. Ghoul. *Enzymatic acylation of flavonoids* *Process Biochemistry* 41 (2006) 2237-2251.
 - [7] T.C. Bruice. *Computational approaches: reaction trajectories, structures and atomic motions. Enzyme reactions and proficiency*. *Chemical Reviews* 106 (2006) 3119-3139.
 - [8] T. Schultz, J. Pleiss, R.D. Schmid. *Stereoselectivity of *Pseudomonas cepacia* lipase towards secondary alcohols : a quantitative model*. *Protein Science* 9 (2000) 1053-1062.
 - [9] S. Tyagi, J. Pleiss. *Biochemical profiling in silico - predicting substrate specificities for a large enzyme families*. *Journal of Biotechnology* 124 (2006) 108-116.
 - [10] M. Cammenberg, K. Hult, S. Park. *Molecular basis for the enhanced lipase-catalyzed N-acetylation of 1-phenylethanamine with methoxyacetate*. *ChemBioChem* 7 (2006) 1745-1749.
 - [11] C. Branneby, P. Carlqvist, K. Hult, T. Brinck, P. Berglund. *Aldol additions with mutant lipase : analysis by experiments and theoretical calculations*. *Journal of Molecular Catalysis B: Enzymatic* 31 (2004) 123-128.
 - [12] M. Svedendahl, P. Carlqvist, C. Branneby, O. Allnér, A. Frise, K. Hult, P. Berglund, T. Brinck. *Direct epoxidation in *Candida antarctica* lipase B studied by experiment and theory*. *ChemBioChem* 9 (2008) 2443-2451.
 - [13] M.A.J. Veld, L. Franson, A.R.A. Palmans, E.W. Meijer, K. Hult. *Lactone size dependent reactivity in *Candida antarctica* lipase B: a molecular dynamics and docking study*. *ChemBioChem* 10 (2009) 1330-1334.
 - [14] E.B. De Oliveira, C. Humeau, L. Chebil, E.R. Maia, F. Dehez, B. Maigret, M. Ghoul, J.M. Engasser. *A molecular modelling study to rationalize the regioselectivity in acylation of flavonoid glycosides catalyzed by *Candida antarctica* lipase B*. *Journal of Molecular Catalysis B: Enzymatic* 59 (2009) 96-105.
 - [15] J. Pleiss, M. Fischer, M. Peiker, C. Thiele, R.D. Schmid. *Lipase engineering database: understanding and exploiting sequence - structure - function relationships*. *Journal of Molecular Catalysis B: Enzymatic* 10 (2000) 491-508.
 - [16] C.H. Hu, T. Brinck, K. Hult. *Ab initio and density functional theory studies of the catalytic mechanism for ester hydrolysis in serine hydrolases*. *International Journal of Quantum Chemistry* 69 (1998) 89-103.
 - [17] S. Tomic, M. Ramek. *Quantum mechanical study of *Burkholderia cepacia* lipase enantioselectivity*. *Journal of Molecular Catalysis B: Enzymatic* 28 (2006) 139-147.
 - [18] C.H. Kwon, J.Y. Jeong, J.W. Kang. *Molecular modelling and experimental verification of lipase-catalyzed enantioselective esterification of racemic naproxen in supercritical carbon dioxide*. *Korean Journal of Chemical Engineering* 26 (2009) 214-219.
 - [19] J.P. Perdew, Y. Wang. *Accurate and simple analytic representation of the electron-gas σ B. Delley. An all-electron numerical method for solving the local density functional for polyatomic molecules*. *Journal of Chemical Physics* 92 (1990) 508-517.
 - [21] Y. Inada, H. Orita. *Efficiency of numerical basis sets for predicting the binding energies of hydrogen bonded complexes: evidence of small basis set superposition error compared to Gaussian basis sets*. *Journal of Computational Chemistry* 29 (2007) 225-232.
 - [22] R.S. Mulliken. *Electronic population analysis on LCAO-MO molecular wave functions I. The* *Journal of Chemical Physics* 23 (1995) 1833-1840.

3. CONTRIBUTION DE L'ARTICLE

L'article porte sur l'application de la Théorie Fonctionnelle de la Densité (DFT) pour comparer la réactivité des hydroxyles 3'-OH de la quercétine, 6''-OH de l'isoquercitrine et 4'''-OH de la rutine, en utilisant la lipase CALB comme biocatalyseur pour leur acétylation. D'après les modèles obtenus dans nos investigations préalables de docking et dynamique moléculaire, ces OH sont, pour chaque flavonoïde, ceux positionnés au plus proche des résidus catalytiques de l'enzyme. Il existe d'autres études ayant appliqué la DFT pour la description électronique des flavonoïdes et la réactivité de substrats divers en présence de lipases. Cependant, l'utilisation de cette méthode quantique dans l'étude de l'acétylation de flavonoïdes catalysée par la CALB est décrite pour la première fois, à notre connaissance.

Les résultats préliminaires ont indiqué que les hydroxyles alcooliques 6''-OH de l'isoquercitrine et 4'''-OH de la rutine forment la liaison ester avec l'acétate, après le transfert de leur proton à l'His224, avec une variation d'énergie admissible pour ce type de réaction (environ +10 kcal.mol⁻¹). Simultanément, la liaison entre l'oxygène de la chaîne latérale de la sérine et le carbone carbonyle de l'acétate (Ser105:O_γ-Ace:C) s'affaiblit, en subissant un étirement et une augmentation de la charge négative sur l'atome Ser105:O_γ. Ainsi, dans ces deux cas, le deuxième intermédiaire tétraédrique est obtenu, suggérant que cette étape de la réaction est faisable. Cet intermédiaire n'est pas obtenu lorsque le proton d'hydroxyle phénolique 3'-OH de la quercétine est transféré vers l'His224. Ceci serait dû à des effets de délocalisation des électrons de l'oxygène phénolique au niveau des cycles aromatiques de la molécule, ce qui n'est pas possible avec les cycles saturés des sucres. Cet atome d'oxygène ne serait donc pas assez nucléophile pour attaquer l'acétate et former l'intermédiaire, indispensable pour que la réaction soit menée à terme. De manière plus générale, ces résultats illustrent l'énorme complexité du processus catalytique : outre des facteurs structuraux liés à la reconnaissance des substrats par le site actif de la lipase (interactions attractives et répulsives, empêchements stériques), la productivité des complexes entre les flavonoïdes et l'acétyl-lipase est aussi influencée par la réactivité chimique des flavonoïdes.

Afin d'avoir une idée plus précise de l'influence des substrats sur la faisabilité de l'acétylation des flavonoïdes, de nombreuses simulations quantiques complémentaires sont encore nécessaires. En particulier, l'étude de la formation du premier intermédiaire tétraédrique (à partir de l'enzyme libre et du vinyle-acétate), et la détermination des états de transition permettraient d'évaluer l'ensemble du profil énergétique de la réaction. Ainsi, l'étape limitante pourrait être identifiée.

Conclusions & perspectives

Au cours de ce travail, nous avons l'objectif d'apporter une meilleure compréhension des aspects structuraux gouvernant les interactions entre flavonoïdes et lipases, afin d'expliquer la régiosélectivité de l'acétylation des ces polyphénols catalysée par ces enzymes. Nous nous sommes focalisés sur quatre cas spécifiques : l'acétylation des flavonoïdes glycosides isoquercitrine et rutine, catalysée par la lipase B de *Candida antarctica* (CALB), et l'acétylation de l'analogue aglycone quercétine, en utilisant cette même lipase et celle de *Pseudomonas cepacia*. Des simulations de docking des flavonoïdes dans les cavités catalytiques des lipases préalablement acétylées ont été exécutées. Des simulations complémentaires de dynamique moléculaire et de chimie quantique ont ensuite été réalisées sur les complexes enzyme-substrats ainsi obtenus. En se basant sur le mécanisme catalytique admis pour les lipases dans le cas des réactions de transesterification, les résultats des simulations ont été analysés afin de trouver d'éventuelles corrélations entre les positions, orientations, interactions, stabilité et réactivité des substrats et les résultats expérimentaux de biosynthèse disponibles. Pour synthétiser et discuter les principaux résultats obtenus, nous allons tenter d'apporter des réponses aux trois questions majeures soulevées au début de cette étude.

(1) Comment les modes d'interaction enzyme-substrats influencent-ils la régiosélectivité observée lors de l'acétylation des flavonoïdes glycosylés isoquercitrine et rutine, catalysée par la lipase B de Candida antarctica (CALB) ?

Les résultats de cette étude ont maintenu l'hypothèse selon laquelle les positions et les orientations des deux substrats (flavonoïde et acétate) dans la cavité catalytique de la CALB sont corrélées à la régiosélectivité observée expérimentalement lors de la réaction d'acétylation : leurs parties aglycones se stabilisent ancrées à l'entrée de la cavité, tandis que leurs parties sucres s'orientent vers les résidus catalytiques, au fond de la poche.

Les complexes enzyme-substrats, obtenus par docking des flavonoïdes dans le site actif de la CALB acétylée, ont été analysés en se basant sur le mécanisme réactionnel admis pour les lipases, en prenant en compte les critères structuraux établis dans d'autres travaux similaires. Ainsi, pour que ces complexes soient considérés comme productifs (susceptibles de mener à la formation d'un produit réactionnel), deux conditions doivent être respectées : d'abord, le carbonyle ($-C=O$) de l'acétate doit être correctement orienté vers les résidus du trou oxyanionique de la lipase (Thr40 et Gln106) pour favoriser l'attaque nucléophile du flavonoïde sur l'atome de carbone carbonyle. Simultanément, un hydroxyle ($-OH$) du flavonoïde doit être positionné dans la région entre les résidus catalytiques His224 et Ser105. Dans cette position, le proton de l'hydroxyle peut être capturé par l'histidine et l'oxygène ainsi déprotoné peut exécuter l'attaque nucléophile sur l'acétate de l'acyle enzyme.

- *Pour l'isoquercitrine, deux orientations possibles ont été trouvées à partir du docking* : dans la première, la partie glycosyidique est placée à l'entrée de la cavité et le cycle A de la partie aglycone est pointé vers le fond de la poche. Dans ce cas, l'acétate n'est pas correctement orienté vers le trou oxyanionique. Les hydroxyles 5-OH et 7-OH du cycle A, qui pointent vers le fond de la poche, ne sont accessibles ni à la sérine ni à l'histidine catalytiques. Cette orientation ne peut donc être considérée productive. Dans la deuxième orientation, la partie aglycone est ancrée à l'entrée de la cavité, par le biais de liaisons hydrogène avec les résidus Val149 et Ile189, tandis que la partie glycosyidique est pointée vers les résidus catalytiques. Dans ce cas, l'acétate présente une orientation correcte vers le trou oxyanionique et seul l'hydroxyle 6''-OH du flavonoïde est positionné à proximité des résidus Ser105 et His224. En conséquence, cette orientation peut être considérée productive, suggérant l'acétylation de ce groupement OH. En effet, expérimentalement, il a été vérifié que l'hydroxyle 6''-OH est le seul groupement de l'isoquercitrine directement acétylé avec la CALB comme biocatalyseur. Les modèles sont donc en accord avec l'expérience.
- *Pour la rutine, une seule orientation possible a été trouvée à partir du docking* : la partie aglycone est ancrée à l'entrée de la cavité de l'enzyme, en formant des liaisons hydrogène avec les résidus Val149 et Leu278. La partie glycosyidique est pointée vers le fond de la cavité, où se situent les résidus catalytiques. Lorsque l'acétate est correctement orienté vers les résidus Thr40 et Gln106, la partie glycosyidique de la rutine adopte une conformation pliée lui permettant de s'adapter à la cavité de la CALB. Le seul hydroxyle du flavonoïde positionné entre la sérine et l'histidine catalytiques est le 4'''-OH. Cette configuration suggère que l'acétylation du groupement 4'''-OH est favorisée. À nouveau, ce modèle corrobore les résultats expérimentaux, qui avaient montré que seul l'hydroxyle 4'''-OH de la rutine est directement acétylé en utilisant la CALB comme biocatalyseur.

Il est important de souligner que ces complexes identifiés comme productifs ont conservé leur configuration globale au cours d'une trajectoire de dynamique moléculaire de 2 ns. Cette stabilité renforce la validité des corrélations qualitatives entre la régiosélectivité de la réaction et les caractéristiques structurales mises en évidence par le docking (orientation de l'acétate et position des hydroxyles du flavonoïde).

(2) Comment les modes d'interaction enzyme-substrats peuvent-ils expliquer le fait que la lipase de *Pseudomonas cepacia* (PCL) catalyse l'acétylation du flavonoïde quercétine tandis que la CALB ne catalyse pas cette réaction ?

Les modèles obtenus dans cette étude indiquent que les différences structurales observées entre la PCL et la CALB au niveau des acides aminés tapissant leurs cavités catalytiques, donnent lieu à des interactions intermoléculaires enzyme-quercétine différentes selon la lipase employée. Le flavonoïde est mieux stabilisé par les interactions formées avec les résidus de la cavité de la PCL que dans le cas de la CALB. Cette différence de stabilisation serait à l'origine de la différence de spécificité de ces deux lipases vis-à-vis de ce flavonoïde.

Lors de l'étude précédente, l'application de simulations de dynamique moléculaire sur les complexes obtenus par la procédure de docking/scoring a permis d'évaluer la stabilité structurale de ces complexes sur une période de temps. Dès lors, plutôt que d'analyser les complexes « statiques » directement issus du docking, nous avons poursuivi l'étude en combinant des simulations de docking et de dynamique moléculaire. Effectivement, les substrats quercétine et acétate adoptent initialement des configurations similaires dans les cavités des deux lipases (cycle A ou cycle B vers le cœur de l'enzyme). Des changements structuraux au sein des complexes n'ont pu être mis en évidence qu'au cours des trajectoires de dynamique :

- *Pour la CALB*, les configurations des substrats subissent des modifications assez prononcées au cours des trajectoires de dynamique moléculaire, notamment dans le cas de l'orientation B (cycle B positionné vers le fond de la poche). Ceci serait dû à la nature des résidus tapissant la cavité catalytique de cette lipase : la plupart d'entre eux sont des acides aminés aliphatiques hydrophobes, incapables de former des interactions attractives suffisamment fortes pour stabiliser le flavonoïde dans le complexe. Le groupement carbonyle de l'acétate présente une tendance à se tourner vers le flavonoïde en établissant des liaisons H avec lui, au lieu de garder une orientation productive, vers les résidus du trou oxyanionique (Thr40 et Gln106). Pour l'orientation A (cycle A positionné vers le fond de la poche) de la quercétine, cette rotation de l'acétate a été observée dès le départ. Pour l'orientation B, elle s'est faite progressivement au cours de la trajectoire. Cette perte de la configuration productive de l'acétyl-CALB peut expliquer l'absence de réaction lorsqu'on essaye d'acétyler la quercétine en présence d'acétate de vinyle, en utilisant cette enzyme comme biocatalyseur.
- *Pour la PCL*, contrairement à la CALB, les trajectoires sont stables en ce qui concerne la configuration des substrats, que le flavonoïde soit dans l'orientation A ou B. En effet, contrairement à la CALB, la cavité catalytique de la PCL contient des résidus aromatiques, qui forment des interactions plus fortes avec les noyaux aromatiques du flavonoïde. Ceci expliquerait la meilleure stabilisation de la quercétine dans la cavité de la PCL par rapport à la CALB. En particulier, dans les orientations A ou B, la quercétine se positionne à proximité des résidus Phe146 et Tyr23. Les chaînes latérales aromatiques de ces résidus tendent à s'approcher de la quercétine au cours de la trajectoire. Pour l'orientation A, le carbonyle de l'acétate s'écarte du trou oxyanionique (Leu17 et Gln88) au début de la trajectoire, en formant une liaison H stable avec l'hydroxyle 7-OH. Il semblerait en effet que cette région de la molécule de quercétine exerce une attraction électrostatique importante sur l'acétate (cette rotation de l'acétate en début de trajectoire a été aussi observée pour cette orientation dans le cas de la CALB). Pour l'orientation B de la quercétine dans le site actif de la PCL, l'acétate garde son orientation catalytiquement productive vers le trou oxyanionique tout au long de la trajectoire. L'hydroxyle 4'-OH reste à proximité des résidus catalytiques His286 et Ser87, suggérant que l'acétylation de ce groupement est possible. Ces modèles corroborent les résultats expérimentaux de biosynthèse : avec l'acétate de vinyle comme donneur d'acyle et la PCL comme catalyseur, le quercétine-4'-acétate a été le seul monoester produit au cours de la réaction.

D'après cette analyse, la présence de résidus aromatiques entourant la cavité de l'enzyme semble favoriser la réaction et être à l'origine de la spécificité enzymatique. Dans la cavité catalytique de CALB, constituée essentiellement de résidus hydrophobes, la quercétine subit des oscillations plus marquées. Ceci favoriserait la rotation de l'acétate, empêchant celui-ci de maintenir sa configuration initiale, productive, vers le trou oxyanionique. De manière analogue, dans la cavité de la PCL, caractérisée par la présence de résidus aromatiques, la quercétine ne subit que de faibles oscillations, présentant une meilleure stabilité et permettant le maintien de l'acétate correctement orienté vis-à-vis du trou oxyanionique. L'hydroxyle 4'-OH du flavonoïde, positionné entre l'histidine et la sérine catalytiques, pourrait ainsi être acylé.

(3) Dans le cas de la CALB, le type de groupement OH de ces trois flavonoïdes (quercétine, isoquercitrine et rutine) qui atteignent les résidus catalytiques peut-il aussi contribuer à la sélectivité ?

Les résultats des calculs quantiques préliminaires, basés sur la Théorie Fonctionnelle de la Densité (DFT), ont suggéré que la propension du flavonoïde à former le deuxième intermédiaire tétraédrique avec la CALB est liée à la nature du groupement hydroxyle qui atteint les résidus catalytiques His224 et Ser105.

En effet, concernant l'hydroxyle le plus proche des résidus catalytiques de l'enzyme pour chacun des trois flavonoïdes, lorsque le proton est transféré à l'atome His224 :N le comportement des systèmes n'est pas le même selon la nature du groupement hydroxyle en question.

- Pour les flavonoïdes glycosylés isoquercitrine et rutine, dont les hydroxyles 6''-OH et 4'''-OH (fonctions alcool) sont positionnés à proximité des résidus catalytiques, une liaison ester stable s'établit entre l'oxygène déprotoné du flavonoïde et le carbone carbonyle de l'acétate de l'acyl enzyme.
- Pour le flavonoïde aglycone quercétine, dont l'hydroxyle 3'-OH (fonction phénol) est proche des résidus catalytiques, cette liaison ester ne se forme pas. Une explication pourrait être la délocalisation des électrons libres de l'oxygène sur le noyau aromatique adjacent, diminuant ainsi son caractère nucléophile. Ce phénomène n'est pas possible avec les cycles saturés des sucres de l'isoquercitrine et de la rutine.

Cette étude a permis de mettre en évidence que, outre la structure de la lipase, la structure du flavonoïde influence aussi la faisabilité de la réaction d'acétylation enzymatique. Ces résultats viennent donc compléter les deux études précédentes et constituent le début d'une étude plus vaste visant à étudier la réactivité chimique des complexes enzyme-substrats impliquant flavonoïdes et lipases. Nonobstant ces premiers résultats, d'autres simulations s'avèrent encore nécessaires pour comprendre plus précisément comment la structure des flavonoïdes affecte la réactivité chimique de leurs hydroxyles. L'étude des propriétés électroniques des trois flavonoïdes pourrait apporter des informations très utiles. Par exemple, la détermination de la forme et de la localisation des orbitales HOMO et LUMO, associée à l'analyse des charges atomiques pour les trois

flavonoïdes (dans leur forme naturelle et déprotonée) pourrait fournir des indications sur la façon dont les électrons se distribuent au sein de la molécule, avant et après arrachage du proton.

Dans l'ensemble, les résultats obtenus au cours de ce travail constituent une première contribution pour améliorer la compréhension, à l'échelle moléculaire, des événements qui influencent la sélectivité de l'acétylation enzymatique des flavonoïdes étudiés. Ces événements dépendent des structures à la fois de l'enzyme et du flavonoïde. En effet, celles-ci sont déterminantes pour parvenir à la conjugaison de trois critères, directement corrélés au résultat de la réaction : la façon dont le flavonoïde se positionne et s'oriente dans la cavité catalytique de l'acétyle-enzyme, la formation ou non d'interactions intermoléculaires favorables à la stabilisation du complexe ES et finalement, la réactivité des groupements hydroxyles des flavonoïdes atteignant les résidus catalytiques de l'enzyme. Davantage de vérifications expérimentales et d'autres simulations complémentaires pourraient rendre les modèles encore plus robustes et contribuer ainsi, par la suite, au développement d'une approche prédictive *in silico* concernant la sélectivité de l'acylation enzymatique de flavonoïdes.

En particulier, quelques perspectives majeures apparaissant à l'issue de ce travail concernant :

Les profils énergétiques complets des réactions d'acétylation

La dernière partie de notre étude, dans laquelle l'approche DFT a été appliquée pour une évaluation préliminaire de la réactivité des hydroxyles des flavonoïdes. Pour cela, nous nous étions appuyés, dans un premier temps, sur l'hypothèse que pour un même donneur d'acyle et une même enzyme seule la réactivité du flavonoïde devrait définir la réactivité du complexe. Nous avons donc comparé seulement deux états stationnaires de la deuxième étape des réactions : le deuxième complexe de Michaelis (flavonoïde + acétyle-CALB) et le deuxième intermédiaire tétraédrique. Cet exercice nous a permis de vérifier que les hydroxyles des sucres (fonctions alcools) sont plus réactifs que ceux de la partie aglycone (fonctions phénol) quand il s'agit de former une liaison ester et ouvre deux voies distinctes - mais complémentaires - d'investigation :

- La première concerne une compréhension plus fondamentale de la réaction est encore à poursuivre. Comme mentionné précédemment, la détermination de la forme et de la localisation des orbitales HOMO et LUMO et l'analyse des charges atomiques des formes naturelles et déprotonés des trois flavonoïdes pourrait fournir des indications sur la manière dont les électrons se redistribuent au sein de la molécule après l'arrachage du proton.
- La deuxième serait d'entreprendre une démarche similaire pour la première étape de la réaction : la formation du premier intermédiaire tétraédrique à partir du premier complexe de Michaelis (acétate de vinyle + CALB libre). Par la suite, la détermination des états de transition permettrait d'obtenir les profils énergétiques complets pour l'acétylation des trois flavonoïdes et ainsi d'identifier, pour chacun d'entre eux, l'étape limitante.

Des vérifications expérimentales additionnelles : la mutagenèse dirigée

En supposant que des mutations ponctuelles au niveau du site actif n'engendreraient pas de réarrangement important de l'architecture globale des deux lipases CALB et PCL, deux vérifications intéressantes pourraient être faites à l'aide de la mutagenèse dirigée.

- Si les résidus aromatiques Tyr23 et Phe146 jouent un rôle décisif dans la stabilisation de ce flavonoïde comme suggéré par les résultats, leur substitution par un résidu aliphatique hydrophobe devrait réduire, voire annuler l'activité d'acylation de la PCL vis-à-vis de la quercétine. L'acide aminé valine semble être un choix raisonnable, car le volume de sa chaîne latérale est proche du volume de celles des résidus enlevés. Les mutants simples Y23V et F146V et le mutant double correspondant (Y23V,F146V) devraient permettre une telle vérification.
- Une démarche analogue pourrait être entreprise pour la CALB : par exemple, si les résidus Ile189 et Ala282 forment des interactions répulsives déstabilisantes avec la quercétine, les changer pour un résidu aromatique comme la phénylalanine, devrait permettre d'améliorer la stabilisation de la quercétine dans la cavité de cette lipase. Ainsi, les mutants simples I189F et A282F et leur analogue double I189F,A282F pourraient constituer de bons candidats pour permettre d'acyler la quercétine en utilisant la CALB.

L'acétylation séquentielle des mono-esters

Les résultats expérimentaux relatifs aux cinétiques réactionnelles ont montré que le mono-ester résultant de l'acétylation régiosélective du flavonoïde peut subir, au bout d'un certain temps réactionnel et sous certaines conditions, une deuxième acétylation, elle aussi régiosélective, et former un diester. Dans la présente étude, nous nous sommes focalisés sur la première étape de la réaction d'acétylation des flavonoïdes, la formation du mono-ester. Une perspective intéressante serait d'étudier la régiosélectivité de la deuxième acétylation. La présence d'un groupement acétate sur la molécule devrait engendrer des différences de positionnement et d'orientation du substrat dans la cavité enzymatique par rapport au flavonoïde d'origine. Pour vérifier cette hypothèse, les mono-esters devraient alors être pris comme ligands pour les simulations de docking dans la cavité de l'acétyl-lipase, puisque ce sont les substrats pour la deuxième acylation.

Acylation des flavonoïdes par d'autres donneurs d'acyle

Comme expliqué dans le début de ce document, le choix de l'acétylation comme réaction modèle a été motivé par l'existence de données expérimentales concernant cette réaction pour les trois flavonoïdes étudiés. Expérimentalement, il a été montré qu'avec d'autres donneurs d'acyle à chaîne plus longue, la vitesse initiale et le rendement de la réaction changent, mais pas sa régiosélectivité. Une autre piste intéressante serait la construction de modèles impliquant d'autres donneurs d'acyle, en particulier des acides gras à chaîne plus longue.

Dans de tels cas, le degré de liberté spatiale des substrats donneurs d'acyle serait plus élevé que dans le cas de l'acétate de vinyle étudié jusqu'ici. Ceci impliquerait une préparation très minutieuse des structures des acyle-enzymes, en appliquant des dynamiques moléculaires plus longues que celles réalisées dans cette étude. Ceci devrait permettre l'obtention de modèles réalistes d'acyle-enzymes, dans lesquels la chaîne acyle, liée à la sérine catalytique, serait correctement disposée et stabilisée dans la cavité catalytique. Ces structures seraient alors utilisées comme cibles de docking pour les flavonoïdes

Modélisation en présence d'autres solvants

Malgré le fait que l'acylation de flavonoïdes soit un bioprocédé mené en milieux non aqueux, l'eau a été utilisée comme solvant explicite lors des simulations. Bien qu'à première vue ce choix semble contradictoire, il se justifie par la connaissance préalable de deux particularités des systèmes étudiés, discutées dans la section 2.2.2 du chapitre II (Matériels et Méthodes de Simulation) : d'une part la présence d'une couche de molécules d'eau entourant les enzymes, vérifiée expérimentalement ; d'autre part la faible flexibilité structurale des -hydrolases en général, et des lipases CALB et PCL en particulier, démontrée récemment par des simulations de dynamique moléculaire. Effectivement, les résultats de simulation obtenus dans ces conditions présentent une bonne corrélation avec les résultats expérimentaux de régiosélectivité. Ceci suggère que les structures des enzymes générées au cours de nos simulations étaient représentatives de celles existant dans les milieux réactionnels. Néanmoins, le développement de paramètres spécifiques pour les solvants utilisés pour l'acylation des flavonoïdes (*tert*-amyl-alcool, acétone et acétonitrile) pourrait, par la suite, permettre la réalisation de simulations sur des systèmes moléculaires encore plus précis.

L'apport possible des simulations hybrides (QM/MM)

Dans ce travail, une procédure rigoureuse de docking/scoring des flavonoïdes dans la cavité des lipases acétylées a été mise en œuvre pour obtenir des complexes ES. La démarche méthodologique pour étudier ces complexes par la suite s'est scindée en deux parties :

- L'analyse des interactions intermoléculaires et de la stabilité structurale des complexes a été faite par une approche de dynamique moléculaire. Ces simulations permettent une appréciation de la stabilité du complexe, sans pour autant donner d'élément sur la réactivité du flavonoïde. En effet, un ligand non réactif bien stabilisé dans la cavité catalytique d'une enzyme pourrait être aussi un bon inhibiteur compétitif pour cette enzyme.
- L'analyse de la propension des groupements hydroxyle à former une liaison ester a été faite par une approche de chimie quantique. Cette approche fournit une indication préalable de la réactivité intrinsèque du substrat, sans pour autant considérer d'éventuels phénomènes de polarisation de celui-ci dû aux acides aminés de la poche. Dûs aux restrictions instrumentales, ces acides aminés sont exclus des mini-systèmes utilisés pour les calculs quantiques.

Une idée intéressante serait l'application de simulations hybrides du type QM/MM pour étudier ces systèmes. Pour rappel, dans cette approche les atomes directement impliqués dans la rupture et la formation de liaisons chimiques sont modélisés au niveau électronique (QM), tandis que les autres parties du système sont traitées par un champ de forces (MM). Ceci permettrait par exemple d'évaluer l'effet des acides aminés de la poche catalytique sur la répartition des électrons du substrat et par suite, sur la réactivité d'un groupement hydroxyle de celui-ci positionné à proximité des résidus catalytiques. Bien entendu, l'utilisation d'une méthode adéquate pour le traitement de l'interface entre les régions QM et MM est indispensable et les parties à traiter au niveau QM doivent être judicieusement choisies. Il est néanmoins important de souligner que la mise au point d'un protocole QM/MM demanderait des investissements conséquents, aussi bien d'un point de vue intellectuel que financier. En conséquence, il est impératif de vérifier si les objectifs précis des études futures justifient l'application de cette méthodologie.

Annexes

A1. LES FONCTIONS DE SCORE UTILISEES POUR LE CONSENSUS SCORE

LigScore1 et LigScore2 [1] : LigScore1 et LigScore2 sont des fonctions « knowledge-based », qui estiment le pK_i du ligand selon une expression du type :

$$pK_i = A - B(vdW) + C(C_{+pol}) - D(Totpol^2) \quad (\text{Eq. A1})$$

Dans cette expression, A , B , C et D sont des constantes dont les valeurs ont été calibrées via des techniques d'extrapolation, en se basant sur l'affinité expérimentale d'un ensemble de 630 complexes protéine-ligand échantillonnés de la PDB ; vdW est l'interaction de van der Waals (potentiel du type Lennard-Jones 9-6) et $Totpol^2 = (C_{+pol})^2 + (C_{-pol})^2$, où C_{+pol} et C_{-pol} sont les surfaces polaires du ligand impliquées dans des interactions attractives et répulsives respectivement. La différence entre les deux variantes de cette fonction de score est que dans LigScore1 le terme $(Totpol^2)$ prend en compte la surface totale du ligand (« total surface »), tandis que dans la deuxième, ce terme est remplacé par $(BuryPol^2)$, se référant seulement à la surface de la partie enfouie dans le site actif (« buried surface »). En réalité, plus la valeur de pK_i est faible, plus d'affinité le ligand présente vis-à-vis du récepteur. Ainsi, le logiciel donne les valeurs de $-pK_i$, pour que l'on puisse associer « valeur plus élevée de score » avec « meilleure affinité ».

PMF [2, 3] : la fonction de score PMF (*Potential of Mean Force*), aussi une fonction du type « knowledge based », se base sur l'analyse de 697 complexes protéine-ligand de la PDB. Pour plusieurs types de paires d'atomes dans ces complexes, une série de potentiels d'interaction dépendants de la distance a été dérivée. Ainsi, les effets enthalpiques et entropiques gouvernant la formation des complexes sont assumés comme inclus implicitement dans une expression du type :

$$PMF = \sum \sum A_{ij}(d_{ij}) \quad (\text{Eq. A2})$$

où d_{ij} est la distance entre les atomes i du ligand et j de la protéine et A_{ij} est une constante dont la valeur dépend de chaque type de paire d'atomes i et j . La distance cut-off adoptée, au delà de laquelle les atomes i et j sont considérés comme non interagissant, est de 9 Å pour les interactions carbone-carbone et de 6 Å pour les toutes les autres interactions.

PLP1 et PLP2 [4, 5] : Les fonctions PLP (*Pairwise Linear Potential*) sont des fonctions de score empiriques, dans laquelle chaque atome ($\neq$ hydrogène) du ligand et de la poche de la protéine reçoivent une des quatre classifications : donneur de liaison hydrogène ou accepteur de liaison hydrogène ou donneur et accepteur de liaison hydrogène à la fois ou non polaire. Ensuite, chaque paire d'atomes qui interagissent sont reçoit une des trois classifications : liaison hydrogène entre un donneur et un accepteur ou paire repulsive entre deux donneur ou deux accepteurs ou autres contacts. Ainsi, l'énergie totale d'interaction peut être conceptuellement exprimée selon l'expression suivante :

$$E_{interaction} = E_{liaison H} + E_{répulsion} + E_{contacts} \quad (\text{Eq. A3})$$

Les termes des liaisons hydrogène et de répulsion sont dérivés de mesures expérimentales et modulés par un facteur prenant en compte les distances et angles d'interaction.

Dans la version PLP2, les termes sont les mêmes que dans PLP1. Néanmoins, pour la mesure des distances et angles d'interaction, les atomes sont assignés comme petits (métaux et F ; rayon = 1,4 Å), moyens (C, O, N ; rayon = 1,8 Å) et grands (S, P, Cl, Br ; rayon = 2,2 Å). Ces valeurs se basent sur des valeurs moyennes obtenues dans des structures cristallines de haute résolution. La prise en compte des rayons atomiques permet estimation plus fine des valeurs de distance et d'angle et, par conséquent, une meilleure précision dans l'estimation de l'énergie d'interaction.

Jain [6] : Cette fonction de score empirique porte le nom de son créateur et se base sur des données mesurées expérimentalement pour un ensemble de complexes protéine-ligand. La valeur du score Jain est donnée par la somme de cinq termes qui sont des approximations pour les interactions lipophiles, les interactions polaires attractives, les interactions polaires répulsives, la solvation de la protéine et du ligand et l'entropie du ligand. Ou, de manière plus simplifiée :

$$E_{interaction} = E_{lipophile} + E_{pol. attractive} + E_{pol. repulsive} + E_{solvation} + S_{ligand} \quad (Eq. A4)$$

Les trois premiers termes sont estimés par une fonction gaussienne pondérée par la distance interatomique. Seuls les atomes du ligand directement en contact avec un atome de la protéine sont considérés dans le calcul. Ceci peut engendrer une surévaluation des contacts de surface, représentant un inconvénient dans les cas où ces contacts ne sont pas le principal facteur gouvernant la formation du complexe.

Consensus scoring : Lorsque plusieurs fonctions de score sont disponibles, une question que l'on peut se poser est laquelle est la plus adaptée pour le type de ligand et le type de protéine étudiés. Souvent, la réponse n'est pas évidente. Dans ce cas, il est judicieux d'utiliser le *Consensus Score*, une technique de classement de complexes qui combine les résultats individuels de plusieurs fonctions de score [7, 8]. Le principe consiste à indiquer les complexes qui apparaissent en tête de classement selon chaque fonction de score individuellement, en déterminant ainsi la fréquence à laquelle chaque complexe est apparaît comme bien classé. Statistiquement, ceci augmente la probabilité de trouver le mode correct d'association protéine-ligand. De fait, des travaux récents ont montré que cette procédure augmente les chances de retrouver par simulation informatique un complexe déterminé expérimentalement par des techniques de diffraction de rayons X [9-12].

A2. L'ALGORITHME INTEGRATEUR DE VERLET-LEAPFROG

L'algorithme de Verlet-Leapfrog a été utilisé dans les simulations de dynamique moléculaire, pour l'intégration de l'équation de mouvement de Newton.

Dans tous les algorithmes d'intégration numérique utilisés en dynamique moléculaire, les positions atomiques (r), les vitesses et les accélérations sont approximées par des séries de Taylor. Considérant un pas d'intégration Δt :

$$r(t + \Delta t) = r(t) + v(t) \cdot \Delta t + 0,5 \cdot a(t) \cdot \Delta t^2 + \dots \quad (\text{Eq. A5})$$

$$v(t + \Delta t) = v(t) + a(t) \cdot \Delta t + 0,5 \cdot b(t) \cdot \Delta t^2 + \dots \quad (\text{Eq. A6})$$

$$a(t + \Delta t) = a(t) + b(t) \cdot \Delta t + \dots \quad (\text{Eq. A7})$$

L'algorithme de Verlet est construit en considérant les positions atomiques à $t + \Delta t$ et $t - \Delta t$, avec les trois premiers termes de la série de Taylor correspondante :

$$r(t + \Delta t) = r(t) + v(t) \cdot \Delta t + 0,5 \cdot a(t) \cdot \Delta t^2 \quad (\text{Eq. A8})$$

$$r(t - \Delta t) = r(t) - v(t) \cdot \Delta t + 0,5 \cdot a(t) \cdot \Delta t^2 \quad (\text{Eq. A9})$$

En sommant ces deux équations, on obtient :

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t) \cdot \Delta t^2 \quad (\text{Eq. A10})$$

Cet algorithme utilise les accélérations au temps t et les positions aux temps t et $t - \Delta t$ pour calculer les nouvelles positions au temps $t + \Delta t$, sans l'utilisation explicite des vitesses atomiques. En conséquence, les calculs sont plus rapides et les besoins en termes d'espace de stockage de données plus modestes. Cependant, cette même caractéristique entraîne des limitations en précision, notamment lorsque les trajectoires à calculer sont longues. Afin de contourner cet inconvénient, Il existe une version améliorée de cet algorithme, appelée **algorithme de Verlet-Leapfrog**. Dans ce cas, les vitesses sont d'abord calculées au temps $t + 0,5 \cdot \Delta t$.

$$v(t + 0,5 \cdot \Delta t) = v(t - 0,5 \cdot \Delta t) + a(t) \cdot \Delta t \quad (\text{Eq. A11})$$

Les nouvelles vitesses sont utilisées pour calculer les positions au temps $t + \Delta t$

$$r(t + \Delta t) = r(t) + v(t + 0,5 \cdot \Delta t) \cdot \Delta t \quad (\text{Eq. A12})$$

Ainsi, on peut dire que les vitesses « sautent » sur les positions, puis les positions « sautent » sur les vitesses et ainsi de suite (d'où le nom de l'algorithme, « saut de grenouille »). On note que les vitesses sont explicitement calculées, ce qui améliore la précision des calculs, même si celles-ci ne sont pas calculées en même temps que les positions.

A3. LA FONCTIONNELLE D'ÉCHANGE-CORRELATION PW91 ET LES FONCTIONS DE BASE NUMÉRIQUES DND

Pour les calculs quantiques, le potentiel d'échange-corrélation PW91 a été approximé par la fonctionnelle de Perdew-Wang, PW91 [14]. Les orbitales atomiques ont été représentées par des fonctions de base numériques incluant des fonctions de polarisation, (Double Numerical with Polarization) [15].

La fonctionnelle PW91 est fonctionnelle de deuxième génération (GGA), qui tient compte de la non homogénéité de la densité électronique des systèmes. Les énergies d'échange (E_x) et de corrélation (E_c) pour n électrons sont calculées selon les équations A13 et A14 :

$$E_x^{\text{PW91}}[n] = - \int d\mathbf{r} n \frac{3k_F}{4\pi} \frac{1 + 0.1965s \sinh^{-1}(7.796s) + (0.274 - 0.151e^{-100s^2})s^2}{1 + 0.1964s \sinh^{-1}(7.796s) + 0.004s^4} \quad (\text{Eq. A13})$$

$$E_c^{\text{PW91}}[n] = \int d\mathbf{r} n (\epsilon_c(\mathbf{r}_s, \zeta) + H(t, r_s, \zeta)) \quad (\text{Eq. A14})$$

Avec :

- $k_F = (3\pi^2 n)^{1/3}$
- $s = |\nabla n| / 2nk_F$
- $t = |\nabla n| / 2ngk_s$
- $g = [(1 + \zeta)^{2/3} + (1 - \zeta)^{2/3}] / 2$
- $k_s = (4k_F / \pi)^{1/2}$

Les fonctions de base numériques DNP incluent des fonctions de polarisation (comparable à la base gaussienne 6-311+G**) [16]. L'inclusion de fonctions de polarisation permettant une meilleure prise en compte des liaisons hydrogène établies dans le système modélisé. La représentation des orbitales pour les éléments H, C, N et O est la suivante :

- H : $1s^{(0)}$, $1s^{(1,3)}$, $2p^{(1,3)}$
- C : $1s^{(0)}$, $2s^{(0)}$, $2p^{(0)}$, $2s^{(2)}$, $2p^{(2)}$, $3d^{(7)}$
- N : $1s^{(0)}$, $2s^{(0)}$, $2p^{(0)}$, $2s^{(2)}$, $2p^{(2)}$, $3d^{(7)}$
- O : $1s^{(0)}$, $2s^{(0)}$, $2p^{(0)}$, $2s^{(2)}$, $2p^{(2)}$, $3d^{(7)}$

où l'exposant (0) signifie que l'orbitale a été générée à partir d'un atome neutre ; (2) indique que l'orbitale a été générée à partir d'un cation +2 ; (1,3) et (7) indique que les orbitales ont été générées à partir de calculs avec un noyau de nombre atomique $Z = 1, 3$ et 7 respectivement.

REFERENCES

- [1] A. Krammer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman. *LigScore : a novel scoring function for predicting binding affinities*. Journal of Molecular Graphics and Modelling 23 (2005) 395-407.
- [2] I. Muegge. *PMF revisited*. Journal of Medicinal Chemistry 49 (2006) 5895-5902.
- [3] I. Muegge, Y.C. Martin. *A general and fast scoring function for protein-ligand interactions : a simplified potential approach*. Journal of Medicinal Chemistry 42 (1999) 791-804.
- [4] D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel, S.T. Freer. *Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming*. (1995).
- [5] D.K. Gehlhaar, D. Bouzida, P.A. Rejto. *Rational drug design: novel methodology and practical applications* in: L. Parrill, M. Rami Reddy (Eds.). American Chemical Society, Washington, DC., 1999, pp. 292-311.
- [6] A.N. Jain. *Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities*. Journal of Computer Aided Molecular Design 10 (1996) 427-440.

- [7] R.D. Clark, A. Strizhev, J.M. Leonard, J.F. Blake, J.B. Matthew. *Consensus scoring for ligand/protein interactions*. Journal of Molecular Graphics and Modelling 20 (2002) 281-295.
- [8] M. Feher. *Consensus scoring for protein-ligand interactions*. Drug Discovery Today 11 (2006) 421-428.
- [9] X.D. Li, T.J. Hou, X.J. Xu. *Comparative studies of 14 binding free energies scoring functions*. Acta Physico - Chimica Sinica 21 (2005) 504-507.
- [10] M. Stahl, M. Rarey. *Detailed analysis of scoring functions for virtual screening*. Journal of Medicinal Chemistry 44 (2001) 1035-1042.
- [11] R. Teramoto, H. Fukunishi. *Surpevised concensus scoring for docking and virtual screening*. Journal of Chemical Informatics Modelling 47 (2007) 526-534.
- [12] R. Wang, Y. Lu, S. Wang. *Comparative evaluation of 11 scoring functions for molecular docking*. Journal of Medicinal Chemistry 46 (2003) 2287-2303.
- [13] P. Rabinowitz, P.J. Davis. *Methods of Numerical Integration* (2007).
- [14] J.P. Perdew, Y. Wang. *Accurate and simple analytic representation of the electron-gas correlation energy*. Physical Review B 45 (1992) 13244-13249.
- [15] B. Delley. *An all-electron numerical method for solving the local density functional for polyatomic molecules*. Journal of Chemical Physics 92 (1990) 508-517.
- [16] Y. Inada, H. Orita. *Efficiency of numerical basis sets for predicting the binding energies of hydrogen bonded complexes: evidence of small basis set superposition error compared to Gaussian basis sets*. Journal of Computational Chemistry 29 (2007) 225-232.

AUTORISATION DE SOUTENANCE DE THESE
DU DOCTORAT DE L'INSTITUT NATIONAL
POLYTECHNIQUE DE LORRAINE

o0o

VU LES RAPPORTS ETABLIS PAR :

Monsieur Serge ANTONCZAK, Professeur, Université de Nice-Sophia Antipolis, Nice

Monsieur Alain MARTY, Professeur, INSA, Toulouse

Le Président de l'Institut National Polytechnique de Lorraine, autorise :

Monsieur DE OLIVEIRA Eduardo Basilio

à soutenir devant un jury de l'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE,
une thèse intitulée :

« Simulations moléculaires appliquées à l'acétylation de flavonoïdes catalysée par des lipases : influence des structures de la lipase et du flavonoïde et sur la régiosélectivité de la bioconversion »

en vue de l'obtention du titre de :

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

Spécialité : **« Procédés biotechnologiques et alimentaires »**

NANCY BRABOIS
2, AVENUE DE LA
FORET-DE-HAYE
BOITE POSTALE 3
F - 54501
VANDŒUVRE CEDEX

Fait à Vandoeuvre, le 24 novembre 2009

Le Président de l'I.N.P.L.,
F. LAURENT



RÉSUMÉ

Les flavonoïdes sont des composés poly-hydroxylés d'origine végétale, connus pour leurs vertus pour la santé. Afin d'obtenir des dérivés plus stables et solubles dans des formulations hydrophobes tout en conservant les activités biologiques des molécules d'origine, une solution consiste à acyler ces composés de manière régiosélective. Ceci peut être accompli en utilisant des lipases comme catalyseurs, en milieu organique. Grand nombre d'études expérimentales sur ces bioprocédés sont disponibles, mais aucune d'entre elles n'apporte d'explication, au niveau moléculaire, de la sélectivité de ces réactions d'acylation. Le but de cette étude est d'appliquer différents outils de simulation moléculaire pour mieux comprendre, au niveau moléculaire, les propriétés de sélectivité de l'acétylation de trois flavonoïdes (quercétine et ses dérivés glycosylés isoquercitrine et rutine), en utilisant les lipases CALB et PCL. D'abord, des simulations de docking ont été appliquées, afin d'obtenir les positions et les orientations les plus probables des flavonoïdes dans la cavité des lipases préalablement acétylées. Ensuite, des simulations de dynamique moléculaire ont été exécutées sur les complexes obtenus par docking, afin d'étudier stabilité structurale des complexes sur une période de temps et notamment la stabilité des interactions enzyme-substrats. Enfin, des simulations basées sur une approche de chimie quantique (DFT) ont été appliquées pour évaluer la réactivité chimique des flavonoïdes dockées dans les complexes. Les premières tendances observées aux cours des simulations ont présenté une bonne corrélation avec les résultats expérimentaux d'acétylation. Globalement, les résultats obtenus ont montré que la sélectivité de ces réactions dépend de l'orientation des substrats (flavonoïde et acétate) dans la cavité catalytique de la lipase, des interactions intermoléculaires stabilisant ces substrats et de la réactivité chimique intrinsèque des groupements OH des flavonoïdes se situant à proximité des résidus catalytiques.

Mots-clé : *Flavonoïde, Lipases, Acétylation, Régiosélectivité, Docking, Dynamique moléculaire, DFT.*

ABSTRACT

Flavonoids are plant-produced polyhydroxylated compounds, well-known for their beneficial health effects. In order to obtain more stable and soluble derivatives for incorporation in hydrophobic formulations without damaging the biological activities of the native molecules, a solution consists to perform a regioselective acylation of these molecules. This can be accomplished by using lipase biocatalysts, in organic media. Several experimental studies dealing with such processes are available, but none of them give any explanation, at the molecular level, for the regioselectivity of such reactions. This study aimed to apply different molecular modelling tools in order to better understand, at the molecular level, the selectivity properties of the acetylation of three flavonoids (quercetin and its glycosylated derivatives isoquercitrin and rutin), by using the lipases CALB and PCL. Firstly, docking simulations were applied, in order to obtain the most probable positions and orientations of the flavonoids in the cavities of acetylated lipases. Then, molecular dynamics simulations were performed, aiming to study the structural stability of the complexes upon a period of time and specially the stability of the enzyme-substrates interactions. Finally, quantum chemical simulations (DFT) were applied to evaluate the chemical reactivity of the flavonoids as docked in the complexes. The trends observed during the simulations were well correlated with previous experimental results on the acetylation reaction of these flavonoids. Overall, the results showed that the selectivity in such reactions depends upon the substrates (flavonoid and acetate) orientations in the enzyme catalytic cavity, the intermolecular interactions that stabilize these substrates and the intrinsic chemical reactivity of the flavonoids OH groups reaching the catalytic residues.

Keywords: *Flavonoid, Lipase, Acetylation, Regioselectivity, Docking, Molecular dynamics, DFT.*