

# UNIVERSITE ASSANE SECK DE ZIGUINCHOR



UFR Sciences et Technologies

## ECOLE DOCTORALE SCIENCES, TECHNOLOGIES ET INGENIERIE

**Laboratoire de Chimie et Physique des Matériaux**

**Département de CHIMIE**

Année 2024

**Thèse de DOCTORAT**

Domaine : Chimie

Spécialité : Chimie du Solide et des Matériaux

Présentée par :

**Ibrahima GOUDIABY**

Etudes Structurales et Fonctionnelles de l'Acide Laurique et des Protéines Lipase B de *Candida Antarctica* (CALB), de la Neuropiline-1 (Nrp-1).

Soutenue publiquement le **17 février 2024** devant un jury composé de :

M. **Claude LECOMTE**, Professeur Emérite, Université de Lorraine (France) / Président

M. **Ibrahima El hadji THIAM**, Professeur Titulaire, Université Cheikh Anta Diop de Dakar / Rapporteur

Mme. **Rokhaya Sylla GUEYE**, Maître de Conférences Agrégée, Université Cheikh Anta Diop de Dakar/ Rapporteuse

M. **Aliou Hamady BARRY**, Professeur Titulaire, Université de Nouakchott (Mauritanie) / Rapporteur

Mme. **Frédérique FAVIER**, Maître de Conférences, Université de Lorraine (France) / Examinatrice

M. **Magatte CAMARA**, Professeur Titulaire, Université Assane Seck Ziguinchor / Directeur de thèse.

M. **Christian JELSCH**, Directeur de Recherche au CNRS, Université de Lorraine (France) / Co-encadrant de thèse.

M. **Abdoulaye GASSAMA**, Professeur Titulaire, Université Assane Seck Ziguinchor / Co-encadrant de thèse

*Travail effectué dans le Laboratoire de Chimie et Physique des Matériaux de l'Université Assane SECK de Ziguinchor-Sénégal et dans le Laboratoire de Cristallographie, Résonance Magnétique et Modélisations (CRM<sup>2</sup>), UMR CNRS 7036, Nancy-France*



## Dédicaces

À mes **chers parents**, dont le soutien inconditionnel et l'amour indéfectible ont été la source même de mon inspiration, de ma patience et de ma persévérance. Votre encouragement constant m'a guidé à travers chaque étape de ce parcours académique et je vous dédie cette thèse avec une profonde gratitude.

À mes chers frères **Amady, Boubacar et Baboucar**, ainsi qu'à mes sœurs **Seynabou, Fatou et Hawa**, vos encouragements, vos conseils et votre présence ont été des piliers sur lesquels j'ai pu m'appuyer tout au long de ce voyage intellectuel. Cette thèse est aussi la vôtre, car elle reflète l'importance de la famille dans ma vie.

À la **famille Française : JUNKE (Etienne JUNKE, Marie-Françoise François JUNKE, Anne-Laure JUNKE, Emmanuelle JUNKE, Noa et Eden ARNAUD, Kaïs NASRAOUI...)**.

Il est des rencontres qui marquent une vie de manière indélébile, et la vôtre en fait incontestablement partie. Votre générosité, votre accueil chaleureux et votre soutien inconditionnel ont transformé mes séjours en France en expériences inoubliables. Cette dédicace est bien plus qu'une simple reconnaissance, c'est une expression profonde de ma gratitude envers chaque membre de la famille cité ci-dessus. Votre hospitalité a été une lumière dans mes moments d'incertitude, une épaule solide sur laquelle m'appuyer, et un foyer où je me sens toujours le bienvenu. À travers cette thèse, je souhaite témoigner de l'impact majeur que vous avez eu sur mon parcours. Votre bonté et votre gentillesse ont agi comme des catalyseurs qui ont façonné ma vision de l'entraide humaine, de la vie et renforcé ma foi en la bienveillance humaine. Puissent ces mots transmettre l'étendue de ma reconnaissance et de ma considération envers vous, la famille JUNKE. Que cette dédicace symbolise l'attachement sincère que je ressens envers votre famille, à laquelle j'appartiens désormais selon vous, et que notre lien perdure au-delà de ces mots, pour toujours.

A **Mado BERGER**, en toi, j'ai trouvé bien plus qu'une amie, mais une véritable étoile qui a illuminé chacun de mes jours lors de mon séjour à Nancy. Ta présence bienveillante, ton soutien inébranlable et ta générosité sans limites ont été des cadeaux précieux qui ont marqué mon expérience de manière profonde. Cette dédicace est une humble tentative de reconnaître l'impact énorme que tu as eu sur ma vie pendant cette période. Par cette thèse, je souhaite témoigner de l'importance de ton soutien.

## Dédicaces et Remerciements

À la **Fondation Rotary Club de Nancy**, votre engagement en faveur de l'éducation a été un moteur puissant dans ma quête du savoir. Votre investissement pour la jeunesse est primordial : grâce à des organisations comme la vôtre, les rêves deviennent réalité, en particulier les miens.

À mon cher **cousin et grand Malick DIATTA**, en toi, j'ai trouvé un confident et un complice de vie. Ta présence chaleureuse et ton soutien indéfectible ont éclairé chacune de mes étapes, y compris celle de cette thèse. À travers les hauts et les bas, tu as toujours été là pour me rappeler l'importance de la persévérance et de la foi en soi, merci infiniment grand.

Je souhaite également dédier cette thèse à la **famille TAMBA** de Coubanao, Votre encouragement et votre aide ont été d'une valeur inestimable. Votre foi en mes capacités m'a constamment poussé à donner le meilleur de moi-même.

Enfin, je dédie cette thèse aux familles **SANE et DIATTA**, qui m'ont soutenu de manière inconditionnelle tout au long de mes années d'études. Je vous dis que votre soutien moral et votre bienveillance ont été des éléments moteurs dans la poursuite de mes objectifs.



# Remerciements

Je tiens à exprimer ma profonde gratitude envers **Allah**, le **Tout-Puissant**, le Miséricordieux, qui m'a accordé la force et la patience nécessaire pour mener à bien cette thèse.

Le travail présenté dans ce manuscrit a été réalisé au sein de deux laboratoires, à savoir le Laboratoire de **Chimie et Physique des Matériaux (LCPM)** de l'Université Assane SECK de Ziguinchor et le Laboratoire de **Cristallographie, Résonance Magnétique et Modélisation (CRM<sup>2</sup>)** de l'Université de Lorraine, sous la direction du **Professeur Magatte CAMARA**, du **Professeur Christian JELSCH** et du **Professeur Abdoulaye GASSAMA**,

Je tiens à exprimer ma vive gratitude envers les membres du jury de ma thèse qui m'honorent en acceptant de juger ce travail :

Tout particulièrement au professeur émérite **Claude LECOMTE** qui me fait l'honneur d'avoir accepté de présider ce jury. Je lui témoigne toute ma reconnaissance pour avoir participé à ma formation.

Merci aux professeurs **Ibrahima El hadji THIAM** et **Aliou Hamady BARRY** et à Mme **Rokhaya Sylla GUEYE**, maître de conférences agrégée, qui me font l'honneur de siéger parmi les membres de mon jury en tant que rapporteurs. Je leur exprime ma profonde gratitude pour l'intérêt qu'ils portent à ce travail et pour l'enseignement qu'ils m'ont prodigué.

De même Merci à **Frédérique FAVIER**, maître de conférences qui me fait l'honneur d'accepter d'évaluer ce travail en tant qu'examinatrice.

Je tiens à exprimer mon profond respect :

**Pr. Magatte CAMARA** pour m'avoir guidé à travers ce véritable travail de recherche au sein du laboratoire **LCPM**. Je le remercie sincèrement pour la qualité de son encadrement, sa disponibilité et l'énergie qu'il a investies pour la finition de cette thèse. Je lui suis également redevable de m'avoir fait partager sa passion pour la Cristallographie.

**Pr. Abdoulaye GASSAMA** pour sa passion pour l'enseignement et son aide envers les doctorants sont indéniables. Son amour pour la recherche, la transmission du savoir et son investissement dans la progression des doctorants ont été palpables à chaque étape de nos échanges. Ses

explications pertinentes et sa capacité à rendre accessible même les concepts les plus complexes m'ont été d'une aide inestimable pour avancer dans mes synthèses de couplage.

**Dr. Christian JELSCH** pour l'accueil exceptionnel que j'ai reçu lors de mes séjours au Laboratoire CRM<sup>2</sup>. Son soutien et son mentorat ont été des éléments clés de mon expérience et de mon apprentissage. Sa générosité dans le partage de son temps, de ses connaissances et de son expertise a créé un environnement où je me suis senti soutenu et encouragé à explorer de nouvelles idées et à développer mes compétences. Sa formation approfondie en cristallisation des protéines a été d'une aide inestimable. De plus, sa rigueur scientifique et son souci du détail m'ont poussé à mener mes synthèses avec une précision accrue. Son expertise dans l'utilisation de différents logiciels cristallographiques m'a ouvert de nouvelles perspectives et m'a donné les compétences techniques essentielles pour mener à bien mes travaux de recherche. Son influence positive sur ma formation restera gravée dans ma mémoire, et je suis honoré d'avoir eu l'opportunité de bénéficier de son expertise et de sa bienveillance.

**Dr. Frédérique FAVIER**, pour son implication exceptionnelle et fructueuse. Son suivi attentif, et le temps qu'elle m'a consacré ont été des éléments cruciaux de mon parcours académique et je tiens ici à lui adresser ma profonde gratitude.

**Dr. Claude DIDIERJEAN**, pour son rôle essentiel et couronné de succès dans ma thèse. Son implication remarquable, sa pédagogie exceptionnelle et son encadrement précieux ont été des facteurs déterminants dans la réalisation de ma thèse. Je tiens à lui adresser mes remerciements les plus sincères. Sa capacité à partager ses connaissances avec clarté m'a permis d'approfondir ma compréhension en biologie structurale. Sa pédagogie extraordinaire a rendu les concepts ardu plus accessibles et a enrichi mon apprentissage de manière significative. Son encadrement attentif et bienveillant a été une source constante de motivation pour moi. Le temps qu'il a généreusement consacré à discuter avec moi et à me guider a été inestimable. Sa disponibilité pour répondre à mes questions et pour fournir des orientations pertinentes a grandement facilité mon avancée dans ma thèse. Les nombreuses réunions fructueuses que nous avons eues ont été des occasions précieuses pour évaluer mes avancées. Ces échanges ont été des moments-clés dans la consolidation de ma thèse.

**Pr. Benoît GUILLOT**, pour son encadrement exceptionnel et ses explications précieuses. Sa guidance et son expertise m'ont été d'une aide précieuse tout au long de mon séjour à Nancy. Je

## Dédicaces et Remerciements

le remercie d'avoir pu bénéficier de ses expertises sur le logiciel Mopro. Merci infiniment, Benoit Guillou, pour votre dévouement et votre soutien.

**Dr. EL-Eulmi BENDEIF**, Les encouragements constants du Dr BENDEIF m'ont motivé à persévérer face aux défis rencontrés tout au long de ce projet.

**Madame Sandrine MATHIOT**, Je souhaite lui exprimer ma profonde gratitude pour sa générosité et sa disponibilité dans le partage de ses connaissances sur la cristallisation des protéines. Les discussions enrichissantes que j'ai eues avec elle dans son bureau ont joué un rôle majeur dans mon apprentissage et ma compréhension de ce domaine, la cristallisation des protéines.

**Dr Emmanuel WENGER**, je tiens à le remercier chaleureusement pour les mesures effectuées.

**Mesdames Anne CLAUSSE et Valérie RICHALET**, secrétaires au Laboratoire CRM<sup>2</sup>

A la direction du Laboratoire CRM<sup>2</sup> pour m'avoir accueilli comme stagiaire.

**Dr. Eva MOCCHETTI, Dr. Julien KAPPELE, Dr. Vedran VUKOVIĆ, Dr Théo Le DUC, Dr. Mathieu SCHWARTZ**, je tiens à vous remercier sincèrement pour l'accueil chaleureux, l'amitié solide et le soutien inestimable que vous m'avez offert pendant mon séjour de recherches au laboratoire CRM<sup>2</sup>. Nos échanges intellectuels et nos discussions passionnantes ont profondément enrichi mon expérience de recherche et ont contribué à forger des souvenirs inoubliables. Travailler à vos côtés a été une expérience mémorable, et je vous suis reconnaissant pour toutes les discussions, les rires et les moments partagés. Bien qu'au fil du temps nos chemins se séparent, je suis convaincu que nos liens d'amitié et de collaboration perdureront. Merci à vous tous pour cette aventure exceptionnelle à Nancy.

Mention spéciale à la famille **JUNKE et Amis**, à la famille **BERGER, FRANÇOIS, ARNAUD, CHEVRIER, Corine HENRY** pour leur soutien, leur aide précieuse et leurs encouragements

Un grand merci à **Evelyne et Lucien LARRET** pour tout ce qu'ils ont fait pour moi.

Je tiens à saluer chaleureusement mon ami de longue date, **Raymond GEORGES** et son épouse, **Nelly**, pour m'avoir appris un vocabulaire auquel je n'aurai pas eu accès sans eux.

Toute ma reconnaissance au **Pr. Pierre BRAVETTI**, doyen honoraire de la faculté d'Odontologie de Nancy, pour ses encouragements et sa grande bienveillance.

## Dédicaces et Remerciements

A la famille SANE de France : **Gnima, Arabietou, Amara, Sékouba** et leurs familles pour leur accueil à chaque fois que j'arrivais à **Paris**.

J'adresse mes sincères remerciements à mes 4 petits : **Dr. Aliou DIATTA, Dr. Philippe HIMBANE, Dr. Mamoudou DIALLO et Monsieur Malang KANDE**, pour leurs conseils précieux et leurs amitiés sans faille depuis la Licence 1 MPC1 à UASZ.

J'adresse mes remerciements sincères à mon ami **Dr. Papa Aly GAYE**, pour ses conseils précieux et encouragements, j'espère vous avoir encore longtemps à mes côtés.

**Pr. Chérif BALDE et Mbaye Ndiagne MBAYE**, je tiens à vous adresser mes remerciements les plus sincères pour vos conseils précieux qui ont joué un rôle déterminant dans le succès de ma thèse.

**Pr. Diouma KOBOR**, directeur de l'Ecole Doctorale Sciences Technologies et Ingénierie et tous ces collaborateurs pour les formations doctorales intéressantes.

**Dr. Soumeya MAZA**, je tiens à remercier le **Dr Soumeya MAZA** pour ses conseils avisés et ses encouragements précieux. Merci infiniment, **Dr Maza**.

**Dr. Mame Seyni Sylla DIOME**, pour ses conseils avisés et sa source inépuisable de motivations. Les moments passés à discuter avec elle m'ont non seulement permis d'acquérir de nouvelles connaissances, mais m'ont également inspiré à dépasser mes limites. Merci infiniment, **Dr. Sylla DIOME**.

**Dr. Insa BADIANE, Dr. Albert Manga BADIANE, Dr. Oumar SAMBOU**, pour leurs précieux conseils et leurs motivations constantes. Je tiens à les remercier chaleureusement pour leur générosité, leur bienveillance et leur investissement dans mon succès.

**Dr. Armel DIATTA, Dr. Anastasie MANGA, Dr. Abdou MBAYE, Dr. Mamadou LO, Dr. Daouda NDIAYE, Dr. Malamine MANSALY, Dr. Yackouba KONTE, Dr. Rokhyatou SECK, Dr. Charlotte SECK, Dr. Aminata THIAM, Dr. Diarra DIATTA,**

**Dr. Mohamed DIEME, Dr. Seydou BA, Dr. Rémi NDIIOUKANE**, pour leurs soutiens, conseils et encouragements.

**Saybatou YAGUE, Penda Seydi DEME, Ngara DIOUF, Moussa TOURE, Ibrahima Paka MANGA, Thierno Abdoulaye BARRY, Salimata CAMARA, Henry DIASSY, Abdoulaye**

## Dédicaces et Remerciements

DIATTA, Eliane Adélaïde MANGA, Siga SAGNE, Emma DIATTA, Amadou DIOUF, Amadou DIEYE, Fallou NGOM et aux autres pour leurs encouragements quotidiens.

Madame Fatoumata Soumaré CAMARA, technicienne du laboratoire de chimie

Mamadou Lamine BADJI, Siré COLY, et leurs enfants ; Siré, Ami et Bassirou pour tout ce qu'ils ont fait et continue à faire pour moi.

À toutes celles et ceux qui ont contribué de près ou de loin à l'aboutissement de cette thèse.

**A mes Parents**

**A mes Frères et sœurs**

---

---

*Sommaire*

---

---



## Table des matières

Abbréviations.....	5
Introduction générale .....	9
Partie A – Généralités sur les réactions en chimie.....	13
Partie B – Généralités sur les protéines et petites molécules utilisées .....	29
Partie C – Objectifs de la thèse .....	95
Partie D – Matériel et méthodes .....	99
Partie E – Résultats et discussions .....	135
Conclusion générale et perspectives de recherche .....	191
Annexes .....	197
Résumé des travaux (français) .....	205
Résumé des travaux (anglais) .....	207





---

---

## *Abréviations*

---

---



ACE-2	Enzyme de conversion de l'angiotensine-2
ADN	Acide désoxyribonucléique
ARNm	Acide ribonucléique messenger
Bis Tris	Bis-(2-hydroxyethyl) imino-tris(hydroxymethyl)methane
CALB	<i>Candida Antarctica</i> Lipase B
COVID-19	Coronavirus Disease-19 (Maladie à coronavirus 19)
DHA	Acide docosahexaénoïque
DMF	Diméthylsulfoxyde
DMSO	N, N-Diméthylformamide
EC	Enzyme Commission
FMN	Flavine mononucleotide
IC <sub>50</sub>	Concentration inhibitrice médiane
MAPK	Mitogen-activated protein kinases
MPD	2-méthyl-2,4-penthanediol
NADP	Nicotinamide adénine dinucléotide phosphate
NAG	<i>N</i> -acétyl- <i>D</i> -glucosamine
NiO <sub>2</sub>	Oxyde de Nickel
Nrp-1	Neuropiline-1
Nrp-2	Neuropiline-2
PDB	Protein Data Bank
PEG	Polyéthylène Glycol
PI3K	Phosphoinositide 3-kinase
PIGF	Facteur de croissance placentaire
RMSD	Root-Mean-Square Deviation
RTK	Recepteur Tيروسine Kinase
SARS-CoV-2	Coronavirus 2 du Syndrome Respiratoire Aigu Sévère
SiO <sub>2</sub>	Oxyde de silicium
SNC	Système nerveux central
TM	Transmembranaire
Tris	2-Amino-2-(hydroxymethyl) propane-1,3-diol.
VEGF	Facteur de croissance endothélial vasculaire
VIH	Virus de l'immunodéficience humaine

---

---

*Introduction*

---

---



Au cœur du monde scientifique, l'exploration incessante des mystères de la nature nous pousse à sonder les profondeurs de la connaissance. Ce travail de thèse est comme un voyage dans le monde de la chimie et de la biologie pour élucider des phénomènes complexes. En quatre parties, ce manuscrit aborde les différentes réactions en chimie, la lipase B de *Candida Antarctica* (CALB), la Neuropiline-1 (Nrp-1), la synthèse des dérivés de l'acide bétulinique, l'acide laurique.

L'objectif initial était d'améliorer la solubilité de l'acide bétulinique dans l'eau, en faisant une modulation au niveau de sa fonction alcool par une réaction de couplage pseudo-peptidique.

La première partie de ce manuscrit nous conduira à travers un rappel bibliographique approfondi. Nous plongeons dans les profondeurs des réactions chimiques en milieu aqueux. Les réactions enzymatiques, catalysées par des biomolécules fascinantes, les enzymes, occuperont également une place centrale dans notre revue bibliographique. Enfin, nous nous plongeons dans le monde des réactions chimiques en milieu solide, où les conditions diffèrent considérablement de celles en solution liquide.

La deuxième partie de notre thèse s'intéressera à l'état de l'art actuel concernant les protagonistes de notre étude : la lipase B de *Candida Antarctica* (CALB), une enzyme d'une grande importance industrielle, et la Neuropiline-1 (Nrp-1), une protéine humaine clé dans la régulation de la croissance cellulaire. De plus, nous explorons les propriétés et les applications potentielles de petites molécules, telles que l'acide bétulinique et l'acide laurique.

La troisième partie de notre travail dévoilera une étude détaillée des matériels et méthodes que nous avons utilisés tout au long de notre thèse.

Enfin, dans la quatrième et dernière partie, nous dévoilerons les résultats issus de nos recherches. Nous présentons les avancées majeures obtenues grâce à la cristallisation des deux protéines étudiées, ainsi les résultats de la synthèse des dérivés de l'acide bétulinique, de l'acide laurique. Les résultats obtenus ont donné lieu à deux publications acceptées et parues dont les textes sont joints à ce manuscrit.





---

---

*- Partie A -  
Généralités sur les réactions  
en chimie*

---

---



## Table des matières

- Partie A - Généralités sur les réactions en chimie .....	13
A - I. Les réactions enzymatiques .....	19
A - II. Les réactions chimiques en milieu solide .....	23
Références Partie A.....	25



Il existe plusieurs types de réactions en chimie. Parmi celles-ci, nous allons nous pencher sur les réactions de couplage, les réactions de couplage pseudo-peptidiques, les réactions enzymatiques et les réactions en milieu solide.

L'estérification a été étudiée pour la première fois en 1861 par Berthelot et Armand Péan de Saint-Gilles. Ils ont démontré que la réaction d'estérification était accompagnée de sa réaction inverse, qui est l'hydrolyse de l'ester. De manière générale, une fonction alcool et une fonction acide carboxylique réagissent ensemble en présence d'un catalyseur acide pour former un ester et de l'eau (Schéma 1).

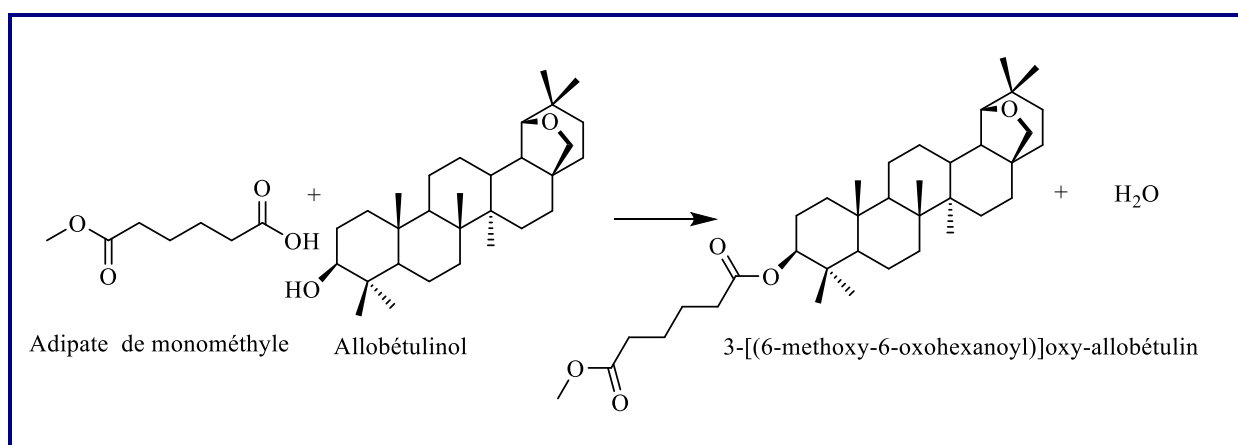


Schéma 1 : Formation d'un ester par le couplage acide carboxylique / alcool

Au cours de ce processus, une molécule d'eau est générée sous une forme quelconque. Comme la réaction a atteint un équilibre avant d'être complète, il est important d'ajuster les conditions réactionnelles afin de favoriser la formation des produits [1]. Si la molécule d'eau formée lors du couplage est retirée du milieu réactionnel, l'équilibre sera déplacé en faveur de la formation des produits. Il existe plusieurs stratégies pour atteindre cet objectif. La première consiste à utiliser l'hexafluorophosphate de benzotriazolyl-N-oxytris(diméthylamino)phosphonium (HOBT) dans des solvants tels que l'acétonitrile, le dichlorométhane, l'acétone ou le diméthylformamide [2] (Schéma 2).

Dans ce cas, l'espèce qui réagit avec l'alcool n'est pas un anhydride mais un ester de benzotriazole, une molécule très réactive (mécanisme à l'annexe 1). Par la suite, l'alcool se substitue au benzotriazole pour former l'ester désiré.

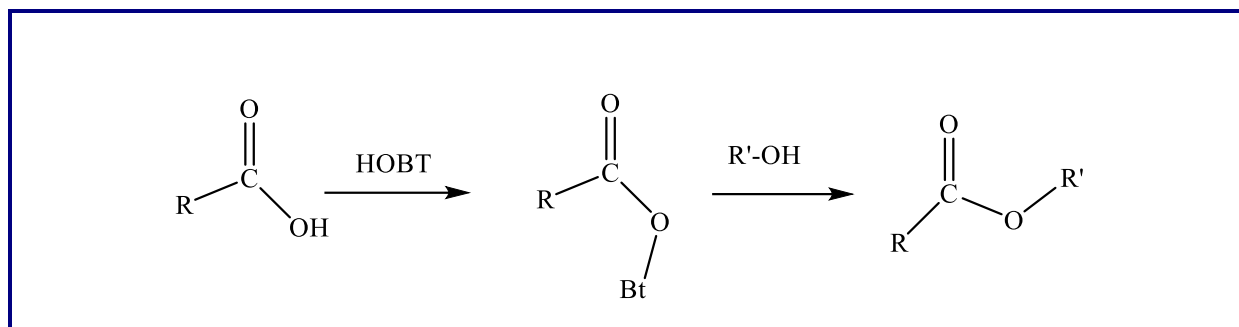


Schéma 2 : Estérification avec HOBT

La seconde stratégie consiste à préparer un chlorure d'acyle, un dérivé très réactif envers les nucléophiles (Schéma 3).

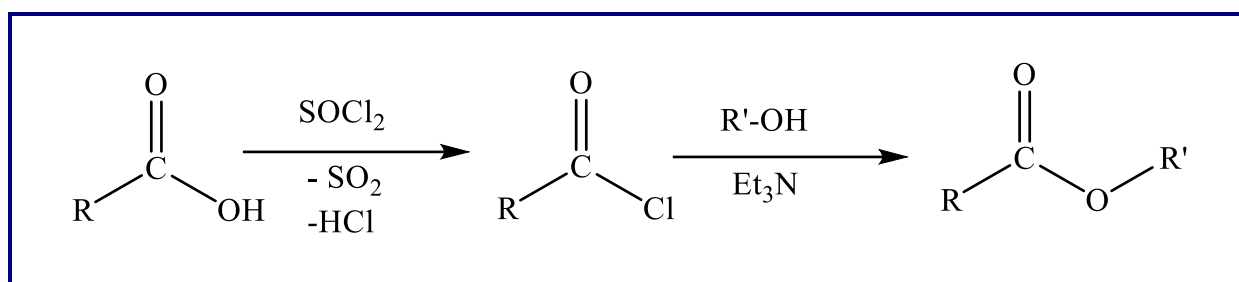


Schéma 3 : Estérification via le chlorure d'acide

Le chlorure de thionyle ( $\text{SOCl}_2$ ) est utilisé pour préparer ce dérivé (mécanisme Annexe I). Au cours de la réaction, du dioxyde de soufre ( $\text{SO}_2$ ) est libéré, ce qui force la réaction vers la formation du produit [2]. L'alcool est ensuite ajouté aux milieux réactionnels avec une base comme la triéthylamine ( $\text{Et}_3\text{N}$ ) qui va piéger l'acide chlorhydrique libéré déplaçant ainsi la réaction vers la formation de l'ester.

La dernière stratégie consiste à utiliser le dicyclohexylcarbodiimide (EDCI) avec le diméthylaminopyridine (DIEA) comme catalyseur [3]. Dans ce cas, deux fonctions acides carboxylique réagissent avec le EDCI pour former l'anhydride correspondant (mécanisme Annexe I). Durant ce processus, une molécule de dicyclohexylurée (DCU = DCC +  $\text{H}_2\text{O}$ ) précipite dans le milieu, forçant ainsi la réaction vers la formation de produits (Schéma 4) [4].

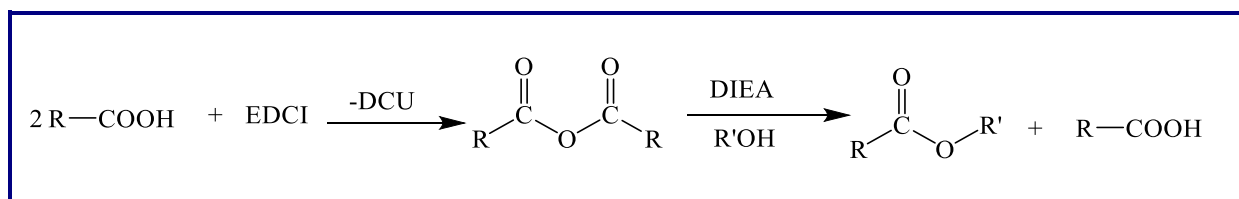


Schéma 4 : Estérification avec EDCI et DMAP

### A - I. LES REACTIONS ENZYMATIQUES

Une enzyme est un catalyseur biologique, généralement une protéine d'une efficacité fonctionnelle remarquable[5]. Elle augmente la vitesse d'une réaction donnée mais sans modifier la constante d'équilibre chimique (conditions thermodynamiques) de la réaction entre le substrat et le produit [6]. Cette propriété est due au fait que l'enzyme peut soit diminuer l'énergie d'activation ( $E_a$ )[7] de l'étape lente, soit rendre possible un trajet réactionnel alternatif, en plusieurs étapes, où l'étape cinétiquement déterminante est plus basse en énergie qu'au cours d'un processus non catalysé (Figure 1). Une accélération de la vitesse de réaction de l'ordre de plusieurs millions de fois est courante dans le cas des réactions catalysées par des enzymes. Les enzymes sont régénérées à l'issue de la réaction qu'elles catalysent. Seul quelques acides aminés participent à la fixation et à la catalyse, les autres servent au maintien du repliement 3D de l'enzyme.

Les réactions catalysées par les enzymes humaines ou de mammifères s'effectuent généralement à un pH proche de 7 et à une température de 37 °C : on parle de pH optimal et de température optimale. L'efficacité catalytique d'une enzyme peut être très fortement réduite lorsque les conditions expérimentales s'éloignent de sa température optimale ou de son pH optimal. Les enzymes de bactéries, de plantes, de champignons opèrent dans la nature dans des gammes de températures.

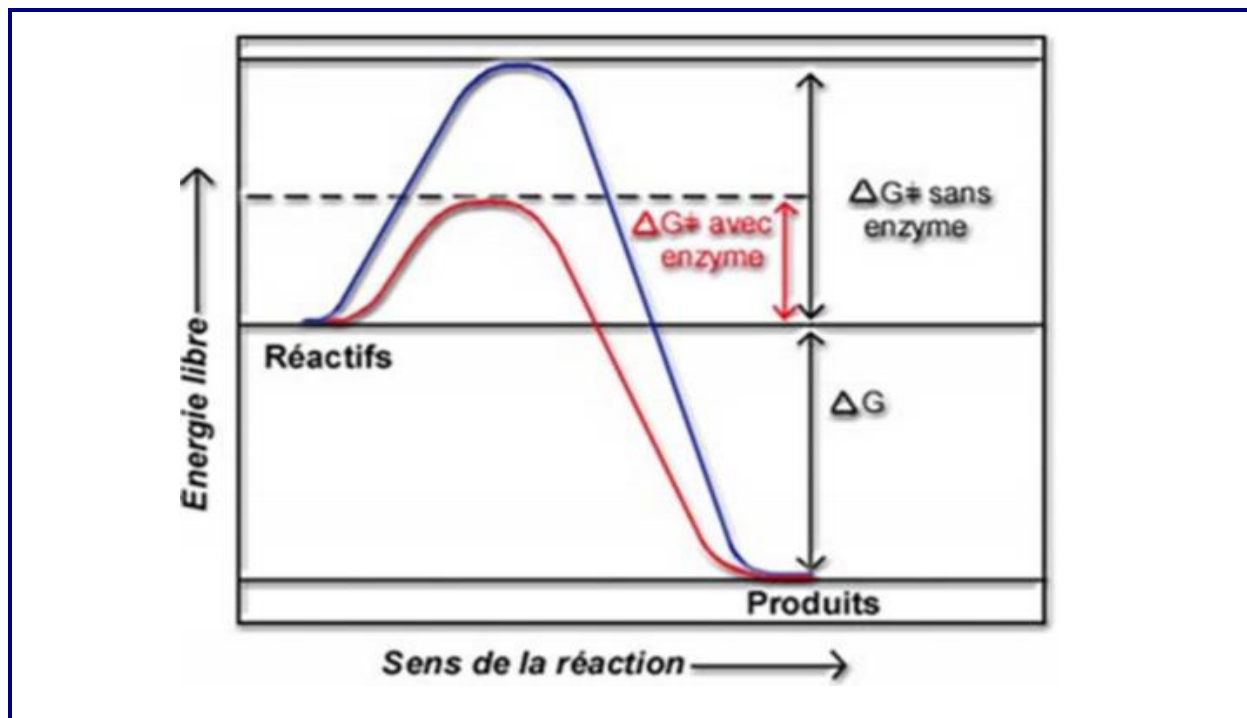


Figure 1 : Variation de l'énergie d'activation  $E_a$  [8]

Les enzymes se distinguent des autres types de catalyseurs par leur très grande spécificité qui découle de leurs structures 3D.

Les enzymes fonctionnent suivant un processus impliquant : (1) la formation initiale d'un complexe enzyme-substrat [E.S], (2) la transformation en plusieurs étapes du complexe enzyme-substrat en un complexe enzyme-produit [E.P] et finalement (3) la libération du produit.

A partir du moment où le complexe [E.S] est formé, ce dernier peut soit redonner les réactifs de départ soit débiter le processus catalytique conduisant au produit. Si, comme c'est souvent le cas, la constante de vitesse de formation du complexe [E.S] ( $k_1$ ) et la constante de vitesse de la formation du complexe [E.P] ( $k_3$ ) sont grandes par rapport à la constante de vitesse de la réaction ( $k_2$ ), alors la transformation du complexe enzyme-substrat en complexe enzyme-produit est l'étape limitante de la réaction. La constante de vitesse globale de la conversion du complexe [E.S] en produit [E+P] est la constante catalytique ( $k_{cat}$ ), aussi appelée « Turn Over Number en anglais » (Figure 2). L'unité d'activité enzymatique est le Katal, qui représente le nombre de moles de substrats que l'enzyme est capable de transformer par unité de temps [9].

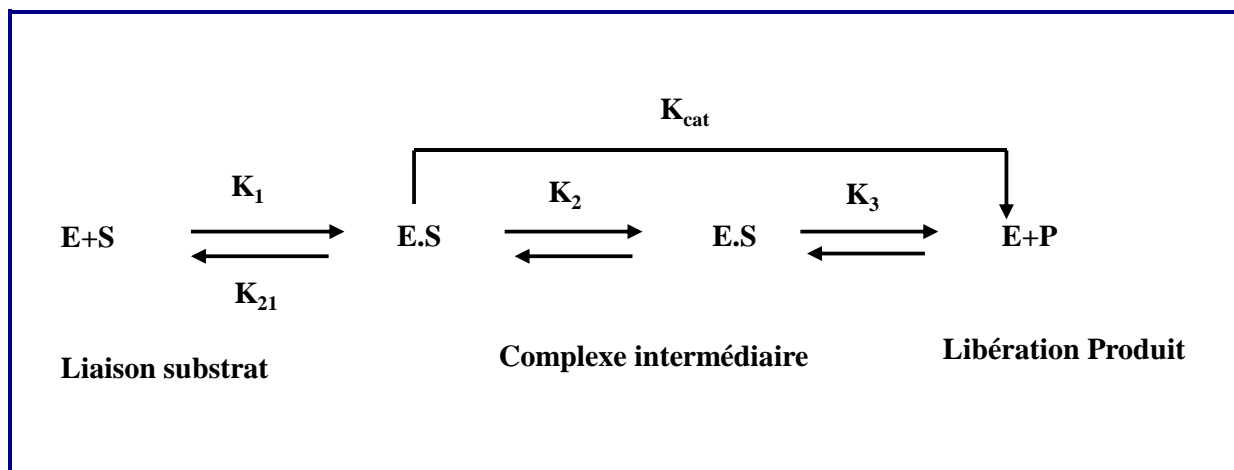


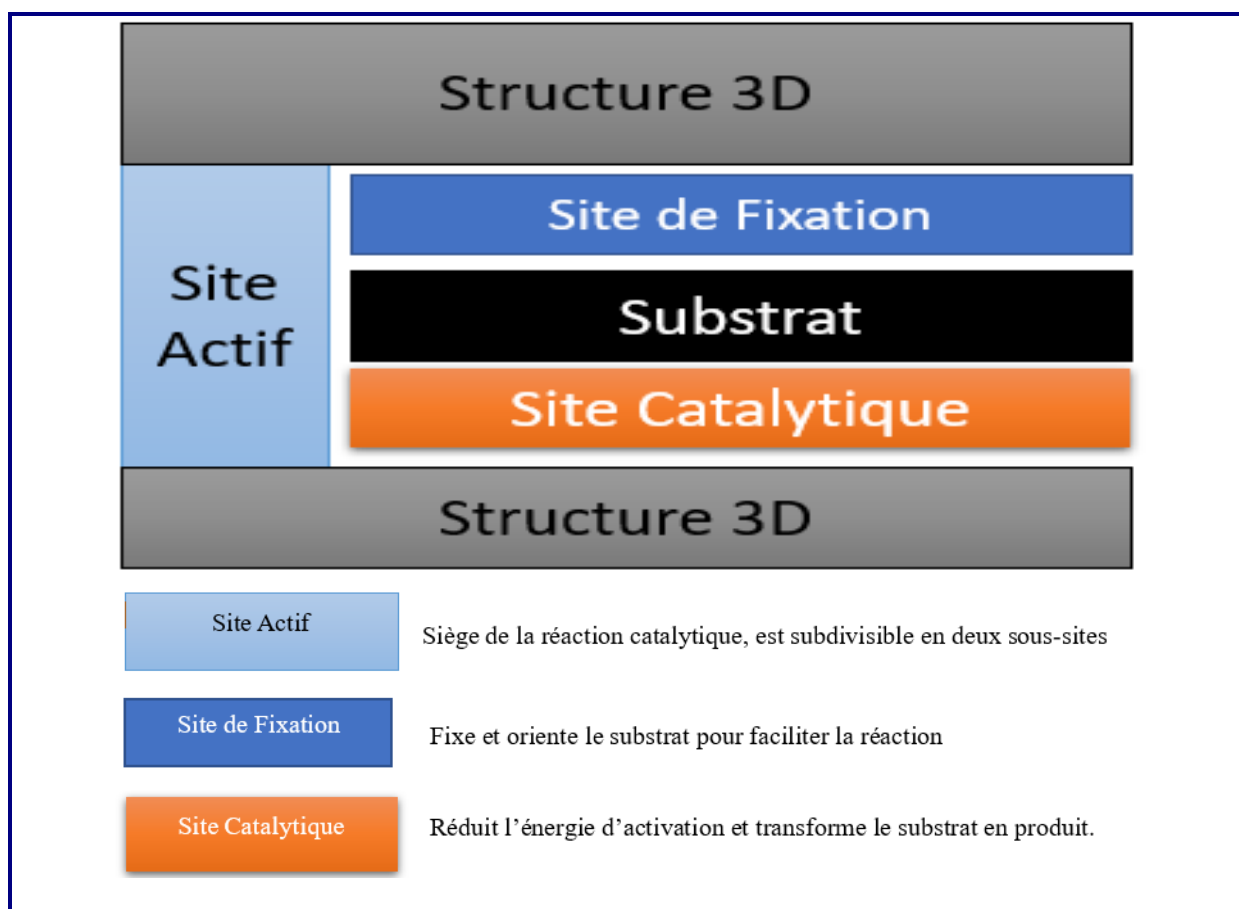
Figure 2 : Réaction enzymatique générale

La fixation du substrat et le processus catalytique ont lieu dans une cavité de l'enzyme qui est appelée site actif [10]. Les enzymes étant constituées d'acides aminés chiraux présentant une grande variété de substituants, le site actif est lui aussi chiral et peut adopter une forme complémentaire à celle du substrat, un peu comme un gant complémentaire à la main [11].



Les molécules sur lesquelles agissent les enzymes sont appelées substrats de réaction plutôt que ligands, et le site de liaison du ligand est appelé site catalytique ou site actif (Figure 3).

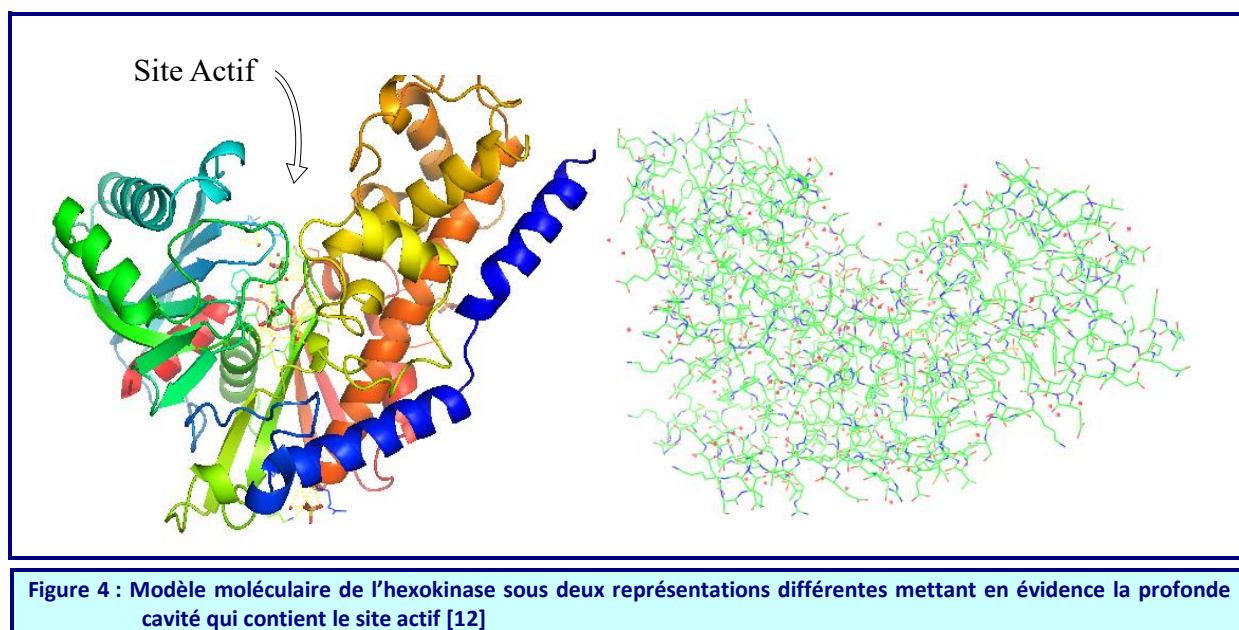
Les acides aminés spécifiques du site de fixation vont interagir avec le substrat pour le lier à l'enzyme avec une certaine affinité ; cette liaison va alors induire une modification locale de la structure 3D de l'enzyme contribuant à abaisser la barrière énergétique et permettant aux acides aminés du site catalytique de transformer le substrat en produits : on parle d'ajustement induit et d'état de transition.



**Figure 3 : Composition d'un site actif d'une enzyme**

Le site actif est tapissé d'acides aminés présentant les propriétés de polarité, d'acidité ou de basicité nécessaire à la catalyse d'une réaction spécifique. De plus, le site actif contient aussi les cofacteurs, ions métalliques ou petites molécules organiques, utiles pour la réaction.

La Figure 4 montre un modèle moléculaire de l'hexokinase, une enzyme qui catalyse la phosphorylation du glucose lors de la première étape de la glycolyse.



Dans l'industrie pharmaceutique, de nombreux principes actifs sont des molécules chirales, dont seul un des deux isomères est biologiquement actif, alors que l'autre forme est au mieux moins efficace, au pire toxique. Citons le cas de l'ibuprofène, dont seule la forme S est active. Dans le cas de la thalidomide, la forme L est un anti nauséeux et la conformation D est responsable de malformations [13]. Les synthèses chimiques classiques conduisent à des mélanges racémiques (R, S ou L, D) dont la séparation est ensuite délicate, coûteuse, voire impossible [9]. Au contraire, de nombreuses réactions enzymatiques permettent d'obtenir des produits optiquement purs, c'est-à-dire dont seule une des deux conformations est présente.

Les enzymes sont classées en six groupes en fonction du type de réaction qu'elles catalysent comme le montre le tableau 1 : (1) les oxydoréductases catalysent les oxydations et les réductions, (2) les transférases catalysent le transfert d'un groupe d'atomes d'un premier substrat vers un second substrat, (3) les hydrolases catalysent l'hydrolyse des esters, amides et autres substrats apparentés, (4) les lyases catalysent l'élimination ou l'addition de petites molécules comme l'eau, (5) les isomérases catalysent les isomérisations et (6) les ligases qui catalysent le couplage entre deux molécules, processus souvent couplé à l'hydrolyse d'adénosine-triphosphate (ATP).

Le nom systématique d'une enzyme comporte deux parties et fini par -ase. La première partie du nom permet d'identifier le substrat de l'enzyme et la seconde partie correspond à la classe d'appartenance (Tableau 1).

Tableau 1 : Classification des Enzymes[11]

Classe	Exemples de sous-classes	Fonctions
(1) Oxydoréductases	Déshydrogénases	Introduction d'une liaison double
	Oxydases	Oxydation
	Réductases	Réduction
(2) Transférases	Kinases	Transfert d'un groupe phosphate
	Transaminases	Transfert d'un groupe aminé
(3) Hydrolases	Lipases	Hydrolyse d'un ester
	Nucléases	Hydrolyse d'un phosphate
	Protéases	Hydrolyse d'un amide
(4) Lyases	Décarboxylases	Élimination du CO <sub>2</sub>
	Déshydratases	Élimination de l'eau
(5) Isomérase	Epimérase	Epimérisation d'un centre de chiralité
(6) Ligases	Carboxylases	Addition de CO <sub>2</sub>
	Synthétases	Formation d'une nouvelle liaison

De nombreuses réactions catalysées par des enzymes, et en particulier celles qui impliquent des oxydations ou des réductions, nécessitent la présence d'un coenzyme, une petite molécule organique qui prend part à la réaction, il nécessite une étape supplémentaire pour revenir à l'état initial et terminer ainsi le cycle catalytique.

## A - II. LES REACTIONS CHIMIQUES EN MILIEU SOLIDE

Une solution solide est composée d'une phase majeure (solvant) et d'une seconde phase mineure (soluté). Elle est souvent formée de cristaux isomorphes, selon les règles de Hume-Rothery (structure et valence cristallines similaires, et taille et électronégativité des métaux similaires) [14] [15]. En revanche, les cristaux non isomorphes peuvent donner naissance à un eutectique [15].

Si deux solides sont fondus ensemble, les liquides qui en résultent sont soit complètement miscibles, soit partiellement miscibles, soit non miscibles. Si les deux liquides sont partiellement ou totalement non miscibles, on peut supposer que leurs formes solides n'auront que peu d'interaction [16]. En revanche, si les deux liquides sont totalement miscibles, les

solides formés lors du refroidissement peuvent présenter l'une des quatre interactions possibles.

Peuvent se former :

- (a) un nouveau composé dont le point de fusion est congruent (conforme à ce qui est attendu ou cohérent avec d'autres caractéristiques de cette substance),
- (b) un nouveau composé dont le point de fusion (péritectique) est incongru (composition chimique de la phase liquide et de la phase solide à la température de fusion n'est pas la même),
- (c) un mélange eutectique [14],
- (d) une solution solide [14].

Dans les systèmes où la solubilité d'un composant dans l'autre à l'état solide est importante, on dit qu'il existe une solution solide. Pour des raisons pratiques, une solubilité supérieure à 5% d'un composant dans l'autre peut être considérée comme une solution solide. Les "solutions" solides sont homogènes, le terme "solution" sera dorénavant utilisé pour désigner une solution solide [17]. Deux types distincts de solutions sont reconnus. Une solution solide interstitielle existe lorsque la molécule d'un composant réside dans les espaces interstitiels du réseau cristallin du second composant. Une solution solide de substitution est formée lorsqu'une molécule d'un composant peut remplacer une molécule du réseau cristallin du second composant. Dans les deux types de solution, les deux composants peuvent ne pas être miscibles dans toutes les proportions. Comme dans les solutions solide-liquide, la solubilité d'un composant dans un autre à l'état solide est souvent limitée [18].

---

*Références Partie A*

---



- [1] F. Grellepois, A. B. Jamaa, et A. Gassama, « Diastereoselective Addition of Organomagnesium and Organolithium Reagents to Chiral Trifluoromethyl *N*-*tert*-Butanesulfinyl Hemiaminals: Diastereoselective Carbanion Additions to Chiral Hemiaminals », *Eur. J. Org. Chem.*, 2013, 6694-6701, **2013**.
- [2] F. Grellepois, « Enantiopure Trifluoromethylated  $\beta^{3,3}$ -Amino Acids: Synthesis by Asymmetric Reformatsky Reaction with Stable Analogues of Trifluoromethyl *N*-*tert*-Butanesulfinylketoimines and Incorporation into  $\alpha$  / $\beta$ -Peptides », *J. Org. Chem.*, 78, 1127-1137, **2013**.
- [3] A. Hassner et V. Alexanian, « Direct room temperature esterification of carboxylic acids », *Tetrahedron Lett.*, 19, 4475-4478, **1978**.
- [4] Y. E. Jad, G. A. Acosta, S. N. Khattab, B. G. de la Torre *et al.*, « Peptide synthesis beyond DMF: THF and ACN as excellent and friendlier alternatives », *Org. Biomol. Chem.*, 13, 2393-2398, **2015**.
- [5] K. M. Koeller et C.-H. Wong, « Enzymes for chemical synthesis », *Nature*, 09, 232-240, **2001**.
- [6] S. M. Thomas, R. DiCosimo, et V. Nagarajan, « Biocatalysis: applications and potentials for the chemical industry », *Trends Biotechnol.*, 20, 238-242, **2002**.
- [7] L. Brochier, Y. Pontie, M. Willson, S. Estrada-Mondaca *et al.*, « Involvement of Deacylation in Activation of Substrate Hydrolysis by *Drosophila* Acetylcholinesterase », *J. Biol. Chem.*, 276, 18296-18302, **2001**.
- [8] W. E. S. Filitti-Wurmser, « L'énergie de formation des complexes dissociables enzyme-substrat et antigène-anticorps », *Biochim Biophys Acta*, 4, 238-243, **1950**.
- [9] Y.-X. Li, A. J. J. Straathof, et U. Hanefeld, « Enantioselective formation of mandelonitrile acetate—investigation of a dynamic kinetic resolution », *Tetrahedron-Asymmetry*, 13, 739-743, **2002**.
- [10] T. Szegletes, W. D. Mallender, P. J. Thomas, et T. L. Rosenberry, « Substrate Binding to the Peripheral Site of Acetylcholinesterase Initiates Enzymatic Catalysis. Substrate Inhibition Arises as a Secondary Effect », *Biochemistry*, 38, 122-133, **1998**.
- [11] J. McMurry et B. Tadhg, "Chimie organique des processus biologiques", 1<sup>ère</sup>-ed., **2006**.

- [12] P. R. Kuser, S. Krauchenco, O. A. C. Antunes, et I. Polikarpov, « The High-Resolution Crystal Structure of Yeast Hexokinase PII with the Correct Primary Sequence Provides New Insights into Its Mechanism of Action », *J. Biol. Chem.*, 275, 20814-20821, **2000**.
- [13] E. Kastritis et M. A. Dimopoulos, « Thalidomide in the treatment of multiple myeloma », *Best Pract Res Clin Haematol.*, 20, 681-699, **2007**.
- [14] S. Cherukuvada et A. Nangia, « Eutectics as improved pharmaceutical materials: design, properties and characterization », *Chem. Commun.*, 50, 906-923, **2014**.
- [15] A. L. Ortiz, F. Sánchez-Bajo, F. L. Cumbreira, et F. Guiberteau, « The prolific polytypism of silicon carbide », *J Appl Crystallogr.*, 46, 242-247, **2013**.
- [16] A. Karaipekli et A. Sari, « Preparation, thermal properties and thermal reliability of eutectic mixtures of fatty acids/expanded vermiculite as novel form-stable composites for energy storage », *J Ind Eng Chem.*, 16, 767-773, **2010**.
- [17] A. H. Goldberg, M. Gibaldi, et J. L. Kanig, « Increasing Dissolution Rates and Gastrointestinal Absorption of Drugs via Solid Solutions and Eutectic Mixtures I », *J Pharm Sci.*, 54, 1145-1148, **1965**.
- [18] A. L. Ortiz, F. Sánchez-Bajo, F. L. Cumbreira, et F. Guiberteau, « The prolific polytypism of silicon carbide », *J Appl Crystallogr.*, 46, 242-247, **2013**.



---

---

*Partie B -  
Généralités sur les protéines  
et petites molécules utilisées*

---

---



## Table des matières

Partie B - Généralités sur les protéines et petites molécules utilisées .....	29
B - I.  La Lipase B de <i>Candida Antarctica</i> (CALB).....	33
B - I.1  Description de la structure de l'enzyme.....	33
B - I.2  La catalyse enzymatique .....	46
B - I.3  Sélectivité de la CALB .....	51
B - I.4  Les peptides acylés .....	53
B - I.5  Stratégie d'acylation des peptides.....	55
B - I.6  Mécanisme d'acylation enzymatique avec CALB.....	57
B - I.7  Applications de CALB.....	58
B - II  La Neupiline-1 (Nrp-1) .....	61
B - II.1  Introduction .....	61
B - II.2  Les Fonctions de la Neupiline-1 .....	65
B - II.3  Structures cristallographiques de Nrp-1 .....	68
B - III.  L'Acide Bétulinique (BTA).....	73
B - III.1  Activités thérapeutiques .....	73
B - III.2  Modifications Chimiques .....	75
B - III.3  La co-solvatation .....	75
B - III.4  La vectorisation .....	75
B - IV.  L'acide laurique .....	76
Références Partie B.....	79



## B - I. LA LIPASE B DE *CANDIDA ANTARCTICA* (CALB)

*Candida*<sup>1</sup> *antarctica*<sup>2</sup> est une levure qui produit deux lipases nommées A et B. Dans la nature, ces lipases catalysent l'hydrolyse d'esters. Classiquement, il s'agit de l'hydrolyse de triacylglycérol en di- et monoglycéride, acides gras et glycérol. Bien que présentant quelques caractéristiques fonctionnelles, la lipase A demeure moins explorée que la lipase B [1]. La lipase B de *Candida antarctica*, usuellement nommée sous la forme acronymique anglo-saxonne « CALB », est une protéine globulaire de type hydrolase d'ester carboxylique (EC 3.1.1.3, triacylglycérol-acylhydrolase). L'enzyme comporte 317 acides aminés, et a un poids moléculaire de 33 KDa. Son point isoélectrique pHi est de 5,8, calculé théoriquement à partir de la séquence primaire<sup>3</sup>. Dans les paragraphes suivants sont tour à tour présentées les structures tridimensionnelles de CALB disponibles dans la « Protein Data Bank » (PDB), obtenues par cristallographie des rayons X; le mécanisme catalytique, s'appuyant sur la triade d'acides aminés sérine-histidine-aspartate ; la sélectivité du site actif de CALB; l'importance de l'acylation ou la modification post-traductionnelle des protéines ou des peptides dans les processus biologiques et la façon dont CALB peut catalyser cette acylation ; et enfin les applications biotechnologiques de CALB.

### B - I.1 Description de la structure de l'enzyme

Une approche fréquente pour déterminer la structure d'une enzyme est la cristallographie par rayons X. La lipase de *Rhizomucor meihei* (espèce de champignons de la famille des Mucoraceae, dont on tire des enzymes qui sont utilisées pour coaguler le lait et les fromages) fut la première structure de lipase étudiée par cette méthode [2]. Elle repose sur la préparation de cristaux de la protéine à étudier. De nombreuses stratégies d'obtention de monocristaux existent actuellement mais le phénomène de cristallisation des macromolécules est encore mal compris [3]. En effet, il n'existe pas de règles pour déterminer les conditions de cristallisation d'une protéine, celles-ci doivent être déterminées expérimentalement, le plus souvent par méthode d'essais-erreurs. Cette étape cruciale franchie, la méthode de cristallographie aux rayons X permet une détermination précise de la structure

---

<sup>1</sup> Du latin *candidus*, signifiant « de couleur blanche » et faisant référence à l'aspect des colonies de levure du genre *Candida*.

<sup>2</sup> Nommée ainsi car elle fût découverte en Antarctique, dans les sédiments du lac Vanda.

<sup>3</sup> [https://web.expasy.org/cgi-bin/compute\\_pi/pi\\_tool](https://web.expasy.org/cgi-bin/compute_pi/pi_tool)

tridimensionnelle de la macromolécule étudiée, le plus souvent à résolution atomique. L'élucidation des structures 3D de protéines au niveau atomique est essentielle pour modéliser et ainsi mieux comprendre comment ces biomolécules interagissent lors d'un processus donné. En particulier, cette méthode permet une description précise du rôle d'une enzyme pour favoriser la réaction chimique de petites molécules organiques.

### **B - I.1.a. Les différentes structures cristallographiques de CALB**

La première structure de la CALB a été élucidée en 1994 avec la méthode de cristallographie aux rayons X (code PDB 1TCA)[4]. Cette structure, résolue à partir d'une forme orthorhombique, présente une résolution de 1,55 Å. L'ensemble des acides aminés de l'enzyme sont observés, tout comme une molécule de *N*-acétyl-*D*-glucosamine (NAG) provenant du milieu de cristallisation, et 286 molécules d'eau. Les structures 1TCB et 1TCC (résolutions respectives de 2,10 et 2,50 Å) déposées simultanément dans la PDB par Uppenberg *et al.* correspondent à une seconde forme cristalline de la CALB, monoclinique, l'enzyme ayant également lié une molécule du détergent octyl  $\beta$ -*D*-glucopyranoside dont la partie aliphatique est observée proche du site actif.

Peu après la structure initiale, Uppenberg *et al.* ont résolu la structure cristallographique de CALB dans deux nouvelles formes cristallines monocliniques, l'une avec le détergent méthylpenta (oxy éthyl) heptadécanoate dont une molécule est localisée dans le site actif (code PDB 1LBT, 2,5 Å de résolution), et l'autre dans laquelle l'enzyme est liée de manière covalente au niveau de la sérine 105 de la triade catalytique avec le *N*-hexylphosphonate-éthyl-ester (1LBS, 2,6 Å de résolution) [5].

Depuis lors, de nombreuses autres formes cristallines de la CALB ont été résolues, amenant le nombre de structures connues à 27. Dans la PDB, il existe à présent plusieurs structures de la CALB co-cristallisée avec différentes molécules. L'ensemble des mailles cristallographiques de la CALB sont répertoriées dans le Tableau 1 ci-après. La fixation de ligand entraîne peu de modifications structurales. Ainsi, par exemple, le code PDB 1LBS correspond à une structure de l'enzyme complexée avec un inhibiteur irréversible : le *N*-hexylphosphonate-éthyl-ester formant une liaison covalente avec la sérine 105 du site actif. Le RMSD, calculé sur les carbones  $\alpha$  entre les structures 1TCA et 1LBS est de 0,33 Å.

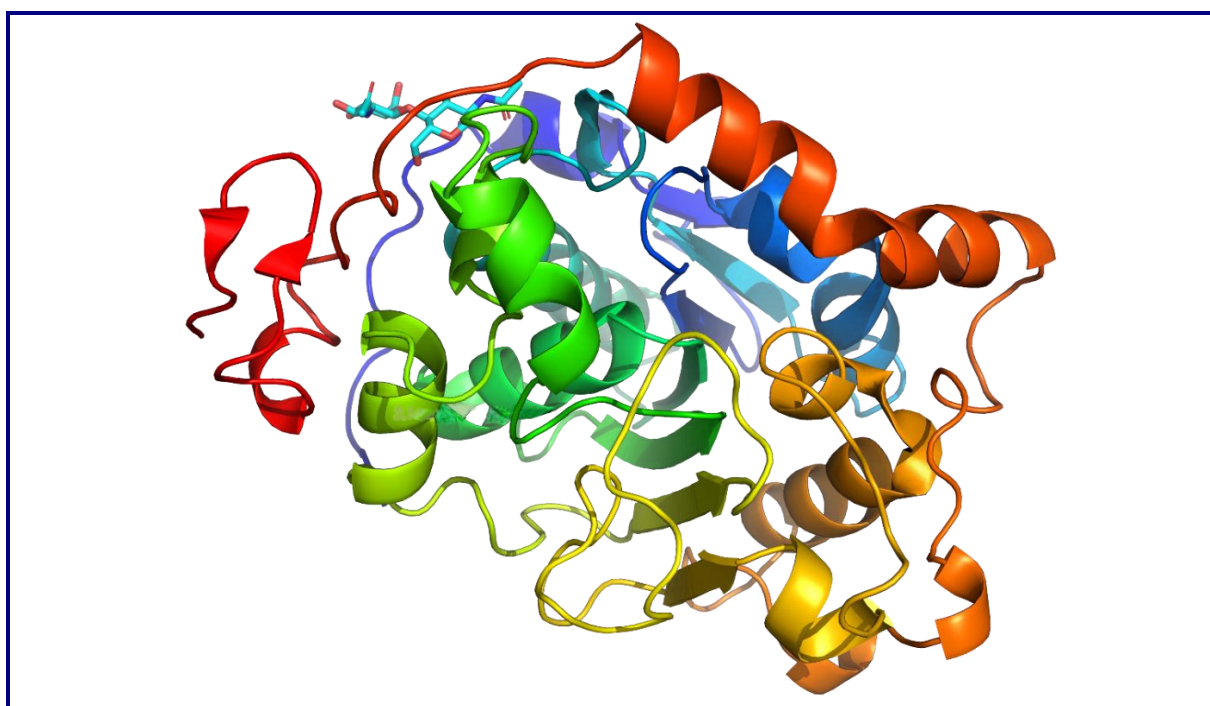
Tableau 1 : Différentes mailles de CALB.										
Paramètres de maille						Groupe d'espace	Z'	Code PDB	Ligands	Références (DOI)
a (Å)	b (Å)	c (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)					
62,1	46,7	92,1	90	90	90	$P2_12_12_1$	2	1TCA	Pas de ligand	10.1016/s0969-2126(00)00031-9 Uppenberg, J <i>et al.</i> <b>1994</b>
69,2	50,5	86,7	90	101,5	90	$P2_1$	2	1TCB	<b>BOG</b> : Octyl bêta-D-glucopyranoside	
67,0	50,5	86,7	90	100,1	90	$P2_1$	2	1TCC	<b>BOG</b> : Octyl bêta-D-glucopyranoside	
45,7	145,9	86,4	90	90	90	$C222_1$	1	4K5Q	-	10.1074/jbc.M113.536045 Xie, Y <i>et al.</i> <b>2014</b>
47,7	81,6	71,6	90	95,8	90	$P2_1$	1	4K6G	<b>EDO</b> : 1,2-Ethanol	
47,7	81,2	71,5	90	95,8	90	$P2_1$	1	4K6H	<b>EDO</b> : 1,2-éthanédiol	
46,7	87,1	138,8	90	90	90	$P2_12_12_1$	1	4K6K	<b>EDO</b> : 1,2-éthanédiol	
46,9	80,4	73,1	90	98,4	90	$P2_1$	1	6J1S	<b>PEG</b> : Ether di(hydroxyéthyle) <b>SO<sub>4</sub></b> : Ion Sulfate <b>EDO</b> : 1,2-éthanédiol	10.1021/jacs.9b02709 Xu, J <i>et al.</i> <b>2019</b>
46,4	80,1	72,9	90	97,9	90	$P2_1$	1	6J1P	<b>PGE</b> : Triéthylène glycol <b>SO<sub>4</sub></b> : Ion sulfate <b>EDO</b> : 1,2-éthanédiol <b>ACT</b> : Ion acétate <b>CL</b> : Ion chlorure	
46,6	80,3	73,2	90	98,2	90	$P2_1$	1	6J1T	<b>B7U</b> : (2s)-2-phényl-n-[(1r)-1-phényléthyl] propanamide <b>PGE</b> : Glycérol triéthylène <b>PEG</b> : Ether di (hydroxyéthyle) <b>SO<sub>4</sub></b> : Ion sulfate <b>EDO</b> : 1,2-éthanédiol <b>CL</b> : Ion Chlorure	
47,1	80,6	73,6	90	98,4	90	$P2_1$	1	6J1R	<b>PEG</b> : Ether di (hydroxyéthyle) <b>SO<sub>4</sub></b> : Ion sulfate <b>EDO</b> : 1,2-éthanédiol	
46,5	80,4	72,5	90	97,5	90	$P2_1$	1	6J1Q	<b>PEE</b> : Ethanesulfonique 4-(2-hydroxyéthyl) -1-piperazine acide <b>1PE</b> : Glycérol pentaéthylène <b>PEG</b> : Ether di (hydroxyéthyle) <b>SO<sub>4</sub></b> : Ion sulfate <b>EDO</b> : 1,2-éthanédiol <b>CL</b> : Ion chlorure	

95,1	50,2	99,5	90	90,6	90	$P_{2,1}$	2	1LBT	<b>T80</b> : heptadécanoate de méthylpenta (oxyéthyl)	10.1021/bi00051a035 Uppenberg, J <i>et al.</i> <b>1995</b>
229,5	95,6	86,8	90	90	90	$C_{2,1}$	2	1LBS	<b>HEE</b> : N-hexylphosphonate ethyl ester	
229,5	95,6	86,8	90	90	90	$C_{2,1}$	2	1LBS	<b>HEE</b> : N-hexylphosphonate ethyl ester	
46,9	155,9	92,5	90	90	90	$P_{2,1}$	1	6ISP	<b>CPQ</b> :N,N-bis (3-d-gluconamidopropyl) deoxycholamide <b>CA</b> : Ion Calcium	10.1038 / s41467-019-11155-3 Cen, Y <i>et al.</i> <b>2019</b>
44,6	132,3	52,4	90	90,1	90	$P_{2,1}$	1	6ISQ	<b>PGE</b> : Glycérol triéthylène <b>EDO</b> : 1,2-éthanédiol <b>IPA</b> : alcool isopropylique <b>ACT</b> : Ion acétate	
52,2	44,5	132,9	90	89,4	90	$P_{2,1}$	1	6ISR	<b>PG4</b> : Glycérol tetraethylene <b>PEG</b> : Ether di (hydroxyéthyle) <b>NI</b> : Ion Nickel (II)	
111,6	111,6	54,9	90	90	120	$P_{3,2,2,1}$	1	3ICV	<b>NAG</b> : 2-acétamido-2-désoxy-bêta-d-glucopyranose <b>BTB</b> : 2- [bis- (2-hydroxy-éthyl) -amino] -2-hydroxyméthyl- propane-1,3-diol	10.1016 / j.jmb.2009.08.008 Qian, Z <i>et al.</i> <b>2009</b>
111,7	111,7	54,8	90	90	210	$P_{3,2,1}$	1	3ICW	<b>NAG</b> : 2-acétamido-2-désoxy-bêta-d-glucopyranose <b>MHH</b> : hydrogène de méthyle (r) -hexylphosphonate <b>PO<sub>4</sub></b> : Ion phosphate	
39,6	48,9	71,6	88,7	97,1	108,4	$P_1$	2	5A6V	<b>XE</b> : Xénon <b>IPA</b> : alcool isopropylique	10.1194 / jlr.M063388 Stauch, B <i>et al.</i> <b>2015</b>
89,4	156,6	138,1	90	90	90	$C_{2,2,1}$	2	6TP8	<b>NTK</b> : Butanoate de 2,3-di (butanoyloxy)propyle <b>DEP</b> : Phosphonate de diéthyle	10.1016 / j.ijbiomac.2020.04.061 Silvestrini, L <i>et al.</i> <b>2020</b>
89,1	89,1	137,2	90	90	120	$P_{6,3,2,2}$	2	4ZV7	Pas de ligand	10.18388 / abp.2015_1065 Strzelczyk, P., <i>et al.</i> <b>2016</b>
123,6	123,6	191,8	90	90	120	$P_{6,5}$	1	3W9B	<b>PE8</b> : 3,6,9,12,15,18,21-heptaotricosane-1,23-diol	10.1021/bm049259q De Diego, T <i>et al.</i> <b>2005</b>
92,4	123,2	150,3	90	96,2	90	$P_{2,1}$	2	5GV5	<b>MSW</b> : acide [(1s) -2- (méthoxycarbonylamino) -1-phényl-éthoxy] - propyl-phosphinique	10.1021 / accatal.6b02192 Park, K <i>et al.</i> <b>2016</b>
91,5	91,5	299,8	90	90	90	$P_{4,3,2,1,2}$	1	2VEO	<b>IUM</b> : Ion uranyl (VI) <b>PG4</b> :Glycérol tetraethylene <b>GOL</b> : glycérol	10.1016 / j.jmb.2007.10.079 Ericsson, DJ <i>et al.</i> <b>2008</b>
39,6	48,9	71,6	88,7	97,1	108,4	$P_1$	2	5A71	<b>IPA</b> : alcool isopropylique <b>K</b> : Ion potassium <b>NA</b> : ion de sodium	10.1194 / jlr.M063388 Stauch, B <i>et al.</i> <b>2015</b>



**B - I.1.b. Le repliement global de la protéine CALB.**

La lipase B de *Candida antarctica* est comme toutes les lipases ; une hydrolase  $\alpha/\beta$  constituée d'une région centrale hydrophobe composée, dans le cas de la CALB, d'un feuillet de sept brins  $\beta$  dont six sont parallèles. Le feuillet est entouré de dix hélices  $\alpha$  disposées sur ses deux faces, et la chaîne polypeptidique se termine par 2 autres brins  $\beta$  adoptant une conformation en épingle à cheveux (en anglais,  $\beta$ -hairpin) (Figure 1). La Figure 2 montre comment ces éléments de structure secondaire s'enchaînent dans la séquence des acides aminés. La Figure 3 montre l'organisation topologique de ces éléments de structure secondaire, agencés autour du feuillet  $\beta$  bien visible sur la Figure 1.



**Figure 1** Structure tridimensionnelle de la CALB (Code PDB : 1TCA).

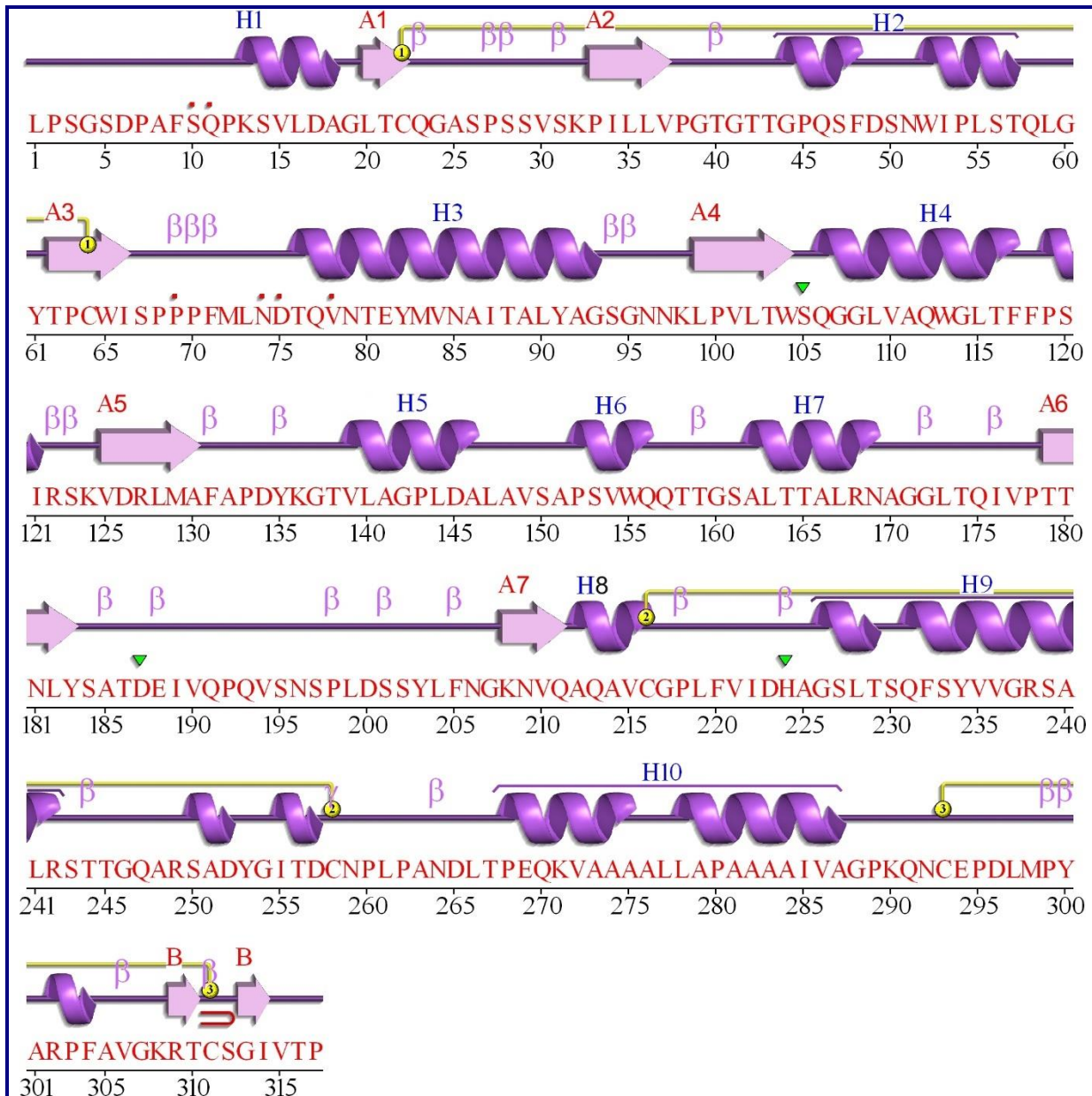
Représentation en mode « cartoon » dans lequel les hélices  $\alpha$  sont figurées par des rubans en spirale et les brins  $\beta$  par des flèches. La structure est colorée en arc-en-ciel du bleu foncé pour l'extrémité N-terminale au rouge à l'extrémité C-terminale. La molécule de N-acétyl-D-glucosamine (NAG) provenant du milieu de cristallisation est représentée en mode bâton.

Le feuillet  $\beta$  central apparaît verticalement, seul son brin 1 tout en haut est dirigé vers le fond, tous les brins suivants lui étant antiparallèles sont dirigés vers l'avant de la figure.

Figure réalisée avec PyMOL.

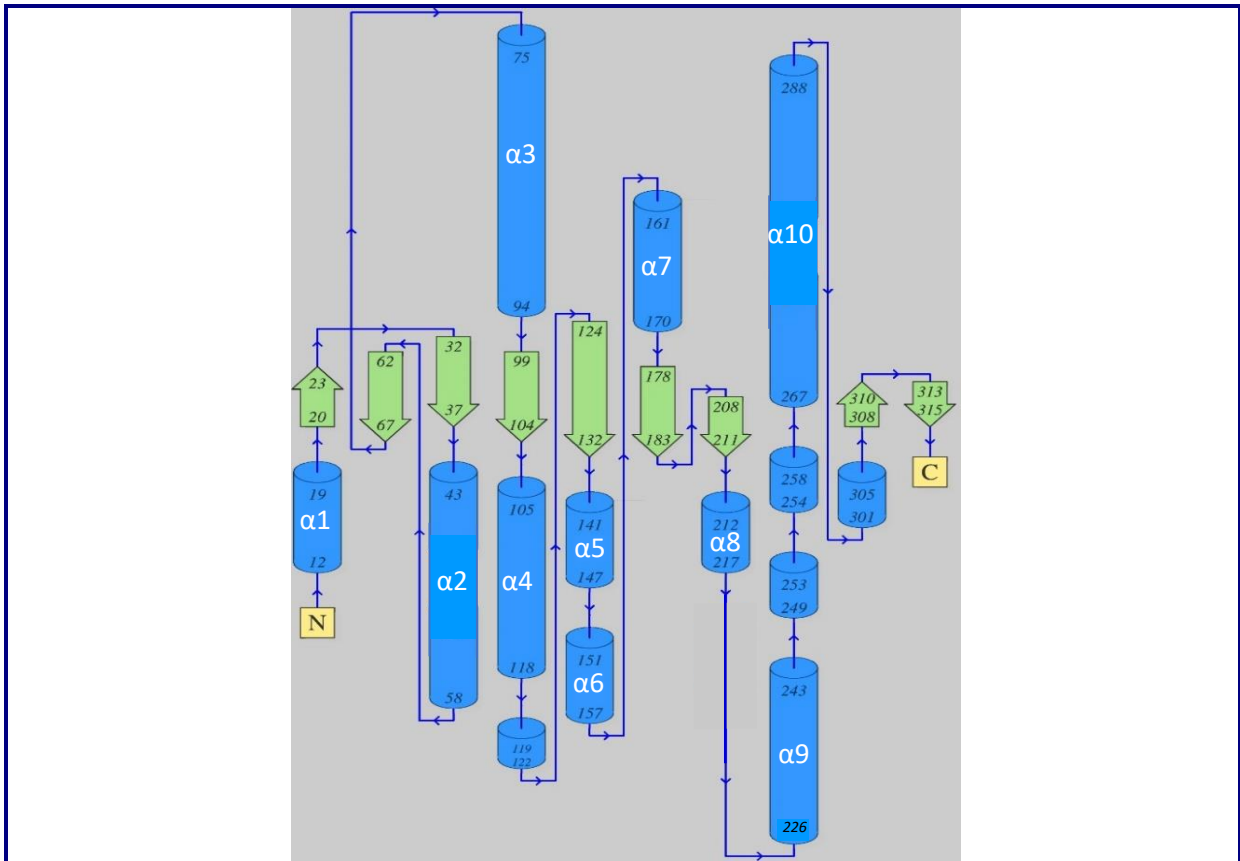
Le diagramme de Ramachandran [6] représenté à la Figure 4 permet d'analyser la conformation ou la distribution des couples d'angles  $\varphi$  et  $\psi$  au sein de la chaîne principale de la CALB de code PDB 1TCA, mettant en évidence les brins  $\beta$  parallèles ( $\varphi = -120^\circ$  et  $\psi = +115^\circ$ ) et les hélices  $\alpha$  ( $\varphi = -60^\circ$  et  $\psi = -50^\circ$ ). La structure de CALB contient 6 cystéines qui

forment les ponts disulfures suivants : Cys22- Cys64, Cys216-Cys258 et Cys293- Cys311. Ces ponts disulfures contribuent à la stabilité de la structure tridimensionnelle de la protéine.



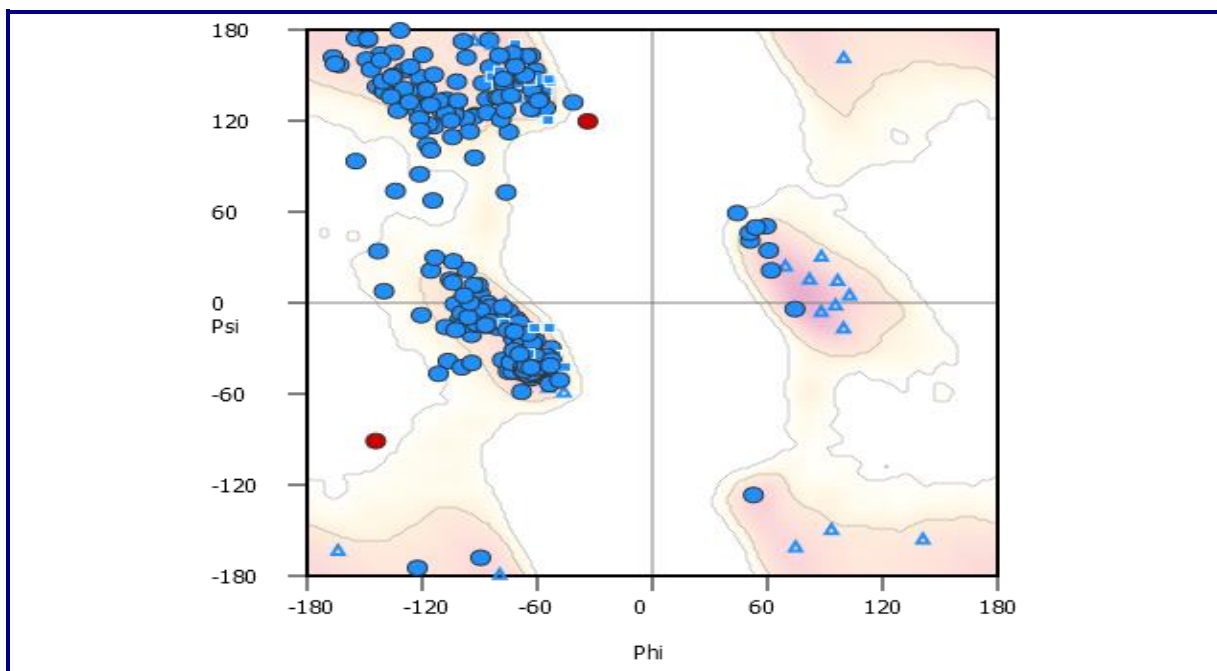
**Figure 2** Éléments de structure secondaire et séquence de la protéine CALB.

Les brins  $\beta$  sont représentés par des flèches numérotées de A1 à A7 pour le feuillet principal et B pour l'épingle à cheveux C-terminale, les hélices  $\alpha$  par des spirales violettes numérotées de H1 à H10. Certaines hélices sont trop courtes pour être prises en considération. Les traits jaunes montrent les ponts disulfures. Les petits triangles verts indiquent les acides aminés impliqués dans la catalyse (voir paragraphe B - 1.2 page 46). Figure adaptée du site <https://www.ebi.ac.uk/pdbsum/> pour le code PDB 1TCA.



**Figure 3** Topologie de la structure de la CALB (code PDB 1TCA).

Les brins  $\beta$  sont représentés par des flèches vertes, les hélices  $\alpha$  par des cylindres bleus. Certaines hélices sont trop courtes pour être prise en considération. La longueur des connexions entre ces éléments ne reflète pas la réalité. Figure tirée du site <https://www.ebi.ac.uk/pdbsum/> pour le code PDB 1TCA.



**Figure 4** Diagramme de Ramachandran (angles  $\phi$  et  $\psi$ ) de la CALB (code PDB 1TCA).

**B - I.1.c. Flexibilité de la structure**

Les protéines sont flexibles. Les changements de conformation peuvent être subtils, reflétant les vibrations moléculaires et les petits mouvements des résidus d'acides aminés dans la protéine. On dit parfois d'une protéine qu'elle "respire". Le facteur d'agitation thermique (aussi appelé facteur *B*) associé à chaque atome constitue un paramètre permettant en première approximation d'apprécier le mouvement des atomes au sein de la structure cristallographique. Il reflète en effet la dispersion des positions de chaque atome dans les différentes copies de la protéine observées dans le cristal. Un exemple de ces facteurs d'agitation thermique des atomes de carbone alpha de la structure de la CALB (code PDB : 1TCA) est présenté sur la Figure 5

constitue un paramètre permettant en première approximation d'apprécier le mouvement des atomes au sein de la structure cristallographique. On observe que les résidus du feuillet  $\beta$  sont moins mobiles tandis que les hélices  $\alpha$  et les extrémités de la séquence sont davantage mobiles.

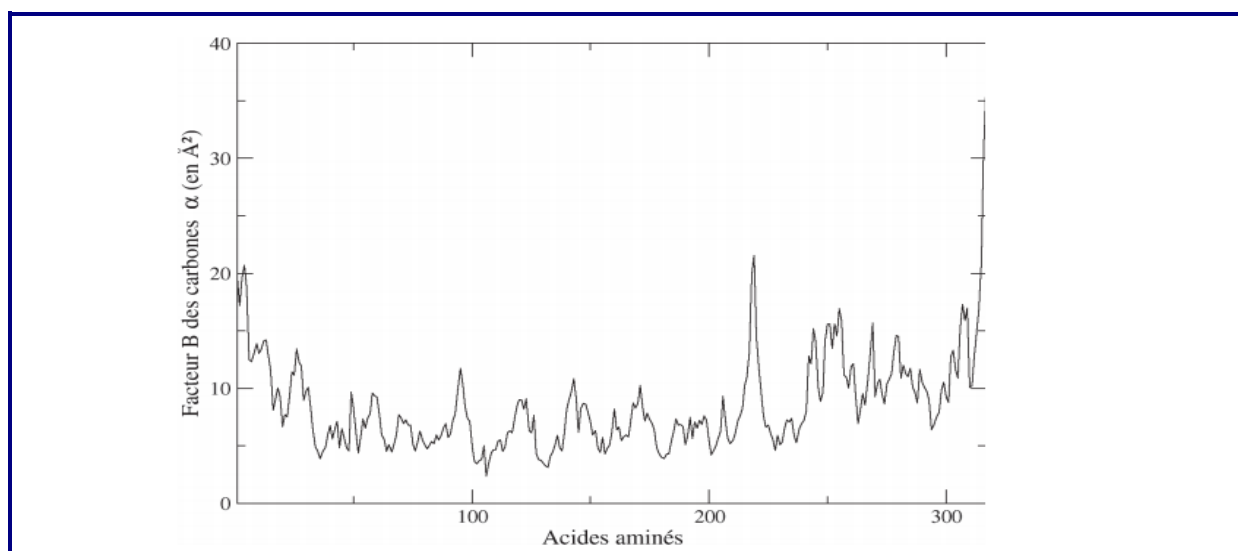


Figure 5 Facteur B de la structure cristallographique (code PDB : 1TCA) de la CALB (Tirée de CHAPUT 2012).

Il est important de noter que dans plusieurs structures, l'hélice  $\alpha 5$  possède une conformation dépliée en boucle (code PDB 1TCC [4]), parfois trop désordonnée pour être observée (code PDB 4K6G). La mobilité d' $\alpha 5$  pourrait lui permettre de recouvrir partiellement la cavité du site actif [7]. D'autre part, une étude a montré que des mutations des séquences autour de l'hélice  $\alpha 5$  avaient un impact significatif sur l'activité et l'énantiosélectivité de l'enzyme [8]. Il a été observé par dynamique moléculaire que l'hélice se déroulait et que la boucle résultante se déplaçait vers l'hélice  $\alpha 10$ , bloquant ainsi l'accès au site actif [8]. Toutefois, aucune donnée expérimentale n'a confirmé ce modèle.

Les hélices  $\alpha 5$  et  $\alpha 10$  semblent d'une importance capitale pour expliquer les propriétés catalytiques de CALB.

La majorité des lipases sont caractérisées par le phénomène d'activation interfaciale [9] : lorsque la concentration micellaire critique<sup>4</sup> est atteinte, les lipases situées à l'interface de la phase eau/lipide voient leur activité d'hydrolyse des acylglycérols fortement augmenter. Ce mécanisme d'activation interfaciale est attribué à l'existence d'une boucle amphiphile venant couvrir le site actif en position fermée dans un milieu aqueux, et s'ouvrant lorsque la lipase est à l'interface eau/lipide. Cette boucle serait formée d'une ou deux hélices  $\alpha$  flexibles. Bien qu' $\alpha 5$  constitue une candidate idéale pour ce rôle du fait de sa proximité avec le site actif, des études ont montré que l'on ne retrouve pas ce phénomène chez CALB [10] [11]. Les travaux publiés par l'équipe de Uppenberg et collaborateurs sur la CALB ont montré l'absence de volet amphiphile [4] [12]. La lipase A de *Candida Antarctica* présente une activation interfaciale, mais de façon moins prononcée que la lipase de *Humicola lanuginosa* [11].

#### **B - I.1.d. La cavité comportant le site actif de la CALB**

Le site actif d'une lipase est une région spatiale tapissée de groupes fonctionnels, appelée encore poche enzymatique.

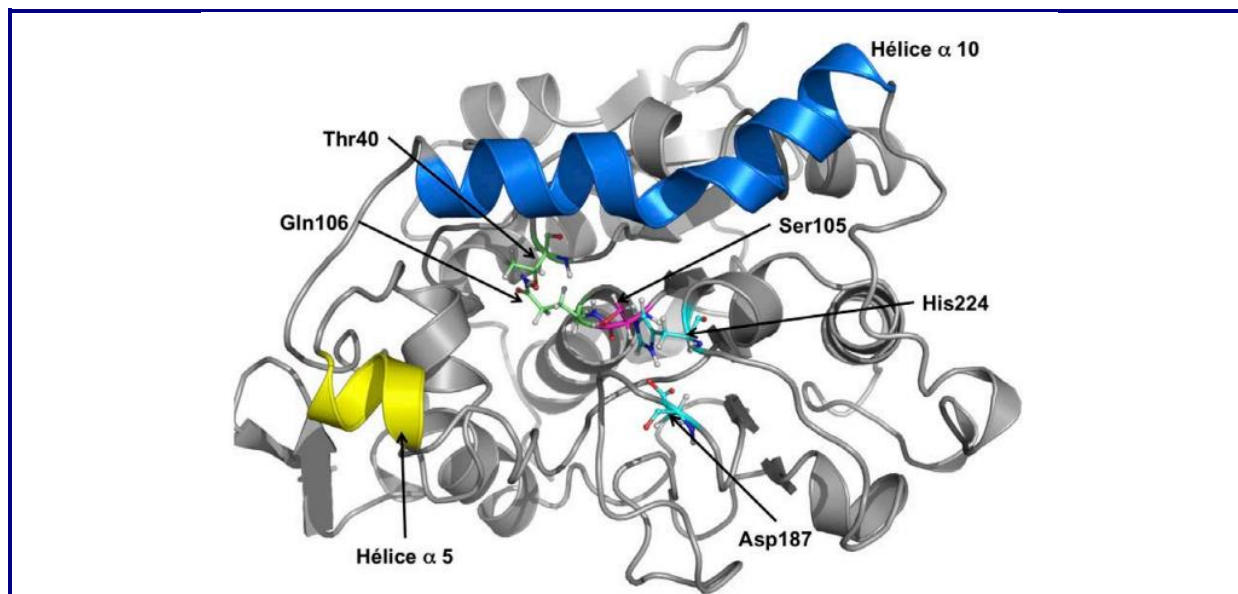
Pour le cas de la CALB, la recherche de cavités à partir des données cristallographiques trouvées dans la littérature a permis de définir une région étroite et profonde [4] (voir ci-dessous). Le site actif de l'enzyme est localisé au fond de cette cavité. Il est composé de deux ensembles (Figure 6): 1) une triade catalytique comprenant les résidus Ser105, His224, Asp187, localisée dans le fond de la cavité; 2) un groupe de 2 résidus constituant le trou oxy-anion<sup>5</sup> (Thr40, Gln106). Ces derniers interviennent dans le mécanisme de catalyse de la CALB en orientant correctement le groupement fonctionnel de l'un des substrats, par la création de 2 ou 3 liaisons hydrogènes. En outre, la cavité de cette enzyme est essentiellement hydrophobe, tapissée par des résidus aliphatiques Leu140, Ala141, Leu144, Val149, Ser150, Ala151, Val154, Ile189, Lys290, Leu278, Ala281, Ala282, Ile285 et Val286

---

<sup>4</sup> Concentration en lipides à partir de laquelle la formation de micelle est spontanée.

<sup>5</sup> Le trou oxyanion est une structure clé dans le site actif des enzymes, qui joue un rôle crucial dans la catalyse des réactions chimiques en stabilisant les charges négatives ou les intermédiaires réactifs formés pendant la réaction enzymatique.

[13] (Figure 7). Seule une petite région autour des résidus catalytiques possède un caractère hydrophile, dû à la présence des résidus Thr40, Asp134 et Gln157.



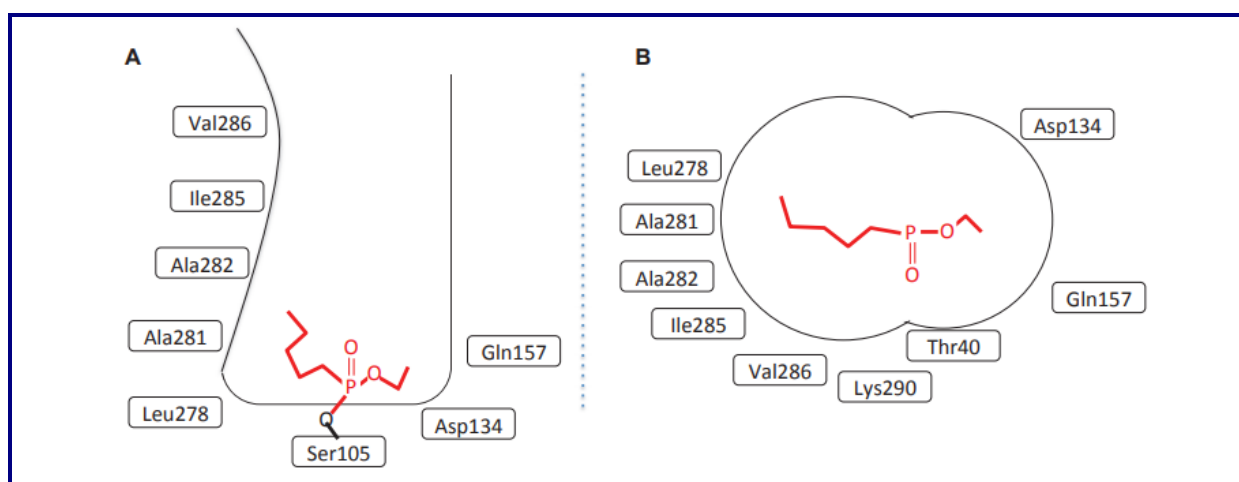
**Figure 6** Les principaux éléments du site actif de la CALB (Code PDB : 1LBS).

Les résidus du trou oxyanionique (Thr40 et Gln106) sont représentés en bâtons de couleur verte. Les résidus de la triade catalytique sont indiqués en bâtons de couleur cyan (His224 et Asp187) et de couleur rose (Ser105). Les deux hélices  $\alpha$  (hélice  $\alpha 5$  et hélice  $\alpha 10$ ) bordant l'entrée du site actif sont également mises en évidence. Pour plus de clarté, le ligand n'est pas représenté.

Dans les structures de la CALB, le site actif est accessible au solvant externe par un canal étroit. Ce canal a une largeur d'environ  $10 \text{ \AA} \times 4 \text{ \AA}$  et une profondeur de  $12 \text{ \AA}$ , mesurée depuis l'Oxygène  $\gamma$  ( $O_\gamma$ ) de la Ser105 jusqu'à la surface de celle-ci. La majeure partie du canal est formée par trois parties de la structure : les hélices  $\alpha 5$  et  $\alpha 10$  et une région en boucle qui projette Ile189 dans le canal [4]. Les parois du canal sont très hydrophobes car elles sont garnies de résidus principalement aliphatiques. Aucune chaîne latérale aromatique n'est présente dans le canal, sauf Trp104, qui précède la sérine catalytique dans la séquence. L'azote de la chaîne latérale de ce résidu établit une liaison hydrogène avec l'oxygène du groupement carbonyle de la chaîne principale de l'histidine (His224) de la triade catalytique dans le site actif et stabilise cette région. Dans d'autres lipases, ce résidu tryptophane est souvent remplacé par une histidine faisant une liaison hydrogène similaire avec l'histidine de la triade catalytique. L'accessibilité du site actif observé dans les différentes formes cristallines ne permet pas d'être certains de la partie de la protéine, le cas échéant, qui fonctionne comme un couvercle pour contrôler l'entrée dans le site actif. Le candidat le plus probable est la courte hélice  $\alpha 5$ , parfois observée dépliée, ce qui suggère une région de grande mobilité qui peut subir des changements de conformation importants pour la liaison aux lipides et la catalyse. Dans la forme ouverte de l'enzyme, une



surface de résidus aliphatiques de l'hélice  $\alpha 5$  qui tapisse le canal menant au site actif a été observée [4]. La caractéristique la plus frappante de cette région est une chaîne latérale d'acide aspartique enfouie (Asp145), qui établit des liaisons hydrogène stabilisantes avec les chaînes latérales de Ser150 et Thr158. La longue hélice à l'extrémité C-terminale de la structure, l'hélice  $\alpha 10$ , est un autre candidat possible pour modifier l'accessibilité du site actif. Cette hélice est dominée par des alanines et d'autres résidus hydrophobes sur tous les côtés et est coudée au milieu par un résidu de proline. Elle n'a pas de liaisons hydrogène avec le reste de la protéine, ce qui suggère qu'elle peut être relativement mobile. Le site actif présente également des facteurs de température de la chaîne principale plus élevés que le reste de la structure. Un changement de conformation de l'hélice  $\alpha 10$  pourrait modifier la taille et la forme du canal du site actif et de la surface environnante de l'enzyme [7].



**Figure 7** Schéma de la poche catalytique de CALB contenant l'inhibiteur compétitif N-ester hexylphosphonate d'éthyle (en rouge) (réalisé à partir de la structure de code PDB 1LBS) (Tirée de FERRARI 2014).

A : Vue de profil. B : Vue de dessus.

Pour compléter la description du site actif, notons une étude de Fuentes (2004)[14] s'intéressant à la plasticité de la CALB et à son rôle dans le mécanisme de catalyse, analysant la distribution des atomes dans l'espace du site actif et le changement de géométrie de la poche catalytique suite à l'insertion d'un sucre et d'une chaîne grasse (C12 : acide laurique). Les résultats de cette étude suggèrent qu'il serait plus juste de considérer la cavité catalytique de CALB comme une unique poche capable de changer sa conformation en fonction des substrats par des réarrangements locaux [15].

### B - I.1.e. Le site actif

Comme c'est généralement le cas pour les lipases (EC 3.1.1.3) et estérases (EC 3.1.1.1), la CALB est une protéase à sérine. Le résidu nucléophile (Ser105) est localisé dans le

coude situé entre le feuillet  $\beta 4$  et l'hélice  $\alpha 4$ . Cette localisation du coude nucléophile est caractéristique des hydrolases  $\alpha/\beta$ .

En plus des acides aminés de la triade catalytique (Ser105, His224, et Asp187), on trouve le Trp104 situé au fond du site actif. Les résidus Thr40 et Gln106 forment le trou oxyanion stabilisant l'intermédiaire réactionnel [16]. On trouve également dans le fond du site actif Ser47, Thr42, et Gly39 limitant la taille d'une poche du site actif appelé « poche stéréospécifique ». L'entrée du site actif est bordée sur un côté par les résidus Leu278, Ala282 et Ile285 faisant partie de l'hélice  $\alpha 10$  (cf. Figure 8 et Figure 9). Sur l'autre bord, c'est le résidu His224 qui limite essentiellement l'accès au site actif.

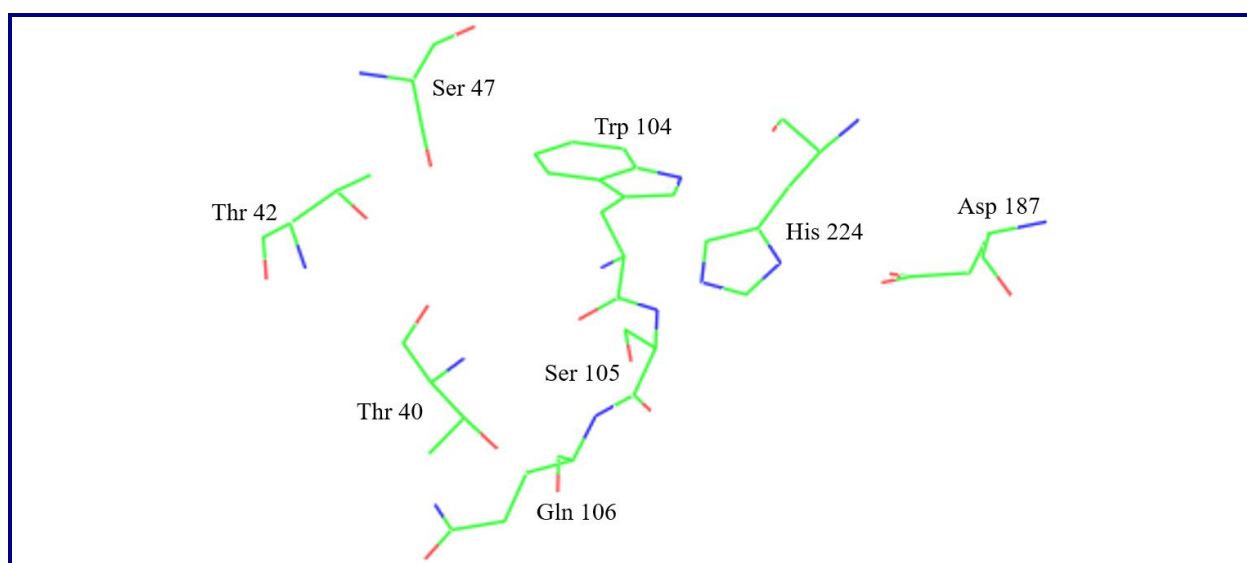


Figure 8 Triade catalytique (S105, H224, D187) de la CALB et acides aminés importants du site actif.

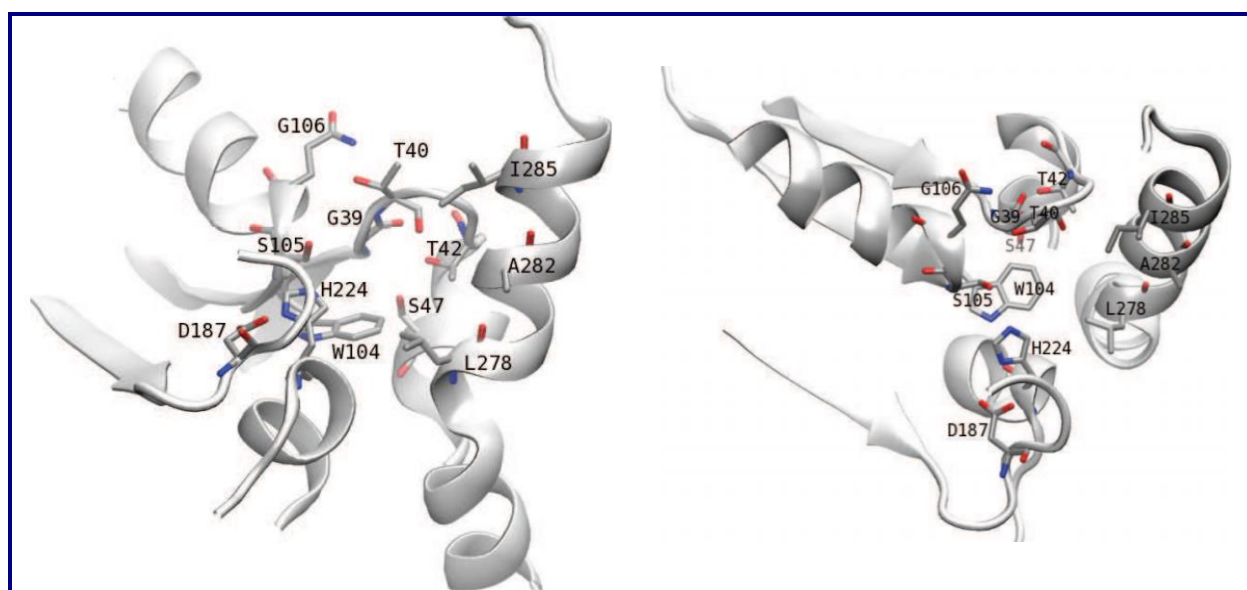


Figure 9 Deux vues du site actif de la CALB.



**B - I.1.f. Comparaison du site actif de la CALB avec celui d'autres enzymes**

A l'instar de CALB, un certain nombre d'enzymes catalysent l'hydrolyse d'acides gras, notamment deux sérines estérases et huit lipases [17]. Leurs architectures et mécanismes d'hydrolyse des esters sont similaires. La géométrie de leurs sites de liaison diffère. Les lipases ont un site de catalyse large et hydrophobe tandis que les estérases comme l'acétylcholinestérase et la bromoperoxydase ont une petite poche pour la partie acyle et cette poche s'ajuste exactement avec leurs substrats naturels. Les lipases et les estérases peuvent être divisées en trois groupes :

- Les lipases telles que celle de *Rhizomucor miehei* qui présente un site actif hydrophobe en crevasse localisé à la surface de la protéine.
- La lipase B de *Candida antarctica*, celle de *Pseudomonas cepacia*, ainsi que celle du pancréas de mammifère et la cutinase de *Fusarium solani* présentent un accès au site actif en entonnoir.
- Enfin, l'accès au site actif de la lipase de *Candida rugosa* forme un long tunnel de 22Å. L'allure des sites actifs de ces différents enzymes est présentée sur les Figure 10 et Figure 11.

Pour toutes les estérases à sérine et lipases, le site de liaison est situé dans une poche à l'extrémité d'un feuillet  $\beta$  central. Chez les lipases, la surface de l'enzyme au bord de la poche est hydrophobe permettant une bonne interaction avec les substrats.

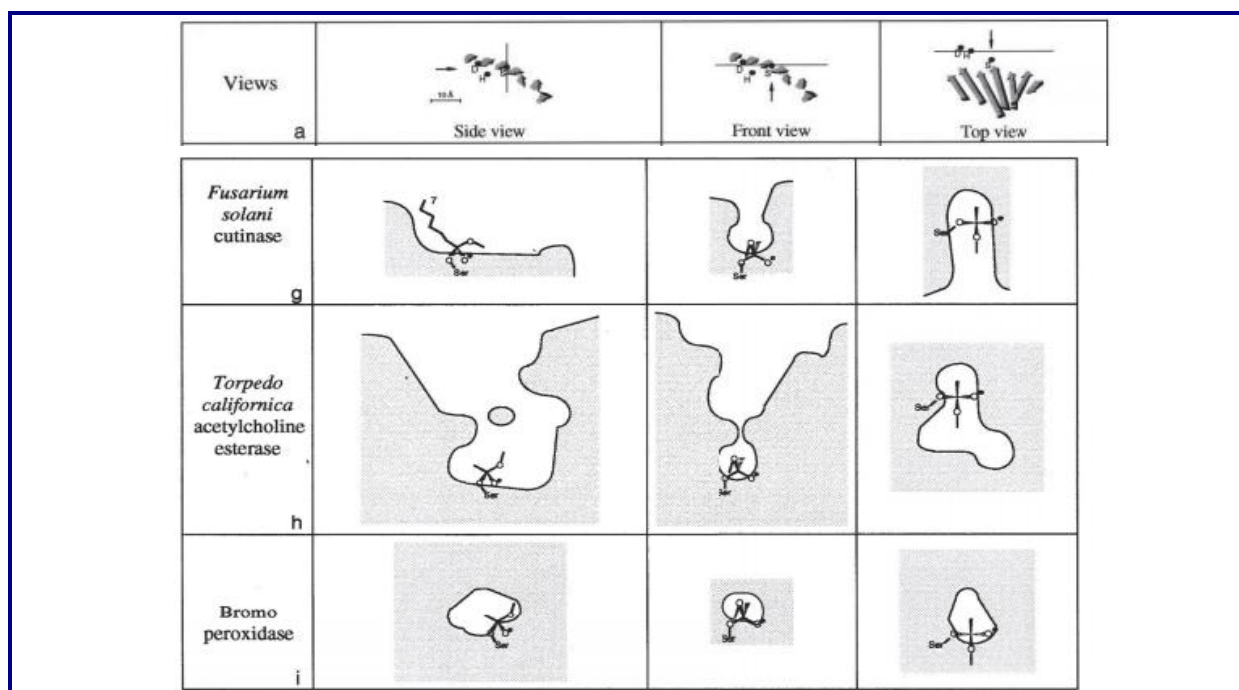


Figure 10 Sites catalytiques de deux estérases et une peroxydase [17], vus sous 3 orientations perpendiculaires.

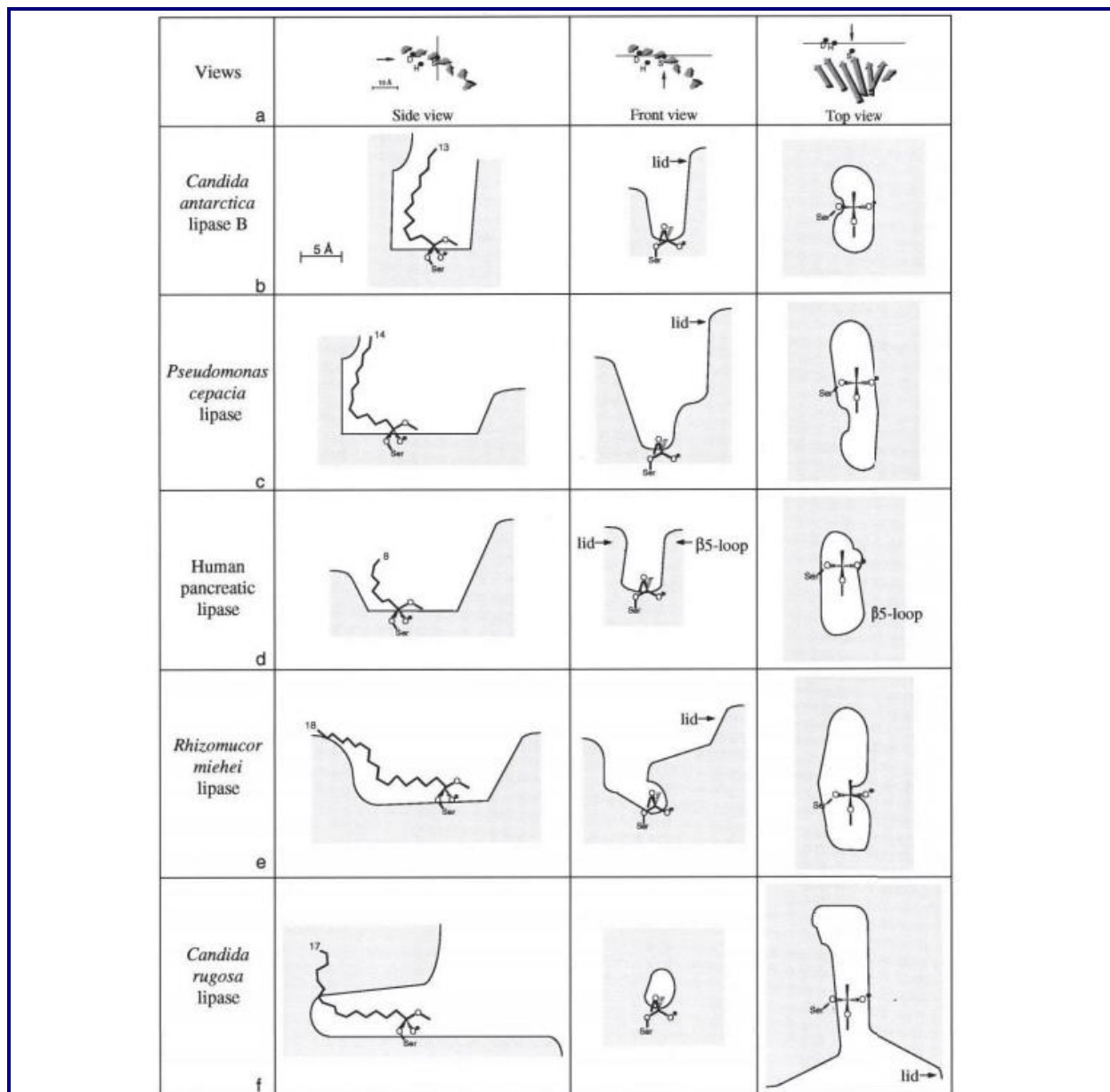


Figure 11 Sites catalytiques de cinq différentes lipases (Tirée de CHAPUT 2012) [17], vus sous 3 orientations perpendiculaires.

## B - I.2 La catalyse enzymatique

Les enzymes sont des protéines capables de catalyser les réactions chimiques et leur effet catalytique est beaucoup plus performant que les systèmes synthétiques. La partie de l'enzyme responsable de l'activité catalytique, le site actif, lie généralement le substrat de manière non-covalente.

L'environnement chiral de ce site permet alors d'avoir une chimio-, régio et stéréosélectivité au cours de la catalyse. Les sites actifs des enzymes contiennent souvent des

composés organiques ou inorganiques autres que des acides aminés, tels que des mono- ou dinucléotides (Nicotinamide-Adénine-Dinucléotide Phosphate...), des vitamines (biotine, vitamine B12, etc.), des cations divalents ( $\text{Fe}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Zn}^{2+}$  ou  $\text{Mg}^{2+}$ ), qui sont essentiels pour l'activité catalytique et qu'on appelle cofacteurs ou coenzymes [18]. Plus de la moitié des catalyseurs enzymatiques sont des métalloenzymes. Les enzymes catalysent des réactions nombreuses et diverses, par exemple des réactions d'hydrolyse (hydrolases), l'addition de groupements fonctionnels (lyases), le transfert de groupes dans une même molécule (isomérases) et la formation de liaisons C-C, C-S, C-O et C-N (ligases), etc.

En chimie, l'application la plus fréquente des enzymes est celle de la résolution cinétique d'un mélange racémique, comme par exemple pour la production d'amines chirales (Schéma 1) [19] [20].

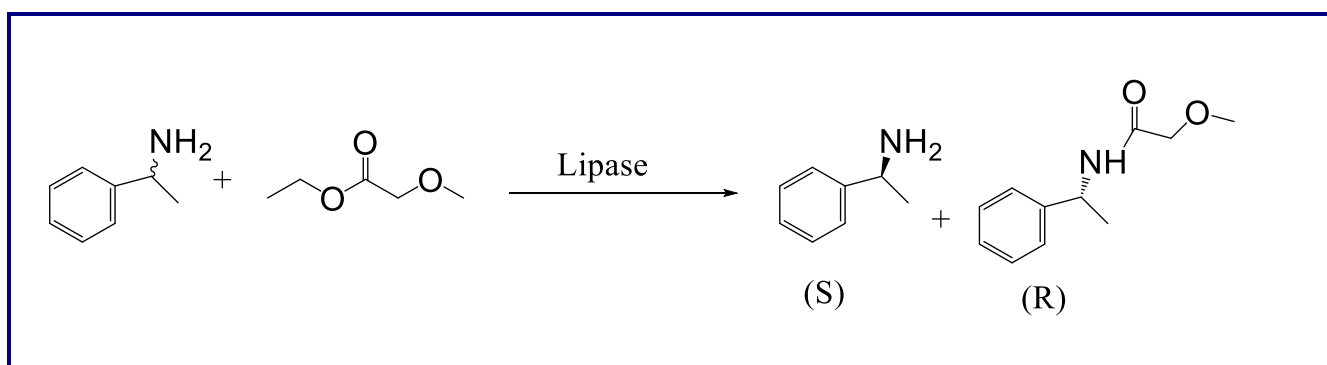


Schéma 1 Résolution cinétique d'une amine racémique à l'aide d'une lipase.

Plusieurs enzymes sont commercialisées pour des transformations énantiosélectives. Cependant, les inconvénients dus à la spécificité du substrat, la stabilité et la formation d'une seule forme énantiomérique sont des facteurs limitants [21]. L'optimisation génétique combinée au criblage à haut débit ont permis de manière significative de mettre en évidence de nouvelles enzymes non naturelles appelées « mutants » (correspondant à des variants de la séquence naturelle de l'enzyme) beaucoup plus efficaces [22]. Il existe plusieurs exemples où l'optimisation génétique apporte une amélioration de l'énantiosélectivité.

Dans le cas de la résolution cinétique d'un ester racémique catalysée par une lipase de *Pseudomonas aeruginosa* (Schéma 2), Reetz a montré qu'à partir de la protéine sauvage dont le facteur de sélectivité  $E$  est de 1.1 en faveur de la forme (S), il est possible après plusieurs cycles de mutations d'améliorer la sélectivité de cette enzyme et d'obtenir deux mutants dont les facteurs de sélectivité  $E$  sont pour l'un supérieur à 51 en faveur de la forme (S) et pour l'autre égal à 30 en faveur de la forme (R) [23] [22].

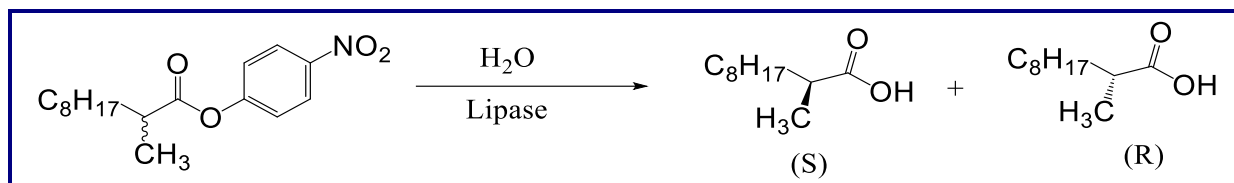


Schéma 2 Résolution cinétique d'un ester racémique par une lipase.

La lipase B de *Candida antarctica* peut être employée pour catalyser des réactions d'hydrolyse, d'estérification, et de transestérification, mais également de thio transestérification, d'aminolyse, d'acylation, de transacylation et d'amidation [24].

### B - I.2.a. Le rôle natif de la CALB

Dans la nature, CALB catalyse des réactions citées ci-avant. Comme de nombreuses lipases, elle assure préférentiellement l'hydrolyse des chaînes courtes en position 1 et 3 des triacylglycérides d'acides gras (Schéma 3). Des travaux indiquent que l'hydrolyse en position 2 est également possible mais beaucoup moins efficace, à l'inverse des enzymes à régiosélectivité stricte tels que la lipase de *Rhizomucor mihei* [25].

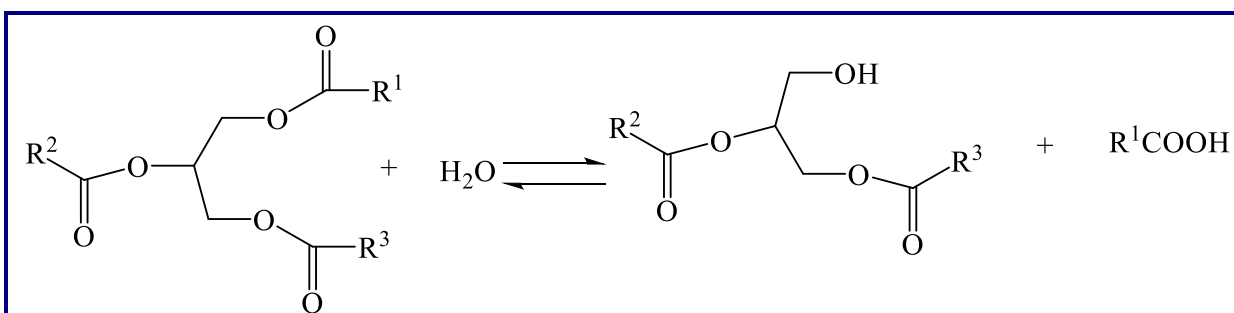


Schéma 3 Hydrolyse d'un triacylglycéride en position 1.

CALB possède une faible activité pour l'hydrolyse des triglycérides à chaînes longues.

### B - I.2.b. Le mécanisme réactionnel de transestérification

La réaction de transestérification impliquant un ester et un alcool secondaire est présentée dans le Schéma 4. Son mécanisme détaillé, de type Bi-Bi Ping-Pong, est présenté dans le Schéma 5 [26] [26].

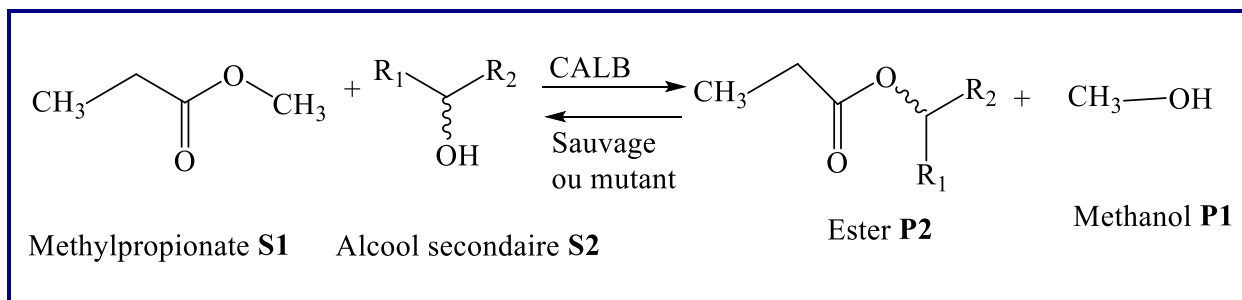


Schéma 4 La réaction de transestérification des alcools secondaires.

Tout d'abord, Ser105 est déprotonée par l'azote N $\epsilon$  de His224, la charge positive du groupe imidazole de l'histidine étant alors stabilisée par le partage d'un proton entre l'azote N $\delta$  de His224 et la fonction carboxylate d'Asp187. L'ester substrat est attaqué par la fonction alcoolate (attaque nucléophile) de Ser105. Puis s'ensuit la formation d'un intermédiaire tétraédrique (TS1) portant une charge négative stabilisée par le trou oxyanion (Thr40 et Gln106), et la libération d'un alcool. Cette première étape dite étape d'acylation, conduit à la formation de l'acyl-enzyme (Schéma 5).

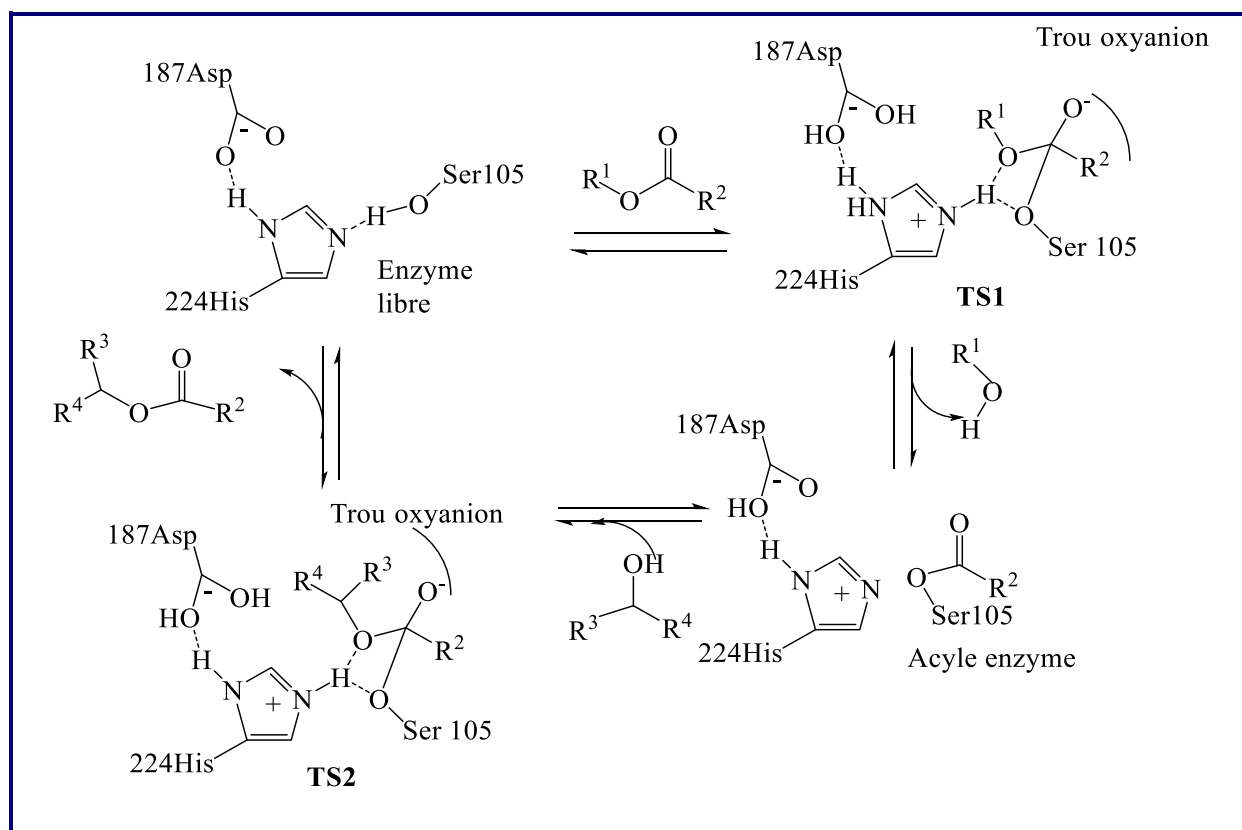


Schéma 5 Mécanisme réactionnel de transestérification de la CALB [27].

La deuxième étape est la dé-acylation de l'enzyme et la libération de l'ester produit. Son mécanisme est analogue à celui de la première étape. Elle fait intervenir l'alcool substrat qui est déprotoné à son tour par His224, la fonction alcoolate du substrat réalise l'attaque

nucléophile de la sérine acylée au niveau de la fonction carbonyle, entraînant la formation d'un second intermédiaire tétraédrique (TS2). Finalement, l'ester produit est libéré et la fonction alcoolate de Ser105 récupère le proton porté par l'azote N $\epsilon$  de His224. Les intermédiaires tétraédriques représentés dans le Schéma 5 sont des espèces chimiques instables dont les énergies sont proches des points hauts des barrières énergétiques associées aux états de transition de la réaction. Les intermédiaires tétraédriques comportent un atome d'oxygène chargé négativement, stabilisé par le trou oxyanion constitué des deux résidus Thr40 et Gln106.

### B - I.2.c. L'intermédiaire tétraédrique

L'équation d'Eyring, a montré que la vitesse de catalyse d'une réaction est fonction de la barrière énergétique à franchir, qui passe par l'état de transition à son sommet. Dans le cadre des alcools secondaires, l'hypothèse la plus probable est de considérer que le passage de la barrière énergétique s'opère au niveau de la formation du second intermédiaire tétraédrique (cf. le mécanisme réactionnel de la transestérification) qui aurait une structure proche de l'état de transition de la réaction [28].

Les travaux de Uppenberg et collaborateurs sont les premiers à s'intéresser au positionnement du substrat dans le site actif [12]. Ils ont étudié par modélisation l'octanoate de 1-phényléthyle. L'orientation du groupe phényle des formes chirales (*R*) et (*S*) vers la surface de la protéine est étudiée. Le méthyle de l'énantiomère (*R*) s'ajuste bien à la poche, tandis que le méthyle de l'énantiomère (*S*) pointe vers Gly39 et Thr40, générant des contraintes stériques. Après minimisation, pour la forme (*S*) : le méthyle bascule dans la poche, et la liaison hydrogène entre l'atome N $\epsilon$  de His224 et l'oxygène de l'alcool libre (O<sub>alcool</sub>) est rompue. On observe donc une difficulté pour la forme chirale (*S*) à « donner » son proton à His224. En 2000, Schulz et collaborateurs ont montré indirectement l'instabilité de l'intermédiaire tétraédrique avec l'énantiomère (*S*) comparé au (*R*) en évaluant la distance de la liaison hydrogène N $\epsilon$ ...O<sub>alcool</sub> (entre l'atome d'oxygène de l'alcool libre dans le site actif et le proton porté par l'azote N $\epsilon$  du groupe imidazole de His224) [29]. Ces calculs montrent que l'instabilité de la liaison hydrogène pour l'énantiomère (*S*) est corrélée avec l'énantiosélectivité, suggérant que la barrière d'énergie d'activation de la réaction est plus haute pour former l'intermédiaire tétraédrique avec l'énantiomère (*S*). Dans un esprit similaire, Nyhlén et collaborateurs ont étudié des alcools secondaires [27]. Leurs travaux ont mis en évidence par modélisation moléculaire l'établissement d'une liaison hydrogène avec Thr40 stabilisant l'intermédiaire tétraédrique. Aussi, ils ont montré la rupture de la liaison hydrogène avec l'His224 pour le (*S*)-

hexan-2-ol par dynamique moléculaire, tandis que la liaison est maintenue pour la forme chirale (*R*) au cours de tout le mécanisme. Ces observations sont en accord avec celles de Schulz et collaborateurs évoquées ci-dessus.

### B - I.3 Sélectivité de la CALB

L'énantiosélectivité est une des propriétés les plus intéressantes de la CALB et elle est au cœur des différentes études menées durant ces dernières décennies [30].

#### B - I.3.a. Chimiosélectivité chez CALB

La chimio-sélectivité est la capacité d'une enzyme à interagir préférentiellement avec un groupe fonctionnel donné.

CALB présente des propriétés de chimiosélectivité. La chimiosélectivité de la CALB vis-à-vis d'un alcool primaire et d'un thiol a été étudiée[24]. CALB montre une sélectivité 105 fois supérieure pour les alcools que pour les thiols dans les réactions de transacylation [24].

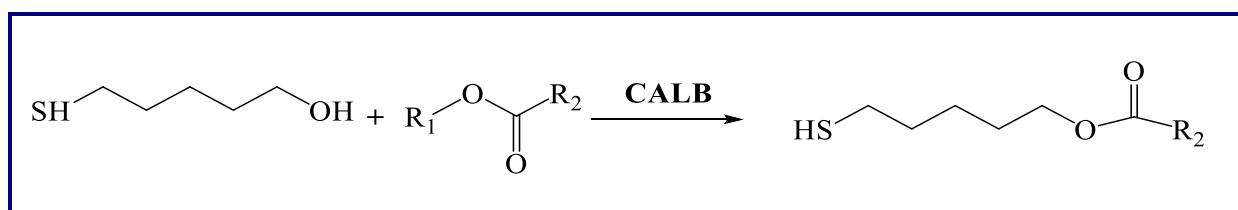


Schéma 6 Réaction d'acylation chimiosélective entre une fonction alcool et une fonction thiol par la lipase B de *Candida antarctica*.

#### B - I.3.b. Régiosélectivité chez CALB

La régiosélectivité qualifie la préférence d'une enzyme à interagir avec un groupe fonctionnel précis parmi plusieurs similaires portés par une molécule de substrat, dirigeant la réaction sur ce groupement fonctionnel.

La lipase B est effectivement douée de régiosélectivité. Outre le fait, qu'elle catalyse préférentiellement l'hydrolyse des acides gras en position 1 ou 3 des triacylglycérols, on trouve de nombreux exemples d'applications des propriétés régiosélectives de CALB. Deux exemples sont présentés ci-dessous. CALB permet l'acylation régiosélective du 3-O-angeloylingenol [31], présentée au Schéma 7. CALB permet d'effectuer également la propanoylation régiosélective des dihydroxybenzènes [32], présentée au Schéma 8.

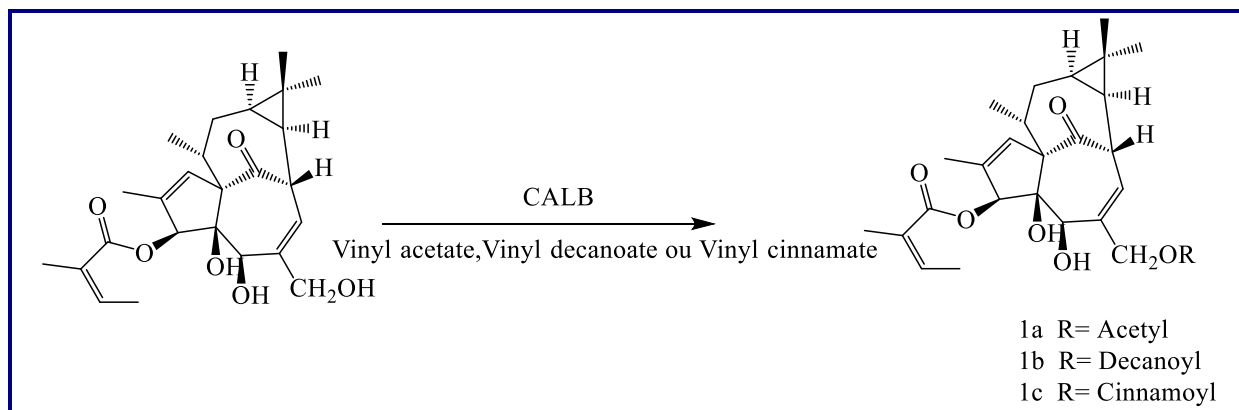
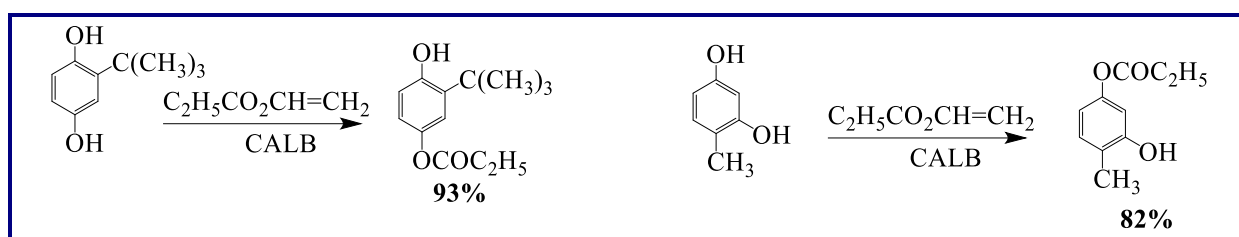
Schéma 7 Acylation régiosélective du 3-O-angeloylingénol (Teng *et al.*, 2009) [31].

Schéma 8 Propanoylation régiosélective de dihydroxybenzènes.

### B - I.3.c. L'orientation spatiale des substrats dans le site actif

La détermination de l'origine de l'énantiosélectivité passe nécessairement par une meilleure compréhension de l'orientation spatiale des substrats dans le site actif, notamment au niveau de l'intermédiaire tétraédrique.

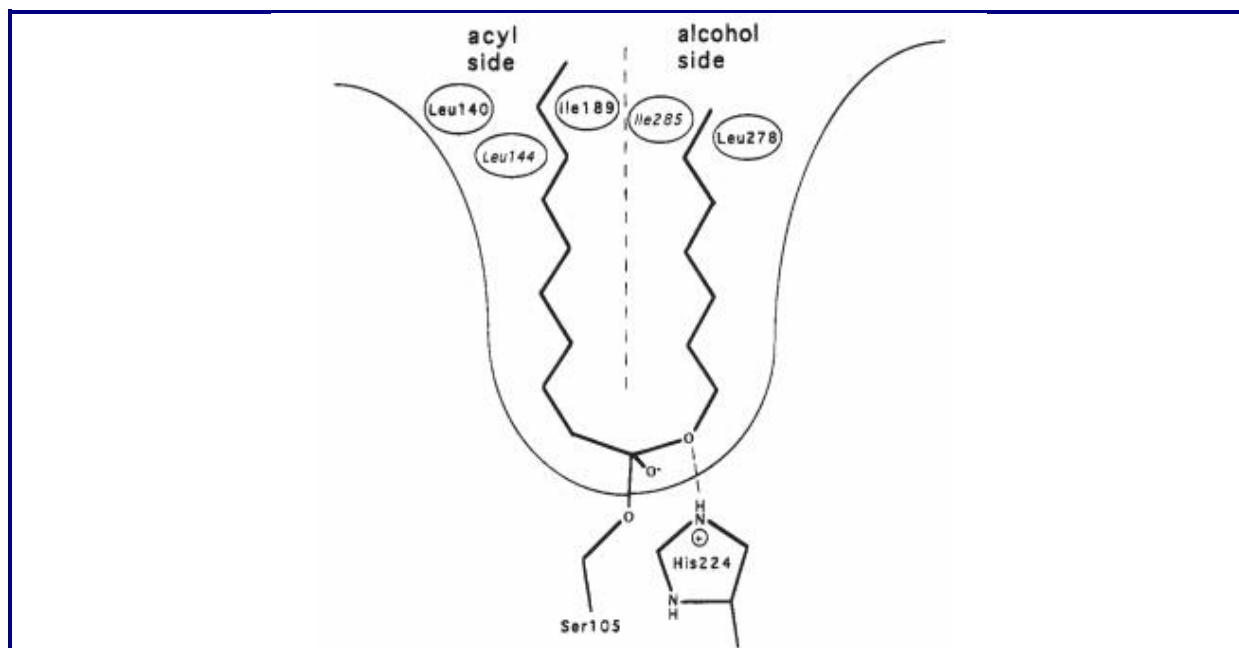


Figure 12 Orientation des parties alcool et acyle dans le site actif de la CALB[12].



Les premiers travaux de Uppenberg et collaborateurs sur la structure cristallographique de la CALB ont permis de déterminer l'orientation des parties acyle et alcool du substrat de type ester [12]. De plus, ces travaux ont mis en exergue la présence de deux zones près de Ser105, séparées par Ile189 et Ile285, l'une destinée à accueillir la partie acyle, et l'autre la partie alcool du substituant. Ceci est représenté sur la Figure 12. La région du site actif accueillant la partie acyle est localisée en direction de l'hélice  $\alpha 5$ .

Les contraintes stériques générées par l'orientation des chaînes latérales des substrats ont été évaluées dans la poche stéréospécifique [33]. Ces travaux ont mis en évidence des contraintes plus importantes au niveau de la chaîne latérale de Ser105, dans la poche stéréospécifique. Un nombre important de contacts établis avec les chaînes latérales des résidus Thr40, Gln157, et Ile189 est identifié de façon qualitative comme l'origine de l'importante énantiosélectivité observée dans des réactions d'aminolyse.

#### **B - I.4 Les peptides acylés**

L'acylation est une réaction au cours de laquelle un groupement acyle est greffé à une molécule, ce groupement étant transféré depuis un agent acylant. Les liaisons ainsi formées peuvent être de natures différentes : amide, ester, thioester, disulfure. Les liaisons amides et esters sont des liaisons covalentes fortes. Tandis que les liaisons thioester ou disulfures sont plus faibles. L'acylation de peptides ou de protéines est une modification post-traductionnelle naturelle. Son rôle reste mal défini, mais elle permet d'augmenter les interactions des peptides avec les membranes cellulaires en changeant leur structure secondaire et/ou leur hydrophobie. La greffe d'un donneur d'acyle sur un peptide peut également moduler son activité biologique.

##### **B - I.4.a. Augmentation de la lipophilie**

La plupart des peptides sont naturellement hydrophiles. Cette caractéristique physico-chimique influe sur l'absorbance, la distribution, l'activité ou l'excrétion de nombreux peptides bioactifs. La greffe d'un acide gras par la réaction d'acylation permet d'adapter de manière souple la polarité des peptides et de les rendre plus hydrophobes. Les lipopeptides sont une famille de peptides possédant une chaîne alkyle de longueur variable. Les lipopeptides ont généralement une forte activité tensioactive et des activités biologiques antimicrobiennes, antivirales ou encore hémolytiques [34]. La présence d'une chaîne grasse sur un peptide permet d'augmenter sa capacité à pénétrer les membranes biologiques [35].

Le peptide BPTI (inhibiteur de la trypsine pancréatique bovine, en anglais « *Bovine Pancreatic Trypsin Inhibitor* ») a une hydrophobie plus marquée et une affinité plus grande avec les membranes cellulaires une fois acylé [36]. L'acylation de ses acides aminés Tyr-Leu par l'acide myristique permet d'une part d'augmenter sa lipophilie et donc d'améliorer son affinité pour les constituants membranaires, et d'autre part d'induire une accumulation dans les tissus cibles de la muqueuse buccale [37]. Le couplage d'un acide gras à un peptide permet donc de faciliter son insertion dans une membrane biologique.

L'association d'un acide aminé ou d'un peptide polaire et d'un composé à longue chaîne non polaire est également utilisée pour construire des structures amphiphiles, permettant l'obtention de molécules ayant une activité de surface élevée [38]. Leurs propriétés permettent leur utilisation dans les produits pharmaceutiques, cosmétiques, ménagers et alimentaires.

#### **B - I.4.b.      Changement de structure secondaire**

L'acylation peptidique avec un acide gras peut provoquer des changements de la structure secondaire du peptide. Ainsi, lorsque la calcitonine de saumon est acylée, des interactions hydrophobes entre la chaîne grasse et le peptide induisent la formation d'une structure en hélice  $\alpha$  [39]. Également, une acylation de la lysine située à la 16<sup>ième</sup> ou 18<sup>ième</sup> position du peptide  $\beta$ -amyloïde humain avec l'acide octanoïque induit la formation d'une structure en feuillet  $\beta$  [40]. La fixation d'une chaîne grasse peut aussi stabiliser le peptide. Plus la chaîne grasse est longue (> C18) plus l'influence du donneur d'acyle sur la structure secondaire du peptide est grande. Mais, plus la chaîne est insaturée et hydrogénée, plus les interactions hydrophobes avec la chaîne peptidique sont faibles, induisant une influence moindre sur la structure peptidique [41].

#### **B - I.4.c.      Influence de la sélectivité d'acylation sur les propriétés des peptides**

Une acylation peut lier le donneur d'acyle à la chaîne principale du peptide ou à la chaîne latérale d'un des résidus. Ce positionnement sur le peptide est un facteur clef à contrôler. En effet, le greffage de la chaîne grasse module l'activité biologique du peptide. Ainsi, l'acylation du groupement amine en position N-terminale du peptide  $\alpha$ -conotoxine-MIII avec l'acide 2-amino-*D,L*-dodécanoïque induit une structuration du peptide en hélice  $\alpha$ . Une acylation sur un résidu au milieu de la chaîne peptidique empêchera la formation de cette hélice et fera perdre l'activité neuroprotectrice de ce peptide [42]. L'extrémité C-terminale de l'endomorphine-1 est responsable du maintien de son activité biologique. Une acylation sur ce

groupement avec de l'acide laurique ou octanoïque diminue fortement l'affinité de ce peptide pour les récepteurs opioïdes alors qu'une acylation sur la fonction N-terminale n'affecte pas cette reconnaissance.

Il est admis qu'une acylation préférentielle du groupement amine en position N-terminale permet généralement d'augmenter l'affinité membranaire des peptides [43]. Cependant, cette remarque n'est pas systématique pour tous les peptides. Ainsi, la di-C10-Leu-encéphaline une fois acylée sur le groupement amine en position N-terminale perd en grande partie son affinité pour ses récepteurs membranaires spécifiques [44]. De plus, une acylation sur ce groupement amine fait perdre au peptide une charge positive. Cette charge, en association avec une séquence riche en lysines et arginines, permet au peptide d'avoir une forte affinité vis-à-vis des membranes, surtout avec les membranes des bactéries qui sont chargées négativement [45].

#### **B - I.4.d. Influence de la chaîne grasse**

La nature de la chaîne grasse peut aussi entraîner une modulation de l'activité biologique. L'agent acylant, par ses propriétés intrinsèques peut améliorer l'activité biologique du peptide auquel il est rattaché. C'est le cas de l'acide docosahexaénoïque (DHA) qui apporte une activité anti-neurodégénérative au peptide auquel il est greffé [46]. La longueur de la chaîne carbonée peut également modifier l'activité ou la biodisponibilité du peptide acylé. Ainsi, des travaux ont montré que plus la chaîne carbonée est longue :

- plus l'activité antibactérienne des lipopeptides est forte [47].
- plus la protection contre la protéolyse est grande. Par exemple, plus la chaîne carbonée de l'agent acylant est longue, plus le temps de demi-vie de l'insuline est grand [46].
- plus un lipopeptide pénètre facilement une membrane lipidique, qui est un modèle utilisé [48] pour représenter une membrane biologique.

Des lipopeptides ayant une chaîne carbonée tronquée présentent, en outre, une perte considérable d'affinité membranaire [49].

#### **B - I.5 Stratégie d'acylation des peptides**

Il existe deux voies principales pour l'acylation de peptides. La première est la voie chimique qui est classiquement utilisée en industrie car bien maîtrisée. La seconde consiste à utiliser des enzymes comme catalyseur.

**B - I.5.a. Acylation peptidique par voie chimique**

L'acylation de biomolécules fonctionnelles par voie chimique est aujourd'hui bien connue.

La réaction la plus utilisée pour une acylation peptidique est la réaction de Schotten-Baumann. Cette réaction a été découverte en 1883 par deux chimistes allemands : Carl Schotten et Eugen Baumann. C'est une méthode de synthèse d'amides (ou d'ester) à partir de chlorure d'acyle et d'amine (ou d'alcool). Dans une première étape, un chlorure d'acyle réagit avec un groupement amine (ou ester) libérant un proton. Lors d'une seconde étape, une base vient capter le proton qui pourrait diminuer le rendement réactionnel. Un coproduit chloré est formé au cours de la réaction lors de la libération de ce proton (Schéma 9).

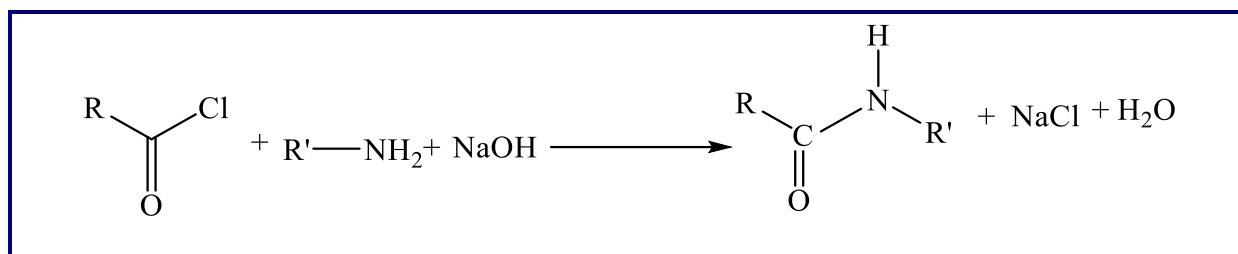


Schéma 9 Schéma général de la réaction de Schotten-Baumann.

L'intérêt des peptides acylés a été mis en évidence par des travaux utilisant cette voie de synthèse, l'acylation peptidique par voie chimique [50] [45]. Ces réactions chimiques sont, de plus, rarement sélectives. L'acylation de molécules polyfonctionnelles comme les acides aminés ou les peptides avec la réaction de Schotten-Baumann engendre des étapes de protection/dé-protection de groupements acylables. Ces étapes supplémentaires entraînent une baisse de rendement et une augmentation des coûts de production. L'utilisation de la réaction de Schotten-Baumann a d'autres inconvénients comme l'utilisation de solvants toxiques (DMF, DMSO, toluène, acétone...) et la formation de coproduits (sels) qui ne répondent plus de nos jours aux exigences en matière de production de molécules pharmaceutiques dans des conditions respectueuses de l'environnement [35].

**B - I.5.b. Acylation peptidique par voie enzymatique**

L'amélioration de la biodisponibilité de peptides et protéines bioactifs utilisés comme médicaments est un challenge extrêmement important. En effet, l'administration des médicaments peptidiques se fait dans 60% des cas par voie orale et les peptides généralement polaires possèdent des difficultés pour passer au travers de la barrière épithélio-intestinale. Ce

qui conduit à une biodisponibilité F de l'ordre de 1-2 % seulement [51] [52]. Les molécules bioactives administrées par voie orale peuvent également être la cible de protéases présentes dans le système digestif. L'acylation constitue une des solutions à ce problème. Il a par exemple été démontré *in vitro* que la longueur de la chaîne carbonée de l'agent acylant joue un rôle dans la protection contre les protéases [53].

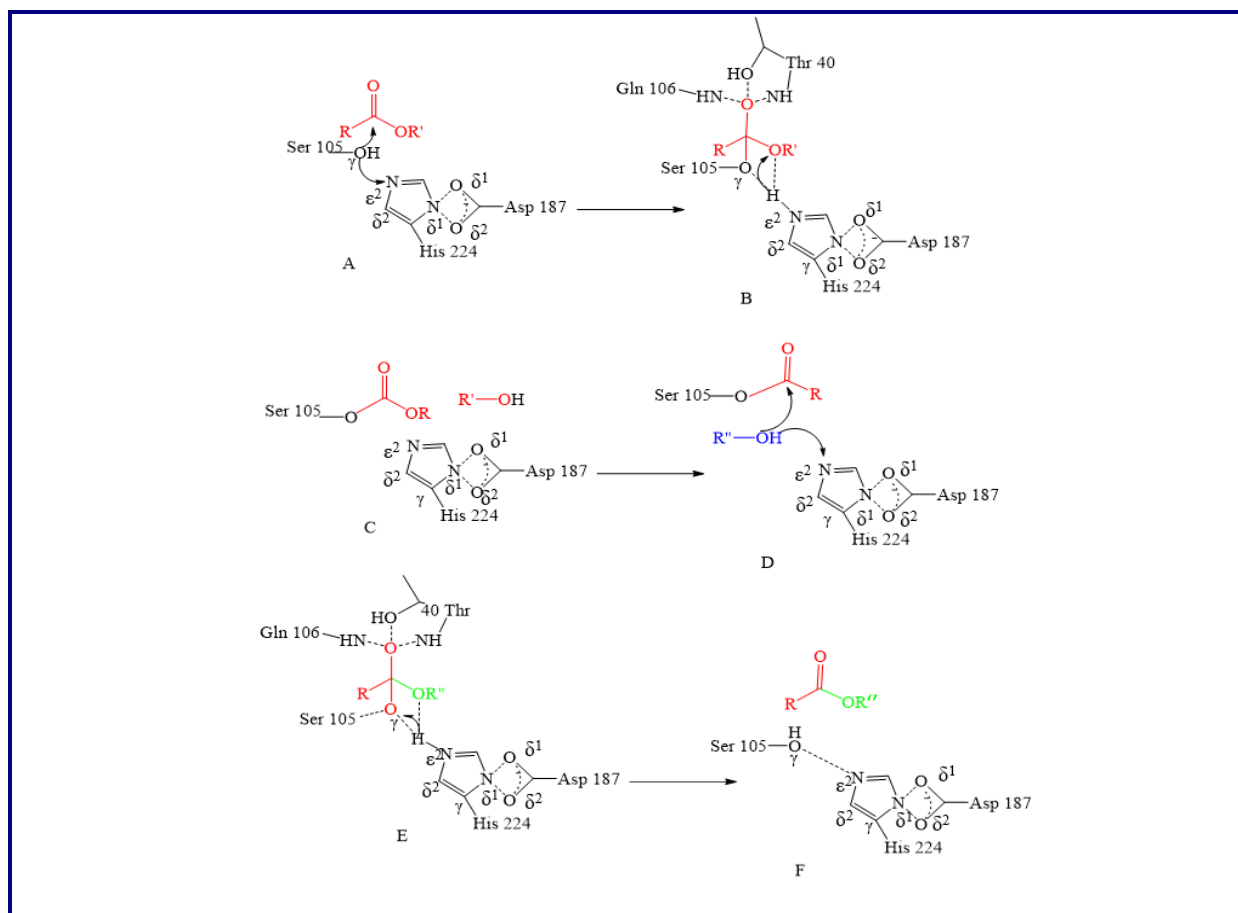
Plusieurs études s'intéressent à l'utilisation d'enzymes pour la synthèse de dérivés acylés d'alcools aminés, d'acides aminés, de peptides ou de carbohydrates [54].

Chaque enzyme a une sélectivité qui lui est propre. Pour catalyser une réaction précise, les biocatalyseurs doivent être choisis en fonction de leur nature, de leur activité et de leur spécificité de substrat. Parmi les enzymes catalysant la réaction d'acylation, on trouve les lipases, les acylases et d'autres enzymes comme les acyl-transférases, les protéases et les carboxyl-estérases.

### **B - I.6 Mécanisme d'acylation enzymatique avec CALB**

Le mécanisme de catalyse par les lipases est proche de celui des sérine-protéases. Il s'agit d'un mécanisme bi bi ping-pong. Il est communément admis que l'état de transition enzyme/substrats/produits des lipases passe par la constitution d'intermédiaires tétraédriques (Fuentes *et al.*, 2004)[14]. Ce mécanisme est détaillé sur le Schéma 10. Il repose sur le même principe que le mécanisme de transestérification présenté dans le Schéma 5 du paragraphe B - I.2.b. Le substrat donneur d'acyle entre dans le site actif et le proton du groupement hydroxyle est transféré à l'atome d'azote N $\epsilon$ 2 du groupement imidazole du résidu His224 (Schéma 10A). Le donneur d'acyle réalise alors une attaque nucléophile sur Ser105 formant ainsi le premier intermédiaire tétraédrique (Schéma 10B). L'oxyanion est orienté et stabilisé par trois liaisons hydrogène établies avec les résidus du trou oxyanion (Thr40 et Gln106). Le coproduit de la réaction de type alcool est ensuite relargué, quittant le complexe alors appelé acyl-enzyme (Schéma 10C). Le substrat accepteur d'acyle entre ensuite dans le site actif et transfère un proton de la fonction à acyler à l'azote N $\epsilon$ 2 de His224 augmentant ainsi sa nucléophilie (Schéma 10D). Ce groupement activé effectue à son tour une attaque nucléophile sur le carbonyle de l'acyl-enzyme formant ainsi le second intermédiaire tétraédrique (Schéma 10E). Le produit acylé est alors relargué, le proton de l'azote N $\epsilon$ 2 de His224 est transféré à Ser105 et la lipase CALB se retrouve à nouveau dans son état initial (Schéma 10F). Le rôle de Asp187 est d'augmenter la basicité de His224 grâce au champ électrique de sa chaîne latérale. Celle-ci

active à son tour le groupement hydroxyle de la chaîne latérale de la sérine catalytique (Ser105) par une liaison hydrogène (Schéma 10F).



**Schéma 10 : Mécanisme de catalyse d'acylation de la CALB (Tirée de FERRARI 2014).**

A : Conformation native de l'enzyme et approche du substrat donneur d'acyle ; B : Formation du 1<sup>er</sup> intermédiaire tétraédrique ; C : Relargage du coproduit et formation de l'acyl-enzyme ; D : Attaque de l'accepteur d'acyle ; E : Formation du 2<sup>nd</sup> intermédiaire tétraédrique ; F : Libération du produit acylé et retour à la conformation initiale de l'enzyme.

## B - I.7 Applications de CALB

La lipase B de *Candida antarctica* (CALB) est une enzyme largement utilisée dans des applications industrielles [55]. Elle est notamment exploitée comme catalyseur pour l'estérification ou la transestérification d'acides gras dans la production de biodiesel [56], ou de glycérols monoacyl [57]. Elle est connue pour rester active dans des solvants organiques aussi bien polaires que non polaires ainsi que dans des milieux pauvres en eau. La Figure 13 montre les réactions catalysées par les lipases, et le Tableau 2 résume plusieurs de leurs applications spécifiques dans différents secteurs industriels. En outre, ces enzymes présentent une régiospécificité, une spécificité de substrat et une stéréospécificité qui sont mises à profit dans de nombreuses applications.

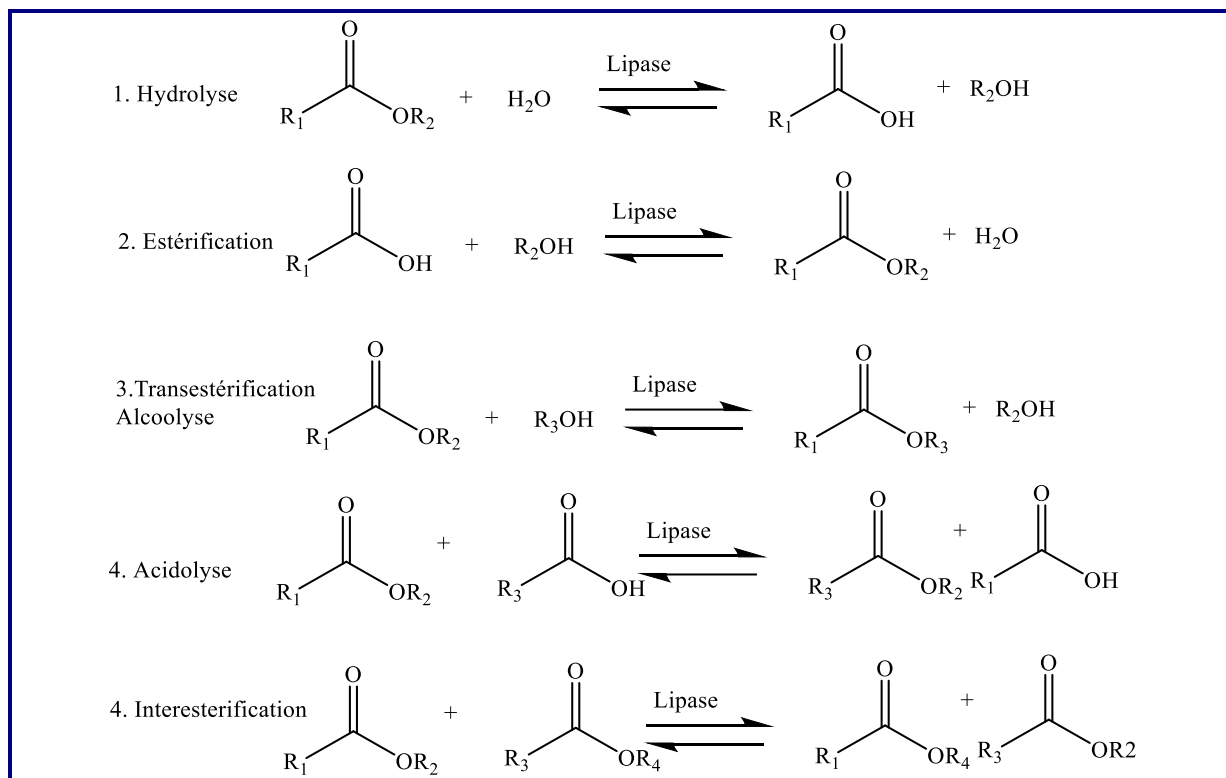


Figure 13 Les différentes réactions catalysées par les lipases.

Tableau 2 : Diverses applications des lipases dans différents secteurs industriels

Industries	Produits / applications	Réf.
Détergent	Élimine les taches de lipides	[58]
Laiterie	Développement des saveurs dans le lait, le fromage et le beurre	[59]
Boulangerie	Amélioration et contrôle du brunissement non enzymatique, augmentation du volume du pain, amélioration de la structure de la mie	[60]
Brasserie et alimentation	Amélioration de la saveur et de la qualité dans des boissons, de la viande et des poissons	[60]
Cuir	Produits en cuir	
Papier	Amélioration de la qualité du papier	[61]
Huile naturelle	Beurre de cacao	
Saveurs et fragrance	Synthèse d'esters aromatiques naturels	[62]
Tensioactif	Utilisation de monoglycérides comme tensioactifs	[9]
Pharmaceutique	Production de médicaments chiraux, résolution de mélange racémique	[63] [64]
Carburant	Biodiesel	[65]
Tensioactifs	Production d'esters mono-acyle de sucres comme tensioactif	[66]

Le processus d'acylation peut être réalisé par des voies enzymatiques ou chimiques (détaillé dans les paragraphes 1.5.a et 1.5.b). Les principaux avantages de l'utilisation d'enzymes au lieu de catalyseurs chimiques sont une grande sélectivité permettant le couplage des chaînes acyles sur des positions spécifiques et la possibilité d'appliquer des conditions de réaction douces qui respectent les exigences de la chimie verte.

Les enzymes hydrolytiques et principalement les lipases (EC 3.1.1.3) sont le plus souvent utilisées pour catalyser de telles réactions en milieu non aqueux, et sont d'ailleurs déjà appliquées à des bioprocédés industriels dans le domaine alimentaire, pharmaceutique et cosmétique [67]. En plus de leur large spécificité de substrat, les lipases ne nécessitent aucune coenzyme. La lipase B de *Candida antarctica* (CALB) a déjà été utilisée dans de nombreux bioprocédés d'acylation appliqués à une large gamme de composés comme les vitamines, les sucres, les acides aminés ou les phénols [68] [69]. Le développement de tels bioprocédés et le contrôle de la sélectivité nécessitent une compréhension approfondie des modes de liaison enzyme/substrat ainsi que des règles moléculaires de la sélectivité. Pour y parvenir, plusieurs études ont fait état de l'utilisation de simulations de modélisation moléculaire, notamment basées sur des méthodologies de docking ou amarrage moléculaire [70] [71] [72]. Dans le cas où la structure de l'enzyme est connue, le docking apparaît comme une méthode efficace pour prédire les orientations préférées des substrats dans la cavité catalytique cible et ensuite pour analyser les interactions entre ces dernières [70] [73]. Les positions issues des calculs de docking sont classées à l'aide de fonctions de score, basées sur l'approximation de l'affinité de liaison [74]. Cette étape de classement représente une étape décisive dans les protocoles de docking qui peut fortement impacter la précision des modèles. De nombreuses études de modélisation moléculaire visant à comprendre la spécificité du substrat ainsi que la sélectivité ont déjà été réalisées sur les réactions catalysées par CALB. Une approche combinant le docking et les simulations de dynamique moléculaire a été appliquée pour étudier la régiosélectivité de CALB lors de l'acétylation des flavonoïdes au niveau moléculaire [75] [71]. De même, l'acylation de peptides à base de lysine a été étudiée et la régiosélectivité de la réaction a été prédite avec succès [76]. La structure de la lysine possède deux groupes amino qui sont enclins à l'acylation.

L'un est situé à l'extrémité de la chaîne latérale de la lysine (position  $\epsilon$ ), et l'autre sur le carbone  $\alpha$ . L'acylation de la lysine catalysée par la lipase de *Rhizomucor miehei* a déjà été étudiée expérimentalement [69]. Les résultats ont montré que l'acylation exclusive du groupe  $\epsilon$ -amino de la lysine mais aucune explication claire n'a été donnée pour cette sélectivité. L'un des objectifs de cette thèse est d'étudier la sélectivité de CALB vis-à-vis des groupements amine de la lysine.



## B - II LA NEUROPILINE-1 (NRP-1)

### B - II.1 Introduction

La Neuropiline-1 (Nrp-1 ou Npn-1) est une glycoprotéine transmembranaire qui a été identifiée pour la première fois comme un antigène exprimé à la surface des neurones de têtards de *Xenopus laevis* [77].

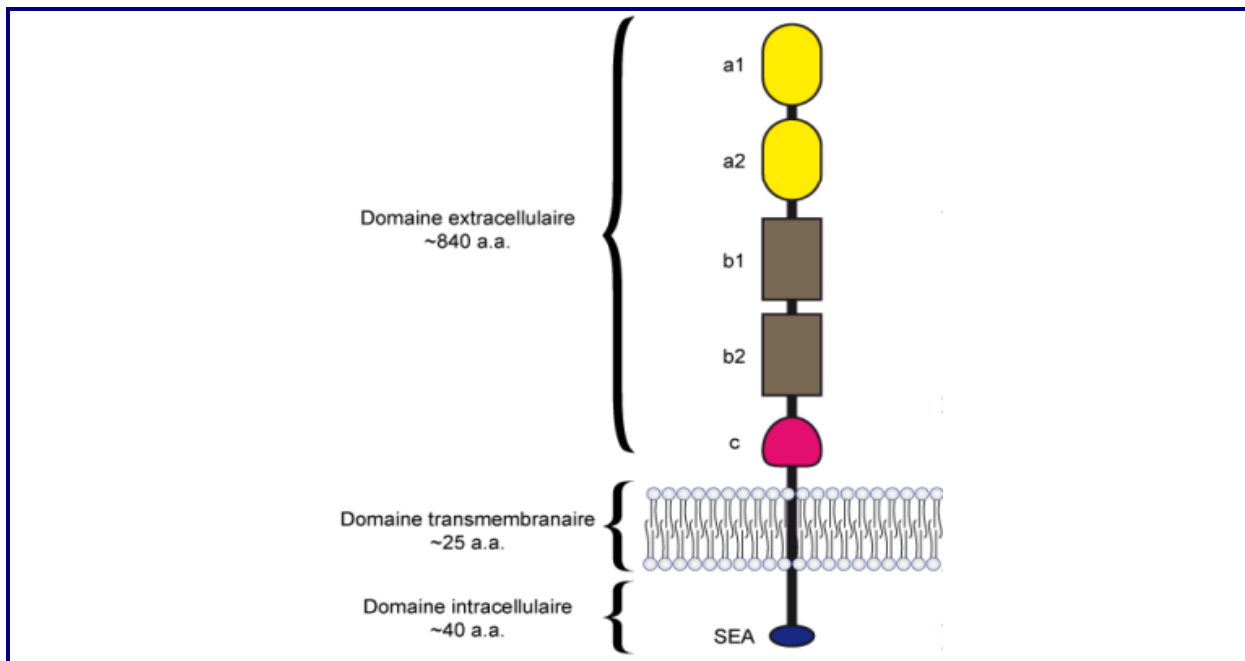


Figure 14 Schéma des différents domaines structurels et fonctionnels de la protéine Nrp-1 (D'après MILPIED 2010).

C'est un récepteur dénué d'activité tyrosine kinase<sup>6</sup> (RTK) de 130 kDa. Chez l'homme, le gène Nrp-1 se situe sur le chromosome 10 bras court bande 12 (10p12), alors que chez la souris, il est localisé sur le chromosome 8. La Neuropiline-1 (Nrp-1) est un corécepteur transmembranaire impliqué dans des interactions avec divers ligands et récepteurs. Un homologue de Nrp-1 a été identifié plus tard et nommé Neuropiline-2 (Nrp-2) qui possède 44% d'homologie en acides aminés avec Nrp-1 (bande 2q34) [78]. L'épissage alternatif de l'ARN<sup>7</sup> produit des formes solubles de Nrp-1 et Nrp-2 dépourvues des domaines transmembranaires et cytosoliques. L'architecture (Figure 14) de Nrp-1 est composée d'une vaste partie extracellulaire modulaire (a1/a2, b1/b2, c) au niveau N-terminal. Ce domaine d'environ 840

<sup>6</sup> Récepteurs transmembranaires qui appartiennent à la famille des récepteurs-enzymes.

<sup>7</sup> Processus par lequel les ARN transcrits à partir de l'ADN génomique peuvent subir des étapes de coupure et ligature qui conduisent à l'élimination de certaines régions dans l'ARN final.

acides aminés est suivi d'un court domaine transmembranaire d'environ 25 résidus et d'un domaine cytoplasmique d'environ 40 résidus. Les domaines a1/a2 et b1/b2 sont indispensables à la liaison des ligands. Une des particularités des neuropilines est de posséder un court domaine intracellulaire (SEA) ne permettant pas d'induire une signalisation, et nécessitant donc une association à des corécepteurs, pour induire la signalisation des ligands spécifiques.

La partie extracellulaire comprend deux domaines CUB<sup>8</sup> (a1/a2) (Complément 1r/s, Uegf - protéine embryonnaire de l'oursin - et BMP1), deux domaines d'homologie aux facteurs de coagulation V/VIII (b1/b2), un domaine MAM (c) (homologue à la protéase méprine) (Figure 15), un simple domaine transmembranaire (TM) et un petit domaine cytoplasmique contenant un motif PDZ<sup>9</sup> suivi d'un triplet de résidus SEA (sérine, acide glutamique, alanine) [79].

### **B - II.1.a. Les ligands de Nrp-1**

Nrp-1 est un corécepteur qui s'associe à de nombreux ligands à la surface cellulaire pour assurer la transduction du signal. Toutefois, il a été montré que Nrp-1 pouvait, par son domaine cytoplasmique, être impliqué dans l'internalisation de l'intégrine  $\alpha 5 \beta 1$  [80]. Les deux principaux ligands de Nrp-1 sont les sémaphorines<sup>10</sup> de classe 3 (Sema3, notamment Sema3A) et le facteur de croissance de l'endothélium vasculaire (FCEV) (en anglais : Vascular Endothelial Growth Factor : VEGF) et en particulier de VEGF-165 [81] (Figure 15). Mais d'autres ligands ont été décrits pour stimuler diverses voies de signalisation comme la progression et l'invasion tumorale. Ainsi Nrp-1 est associée à de nombreuses fonctions. Nrp-1 intervient lors du développement embryonnaire du système nerveux et cardiovasculaire et chez l'adulte, notamment en conditions pathologiques (rôle dans l'angiogenèse tumorale) [82] [83]. En plus de son rôle crucial dans le développement des systèmes nerveux et cardiovasculaires, Nrp-1 est impliquée dans des processus physiopathologiques. Dans le système immunitaire,

---

<sup>8</sup> Un domaine CUB est un motif structural, d'environ 110 résidus, présent presque uniquement dans des protéines extracellulaires associées à la membrane et plasmiques

<sup>9</sup> « PDZ est un acronyme pour trois noms de protéines possédant le motif : Post-synaptic density protein 95 (PSD-95), drosophila disc large tumor suppressor (Dlg1) et zona occludens 1 (ZO-1)

<sup>10</sup> Constituent une famille de protéines dont le rôle est de guider la croissance des axones.

Nrp-1 participe aux interactions entre les lymphocytes T et les cellules dendritiques<sup>11</sup>, transmet les effets immunorégulateurs de Sema3A sur les lymphocytes T. De plus, il est exprimé par d'autres types cellulaires comme les cellules des îlots pancréatiques<sup>12</sup>, des hépatocytes et les cellules tumorales[84].

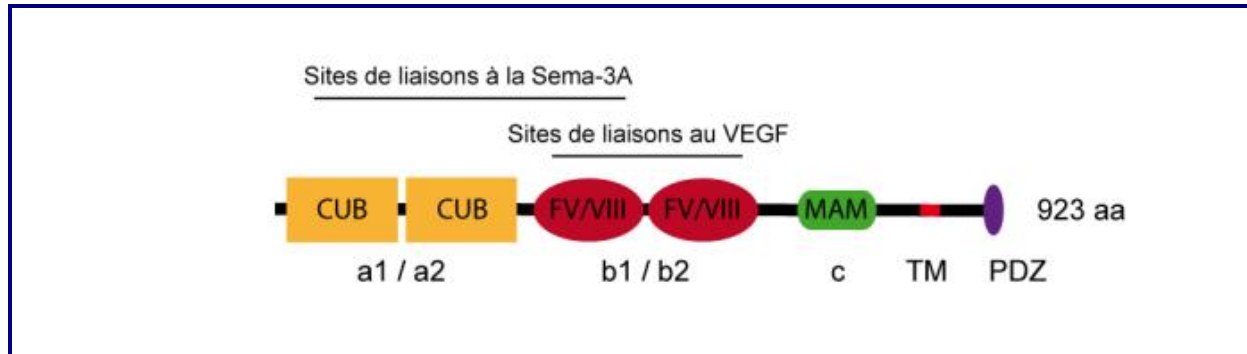


Figure 15 Schéma de fixation des différents ligands (Sema 3A, VEGF) sur Nrp-1 (Tirée de RENAND 2011).

### B - II.1.b. Les sémaphorines

Les sémaphorines sont une famille de protéines solubles et transmembranaires retrouvées chez les vertébrés et invertébrés, et même les virus, mais absentes chez métazoaires primitifs les protozoaires et les plantes. On distingue 8 classes de sémaphorines : la classe V (sémaphorines virales), les classes 1 et 2 (chez les invertébrés), et les classes 3 à 7 (chez les vertébrés). Les sémaphorines de classe 3 (Sema3) sont des protéines solubles d'environ 90 kDa contenant un domaine structurel commun à toutes les sémaphorines (domaine Sema) en position N-terminale, ainsi qu'un domaine de type immunoglobuline, et une région basique en position C-terminale [85] [82]. On compte 7 membres dans cette famille de sémaphorines (Sema3A-G).

Parmi ces membres, Sema3A, Sema3B, Sema3C, Sema3E et Sema3F se fixent à Nrp-1, alors que Sema3B, Sema3C, Sema3F et Sema3G se fixent à Nrp-2. Les couples ligand/récepteur les plus étudiés dans ce système sont Sema3A/Nrp-1 [79]. Le domaine Sema de Sema3A se fixe aux domaines a1 et a2 (domaines CUB) de Nrp-1, alors que la queue C-terminale basique de Sema3A interagit avec le domaine b1 [82]. Le domaine immunoglobuline

<sup>11</sup> Cellules dendritiques sont des leucocytes spécialisés dans la présentation antigénique aux lymphocytes.

<sup>12</sup> Les îlots pancréatiques, ou îlots de Langerhans (du nom de leur découvreur), sont formés de cellules endocrines capables de synthétiser des hormones : l'insuline, le glucagon etc.

et la queue basique de Sema3A sont aussi impliqués dans la formation d'homodimères de Sema3A, forme sous laquelle ce ligand est efficace pour induire ses effets sur les cellules Nrp-1<sup>+</sup> [86].

Comme indiqué précédemment, Nrp-1 ne possède pas, dans sa partie intracytoplasmique, de domaine fonctionnel capable de traduire un signal [87]. Les effets des ligands de Nrp-1 dépendent donc de la formation de complexes avec des corécepteurs capables, eux, de transmettre un signal intracellulaire. Dans le cas des Sema3, les corécepteurs impliqués dans la signalisation appartiennent à la famille des plexines<sup>13</sup>. Les plexines sont distribuées en 4 familles (PlexinA-D). Les plexines impliquées dans la signalisation des Sema3 sont principalement les PlexinA, dont on compte 4 membres (PlexinA1-4) [88].

### **B - II.1.c. Les Facteurs de Croissance de l'Endothélium Vasculaire (VEGF)**

Les protéines de la famille du facteur de croissance de l'endothélium vasculaire ou (« vascular endothelial growth factor en anglais ») VEGF comptent 5 membres : VEGF-A-D et le facteur de croissance placentaire PlGF (« placental growth factor en anglais ») [89]. Ces protéines sont toutes impliquées dans la vasculogénèse (formation de nouveaux vaisseaux sanguins à partir de cellules souches hématopoïétiques) et dans l'angiogénèse<sup>14</sup> pendant le développement embryonnaire et à l'âge adulte. Le membre le plus abondamment retrouvé, et étudié, est le VEGF-A, souvent appelé simplement VEGF. L'épissage<sup>15</sup> alternatif du gène du VEGF donne naissance à différentes isoformes de tailles différentes : VEGF121, VEGF145, VEGF165, VEGF189, VEGF206. Les VEGF sont des homodimères anti-parallèles reliés par deux ponts disulfures [90]. L'isoforme la plus abondante et la plus active du VEGF est le VEGF-165, qui se fixe à Nrp-1 par les segments codés par les exons 7 et 8 [90]. Bien qu'il ait longtemps été considéré que seul l'exon 7 pouvait conférer au VEGF la capacité de lier Nrp-1, plusieurs études récentes ont montré que l'exon 8 conférait aussi cette propriété aux VEGF. En effet, le VEGF-121, qui ne contient pas l'exon 7, peut se lier à Nrp-1 par l'exon 8 [91].

---

<sup>13</sup> Protéine qui agit comme un récepteur pour les protéines de signalisation de la famille des sémaphorines.

<sup>14</sup> Bourgeonnement de nouveaux vaisseaux sanguins à partir de vaisseaux préexistants.

<sup>15</sup> L'épissage alternatif est un processus qui permet, à partir d'une séquence génomique unique, de produire plusieurs ARNm correspondant à des protéines distinctes.

De même, un petit peptide naturel homologue à la séquence codée par l'exon 8, la tuftsine<sup>16</sup>, se fixe à Nrp-1 [92]. Les domaines b1 et b2 sont impliqués dans la fixation des différentes isoformes de VEGF sur Nrp-1 [81]. Ainsi, la partie C-terminale de l'exon 8 du VEGF va fixer le domaine b1 de la Nrp1, et la liaison de l'exon 7 va permettre de stabiliser le complexe [93].

La signalisation du VEGF sur les cellules Nrp-1<sup>+</sup> (notamment les cellules endothéliales) nécessite, dans la plupart des cas, la présence du récepteur à activité tyrosine kinase VEGFR2 (aussi appelé KDR) [79]. Les récepteurs de cette famille (VEGFR1-3) sont des protéines transmembranaires contenant, dans leur partie extracellulaire deux sous-unités à activité tyrosine kinase [79]. Le VEGFR2 est capable de fixer VEGF en l'absence de Nrp-1. Cependant, de nombreuses réponses des cellules endothéliales au VEGF nécessitent l'association de VEGFR2 avec Nrp-1. L'association de ces deux récepteurs renforce la signalisation (phosphorylation de tyrosines, activation des voies PI3K<sup>17</sup> et de certaines voies MAPK<sup>18</sup>) du VEGFR2 [79].

## **B - II.2 Les Fonctions de la Neuropiline-1**

### **B - II.2.a. Rôle dans les systèmes nerveux et vasculaire**

L'implication de la Nrp-1 dans le développement cardiovasculaire et nerveux est fonction de son expression cellulaire et des ligands présents. En effet, tandis que le complexe Sema-3A/Nrp-1/Plexin-A intervient dans la « guidance » des neurones, le complexe VEGF/VEGFR2 est quant à lui impliqué dans l'angiogenèse [94]. La Nrp-1 est indispensable pour induire la signalisation de Sema3A sur les neurones. En culture *in vitro*, Sema3A a une action répulsive sur la croissance des axones en induisant un « collapsus » au niveau du cône de croissance [95]. Sema3A a aussi un rôle sur les dendrites des neurones. Ceci permet une organisation structurée et complexe du système nerveux au cours de son développement. Chez l'adulte, il est montré que le couple Sema3A/Nrp-1 a une importance dans la régénération et

---

<sup>16</sup> Tuftsine, tétrapeptide (Thr-Lys-Pro-Arg) située dans le domaine Fc de la chaîne lourde de l'immunoglobuline.

<sup>17</sup> <https://www.anticorps-enligne.fr/signalisation-pi3k-akt-pathway-4/>.

<sup>18</sup> Les MAPK sont un ensemble de protéines kinases nécessaires à l'induction de la mitose dans les cellules eucaryotes

l'invasion des nerfs. En effet, au niveau d'une lésion nerveuse périphérique une accumulation de Sema3A empêche la cicatrisation et la régénération nerveuse [96].

Sur les cellules endothéliales la Nrp-1, associée au VEGFR2, est impliquée dans la migration cellulaire [90]. L'utilisation d'anticorps bloquant Nrp-1 pour la fixation du VEGF inhibe la migration des cellules endothéliales, *in vitro*, et la néovascularisation, *in vivo* [97]. Bien que l'utilisation d'anticorps bloquant la fixation de Sema3A ne semble pas affecter significativement la migration des cellules endothéliales, d'autres études montrent que la Sema3A induit une chimio-répulsion et inhibe l'adhésion des cellules endothéliales de porc (PAE) et humaines [97] [98]. Nrp-1 semble aussi jouer un rôle dans l'adhésion des cellules endothéliales à la matrice extracellulaire et ce, de manière dépendante ou indépendante de ses ligands habituels [99] [100] [80]. Cependant, des études proposent une action inhibitrice de Sema3A sur l'adhésion des cellules endothéliales en inhibant l'activité des intégrines<sup>19</sup>[101]. Nrp-1 et ses ligands sont donc impliqués à différents niveaux sur la mobilité des cellules endothéliales.

#### **B - II.2.b. Rôle dans le cancer**

Nrp-1 est retrouvée sur différents types de cellules cancéreuses et est impliquée dans la migration, la survie et la prolifération tumorale par VEGF [102] [103]. La sur- ou la sous-expression de Nrp-1 est souvent associée à un mauvais pronostic et au potentiel métastatique de la tumeur [104] [105]. Naturellement, Nrp-1 est devenue une cible anti-tumorale et l'utilisation de peptides inhibiteurs de la Nrp-1 induit l'apoptose de certaines cellules cancéreuses [106] [107]. D'autre part, l'utilisation d'anticorps bloquant Nrp-1, en association avec des anticorps anti-VEGF, permet d'inhiber la vascularisation de la tumeur et ainsi de diminuer sa croissance tumorale, et c'est indépendamment de l'expression de Nrp-1 par les cellules cancéreuses [97]. Nrp-1 est aussi utilisée pour favoriser une meilleure pénétration des médicaments au niveau de la tumeur [108]. Ainsi, l'utilisation de peptides fixant la Nrp-1, co-administré avec des agents actifs, permet une pénétration au niveau des vaisseaux sanguins et une atteinte de la tumeur plus en profondeur. Contrairement au VEGF, les sémaphorines sont plutôt décrites comme ayant un rôle anti-tumoral. Différentes sémaphorines de classe III, dont Sema3A, ont la capacité de bloquer la migration des cellules tumorales [109] [110]. De plus,

---

<sup>19</sup> Les intégrines sont des récepteurs d'adhésion cellulaire, ou des protéines transmembranaires dont l'une des extrémités interagit en général avec des protéines de la matrice extracellulaire.

la délétion des gènes Sema3B et Sema3F dans certains cancers suggère que les sémaphorines de classe III font partie des gènes répresseurs de tumeurs [111]. Cependant, certaines études cliniques montrent une corrélation entre le haut niveau de Sema3A et la progression tumorale [112].

### **B - II.2.c. Rôle de Nrp-1 dans le SRAS-COV-2 (COVID-19)**

L'infection par le coronavirus-2 (SARS-CoV-2) du syndrome respiratoire aigu sévère (SRAS) est à l'origine d'une nouvelle maladie infectieuse virale, appelée maladie à coronavirus 2019 (COVID-19) [113]. Après l'apparition initiale de cas de COVID-19 à la fin de l'année 2019 en Chine, cette maladie a atteint le statut de pandémie en quelques mois, avec des taux d'infection et de mortalité en augmentation rapide. Ainsi, en 2020, plus de 118000 de cas confirmés et plus de 4 000 décès ont été signalés dans le monde [114]. Le COVID-19 est le plus souvent asymptomatique ou bénin et se manifeste principalement comme une infection des voies respiratoires inférieures qui se transmet essentiellement par les gouttelettes d'air [115]. De nouvelles données indiquent que le SRAS-CoV-2 peut également envahir et attaquer le système nerveux central (SNC), entraînant un large répertoire de symptômes et de complications neurologiques [116] [117]. C'est pourquoi les recherches se concentrent de plus en plus sur l'identification des médiateurs qui facilitent l'infection par le SRAS-CoV-2 dans différents organes/tissus humains. Il est désormais bien établi que l'entrée du SRAS-CoV-2 dans les cellules est facilitée par ses protéines de surface (ou protéine spike S), notamment par une liaison à l'enzyme de conversion de l'angiotensine-2 (ACE-2) [118].

Des données récentes suggèrent Nrp1 pourrait également être impliquée dans l'infection par le SRAS-CoV-2 [119]. Deux études précliniques ont indiqué que Nrp-1 pourrait jouer un rôle en tant que nouveau médiateur de l'infection par le SRAS-CoV-2 [120]. En effet, dans la première de ces études, Nrp-1 a été identifiée comme un facteur de l'infection par le SRAS-CoV-2, puisque, outre l'ACE-2, les protéines spike (S) de ce coronavirus peuvent également se lier à Nrp-1 *in vitro* [121]. Des résultats similaires ont été rapportés par une seconde étude dans laquelle l'analyse neuropathologique des autopsies de COVID-19 a en outre révélé que le SRAS-CoV-2 avait infecté des cellules Nrp-1 positives (Nrp-1<sup>+</sup>) dans l'épithélium et le bulbe olfactifs [120]. Sur la base de ces résultats, Nrp-1 peut être impliquée dans le COVID-19.

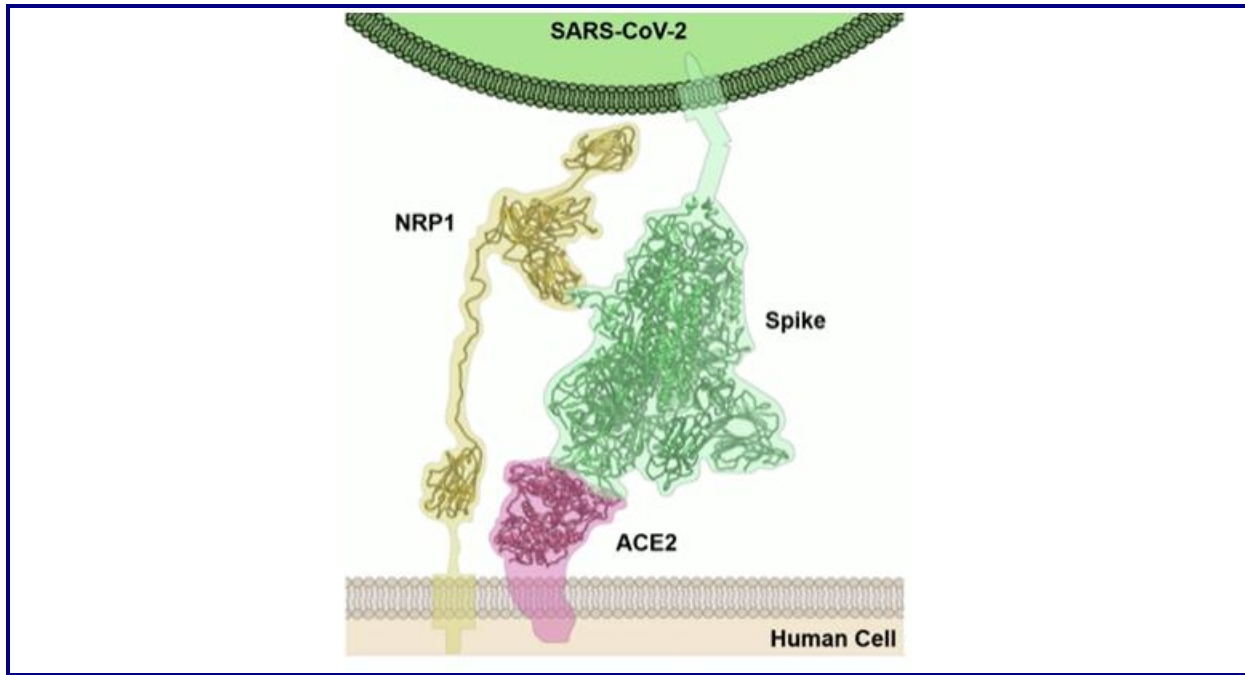


Figure 16 Interaction entre le virus et les récepteurs (Nrp-1 et ACE-2)[122].

### B - II.3 Structures cristallographiques de Nrp-1

La structure cristallographique de Nrp-1 dans sa totalité n'a pas été déterminée à ce jour. Le domaine b1 représente la première structure d'une partie de Nrp-1 résolue (Figure 17) [123]. Ce domaine est de première importance car impliqué dans la plupart des interactions réalisées par Nrp-1, notamment avec les ligands principaux Sema3A et VEGF. Cette structure fournissait une base pour comprendre les principes généraux de la reconnaissance moléculaire de Nrp-1 dans divers processus comme le guidage de l'axone, l'adhésion cellulaire etc. Depuis de nombreuses autres structures ont été déterminées, des structures de plusieurs domaines, des complexes d'un ou plusieurs domaines avec divers ligands naturels mais aussi des complexes avec divers cibles thérapeutiques (voir Tableau 3). Dans le cadre de cette thèse, les tentatives de cristallisation ont été uniquement réalisées avec le domaine b1 qui sera le seul décrit d'un point de vue structure.



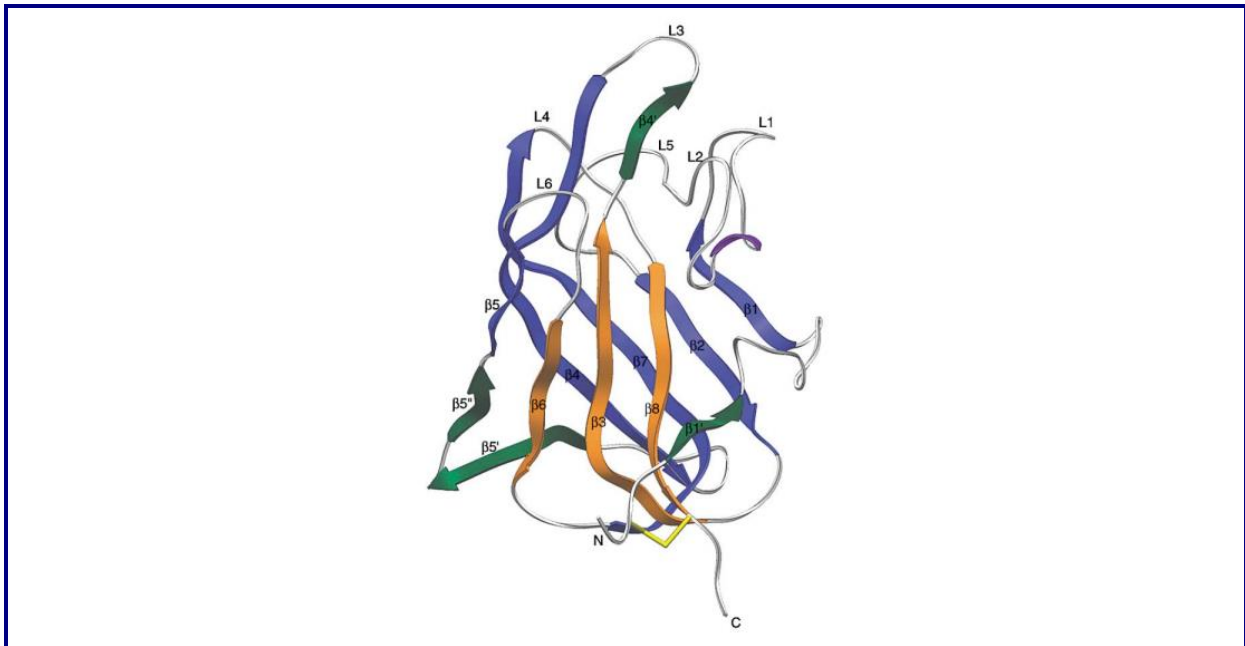


Figure 17 Représentation schématique de Nrp-1 ( Tirée de Lee 2003)[123]

Les domaines b1 et b2 de Nrp-1 sont des domaines homologues à la discoïdine. Le domaine b1 se présente sous la forme d'un tonneau  $\beta$  distordu « Jelly roll » à 8 brins. Le feuillet ( $\beta 1$ - $\beta 2$ - $\beta 7$ - $\beta 4$ - $\beta 5$ ) est plaqué contre le feuillet ( $\beta 3$ - $\beta 6$ - $\beta 8$ ). Le domaine b1 possède quelques brins supplémentaires ( $\beta 1'$ ,  $\beta 4'$ ,  $\beta 5'$  et  $\beta 5''$ ) dont certains impliqués dans deux coudes bêta (Figure 17). Six régions de boucles juxtaposées (L1-L6) s'étendent au niveau d'un sommet du tonneau bêta. Ces boucles délimitent un profond sillon, qui s'étend entre les boucles L1/L5 et les boucles L2-L4. La boucle la plus proéminente (L3) adopte une conformation de virage bêta étendu et est légèrement étirée au-dessus de la fente vers L1/L5. L6 est positionnée directement derrière L3 et se tasse contre elle (Figure 17).

La structure tridimensionnelle du domaine b1 de Nrp-1 permet de visualiser et de réconcilier la façon dont ce domaine unique peut fonctionner dans une interaction de liaison spécifique avec des ligands protéiques de structures apparemment diverses [124]. Elle met en évidence les régions générales de la molécule impliquées dans la liaison de la sémaphorine sécrétée et du ligand VEGF-165 [123]. La structure des domaines b1 et b2 en complexe avec la tuftsine révèle une interaction avec la pointe du domaine b1. Pour rappel, la tuftsine est un térapeptide (TKPR) immunostimulateur qui est très similaire à l'extrémité C-terminale de VEGF-165 (DKPRR) et qui rentre en compétition avec ce dernier pour interagir avec Nrp-1. L'arginine terminale du térapeptide est primordial dans l'interaction[125]. Sa partie aliphatique interagit avec la tyrosine 297 de la boucle L1 et le groupement guanidinium fait un

pont salin avec la fonction carboxylate de l'aspartate 320 de la seconde boucle (L2)[125]. A des fins thérapeutiques, de nombreuses études ont tenté d'inhiber l'interaction entre VEGF-165 et Nrp-1 sur la base du complexe Nrp-1/tuftsine. Ainsi, de nombreuses molécules peptidomimétiques ont été développées et les structures cristallographiques révèlent des inhibiteurs fixés à la pointe de Nrp-1, au niveau des boucles et impliquant souvent l'aspartate 320 [72] [126].

Figure 14 Schéma des différents domaines structuraux et fonctionnels de la protéine Nrp-1											
Paramètres de mailles						Groupe d'espace	Z'	Code PDB	Domaines	Ligands	Références (DOI)
a (Å)	b (Å)	c (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)						
41,0	89,6	41,6	90	97,7	90	$P2_1$	2	3I97	Humaine-b1-EG002229	<b>8DR</b> : Acide (S)-2-(3-(benzo [c] [1,2,5] thiadiazole-4-sulfonamido) thiophène-2-carboxamido)-5-guanidinopentanoïque <b>GOL</b> : Glycérol	10.1021 / jm901755g Jarvis, A., <i>et al.</i> , <b>2010</b>
115,0	115,0	50,9	90	90	90	$P4_3$	2	4DEQ	Humaine-b1-VEGFA (exon 7/8)	<b>PO4</b> : Ion phosphate	10.1074 / jbc.M111.331140 Parker, MW., <i>et al.</i> , <b>2012</b>
65,9	67,0	74,7	90	90	90	$P2_12_12_1$	1	2QQI	Humaine-b1b2	<b>GOL</b> : Glycérol	10.1038 / sj.emboj.7601906 Appleton, BA., <i>et al.</i> , <b>2007</b>
214,1	214,1	45,5	90	90	120	$H3$	1	2QQN	Humaine-b1- Avec un Fab bloquant le VEGF	<b>EDO</b> : 1,2-Ethanédiol	
53,2	68,2	66,6	90	102,1	90	$P2_1$	1	2QQM	Humaine-a2b1b2	<b>EDO</b> : 1,2-Ethanédiol <b>CA</b> : Ion Calcium	
39,7	39,7	91,8	90	90	90	$P4_1$	1	6FMF	Humaine-b1-EG01377	<b>DUE</b> : (2{S})-2-[[3-[[5-[4-(aminométhyl)phényl]-1-benzofuran-7-yl] sulfonyl amino]thiophén-2-yl] carbonylamino]-5-carbamimidamido-acide pentanoïque <b>TFA</b> : Acide trifluoroacétique	10.1021 / acs.jmedchem.8b00210 Powell, J., <i>et al.</i> , <b>2018</b>
43,4	43,4	91,2	90	90	90	$P4_1$	1	6FMC	Humaine-b1-EG01377	<b>DUE</b> : (2{S})-2-[[3-[[5-[4-(aminométhyl)phényl]-1-benzofuran-7-yl] sulfonyl amino]thiophén-2-yl] carbonylamino]-5-carbamimidamido-acide pentanoïque	
62,8	62,8	87,9	90	90	90	$P4_12_12$	1	4RN5	Humaine-b1	<b>GOL</b> : Glycérol <b>ZN</b> : Ion zinc <b>ACT</b> : Acétate ion	Allerston, C.K., <i>et al.</i> , <b>2015</b>
62,3	62,3	86,0	90	90	90	$P4_12_12$	1	5C7G	Humaine-b1-bicine	<b>BCN</b> : Bicine <b>NA</b> : Ion de Sodium	Richard, M., <i>et al.</i> , <b>2016</b>

41,5	88,9	41,7	90	91,61	90	P2 <sub>1</sub>	2	5IYY	Humaine-b1-analogues d'Arg	<b>R40</b> : 2- [(benzyloxy) carbonyl] -L-arginine	10.1111 / febs.14405 Mota, F., <i>et al.</i> , <b>2018</b>
40,8	89,6	41,4	90	98,57	90	P2 <sub>1</sub>	2	5JGQ	Humaine-b1-analogues d'Arg	<b>6JY</b> : N-2- (benzèncarbonyl) -l-arginine <b>DMS</b> : Sulfoxyde de diméthyle	
40,7	89,2	41,6	90	98,71	90	P2 <sub>1</sub>	2	5JGI	Humaine-b1-composés M45	<b>AAG</b> : -Alpha-1-acétyl-arginine	
40,4	90,8	41,0	90	98,27	90	P2 <sub>1</sub>	2	5JHK	Humaine-b1-Arg6	<b>6K8</b> : N- (benzèncarbonyl) glycy-l-arginine	
40,1	88,9	41,0	90	96,73	90	P2 <sub>1</sub>	2	5IJR	Humaine-b1-Arg1	<b>HRG</b> : L-homoarginine <b>DMS</b> : Sulfoxyde de diméthyle	
41,6	90,5	81,9	90	99,64	90	P2 <sub>1</sub>	2	5JIX	Humaine-b1-Arg5	<b>RE5</b> : N-2- (tert-butoxycarbonyl) -L-arginine <b>DMS</b> : Sulfoxyde de diméthyle	
62,8	62,8	88,4	90	90	90	P4 <sub>1,2,2</sub>	1	1KEX	Humaine-b1	Pas de ligand	10.1016 / s0969-2126 (02) 00941-3 Lee, <i>et al.</i> , <b>2003</b>
89,9	89,9	108,3	90	90	90	P2 <sub>1,2,2,1</sub>	4	7JJC	Humaine-b1-peptide SRAS-CoV-2	<b>PEG</b> : Ether di (hydroxyéthyle) <b>GOL</b> : glycérol	10.1126 / science.abd3072 Daly, JL., <i>et al.</i> , <b>2020</b>
38,9	40,0	97,6	90	90	90	P2 <sub>1,2,2,1</sub>	1	6TKK	Humaine-b1-peptide de VEF186	<b>EDO</b> : 1,2-Ethanédiol	Eldrid, C., <i>et al.</i> , <b>2020</b>
51,0	64,7	102,3	90	90	90	P2 <sub>1,2,2,1</sub>	1	2ORZ	Rat-b1b2-tufsin	Pas de ligands	.1073 / pnas.07000431010 4 Vander Kooi, CW., <i>et al.</i> , <b>2007</b>
38,1	92,2	120,1	90	90	90	P2 <sub>1,2,2,1</sub>	1	2ORX	Rat-b1b2		
191,6	293,6	252,2	90	106,4	90	C2 <sub>1</sub>	1	4GZA	Souris- a1a2b1b2-sema3A-PlxnA2	<b>CA</b> : Ion calcium	10.1038 / nsmb.2416 Janssen, BJC., <i>et al.</i> , <b>2012</b>
245,4	245,4	47,9	90	90	120	P321	1	4GZ9	Souris-a1a2b1b2	<b>NAG</b> : 2-Acétamido-2-désoxy-bêta-d-glucopyranose <b>SO4</b> : Ion sulfate <b>EDO</b> : 1,2-Éthanédiol <b>CA</b> : Ion calcium	

### B - III. L'ACIDE BÉTULINIQUE (BTA)

Les lupanes sont une famille de triterpènes pentacycliques dont font partie, l'acide bétulinique (Figure 18). Cette molécule a comme particularité un cycle E à cinq membres, un groupe isopropylène en position 19 et une fonction alcool en position 3 [127]. La configuration tridimensionnelle de l'acide bétulinique montre une série de cycles orientés dans un même plan avec 4 groupements méthyles et une fonction acide carboxylique en position axiale, tandis que le groupement isopropylène, la fonction alcool en position 3 et la fonction méthyle en position 4 sont en position équatoriale [128].

L'acide bétulinique ou l'acide 3(3-hydroxy-lup-20 (29) -en-28-oïque est sans aucun doute le composé le plus étudié de la série des triterpènes pentacycliques de la famille du lupane. En effet, les nombreuses revues de littérature publiées sur le sujet durant les dernières années démontrent bien l'intérêt de la communauté scientifique envers ce produit naturel [129].

L'acide bétulinique, possède une activité anticancéreuse in vitro sur un large éventail de tumeurs.

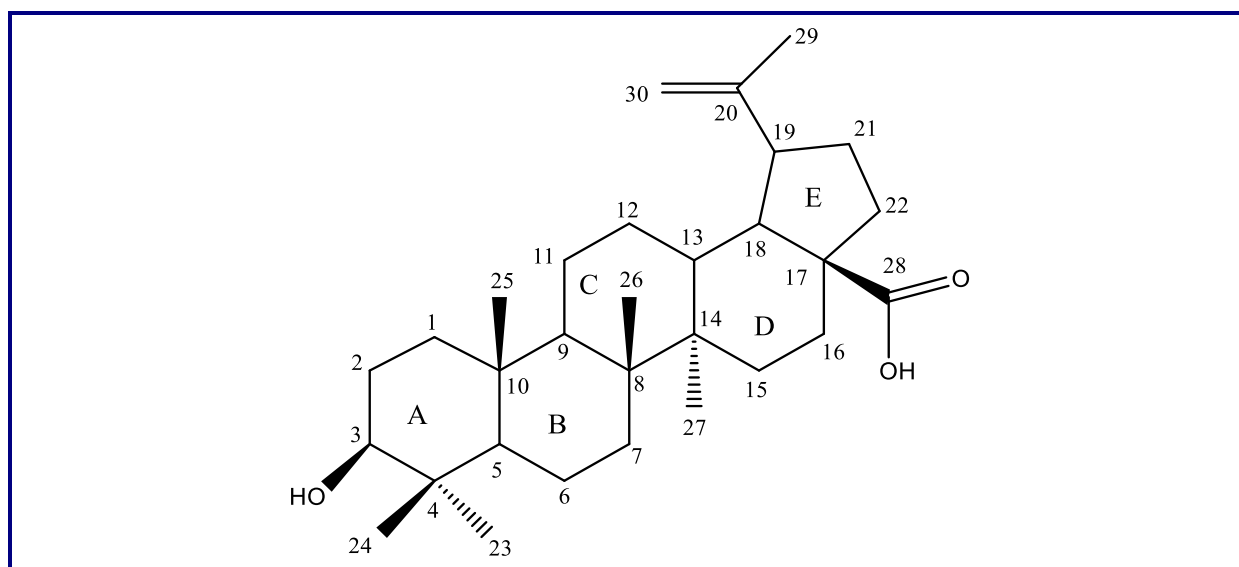


Figure 18 Structure de l'acide bétulinique.

#### B - III.1 Activités thérapeutiques

Les activités biologiques des triterpènes ont été résumées dans plusieurs articles [130] [131].

**B - III.1.a. Activités antitumorales**

L'intérêt pour les dérivés du Lupane est principalement dû à leurs activités antitumorales. L'un des composés les plus spectaculaires est sans doute l'acide bétulinique. En effet, son activité antitumorale ( $IC_{50}$  : 1.1-4.8  $\mu\text{g/mL}$ ) a d'abord été démontrée sur des mélanomes humains, un type de cancer de la peau [132]. Dans cette étude, l'acide bétulinique montre une sélectivité très forte envers les mélanomes, par rapport à d'autres lignées de cellules cancéreuses. De plus, un test préliminaire *in vivo* a été effectué sur des souris athymiques infectées par des mélanomes humains. Tous les animaux survivent avec un traitement à l'acide bétulinique, même ceux chez qui le cancer avait déjà évolué [133]. Il est intéressant de noter que des souris saines ont été traitées avec de très fortes doses d'acide bétulinique sans observer une quelconque toxicité [134]. D'ailleurs, les traitements de la leucémie, des lymphomes, des cancers de la prostate, du poumon et des ovaires par l'acide bétulinique et ses dérivés sont protégés par un brevet [135] [136].

**B - III.1.b. Autres activités thérapeutiques**

En plus des activités antitumorales et anti-VIH, l'acide bétulinique démontre plusieurs autres activités biologiques. Le parasite *Plasmodium falciparum* (malaria) est l'agent causant l'une des maladies les plus meurtrières au monde [137] [138] [139]. Des tests *in vitro* démontrent une activité pour l'acide bétulinique ( $IC_{50}$  = 10.46  $\mu\text{g/mL}$ ), mais très faible en comparaison avec la chloroquine ( $IC_{50}$  = 0.005  $\mu\text{g/mL}$ ) [140]. D'autres études démontrent que le mécanisme d'action de l'acide bétulinique est différent de celui de la chloroquine puisque des parasites résistants à ce dernier sont tout de même sensibles à l'acide bétulinique [140].

D'après une revue de littérature très poussée sur le mécanisme d'action de l'acide bétulinique, la faible biodisponibilité du composé représente sa seule limitation [141]. En effet, l'acide bétulinique n'est pas soluble dans l'eau, de sorte qu'il s'accumule dans les organes riches en graisse. Les mélanomes qui sont présents dans la peau sont donc très peu exposés à l'acide bétulinique. La modification chimique, la co-solvation et la vectorisation sont des solutions envisagées pour augmenter l'hydrosolubilité de l'acide bétulinique [142].

**B - III.2 Modifications Chimiques**

Trois cas sont à mettre en relief ici :

- Soit la molécule est modifiée afin d'obtenir une entité nouvelle, elle-même active à la façon de la molécule d'origine mais avec des propriétés physiques et chimiques différentes. Dans ce but, de nombreuses équipes ont greffé, sur la fonction hydroxyle et carboxyle de l'acide bétulinique, des substituants divers [143].
- Soit la molécule est utilisée sous forme de son sel. Ainsi l'acide bétulinique sous la forme de cholinium de bétulinate est 100 fois plus soluble que la molécule initiale dans une solution à 0,1% de DMSO. Son potentiel anticancéreux sur les cellules MCF7<sup>20</sup> et A375 est multiplié par 5 [144].
- Soit la molécule est transformée de façon à obtenir une forme transitoire appelée prodrogue dont la biotransformation après administration permet au composé initial d'exercer un effet pharmacologique [145].

**B - III.3 La co-solvation**

La solvation est le phénomène physico-chimique observé lors de la dissolution d'un composé chimique dans un solvant. Une espèce solvatée est entourée par un « bouclier » de molécules de solvant. Dans le cas de la co-solvation, un premier solvant (le co-solvant) forme ce « bouclier » de molécules de solvant. Ce sont les propriétés de l'entité formée par la molécule à solubiliser et ce co-solvant qui permettront d'obtenir une solubilisation dans le solvant visé. Dans le cas des acides hydroxy pentacycliques triterpénoïques, ces co-solvants doivent être un solubilisant de ces acides et solubles dans l'eau. Plusieurs molécules servant de co-solvant en thérapeutique humaine ont été citées: le diméthylacetamide, le diméthylsulfoxyde, l'éthanol, la glycérine, le N-méthyl-2-pyrrolidone, les PEG 300 et 400, le poloxamer 407, le propylène-glycol [146]. Le DMSO, largement utilisé avec les acides hydroxy pentacycliques triterpénoïques pose de nombreux problèmes liés à sa simple toxicité [147].

**B - III.4 La vectorisation**

La vectorisation met en jeu des vecteurs, de type soit particuliers (nano-sphères, microsphères, liposomes), soit moléculaires (virus, polymères, cyclodextrines, anticorps), soit

---

<sup>20</sup> Est le nom de la lignée de cellules tumorales mammaires la plus utilisée dans les laboratoires de recherche sur le cancer du sein

vivants (bactéries, cellules), qui permettent de transporter des principes actifs. Les cyclodextrines sont le plus souvent utilisées pour former des complexes supramoléculaires avec le principe actif à vectoriser. Leur utilisation pour former des nanoparticules, des liposomes ou des microsphères de cyclodextrine se développe de plus en plus [148] [149].

Trois formes principales peuvent exister en vectorisation :

- Les liposomes, sont des systèmes principalement fondés sur les propriétés amphiphiles des molécules qui les composent. Ces assemblages forment des monocouches ou des bicouches capables d'emprisonner les principes actifs à transporter tout en restant solubles dans des phases principalement aqueuses. Ces approches ont été utilisées pour vectoriser certains polyphénols comme le trans-resvératrol avec une visée de traitement anticancéreux [150] [151].
- Les nanoparticules, sont actuellement très utilisées dans de nombreuses formulations médicamenteuses ou cosmétiques. Les plus sûres sont les nanoparticules « métalliques » ( $\text{TiO}_2$  –  $\text{NiO}_2$ ...) alors que les nanoparticules à base de « silice » ( $\text{SiO}_2$ ) semblent beaucoup plus cytotoxiques [152]. Les nanoparticules de cyclodextrines sont une piste pleine de promesse pour l'augmentation des capacités de transport des nanoparticules. Ceci tout en augmentant leur hydrosolubilité et en stabilisant les principes actifs transportés. [153] [148].
- Les cyclodextrines, offrent quant à elles la possibilité de transporter le principe actif par l'inclusion de ce dernier dans des navettes moléculaires. Plusieurs cyclodextrines peuvent même être assemblées afin de vectoriser des molécules de taille supérieure à celle de leur cavité [154].

La vectorisation semble donc une voie d'avenir pour la mise au point de traitement efficace contre des maladies encore peu ou mal soignées de nos jours.

#### **B - IV. L'ACIDE LAURIQUE**

L'acide laurique est un acide n-alkyl monocarboxylique à douze carbones (Figure 19) présent naturellement dans certaines huiles végétales et dans le lait de certains mammifères [155] [156] [157] [158]. L'influence de l'acide laurique sur la santé et le régime alimentaire de l'homme, ainsi que son activité antimicrobienne bien connue ont fait de cet acide le sujet d'un grand nombre d'études [159]. La plupart de ces études portent sur les propriétés et le comportement de l'acide laurique en solution. A l'état solide, il présente un polymorphisme



avec 5 phases connues [160] [161] [162] [163] [164] [165]. La phase C [166], qui se produit immédiatement avant la fusion de l'acide laurique, est particulièrement intéressante, en partie

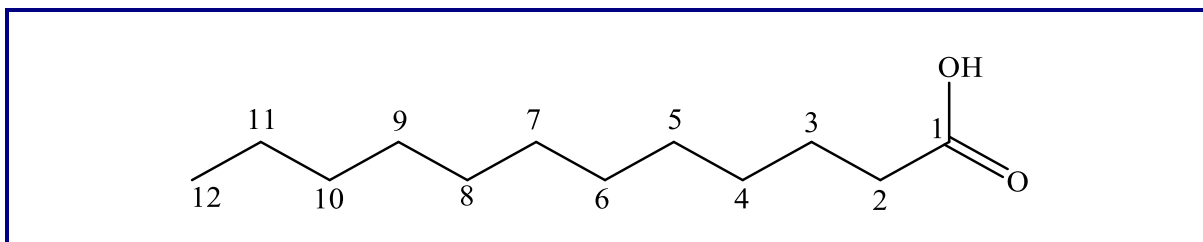


Figure 19 Structure de l'acide laurique

parce que celui-ci a une enthalpie de fusion élevée et peut donc être utilisé comme matériau de stockage d'énergie thermique [167][168]. Lorsque le nombre de carbones de ces acides augmente, on observe une augmentation inhabituelle et non monotone du point de fusion, les acides numérotés pairs fondant à des températures plus élevées [163]. Les densités, les enthalpies de sublimation et les solubilités présentent également ce comportement, les acides numérotés pairs affichant des valeurs plus élevées [163].



---

---

*Références Partie B*

---

---



- [1] P. Domínguez de María, C. Carboni-Oerlemans, B. Tuin, G. Bargeman, A. van der Meer, et R. van Gemert, « Biotechnological applications of *Candida antarctica* lipase A : State-of-the-art », *Mol. Catal. B Enzym.*, 37, 36-46, **2005**.
- [2] K. Brady et R. H. Abeles, « Inhibition of chymotrypsin by peptidyl trifluoromethyl ketones: determinants of slow-binding kinetics », *Biochemistry*, 29, 7608-7617, **1990**.
- [3] A. Mcpherson, « Current approaches to macromolecular crystallization », *Eur J Biochem.*, 189, 1-23, **1990**.
- [4] J. Uppenberg, M. T. Hansen, S. Patkar, et T. A. Jones, « The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica* », *struc.*, 2, 293-308, **1994**.
- [5] P. Fickers, F. Fudalej, MT Le Dall, S. Casaregola, C. Gaillardin *et al.*, « Identification and characterisation of LIP7 and LIP8 genes encoding two extracellular triacylglycerol lipases in the yeast *Yarrowia lipolytica* », *Fungal Genet. Biol.*, 42, 264-274, **2005**.
- [6] G. N. Ramachandran, C. Ramakrishnan, et V. Sasisekharan, « Stereochemistry of Polypeptide Chain Configurations », *Mol. Biol.*, 07, 95-99, **1963**.
- [7] J. Kumaresan, T. Kothai, et B. S. Lakshmi, « *In silico* approaches towards understanding CALB using molecular dynamics simulation and docking », *Mol Simul.*, 37, 1053-1061, **2011**.
- [8] M. Skjøt, L. De Maria, R. Chatterjee, A. Svendsen *et al.*, « Understanding the Plasticity of the  $\alpha/\beta$  Hydrolase Fold: Lid Swapping on the *Candida antarctica* Lipase B Results in Chimeras with Interesting Biocatalytic Properties », *ChemBioChem.*, 10, 520-527, **2009**.
- [9] R. D. Schmid et R. Verger, « Lipases: Interfacial Enzymes with Attractive Applications », *Angew. Chem. Int. Ed.*, 37, 26, **1998**.
- [10] P. L. A. Overbeeke, C. Govardhan, N. Khalaf, J. A. Jongejan, et J. J. Heijnen, « Influence of lid conformation on lipase enantioselectivity », *J. Mol. Catal., B Enzym.*, 10, 385-393, **2000**.
- [11] M. Martinelle, M. Holmquist, et K. Hult, « On the interfacial activation of *Candida antarctica* lipase A and B as compared with *Humicola lanuginosa* lipase », *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids.*, 1258, 272-276, **1995**.
- [12] J. Uppenberg, S. Patkar, V. Waagen, T. A. Jones *et al.*, « Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols », *Biochem.*, 34, 16838-16851, **1995**.

- [13] P. Trodler et J. Pleiss, « Modeling structure and flexibility of *Candida antarctica* lipase B in organic solvents », *BMC Struct Biol.*, 8, 9, **2008**.
- [14] P. Fuentes-Prior et G. S. Salvesen, « The protein structures that shape caspase activity, specificity, activation and inhibition », *Biochem. J.*, 384, 201-232, **2004**.
- [15] G. Fuentes, A. Ballesteros, et C. S. Verma, « Enthalpie and Entropie Contributions in the Transesterification of Sucrose: Computational Study of Lipases and Subtilisin », *J. Biomol. Struct. Dyn.*, 25, 145-155, **2007**.
- [16] T. Szegletes, W. D. Mallender, P. J. Thomas, et T. L. Rosenberry, « Substrate Binding to the Peripheral Site of Acetylcholinesterase Initiates Enzymatic Catalysis. Substrate Inhibition Arises as a Secondary Effect », *Biochimie*, 38, 122-133, **1999**.
- [17] J. Pleiss, M. Fischer, et R. D. Schmid, « Anatomy of lipase binding sites: the scissile fatty acid binding site », *Chem. Phys. Lipids.*, 93, 67-80, **1998**.
- [18] T. Bugg, "Introduction to enzyme and coenzyme chemistry", 2<sup>nd</sup>-ed., **2008**.
- [19] D. R. Yazbeck, C. A. Martinez, S. Hu, et J. Tao, « Challenges in the development of an efficient enzymatic process in the pharmaceutical industry », *Tetrahedron-Asymmetry*, 15, 2757-2763, **2004**.
- [20] M. I. Youshko, F. van Rantwijk, et R. A. Sheldon, « Enantioselective acylation of chiral amines catalysed by aminoacylase I », *Tetrahedron-Asymmetry*, 12, 3267-3271, **2001**.
- [21] H. E. Schoemaker, D. Mink et M. G. Wubbolts, « Dispelling the Myths--Biocatalysis in Industrial Synthesis », *Science*, 299, 1694-1697, **2003**.
- [22] K. A. Powell, S. W. Ramer, M. B. Tobin, P. F. Longchamp, et G. W. Huisman, « Directed Evolution and Biocatalysis », *Angew. Chem. Int. Ed.*, p. 12, **2001**.
- [23] M. T. Reetz, « Combinatorial and Evolution-Based Methods in the Creation of Enantioselective Catalysts », *Chem. Int. Ed.*, 40, 284-310, **2001**.
- [24] C. Hedfors, K. Hult, et M. Martinelle, « Lipase chemoselectivity towards alcohol and thiol acyl acceptors in a transacylation reaction », *J. Mol. Catal., B Enzym.*, 66, 120-123, **2010**.
- [25] J. Pfeffer, A. Freund, R. B. Rhlid, C.E. Hansen, M. Reuss *et al.*, « Highly Efficient Enzymatic Synthesis of 2-Monoacylglycerides and Structured Lipids and their Production on a Technical Scale », *Lipids*, 42, 947-953, **2007**.
- [26] M.-P. Bousquet-Dubouch, M. Graber, N. Sousa, S. Lamare, et M.-D. Legoy, « Alcoholysis catalyzed by *Candida antarctica* lipase B in a gas/solid system obeys a Ping Pong Bi Bi mechanism with competitive inhibition by the alcohol substrate and water », *Biochim. biophys. acta, Prot. struct. mol.*, 1550, 90-99, **2001**.

- [27] J. Nyhlén, B. Martín-Matute, A. G. Sandström, M. Bocola, et J.-E. Bäckvall, « Influence of  $\delta$ -Functional Groups on the Enantiorecognition of Secondary Alcohols by *Candida antarctica* Lipase B », *ChemBioChem.*, 9, 1968-1974, **2008**.
- [28] A. Warshel, G. Naray-Szabo, F. Sussman, et J. K. Hwang, « How do serine proteases really work? », *Biochemistry*, 28, 3629-3637, **1989**.
- [29] T. Schulz, J. Pleiss, et R. D. Schmid, « Stereoselectivity of *Pseudomonas cepacia* lipase toward secondary alcohols: A quantitative model », *Protein Sci.*, 9, 1053-1062, **2000**.
- [30] R. J. Kazlauskas, A. N. E. Weissfloch, A. T. Rappaport, et 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* », *J. Org. Chem.*, 56, 2656-2665, **1991**.
- [31] R. W. Teng, D. McManus, J. Aylward, S. Ogbourne, J. Johns, P. Parsons *et al.*, « Regioselective acylation of 3-O-angeloylingenol by *Candida antarctica* Lipase B », *Fitoterapia*, 80, 233-236, **2009**.
- [32] T. Miyazawa, M. Hamada, R. Morimoto, T. Murashima, et T. Yamada, « Highly regioselective propanoylation of dihydroxybenzenes mediated by *Candida antarctica* lipase B in organic solvents », *Tetrahedron Lett.*, 49, 175-178, **2008**.
- [33] E. García-Urdiales, N. Ríos-Lombardía, J. Mangas-Sánchez, V. Gotor-Fernández, et V. Gotor, « Influence of the Nucleophile on the *Candida antarctica* Lipase B-Catalysed Resolution of a Chiral Acyl Donor », *Chem. Eur. J. of Chem. Bio.*, 10, 1830-1838, **2009**.
- [34] D. Vollenbroich, M. Özel, J. Vater, R. M. Kamp, et G. Pauli, « Mechanism of Inactivation of Enveloped Viruses by the Biosurfactant Surfactin from *Bacillus subtilis* », *Biologicals*, 25, 289-297, **1997**.
- [35] J. Feuillard, M.-C. Jacob, F. Valensi, M. Maynadie, F. Brignole-Baudouin, et E. Kuhlein, « Clinical and biologic features of CD42CD562 malignancies », *Blood*, 99, 9, **2002**.
- [36] D. E. Goll, V. F. Thompson, R. G. Taylor, et A. Ouali, « The calpain system and skeletal muscle growth », *Can. J. Anim. Sci.*, 78, 503-512, **1998**.
- [37] F. Veuillez, J. Deshusses, et P. Buri, « Synthesis and characterization of an acylated dipeptide (Myr-Trp-Leu) with modi®ed transmucosal transport properties », *Eur J Pharm Biopharm.*, 48,21-26, **1999**.
- [38] J. L. Zambonino Infante, C. L. Cahu, et A. Peres, « Partial Substitution of Di- and Tripeptides for Native Proteins in Sea Bass Diet Improves *Dicentrarchus labrax* Larval Development », *J. Nutr.*, 127, 608-614, **1997**.

- [39] F. Ferrari, « Etude de la sélectivité d'acylation enzymatique de peptides : prédiction de la sélectivité de la lipase B de *Candida antarctica* par modélisation moléculaire et recherche de nouvelles enzymes spécifiques de type aminoacylases », Université de Lorraine, **2014**.
- [40] I. M. Qahwash, A. Boire, J. Lanning, T. Krausz, P. Pytel, et S. C. Meredith, « Site-specific Effects of Peptide Lipidation on  $\beta$ -Amyloid Aggregation and Cytotoxicity », *J. Biol. Chem.*, 282, 36987-36997, **2007**.
- [41] L. Zhang et G. Bulaj, « Converting Peptides into Drug Leads by Lipidation », *CMC*, 19, 1602-1618, **2012**.
- [42] J. T. Blanchfield, J.L. Dutton, R.C. Hogg *et al.*, « Synthesis, Structure Elucidation, in Vitro Biological Activity, Toxicity, and Caco-2 Cell Permeability of Lipophilic Analogues of  $\alpha$ -Conotoxin MII », *J. Med. Chem.*, 46, 1266-1272, **2003**.
- [43] D. Avrahami et Y. Shai, « A New Group of Antifungal and Antibacterial Lipopeptides Derived from Non-membrane Active Peptides Conjugated to Palmitic Acid », *J. Biol. Chem.*, 279, 12277-12285, **2004**.
- [44] C. Wang, H. Daimon, T. Onodera, T. Koda, et S. Sun, « A General Approach to the Size- and Shape-Controlled Synthesis of Platinum Nanoparticles and Their Catalytic Reduction of Oxygen », *Angew. Chem.*, 120, 3644-3647, **2008**.
- [45] D. Zweytick, G. Pabst, P. M. Abuja, A. Jilek, S. E. Blondelle, J. Andrä, R. Jerala, D. Monreal, G. M. de Tejada, K. Lohner, « Influence of N-acylation of a peptide derived from human lactoferricin on membrane selectivity », *Biochim. Biophys. Acta*, 1758, 1426-1435, **2006**.
- [46] C. Malaplate-Armand, S.F. Béchard, I. Youssef, V. Koziel, I. Sponne, B. Kriem, B. Leininger-Müller, J. L. Olivier, T. Oster, T. Pillot, « Soluble oligomers of amyloid- $\beta$  peptide induce neuronal apoptosis by activating a cPLA2-dependent sphingomyelinase-ceramide pathway », *Neurobiol. Dis.*, 23, 178-189, **2006**.
- [47] R. Howard, R. Howard, R. McShane, J. Lindsay *et al.*, « Donepezil and Memantine for Moderate-to-Severe Alzheimer's Disease », *N Engl J Med.*, 366, 893-903, **2012**.
- [48] M. Eeman, A. Berquand, Y. F. Dufrière, M. Paquot, S. Dufour, et M. Deleu, « Penetration of Surfactin into Phospholipid Monolayers: Nanoscale Interfacial Organization », *Langmuir*, 22, 11337-11345, **2006**.
- [49] M.S. Diamond, D. E. Staunton, A. R. de Fougères, S. A. Stacker, J. Garcia-Aguilar, M. L. Hibbs, and T. A. Springer, « ICAM-1 (CD54): A Counter-Receptor for Mac-1 (CD11b/CD18) », *J Cell Biol.*, 111, 3129-3139, **1990**.



- [50] F. Veuillez, Y. N. Kalia, Y. Jacques, J. Deshusses, et P. Buri, « Factors and strategies for improving buccal absorption of peptides », *Buri, Eur J Pharm Biopharm.*, 51, 93-109, **2001**.
- [51] J. H. Hamman, G. M. Enslin, et A. F. Kotz??, « Oral Delivery of Peptide Drugs: Barriers and Developments », *BioDrugs*, 19, 165-177, **2005**.
- [52] A. A. Mangoni et S. H. D. Jackson, « Age-related changes in pharmacokinetics and pharmacodynamics: basic principles and practical applications: Age-related changes in pharmacokinetics and pharmacodynamics », *Br. J. Clin. Pharmacol.*, 57, 6-14, **2003**.
- [53] A. Ginsburg et H. K. Schachman, « Studies on the Enzymatic Breakdown of Proteins », *J. Biol. Chem.*, 235, 108-114, **1960**.
- [54] M. Ardhaoui *et al.*, « Effect of acyl donor chain length and substitutions pattern on the enzymatic acylation of flavonoids », *J. Biotechnol.*, 110, 265-272, **2004**.
- [55] E. M. Anderson, K. M. Larsson, et O. Kirk, « One Biocatalyst—Many Applications: The Use of *Candida Antarctica B*-Lipase in Organic Synthesis », *Biocatal. Biotransformation*, 16, 181-204, **1998**.
- [56] A. Ghanem, « Trends in lipase-catalyzed asymmetric access to enantiomerically pure/enriched compounds », *Tetrahedron*, 63, 1721-1754, **2007**.
- [57] M. L. Verma, C. J. Barrow, et M. Puri, « Nanobiotechnology as a novel paradigm for enzyme immobilisation and stabilisation with potential applications in biodiesel production », *Appl. Microbiol. Biotechnol.*, 97, 23-39, **2013**.
- [58] F. Hasan, A. A. Shah, S. Javed, et A. Hameed, « Enzymes used in detergents: Lipases », *Afr. J. Biotechnol.*, 9, 31, **2010**.
- [59] R. G. Arnold, K. M. Shahani, et B. K. Dwivedi, « Application of Lipolytic Enzymes to Flavor Development in Dairy Products », *J. Dairy Sci.*, 58, 1127-1143, **1975**.
- [60] F. Hasan, A. A. Shah, et A. Hameed, « Industrial applications of microbial lipases », *Enzyme Microb. Technol.*, 39, 235-251, **2006**.
- [61] S. Bloomer, P. Adlercreutz, et B. Mattiasson, « Triglyceride interesterification by lipases. 1. Cocoa butter equivalents from a fraction of palm oil », *J Am Oil Chem Soc.*, 67, 519-524, **1990**.
- [62] E. A. Snellman et R. R. Colwell, « *Acinetobacter* lipases: molecular biology, biochemical properties and biotechnological potential », *Ind. Microbiol. Biotechnol.*, 31, 391-400, **2004**.
- [63] V. Gotor-Fernández, R. Brieva, et V. Gotor, « Lipases: Useful biocatalysts for the preparation of pharmaceuticals », *J. Mol. Catal., B Enzym.*, 40, 111-120, **2006**.

- [64] A. Ghanem et H. Y. Aboul-Enein, « Application of lipases in kinetic resolution of racemates », *Chirality*, 17, 1-15, **2005**.
- [65] A. Gog, M. Roman, M. Toşa, C. Paizs, et F. D. Irimie, « Biodiesel production using enzymatic transesterification – Current state and perspectives », *Renew. Energ.*, 39, 10-16, **2012**.
- [66] S. Tarahomjoo et I. Alemzadeh, « Surfactant production by an enzymatic method », *Enzyme Microb. Technol.*, 33, 33-37, **2003**.
- [67] B. M. Nestl, B. A. Nebel, et B. Hauer, « Recent progress in industrial biocatalysis », *Curr Opin Chem Biol.*, 15, 187-193, **2011**.
- [68] F. Ganske et U. T. Bornscheuer, « Lipase-Catalyzed Glucose Fatty Acid Ester Synthesis in Ionic Liquids », *Org. Lett.*, 7, 3097-3098, **2005**.
- [69] E. L. Soo, A. B. Salleh, M. Basri, R. N. Z. A. Rahman, et K. Kamaruddin, « Response surface methodological study on lipase-catalyzed synthesis of amino acid surfactants », *Process Biochemistry*, 39, 1511-1518, **2004**.
- [70] A. R. Leach, B. K. Shoichet, et C. E. Peishoff, « Prediction of Protein–Ligand Interactions. Docking and Scoring: Successes and Gaps », *J. Med. Chem.*, 49, 5851-5855, **2006**.
- [71] E. B. De Oliveira, C. Humeau, L. Chebil, E.R. Maia *et al.*, « A molecular modelling study to rationalize the regioselectivity in acylation of flavonoid glycosides catalyzed by *Candida antarctica* lipase B », *J. Mol. Catal., B Enzym.*, 59, 96-105, **2009**.
- [72] M. Richard, A. Chateau, C. Jelsch, C. Didierjean *et al.*, « Carbohydrate-based peptidomimetics targeting neuropilin-1: Synthesis, molecular docking study and in vitro biological activities », *Bioorg. Med. Chem.*, 24, 5315-5325, **2016**.
- [73] E. Yuriev et P. A. Ramsland, « Latest developments in molecular docking: 2010-2011 in review: Latest developments in molecular docking », *J. Mol. Recognit.*, 26, 215-239, **2013**.
- [74] L. Dettori, C. Jelsch, Y. Guiavarc'h, S. Delaunay, X. Framboisier, I. Chevalot, C. Humeau, « Molecular rules for selectivity in lipase-catalysed acylation of lysine », *Process Biochem.*, 74, 50-60, **2018**.
- [75] B. Christelle, B. D. O. Eduardo, C. Latifa, M. Elaine-Rose, M. Bernard, R.H. Evelyne, G. Mohamed, E. J. Marc, H. Catherine, « Combined docking and molecular dynamics simulations to enlighten the capacity of *Pseudomonas cepacia* and *Candida antarctica* lipases to catalyze quercetin acetylation », *J. Biotechnol.*, 156, 203-210, **2011**.

- [76] F. Ferrari, C. Parisb, B. Maigretc, C. Bidouil, S. Delaunay, C. Humeau, I. Chevalot, « Molecular rules for chemo- and regio-selectivity of *Candida antarctica* lipase B in peptide acylation reactions », *J. Mol. Catal., B Enzym.*, 101, 122-132, **2014**.
- [77] S. Takagi, T. Hirata, K. Agata, M. Mochii, G. Eguchi, et H. Fujisawa, « The A5 antigen, a candidate for the neuronal recognition molecule, has homologies to complement components and coagulation factors », *Neuron*, 7, 295-307, **1991**.
- [78] B. A. Appleton, P. Wu, J. Maloney, Y. Wu *et al.*, « Structural studies of neuropilin/antibody complexes provide insights into semaphorin and VEGF binding », *EMBO J.*, 26, 4902-4912, **2007**.
- [79] C. Pellet-Many, P. Frankel, H. Jia, et I. Zachary, « Neuropilins: structure, function and role in disease », *Biochem. Eng. J.*, 411, 211-226, **2008**.
- [80] D. Valdembri, P. T. Caswell, K. I. Anderson, I. König *et al.*, « Neuropilin-1/GIPC1 Signaling Regulates  $\alpha 5\beta 1$  Integrin Traffic and Function in Endothelial Cells », *PLoS Biol.*, 7, 1000025, **2009**.
- [81] C. Gu, B. J. Limberg, G. B. Whitaker, B. Perman, D. J. Leahy, J. S. Rosenbaum *et al.*, « Characterization of Neuropilin-1 Structural Features That Confer Binding to Semaphorin 3A and Vascular Endothelial Growth Factor 165 », *J. Biol. Chem.*, 277, 18069-18076, **2002**.
- [82] Z. He et M. Tessier-Lavigne, « Neuropilin Is a Receptor for the Axonal Chemorepellent Semaphorin III », *Cell.*, 90, 739-751, **1997**.
- [83] M. S. Hirsch et C. Wasmuth, « British HIV Association guidelines for the treatment of HIV-1-infected adults with antiretroviral therapy 2008 », *Med Mal Infect.*, 47, 266-285, **2008**.
- [84] W.S. Chang, C Y. Kang, H. Yagita, Y, S. Kim *et al.*, « Cutting Edge: Programmed Death-1/Programmed Death Ligand 1 Interaction Regulates the Induction and Maintenance of Invariant NKT Cell Anergy », *J Immunol.*, 181, 6707-6710, **2008**.
- [85] A. L. Kolodkin, D. J. Matthes, et C. S. Goodman, « The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules », *Cell.*, 75, 1389-1399, **1993**.
- [86] S. Takagi, T. Tsuji, T. Amagai, T. Takamatsu, et H. Fujisawa, « Specific cell surface labels in the visual centers of *Xenopus laevis* tadpole identified using monoclonal antibodies », *Dev. Biol.*, 122, 90-100, **1987**.

- [87] L. Dufossé, P. Galaup, S. M. Arad, G. A. Ravishankar *et al.*, « Microorganisms and microalgae as sources of pigments for food use: a scientific oddity or an industrial reality? », *Trends Food Sci Technol.*, 16, 389-406, **2005**.
- [88] T. Takahashi, T. Takahashi, A. Fournier, F. Nakamura, L. Wang, *et al.*, « Plexin-Neuropilin-1 Complexes Form Functional Semaphorin-3A Receptors », *Cell.*, 99, 59-69, **1999**.
- [89] B. Blouw, H. Song, T. Tihan, J. Bosze *et al.*, « The hypoxic response of tumors is dependent on their microenvironment », *Cancer Cell.*, 4, 133-146, **2003**.
- [90] S. Soker, H. Fidler, G. Neufeld, et M. Klagsbrun, « Characterization of Novel Vascular Endothelial Growth Factor (VEGF) Receptors on Tumor Cells That Bind VEGF165 via Its Exon 7-encoded Domain », *Biol. Chem.*, 271, 5761-5767, **1996**.
- [91] Q. Pan, Y. Chathery, Y. Wu, N. Rathore, R.K. Tong *et al.*, « Neuropilin-1 Binds to VEGF121 and Regulates Endothelial Cell Migration and Sprouting », *Biol. Chem.*, 282, 24049-24056, **2007**.
- [92] M. A. von Wronski, N. Raju, R. Pillai, N.J. Bogdan *et al.*, « Tuftsin Binds Neuropilin-1 through a Sequence Similar to That Encoded by Exon 8 of Vascular Endothelial Growth Factor », *Biol. Chem.*, 281, 5702-5710, **2006**.
- [93] H. Xiang, G.F. Wei; S. Jia, J. Huang, X. Miao, Z. Zhou *et al.*, « Microbial communities in the larval midgut of laboratory and field populations of cotton bollworm (*Helicoverpa armigera*) », *Can. J. Microbiol.*, 52, 1085-1092, **2006**.
- [94] B. Gu et J. Chen, « Enhanced microbial reduction of Cr (VI) and U(VI) by different natural organic matter fractions », *Geochim. Cosmochim. Acta*, 67, 3575-3582, **2003**.
- [95] A. L. Kolodkin, D. V. Levengood, E. G. Rowe, Y.-T. Tai, R. J. Giger, et D. D. Ginty, « Neuropilin Is a Semaphorin III Receptor », *Cell.*, 90, 753-762, **1997**.
- [96] M. Kompis, H. Pasterkamp, et G. R. Wodicka, « Acoustic Imaging of the Human Chest », *Chest.*, 120, 1309-1321, **2001**.
- [97] B. Zhang, Q. Wang, K. Wang, X. Pan, F. Lui, T. Guo *et al.*, « Identification of cotton microRNAs and their targets », *Gene*, 397, 26-37, **2007**.
- [98] G. Serini, D. Valdembri, S. Zanivan, G. Morterra, C. Brukhardt *et al.*, « Class 3 semaphorins control vascular morphogenesis by inhibiting integrin function », *Nature*, 424, 391-397, **2003**.
- [99] D. M. Goeres, L. R. Loetterle, M. A. Hamilton, R. Murga, D. W. Kirby, et R. M. Donlan, « Statistical assessment of a laboratory method for growing biofilms », *Microbiology*, 151, 757-762, **2005**.

- [100] C. W. Smith, « Endothelial adhesion molecules and their role in inflammation », *Can. J. Physiol. Pharmacol.*, 71, 76-87, **1993**.
- [101] C. S. Mantsounga, « Neuro-angiopathie diabétique : Rôle des molécules de guidance neuronale », Université Paris Descartes, **2015**.
- [102] R. E. Bachelder, A. Crago, J. Chung, M.A. Wendt, L.M. Shaw *et al.*, « Vascular Endothelial Growth Factor Is an Autocrine Survival Factor for Neuropilin-expressing Breast Carcinoma Cells », *Cancer Res.*, 61, 5736-5740, **2001**.
- [103] J. Miao, J. Kirz, et D. Sayre, « The oversampling phasing method », *Acta Crystallogr D Biol Crystallogr.*, 56, 1312-1315, **2000**.
- [104] M. Riebe, H. Haffner, C.F. Roos, W. Hansel, J. Benhelm *et al.*, « Deterministic quantum teleportation with atoms », *Nature*, 429, 734-737, **2004**.
- [105] A. Latil, I. Bièche, S. Pesche, A. Valérie, G. Fournier *et al.*, « VEGF overexpression in clinically localized prostate tumors and neuropilin-1 overexpression in metastatic forms », *Int. J. Cancer*, 89, 167–171, **2000**.
- [106] D. B. Barr, R. Allen, A.O. Olsson, R. Bravo *et al.*, « Concentrations of selective metabolites of organophosphorus pesticides in the United States population », *Environ. Res.*, 99, 314-326, **2005**.
- [107] T.-M. Karjalainen et D. Snelders, « Designing Visual Recognition for the BrandÃ », *J Prod Innov Manage*, 7, 6–22, **2010**.
- [108] K. N. Sugahara, T. Teesalu, P.P. Karmali, V.R. Kotamraju, L. Agemy *et al.*, « Coadministration of a Tumor-Penetrating Peptide Enhances the Efficacy of Cancer Drugs », *Science*, 328, 1031-1035, **2010**.
- [109] B. Kigel, A. Varshavsky, O. Kessler, et G. Neufeld, « Successful Inhibition of Tumor Development by Specific Class-3 Semaphorins Is Associated with Expression of Appropriate Semaphorin Receptors by Tumor Cells », *PLoS ONE.*, 3, 3287, **2008**.
- [110] B. Chaudhary, Y. S. Khaled, B. J. Ammori, et E. Elkord, « Neuropilin 1: function and therapeutic potential in cancer », *Cancer Immunol. Immunother.*, 63, 81-99, **2014**.
- [111] A. Turpin, « Etude des genes reprimés par le recepteur aux androgenes dans les cancers de la prostate resistants a la castration et leur evolution neuroendocrine », Université de Lille, **2021**.
- [112] J. M. Vieira, Q. Schwarz, et C. Ruhrberg, « Selective requirements for NRP1 ligands during neurovascular patterning », *Development*, 134, 1833-1843, **2007**.

- [113] P. Domingo, I. Mur, V. Pomar, H. Corominas, J. Casademont, et N. de Benito, « The four horsemen of a viral Apocalypse: The pathogenesis of SARS-CoV-2 infection (COVID-19) », *EBioMedicine*, 58, 102887, **2020**.
- [114] L. Cavée et G. Kaltenbach, « Épidémiologie de la COVID-19, focus sur le pôle de gériatrie des hôpitaux universitaires de Strasbourg », *NPG Neurol. - Psychiatr. - Geriatr.*, 21, 289-296, **2021**.
- [115] A. H. Shafaghi, F. Rokhsar Talabazar, A. Koşar, et M. Ghorbani, « On the Effect of the Respiratory Droplet Generation Condition on COVID-19 Transmission », *Fluids*, 5, 113, **2020**.
- [116] H. Flis-Richard et F. Verdonk, « Atteintes neurologiques dans l'infection au SARS-CoV-2 (COVID-19) », *Le Praticien en Anesthésie Réanimation*, 24, 186-189, **2020**.
- [117] J.-P. Changeux, Z. Amoura, F. A. Rey, et M. Miyara, « Une hypothèse nicotinique pour Covid-19 avec des implications préventives et thérapeutiques », *Biologies*, 343, 33-39, **2020**.
- [118] M. Hoffmann, K. Weber, S. Schroeder, N. Kruger *et al.*, « SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor », *Cell.*, 181, 271-280, **2020**.
- [119] L. Cantuti-Castelvetri, R. Ojha, L.D. Pedro, M. Djannatian *et al.*, « Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity », *Science*, 370, 856-860, **2020**.
- [120] Y. Lu, X. Li, D. Geng, N. Mei, P.Y. Wu, C.C. Huang *et al.*, « Cerebral Micro-Structural Changes in COVID-19 Patients – An MRI-based 3-month Follow-up Study », *EClinicalMedicine*, 25, 100484, **2020**.
- [121] J. L. Daly, B. Simonetti, K. Klein, K.E. Chen, M.K. Williamson *et al.*, « Neuropilin-1 is a host factor for SARS-CoV-2 infection », *Science*, 370, 861-865, **2020**.
- [122] C. C. Lee, A. Kreuzsch, D. McMullan, K. Ng, et G. Spraggon, « Crystal Structure of the Human Neuropilin-1 b1 Domain », *struc.*, 11, 99-108, **2003**.
- [123] T. Yelland et S. Djordjevic, « Crystal Structure of the Neuropilin-1 MAM Domain: Completing the Neuropilin-1 Ectodomain Picture », *Struc.*, 24, 2008-2015, **2016**.
- [124] C. W. Vander Kooi, M. A. Jusino, B. Perman, D. B. Neau, H. D. Bellamy, et D. J. Leahy, « Structural basis for ligand and heparin binding to neuropilin B domains », *Proc. Natl. Acad. Sci.*, 104, 6152-6157, **2007**.
- [125] J. Powell, F. Mota, D. Steadman, C. Soudy *et al.*, « Small Molecule Neuropilin-1 Antagonists Combine Antiangiogenic and Antitumor Activity with Immune



- Modulation through Reduction of Transforming Growth Factor Beta (TGF $\beta$ ) Production in Regulatory T-Cells », *J. Med. Chem.*, 61, 4135-4154, **2018**.
- [126] R. A. Hill et J. D. Connolly, « Triterpenoids », *Nat. Prod. Rep.*, 30, 1028-1065, **2013**.
- [127] M. Erba, A. Cherubini, V. V. Parail, E. Springmann, et A. Taroni, « Development of a non-local model for tokamak heat transport in L-mode, H-mode and transient regimes », *Plasma Phys. Control. Fusion.*, 39, 261-276, **1997**.
- [128] D. A. Eiznhamer, M. T. Flavin, G. J. Jesmok, J. F. Borgia *et al.*, « Effective attenuation of endotoxin-induced acute lung injury by 2,3-diacetyloxybenzoic acid in two independent animal models », *Pulm Pharmacol Ther.*, 17, 105-110, **2004**.
- [129] S. Sucharita, J. Idiculla, M. Vaz, G. Bantwal, et V. Ayyar, « Autonomic nervous system function in type 2 diabetes using conventional clinical autonomic tests, heart rate and blood pressure variability measures », *Indian J Endocr Metab.*, 15, 198, **2011**.
- [130] S. B. Mahato et G. Roy, « Review article number 67 », *Phytochimie*, 31, 2199-2249, **1992**.
- [131] S. Alakurtti, T. Mäkelä, S. Koskimies, et J. Yli-Kauhaluoma, « Pharmacological properties of the ubiquitous natural product betulin », *Eur J Pharm Sci.*, 29, 1-13, **2006**.
- [132] M. Pretolani et M. Goldman, « IL-10: a potential therapy for allergic inflammation? », *Immunology*, 18, 278-279, **1997**.
- [133] S. Ogura, S. Kashikawa, T. Miura, et I. Shioya, « Meta queries in parallel environment for databases », in *2001 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (IEEE Cat. No.01CH37233)*, Victoria, BC, Canada: IEEE, 5,707-710, **2001**.
- [134] R. A. Nagourney, « Garlic: Medicinal Food or Nutritious Medicine? », *J. Med. Food.*, 1, 13-28, **1998**.
- [135] A. Gartner, S. Milstein, S. Ahmed, J. Hodgkin, et M. O. Hengartner, « A Conserved Checkpoint Pathway Mediates DNA Damage-Induced Apoptosis and Cell Cycle Arrest in *C. elegans* », *Mol. Cell.*, 5, 435-443, **2000**.
- [136] R. Seck, M. Mansaly, A. Gassama, C. Cavé, et S. Cojean, « Synthesis and antimalarial activity of cinnamic acid derivatives », *Eur J Pharm Sci.*, 6, 453-454, **2019**.
- [137] N. Dan et S. Bhakat, « New paradigm of an old target: An update on structural biology and current progress in drug design towards plasmepsin II », *Eur. J. Med. Chem.*, 95, 324-348, **2015**.

- [138] B. R. Williamson, « Sources and Deposition Processes Linking Atmospheric Chemistry and Firn Records from Four Glacier Accumulation Zones in the McMurdo Dry Valleys, Antarctica », Université de Maine, **2006**.
- [139] G. Bringmann, G. Lang, K. Maksimenka, A. Hamm *et al.*, « Gephyromycin, the first bridged angucyclinone, from *Streptomyces griseus* strain NTK 14 », *Phytochemistry*, 66, 1366-1373, **2005**.
- [140] G. Stry, I. Klein, S. Kohlhofer, F. Koszik *et al.*, « Plasmacytoid dendritic cells express TRAIL and induce CD42 T-cell apoptosis in HIV-1 viremic patients », *Immunobiologie*, 114, 11, **2009**.
- [141] A. K. Chatterjee, J. P. Morgan, M. Scholl, et R. H. Grubbs, « Synthesis of Functionalized Olefins by Cross and Ring-Closing Metatheses », *J. Am. Chem. Soc.*, 122, 3783-3784, **2000**.
- [142] X.-H. N. Xu, W. J. Brownlow, S. V. Kyriacou, Q. Wan, et J. J. Viola, « Real-Time Probing of Membrane Transport in Living Microbial Cells Using Single Nanoparticle Optics and Living Cell Imaging », *Biochemistry*, 43, 10400-10413, **2004**.
- [143] D. Thibeault, « Évaluation de l'impact de modifications chimiques sur le potentiel anticancéreux de produits naturels de la forêt boréale québécoise », Université du Québec, **2008**.
- [144] M. Drag-Zalesinska, J. Kulbacka, J. Saczko *et al.*, « Esters of betulin and betulinic acid with amino acids have improved water solubility and are selectively cytotoxic toward cancer cells », *Bioorg. Med. Chem. Lett.*, 19, 4814-4817, **2009**.
- [145] O. Díaz, P. Bégin, B. Torrealba, E. Jover, et C. Lisboa, « Effects of noninvasive ventilation on lung hyperinflation in stable hypercapnic COPD », *Eur Respir J.*, 20, 1490-1498, **2002**.
- [146] P. F. Rebillot, H. M. Badran, G. Blalock, S. M. Bradbury, J. H. Buckley *et al.*, « Multiwavelength Observations of the Blazar Markarian 421 in 2002 December and 2003 January », *Astrophys. J.*, 641, 740-751, **2006**.
- [147] Y. Huo, J. Zhang, M. Miao, et Y. Jin, « Solvothermal synthesis of flower-like BiOBr microspheres with highly visible-light photocatalytic performances », *Appl. Catal. B: Environ.*, 112, 334-341, **2012**.
- [148] G. C. Reddy, P. Balasubramanyam, N. Salvanna, et B. Das, « Copper-Mediated C-H Activation of 1,3,4-Oxadiazoles with 1,1-Dibromo-1-alkenes Using PEG-400 as a Solvent Medium: Distinct Approach for the Alkynylation of 1,3,4-Oxadiazoles », *Eur. J. Org. Chem.*, 2012, 471-474, **2012**.



- [149] S. Savonitto *et al.*, « Early Aggressive Versus Initially Conservative Treatment in Elderly Patients With Non–ST-Segment Elevation Acute Coronary Syndrome », *Cardiovascular Interventions*, 5, 906-916, **2012**.
- [150] N. Skalko-Basnet, Z. Pavelic, et M. Becirevic-Lacan, « Liposomes Containing Drug and Cyclodextrin Prepared by the One-Step Spray-Drying Method », *Drug Dev Ind Pharm.*, 26, 1279-1284, **2000**.
- [151] D. Napierska *et al.*, « Oxidative Stress Induced by Pure and Iron-Doped Amorphous Silica Nanoparticles in Subtoxic Conditions », *Chem. Res. Toxicol.*, 25, 828-837, **2012**.
- [152] L. Cavin, V. Rabolli, L.C.J. Thomassen *et al.*, « L'évolution paléoenvironnementale des faunes de poissons du Crétacé supérieur du bassin du Tafilalt et des régions avoisinantes (Sud-Est du Maroc): implications paléobiogéographiques », *C. R. Acad. Sci. Paris ou ASCR.*, 333, 677-683, **2001**.
- [153] S. Ochoa, J. Usall, J. Cobo, X. Labad, et J. Kulkarni, « Gender Differences in Schizophrenia and First-Episode Psychosis: A Comprehensive Literature Review », *Schizophr Res Treatment.*, 2012, 1-9, **2012**.
- [154] J. Metz et M. Lassner, « Reprogramming of Oil Synthesis in Rapeseed: Industrial Applications », *Ann NY Acad Sci.*, 792, 82-90, **1996**.
- [155] K. G. Berger et N. A. Idris, « Formulation of zero- *trans* acid shortenings and margarines and other food fats with products of the oil palm », *J Am Oil Chem Soc.*, 82, 775-782, **2005**.
- [156] F. M. Dayrit, « The Properties of Lauric Acid and Their Significance in Coconut Oil », *J Am Oil Chem Soc.*, 92, 1-15, **2015**.
- [157] A. Afiq, « Date seed and date seed oil », *Food Res. Int. J.*, 9, 2035-2043, **2013**.
- [158] H. Thormar et H. Hilmarsson, « The role of microbicidal lipids in host defense against pathogens and their potential as therapeutic agents », *Chem. Phys. Lipids.*, 150, 1-11, **2007**.
- [159] T. R. Lomer, « The crystal and molecular structure of lauric acid (form A) », *Acta Cryst.*, 16, 984-988, **1963**.
- [160] T. R. Lomer et R. M. Spanswick, « A new crystalline form of lauric acid », *Acta Cryst.*, 14, 312-313, **1961**.
- [161] S. Abrahamsson, G. Larsson, et E. von Sydow, « The crystal structure of the monoclinic form of *n*-hexadecanol », *Acta Cryst.*, 13, 770-774, **1960**.
- [162] E. Moreno-Calvo, G. Gbabode, R. Cordobilla, T. Calvet, M. A. Cuevas-Dirarte, P. Negrine *et al.*, « Competing Intermolecular Interactions in the High-Temperature Solid

- Phases of Even Saturated Carboxylic Acids ( $C_{10}H_{19}O_2H$  to  $C_{20}H_{39}O_2H$ ) », *Chem. Eur. J.*, 15, 13141-13149, **2009**.
- [163] A. D. Bond, « On the crystal structures and melting point alternation of the n-alkyl carboxylic acids », *New J. Chem.*, 28, 104-114, **2004**.
- [164] V. Vand, WM Morley et TR Lomer, « La structure cristalline de l'acide laurique », *Acta Cryst.*, 4, 324-329, **1951**.
- [165] S. Takamizawa et Y. Takasaki, « Versatile Shape Recoverability of Odd-Numbered Saturated Long-Chain Fatty Acid Crystals », *Cryst. Growth Des.*, 19, 1912-1920, **2019**.
- [166] A. Sari et K. Kaygusuz, « Thermal performance of palmitic acid as a phase change energy storage material », *Energy Convers. Manag.*, 43, 863-876, **2002**.
- [167] A. Sari et A. Karaipekli, « Thermal conductivity and latent heat thermal energy storage characteristics of paraffin/expanded graphite composite as phase change material », *Appl. Therm. Eng.*, 27, 1271-1277, **2007**.
- [168] A. Sari et A. Karaipekli, « Thermal conductivity and latent heat thermal energy storage characteristics of paraffin/expanded graphite composite as phase change material », *Appl. Therm. Eng.*, 27, 1271-1277, **2007**.

---

---

*- Partie C -*  
*Objectifs de la thèse*

---

---



## Objectifs de la thèse

Au début, l'objectif fondamental de cette thèse était d'améliorer la solubilité du BTA dans l'eau, en faisant une modulation au niveau de sa fonction alcool par une réaction de couplage pseudo peptidique avec le 2-chlorocinnamoyl ou et le chlorure de phénoxyacétyle. Par la suite, le dessein consistait à cristalliser ces dérivés et, en définitive, à examiner en détail les structures cristallines.

En parallèle à ces recherches, des cristaux de laurate de sodium, acide laurique (NaLLA), ont été obtenus et leur structure a été résolue par diffractométrie des rayons X sur monocristal. Pour mieux appréhender les interactions atomiques, l'approche de la surface de Hirshfeld ainsi que l'analyse des contacts d'enrichissement ont été employés.

Cependant, l'orientation de la thèse s'est également tournée vers le domaine de la cristallographie des protéines. Au cours de cette réorientation, deux protéines ont été soumises à une étude approfondie.

Dans un premier volet, une attention particulière a été portée à la lipase B de *Candida Antarctica* (CALB). Cette enzyme suscite un vif intérêt dans le domaine des biotechnologies car elle est capable de catalyser une variété de réactions différentes pour la synthèse de composés à haute valeur ajoutée. Initialement, une cocrystallisation a été entreprise en présence de la lysine, visant à mieux appréhender les mécanismes de fixation de cette dernière sur le site actif de la protéine. L'objectif sous-jacent était l'optimisation de l'activité enzymatique de la lipase vis-à-vis de la lysine comme substrat. Une seconde phase d'étude a été menée avec la rutine qui est un autre substrat potentiel appartenant à la famille des composés phénoliques.

La seconde protéine d'intérêt a été la Neuropiline1 humaine (Nrp1). Cette protéine transmembranaire est impliquée dans de nombreux processus biologiques tels que le développement embryonnaire et la croissance des vaisseaux sanguins. Nrp1 est une cible thérapeutique potentielle car elle est supposée jouer un rôle dans le développement de certains cancers en facilitant la vascularisation tumorale. L'étude cristallographique a été réalisée sur le fragment b1 de Nrp1 (Nrp1-b1) car le fragment est soluble et renferme le site de liaison du facteur de croissance endothélial vasculaire (VEGF). Ce site est la principale cible

thérapeutique pour inhiber l'action de Nrp1. Dans les cristaux de la forme native de Nrp1-b1 ce site de liaison est obstrué par l'empilement cristallin. Il n'est donc pas possible d'utiliser les cristaux de la forme native pour tester des inhibiteurs ciblant le site liaison du VEGF. Dans le but de dégager ce site, des mutations ont été introduites au niveau d'acides aminés chargés situés à la surface de la protéine. Cela a permis à la formation d'un nouvel empilement cristallin, laissant libre le site de liaison du VEGF. Dans une perspective plus analytique, le variant a été cocrystallisé avec le peptide de séquence primaire KDKPPR. Les interactions atomiques ont été minutieusement analysées en recourant à de nouvelles approches basées sur des modèles de densité électronique multipolaire. Afin de pousser plus en avant l'analyse, des simulations de dynamique moléculaire ont été mises en œuvre, permettant de comparer la dynamique du variant et de la forme native mais aussi de la forme libre et de la forme en complexe avec le peptide KDKPPR.

---

---

*- Partie D -  
Matériel et Méthodes*

---

---





## Table des matières

- Partie D - Matériel et Méthodes .....	99
D - I  Cristallogénèse des macromolécules biologiques .....	103
D - I.1  Introduction .....	103
D - I.2  Les paramètres physico-chimiques .....	104
D - I.3  Les principaux agents précipitants .....	104
D - I.4  Techniques de cristallisation des protéines utilisées durant la thèse.....	105
D - I.5  La recherche des conditions de cristallisation.....	107
D - I.6  L'optimisation .....	108
D - I.7  L'ensemencement.....	110
D - I.8  Interprétation de la cristallisation à l'aide du diagramme de phase .....	111
D - I.9  Echantillons protéiques cristallisés .....	112
D - II  Cristallogénèse des petites molécules.....	116
D - II.1  Méthode de diffusion de vapeur ou d'évaporation de solvant .....	116
D - II.2  Méthode par Refroidissement lent .....	117
D - II.3  Petites molécules étudiées durant la thèse.....	117
D - III  Congélation des cristaux de protéine, tests de diffraction, collecte des données de diffraction et traitement des données .....	117
D - III.1  Cryoprotection et Congélation des cristaux .....	117
D - III.2  Tests de diffraction.....	118
D - III.3  Collecte des données de diffraction .....	119
D - III.4  Traitement des données .....	120
D - IV  Méthode de phasage par remplacement moléculaire.....	122
D - IV.1  Introduction au problème de la phase .....	122
D - IV.2  Coefficient de Matthews .....	123
D - IV.3  Principe du remplacement moléculaire .....	124
D - V  Construction, affinement et validation du modèle.....	124
D - V.1  Construction avec le logiciel COOT .....	124
D - V.2  Affinement par maximum de vraisemblance avec le logiciel PHENIX .....	126
D - V.3  Validation du modèle avec MOLPROBITY et PDB-Validation.....	127
Références Partie D.....	129



Cette partie décrit en détail les méthodes utilisées pour obtenir les structures cristallographiques étudiées au cours de la thèse, en mettant l'accent sur les techniques de cristallogenèse. La plupart de ces méthodes sont classiques et largement documentées dans la littérature, que ce soit dans des livres ou des revues scientifiques comme l'ouvrage « Crystallography Made Crystal Clear » de Gale RHODES, 3<sup>ième</sup> Edition (2006) ou la revue de Claude Sauter et Richard Giegé (2001). Nous avons choisi de répertorier les méthodes les plus couramment utilisées, en fournissant parfois une brève description de chacune d'entre elles. Les deux premières parties du texte traitent de la cristallogenèse des macromolécules biologiques, tandis que les deux suivantes abordent celle des petites molécules. Les paragraphes suivants présentent les méthodes de mesure en biocristallographie et leur utilisation pour obtenir un modèle cristallographique. La détermination d'une structure tridimensionnelle d'une macromolécule biologique par diffraction des rayons X se déroule en plusieurs étapes. Tout d'abord, il est nécessaire d'obtenir une quantité suffisante de la macromolécule, de manière pure et homogène (quelques milligrammes). Ensuite, la protéine est cristallisée en utilisant des agents précipitants adaptés. Un monocristal est ensuite exposé à un faisceau de rayons X pour obtenir son signal de diffraction. L'analyse des intensités diffractées permet de générer des cartes de densité électronique, à condition que le problème de la phase soit résolu. L'amélioration des phases permet d'obtenir un modèle tridimensionnel de la macromolécule d'intérêt. Plusieurs ouvrages détaillent ces méthodes, tels que "Principles of protein X-ray Crystallography" de Jan DRENTH (1994). Chacune de ces étapes peut présenter des difficultés, voire conduire à des échecs. De nos jours, la capacité à cristalliser une protéine ou à obtenir des phases initiales reste souvent un défi dans la détermination d'une structure cristallographique. Il convient de noter que les modèles générés par l'intelligence artificielle sont maintenant de qualité suffisante pour être utilisés en remplacement moléculaire dans la résolution du problème de la phase [1].

## **D - I CRISTALLOGENESE DES MACROMOLECULES BIOLOGIQUES**

### **D - I.1 Introduction**

La cristallogenèse est une étape indispensable pour déterminer la structure cristallographique d'une macromolécule biologique. Le principe de la cristallisation d'une macromolécule ne diffère pas significativement de celui des petites molécules organiques. Il implique la transition d'un état dispersé en solution vers un état solide ordonné. Un cristal est

le résultat d'un arrangement régulier périodique d'atomes, de molécules ou d'assemblages moléculaires [2]. La production d'un cristal de macromolécule biologique est un processus complexe et multifactoriel. Le processus de cristallisation dépend à la fois de facteurs biologiques tels que la stabilité conformationnelle, ainsi que de variables physico-chimiques comme les additifs, la force ionique et le pH. Dans cette partie, nous détaillerons les paramètres physico-chimiques qui influencent la cristallisation des macromolécules biologiques, ainsi que les principaux agents précipitants et les techniques de cristallisation des protéines utilisées au cours de cette thèse.

### **D - I.2 Les paramètres physico-chimiques**

De nombreux paramètres influencent la solubilité des protéines, et ces paramètres varient d'une protéine à l'autre. Afin de cristalliser une protéine, il est nécessaire de la placer dans un état de sursaturation en modifiant les paramètres suivants :

- La concentration de la protéine et son degré de pureté : il est important d'éliminer les contaminants et les structures hétérogènes (secondaires, tertiaires ou quaternaires).
- Les agents chimiques qui agissent sur la stabilité et/ou la conformation de la protéine : cela peut inclure des ions, des tampons, des ligands et des détergents.
- Le temps écoulé depuis la purification et les conditions de stockage de la protéine, car elles peuvent entraîner des dégradations.
- Le pH, qui détermine la charge nette de la protéine et la distribution des charges en surface.
- La température, qui affecte à la fois la solubilité, la stabilité de la protéine et la cinétique de cristallisation.
- Les agents précipitants et les additifs.

Il est à noter que ces paramètres sont spécifiques à chaque protéine et doivent être optimisés expérimentalement pour favoriser la cristallisation.

### **D - I.3 Les principaux agents précipitants**

L'agent précipitant joue un rôle crucial dans le processus de cristallisation, car il influence directement et principalement la solubilité de la protéine. Les agents les plus couramment utilisés peuvent être classés en trois familles distinctes en fonction de leurs effets sur les interactions protéine-solvant et protéine-protéine :

- Les polymères, tels que le polyéthylène glycol (PEG), agissent en excluant le solvant autour de la protéine.
- Les sels non chaotropiques (qui ne dénaturent pas les protéines) modifient la force ionique et augmentent les interactions hydrophobes en excluant le solvant. Exemples : le sulfate d'ammonium, les acétates ou les chlorures.
- Les solvants organiques réduisent la constante diélectrique du milieu. Exemples : le 2-méthyl-2,4-penthanediol (MPD), l'isopropanol, le dioxane, etc.

Pour bien contrôler la solubilité de la protéine, il est souvent nécessaire de combiner différents agents précipitants ou d'ajouter des additifs tels que des détergents ou des alcools. Cette approche permet d'optimiser les conditions de cristallisation.

#### **D - I.4 Techniques de cristallisation des protéines utilisées durant la thèse**

Pour cristalliser les protéines, la méthodologie repose sur deux étapes distinctes. La première étape consiste à trouver les conditions de cristallisation en utilisant une approche probabiliste. Cela implique de tester un grand nombre de solutions de cristallisation, avec des compositions variables en termes de contenu, de concentration et de pH. Ces solutions sont généralement disponibles sous forme de kits commerciaux qui offrent une large gamme de conditions de cristallisation. Deux méthodes couramment utilisées sont la diffusion de vapeur en gouttes suspendues ou assises et la méthode micro batch.

La seconde étape consiste à optimiser les conditions de cristallisation afin d'obtenir des cristaux de plus grande taille. Cela peut être réalisé en ajustant certains paramètres tels que la concentration de la protéine, la composition de la solution de cristallisation et les conditions environnementales (température, pH, etc.). L'optimisation permet d'améliorer la qualité des cristaux, ce qui facilite leur analyse structurale ultérieure.

##### **D - I.4.a. La diffusion de vapeur en gouttes suspendues ou assises :**

La technique la plus couramment utilisée actuellement est la diffusion de vapeur en gouttes suspendues ou assises [3]. Cette méthode repose sur les propriétés thermodynamiques qui entraînent un équilibre de concentration en agents précipitants entre deux solutions enfermées dans un environnement hermétiquement clos, grâce à la diffusion de la vapeur de solvant [4]. Elle est particulièrement adaptée aux expériences réalisées avec de petits volumes (entre 1 et 5  $\mu$ l). Au début de l'expérience, quelques  $\mu$ l (voire quelques nl) de protéine sont

mélangés à quelques  $\mu\text{l}$  d'agent précipitant, à proximité de la zone de sursaturation. La goutte formée est ensuite placée dans une enceinte hermétiquement scellée à l'aide de graisse minérale ou d'un joint synthétique, tout en étant à proximité d'un volume important de la solution précipitante (Figure 1).

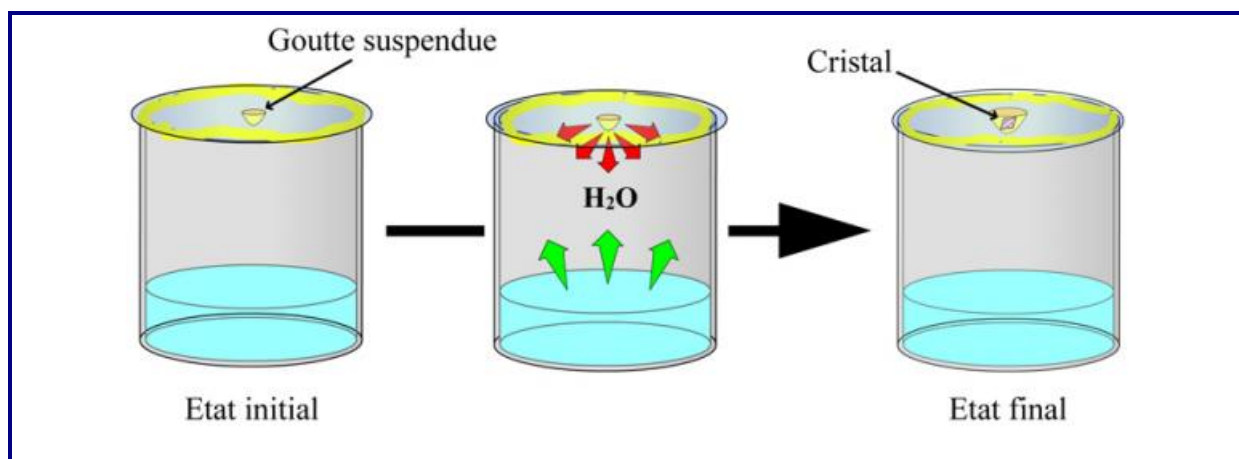


Figure 1 Technique de la goutte suspendue (Tirée de DUPUY 2005) [5]

Le retour à l'équilibre des concentrations se produit lentement par évaporation de l'eau (diffusion de la vapeur) de la goutte vers le réservoir. Cette évaporation entraîne une diminution du volume de la goutte, ce qui provoque simultanément une augmentation des concentrations en agent précipitant et en protéine. Le rôle du réservoir est simplement de maintenir une pression de vapeur constante. Au fil du temps, les concentrations de tous les constituants dans la goutte évoluent. Le réservoir agit pour concentrer progressivement la solution, ce qui peut favoriser la cristallisation.

#### D - I.4.b. Microbatch :

La méthode du microbatch a été initialement décrite comme une technique dans laquelle une petite goutte d'échantillon de protéine, combinée à l'agent de cristallisation choisi, est déposée sous une couche d'huile quasi hermétique [6] (Figure 2). Tous les réactifs impliqués dans le processus de cristallisation sont présents à des concentrations spécifiques.

Cette méthode présente plusieurs avantages, notamment la protection de la protéine contre les contaminants et l'oxydation, ainsi que la création d'un environnement propice à la cristallisation. Elle permet également l'utilisation de différentes huiles avec des propriétés variées. En effet, l'huile utilisée dans cette technique joue un rôle important dans le phénomène

de cristallisation en influençant l'évaporation de la goutte et en modifiant ainsi l'état initial de la goutte. L'huile peut également affecter la vitesse d'évaporation et, par conséquent, la cinétique de la cristallisation. Actuellement, deux types d'huiles sont utilisés. La paraffine est quasi hermétique alors que l'huile de silicone autorise la diffusion de vapeurs. Un mélange de ces deux huiles peut entraîner une diffusion partielle de l'eau, et la proportion de ces deux huiles peut faire varier le taux de cristallisation[6]. Il convient de noter que l'huile de silicone est un polymère d'unités diméthylsiloxane  $(\text{Si}(\text{CH}_3)_2\text{O})_n$  qui accélère considérablement la cristallisation par rapport à l'huile de paraffine. La technique qui consiste à faire varier la proportion entre l'huile de paraffine et l'huile de silicone est une variante de la technique microbatch. Cette variante est appelée "microbatch modifié" ou "diffusion par microbatch"[6]. La méthode du microbatch est particulièrement adaptée à l'automatisation par robot. La figure 3 présente le type de robot utilisé pour cette technique. Le microbatch est la méthode la plus ancienne et la plus simple, dans laquelle les molécules sont mélangées avec un agent de cristallisation à une certaine concentration au début de l'expérience[6]. Auparavant, le volume des essais se situait généralement dans la plage de 50  $\mu\text{L}$  à plusieurs mL. Les robots de cristallisation permettent aujourd'hui d'effectuer des essais avec des volumes réduits. Des volumes de gouttes de 500 à 100 nanolitres sont couramment utilisés.

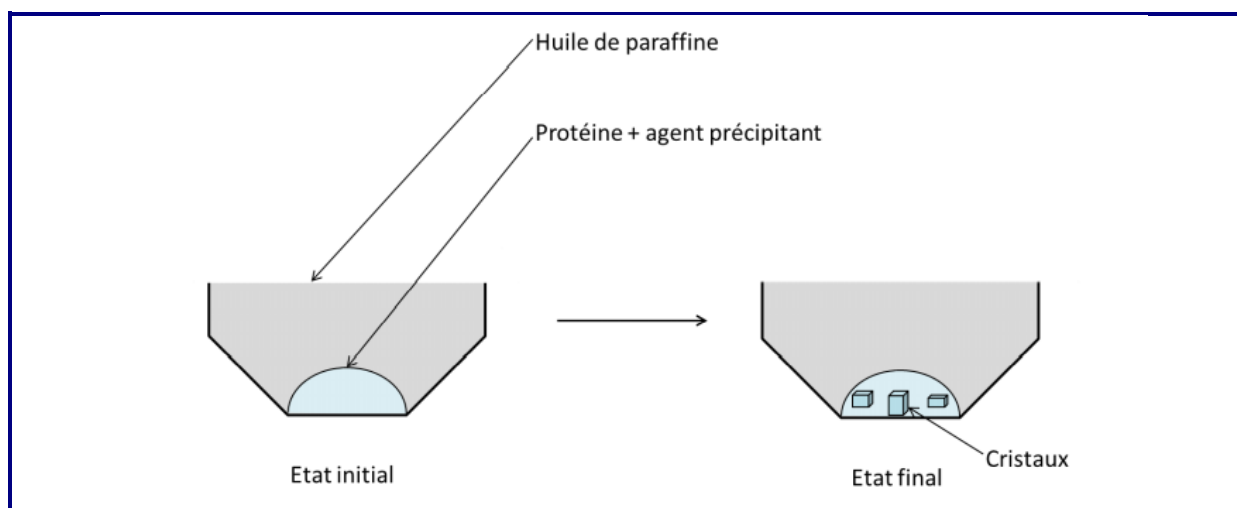


Figure 2 Technique Microbatch (Tirée de SCHWARTZ 2018)[7]

### D - I.5 La recherche des conditions de cristallisation

Les conditions de cristallisation publiées et le plus souvent stockées dans des banques de données comme la Protein Data Bank (PDB, <https://www.rcsb.org/>) ou la BMCD (Biological Macromolecule Crystallization Database <http://xpdb.nist.gov:8060/BMCD/>, Gillilan *et al.*, 1988) ou plus récemment la MPCD (Marseille Protein Crystallization Database

<http://www.crmcn.univ-mrs.fr/mpcd>), ont conduit à la commercialisation de matrices éparées de conditions de cristallisation (sparse matrices en anglais). L'utilisation de ces matrices avec les robots de cristallisation a considérablement facilité l'obtention des cristaux.

Dans le cadre de cette thèse, des matrices provenant des sociétés Molecular Dimension (matrices Structure Screen et Morpheus) et Jena Bioscience (matrices JBS Classic, JCSG, Wizard) ont été utilisées. Les criblages ont été réalisés à l'aide du robot de pipetage Oryx 8 de Douglas Instruments (Figure 3). Ce robot a été utilisé avec des plaques de cristallisation de 96 puits et a la capacité de dispenser des gouttes hautement reproductibles de 100 nL à 10  $\mu$ L au lieu d'1  $\mu$ L à quelques dizaines de  $\mu$ L en mode manuel [8].



**Figure 3** Exemple de matériel utilisé pour le criblage des conditions de cristallisation. Robot de cristallisation de type Oryx 8, disponible au laboratoire CRM2

### D - I.6 L'optimisation

Le plus souvent, les cristaux obtenus lors du criblage avec les matrices éparées ne permettent pas encore une diffraction de bonne qualité. Une étape d'optimisation est alors nécessaire. Le but de cette étape est d'ajuster finement les conditions afin de réduire la germination, d'augmenter la taille des cristaux, d'améliorer leur apparence, d'éliminer les problèmes de macle et d'accroître l'ordre cristallin [9]. Différentes stratégies peuvent être mises en œuvre pour limiter le taux de nucléation ou ralentir la croissance cristalline :

- Le « scale-up » : dans de nombreux cas, des cristaux plus gros peuvent être obtenus en augmentant simplement la quantité des produits utilisés (volumes des gouttes), moyennant quelques ajustements des conditions.



- La modification du rapport volumique entre la solution de protéine et le réservoir permet de modifier la vitesse de concentration et d'atteindre un niveau de sursaturation différent.
- La méthode de perturbation : cette approche consiste à modifier les conditions initiales de cristallisation (pH, concentrations de la protéine et/ou de l'agent précipitant) pour les déplacer vers des conditions plus favorables. Cela peut être également réalisé en ajoutant divers types d'additifs chimiques tels que des agents de solubilisation (détergents non ioniques, dioxane...), des sels chaotropiques (urée, thiocyanate de potassium...), des sels métalliques ( $Zn^{2+}$ ,  $Ni^{2+}$ ...), des composés agissant sur l'état redox de la protéine (dithiothréitol, ferricyanure...).
- L'ensemencement : cette méthode permet de ralentir la croissance cristalline et d'éviter les contraintes de germination.
- La cristallisation en gel : la cristallisation dans les gels de silice ou d'agarose peut s'avérer efficace pour réduire le taux de germination, augmenter la taille et améliorer la qualité de diffraction des cristaux[10]. Les gels permettent d'éviter les effets de convection qui perturbent la formation des cristaux.
- Le contrôle de la diffusion de vapeur : différentes techniques permettent de modifier les échanges de vapeur entre la goutte et le réservoir. Dans cette étude, une couche d'huile est placée à la surface du réservoir (Figure 4). Cette barrière limite la diffusion des vapeurs d'eau, ce qui agit directement sur la cinétique de croissance des cristaux [11]. Cette technique permet de réduire le nombre de cristaux en croissance et d'augmenter leur taille.

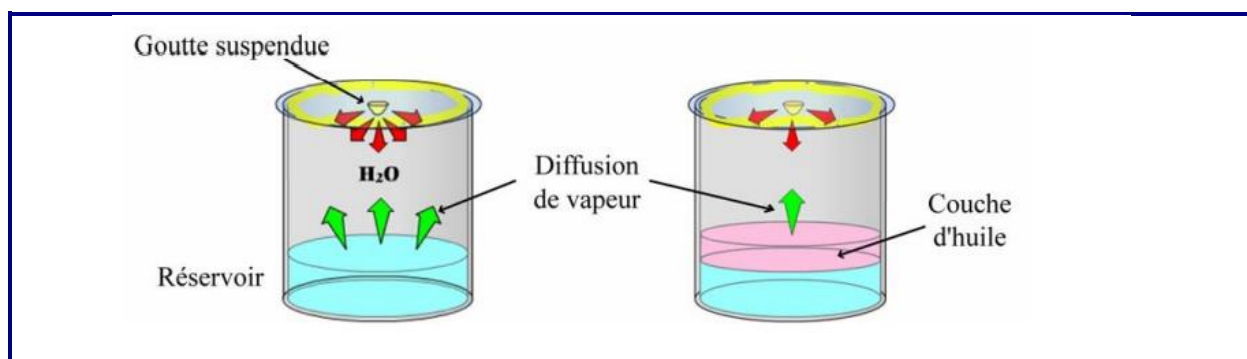


Figure 4 Limitation de la diffusion de vapeur grâce à l'emploi d'une barrière minimale constituée d'un mélange d'huile plus ou moins perméable à l'eau. (Tirée de DUPUY 2005)[5]

### D - I.7 L'ensemencement

Lorsque les cristaux sont trop petits, il est possible d'utiliser la technique d'ensemencement. En effet, la germination demande plus de sursaturation que la croissance cristalline. Cette méthode consiste à utiliser les germes microcristallins dès le début de l'expérience, obtenus en forte sursaturation, pour ensemer en faible sursaturation. Cela peut faciliter la détermination de nouvelles conditions de cristallisation, à condition que les concentrations ne conduisent pas à la fonte immédiate des germes (sous-saturation).

Les germes peuvent être soit quelques monocristaux (macro-ensemencement ou macroseeding), soit un broyat de cristaux (micro-ensemencement ou microseeding). En permettant de contrôler le nombre de germes ensemencés, le macro-ensemencement favorise la croissance de quelques grands cristaux plutôt que de nombreux petits cristaux. En dissociant les étapes de germination et de croissance, l'ensemencement permet de faire pousser les cristaux dans des conditions plus variées et moins saturées. En effet, les cristaux ont souvent une croissance plus lente et plus régulière en faible saturation.

Plusieurs techniques d'ensemencement ont été recensées par De-Yu-Zhu *et al.*[12]:

- Macro-ensemencement : Cette méthode consiste à introduire un seul cristal (de 5 à 50  $\mu\text{m}$ ) dans une solution de protéine pré-équilibrée. Cela nécessite que le cristal soit préalablement nettoyé afin de présenter des faces non contaminées, prêtes à croître.
- Micro-ensemencement : Des cristaux microscopiques (micro-semences) sont introduits dans une solution de protéine. Ces micro-semences sont créées en broyant un cristal qui ne convient pas aux expériences de diffraction. Le cristal est cassé à l'aide d'une bille en verre, créant des microcristaux qui adhèrent à un filament. Ensuite, ce filament est introduit dans des solutions de protéine pré-équilibrées. Ces microcristaux agiront comme des germes qui se développeront dans la nouvelle solution.

Le micro-ensemencement est considéré comme étant à la fois facile et efficace, et est généralement testé en premier. Cependant, il est difficile de contrôler précisément le nombre de micro-semences introduites dans la solution.

Le macro-ensemencement est un outil puissant pour augmenter la taille des cristaux. Cependant, il implique des manipulations délicates : il est nécessaire de retirer le cristal de sa

solution initiale, de le laver et de l'introduire dans une nouvelle solution. Pendant ces manipulations, il existe un risque de dommage pour le cristal.

### **D - I.8 Interprétation de la cristallisation à l'aide du diagramme de phase**

Un outil important dans la cristallisation des protéines est le diagramme de phase[13]. Celui-ci présente l'état d'une molécule en fonction de variables pertinentes du système. Pour une solution de protéine, ces variables peuvent inclure la concentration de la protéine dans la solution, la température et les caractéristiques du solvant (pH, force ionique, concentration, nature des tampons et additifs, etc.). La forme la plus courante du diagramme pour les protéines est le diagramme à deux dimensions qui représente la concentration de protéine en fonction d'un des paramètres de la solution (les autres paramètres étant maintenus constants)[14].

Dans ce diagramme (Figure 5), la courbe de solubilité sépare le domaine de sous-saturation de la protéine du domaine de sursaturation [15].

- La zone métastable indique la région où les noyaux se forment dans des délais raisonnables.
- La zone dite "labile" représente le domaine dans lequel les noyaux apparaissent spontanément. En revanche, si la sursaturation est trop élevée, des structures désordonnées, des agrégats ou des précipités peuvent se former.
- Cette zone de "précipitation" est défavorable à la formation de cristaux, car les précipités se forment plus rapidement que les cristaux eux-mêmes.

Cependant, la position de ces zones est également régie par des phénomènes cinétiques et est donc imprécise (contrairement à la courbe de solubilité).

Dans la région de sous-saturation, les cristaux se dissolvent, tandis qu'au-dessus de la courbe de solubilité, ils croissent.

Dans le cas du procédé batch, le mélange doit être placé dans la zone de nucléation, où des germes microcristallins peuvent se former. Si les conditions sont réunies, les molécules de protéines s'assemblent de manière ordonnée en noyaux, ce qui fait diminuer la concentration en protéine. Les cristaux se développent ensuite dans la zone de croissance jusqu'à atteindre leur taille maximale, à partir de laquelle le système est en équilibre.

Dans le cas de la diffusion de vapeur, que ce soit avec la technique de la goutte suspendue ou de la goutte assise, le système devient plus complexe car il est composé de deux

compartiments. Au début de l'expérience, la goutte contenant la protéine et l'agent précipitant est moins concentrée en agent précipitant que le réservoir. Les échanges de vapeur entre les deux compartiments vont entraîner une concentration croissante des macromolécules dans la goutte. En effet, l'eau s'évapore du compartiment le moins concentré (la goutte) vers le réservoir plus concentré pour équilibrer les concentrations en agent précipitant au fil du temps, et cela augmente également la concentration en protéine dans la goutte. Cette dynamique vise à favoriser la formation de germes microcristallins dans la zone de nucléation si les conditions sont adaptées, puis à faire croître les cristaux dans la zone de croissance.

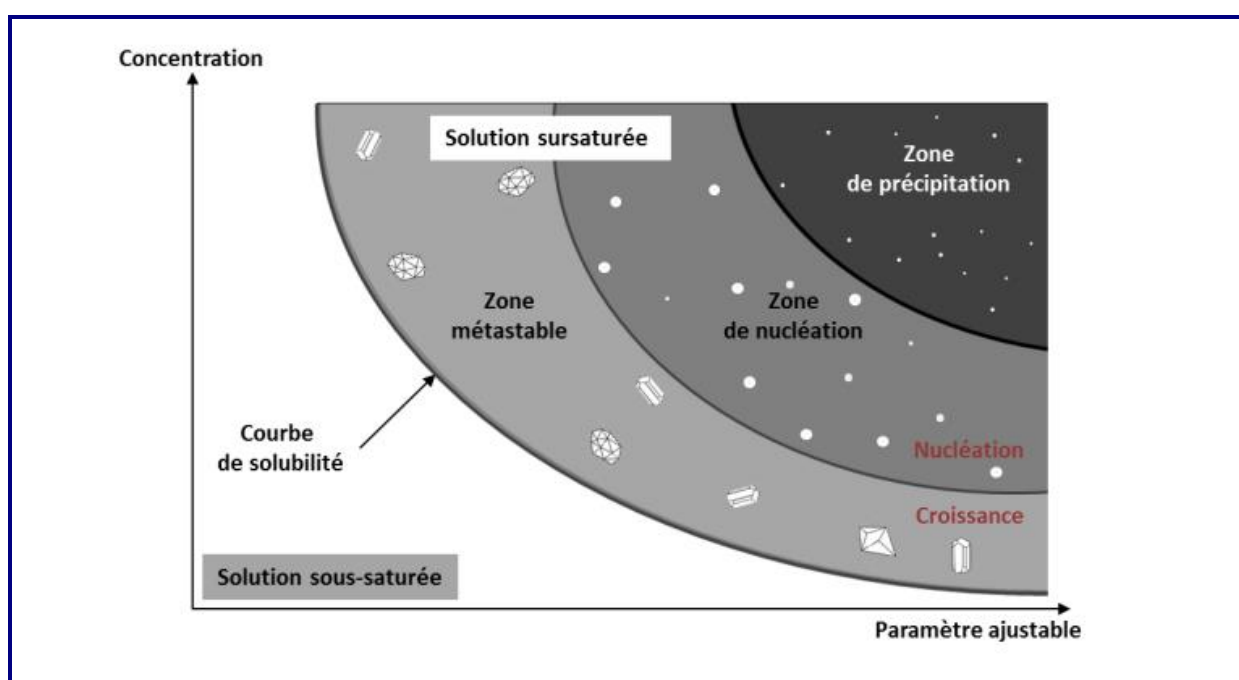


Figure 5 Diagramme de phase (Tirée de GERARD 2017)[16].

## D - I.9 Echantillons protéiques cristallisés

### D - I.9.a. Cristallisation de la protéine CALB en complexe avec la lysine ou la rutine

#### Méthodologie de cristallisation de la protéine CALB :

La lipase B de *Candida antarctica* (CALB) a été commandée chez Novozymes. L'échantillon contient de l'eau à 43,7 % (w/w), du sorbitol à 25 % (w/w), du glycérol à 25 % (w/w), de la lipase à 6 % (w/w), du benzoate de sodium à 0,20 % (w/w) et du sorbate de potassium à 0,10 % (w/w). La protéine a été utilisée en suivant les protocoles décrits ci-dessous.

Pour tester les conditions de cristallisation de la protéine CALB en complexe avec la lysine (CALB-Lysine), nous avons consulté plusieurs publications relatives à la cristallisation de cette protéine, dont l'une indiquait la composition d'une solution de cristallisation [17] [18]. Nous avons également utilisé le robot Oryx8 ainsi que des kits commerciaux (Structure Screen, Wizard, JBScreen, JCSG et Morpheus). Ainsi, nous avons découvert une nouvelle condition de cristallisation pour CALB-Lysine, qui consiste en une goutte composée de 40% (v/v) de 2-Propanol, 100 mM d'imidazole/acide chloridrique à pH 6,5 et 15% (w/v) de PEG 8000 (condition Wizard4-T45). Ensuite, nous nous sommes intéressés à cette goutte et avons réalisé une série d'optimisations autour de cette condition en utilisant la méthode de la goutte suspendue ou le microbatch.

Le robot Oryx8 a de nouveau été utilisé avec le kit JCSG (Joint Centre for Structural Genomics) pour rechercher des conditions de cristallisation de CALB en complexe avec la rutine. Des cristaux ont été obtenus en mélangeant 0,6  $\mu$ L d'échantillon (CALB + rutine) et 0,3  $\mu$ L du kit de cristallisation contenant 100 mM de citrate de sodium à pH 4,0 et 800 mM de sulfate d'ammonium (condition B01a-JCSG). Une fois les cristaux découverts, une étape d'optimisation autour de cette condition de cristallisation a été effectuée en utilisant la méthode de la goutte suspendue ou la méthode microbatch.

#### **Préparation de la solution de la rutine**

Typiquement, 10 mg de rutine ont été dissous dans 100  $\mu$ L (80  $\mu$ L d'eau + 20  $\mu$ L d'isopropanol), ce qui équivaut à une concentration de 10 mg / 0,1 mL ou 100 g/L = 0,16 mol /L. Ensuite, la solution a été chauffée à une température comprise entre 30 et 40 °C. En raison de l'insolubilité totale du composé dans le solvant, nous avons procédé à une centrifugation, puis nous avons récupéré le surnageant.

Voici ci-dessous quelques exemples de cristallisation de la CALB.

#### **Optimisation par diffusion de vapeur en goutte suspendue**

Une boîte Linbro comprenant 24 puits de quelques millilitres a été utilisée pour réaliser les expériences manuellement (Figure 6A). En utilisant une seringue contenant de la graisse silicone, un cercle entourant les 24 puits (Figure 6B) a été appliqué, créant ainsi un joint entre la lamelle et le puits afin de fermer hermétiquement le système.

Ensuite, nous avons ajouté 400  $\mu$ L de solution de cristallisation dans ces mêmes puits et préparé 24 lamelles correspondant aux 24 puits.

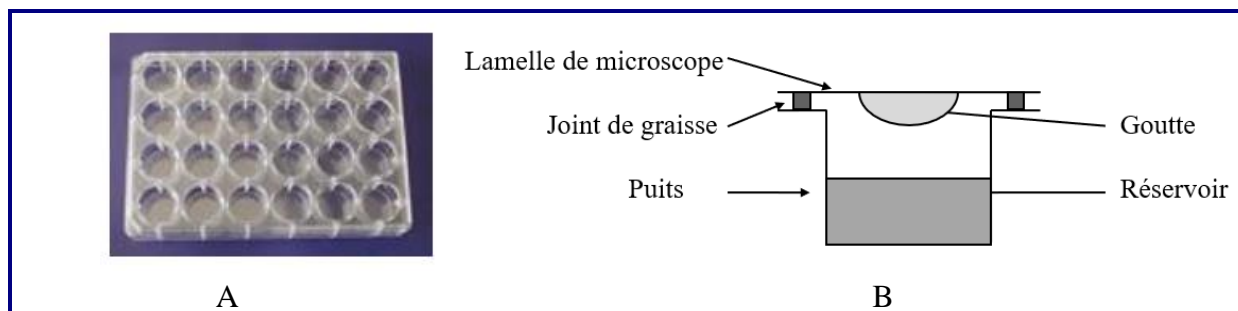


Figure 6 Photographie d'une boîte Linbro (A), Schéma de l'un des puits de la boîte Linbro (B).

- Optimisation de la condition CALB-Lysine

Pour lancer l'optimisation, des solutions basées sur la condition Wizard4-T45 ont été préparées, composées de 35%, 40%, ou 43% (v/v) de 2-propanol, 100 mM d'imidazole/acide chloridrique avec un pH variable (6.2, 6.5, 6.6 ou 7.0), et une variation du pourcentage de PEG 8000 entre 15% et 18% (w/v) (Tableau 1). La solution de Lysine/HCl utilisée est une solution commerciale de 1M.

Les gouttes des lamelles associées à ces puits contiennent :

- Colonne 1 et 2 : 2  $\mu$ L de (protéine + Lysine) et 1.2  $\mu$ L de solution du puits 1.
- Colonne 3 et 4 : 2  $\mu$ L de (protéine + Lysine) et 1.2  $\mu$ L de solution du puits 3.
- Colonne 5 et 6 : 2  $\mu$ L de (protéine + Lysine) et 1.2  $\mu$ L de solution du puits 5.

Tableau 1 Conditions de cristallisation du cristal CALB-Lys

PEG(8000)	15%	18%	15%	18%	15%	18%
2-propanol	35 %	35%	40%	40%	43%	43%
Prot + Lys						
pH 6.2						
pH 6.5						
pH 6.6						
pH 7						
	1	2	3	4	5	6

- Optimisation de la condition CALB-Rutine

Pour optimiser les cristaux obtenus lors du screening de CALB-Rutine, nous avons préparé une solution de cristallisation basée sur la condition suivante : 750 mM de sulfate d'ammonium (SA), 100 mM de citrate de sodium (CS) avec un pH de 4.2. La goutte de la

lamelle associée à ce puits contient 0.9  $\mu\text{L}$  de (protéine + Rutine), 0.45  $\mu\text{L}$  de screen B01a-JCSG et 0.33  $\mu\text{L}$  d'eau contenant de l'azide (Tableau 2).

700 mM SA CS pH 4.0	750 mM SA CS pH 4.0	800 mM SA CS pH 4.0	800 mM SA CS pH 4.0	800 mM SA CS pH 4.0	850 mM SA CS pH 4.0
700 mM SA CS pH 4.0	750 mM SA CS pH 4.0	800 mM SA CS pH 4.0	800 mM SA CS pH 4.0	800 mM SA CS pH 4.0	850 mM SA CS pH 4.0
750 mM AS CS pH 3.8	800 mM AS CS pH 3.8	850 mM AS CS pH 3.8	750 mM AS CS pH 4.0	800 mM AS CS pH 4.0	850 mM AS CS pH 4.0
750 mM AS CS pH 4.2	800 mM AS CS pH 4.2	850 mM AS CS pH 4.2	750 mM AS CS pH 4.0 4% glycérol	800 mM AS CS pH 4.0 4% glycérol	850 mM AS CS pH 4.0 4% glycérol

### Cristallisation par Microbatch

Pour augmenter les chances d'obtenir des cristaux CALB-Lysine, nous avons utilisé une deuxième méthode de cristallisation expliquée au paragraphe D - I.4: la méthode microbatch, en utilisant une plaque microbatch composée de 72 micropuits (notés de 1A à 12F) (Figure 7A). Le processus de cristallisation que nous avons suivi est le suivant : nous avons rempli les puits 1A à 8F avec une goutte d'huile de paraffine en suivant une ligne, à l'aide d'une pipette en plastique de 3 mL, de manière à former un dôme d'huile sur chaque puits.

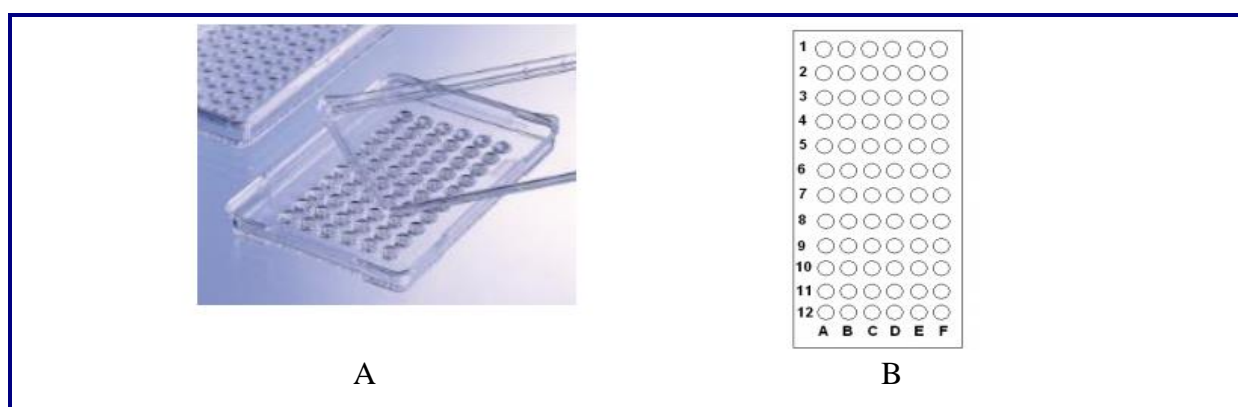


Figure 7 Plaque microbatch (A), Schéma d'une plaque micobatch (B).

À l'aide d'une micropipette, nous avons prélevé 1  $\mu\text{L}$  de chacune des 48 premières solutions commerciales du kit de cristallisation JCSG comprenant 96 conditions de cristallisation, et nous les avons déposées sous la couche d'huile (2  $\mu\text{L}$ ) de chacun des 48 puits

de la plaque. La plaque a été centrifugée à 3220 g pendant 1 minute afin de permettre la fusion de la goutte de solution commerciale avec la goutte de protéine contenant la Lysine au fond du puits (formant ainsi une sorte d'entonnoir), et pour que l'huile ferme hermétiquement ce mélange des deux gouttes (ce qui évite l'évaporation de la protéine et de la solution commerciale). Ensuite, la plaque a été placée dans la chambre à 4°C.

#### **D - I.9.b. Cristallisation d'un variant du fragment Nrp1-b1 de Nrp-1**

##### **Méthodologie de cristallisation du variant de Nrp1-b1**

L'échantillon utilisé a été produit sur la plateforme de biologie structurale intégrée de l'IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire de Strasbourg). Le variant contient 6 mutations : Glu277Lys, Glu285Lys, Asp289Lys, Glu367Lys, Lys373Glu, Lys397Glu). La protéine purifiée se trouve à une concentration de 63 mg/ml dans une solution 50 mM Tris/HCl, pH 8, 300 mM NaCl, 20 mM imidazole. Le variant a été utilisé en suivant les protocoles décrits ci-dessous. Pour tester les conditions de cristallisation du variant de NRP1-b1 avec le peptide KDKPPR, nous avons consulté plusieurs publications relatives à la cristallisation de cette protéine, dont l'une indiquait la composition d'une solution de cristallisation [19] [20] [21]. Nous avons également utilisé un robot de pipetage Oryx8 (Douglas Instruments) avec des kits commerciaux disponibles au laboratoire pour la recherche de conditions. Les essais de cristallisation de l'hexavariant ont été réalisés en présence de l'hexapeptide KDKPPR à 20 °C. Les cristaux sont apparus avec une solution de JCSG MD1-37 composée de citrate d'ammonium 0,2 M, de bis-tris 0,1 M, pH 5,5 et de 25% de PEG 3350 (w/v).

## **D - II CRISTALLOGENESE DES PETITES MOLECULES**

Selon Bernhard Spingler *et al.* (2012), les méthodes de cristallisation des petites molécules peuvent être réparties en cinq catégories [22]: diffusion de vapeur ou évaporation, refroidissement, technique de couches et gel. Dans la suite, Nous détaillerons les méthodes les plus couramment utilisées.

### **D - II.1 Méthode de diffusion de vapeur ou d'évaporation de solvant**

Une des méthodes de cristallisation d'une molécule consiste à utiliser un solvant suivi de son évaporation. La molécule (sous forme de poudre) est placée dans de petits tubes à essai auxquels on ajoute le solvant approprié. Après avoir homogénéisé la solution soluté-solvant,



soit en utilisant une plaque chauffante, soit par agitation simple, on recouvre le tube d'un film de paraffine que l'on perce pour favoriser une évaporation lente du solvant. Cette méthode permet une évaporation progressive du solvant et favorise ainsi la formation de cristaux de la molécule.

### **D - II.2 Méthode par Refroidissement lent**

Une autre méthode consiste à cristalliser la molécule dans un solvant chauffé. Dans un bécher, la molécule est placée avec le solvant et la solution est homogénéisée à l'aide d'une plaque chauffante. Ensuite, la solution soluté-solvant chauffée est transférée dans une boîte en polystyrène contenant une bouteille d'eau chaude. Cette technique permet un refroidissement lent de la solution, favorisant ainsi la formation des cristaux.

### **D - II.3 Petites molécules étudiées durant la thèse**

Dans cette thèse, l'acide bétulinique a été utilisé pour être couplé avec deux molécules, à savoir le chlorure de phénoxyacétyle et le 2-chlorocinnamoyl. Par la suite, l'acide laurique, un acide gras de type C12, a été cristallisé et a fait l'objet d'études approfondies.

## **D - III CONGELATION DES CRISTAUX DE PROTEINE, TESTS DE DIFFRACTION, COLLECTE DES DONNEES DE DIFFRACTION ET TRAITEMENT DES DONNEES**

L'étape suivant l'obtention de cristaux de macromolécules biologiques consiste à tester leur pouvoir de diffraction sous un rayonnement X. Cette capacité de diffraction est essentielle pour obtenir des données exploitables permettant de construire un modèle moléculaire. Cette étape permet également de vérifier si les cristaux obtenus sont bien des cristaux de protéines et non des cristaux de sels qui présentent un très bon pouvoir de diffraction et des paramètres métriques de réseau significativement plus petits (i.e.  $a$ ,  $b$ , et  $c$ ).

### **D - III.1 Cryoprotection et Congélation des cristaux**

Lors de l'enregistrement des données de diffraction, les cristaux sont exposés plus ou moins longtemps aux rayons X et sous un rayonnement plus ou moins brillant. Ces rayonnements entraînent la dégradation des cristaux, principalement en générant des radicaux libres (radiolyse de l'eau) et des phénomènes d'ionisation. Avec l'utilisation de sources de rayonnement de plus en plus intenses, il est devenu essentiel de protéger les cristaux contre cette dégradation. La congélation à des températures inférieures à 140 K (dans notre cas,

généralement 100 K) permet de réduire la propagation des radicaux libres dans le cristal. Cela évite également la congélation de l'eau sous forme cristalline, ce qui provoque l'apparition de taches parasites sur les images de diffraction. Cette méthode facilite également la conservation et le transport des cristaux.

Les cristaux des macromolécules biologiques contiennent une quantité significative d'eau (le plus souvent entre 30 et 80%), ce qui peut conduire à la formation de glace cristalline lors de la congélation. Cette formation de glace peut entraîner une augmentation du volume du cristal, ce qui peut le détériorer, voire le détruire. De plus, l'eau entourant le cristal peut également cristalliser lors de la congélation, ce qui peut perturber les images de diffraction en générant des anneaux de glace à des résolutions caractéristiques de 3,7 Å, 2,2 Å et 1,9 Å [23].

Pour éviter cela, il est souvent nécessaire d'ajouter un cryoprotectant par le biais d'une technique de trempage et d'échange[24]. Différents cryoprotectants peuvent être utilisés, tels que le glycérol, le saccharose, le méthylpentanediol (MPD) ou le polyéthylène glycol 400 (PEG 400). L'utilisation d'un cryoprotectant permet d'empêcher la formation de glace cristalline au profit d'un état vitreux amorphe, qui est plus transparent aux rayons X.

Dans notre cas, lors de la cristallisation de la CALB-Lysine, la solution mère contenait 40% d'isopropanol, ce qui la rendait déjà cryoprotectante. Par conséquent, il n'a pas été nécessaire d'utiliser un cryoprotectant supplémentaire (Figure 8).

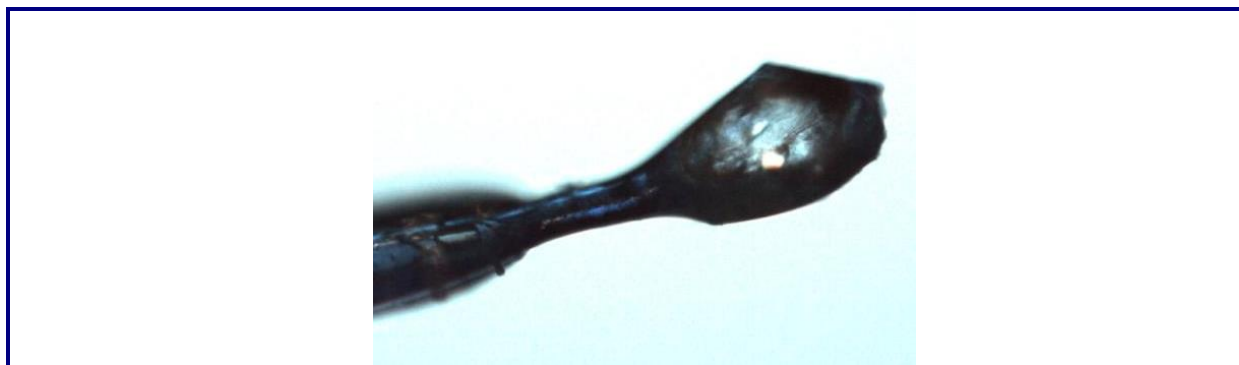


Figure 8 Cristal de CALB-Lys monté sur une Cryoloop et soumis à un flux d'azote.

### D - III.2 Tests de diffraction

Les tests de diffraction ont été réalisés à l'aide du diffractomètre automatique SuperNova (Agilent) à 4 cercles (oméga, phi, kappa et thêta), équipé d'une microsource au cuivre et d'un détecteur CCD (dispositif à transfert de charge). Cet appareil fait partie du parc de diffractomètres de la plateforme de mesures de diffraction X, gérée par le laboratoire CRM2 (Figure 9). Tous les cristaux ont été systématiquement prélevés à l'aide d'une boucle de nylon

(CryoLoop de chez Hampton Research ou Molecular Dimensions). Un cliché de diffraction préalable sans cristal dans la boucle est réalisé pour vérifier la capacité de la solution mère à vitrifier. Aucun anneau de diffraction de glace microcristalline ne doit être présent sur le cliché. Le cristal monté doit être centré sur le faisceau X à l'aide d'une tête goniométrique. Pour un test classique, le cristal est exposé pendant 20 secondes au flux de rayons X avec une oscillation de  $1^\circ$  autour de l'axe oméga du diffractomètre, qui est perpendiculaire à la direction du faisceau. Ensuite, un deuxième cliché est réalisé avec les mêmes paramètres, mais le cristal est tourné de  $90^\circ$  par rapport à l'axe oméga pour vérifier l'isotropie du signal de diffraction. Si les clichés de diffraction sont satisfaisants, deux options sont possibles. Si la résolution est suffisante (au moins  $4 \text{ \AA}$ ), une recherche de la maille peut être initiée en prenant plusieurs clichés pour indexer le réseau réciproque du cristal et caractériser ses paramètres de maille. À la fin de cette expérience, une collecte des données de diffraction peut être planifiée au laboratoire. Sinon, le cristal est conservé dans de l'azote liquide pour être mesuré ultérieurement sur une ligne synchrotron où le faisceau de rayons X est plus intense.



**Figure 9** Diffractomètre supernova équipé d'un détecteur CCD Atlas

Le diffractomètre provient de la société Agilent et est installé sur la plate-forme de mesures de diffraction X de l'Université de Lorraine (tirée de [crm2.univ-lorraine.fr](http://crm2.univ-lorraine.fr)).

### D - III.3 Collecte des données de diffraction

Classiquement au laboratoire, une stratégie de collecte est calculée à l'aide du logiciel CrysAlisPro (CrysAlisPro, Oxford Diffraction/Agilent Technologies UK Ltd, Yarnton, England), en utilisant les paramètres de maille obtenus lors d'une expérience préliminaire (Figure 10). La stratégie est optimisée pour collecter un maximum de pics de Bragg (taches de diffraction) dans un temps minimum, en tenant compte des résolutions, complétude et redondance souhaitées. La symétrie du cristal et l'estimation de son pouvoir diffractant sont

prises en compte dans le calcul. À cet effet, le programme CrysalisPro détermine une plage angulaire à couvrir pour différentes orientations du cristal. En se basant sur les données collectées lors de l'expérience préliminaire, le programme propose l'angle d'oscillation et le temps d'exposition pour chaque image. La distance entre le cristal et le détecteur est optimisée "manuellement" afin de mesurer les données à la meilleure résolution possible tout en obtenant des taches bien séparées. Le temps d'acquisition est également ajusté. La durée d'une collecte, qui varie en fonction de la symétrie du cristal, dure généralement une dizaine d'heures au maximum. Dans ce projet, les collectes ont été effectuées sur les lignes de lumière PROXIMA ([Synchrotron-soleil.fr/fr/lignes-de-lumière/proxima-2a](http://Synchrotron-soleil.fr/fr/lignes-de-lumière/proxima-2a)) du synchrotron Soleil de Paris mais aussi sur des lignes de lumières du synchrotron ESRF de Grenoble (<https://www.esrf.fr/>).

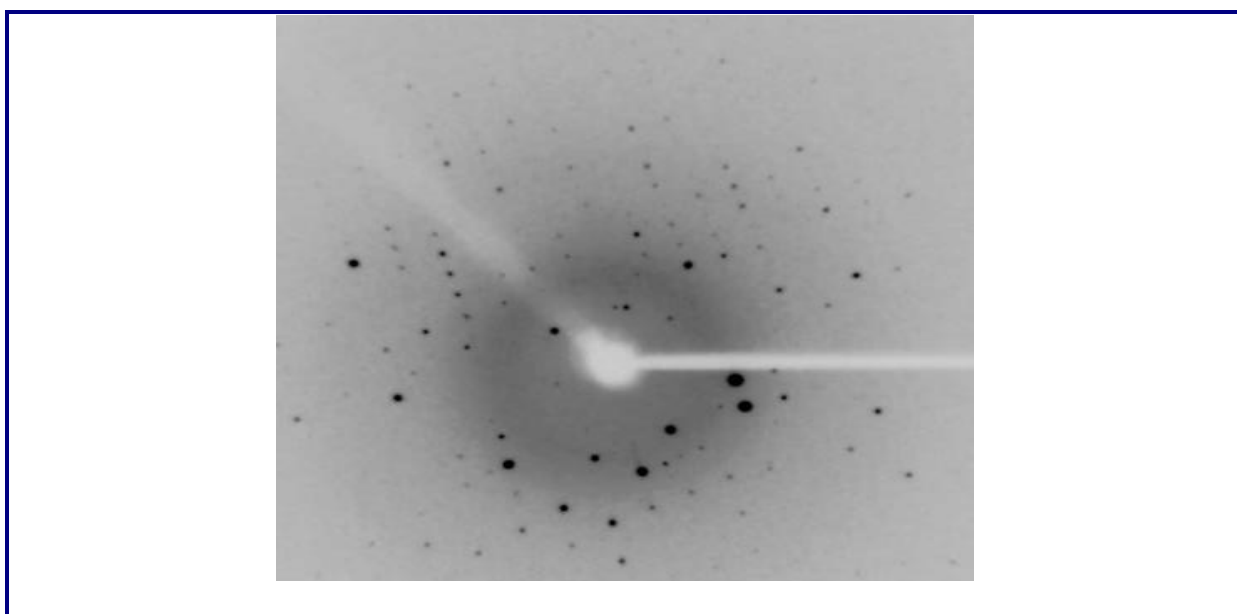


Figure 10 Cliché de diffraction d'un cristal de CALB-Lys maclé mesuré au synchrotron

#### D - III.4 Traitement des données

À l'issue d'une collecte de données cristallographiques, une série de fichiers d'images contenant les clichés de diffraction est obtenue. La prochaine étape consiste à traiter ces images afin d'indexer et d'extraire les intensités des différents pics de Bragg mesurés, également appelés réflexions. Pour le traitement des données collectées au synchrotron, nous avons utilisé le logiciel XDS (Kabsch, 2010). XDS est une suite de programmes fonctionnant sous Linux et MacOS[25]. À partir des fichiers images, XDS est capable d'effectuer les tâches suivantes :

- XYCORR : Application de facteurs de correction à chaque pixel pour corriger les défauts du détecteur.
- INIT : Estimation du niveau de bruit.

- COLSPOT : Identification des pics les plus intenses.
- IDXREF : Indexation des pics et premières estimations des paramètres de maille, du groupe ponctuel, de l'orientation du cristal et de la mosaïcité.
- DEFFPIX : Identification des régions "utiles" sur le détecteur et mise à l'écart des régions de mauvaise qualité.
- XPLAN : Établissement d'une stratégie de collecte à partir des informations. Cette étape est facultative.
- INTEGRATE : Estimation de l'intensité de chaque réflexion dans les régions définies par DEFFPIX.
- CORRECT : Application de facteurs de correction aux intensités des réflexions et à leurs écarts types.

A la fin du premier tour d'intégration des données, des indicateurs statistiques sont utilisés pour évaluer la qualité du jeu de données et estimer la résolution, permettant ainsi, le cas échéant, d'affiner l'intégration[26]. Les indicateurs utilisés sont les suivants :

- Le rapport moyen entre l'intensité et l'écart-type de la mesure ( $I/\text{sig}(I)$ ).
- Le facteur  $R_{\text{merge}}$ , qui évalue le désaccord entre l'intensité moyenne d'une réflexion  $h, k, l$  et les intensités de chaque mesure de la réflexion ou de ses équivalentes par symétrie.
- Le facteur  $R_{\text{meas}}$ , similaire au facteur  $R_{\text{merge}}$ , mais corrigé pour prendre en compte la redondance.
- La complétude, qui représente la quantité de pics de Bragg mesurés par rapport au maximum mesurable à une résolution donnée.
- La redondance, qui correspond au nombre moyen de mesures pour chaque réflexion indépendante.
- Le facteur de corrélation de Pearson, appelé  $CC_{1/2}$ , qui mesure la corrélation entre deux ensembles de données de diffraction divisés aléatoirement.

A partir de ce facteur, Karplus *et al.* ont défini un facteur  $CC^*$  comparable au facteur de corrélation entre les facteurs de structure observés et calculés[27]. Ainsi, une dernière tranche de résolution avec un  $CC^*$  supérieur ou égal à 50% contient des informations utiles pour l'affinement du modèle. Les réflexions sont ensuite mises à l'échelle et réduites à l'aide du programme XSCALE (Kabsch, 2010), qui permet également de fusionner plusieurs jeux de données. XDSCONV est utilisé pour générer un fichier au format mtz contenant l'ensemble

des réflexions associées à leur intensité et à leur déviation standard[25]. Le programme Pointless a également été utilisé pour optimiser la recherche du réseau et du groupe d'espace[28].

## D - IV METHODE DE PHASAGE PAR REMPLACEMENT MOLECULAIRE

### D - IV.1 Introduction au problème de la phase

A la fin du traitement des données, l'information relative à la structure est contenue dans la liste des intensités enregistrées. Pour décrire la structure de l'objet contenu dans la maille, on utilise la relation :

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_h \sum_k \sum_l F_{hkl} e^{-2i\pi(hx+ky+lz)}$$

Avec :

- $\rho$  la densité électronique en tout point du cristal de coordonnées fractionnaires  $x$ ,  $y$  et  $z$ .
- $V$  le volume de la maille du cristal.
- $h$ ,  $k$  et  $l$  les coordonnées des nœuds du réseau réciproque.
- $F_{hkl}$  le facteur de structure décrivant l'onde diffractée par la maille dans une direction définie par les indices  $h$ ,  $k$  et  $l$ .

Néanmoins, seuls les modules  $|F_{hkl}|$  sont connus à partir des intensités mesurées, tandis que l'information sur la phase de ces ondes est perdue. Pour retrouver la densité électronique, il est nécessaire de disposer des phases. La méthode du remplacement isomorphe consiste à introduire des atomes lourds facilement localisables dans la maille en raison de leur grand facteur de diffusion, afin d'en déduire ces phases. Une autre méthode repose sur l'utilisation de diffuseurs anomaux, tels que la sélénométhionine incorporée dans la protéine, qui présente une diffusion particulière (diffusion anormale) dépendant de la longueur d'onde utilisée.

En contraste avec ces méthodes expérimentales, appelées méthodes de perturbation, la méthode du remplacement moléculaire consiste à utiliser la structure déjà déterminée d'une molécule similaire à la molécule cible pour obtenir des phases initiales partiellement correctes. L'objectif du remplacement moléculaire est de placer dans la maille du cristal à caractériser une molécule dont la structure est connue et similaire, qui servira de point de départ pour

obtenir les phases initiales. Ces phases sont approximatives, mais suffisantes pour commencer la reconstruction progressive du modèle. Dans la mesure où de nombreuses structures de CALB étaient déjà déposées dans la Protein Data Bank (PDB) au début du projet, le remplacement moléculaire se présentait comme la méthode de choix pour résoudre le problème de la phase. Il en a été de même pour le domaine b1 de Nrp-1.

#### D - IV.2 Coefficient de Matthews

Parallèlement à la recherche d'une protéine de structure similaire, il est nécessaire de caractériser le contenu du cristal étudié. La maille élémentaire est le motif de base qui, par translation répétée, constitue le cristal. Selon la symétrie du réseau, la maille élémentaire est subdivisée en unités asymétriques qui représentent le plus petit volume du cristal ne contenant pas d'éléments liés par une symétrie cristallographique. Cependant, il est possible que différents objets liés par une symétrie non cristallographique soient présents dans la maille (par exemple, un homodimère lié par une symétrie axiale). Le contenu de la maille, représenté par  $Z_a$ , qui est le nombre de chaînes, peut être estimé à l'aide du coefficient de Matthews ( $V_M$ ), exprimé par la formule suivante[29] :

$$V_M = \frac{V}{M \cdot n \cdot z_a}$$

Avec :

- $V$  le volume de la maille élémentaire
- $M$  la masse moléculaire de la protéine
- $n$  le nombre de positions équivalentes
- $Z_a$  le nombre de molécule dans l'unité asymétrique

Sur la base de 116 cristaux de protéines, B. W. Matthews a observé que le coefficient  $V_M$  variait de 2 à 4 Å<sup>3</sup>.Da<sup>-1</sup>, ce qui correspond à un taux de solvant d'environ 27 à 65 % en utilisant un volume spécifique de 0,74 cm<sup>3</sup>.g<sup>-1</sup> pour la protéine[30]. Ainsi, il est possible d'estimer le nombre de molécules à rechercher et à placer dans l'unité asymétrique en connaissant la masse moléculaire de la protéine et la symétrie du cristal. La résolution peut également être un paramètre à prendre en compte, car une analyse statistique portant sur près de 16 000 structures cristallographiques a montré que les cristaux présentant la résolution la plus élevée ont les plus faibles valeurs de  $V_M$  [29].

### **D - IV.3 Principe du remplacement moléculaire**

Après avoir sélectionné une structure de protéine similaire dans la PDB (idéalement avec un pourcentage d'identité de séquence supérieur ou égal à 30 %), il peut être utile de préparer le modèle de structure en substituant tous les résidus différents de la protéine d'intérêt par des résidus alanine. La méthode du remplacement moléculaire vise à créer un cristal virtuel de la protéine modèle similaire au cristal de la protéine étudiée[31]. Pour ce faire, les programmes de remplacement moléculaire utilisés fonctionnent à partir des cartes de Patterson des molécules[32] [33]. Une carte de Patterson est une carte centrosymétrique représentant tous les vecteurs intra- et intermoléculaires par des pics de positions et d'intensités variables. Elle est générée par transformée de Fourier à partir des modules au carré des facteurs de structure et peut être calculée directement après l'enregistrement des clichés de diffraction.

Dans un premier temps, la fonction de rotation permet de déterminer l'opérateur (composé de trois angles) à appliquer au modèle pour l'orienter correctement dans la maille. En pratique, le logiciel superpose les cartes de Patterson du modèle et de la protéine cible sous différentes orientations afin de les faire coïncider au mieux, en tenant compte des vecteurs intramoléculaires uniquement (avec une sphère de diamètre inférieur à celui de la protéine).

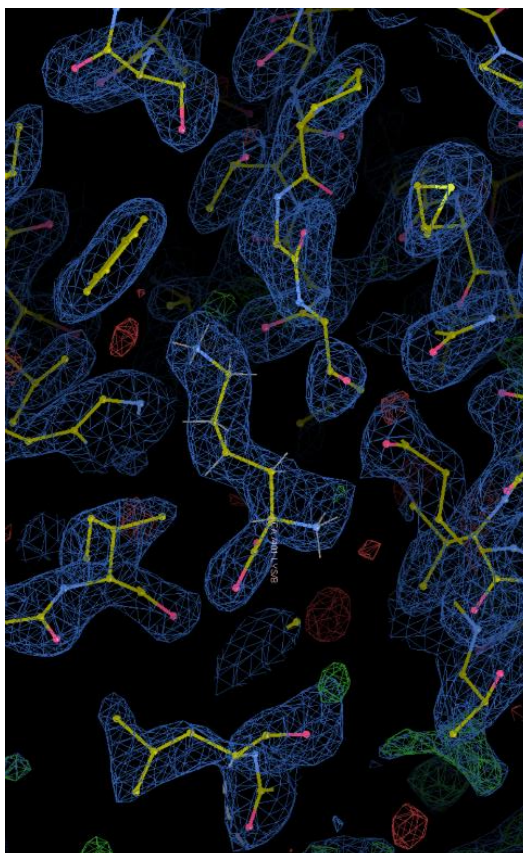
Dans un second temps, on recherche un maximum de corrélation entre la carte de Patterson du cristal et celle du modèle grâce à la fonction de translation. Celle-ci prend en compte les vecteurs intermoléculaires et détermine la translation à appliquer au modèle préorienté dans la maille du cristal. À l'issue du remplacement moléculaire et en fonction de l'algorithme utilisé, différents scores sont calculés pour évaluer la fiabilité du résultat. De plus, une inspection visuelle vérifie la cohérence de l'empilement de la structure modèle dans la maille du cristal de la protéine d'intérêt. Si les molécules du modèle se chevauchent, la solution du remplacement moléculaire n'est pas physiquement réaliste et doit être rejetée. Dans le cas d'un remplacement moléculaire réussi, ce cristal virtuel du modèle permet de calculer les modules et les phases qui, associés aux modules des facteurs de structure  $F_{\text{c}}(\mathbf{hkl})$  du cristal étudié, permettent d'obtenir différentes cartes de densité électronique de la molécule étudiée. Ensuite, une reconstruction du modèle est effectuée.

## **D - V CONSTRUCTION, AFFINEMENT ET VALIDATION DU MODELE**

### **D - V.1 Construction avec le logiciel COOT**



Après l'étape de phasage par le remplacement moléculaire, il est nécessaire de reconstruire la structure de la protéine d'intérêt en utilisant les cartes de densités électroniques. Cette reconstruction a été effectuée à l'aide du logiciel graphique Coot ("Crystallographic Oriented Object Toolkit"[34]). Ce logiciel intuitif offre différentes fonctionnalités, telles que l'affinement local des positions des atomes dans la densité, la mutation des chaînes latérales, la régularisation de la stéréochimie, etc. L'objectif est d'obtenir un modèle avec une stéréochimie correcte, en cohérence avec les données enregistrées (à partir des cartes de densité électronique). Le logiciel permet également d'ajouter les molécules d'eau liées à la protéine, les ions, ainsi que les éventuels ligands de la protéine. Deux types de cartes de densité électronique "différence" sont utilisées conjointement. La carte  $F_o-F_c$  est obtenue en soustrayant le facteur de structure  $F_c$  "calculé" à partir du modèle du facteur de structure  $F_o$  "observé" lors de l'expérience. Cette carte met en évidence les erreurs du modèle sous la forme de pics positifs (le plus souvent coloriés en vert, Figure 11) là où des électrons sont manquants, et de pics négatifs (le plus souvent coloriés en rouge) dans les régions du modèle comportant un excès d'électrons.



**Figure 11** Région du site catalytique de la structure de la CALB-Lys au cours de l'étape de reconstruction de la densité électronique.

Carte  $2F_o-F_c$  (contourée à  $1.2 \sigma$ ) en bleu et carte  $F_o-F_c$  (contourée à  $3.0 \sigma$ ) en rouge et en vert.

La carte Fo-Fc, qui met l'accent sur les erreurs du modèle, présente deux limitations. D'une part, elle ne ressemble pas à la surface moléculaire réelle, et d'autre part, elle est très bruitée, ce qui la rend difficile à interpréter lorsque le modèle comporte de nombreuses erreurs. La deuxième carte, appelée carte différence 2Fo-Fc, est beaucoup plus intuitive car elle peut être interprétée comme une carte représentant la localisation réelle des électrons (surface bleue, Figure 11). Cependant, ces cartes sont souvent difficiles à interpréter car elles sont biaisées par les erreurs sur les positions atomiques, les intensités mesurées et les phases. Il est possible de prendre en compte ces erreurs en calculant des cartes pondérées, également appelées cartes sigmaA, comme décrit dans l'ouvrage de Sherwood et Cooper, 2010. Ces cartes sont du type mFo-DFc ou 2mFoDFc, où m représente la qualité des phases pour chaque réflexion (figure de mérite) et D peut être considéré comme la fraction correcte du modèle.

### D - V.2 Affinement par maximum de vraisemblance avec le logiciel PHENIX

Après plusieurs étapes de reconstruction, une étape d'affinement de la structure peut être entreprise. Cette étape vise à affiner numériquement les coordonnées atomiques du modèle ainsi que les facteurs d'agitation de chaque atome, en tenant compte à la fois des données initiales et de la stéréochimie du modèle, à travers l'application de contraintes sur les longueurs de liaisons, les angles de liaison, la planéité, les angles de torsion, etc. Le logiciel utilisé pour l'affinement de la structure de CALB-Lysine a été phenix.refine de la suite PHENIX [35] qui fonctionne en utilisant la méthode du maximum de vraisemblance [35]. Le concept du maximum de vraisemblance repose sur la probabilité de mesurer les facteurs de structure observés  $F_{obs}$ , compte tenu du modèle à affiner. La recherche du maximum de cette probabilité est ensuite effectuée lors de l'affinement. Pour surveiller le bon déroulement de l'affinement, deux indicateurs sont utilisés : les facteurs  $R$  et  $R_{free}$ .

Le facteur  $R$  (également appelé  $R_{cryst}$ ) est un facteur d'accord exprimé selon la relation :

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}|}{\sum_{hkl} |F_{obs}|}$$

$F_{obs}$  Facteur de structure observé

$F_{calc}$  Facteur de structure calculé

$k$  Facteur d'échelle

Ce facteur  $R$  mesure le désaccord entre les modules des facteurs de structure dérivés du modèle et les modules des facteurs de structure mesurés expérimentalement. En général, ce facteur a une valeur d'environ 40 % au début de l'affinement, mais il diminue à 20 % voire moins vers la fin du cycle de construction/affinement si la résolution est suffisante.

Le deuxième facteur,  $R_{\text{free}}$ , est calculé de la même manière que le facteur  $R$ , mais à partir d'une petite portion (habituellement 5 %) du jeu de données complet qui n'est pas utilisée pendant l'affinement. Cet indicateur non biaisé permet d'éviter un sur-affinement des données. Ainsi, la qualité du modèle s'améliore lorsque les facteurs  $R$  et  $R_{\text{free}}$  diminuent de concert au cours de l'affinement.

En pratique, plusieurs cycles d'affinement et de reconstruction manuelle sont nécessaires pour obtenir une structure correctement affinée. Lorsque les facteurs  $R$  et  $R_{\text{free}}$  ne diminuent plus, que le modèle est correct d'un point de vue stéréochimique et que l'affinement n'améliore plus les cartes de densité, on considère que la structure est finalisée.

### **D - V.3 Validation du modèle avec MOLPROBITY et PDB-Validation**

La dernière étape de la résolution d'une structure consiste à valider la stéréochimie du modèle et la concordance entre les données expérimentales et le modèle. Le programme MOLPROBITY (Davis *et al.*, 2004) est intégré à PHENIX et permet de surveiller en continu le bon déroulement du processus après chaque étape de construction/affinement[36]. Ce module permet notamment de vérifier les points suivants :

- Diagramme de Ramachandran pour évaluer les angles de torsion  $\varphi$  et  $\psi$ .
- Rotamères préférentiels
- Les contraintes stériques entre atomes voisins
- Chiralité
- Planarité de groupements comme la liaison amide
- Corrélation entre le modèle et la densité électronique

L'ensemble de ces paramètres est évalué à travers un score global qui devrait être aussi bas que possible (généralement inférieur à 2) pour une structure correctement affinée. Ce score reflète la résolution cristallographique à laquelle on pourrait s'attendre pour ces paramètres[37].

A l'issue de cette première évaluation, le modèle peut ensuite être vérifié en utilisant le serveur en ligne PDB-Validation Server (<https://validate-rcsb-2.wwpdb.org/>), qui évalue la

structure et les données à travers le processus de dépôt dans la PDB. Si nécessaire, des modifications peuvent être apportées à la structure avant le dépôt définitif. Le serveur de validation vérifie globalement les mêmes paramètres que MOLPROBITY, mais il est complémentaire car il compare la structure nouvellement produite aux autres structures déjà déposées dans la PDB. Cela permet, en particulier, d'évaluer la qualité du modèle par rapport aux structures de résolution similaire.

Après avoir décrit les matériels et les méthodes utilisés dans cette étude, nous allons maintenant examiner les résultats obtenus et les interpréter dans le contexte de nos objectifs

---

*Références Partie D*

---



**Ouvrages:**

Crystals, X-rays and Proteins, Sherwood D & Cooper J, **2010**.

Principles of protein X-ray Crystallography, DRENTH J **1994**.

Crystallography Made Crystal Clear, RHODES G, 3<sup>ème</sup> Edition **2006**

**Articles de revue:**

- [1] J. Jumper, A. Evans, A. Pritzel, T. Green, M. Figumoy *et al.*, « Highly accurate protein structure prediction with AlphaFold », *Nature*, 596, 583-589, **2021**.
- [2] A. Ducruix et R. Giege, "Crystallization of Nucleic Acids and Proteins: A Practical Approach", 121-148, 2<sup>nd</sup> -éd., **1999**.
- [3] A. Wlodawer et K. O. Hodgson, « Crystallization and Crystal Data of Monellin », *Proc. Natl. Acad. Sci.*, 72, 398-399, **1975**.
- [4] N. E. Chayen, « Recent advances in methodology for the crystallization of biological macromolecules », *J. Cryst. Growth.*, 198, 649-655, **1999**.
- [5] J. Dupuy, « Etude Cristallographique de la proteine Irp1 (iron regulatory protein 1), un regulateur de l'homeostasie du Fer. », Université Joseph Fourier - Grenoble I, 54, **2005**.
- [6] A. Darcy, M. Sweeney, A. Haber, « Practical aspects of using the microbatch method in screening conditions for protein crystallization », *Methods*, 34, 323-328, **2004**.
- [7] Mathieu Schwartz, « Diversité structurale des Glutathion Transférases fongiques des classes Oméga et Xi et identification de leurs ligands par des approches cristallographiques », Université de Lorraine, 79, **2018**.
- [8] D. M. Hilty, D. C. Ferrer, M. B. Parish, B. Johnston, E. J. Callahan, et P. M. Yellowlees, « The Effectiveness of Telemental Health: A 2013 Review », *Telemed e-Health.*, 19, 444-454, **2013**.
- [9] A. McPherson et B. Cudney, « Optimization of crystallization conditions for biological macromolecules », *Acta Crystallogr. F: Struct. Biol. Commun.*, 70, 1445-1467, **2014**.
- [10] A. McPherson, « Two approaches to the rapid screening of crystallization conditions », *J. Cryst. Growth.*, 122, 161-167, **1992**.
- [11] N. E. Chayen, « A novel technique to control the rate of vapour diffusion, giving larger protein crystals », *J Appl Crystallogr.*, 30, 198-202, **1997**.
- [12] D.-Y. Zhu, Y.-Q. Zhu, Y. Xiang, et D.-C. Wang, « Optimizing protein crystal growth through dynamic seeding », *Acta Crystallogr D Biol Crystallogr.*, 61, 772-775, **2005**.

- [13] N. Asherie, « Protein crystallization and phase diagrams », *Methods*, 34, 266-272, **2004**.
- [14] E. E. G. Saridakis, P. D. Shaw Stewart, L. F. Lloyd, et D. M. Blow, « Phase diagram and dilution experiments in the crystallization of carboxypeptidase G2 », *Acta Crystallogr D Biol Crystallogr.*, 50, 293-297, **1994**.
- [15] C. Sauter, B. Lorber, D. Kern, J. Cavarelli, D. Moras, et R. Giegé, « Crystallogenesi studies on yeast aspartyl-tRNA synthetase: use of phase diagram to improve crystal quality », *Acta Crystallogr D Biol Crystallogr.*, 55, 149-156, **1999**.
- [16] C. J.J. GERARD, « Plateforme microfluidique pour l'optimisation des conditions de cristallisation des protéines », Université D'Aix-Marseille, **2017**.
- [17] J. Uppenberg, M. T. Hansen, S. Patkar, et T. A. Jones, « The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica* », *Struc.*, 2, 293-308, **1994**.
- [18] P. Strzelczyk, G. D. Bujacz, P. Kielbasiński, et J. Błaszczuk, « Crystal and molecular structure of hexagonal form of lipase B from *Candida antarctica*. », *Acta Biochim Pol.*, 63, 103-109, **2015**.
- [19] A. Jarvis, C.K. Allerston, H. J. B. Herzog *et al.*, « Small Molecule Inhibitors of the Neuropilin-1 Vascular Endothelial Growth Factor A (VEGF-A) Interaction », *J. Med. Chem.*, 53, 2215-2226, **2010**.
- [20] B. A. Appleton, P. Wu, J. Maloney, J. P. Yin *et al.*, « Structural studies of neuropilin/antibody complexes provide insights into semaphorin and VEGF binding », *EMBO J.*, 26, 4902-4912, **2007**.
- [21] A. Novoa, N. Pellegrini-Moïse, D. Bechet, M. Barberi-Heyob, et Y. Chapleur, « Sugar-based peptidomimetics as potential inhibitors of the vascular endothelium growth factor binding to neuropilin-1 », *Bioorg. Med. Chem.*, 18, 3285-3298, **2010**.
- [22] B. Spingler, S. Schnidrig, T. Todorova, et F. Wild, « Some thoughts about the single crystal growth of small molecules », *CrystEngComm.*, 14, 751-757, **2012**.
- [23] M. S. Chapman et T. Somasundaram, « De-icing: recovery of diffraction intensities in the presence of ice rings », *Acta Crystallogr D Biol Crystallogr.*, 66, 741-744, **2010**.
- [24] J. Pflugrath, « Macromolecular cryocrystallography? methods for cooling and mounting protein crystals at cryogenic temperatures », *Methods*, 34, 415-423, **2004**.
- [25] W. Kabsch, « XDS », *Acta Crystallogr D Biol Crystallogr.*, 66, 125-132, **2010**.
- [26] A. Wlodawer, W. Minor, Z. Dauter, et M. Jaskolski, « Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination », *FEBS J.*, 280, 5705-5736, **2013**.



- [27] P. A. Karplus et K. Diederichs, « Linking Crystallographic Model and Data Quality », *Science*, 336, 1030-1033, **2012**.
- [28] P. R. Evans et G. N. Murshudov, « How good are my data and what is the resolution? », *Acta Crystallogr D Biol Crystallogr.*, 69, 1204-1214, **2013**.
- [29] K. A. Kantardjieff et B. Rupp, « Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals », *Protein Sci.*, 12, 1865-1871, **2003**.
- [30] M. Blaber, X. Zhang, et B. W. Matthews, « Structural Basis of Amino Acid Helix Propensity », *Science*, 260, 1637-1640, **1993**.
- [31] I. Müller, « Bringing order to disorder: structural and functional analysis for development of strategies to target MYC », Karolinska Institutet, Stockholm, **2014**.
- [32] A. Vagin et A. Teplyakov, « Molecular replacement with *MOLREP* », *Acta Crystallogr D Biol Crystallogr.*, 66, 22-25, **2010**.
- [33] A. J. McCoy, « Solving structures of protein complexes by molecular replacement with *Phaser* », *Acta Crystallogr D Biol Crystallogr.*, 63, 32-41, **2007**.
- [34] P. Emsley et K. Cowtan, « *Coot*: model-building tools for molecular graphics », *Acta Crystallogr D Biol Crystallogr.*, 60, 2126-2132, **2004**.
- [35] P. D. Adams, F. Dimaio, N. Echols, J. J. Headd *et al.*, « PHENIX: a comprehensive Python-based system for macromolecular structure solution », *Acta Crystallogr D Biol Crystallogr.*, 66, 213-221, **2010**.
- [36] I. W. Davis, L. W. Murray, J. S. Richardson, et D. C. Richardson, « MOLPROBITY: Structure validation and all-atom contact analysis for nucleic acids and their complexes », *Nucleic Acids Res.*, 32, 615-619, **2004**.
- [37] V. B. Chen, W. Arendall, J. Headd, D. Keedy *et al.*, « MolProbity: all-atom structure validation for macromolecular crystallography », *Acta Crystallogr D Biol Crystallogr.*, 66, 12-21, **2010**.



---

---

*- . Partie E -  
Résultats et Discussions*

---

---

## Table des matières

- Partie E - Résultats et Discussions .....	135
E - I Les protéines : la Lipase B (CALB) de <i>Candida Antarctica</i> et la Neuropiline-1 (Nrp-1) humaine.....	137
E - I.1 La Lipase B (CALB) de <i>Candida Antarctica</i> .....	137
E - I.2 Neuropilin-1-peptide KDKPPR.....	143
E - II Etude des petites Molécules .....	169
E - II.1 Dérivés de l'acide Bétulinique (BTA) .....	169
E - II.2 L'acide Laurique .....	173
Références partie E .....	187

Cette dernière partie rend compte des résultats obtenus pour les protéines et les petites molécules cristallisées.

## **E - I LES PROTEINES : LA LIPASE B (CALB) DE *CANDIDA ANTARCTICA* ET LA NEUROPILINE-1 (NRP-1) HUMAINE**

### **E - I.1 La Lipase B (CALB) de *Candida Antarctica***

#### **E - I.1.a. Cristallisation du complexe CALB+Rutine**

##### **i) Résultats de Cristallisation**

Les résultats de la cristallisation du complexe CALB+Rutine sont illustrés dans la Figure 1.

Dans une première étape, la boîte "CALB+Rutine (0,6 µL) – B01a-JCSG (0,3 µL)" a été utilisée pour reproduire les conditions de cristallisation utilisées par Uppenberg *et al.* (1994). La solution de cristallisation utilisée était une solution provenant du kit commercial B01a-JCSG, contenant 100 mM de citrate de sodium à pH 4,0 et 800 mM de sulfate d'ammonium. Deux jours plus tard, de petits cristaux polycristallins de forme hexagonale ont été obtenus. Cette première boîte a servi de point de départ pour optimiser la condition de cristallisation.

La stratégie d'optimisation a consisté à modifier les concentrations en sulfate d'ammonium (750 mM à 800 mM) et le pH du tampon citrate (4,0 à 4,2) dans les réservoirs. En ce qui concerne les gouttes, des mélanges solution protéique/solution de cristallisation B01a-JCSG/eau ont été réalisés en faisant varier les proportions. Des cristaux de forme hexagonale, de taille convenable ont été obtenus pour une concentration en sulfate d'ammonium de 800 mM et un pH de 4,0. La goutte de la lamelle associée à ce puits contenait 0,9 µL de CALB+Rutine, 0,45 µL de screen B01a-JCSG et 0,33 µL d'eau contenant de l'azide. Cette seconde étape a donc permis d'optimiser la taille des cristaux (Figure 1).

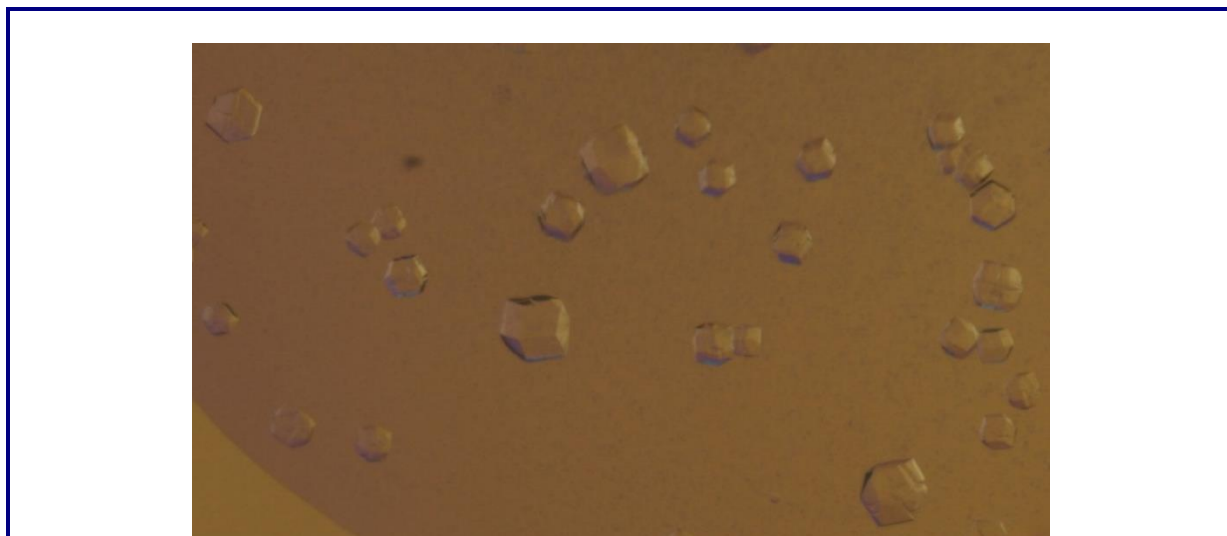


Figure 1 : Photographie des cristaux du complexe CALB+Rutine

ii) Essais de diffraction

L'étape suivante après l'obtention des cristaux de protéine consiste à tester leur pouvoir diffractant. Ce paramètre est fondamental pour obtenir des données exploitables afin de construire un modèle. Cette étape permet également de constater que les résultats obtenus sont bien des cristaux protéiques et non des cristaux de sels. En effet les clichés de diffraction d'une macromolécule et d'un sel se distinguent facilement. Les tests de diffraction des cristaux ont été réalisés sur la plateforme de mesures de diffraction et de diffusion des rayons X (PMD2X) du CRM<sup>2</sup>. L'appareil utilisé a été le diffractomètre automatique SuperNova (Rigaku) équipé d'une micro-source au cuivre et d'un détecteur CCD (Charge Coupled Device). Tous les cristaux ont été systématiquement prélevés à l'aide d'une CryoLoop (Hampton Research ou Molecular Dimensions). Cependant, ces tests ont révélé un pouvoir diffractant faible, voire inexistant, ce qui a conduit à l'arrêt des étapes suivantes de caractérisation.

**E - I.1.b. Cristallisation de la protéine CALB+Lysine**

i) Résultats de Cristallisation

Les résultats de la cristallisation CALB+Lysine sont illustrés dans la Figure 2.

Dans une première étape, la boîte "CALB+Lysine (2  $\mu$ L) – Wizard4-T45 (0,3  $\mu$ L)" a été utilisée pour reproduire les conditions de cristallisation utilisées par Stauch *et al.* (2015). La concentration de la solution de Lysine/HCl était à 1M. La solution de cristallisation utilisée

était une solution provenant du kit commercial Wizard4-T45, contenant 40% (v/v) de 2-Propanol, 100 mM d'imidazole/acide chlorhydrique à pH 6,5 et 15% (w/v) de PEG 8000. Deux jours plus tard, de petits cristaux polycristallins ont été obtenus. Cette première boîte a servi de point de départ pour obtenir des cristaux monocristallins et d'une taille suffisante pour les expériences de diffraction.

La stratégie d'optimisation a consisté à varier la concentration du 2-propanol (35% => 43%) et en ajustant le pH de l'imidazole/acide chloridrique (6,2 =>7,0) dans les réservoirs. En ce qui concerne les gouttes, des mélanges solution protéique/solution des puits ont été réalisés en faisant varier les solutions prises dans les différents puits. Des cristaux sous forme de plaquette, de taille convenable ont été obtenus pour une concentration en imidazole/acide chlorhydrique de 1 M et un pH de 6,2. La goutte de la lamelle associée à ce puits contenait 2  $\mu$ L de CALB+Lysine/HCl et 1.2  $\mu$ L de solution du puits. Cette seconde étape a donc permis d'optimiser la taille des cristaux (Figure 2).



Figure 2 : Micro-Photographie des cristaux maclés CALB+Lys

## ii) Essais de diffraction

Le diffractomètre utilisé pour les essais de diffraction des cristaux CALB+Rutine a été le même que celui utilisé pour les cristaux CALB+Lysine. Un cliché de diffraction préliminaire a été réalisé en utilisant uniquement la liqueur mère afin de vérifier si elle convenait. Aucun anneau de diffraction provenant de la glace microcristalline ne doit être visible sur le cliché. Cela a été le cas. Le cristal monté a été centré sur le faisceau X à l'aide d'une tête goniométrique. Le cristal a été exposé pendant 20 secondes au flux de rayons X avec une oscillation de 1° autour de l'axe oméga du diffractomètre, perpendiculaire à la direction du

faisceau. Ensuite, un deuxième cliché a été enregistré où le cristal a été tourné de 90° par rapport à l'axe oméga pour vérifier l'isotropie du signal de diffraction. Les clichés étant satisfaisants (résolution meilleure que 4 Å), une recherche de paramètres de maille a été tentée en prenant plusieurs clichés afin d'indexer le réseau réciproque du cristal. À la fin de cette expérience, une collecte de données de diffraction a été planifiée au synchrotron SOLEIL (Palaiseau, France) sur la ligne de lumière PROXIMA-2. Pour cela, le cristal a été conservé dans un cryo-conservateur contenant de l'azote liquide.

### iii) Mesures

Les collectes ont été réalisées sur les lignes de lumière PROXIMA-2A (<https://www.synchrotron-soleil.fr/fr/lignes-de-lumiere/proxima-2a>) du synchrotron Soleil. Cette ligne de lumière, située à la sortie d'un onduleur, est équipée d'un détecteur de rayons X Eiger-X 16M (Dectris Ltd.). La brillance du rayonnement synchrotron (au moins 1 milliard de fois plus intense que celle de la micro-source de laboratoire) et le détecteur « shutter less » (sans fermeture d'obturateur entre deux clichés de diffraction) permettent la réalisation de collectes de données en quelques minutes voire moins. Les statistiques de la meilleure collecte sont résumées Tableau 1 et la Figure 3 présente un cliché de diffraction.

**Tableau 1 : Données de diffraction du cristal CALB+Lysine**

	CALB+Lys		
Synchrotron Beamline	PROXIMA2-SOLEIL		
Space group	$P 2_1$		
$a, b, c$	45.33 107.49 68.25		
$\alpha, \beta, \gamma$	90.0 90.0 90.0		
Resolution	45.33 – 1.87		
Resolution ring	45.33 – 1.87	45.33 – 8.97	1.91 – 1.87
$R_{\text{merge}}$ (%)	15.7	4.8	1.349
$R_{\text{meas}}$ (%)	17.1	5.3	1.477
$R_{\text{pim}}$ (%)	6.6	2.1	58.6
Nb. Obs	352847	3305	14582
Nb. Unique	52824	507	2634
$I/\sigma(I)$	8.3	24.9	1.2
CC1/2 (%)	99.4	99.5	63.4
Completeness	98.1	99.4	76.0
Multiplicity	6.7	6.5	5.5



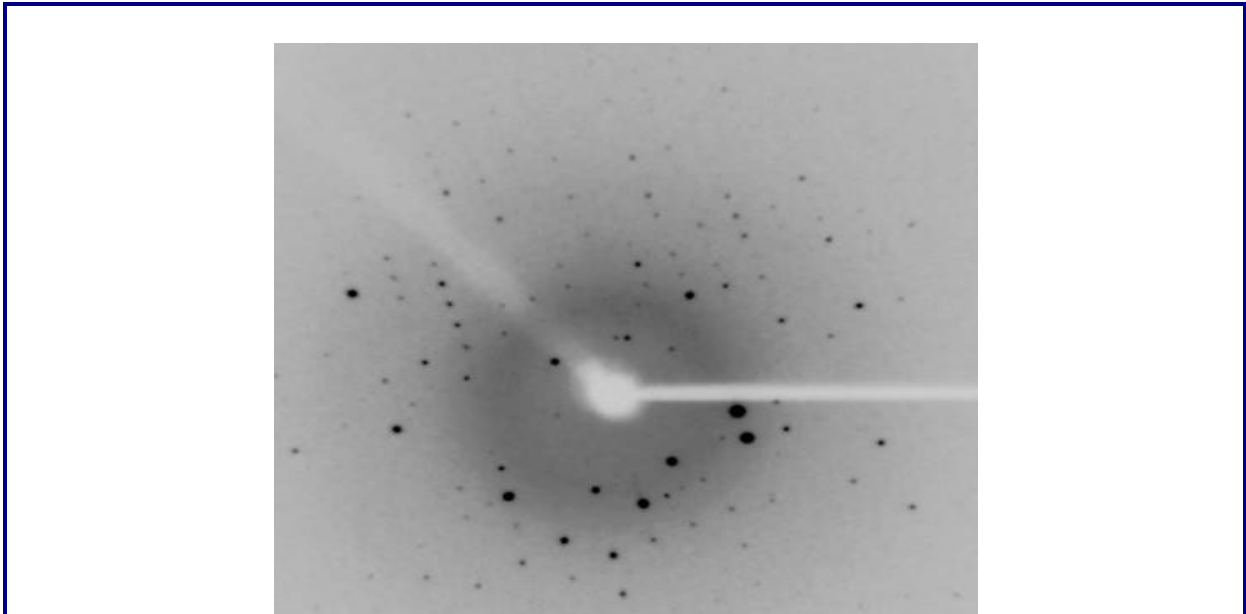


Figure 3 : Cliché de diffraction du cristal maclé de CALB+Lys mesuré au synchrotron Soleil

#### iv) Problème de macle

La définition d'une macle selon [1] est la suivante : "Une macle est un édifice cristallin hétérogène formé de l'association de deux ou plusieurs cristaux homogènes de même nature orientés suivant des lois déterminées." Une macle, également appelée "Crystal Twinning" en anglais, se produit lors de la croissance des cristaux. Les clichés de diffraction des cristaux maclés sont souvent difficiles à interpréter. Parfois, il est impossible de résoudre la structure car les lois de macle sont indéterminées. Leur détermination permet la résolution des structures, mais l'affinement reste souvent difficile.

Les macles peuvent généralement être classées de la manière suivante :

#### ❖ D'après leur genèse:

- macles de croissance, qui se forment pendant la croissance cristalline, soit dans les premières étapes, soit par accolement tardif de cristaux ayant déjà atteint une taille considérable.
- macles de transformation, qui se forment suite à une transition de phase dans laquelle la symétrie du cristal baisse et que, dans sa structure, des domaines à différentes orientations vont se former.
- macles mécaniques, qui se forment suite à une action mécanique, notamment une pression orientée le long d'une direction.

## ❖ Classement morphologique [2]:

Le mode de jonction des cristaux permet de définir trois grandes familles de macles.

- macle par rotation ou axiale, lorsque l'élément de macle est une rangée (axe de macle);
- macle par réflexion ou par pénétration, lorsque l'élément de macle est un plan réticulaire (plan de macle);
- macle par inversion, lorsque l'élément de macle est un point;

## ❖ Classement structurale des macles par Georges Friedel :

Sur la base des valeurs de l'indice de macle ( $n$ ) et de l'obliquité ( $\omega$ ), Friedel a classé les macles en quatre catégories principales.

**Tableau 2 : Classification des macles selon les valeurs de l'indice macle et de l'obliquité**

Indice de macle  $n = \frac{V_{\text{maille maclée}}}{V_{\text{maille individuel}}}$  qui est le nombre avec lequel il faut multiplier le volume de la maille du cristal de base (réseau Rc) pour obtenir la maille du réseau de macle (Rm).

Obliquité de la macle  $\omega$  : c'est la déviation par rapport à la perpendicularité exacte de l'axe ou du plan de macle ( $\omega \leq 3-4^\circ$ ).

	$n = 1$	$n > 1$
$\omega = 0$	macle par méridrie	macle par méridrie réticulaire
$\omega > 0$	macle par pseudo-méridrie	macle par pseudo-méridrie réticulaire

Dans le cocrystal CALB+Lys, une macle par rotation de la forme  $[h -k -l]_{180^\circ}$  a été identifiée lors de l'analyse des données, et cette loi de macle correspond à une macle par pseudo-méridrie. Le paramètre statistique développé par J. Padilla et T. Yeates a permis de révéler sans ambiguïté le problème de macle [3]. Ce paramètre consiste à comparer localement des intensités de réflexions choisies de manière appropriée. Un cristal non maclé a un paramètre de 0.5 alors qu'un cristal maclé avec une fraction de 0.5 a un paramètre de 0.375. Les données de diffraction de CALB+Lys présentent un paramètre de 0.39 indiquant une fraction de macle d'environ 30%. Par ailleurs ce problème de macle avait déjà été évoqué dans l'analyse de cristaux de CALB [4] [5]. Les cristaux de protéines maclés posent souvent des défis lors de leur analyse structurale par diffraction des rayons X, car la macle perturbe les informations

obtenues à partir des données de diffraction, comme cela a été le cas dans le cocrystal CALB+Lys. L'affinement a été réalisé à l'aide du logiciel PHENIX. Il n'a pas été possible de déposer la structure affinée dans la Protein Data Bank car il y avait un problème de gestion de la macle dans le logiciel de la PDB. Les cartes de densité électronique peuvent être légèrement bruitées dans les zones agitées de la protéine. Toutefois, les cartes de densité électronique semblent indiquer la présence de lysine dans le site actif, avec le groupement amine de la chaîne latérale à proximité de la triade catalytique. Ce résultat est en accord avec les conclusions de la récente publication de Dettori, Jelsch et collaborateurs [6]. Ces travaux montrent que CALB acyle quasi exclusivement le groupement amine de la chaîne latérale de la lysine en présence d'acide laurique, et les études *in silico* corroborent ce résultat expérimental.

L'analyse préliminaire réalisée dans le cadre de cette thèse semble indiquer qu'il est possible d'obtenir un complexe entre CALB et la lysine par co-cristallisation. Malheureusement, les cristaux sont maclés. Des données de diffraction de meilleure qualité permettraient peut-être la modélisation de la macle, en d'autres termes, un affinement cristallographique plus précis pour interpréter avec confiance les cartes de densité électronique. De nouvelles conditions de cristallisation pourraient également être explorées dans le but d'obtenir des cristaux non maclés.

## **E - I.2 Neuropilin-1-peptide KDKPPR**

### **E - I.2.a. Résumé de l'article**

La neuropiline-1 (NRP-1), un corécepteur de surface cellulaire de nombreux facteurs de croissance et d'autres molécules de signalisation, a depuis longtemps été au centre de l'attention en raison de son association avec le développement et la progression de plusieurs types de cancers. Plus récemment, son rôle en tant que récepteur pour l'infection par le SARS-CoV-2 (virus du COVID-19) a suscité un vif intérêt. La principale cible thérapeutique est une poche située dans le domaine b1 de NRP-1, dans laquelle les régions C-terminales des isoformes du facteur de croissance de l'endothélium vasculaire A se fixent. Dans la structure cristalline du fragment b1 humain natif, ce site de liaison au VEGF est obstrué par l'empilement cristallin, ce qui empêche la liaison des ligands. Six acides aminés chargés situés à la surface de la protéine ont été mutés afin de perturber l'empilement cristallin et d'obtenir une nouvelle forme cristalline. La structure du fragment b1 muté complexé avec le peptide KDKPPR a été

déterminée par cristallographie aux rayons X. Récemment, ce peptide a été combiné à un photosensibilisateur et à un agent de contraste afin de cibler spécifiquement NRP-1 pour la détection et le traitement du glioblastome, un cancer du cerveau agressif, par thérapie photodynamique (PDT).

Le variant électrostatique a cristallisé dans une nouvelle forme cristalline avec le site de liaison au VEGF exposé au solvant et occupé par les trois résidus C-terminaux du peptide KDKPPR. Les interactions atomiques ont été analysées à l'aide de l'interface de Hirshfeld du complexe peptide/protéine. Des simulations de dynamique moléculaire (MD) ont été effectuées pour analyser plus en détail la liaison du peptide et les mouvements des protéines de type sauvage et mutant. La dynamique moléculaire (MD) du peptide suggère que le résidu arginine est le principal point d'ancrage du peptide dans la poche de liaison du fragment b1 de NRP1-1.

**E - I.2.b.      Publication N° 1 :**

New crystal form of Human Neuropilin-1 b1 fragment with six electrostatic mutations complexed with KDKPPR peptide ligand.

## Article

# New Crystal Form of Human Neuropilin-1 b1 Fragment with Six Electrostatic Mutations Complexed with KDKPPR Peptide Ligand

Ibrahima Goudiaby<sup>1,2</sup>, Thérèse E. Malliavin<sup>3</sup> , Eva Mocchetti<sup>1</sup>, Sandrine Mathiot<sup>1</sup>, Samir Acherar<sup>4</sup> , Céline Frochot<sup>5</sup> , Muriel Barberi-Heyob<sup>6</sup>, Benoît Guillot<sup>1</sup>, Frédérique Favier<sup>1</sup> , Claude Didierjean<sup>1,\*</sup>  and Christian Jelsch<sup>1,\*</sup> 

<sup>1</sup> Université de Lorraine, CNRS, CRM2, F-54000 Nancy, France; i.goudiaby1592@zig.univ.sn (I.G.); eva.mocchetti@univ-lorraine.fr (E.M.); benoit.guillot@univ-lorraine.fr (B.G.)

<sup>2</sup> Université Assane Seck de Ziguinchor, Laboratoire de Chimie et de Physique des Matériaux (LCPM), 523 Ziguinchor, Senegal

<sup>3</sup> Université de Lorraine, CNRS, LPCT, F-54000 Nancy, France; therese.malliavin@univ-lorraine.fr

<sup>4</sup> Université de Lorraine, CNRS, LCPM, F-54000 Nancy, France

<sup>5</sup> Université de Lorraine, CNRS, LRGP, F-54000 Nancy, France

<sup>6</sup> Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France; muriel.barberi@univ-lorraine.fr

\* Correspondence: claudedidierjean@univ-lorraine.fr (C.D.); christian.jelsch@univ-lorraine.fr (C.J.)

**Abstract:** Neuropilin 1 (NRP1), a cell-surface co-receptor of a number of growth factors and other signaling molecules, has long been the focus of attention due to its association with the development and the progression of several types of cancer. For example, the KDKPPR peptide has recently been combined with a photosensitizer and a contrast agent to bind NRP1 for the detection and treatment by photodynamic therapy of glioblastoma, an aggressive brain cancer. The main therapeutic target is a pocket of the fragment b1 of NRP1 (NRP1-b1), in which vascular endothelial growth factors (VEGFs) bind. In the crystal packing of native human NRP1-b1, the VEGF-binding site is obstructed by a crystallographic symmetry neighbor protein, which prevents the binding of ligands. Six charged amino acids located at the protein surface were mutated to allow the protein to form a new crystal packing. The structure of the mutated fragment b1 complexed with the KDKPPR peptide was determined by X-ray crystallography. The variant crystallized in a new crystal form with the VEGF-binding cleft exposed to the solvent and, as expected, filled by the C-terminal moiety of the peptide. The atomic interactions were analyzed using new approaches based on a multipolar electron density model. Among other things, these methods indicated the role played by Asp320 and Glu348 in the electrostatic steering of the ligand in its binding site. Molecular dynamics simulations were carried out to further analyze the peptide binding and motion of the wild-type and mutant proteins. The simulations revealed that specific loops interacting with the peptide exhibited mobility in both the unbound and bound forms.

**Keywords:** Neuropilin 1; variant; ligand; X-ray crystallography; molecular dynamics simulation; Hirshfeld interface; electrostatic influence



**Citation:** Goudiaby, I.; Malliavin, T.E.; Mocchetti, E.; Mathiot, S.; Acherar, S.; Frochot, C.; Barberi-Heyob, M.; Guillot, B.; Favier, F.; Didierjean, C.; et al. New Crystal Form of Human Neuropilin-1 b1 Fragment with Six Electrostatic Mutations Complexed with KDKPPR Peptide Ligand.

*Molecules* **2023**, *28*, 5603. <https://doi.org/10.3390/molecules28145603>

Academic Editor: Chojiro Kojima

Received: 23 June 2023

Revised: 15 July 2023

Accepted: 20 July 2023

Published: 24 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

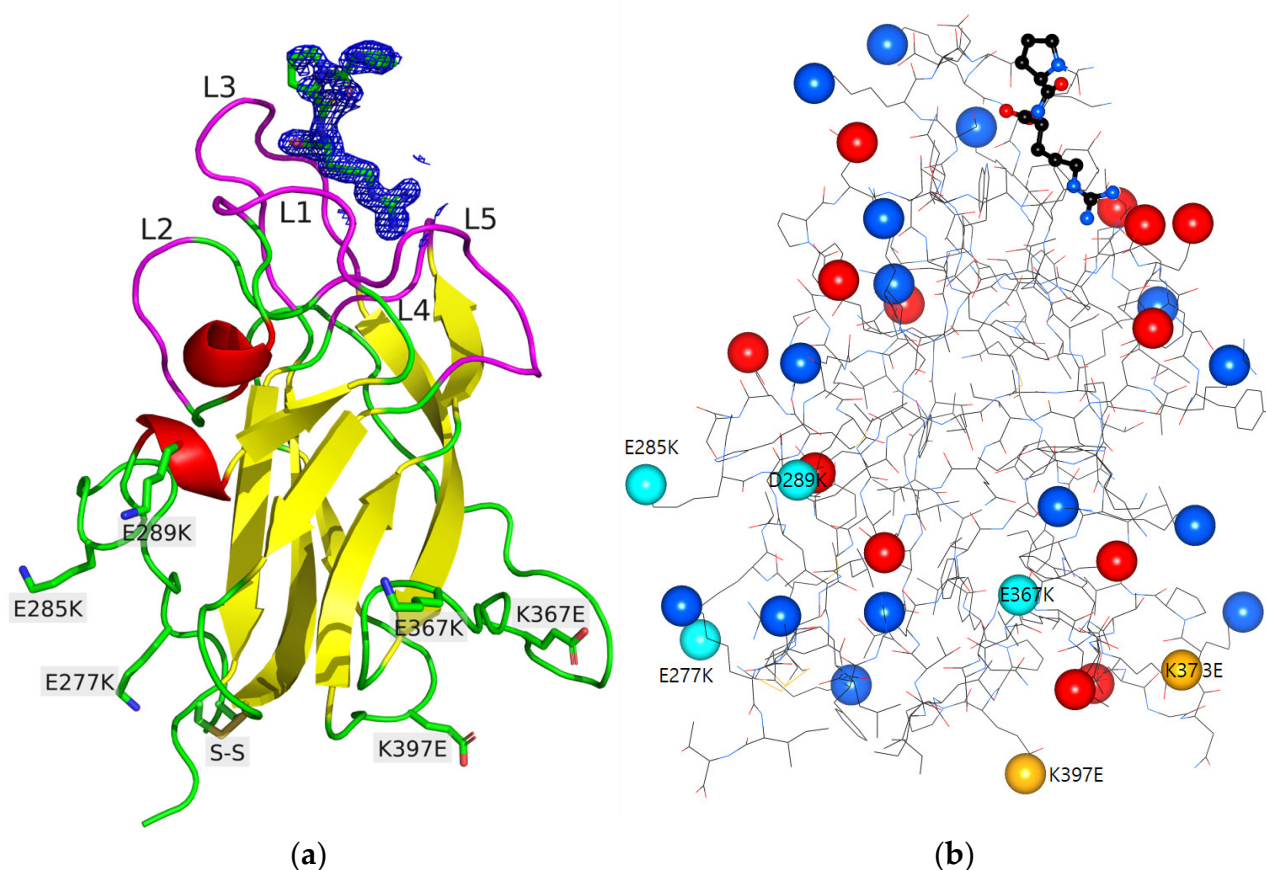
Neuropilins (NRP1 and NRP2) are type I single-pass transmembrane glycoproteins expressed in all vertebrates and have many physiological roles. They act as co-receptors of a range of growth factors and other signaling molecules. A recent cryo-electron microscopy structure highlights the role of NRP1 in a ternary complex with a semaphorin protein and a plexin receptor, which altogether mediates signaling in neuronal axon guidance and other processes [1]. NRP1 also forms a ternary complex with vascular endothelial growth factor A 165 (VEGF-A165) and the receptor VEGFR2 [2]. This complexation is associated

with intracellular signaling, mitogenesis, cell migration and angiogenesis [3]. Research has shown that NRP1 plays a significant role in the development and progression of various cancer types [4] and also more recently in the infectivity of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [5]. In particular, this transmembrane receptor has been suggested as a molecular therapeutic target for glioblastoma, with overexpression mainly due to endothelial cells of angiogenic phenotype and associated pro-tumor macrophages, both of which are linked to an unfavorable prognosis [6]. A recent review outlines the various functions of NRP1 in the context of cancer treatments [7].

The membrane protein NRP1 (as NRP2) contains a large N-terminal extracellular region (~850 amino acids (AA)), a single transmembrane domain (~25 AA) and a short C-terminal cytoplasmic domain (~45 AA) [8]. The ectodomain consists of five independent domains, where the first four (a1, a2, b1 and b2 domains) are involved in ligand binding, while the role of the last one (c domain) is still under debate (oligomerization, NRP1 homodimerization, etc.) [1]. The interdomain linkers are also important in the heterocomplex formation as spacers [1]. The determination of the crystal structure of the b1 domain provided the first structural insight at the atomic-level into NRP1 [9]. The interaction of NRP1 with VEGF has been extensively studied. Briefly, the NRP1-b1 domain folds into a distorted jelly roll barrel motif that is composed of two beta-sheets [9] (Figure 1a). The strands are connected by loops of varying length. The bottom of the beta-barrel core exhibits a triangular shape that contains an intramolecular disulfide bridge. At the top of the beta-barrel core, the loops are divided in six loop regions (L1–L6) to delimit the pocket of the positively charged tail of VEGF. The C-terminal arginine residue of VEGF-A165 is buried in this pocket and is a key feature of the binding. Numerous atomic experimental structures have been determined to characterize and/or inhibit the interaction of NRP1-b1 and VEGF using peptides or arginine derivatives as well as a fusion protein [10–13].

We have developed peptides combined with a photosensitizer to target NRP1 in the context of photodynamic therapy (PDT) to detect and treat glioblastoma [14–16]. Recently, a nanoparticle was designed that combines KDKPPR motif as a targeting peptide, porphyrin as photosensitizer and gadolinium chelate as contrast agent. This nanoparticle, called AGuIX@PS@KDKPPR, enables the detection of tumor tissue by magnetic resonance imaging and treatment by PDT [6,17,18]. The affinity of the nanoparticle for human NRP1 was validated, and it was found to be ten times lower than that of the free peptide ( $K_D$  of 4.7  $\mu\text{M}$  for AGuIX@PS@KDKPPR and  $K_D$  of 0.5  $\mu\text{M}$  for KDKPPR) [17].

In this study, the KDKPPR peptide was synthesized, and its molecular interactions with NRP1-b1 fragment were investigated by X-ray crystallography and molecular dynamics (MD) simulations. Previously, we have attempted to co-crystallize NRP1-b1 with a carbohydrate-based peptidomimetic [19]. However, we constantly obtained tetragonal crystals that were isomorphous to the crystals of the unbound protein. These crystals are unsuitable for obtaining structures of complexes because the site that binds the C-terminal tail of VEGF is obstructed by symmetry-related protein molecules [9]. We have used site-directed mutagenesis to modify the repartition of charges on the surface of NRP1-b1 to induce changes in the crystal packing. This approach is an efficient tool for crystallizing a protein in a new form and facilitating, for example, the formation of a protein–ligand complex in the crystal through co-crystallization or soaking techniques [20]. In this study, we successfully co-crystallized the KDKPPR peptide with a hexavariant of NRP1-b1 (Glu277Lys, Glu285Lys, Asp289Lys, Glu367Lys, Lys373Glu, Lys397Glu). An original crystal form was obtained where the VEGF-binding pocket is filled by the KDKPPR peptide and is located in large spaces connected by solvent channels in the crystal packing. The structure of the NRP1-b1/KDKPPR complex was analyzed by MD simulations and innovative tools based on a multipolar electron density model [21].



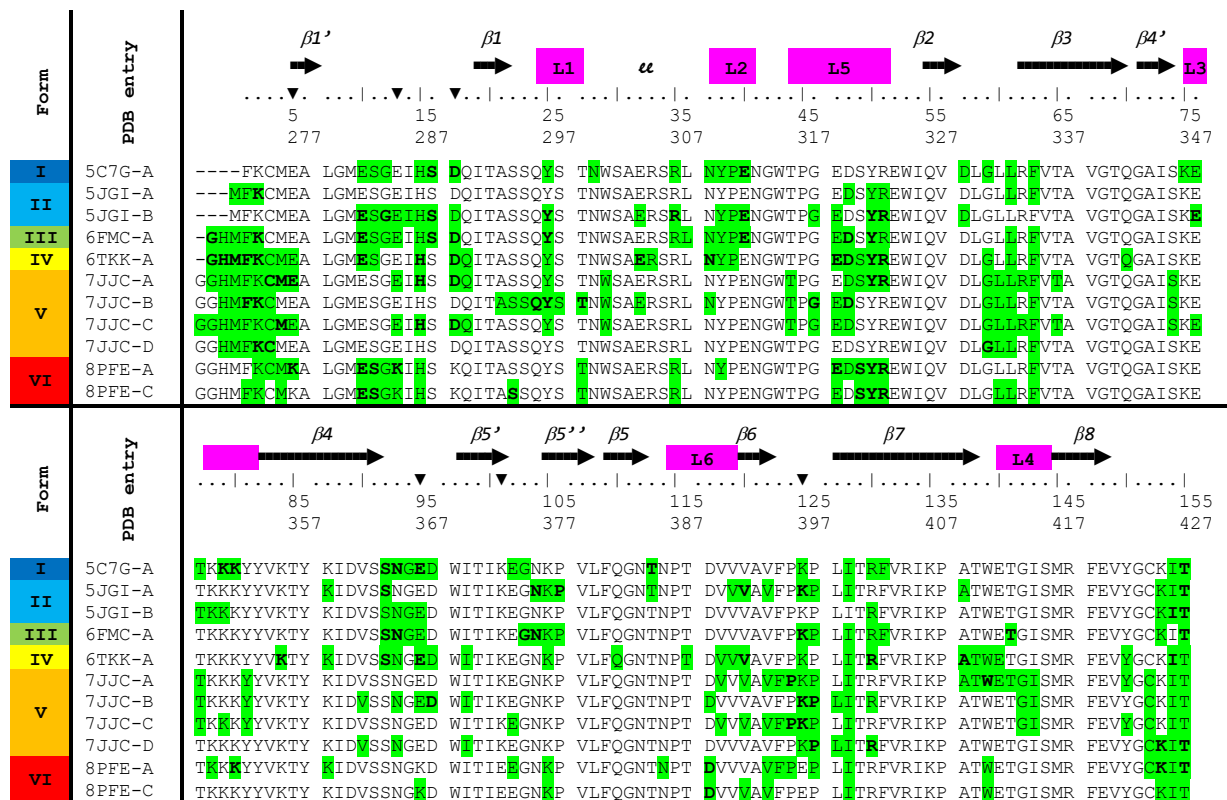
**Figure 1.** (a) Ribbon view of the crystal structure of NRP1-b1 hexavariant. The six mutations and the disulfide bridge Cys275-Cys424 are shown as sticks and labelled. The loops L1–L5 that line the VEGF-binding pocket are highlighted in magenta and labelled. The PPR moiety of KDKPPR peptide is shown as sticks with refined  $2mFo-DFc$  electron density contoured at  $1.0 \sigma$ . (b) Repartition of the charged residues in NRP1-b1 hexavariant. The positive (blue) and negative (red) charges are shown as spheres on the protein structure. The mutated residues with change in charge are labeled and are shown in light blue and light yellow. The PR residues of KDKPPR peptide are shown on the top of the figure.

## 2. Results and Discussion

### 2.1. Design of the NRP1-b1 Hexavariant

The charge distribution on the surface of NRP1-b1 was significantly altered in order to promote the formation of a new crystal form, based on a visual inspection of the model. The point mutations were chosen to be far away from the VEGF-binding pocket to minimize disruption of the peptide-binding site (Figure 1b). Specifically, six charged residues on the protein surface were mutated to residues of opposite charge: Glu277Lys, Glu285Lys, Asp289Lys, Glu367Lys, Lys373Glu and Lys397Glu. Mutation of these residues did not lead to isoelectric conservation. In fact, the estimated isoelectric point of the hexavariant was 1 unit higher than that of the native protein, with a value of 9.2 for the hexavariant and 8.0 for the wild type. These choices resulted in the creation of a highly electropositive region around the Glu285Lys, Asp289Lys and Glu277Lys mutations, consisting of seven positively charged residues (Lys274, Lys277, Lys285, Lys289, Arg334, Arg402 and Lys425). Three of them (Lys277, Lys285 and Lys425) form hydrogen bonds with symmetry-related molecules in the crystal of the hexavariant (see below, Figure 2 and Table S1).





**Figure 2.** Highlights (in green) of the NRP1-b1 residues involved in contacts with a neighboring monomer in the crystal forms I to VI. Form VI corresponds to the NRP1-b1 hexavariant. Contacts are defined as residues with a proximity of less than 4 Å. Residues in bold characters forms intermolecular hydrogen bonds. The positions of the mutations have been highlighted with triangles above the two sets of residue numbering.

## 2.2. Crystal Structure of the NRP1-b1 Hexavariant

### 2.2.1. Description of the Structure

The hexavariant (Glu277Lys, Glu285Lys, Asp289Lys, Glu367Lys, Lys373Glu, Lys397Glu) of the NRP1-b1 domain crystallized in the  $P3_221$  space group with a novel packing arrangement. The asymmetric unit contained two chains (A and C), each with a KDKPPR peptide (chains B and D) in its VEGF-binding site, plus 377 water molecules and an acetate ion. Only the last three residues of the peptide (PPR) in both monomers were included in the refined structure (Figure 1a). Positive residual peaks in the difference electron density maps persisted around the first proline residue of the KDKPPR peptide. These peaks were slightly stronger in monomer D. We made several attempts to improve the final  $2mFo-DFc$  electron density map, such as modeling an additional residue in the peptide or modeling alternative conformations for the first proline residue. However, no satisfactory model has emerged from any of these efforts. The two monomers were nearly identical, with an overall coordinate root mean square deviation (RMSD, Å) of 0.30 Å on the 154 C $\alpha$  atoms common to both chains. Slight conformational differences were observed at the very first and last residues of the two monomers, due to their different involvement in packing contacts. This probably also explained the differences observed at the neighboring disulfide bridge Cys275-Cys424 of the two monomers. Indeed, in monomer A, it showed two alternative conformations, one of which was rather ill-defined, whereas only one conformation was observed in monomer C. The same argument concerning packing effects probably applied for the slight differences observed in some side chain conformations (Phe335) and sometimes in some main chain regions (Glu374-Pro378, Pro396-Pro398). The mutations did not affect the protein fold, since the current hexavariant model showed an overall



RMSD of 0.51 Å with the wild-type structure (PDB entry 1KEX, [9]), compared to a 0.30 Å RMSD between the hexavariant independent monomers (A and C). Four of the mutations (Glu277Lys, Glu285Lys, Asp289Lys and Lys397Glu) resulted in no apparent change in the main chain fold. On the contrary, Glu367Lys and Lys373Glu were located in regions with larger observed displacements: the C $\alpha$  atom of residue 367 underwent a 1.75 Å shift due to an overall movement of region Ser363-Trp369, while the  $\psi$  angle of Gly375 rotated 180° in the rearrangement of the Glu373-Pro378 loop. Both of these observations are most likely the consequence of the change in crystal packing and protein–protein contacts (Figure 2), rather than the direct influence of the mutations on the polypeptide conformation.

### 2.2.2. Crystal Packing and Intermolecular Contacts

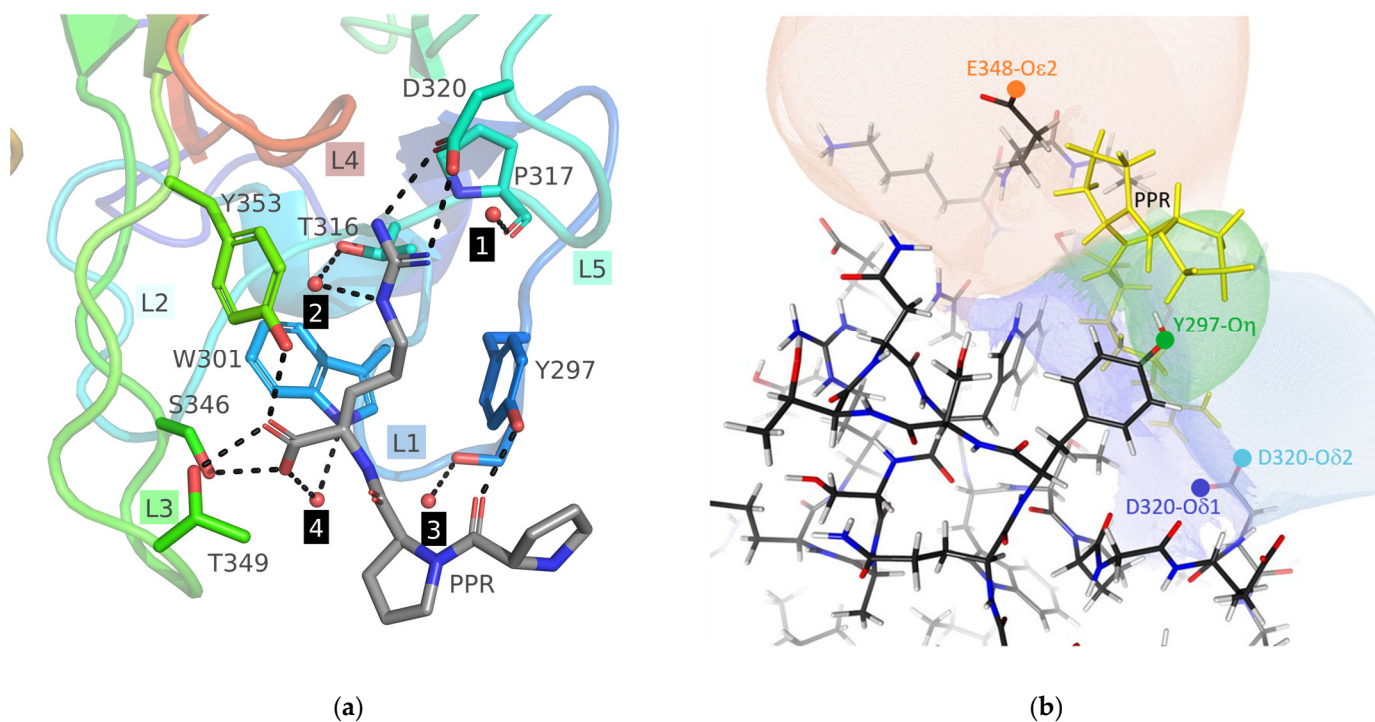
The mutations induced a new crystal packing in which a few introduced residues (Lys277 and Lys285 in chain A) were involved in modified intermolecular interactions. The two independent hexavariant models (chains A and C) have similar molecular environments. Indeed, more than half of the contacts are identical in both chains (Table S1). The KDKPPR peptide is not involved in the crystal cohesion. The hexavariant crystal contains large solvent spaces (volume of approximately 30,000 Å<sup>3</sup>, Figure S1) into which it seems possible for small molecules to diffuse because they are connected by solvent channels. The N-terminal parts of the two independent KDKPPR peptides (KDK moiety, chains B and D) are located in the bulk solvent, while their C-terminal parts are tightly bound in the pocket that would welcome the C-terminal tail of VEGF, in chains A and C, respectively (see below). The electron density for the N-terminal regions of chains B and D was too weak to build a model, probably because these regions exhibited dynamic disorder.

We compared the hexavariant crystal form with those of the NRP1-b1 domain available in the Protein Data Bank [22]. Five structures of NRP1-b1 with distinct unit cell parameters were found (Figure 2, Table S2). The asymmetric units contain one to four independent chains. The crystal structure of the tetragonal form I represents the unbound state of NRP1-b1, in which the VEGF-binding site is obstructed by symmetry-related molecules. All other crystal forms were obtained by co-crystallization of NRP1-b1 with arginine or close derivatives (crystal forms II to V). The intermolecular environment analysis revealed that crystal forms II and III have similar packings (*P*2<sub>1</sub> space group with two independent chains and *P*4<sub>1</sub> space group with one independent chain, respectively) (Figure S2). We also noticed that in all cases, at least one independent ligand was positioned at the interface between two adjacent NRP1-b1 monomers (Figure S3). Their presence was probably necessary for the cohesion of these crystal packings. That is why we worked with a variant, trying to find a new crystal packing where the ligand only contacts the protein to which it is bound.

## 2.3. Protein Ligand Interaction

### 2.3.1. Description of Ligand Binding

The KDKPPR peptide was bound in the VEGF-binding pocket formed by loops L1 to L5 of NRP1-b1, with its C-terminal arginine positioned in a very similar manner to what was described in detail by [11]. Briefly, its guanidine group forms a salt bridge involving two hydrogen bonds with Asp320 (L5), while the aliphatic part of its side chain is stacked between the phenyl rings of Tyr297 (L1) and Tyr353 (L3) (Figure 3a). The arginine residue of the ligand is also tightly bound at the main chain by residues of L3, with one of its carboxylate oxygen atoms hydrogen-bonded to the side chain hydroxyl groups of Thr349 and Tyr353 and the second one to the lateral chain of Ser346. The last strong anchoring of the KDKPPR peptide arises from the main chain carbonyl group of the first proline, which forms a hydrogen bond with the hydroxyl group of Tyr297 (L1). The first three residues of the peptide apparently found no preferred binding to NRP1-b1 that could have fixed them in a given conformation and made them visible in the electron density.



**Figure 3.** (a) Structure of the binding site of NRP1-b1 hexavariant in complex with the KDKPPR peptide. The pocket is mainly composed of five loops (L1 to L5), which are respectively colored blue, cyan, green, red and turquoise. The crystallographic model of the peptide includes only the PPR moiety, represented as sticks. The NRP1 residues in the close proximity of the peptide are also depicted as sticks. The four structural water molecules are highlighted as spheres, while hydrogen bonds are illustrated as dashed sticks. Various labels are provided to enhance clarity, indicating loops, peptide, residues and water molecules. (b) Nucleophilic Influence Zones associated with the oxygen atoms Asp320-O $\delta$ 1 (dark blue), Asp320-O $\epsilon$ 2 (light blue), Tyr297-O $\eta$  (green) and Glu348-O $\epsilon$ 2 (major conformer, orange) in the vicinity of the ligand-binding site of monomer A. The corresponding atomic nucleophilic sites are indicated by colored circles, and the PPR moiety of the KDKPPR peptide is highlighted in yellow.

Ordered water molecules were checked and found identical to those discussed by [11]. Four structural water molecules in the peptide-binding site (HOH A690, A666, A717 and B101 for chain A and HOH C570, C549, C596 and D101 for chain B) were included in calculation of the contacts enrichment ratio (see below, Figure 3a).

### 2.3.2. Electrostatic Influence of the Protein on the Peptide

The VEGF-binding site of NRP1-b1 is occupied by the PPR moiety of the KDKPPR peptide, with its electrophilic C-terminal arginine residue forming a salt bridge with Asp320. For this reason, it is interesting to study the electrostatic influences from the whole NRP1 protein on the peptide from the point of view of Nucleophilic Influence Zones (NIZ) [23] (Figure 3b). A NIZ represents the volume containing all the electric field lines converging to a specific nucleophilic site, often an oxygen atom in proteins. Therefore, an electrophilic ligand within this space, like a positive charged entity, experiences attractive electrostatic forces directed toward the corresponding nucleophilic site. NIZs associated with the relevant oxygen atoms involved in the binding of the PPR moiety were calculated excluding ligand atoms and using CHARGER program [24].

As anticipated, the NH<sub>2</sub> groups of the arginine guanidinium of the KDKPPR peptide are influenced by their respective hydrogen-bonded oxygen atoms of Asp320 (dark blue surface for O $\delta$ 1 atom and light blue surface for O $\delta$ 2 atom) (Figure 3b). The NIZ of the Glu348-O $\epsilon$ 2 atom (orange surface) covers the position of one peptide proline residue, while

the NIZ of the Tyr297-O $\eta$  atom (green surface) encompasses the majority of the other peptide proline residue. Hence, these NIZs illustrate the forces acting on both proline residues, specifically exerting an attraction on their hydrogen atoms, thus stabilizing the ligand conformation by pulling one proline residue towards Glu348 and the other towards Tyr297.

Given the close proximity of the solvent dielectric medium and the distances between the proline residues involved and the generators of the discussed NIZs (i.e., Glu348-O $\epsilon$ 2 and Tyr297-O $\eta$ ), it is likely that the electrostatic stabilization is relatively weak but cannot be disregarded. Electrostatic forces, being long-range interactions, play a role, and the presence of numerous charged hydrogen atoms in the pyrrolidine rings of prolines facilitates favorable interactions with the negatively charged oxygen atoms of the side chains of Glu348 and Tyr297, supporting our interpretation.

However, there are other factors contributing to the stabilization of the KDKPPR peptide that deserve attention. One such factor is the strong hydrogen bond between the hydroxyl group of Tyr297 and the carbonyl oxygen atom of the first proline residue in KDKPPR (Figure 3a). Furthermore, a stabilizing van der Waals contact can be inferred from the proximity of the Glu348-C $\gamma$  hydrogen atoms and the C $\gamma$  hydrogen atoms in the second proline residue of the KDKPPR peptide.

Furthermore, since electrostatic influences participate in the driving of ligand diffusion across the solvent toward a protein-binding site [25], the NIZs also provide insights on the electrostatic forces originating from the protein residues and directing the ligand during its approach. Here, the NIZs of Glu348-O $\epsilon$ 2 and Asp320-O $\delta$ 2 atoms extend beyond the protein surface, suggesting their potential contribution to the electrostatic steering effect that facilitates electropositive ligand fixation in the VEGF-binding site of NRP1-b1 (Figure 3b).

### 2.3.3. Hirshfeld Surface and Contacts Enrichment Ratio

The Hirshfeld surface [26] between the peptide and the protein was calculated with MoProViewer software [27]. The Hirshfeld surface allows the analysis and visualization of intermolecular interactions. The contact enrichment ratio  $E_{XY}$  between chemical species X and Y is obtained by comparing the actual  $C_{XY}$  contacts with those calculated as if all contact types had the same probability of forming [28]. The equiprobable proportions  $R_{XY}$  are derived by probability products from the chemical proportions on the Hirshfeld surface. An  $E_{XY}$  enrichment ratio greater than unity for a particular contact between chemical species X...Y indicates that these are over-represented. The chemical nature of the contacts and their enrichment in the complex of NRP1-b1 with KDKPPR peptide are shown in Table 1. The proportions of contact types are very similar in the two independent monomers of NRP1-b1 (correlation of  $C_{XY}$  contact type proportions of 99.9%).

The less polar Hc hydrogen atoms bonded to carbon were distinguished from the more electropositive Ho/n atoms bound to oxygen or nitrogen. Four structural water molecules in monomer A (and their equivalent in monomer B, see above) in the binding cleft were kept and attributed to the protein in the complex. Obviously, the O...Ho/n hydrogen bonds are strongly attractive from an electrostatic point of view and are overrepresented ( $E = 2.58$ , Table 1). Representing 17.4% of the interaction surface, they are recognized as the most favored contacts. This concerns interactions between C=O and COO $^-$  acceptors and N-H, NH $_2$  and O-H hydrogen bond donors. The O...Hc weak hydrogen bond contacts represent 15.3% of the interaction surface and can be considered as weakly favored contacts, with an enrichment ratio of  $E = 1.20$ . Some enrichment ratios close to zero concern the O...O and Hn/o...Hn/o contacts, which are absolutely avoided in the protein/ligand complex because they concern repulsive self-contacts between charged species.

Occupying the largest contact area (21.3%), non-polar Hc...C contacts are significantly over-represented ( $E = 1.80$ ) and consist in particular of C-H... $\pi$  interactions involving the aromatic rings of Tyr239, Tyr183 and Tyr187. The Hc...Hc contacts represent 10.3% of the surface and can be considered as weakly disfavored contacts, as they present an enrichment ratio lower than unity ( $E = 0.81$ ). The Ho/n...W (oxygen atoms of water

molecules) contacts involving the four structural water molecules represent 6.9% of the interaction surface and can be considered as significantly favored contacts, with  $E = 1.88$ . The four water molecules interact essentially with Ho/n and secondarily with Hc atoms.

**Table 1.** Statistical analysis of intermolecular contacts on the Hirshfeld interface between the PPR moiety of the KDKPPR peptide and NRP1-b1.

Atom Type	C	Hc	N	Ho/n	O	W <sup>#</sup>
surface_peptide	8.8	<b>42.7</b>	5.5	<b>22.9</b>	<b>20.1</b>	0
surface_protein	<b>21.4</b>	<b>30.0</b>	1.2	<b>15.6</b>	<b>15.9</b>	<b>16.0</b>
C	1.4	<b>21.3</b>	1.3	3.4	0.4	1.0
Hc		<b>10.3</b>	3.5	7.2	<b>15.3</b>	4.8
N			0	0.5	0.2	1.3
Ho/n	$C_{XY}$	(%)		1.6	<b>17.4</b>	6.9
O					0.3	2.0
C	0.75	<b>1.80</b>	1.02	0.54	0.08	0.73
Hc		0.81	1.62	0.53	<b>1.20</b>	0.70
N			0	0.40	0.17	1.44
Ho/n	$E_{XY}$			0.44	<b>2.58</b>	<b>1.88</b>
O					0.09	0.64
		Hphob	Hphil	Hphob * Hphil		
surface %	peptide	57.1	43.0			
surface %	protein	52.6	47.4			
contacts	%	37.8	28.2	34.0		
enrichment		1.26	1.38	0.69		

\* The Hirshfeld surface was limited to the regions where the electron density is larger than  $0.0013 \text{ e}/\text{\AA}^3$  in order to omit the peptide surface exposed to the solvent. <sup>#</sup> Oxygen atom of water molecules. The second and third rows show the chemical content on the Hirshfeld surface. The next rows show the %  $C_{XY}$  of the contact types on the surface, followed by their enrichment ratios. The major surface components, the  $C_{XY}$  contacts and the significantly enriched contacts ( $E > 1$ ) are highlighted in bold characters. In the lower part of the table, the atoms are grouped into hydrophobic (Hphob) and hydrophilic (Hphil) atoms.

The hydrophobic and hydrophilic atoms were regrouped in order to analyze the interactions between the two subgroups. The N contact surface occurs on  $sp^2$  peptide and guanidinium nitrogen atoms (N without electron lone pair) and was considered hydrophobic together with the C and Hc atoms. The peptide and protein interaction surfaces are constituted by more hydrophobic (57.1 and 52.6%, respectively) than hydrophilic atoms. The protein/ligand complex shows an enrichment of contacts between hydrophilic atoms ( $E = 1.26$ ) and between hydrophobic atoms ( $E = 1.38$ ). On the other hand, despite the mild enrichment of the weak hydrogen bonds of O...Hc, the cross contacts Hphob × Hphil are strongly under-represented ( $E = 0.69$ ), which indicates a good partitioning of hydrophilic and hydrophobic contacts.

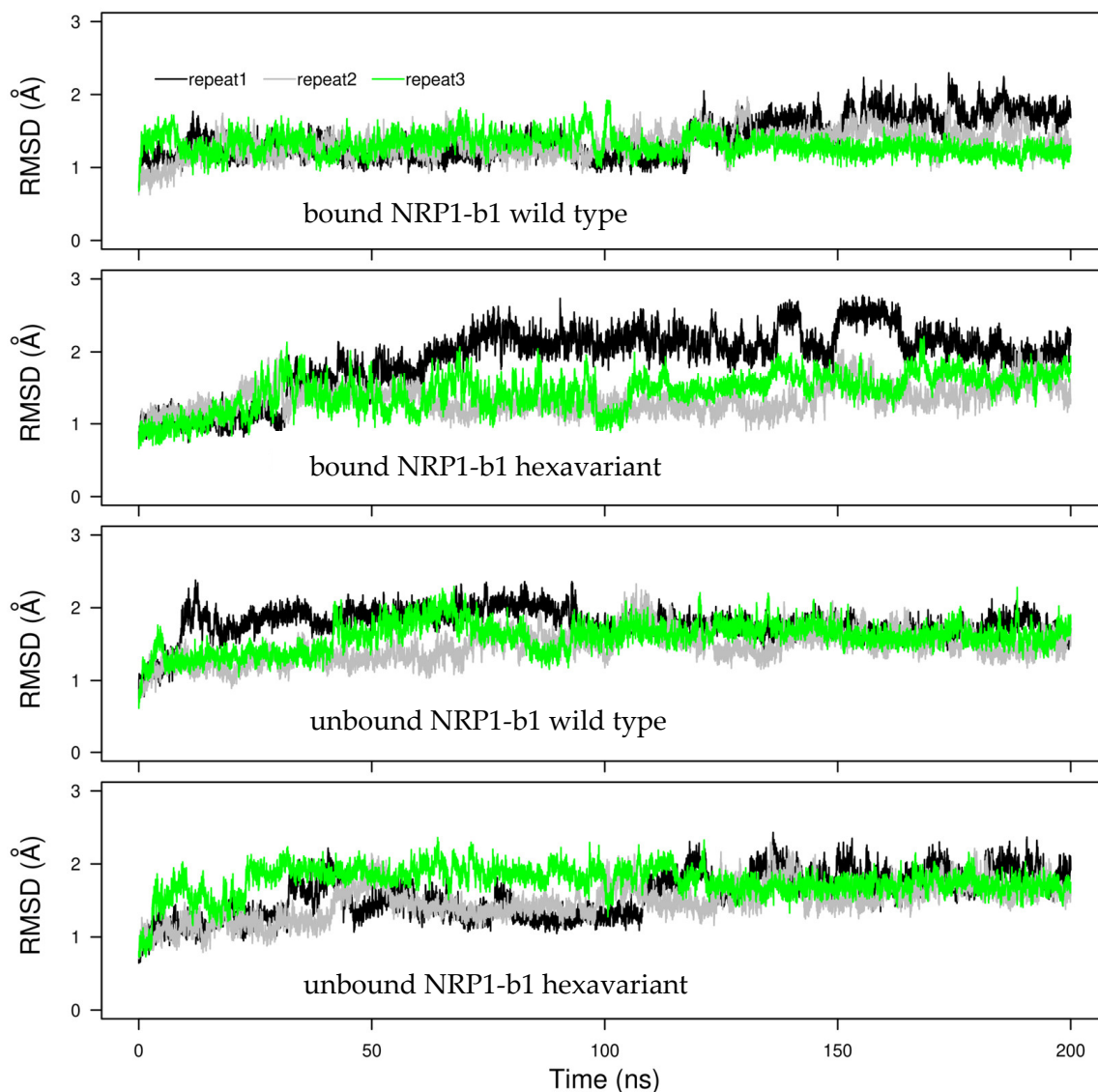
In summary, the protein/ligand complex is mainly maintained by over-represented strong O...Ho/n interactions, which correspond notably to the salt bridge anchoring Asp206 and the arginine residue of the peptide. Secondarily, more moderately enriched interactions also play an important role, such as weak C-H...O hydrogen bonds and hydrophobic contacts, notably between Hc and C atoms. The enrichment values agree with trends found in studies of interactions in several families of oxygenated and nitrogenated hydrocarbon molecules [28,29]. The strong hydrogen bonds such as O...Ho/n are significantly enriched; in the case of small-molecule crystals, the over-representation reaches even larger values beyond 10.

Concerning the weak C-H...O hydrogen bonds, they tend to occur in a moderately under-represented way in crystal structures of small molecules containing both strong H-bond donors and acceptors (such as alcohols for example), due to the competition of strong H-bonds. On the contrary, in the present protein/peptide interface, they appear slightly enriched. This can be explained by the excess of strong H-bond acceptors on the protein ( $S_O + S_W = 31.9\%$ ) compared to  $S_{Ho/n} = 22.9\%$  of strong H-bond donors on the peptide.



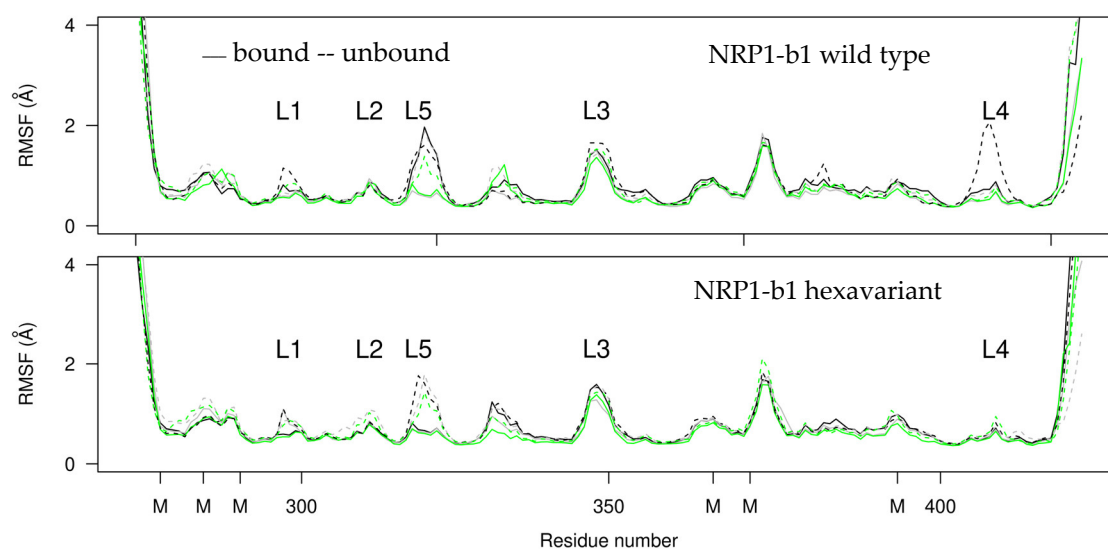
#### 2.4. Molecular Dynamics

Coordinate RMSD of NRP1-b1 was monitored along the molecular dynamics (MD) trajectories and stabilized after 50 ns in a range of 1–2 Å. Similar coordinate drift was observed for the bound and unbound proteins, as well as for the mutated and wild-type sequences. Only one copy of the mutated protein in complex with the peptide displays slightly larger coordinate drift (Figure 4).



**Figure 4.** Coordinate RMSD (Å) calculated on the backbone heavy atoms of NRP1-b1 with respect to the initial X-ray crystallographic structure. The curves measured on the triplicated trajectories are colored in black, green and gray, respectively.

The atomic root mean square fluctuations (RMSFs, Å, Figure 5) of NRP1-b1 display quite superimposed profiles for all MD trajectories, with the unbound protein (dashed line) showing more mobile regions in particular for the loop of residues 317–322, which corresponds to the loop L5 of the VEGF-binding pocket. The very similar fluctuation profiles for the variant and WT forms show that the mutations introduced to alter crystal packing do not introduce a major bias in the dynamics of the protein. The N-terminal region of the KDKPPR peptide interacting with NRP1-b1 displays large fluctuations along the trajectories, in agreement with the invisible electronic density for this part of the peptide.



**Figure 5.** Atomic root mean square fluctuations (RMSFs, Å) calculated along the molecular dynamics (MD) trajectories by superimposing the heavy backbone atoms of NRP1-b1 on the corresponding atoms in the initial crystal structure. The upper panel corresponds to the trajectories recorded on the WT protein, while the lower panel corresponds to the trajectories recorded on the NRP1-b1 hexavariant used to determine the crystal structure. The curves measured on the triplicated trajectories are colored black, green and grey, respectively. They are plotted as solid and dashed lines for NRP1-b1 in complex with the peptide and for the unbound form, respectively. The positions in the sequence of the mutated residues are marked with the letter ‘M’. The NRP1 loops interacting with the peptide are labelled as defined in the text.

The root mean square thermal displacements (RMSTDs) of the C $\alpha$  atoms in the crystal structure were derived from the thermal parameters using the formula  $RMSTD = \sqrt{B/8\pi^2}$ . This estimation is meaningful because we obtained a high-resolution structure (1.35 Å, Table 2) [30]. The values were then compared to the average RMSF obtained from three MD simulations of NRP1-b1 hexavariant in complex with the peptide (Figure S4). RMSTD and RMSF show similar profiles, indicating a general agreement between both indicators. However, the values of RMSF are noticeably larger (up to 1.7 Å) in the most dynamic regions of the polypeptide chain, while RMSTDs consistently remain below 0.75 Å. One plausible explanation for this discrepancy is that proteins in the crystalline state typically exhibit reduced mobility compared to their counterparts in solution. Overall, RMSF and RMSTD show a correlation coefficient of 0.735, indicating a moderate correlation between the two parameters. The values of RMSF are noticeably larger (up to 1.7 Å) in the most dynamic region of the polypeptide chain, while RMSTDs consistently remain below 0.75 Å. Several MD simulations of the literature [31–34] have also shown larger fluctuations in solution compared to the crystal environment. Our observations agree with these references.

Looking more closely at the RMSF profiles (Figure 5), the loops interacting with the peptide all display fluctuation peaks whether the peptide is absent or present. The loop 347–350 (L3 loop) bearing Glu348 displays the same mobility. By contrast, a group of loops clustered on the other side of the binding side, the loop 296–300 (L1) bearing Tyr297 and Asn300, the loop 310–313 (L2) bearing Glu312 and the loop 318–322 (L5) bearing Glu319, are more mobile in the absence than in the presence of the ligand. Somehow, the two sides of the binding pocket behave differently with respect to the ligand. One moiety (L1, L2 and L5) stabilizes upon binding, while the other moiety (L3) retains the same mobility. A recent MD study by Alshawaf et al. [35] found similar RMSF profiles in NRP1-b1 when complexed with specialized metabolites 3-*O*-methylquercetin and esculetin. Specifically, the fluctuations in the esculetin/NRP1-b1 complex were similar to those observed in our study of the unbound form, while the 3-*O*-methylquercetin/NRP1-b1 complex was more

similar to our complex with the peptide. Notably, the study found that 3-*O*-methylquercetin had a more favorable energy of interaction with NRP1-b1 than esculetin [35]. Our “bound” RMSF profile may be indicative to a state of NRP1-b1 that allows stable interactions with a ligand.

**Table 2.** Statistics of X-ray diffraction data collection and model refinement.

<b>Data Collection</b>	
Diffraction source	ESRF FIP2-BM07
Wavelength (Å)	0.9795
Space group	<i>P</i> 3 <sub>2</sub> 2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	59.77, 59.77, 174.60
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 120
Resolution range (Å)	44.53–1.35 (1.37–1.35) <sup>1</sup>
Total number of measured intensities	696,189 (16,486) <sup>1</sup>
Number of unique reflections	80,230 (3895) <sup>1</sup>
Average redundancy	8.7 (4.2) <sup>1</sup>
Mean <i>I</i> /sig( <i>I</i> )	28.7 (1.9) <sup>1</sup>
Completeness (%)	99.7 (95.7) <sup>1</sup>
<i>R</i> <sub>merge</sub> <sup>2</sup> ; <i>R</i> <sub>meas</sub> <sup>3</sup>	0.031 (0.681) <sup>1</sup> ; 0.033 (0.779) <sup>1</sup>
CC <sub>1/2</sub> <sup>4</sup>	1.00 (0.69) <sup>1</sup>
Wilson <i>B</i> -factor (Å <sup>2</sup> )	17.9 (Aimless)/21.14 (Buster)
<b>Refinement and structure</b>	
Resolution range (Å)	19.57–1.35 (1.36–1.35) <sup>1</sup>
Number of reflections	80,203 (1605) <sup>1</sup>
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> <sup>5</sup>	0.1942/0.2100 (0.2826/0.2760) <sup>1</sup>
Correlation <i>F</i> <sub>o</sub> – <i>F</i> <sub>c</sub> / <i>F</i> <sub>o</sub> – <i>F</i> <sub>c</sub> <sub>free</sub>	0.966/0.963
Total number of atoms	2910
Average <i>B</i> factor (Å <sup>2</sup> )	25.55
<b>Model quality</b>	
RMSZ bond lengths <sup>6</sup>	1.28
RMSZ bond angles <sup>6</sup>	1.14
Ramachandran favored (%)	97.5
Ramachandran allowed (%)	2.4
Rotamer outliers (%)	1.8
Clash-score <sup>7</sup>	10

<sup>1</sup> Values in parentheses are for the highest resolution shell; <sup>2</sup>  $R_{merge} = \frac{\sum_h \sum_i |I_{hi} - \langle I_h \rangle|}{\sum_h \sum_i \langle I_h \rangle}$ ; <sup>3</sup>  $R_{meas} = \frac{\sum_h \sum_i \left( \frac{n_h}{n_h - 1} \right)^{1/2} |I_{hi} - \langle I_h \rangle|}{\sum_h \sum_i \langle I_h \rangle}$  (with *I*<sub>hi</sub> being the intensity of an individual observation of the reflection *h* and  $\langle I_h \rangle$  being the average of all symmetry-related or replicate observations); <sup>4</sup> CC<sub>1/2</sub> is the correlation coefficient of the mean intensities between two random half-sets of data. <sup>5</sup>  $R_{work} = \frac{\sum_h ||F_o| - |F_c||}{\sum_h |F_o|}$ , 95% of the reflections, *R*<sub>free</sub> same formula (5% of the reflections) (*F*<sub>o</sub> and *F*<sub>c</sub> observed and calculated structure factors, respectively). <sup>6</sup> RMSZ: root mean square Z-score. <sup>7</sup> The MolProbity clash-score is the number of serious clashes per 1000 atoms.

Comparing the fluctuation profiles of the WT and of the mutated sequences of NRP1-b1 (Figure 5), only one major difference can be noticed: the loop 411–416 (L4), interacting with the peptide, displays a fluctuation peak for one of the unbound trajectories on the WT sequence, whereas this loop is quite rigid in all trajectories recorded for the modified sequence.

The loops 282–289 and 373–378, located in the bottom of the structure, display high mobility in all conditions. The loop 282–289, containing the charged and polar sequence ESGEIHSD, becomes more mobile in the absence of peptide. The sequence ESGEIHSD (residues 282–289) is located at the surface of b1 domain close to the surface of a2 domain in NRP1 structures, which contain a2, b1 and b2 domains (PDB entry 2QQM, [36]) or a1, a2, b1 and b2 domains (PDB entry 4GZ9, [37]). A similar configuration is also visible in the more recent cryo-EM structure of the Sema3A/PlexinA4/NRP1 tripartite complex [1]. As

the b1 and a2 interface do not form direct contact in any of these structures, it is difficult to speculate on the precise functional effect of the mobility of the region, but this mobility may have an influence on the propagation of conformational signals during the physiological processes.

### 3. Materials and Methods

#### 3.1. Protein Production and Purification

The gene of NRP1-b1 hexavariant (residues Met272 to Thr427 of NRP1 with mutations Glu277Lys, Glu285Lys, Asp289Lys, Glu367Lys, Lys373Glu, Lys397Glu) was cloned into pET15b-NRP16mut-6His-3C (Novagen, Pretoria, South Africa) and expressed in *Escherichia coli* after induction at 18 °C. Cells were grown in Terrific-Broth at 18 °C to optical density (OD) = 0.36. After 8 h 25 min at 18 °C, they were induced with 0.2 mM Isopropyl  $\beta$ -D-1 thiogalactopyranoside. After growth at 18 °C for 10 h 25 min to OD = 0.7, cells were harvested by centrifugation, lysed and centrifuged. Proteins were purified on HIS-Select nickel affinity resin (Sigma-Aldrich, St. Louis, MO, USA) in 50 mM Tris (pH 8.0) and 300 mM NaCl with a 250 mM imidazole gradient. The protein was further purified by gel filtration using a Superdex75 HiLoad 16/60 column (GE Healthcare, Piscataway, NJ, USA) equilibrated and run in buffer A (50 mM Tris, pH 8, 300 mM NaCl, 20 mM imidazole). Analytical gel filtration experiments were performed using a Superdex 75 10/16 column (GE Healthcare) in buffer A. The protein was concentrated to 63 mg/mL by centrifugation on an Amicon ultrafiltration unit with a 10 kDa molecular weight cutoff, in a solution of Tris/HCl pH 8, NaCl 50 mM.

#### 3.2. KDKPPR Synthesis on Solid Phase

The KDKPPR peptide was synthesized using the automated ResPepXL peptide synthesizer, with a Fmoc/tBu methodology. The side chains of arginine, lysine and aspartic acid were protected by Pbf, OtBu and Boc groups. A Fmoc-Arg(Pbf)-Wang resin swelled in DCM was used. The Fmoc group was removed by a piperidine solution (20% in DMF), and this step was performed two times (the first for 4 min and the second for 7 min). Then, the next AA was grafted by adding an excess of Fmoc-AA-OH (6 eq), HBTU (5 eq), NMP (3 eq) and NMM (10 eq) in DMF, and this step was repeated two times for 18 min. A last step of capping, using a solution of acetic anhydride (5% in DMF), was performed for 5 min to trap all amino functions that did not react. Deprotection, coupling and capping steps were repeated until the end of the synthesis of the peptide. After a last Fmoc deprotection, the resin was dried under vacuum and then cleaved (with full deprotection of lateral chains) using TFA/TIPS/water (92.5/2.5/5, *v/v/v*) for 2 h. The acidic resin was filtered and washed with DCM and EtOH. The filtrate was dried under vacuum, and the compound was precipitated in diethylether by centrifugation. TSK gel Amide-80 column was used for HILIC purification of KDKPPR peptide using acetonitrile/water (0.1% TFA; 95/5 (*v/v*) to 55/45 (*v/v*) gradient) for 15 min, followed by an isocratic elution (0.1% TFA; 55/45, *v/v*) for 10 min at a flow of 12 mL/min ( $R_t = 20.8$  min). KDKPPR was isolated as a white powder with a yield of 64% and a purity of 95% (UV-vis detection at 214 nm).

#### 3.3. Crystallization

Crystallization was conducted by sitting-drop vapor diffusion method. The reservoir solutions were prepared by mixing 0.3  $\mu$ L of commercial reservoir solutions (screens) and 0.3  $\mu$ L of protein solution. The crystallization trials of NRP1-b1 hexavariant were carried out in the presence of the KDKPPR hexapeptide at 20 °C. Crystals appeared with a JCSGplus solution composed of 0.2 M ammonium citrate, 0.1 M bis-tris pH 5.5 and 25% *w/v* PEG 3350.

#### 3.4. X-ray Diffraction Data Collection and Crystal Structure Determination

Crystals of NRP1-b1 hexavariant that appeared suitable for X-ray diffraction data collection were quickly soaked in their mother liquor supplemented with 20% glycerol (*v/v*),



before flash freezing in a nitrogen stream at 100 K. Preliminary X-ray diffraction experiments were carried out in-house on an Agilent SuperNova diffractometer (Oxford Diffraction, Oxford, UK) equipped with a CCD detector, and high-resolution data were further collected by ESRF synchrotron on beamline BM07 (Grenoble, France). The data set was indexed and integrated with XDS [38], scaled, and merged with Aimless [39] from the CCP4 suite [40]. The atomic structure was solved by molecular replacement using MOLREP [41] with the coordinates of NRP1-b1 wild type (PDB code 5C7G, [19]) as the search model. The structure was manually adjusted with Coot [42] and refined with Buster [43]. Structure validation was performed with MolProbity [44] and the wwPDB validation server (<http://validate.wwpdb.org>, (accessed on 22 december 2022)). Diffraction data and refinement statistics are shown in Table 2. Figures of the protein structures were generated with Pymol (Schrödinger LLC, New York, NY, USA), MoProViewer [27] and Ligplot+ [45], and cleft volume calculations were performed with 3V [46]. Coordinates and structure factors were deposited in the Protein Data Bank (PDB ID: 8PFE, DOI:10.2210/pdb8pfe/pdb).

### 3.5. Nucleophilic Influence Zones

The Nucleophilic Influence Zones were calculated from the electrostatic potential using Charger module of MoProviewer [24,27]. The electrostatic potential was generated from an electron density model, based on transferred multipolar parameters of the ELMAM2 library [21].

### 3.6. Hirshfeld Analysis

MoProViewer software [27] was used to investigate the intermolecular interactions and the contacts enrichment on the Hirshfeld interface between the protein molecules and the ligand peptides. The intermolecular interactions were evaluated by computing the enrichment ratios (Table 1) in order to highlight which contacts are favored. The enrichment values are obtained as the ratio between the proportions of actual contacts  $C_{XY}$  and the equiprobable (random) contacts  $R_{XY}$ , the latter being obtained by probability products ( $R_{XY} = S_X S_Y$ ).

Contacts X...Y, which are over-represented with respect to the share of X and Y chemical species on the Hirshfeld surface, have enrichments larger than unity. They are likely to represent interactions that are attractive from an electrostatic point of view and shall be the driving force in the complex formation [28]. Interactions between atoms that have electric charges of the same sign are repulsive and are generally under-represented ( $E < 1$ ).

### 3.7. Molecular Dynamics

The protein and peptide chains C and D were selected from the X-ray crystallographic structure. The hydrogen atoms were added, and the flip of side chains was optimized using the Molprobity server [44]. The NRP1-b1 crystal structure with six mutations was unmodified, whereas for the wild-type (WT) system, the mutations Lys277Glu, Lys285Glu, Lys289Asp, Lys367Glu, Glu373Lys and Glu397Lys were introduced to return to the WT sequence of NRP1 for both protein sequences. The bound and unbound systems were simulated.

For each previously described system, the protein was embedded in a water box. Sodium and chloride counterions were added to obtain an ionic concentration of 0.15 M. The total number of atoms was about 31,000 in both cases. All MD simulations were performed using NAMD 2.14 [47], with the CHARMM36 force field [48] for protein and the TIP3P model for water [49]. A cutoff of 12 Å and a switching distance of 10 Å were used for non-bonded interactions, while long-range electrostatic interactions were calculated with the Particle Mesh Ewald (PME) method [50]. The RATTLE algorithm [51] was used to keep rigid all covalent bonds involving hydrogen atoms, enabling a time step of 2 fs. At the beginning of each trajectory, the system was minimized for 20,000 steps, and it was then heated up gradually from 0 K to 310 K in 31,000 integration steps. Finally, the system was

equilibrated for 1 ns in the NPT ensemble at 310 K. During the equilibration stage, the C $\alpha$  atoms were kept fixed. Simulations were then performed in the NPT ensemble ( $P = 1$  bar,  $T = 310$  K), with all atoms free to move. Atomic coordinates were saved every 10 ps. For each trajectory, 200 ns of production and the trajectories were triplicated for a cumulative trajectory duration of 3  $\mu$ s.

#### 4. Conclusions

In this study, we have reported the crystal structure of a hexavariant of the domain b1 of human NRP1 in complex with the KDKPPR peptide. The mutant was designed to modify the monomer assembly observed in the crystal packing of the unbound form [9,19], in which the VEGF-binding pocket of NRP1-b1 is inaccessible. Molecular dynamic trajectories permitted investigating the differences in the structures of the wild type and variant. Both structures produced similar internal flexibility and protein/peptide interaction. This showed that the ability of the protein to bind small ligands was not affected by the designed mutations. As part of our future search for ligands, we need to check that the dissociation constant ( $K_D$ ) of the molecules tested is the same for mutated and wild-type NRP1-b1.

The NRP1-b1 hexavariant crystallized in a new crystal form, in which the KDKPPR peptide was not involved in the cohesion of the solid state. In the crystal, the peptide-binding site was observed to communicate with a solvent cavity large enough to diffuse small molecules. Therefore, ligand soaking in the crystal of the unbound form of NRP1-b1 hexavariant could be considered as a strategy to prepare NRP1-b1 complexes with peptides that target the pocket where VEGF binds.

The structure of the NRP1-b1 hexavariant in complex with the KDKPPR peptide was analyzed with two original tools. First, the Nucleophilic Influence Zones (NIZ) of the ligand-binding cleft were analyzed. They revealed two additional residues (Tyr297 and Glu348) as probable attractors of the ligand electrophilic groups and two residues (Asp320 and Glu348) in the electrostatic steering of the ligand in its binding site. Secondly, the enrichment of contacts was calculated to analyze the interactions between the protein and the peptide. This metric has provided valuable insights into the diversity and specificity of the protein/ligand interaction. The complex was mainly stabilized by a notable presence of strong N-H...O and O-H...O hydrogen bonds, which were crucial due to the loop-rich nature of the VEGF-binding site. Indeed, these loops exhibited mobility both in the unbound and bound forms, as suggested by the MD simulation.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/molecules28145603/s1>, Figure S1. Large aqueous cavity in the crystal structure of NRP1-b1 hexavariant; Figure S2: Comparison of the packing of NRP1-b1 in crystal forms II (a) and III (b); Figure S3: Environment of the ligand in crystal forms II, III, IV and V; Figure S4: Comparison of atomic root mean square fluctuations (RMSFs) and root mean square thermal displacements (RMSTDs); Table S1: Intermolecular hydrogen bonds between NRP1-b1 chains in hexavariant crystal; Table S2: Crystal forms of NRP1-b1 fragment in the unbound form or in complex with small ligands.

**Author Contributions:** Conceptualization, C.D., C.F., C.J., F.F., S.A. and T.E.M.; methodology, C.D., C.J. and T.E.M.; formal analysis, I.G., S.A., B.G., E.M., F.F., C.D., C.J., S.M. and T.E.M.; investigation, I.G., S.A., B.G., E.M., F.F., C.D., C.J., S.M. and T.E.M.; writing—original draft preparation, I.G., T.E.M., F.F., C.D. and C.J.; writing—review and editing, C.F., I.G., S.A., B.G., E.M., F.F., C.D., C.J., S.M., M.B.-H. and T.E.M.; supervision, C.D. and C.J.; project administration, C.D. and C.J.; funding acquisition, C.D. and C.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** PDB data (8PFE) are made freely available by the wwPDB (<https://www.wwpdb.org/> (accessed on 21 July 2023)).

**Acknowledgments:** The authors appreciated the access to the “Plateforme de mesures de diffraction X” of the Université de Lorraine. We acknowledge ESRF (Grenoble, France) for providing synchrotron radiation facilities, and we thank the staffs of BM07 beamlines for assistance.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Sample Availability:** Not applicable.

## References

1. Lu, D.; Shang, G.; He, X.; Bai, X.C.; Zhang, X. Architecture of the Sema3A/PlexinA4/Neuropilin tripartite complex. *Nat. Commun.* **2021**, *12*, 3172. [[CrossRef](#)] [[PubMed](#)]
2. Soker, S.; Miao, H.-Q.; Nomi, M.; Takashima, S.; Klagsbrun, M. VEGF165 mediates formation of complexes containing VEGFR-2 and neuropilin-1 that enhance VEGF165-receptor binding. *J. Cell. Biochem.* **2002**, *85*, 357–368. [[CrossRef](#)] [[PubMed](#)]
3. Soker, S.; Takashima, S.; Miao, H.Q.; Neufeld, G.; Klagsbrun, M. Neuropilin-1 Is Expressed by Endothelial and Tumor Cells as an Isoform-Specific Receptor for Vascular Endothelial Growth Factor. *Cell* **1998**, *92*, 735–745. [[CrossRef](#)]
4. Chaudhary, B.; Khaled, Y.S.; Ammori, B.J.; Elkord, E. Neuropilin 1: Function and therapeutic potential in cancer. *Cancer Immunol. Immunother.* **2014**, *63*, 81–99. [[CrossRef](#)]
5. Daly, J.L.; Simonetti, B.; Klein, K.; Chen, K.-E.; Williamson, M.K.; Antón-Plágaro, C.; Shoemark, D.K.; Simón-Gracia, L.; Bauer, M.; Hollandi, R.; et al. Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **2020**, *370*, 861–865. [[CrossRef](#)]
6. Lerouge, L.; Gries, M.; Chateau, A.; Daouk, J.; Lux, F.; Rocchi, P.; Cedervall, J.; Olsson, A.K.; Tillement, O.; Frochot, C.; et al. Targeting Glioblastoma-Associated Macrophages for Photodynamic Therapy Using AGuIX((R))-Design Nanoparticles. *Pharmaceutics* **2023**, *15*, 997. [[CrossRef](#)] [[PubMed](#)]
7. Liu, S.D.; Zhong, L.P.; He, J.; Zhao, Y.X. Targeting neuropilin-1 interactions is a promising anti-tumor strategy. *Chin. Med. J.* **2020**, *134*, 508–517. [[CrossRef](#)]
8. Pellet-Many, C.; Frankel, P.; Jia, H.; Zachary, I. Neuropilins: Structure, function and role in disease. *Biochem. J.* **2008**, *411*, 211–226. [[CrossRef](#)]
9. Lee, C.C.; Kreuzsch, A.; McMullan, D.; Ng, K.; Spraggon, G. Crystal Structure of the Human Neuropilin-1 b1 Domain. *Structure* **2003**, *11*, 99–108. [[CrossRef](#)]
10. Powell, J.; Mota, F.; Steadman, D.; Soudy, C.; Miyachi, J.T.; Crosby, S.; Jarvis, A.; Reisinger, T.; Winfield, N.; Evans, G.; et al. Small Molecule Neuropilin-1 Antagonists Combine Antiangiogenic and Antitumor Activity with Immune Modulation through Reduction of Transforming Growth Factor Beta (TGFbeta) Production in Regulatory T-Cells. *J. Med. Chem.* **2018**, *61*, 4135–4154. [[CrossRef](#)]
11. Mota, F.; Fotinou, C.; Rana, R.R.; Chan, A.W.E.; Yelland, T.; Arooz, M.T.; O’Leary, A.P.; Hutton, J.; Frankel, P.; Zachary, I.; et al. Architecture and hydration of the arginine-binding site of neuropilin-1. *FEBS J.* **2018**, *285*, 1290–1304. [[CrossRef](#)] [[PubMed](#)]
12. Parker, M.W.; Xu, P.; Li, X.; Vander Kooi, C.W. Structural basis for selective vascular endothelial growth factor-A (VEGF-A) binding to neuropilin-1. *J. Biol. Chem.* **2012**, *287*, 11082–11089. [[CrossRef](#)] [[PubMed](#)]
13. Jarvis, A.; Allerton, C.K.; Jia, H.; Herzog, B.; Garza-Garcia, A.; Winfield, N.; Ellard, K.; Aqil, R.; Lynch, R.; Chapman, C.; et al. Small molecule inhibitors of the neuropilin-1 vascular endothelial growth factor A (VEGF-A) interaction. *J. Med. Chem.* **2010**, *53*, 2215–2226. [[CrossRef](#)]
14. Kamarulzaman, E.E.; Vanderesse, R.; Gazzali, A.M.; Barberi-Heyob, M.; Boura, C.; Frochot, C.; Shawkataly, O.; Aubry, A.; Wahab, H.A. Molecular modelling, synthesis and biological evaluation of peptide inhibitors as anti-angiogenic agent targeting neuropilin-1 for anticancer application. *J. Biomol. Struct. Dyn.* **2017**, *35*, 26–45. [[CrossRef](#)]
15. Kamarulzaman, E.E.; Gazzali, A.M.; Acherar, S.; Frochot, C.; Barberi-Heyob, M.; Boura, C.; Chaimbault, P.; Sibille, E.; Wahab, H.A.; Vanderesse, R. New Peptide-Conjugated Chlorin-Type Photosensitizer Targeting Neuropilin-1 for Anti-Vascular Targeted Photodynamic Therapy. *Int. J. Mol. Sci.* **2015**, *16*, 24059–24080. [[CrossRef](#)] [[PubMed](#)]
16. Bechet, D.; Mordon, S.R.; Guillemin, F.; Barberi-Heyob, M.A. Photodynamic therapy of malignant brain tumours: A complementary approach to conventional therapies. *Cancer Treat. Rev.* **2014**, *40*, 229–241. [[CrossRef](#)]
17. Gries, M.; Thomas, N.; Daouk, J.; Rocchi, P.; Choulier, L.; Jubreaux, J.; Pierson, J.; Reinhard, A.; Jouan-Hureauux, V.; Chateau, A.; et al. Multiscale Selectivity and in vivo Biodistribution of NRP-1-Targeted Theranostic AGuIX Nanoparticles for PDT of Glioblastoma. *Int. J. Nanomed.* **2020**, *15*, 8739–8758. [[CrossRef](#)]
18. Thomas, E.; Colombeau, L.; Gries, M.; Peterlini, T.; Mathieu, C.; Thomas, N.; Boura, C.; Frochot, C.; Vanderesse, R.; Lux, F.; et al. Ultrasmall AGuIX theranostic nanoparticles for vascular-targeted interstitial photodynamic therapy of glioblastoma. *Int. J. Nanomed.* **2017**, *12*, 7075–7088. [[CrossRef](#)]
19. Richard, M.; Chateau, A.; Jelsch, C.; Didierjean, C.; Manival, X.; Charron, C.; Maignet, B.; Barberi-Heyob, M.; Chapleur, Y.; Boura, C.; et al. Carbohydrate-based peptidomimetics targeting neuropilin-1: Synthesis, molecular docking study and in vitro biological activities. *Bioorgan. Med. Chem.* **2016**, *24*, 5315–5325. [[CrossRef](#)]
20. Jelsch, C.; Longhi, S.; Cambillau, C. Packing forces in nine crystal forms of cutinase. *Proteins Struct. Funct. Genet.* **1998**, *31*, 320–333. [[CrossRef](#)]
21. Domagala, S.; Fournier, B.; Liebschner, D.; Guillot, B.; Jelsch, C. An improved experimental databank of transferable multipolar atom models—ELMAM2. Construction details and applications. *Acta Crystallogr. A* **2012**, *68*, 337–351. [[CrossRef](#)] [[PubMed](#)]

22. wwPDB consortium. Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **2019**, *47*, D520–D528. [[CrossRef](#)] [[PubMed](#)]
23. Mata, I.; Molins, E.; Espinosa, E. Zero-Flux Surfaces of the Electrostatic Potential: The Border of Influence Zones of Nucleophilic and Electrophilic Sites in Crystalline Environment. *J. Phys. Chem. A* **2007**, *111*, 9859–9870. [[CrossRef](#)] [[PubMed](#)]
24. Vuković, V.; Leduc, T.; Jelić-Matošević, Z.; Didierjean, C.; Favier, F.; Guillot, B.; Jelsch, C. A rush to explore protein–ligand electrostatic interaction energy with Charger. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77*, 1292–1304. [[CrossRef](#)] [[PubMed](#)]
25. Wade, R.C.; Gabdouliline, R.R.; Lüdemann, S.K.; Lounnas, V. Electrostatic steering and ionic tethering in enzyme–ligand binding: Insights from simulations. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5942–5949. [[CrossRef](#)]
26. Hirshfeld, F.L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129–138. [[CrossRef](#)]
27. Guillot, B.; Enrique, E.; Huder, L.; Jelsch, C. MoProViewer: A tool to study proteins from a charge density science perspective. *Acta Crystallogr. Sect. A* **2014**, *70*, C279. [[CrossRef](#)]
28. Jelsch, C.; Ejsmont, K.; Huder, L. The enrichment ratio of atomic contacts in crystals, an indicator derived from the Hirshfeld surface analysis. *IUCrJ* **2014**, *1*, 119–128. [[CrossRef](#)]
29. Jelsch, C.; Bibila Mayaya Bisseyou, Y. Atom interaction propensities of oxygenated chemical functions in crystal packings. *IUCrJ* **2017**, *4*, 158–174. [[CrossRef](#)]
30. Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M.T. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.* **2019**, *119*, 1626–1665. [[CrossRef](#)]
31. Ahlstrom, L.S.; Miyashita, O. Packing interface energetics in different crystal forms of the  $\lambda$  Cro dimer. *Proteins Struct. Funct. Bioinform.* **2014**, *82*, 1128–1141. [[CrossRef](#)]
32. Ahlstrom, L.S.; Vorontsov, I.I.; Shi, J.; Miyashita, O. Effect of the Crystal Environment on Side-Chain Conformational Dynamics in Cyanovirin-N Investigated through Crystal and Solution Molecular Dynamics Simulations. *PLoS ONE* **2017**, *12*, e0170337. [[CrossRef](#)] [[PubMed](#)]
33. Janowski, P.A.; Liu, C.; Deckman, J.; Case, D.A. Molecular dynamics simulation of triclinic lysozyme in a crystal lattice. *Protein Sci.* **2016**, *25*, 87–102. [[CrossRef](#)] [[PubMed](#)]
34. Kuzmanic, A.; Pannu, N.S.; Zagrovic, B. X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nat. Commun.* **2014**, *5*, 3220. [[CrossRef](#)] [[PubMed](#)]
35. Alshawaf, E.; Hammad, M.M.; Marafie, S.K.; Ali, H.; Al-Mulla, F.; Abubaker, J.; Mohammad, A. Discovery of natural products to block SARS-CoV-2 S-protein interaction with Neuropilin-1 receptor: A molecular dynamics simulation approach. *Microb. Pathog.* **2022**, *170*, 105701. [[CrossRef](#)]
36. Appleton, B.A.; Wu, P.; Maloney, J.; Yin, J.; Liang, W.-C.; Stawicki, S.; Mortara, K.; Bowman, K.K.; Elliott, J.M.; Desmarais, W.; et al. Structural studies of neuropilin/antibody complexes provide insights into semaphorin and VEGF binding. *EMBO J.* **2007**, *26*, 4902–4912. [[CrossRef](#)]
37. Janssen, B.J.C.; Malinauskas, T.; Weir, G.A.; Cader, M.Z.; Siebold, C.; Jones, E.Y. Neuropilins lock secreted semaphorins onto plexins in a ternary signaling complex. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1293–1299. [[CrossRef](#)]
38. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132. [[CrossRef](#)]
39. Evans, P.R.; Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204–1214. [[CrossRef](#)]
40. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 235–242. [[CrossRef](#)]
41. Vagin, A.; Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 22–25. [[CrossRef](#)]
42. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501. [[CrossRef](#)]
43. Smart, O.S.; Womack, T.O.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. Exploiting structure similarity in refinement: Automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. Sect. D* **2012**, *68*, 368–380. [[CrossRef](#)] [[PubMed](#)]
44. Williams, C.J.; Headd, J.J.; Moriarty, N.W.; Prisant, M.G.; Videau, L.L.; Deis, L.N.; Verma, V.; Keedy, D.A.; Hintze, B.J.; Chen, V.B.; et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018**, *27*, 293–315. [[CrossRef](#)] [[PubMed](#)]
45. Laskowski, R.A.; Swindells, M.B. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **2011**, *51*, 2778–2786. [[CrossRef](#)]
46. Voss, N.R.; Gerstein, M. 3V: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010**, *38*, W555–W562. [[CrossRef](#)]
47. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [[CrossRef](#)]
48. Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.M.; Mittal, J.; Feig, M.; MacKerell, A.D., Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone  $\phi$ ,  $\psi$  and Side-Chain  $\chi_1$  and  $\chi_2$  Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273. [[CrossRef](#)] [[PubMed](#)]

49. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [[CrossRef](#)]
50. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [[CrossRef](#)]
51. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

# Supplementary Materials

## New crystal form of Human Neuropilin-1 b1 fragment with six electrostatic mutations complexed with KDKPPR peptide ligand

Goudiaby et al.

### Table of contents

<b>Table S1.</b> Intermolecular hydrogen bonds between NRP1-b1 chains in hexavariant crystal... 2	2
<b>Table S2.</b> Crystal forms of NRP1-b1 fragment in the unbound form or in complex with small ligands..... 3	3
<b>Figure S1.</b> Large aqueous cavity in the crystal structure of NRP1-b1 hexavariant. .... 4	4
<b>Figure S2.</b> Comparison of the packing of NRP1-b1 in crystal forms II (a) and III (b). .... 5	5
<b>Figure S3.</b> Environment of the ligand in crystal forms II, III, IV and V. .... 6	6
<b>Figure S4.</b> Comparison of atomic root mean square fluctuations (RMSF) and root mean square thermal displacements (RMSTD) ..... 7	7

**Table S1.** Intermolecular hydrogen bonds between NRP1-b1 chains in hexavariant crystal.

Residue 1	Residue 2	Distance (Å)
NZ-K277(A)	OXT-T427(A) <sup>i</sup>	2.67
O-E282(A)	OH-Y322(A) <sup>ii</sup>	2.56
O-S283(A)	OG-S321(A) <sup>ii</sup>	2.78
NZ-K285(A)	OE1-E319(A) <sup>ii</sup>	2.63
NH1-R323(A)	OD1-D389(C) <sup>iii</sup>	2.85
NH2-R323(A)	OD2-D389(C) <sup>iii</sup>	2.85
NZ-K352(A)	O-S294(C) <sup>iv</sup>	3.05
OD1-D389(A)	NH1-R323(C) <sup>iv</sup>	2.77
OD2-D389(A)	NH2-R323(C) <sup>iv</sup>	2.70
N-K425(A)	O-T427(A) <sup>i</sup>	2.76
O-K425(A)	N-T427(A) <sup>i</sup>	2.90
O-E282(C)	OH-Y322(C) <sup>v</sup>	2.58
O-S283(C)	OG-S321(C) <sup>v</sup>	2.99

**Symmetry Codes :** (i) -1+Y, 1+X, -Z (ii) -1+Y, X, -Z (iii) Y, 1+X, -Z (iv) 1+X-Y, 1-Y, -Z+1/3 ;  
(v) 1+X-Y, 1-Y, -Z+1/3



**Table S2.** Crystal forms of NRP1-b1 fragment in the unbound form or in complex with small ligands

Form	Z'	Space group	<i>a, b, c</i> (Å), $\alpha, \beta, \gamma$ (°)	Res. (Å)*	PDB entries
I	1	<i>P4<sub>1</sub>2<sub>1</sub>2</i>	62.3, 62.3, 86.0, 90, 90, 90	1.45	<b>5C7G</b> <sup>1</sup> , <b>4RN5</b> <sup>2</sup> , <b>1KEX</b> <sup>3</sup> ,
II <sup>#</sup>	2	<i>P2<sub>1</sub></i>	40.7, 89.2, 41.6, 90, 99, 90	1.38	<b>5JGI</b> <sup>4</sup> , <b>5J1X</b> <sup>5§</sup> , <b>5IJR</b> <sup>6</sup> , <b>5JHK</b> <sup>7</sup> , <b>5JGQ</b> <sup>8</sup> , <b>5IYY</b> <sup>9</sup> , <b>3I97</b> <sup>10</sup>
III <sup>#</sup>	1	<i>P4<sub>1</sub></i>	43.4, 43.4, 91.2, 90, 90, 90	0.9	<b>6FMC</b> <sup>11#</sup>
IV	1	<i>P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></i>	38.9, 40.0, 97.6, 90, 90, 90	1.06	<b>6TKK</b> <sup>12</sup>
V	4	<i>P2<sub>1</sub>2<sub>1</sub>2</i>	89.9, 89.9, 108.3, 90, 90, 90	2.36	<b>7JJC</b> <sup>13</sup>
VI	2	<i>P3<sub>2</sub>2<sub>1</sub></i>	59.8, 59.8, 174.6, 90, 90, 120	1.35	<b>8PFE</b> <sup>14</sup>

\* Highest resolution obtained in a crystal form; § the crystal form of 5J1X is slightly different from that of form II. The *c* parameter has doubled, the space group remains *P2<sub>1</sub>* and *Z'* is equal to 4; # The packing of NRP1-b1 domain is identical in crystal forms II and III (see Figure S2).

Ligands observed in the VEGF pocket:

<sup>1</sup> Bicine; <sup>2</sup> Acetate ion; <sup>3</sup> No ligand; <sup>4</sup> *N*-alpha-l-acetyl-arginine, isomeric SMILES:

(CC(=O)N[C@@H](CCCNC(=N)N)C(=O)O); <sup>5</sup> *N*-2-(*tert*-butoxycarbonyl)-L-arginine, isomeric

SMILES: ([H]/N=C(/N)\NCCC[C@@H](C(=O)O)NC(=O)OC(C)(C)C); <sup>6</sup> L-homoarginine,

isomeric SMILES: (C(CCNC(=N)N)C[C@@H](C(=O)O)N); <sup>7</sup> *N*-(benzenecarbonyl)glycyl-L-

arginine, isomeric SMILES: [H]/N=C(/N)\NCCC[C@@H](C(=O)O)NC(=O)CNC(=O)c1ccccc1; <sup>8</sup>

*N*-2~-(benzenecarbonyl)-L-arginine, isomeric SMILES:

[H]/N=C(/N)\NCCC[C@@H](C(=O)O)NC(=O)c1ccccc1; <sup>9</sup> *N*-2~-[ (benzyloxy)carbonyl]-L-

arginine, isomeric SMILES: [H]/N=C(/N)\NCCC[C@@H](C(=O)O)NC(=O)OCc1ccccc1; <sup>10</sup> (*S*)-

2-(3-(benzo[*c*][1,2,5]thiadiazole-4-sulfonamido)thiophene-2-carboxamido)-5-

guanidinopentanoic acid, isomeric SMILES

[H]/N=C(/N)\NCCC[C@@H](C(=O)O)NC(=O)c1c(ccs1)NS(=O)(=O)c2cccc3c2nsn3; <sup>11</sup> (2~{*S*})-2-

[[3-[[5-[4-(aminomethyl)phenyl]-1-benzofuran-7-yl]sulfonylamino]thiophen-2-

yl]carbonylamino]-5-carbamimidamido-pentanoic acid, isomeric SMILES:

[H]/N=C(/N)\NCCC[C@@H](C(=O)O)NC(=O)c1c(ccs1)NS(=O)(=O)c2cc(cc3c2occ3)c4ccc(cc4)

CN; **12** Ace-Arg-Pro-Gln-Pro-Arg; **13** Asn-Ser-Pro-Arg-Arg-Ala-Arg; **14** Lys-Asp-Lys-Pro-

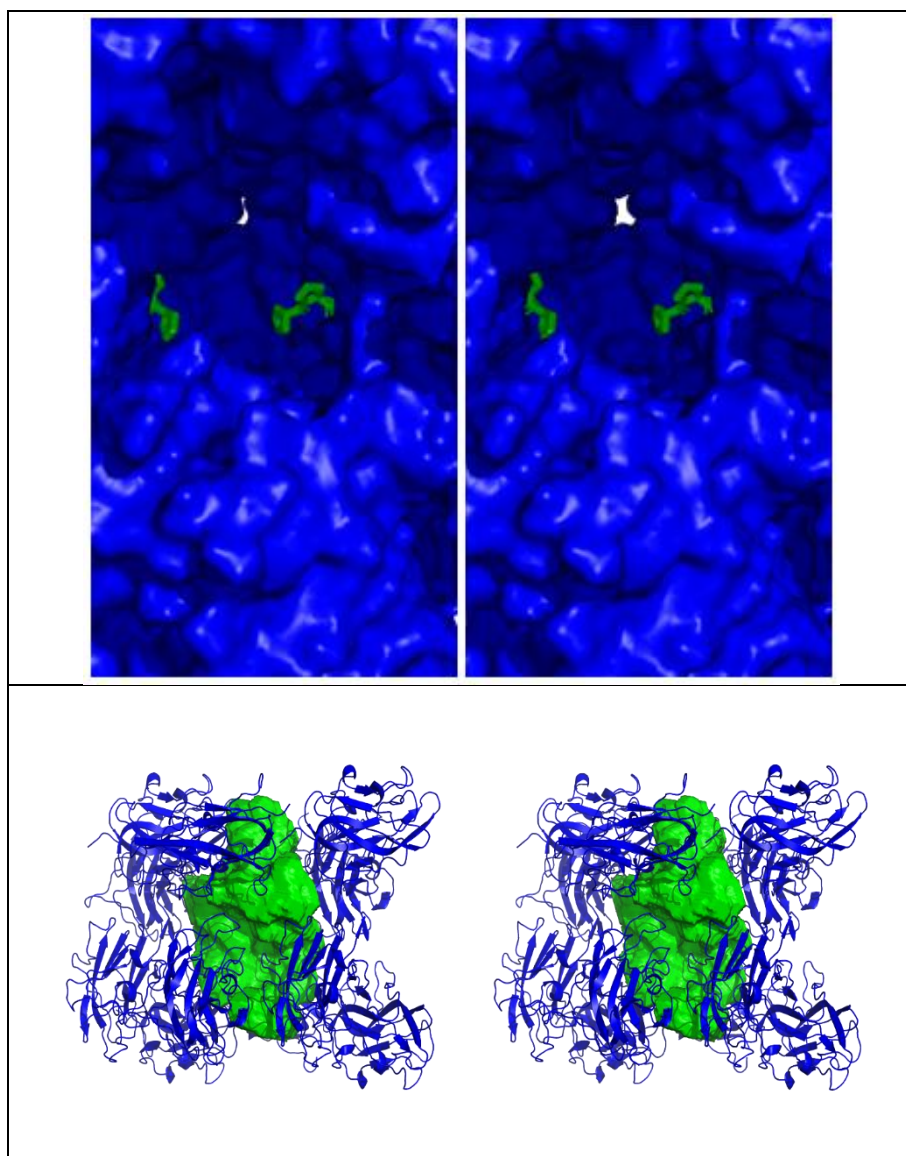
Pro-Arg.



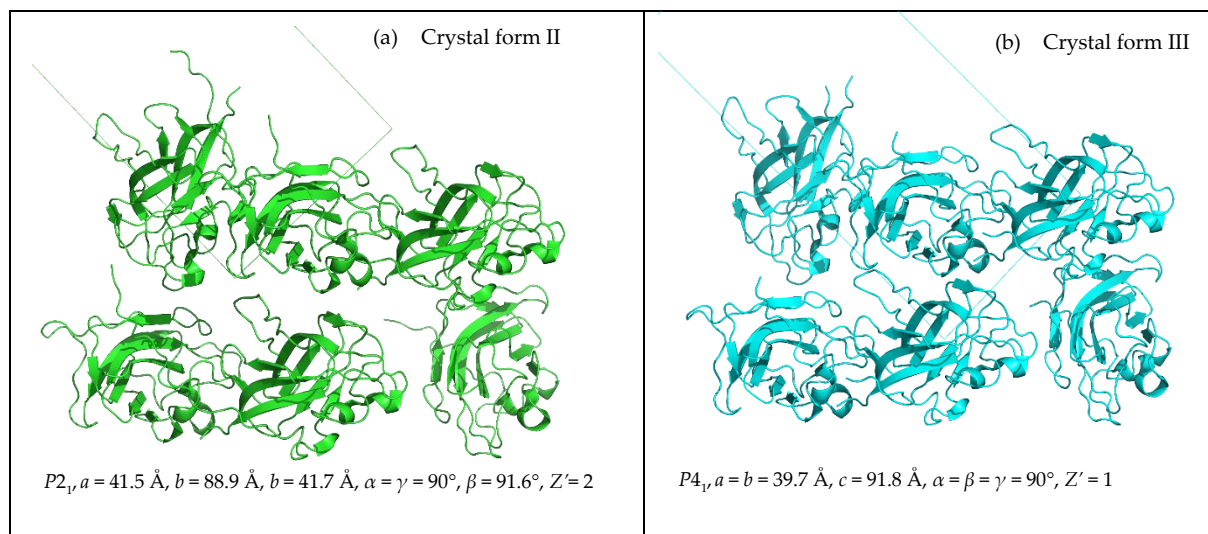
**Figure S1.** Large aqueous cavity in the crystal structure of NRP1-b1 hexavariant.

Top: Stereoview of a large aqueous cavity. A cavity contains two peptides. The electron density map indicates that the KDK moiety of the peptide is not highly ordered within this cavity. The NRP1-b1 chains are shown as a surface and the PPR moiety of the peptides is depicted as sticks.

Bottom: Stereo-view that highlights the volume of the cavity. A volume of approximately 30,000 Å<sup>3</sup> was estimated using 3V (Voss and Gerstein, 2010, NAR, DOI 10.1093/nar/gkq395)

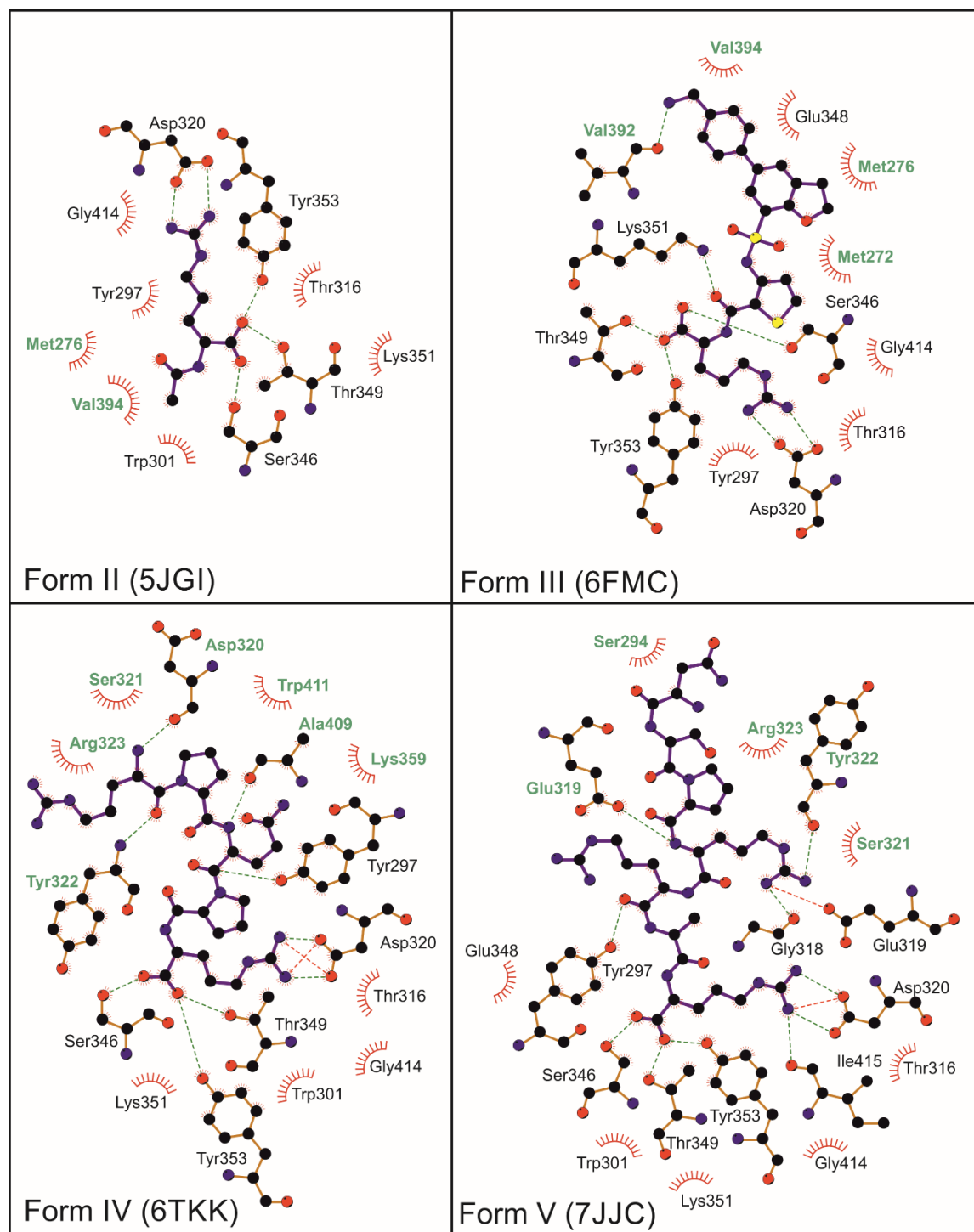


**Figure S2.** Comparison of the packing of NRP1-b1 in crystal forms II (a) and III (b). The figure shows that both crystal forms have the same molecular packing pattern. NRP1-b1 chains are presented in ribbon mode. The left and right parts were constructed using PDB entries 5JGI and 6FMC, respectively.



**Figure S3.** Environment of the ligand in crystal forms II, III, IV and V.

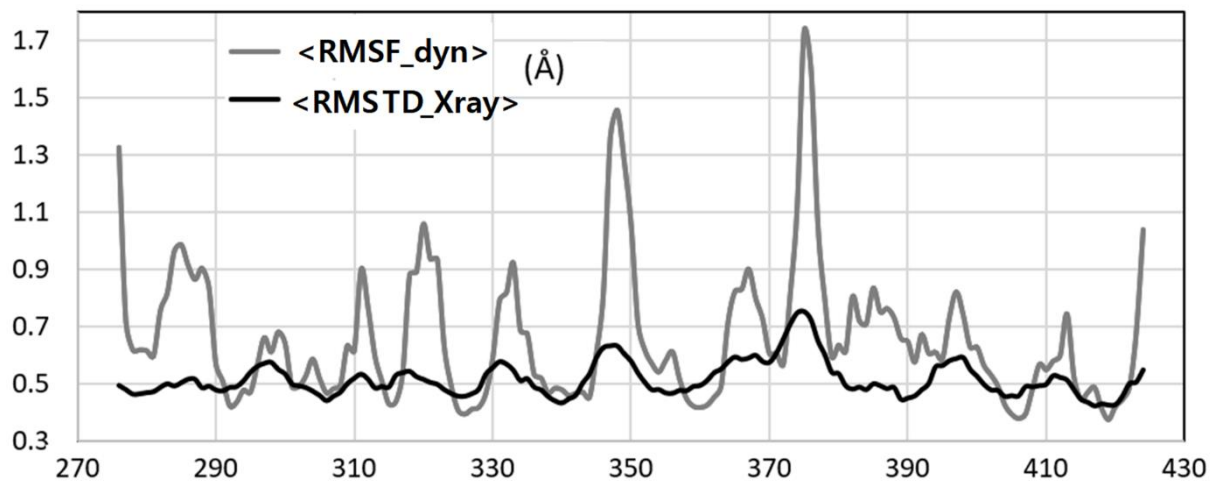
The figures depict ligands within the VEGF-binding pocket. In all cases, at least one ligand is involved in maintaining the cohesion of the crystal packing. The bonds of the ligands are colored in blue. The labels of the symmetry-related molecules are colored in green and displayed in bold characters. Hydrogen bonds are illustrated as dashed lines. The figures were generated using Ligplot+ (Laskowski & Mark, 2011, *J. Chem. Inf. Model.*, DOI 10.1021/ci200227u).



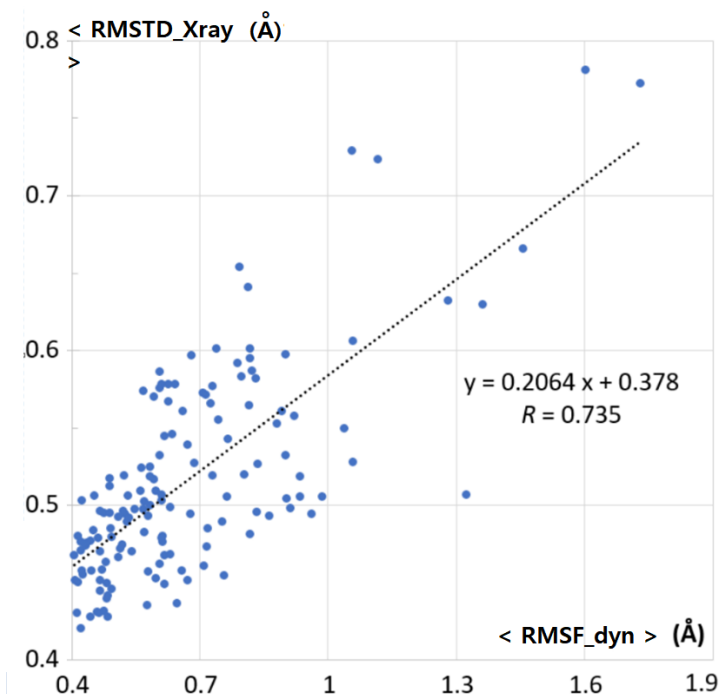
**Figure S4.** Comparison of atomic root mean square fluctuations (RMSF) and root mean square thermal displacements (RMSTD)

RMSF and RMSD of the  $C\alpha$  atom positions averaged, respectively, over the 3 molecular dynamics simulations of NRP1 (holo, 6mut) and over the 2 monomers of the crystal structure. The RMSTD values are derived from the thermal parameters  $B$ . (a) values as a function of residue numbers, (b) scatterplot of (RMSF,RMSD).

(a)



(b)



## E - II ETUDE DES PETITES MOLECULES

### E - II.1 Dérivés de l'acide Bétulinique (BTA)

L'acide bétulinique est une molécule extraite de l'écorce de bouleau qui est insoluble dans l'eau. Les applications pharmaceutiques d'une telle substance sont bien sûr limitées en raison de son insolubilité dans l'eau. Cela rend difficile son absorption et sa diffusion dans l'organisme, ce qui entraîne une faible biodisponibilité générale de cette molécule. De nombreuses études soulignent ce fait comme la principale limitation à son utilisation en tant qu'agent thérapeutique. Pour améliorer sa biodisponibilité, des dérivés de l'acide bétulinique ont été synthétisés avec du 2-chlorocinnamoyl et du chlorure de phénoxyacétyle. Cette partie se concentrera sur trois aspects : tout d'abord, la synthèse des dérivés bétuliniques, puis leur caractérisation par spectroscopie IR, et enfin la cristallisation. L'un des objectifs de cette partie est d'obtenir des cristaux des dérivés de l'acide bétulinique afin de les caractériser.

#### E - II.1.a. Couplage de l'acide bétulinique avec du 2-chlorocinnamoyl

La réaction d'estérification de l'acide bétulinique **01** avec le 2-chlorocinnamoyl **02**, en présence de la triéthylamine dans le dichlorométhane, a donné le composé **03** avec un rendement de 98% (Schéma 1), après purification par chromatographie sur gel de silice.

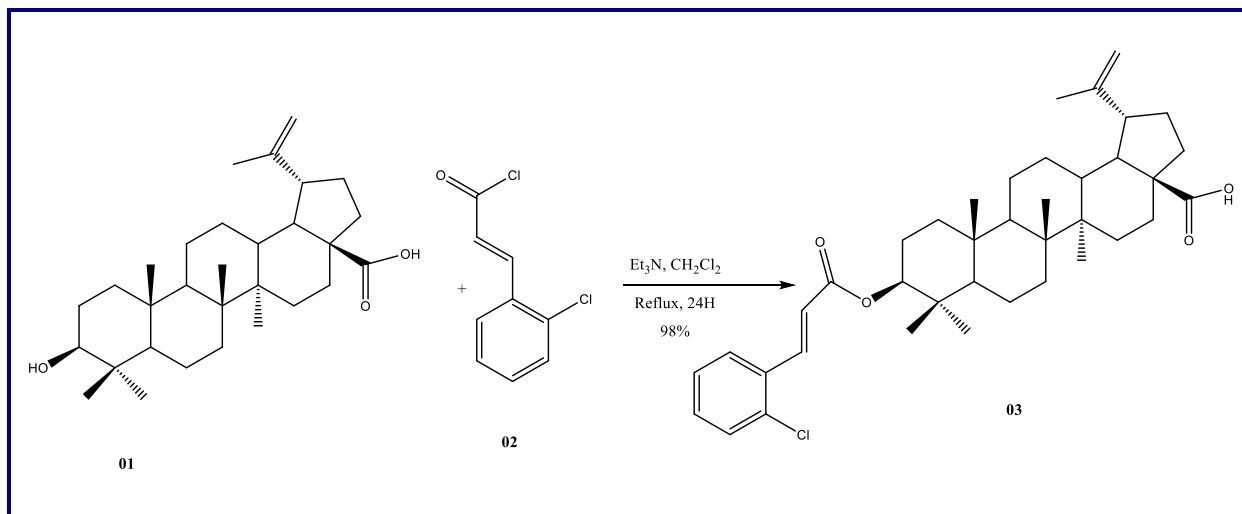


Schéma 1 : Synthèse du composé 03

#### ➤ Spectroscopie IR

Le composé **03** a été caractérisé par spectroscopie infrarouge (IR) afin de vérifier la formation de la fonction ester avant l'étape de cristallisation.

Sur le spectre IR (Figure 4), nous observons la présence d'un groupement hydroxyle identifiable par une large bande à  $3271\text{ cm}^{-1}$ , qui correspond au OH caractéristique de l'acide. Nous notons également une bande intense autour de  $1682\text{ cm}^{-1}$ , correspondant au CO de l'acide, ainsi qu'un pic à environ  $1621\text{ cm}^{-1}$ , correspondant au CO de l'ester. Ces observations confirment la formation du composé **03**.

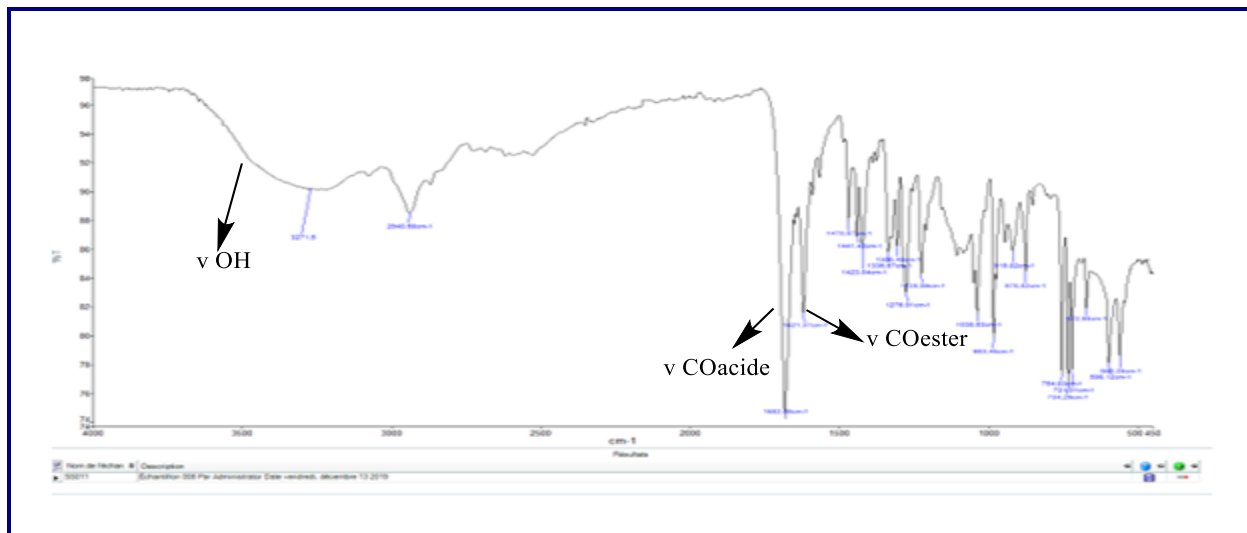


Figure 4 : Spectre IR du composé **03**

### E - II.1.b. Couplage de l'acide bétulinique avec du chlorure de phénoxyacétyle

La réaction d'estérification de l'acide bétulinique **01** avec le chlorure de phénoxyacétyle **04** en présence de la triéthylamine dans le dichlorométhane a donné le composé **05** avec un rendement très modeste de 30% (Schéma 2), après purification par chromatographie sur gel de silice. Le composé **05** a été caractérisé par spectroscopie infrarouge (IR).

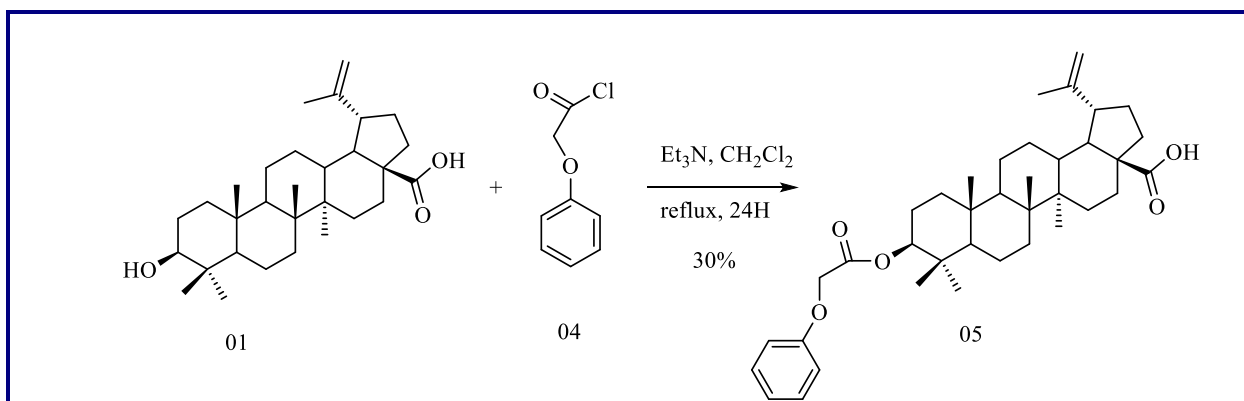


Schéma 2 : Synthèse du composé **05**

### ➤ Spectroscopie IR

Sur le spectre IR (Figure 5), nous observons la présence d'un groupement hydroxyle identifiable par une bande à  $3397\text{ cm}^{-1}$ , qui correspond au OH caractéristique de l'acide. Nous notons également une bande intense autour de  $1681\text{ cm}^{-1}$ , correspondant au CO de l'acide, ainsi qu'un pic à environ  $1727\text{ cm}^{-1}$ , correspondant au CO de l'ester. Ces observations confirment la formation du composé **05**.

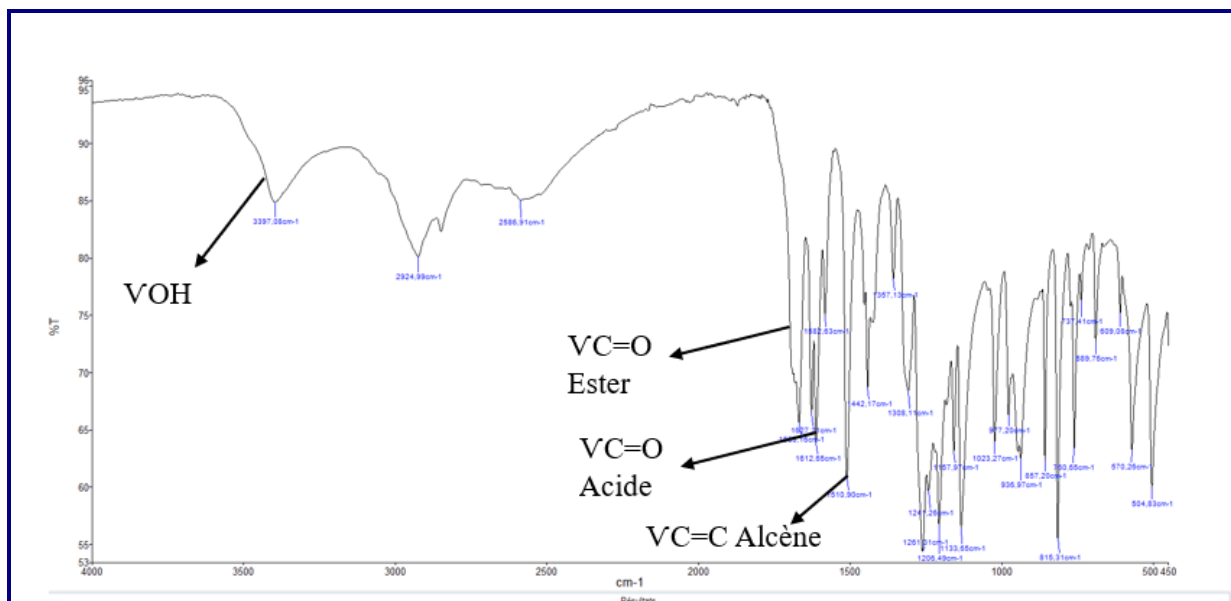


Figure 5 : Spectre IR du composé 05

Après la synthèse de ces dérivés, nous avons testé leur solubilité dans différents solvants (eau, diéthylamine, éthylène diamine, isopropanol, acétonitrile et méthanol) avant de procéder à la cristallisation. Nous avons constaté une solubilité, ce qui suggère que la synergie entre l'acide bétulinique et les molécules utilisées a probablement entraîné la solubilité des dérivés synthétisés [7] [8].

#### E - II.1.c. Cristallisation des dérivés de l'acide bétulinique

Afin d'obtenir des cristaux d'une taille appropriée pour la diffraction des rayons X, nous avons utilisé deux méthodes de cristallisation :

##### i) Evaporation lente

La molécule, sous forme de poudre (50 mg), a été répartie dans 4 flacons. Ensuite, du solvant a été ajouté à chacun de ces flacons (Tableau 3). Après la dissolution complète de la poudre, un film de paraffine a été délicatement placé sur chaque flacon, en y pratiquant une petite

perforation pour favoriser une évaporation lente à température ambiante. Quelques choix de solvants sont répertoriés dans le tableau suivant

**Tableau 3 : Solvants utilisés pour la cristallisation des dérivés de l'acide bétulinique**

Flacons + Molécule	Solvant
Flacon 1 + 05, 03	Eau
Flacon 2 +03	Diéthylamine
Flacon 3 +05	Ethylène diamine
Flacon 4 + 03	Méthanol

#### ii) Refroidissement lent

Afin de diversifier les méthodes de cristallisation et d'optimiser les chances d'obtenir des cristaux, la méthode de refroidissement lent a été utilisée.

Trois flacons ont été utilisés pour la cristallisation des dérivés :

- Dans le premier flacon, la molécule 05 (50 mg) a été mise en contact avec un volume minimal d'acétonitrile, puis chauffée à 30°C.
- Dans le deuxième flacon, la molécule 03 (50 mg) a été mise en contact avec de l'isopropanol, puis chauffée à 30°C.
- Dans le troisième flacon, la molécule 03, 05 (50 mg) a été mise en contact avec de l'eau, puis chauffée environ 20°C.

Ces trois flacons ont été placés dans une boîte en polystyrène, à l'intérieur de laquelle des flacons d'eau chaude ont été disposés. Ainsi, le refroidissement peut s'effectuer lentement sur une journée.

#### **E - II.1.d. Conclusion**

Ces différentes techniques utilisées n'ont pas donné de résultats satisfaisants en ce qui concerne la cristallisation des dérivés de l'acide bétulinique synthétisés, en raison notamment des petites quantités disponibles (quelques milligrammes).



**E - II.2 L'acide Laurique****E - II.2.a. Résumé de l'article**

Les acides gras supérieurs de formule  $C_nH_{2n-1}OOH$  sont composés de longues chaînes aliphatiques saturées avec un groupe carboxylique à une extrémité. Des cristaux de laurate de sodium et d'acide laurique (NaLLA) ont été obtenus et leur structure a été déterminée par diffraction des rayons X sur un monocristal. La nouvelle forme cristalline cristallise dans le système monoclinique avec le groupe d'espace  $P2_1/c$ . L'unité asymétrique contient deux molécules, une d'acide laurique et une de laurate, qui sont indépendantes, dont les groupes carboxyliques/carboxylates sont liés par une liaison hydrogène O-H...O à faible barrière (low barrier hydrogen bond). Les liaisons hydrogène à faible barrières d'énergie (LBHB) se produisent lorsque les pKa des deux hétéroatomes sont proches, permettant un partage plus équitable de l'hydrogène entre eux. Ce partage de l'hydrogène conduit à la formation de liaisons hydrogène particulièrement courtes et fortes.

Les deux molécules de laurate/laurique sont dans une configuration tête-à-tête, et les chaînes hydrophobes allongées sont parallèles le long de l'axe  $\vec{b}$ . L'atome d'hydrogène du groupe carboxylique s'est révélé désordonné, lié de manière non symétrique à chacun des deux groupes carboxylates. Le caractère non symétrique de la liaison hydrogène est attribuable à la présence de deux molécules d'acides gras indépendantes dans l'unité asymétrique, en accord avec les différentes longueurs des quatre liaisons C-O présentes dans la structure moléculaire. La structure cristalline a été analysée en termes d'interactions par la surface de Hirshfeld. L'empilement cristallin est stabilisé par des liaisons hydrogène et des interactions ioniques O...Na dans la couche hydrophile, ainsi que par des contacts C-H...H-C dans les couches hydrophobes, qui sont les contacts majeurs les plus enrichis.

**E - II.2.b. Publication N° 2 :**

Case of Charge-Assisted Hydrogen Bonding in the Crystal Structure of Sodium Laurate, Lauric Acid.



# Case of Charge-Assisted Hydrogen Bonding in the Crystal Structure of Sodium Laurate, Lauric Acid

Ibrahima Goudiaby<sup>1,2</sup> · Benoît Guillot<sup>1</sup> · Emmanuel Wenger<sup>1</sup> · Sarra Soudani<sup>3</sup> · Cherif ben Nasr<sup>3</sup> · Magatte Camara<sup>2</sup> · Abdoulaye Gassama<sup>2</sup> · Christian Jelsch<sup>1</sup>

Received: 31 January 2022 / Accepted: 22 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Crystals of Sodium Laurate, Lauric Acid (NaLLA) were obtained and the structure was determined by single-crystal X-ray diffraction. The new crystal form is monoclinic of space group  $P2_1/c$ . The asymmetric unit contains two independent laurate molecules whose carboxylic/carboxylate groups are linked by a low barrier O-H...O hydrogen bond. Two lauric/laurate molecules are in a head-to-head configuration and the elongated hydrophobic chains are parallel to the long  $b$  axis. The carboxylic hydrogen atom was found to be disordered, bound on each of the two carboxylate groups in an unsymmetrical way. The non-symmetrical character of the hydrogen bond is related to the presence of two independent fatty acid molecules in the asymmetric unit and is in accordance with the different lengths of the four C-O bonds present in the molecular structure. The crystal structure was analyzed in terms of interactions on the Hirshfeld surface. The packing is stabilized by hydrogen bonds and O...Na ionic interactions in the hydrophilic layer and by C-H...H-C contacts in the hydrophobic layers which are the most enriched major contacts.

**Keywords** Fatty acid · Crystal structure · Low barrier hydrogen bond

## Introduction

Higher fatty acids of the formula  $C_nH_{2n-1}OOH$  are composed of long, saturated aliphatic chains with a carboxylic group at one of their extremities. Lauric acid  $C_{12}H_{23}OOH$ , or dodecanoic acid, occurs naturally in some vegetable oils, notably coconut oil [1, 2] and in the milk of some mammals [3]. The influence of lauric acid on human health and diet, as well as its well-known antimicrobial activity, have made this acid the subject of a large number of studies [4, 5].

There are several studies on the properties and behaviour of lauric acid or laurate salts in solution, different phases [6, 7] and as liquid crystal.

[8]. There is also interest in the solid state of lauric acid. The normal chain monocarboxylic acids with even number of carbon atoms were known to exist in three main crystalline forms which have been called A, B and C by [9], and respectively  $\gamma$ ,  $\beta$  and  $\alpha$  by [10]. The different forms can be distinguished from each other by the value of the long spacing which, for a given acid, is greatest for the  $\gamma$ -form and smallest for the  $\alpha$ -form. Long chain aliphatic compounds adopt even more solid forms, with different configurations of the packing of the molecules in and between the layers [11].

Saturated carboxylic acids show polymorphism in the solid state. At least seven solid phases exist for the even members [12], called A2, Asuper, Bo, Bm, Eo, Em and C and five for the odd-numbered [13] called A', B', C', C'' and D'.

Generally, the crystal structures of these forms consist of molecules packed in bilayers containing two molecules. The carbon chains saturated are in the trans conformation (except for the B form and some triclinic forms, where the

✉ Christian Jelsch  
christian.jelsch@univ-lorraine.fr

<sup>1</sup> Université de Lorraine, CNRS, CRM2, F-54000 Nancy, France

<sup>2</sup> Groupe Matériaux Inorganiques : Chimie Douce et Cristallographie, LCPM, Assane SECK University of Ziguinchor, Ziguinchor, Senegal

<sup>3</sup> Laboratoire de Chimie des Matériaux, Faculté des Sciences de Bizerte, Université de Carthage, 7021 Zarzouna, Tunisia

molecules adopt a left-hand conformation), and with their carboxyl groups forming dimers through a typical  $R_2^2$  (8) hydrogen bonding system [14]. Two types of molecular conformations of lauric acid exist in the crystal structure. One is the ordinary type, as seen in the C-form or the  $A_1$  form belonging to the A form, while in the other one the carboxyl group is rotated about the C1-C2 bond [15].

The structures of several polymorphic crystal forms of lauric acid and laurate salts have been described. The first crystal structure reported was that of strontium laurate, in  $P2_1/n$  monoclinic space group [16]. In the same year, the crystal structures of several C2/n silver fatty acids, including laurate, were analyzed and turned out to be all triclinic of space group  $P\bar{1}$  [17].

The form C of lauric acid crystal structure was determined from monoclinic crystals of space group  $P2_1/a$  [18]. Crystals of the  $\gamma$  form of lauric acid were obtained by [19] and the dimensions of the monoclinic unit cell were determined. A new crystal form (not of form A, B, nor C) of lauric acid was obtained by [20] and is triclinic. Form  $A_1$  of lauric acid (triclinic  $P\bar{1}$ ) was later determined by [21].

Five crystal structures of lauric acid are available in the Cambridge Structural Database (CSD-21) [22], corresponding to four polymorphs. Lomer [20, 21] deposited two copies of the triclinic  $P\bar{1}$  structure. The form A crystal structure has been deposited [23] as well as the A-super form [15]. The latest crystal structure was deposited in 2009 and is monoclinic [24].

Only unit cells of crystals of laurate salts are available in the CCDC for silver, strontium and potassium (2 polymorphs) while the crystal structure of the salt with lead (II) [25] is available in the COD [26]. In the present paper, we report a new crystal structure of lauric acid containing two fatty acid molecules for one sodium cation in the asymmetric unit.

## Materials and methods

### Crystallization

The crystals were obtained in an attempt to co-crystallize lysine with sodium laurate. A stoichiometric quantity of lauric acid, lysine HCl and sodium hydroxide were dissolved in a minimal volume of ethyl acetate by stirring at 35 °C. All compounds were purchased from Sigma-Aldrich. The solution was at first left to cool to room temperature and then to 4 °C overnight.

### Crystal data collection

A single crystal was carefully selected under a polarizing microscope in order to perform its structural analysis. X-ray diffraction data were collected on a Rigaku SuperNova diffractometer [27] at 100 K, using graphite mono-chromated Cu  $K\alpha$  radiation. The crystal structure was solved using a direct method with the SHELXT program [28] and initially refined by successive differential Fourier syntheses and a full-matrix least-squares procedure using the SHELXL program [29]. Crystal data and experimental parameters used for the intensity data collection are summarized in Table 1.

### Crystal structure refinement

The crystal structure was then refined with software MoPro [30] using a multipolar atom model. The electron density was transferred from the ELMAM2 electron density database [31]. When using the multipolar atom model instead of the spherical one, the root mean square Fourier residual electron density diminished from 0.105 to 0.098  $e \text{ \AA}^{-3}$  and the  $wR^2(I)$  factor from 13.3 to 10.5%. Initially the sodium cation position was modelled as an oxygen water atom which refined to an occupancy of 1.3. Therefore, a sodium cation was assigned to that position. The crystal structure has been deposited at the CSD (Deposition Number 2,143,409).

**Table 1** Crystal data and structure parameters of the title compound

Chemical formula	$\text{CH}_3(\text{CH}_2)_{10}\text{COOH}$ , $\text{CH}_3(\text{CH}_2)_{10}\text{COO}^-$ , $\text{Na}^+$
MW (g.mol <sup>-1</sup> )	422.601
Crystal system, space group	Monoclinic, $P2_1/c$
Temperature (K)	100(1)
<i>a</i> , <i>b</i> , <i>c</i> (Å)	35.3860 (2), 7.0683 (6), 9.8952(6)
$\beta$ (°)	93.5200 (6)
<i>V</i> (Å <sup>3</sup> ); <i>Z</i>	2470.307 (6); 4
<b>Data collection</b>	
Radiation type	Cu $K\alpha$
Wavelength $\lambda$ (Å)	1.54184
Crystal size (mm)	0.041, 0.068, 0.178
Diffractometer	Bruker SMART X2S benchtop
Absorption correction	Multi-scan (SADABS; Bruker, 2016)
<i>T</i> <sub>min</sub> , <i>T</i> <sub>max</sub>	0.694, 1.000
No. of measured reflections	15 823
<i>R</i> <sub>int</sub> (%)	6.3
$\sin\theta_{\text{max}}/\lambda$ (Å <sup>-1</sup> )	0.63
<b>Refinement</b>	
$wR^2[F^2 > 2\sigma(F^2)]$ , $wR(F^2)$ , <i>gof</i>	0.089, 0.105, 1.03
$R[F > 4\sigma(F)]$ , <i>R</i> ( <i>F</i> )	0.052, 0.073
No. of reflections	4906
No. of parameters	455
$\Delta\rho_{\text{max}}$ , $\Delta\rho_{\text{min}}$ , rms ( $e \text{ \AA}^{-3}$ )	0.45, -0.44, 0.098

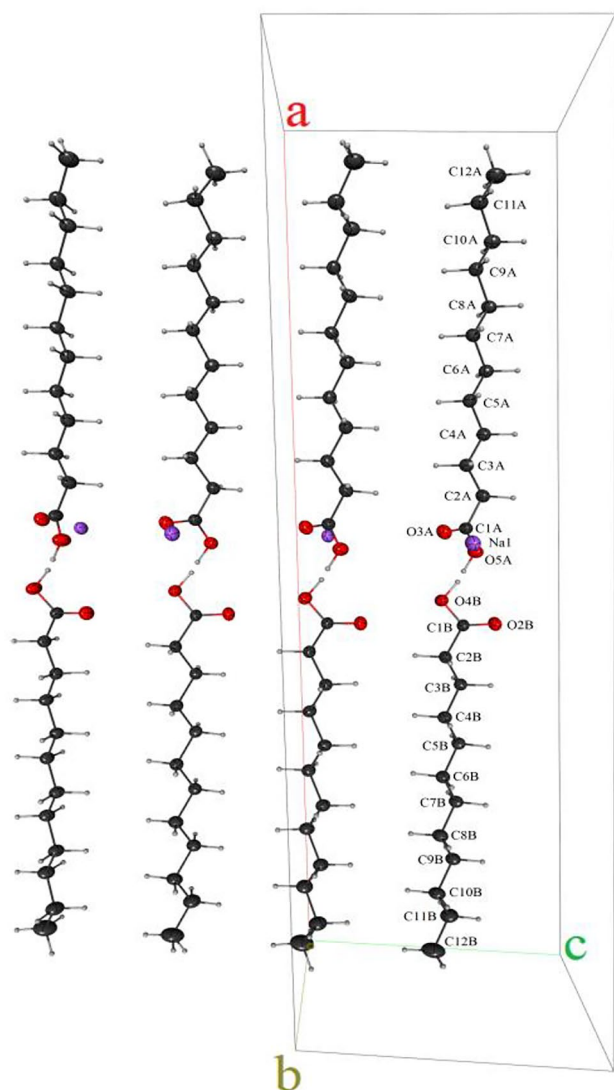
## Results and discussion

### The low barrier hydrogen bond

In the NaLLA crystal structure, a single hydrogen bond between O4 and O3A atoms, is linking together the two fatty acid molecules of the asymmetric unit and displays interesting peculiarities (Fig. 1).

At first, the short O4B...O5A distance  $d_{O...O} = 2.452(3)$  Å obviously classifies this interaction among the strong intermolecular hydrogen bonds [32]. The O...O distance is significantly shorter than the ones found in odd and even-numbered fatty acid crystal structures. The shortest reported

O...O distance is 2.54(4) Å, seen in  $C_{13}H_{25}OOH$  crystal structure [11], but measured by X-ray diffraction experiments at room temperature (versus  $T = 100$  K in the present study). Short ( $d_{O...O} < 2.65$  Å) or very short ( $d_{O...O} < 2.5$  Å) intermolecular O-H...O hydrogen bonds are often observed between pairs of carboxylic acids or between interacting carboxylic acids – carboxylate groups as seen in a search of the CSD version 2021. Also, O...O distances as short as 2.5 Å can be seen in intramolecular H-bond cases [33], where resonance-assisted hydrogen bonds (RAHB) are modeled against diffraction data as symmetrical, displaying a shared hydrogen atom which can be found located at equal distance from the two oxygen atoms [34].



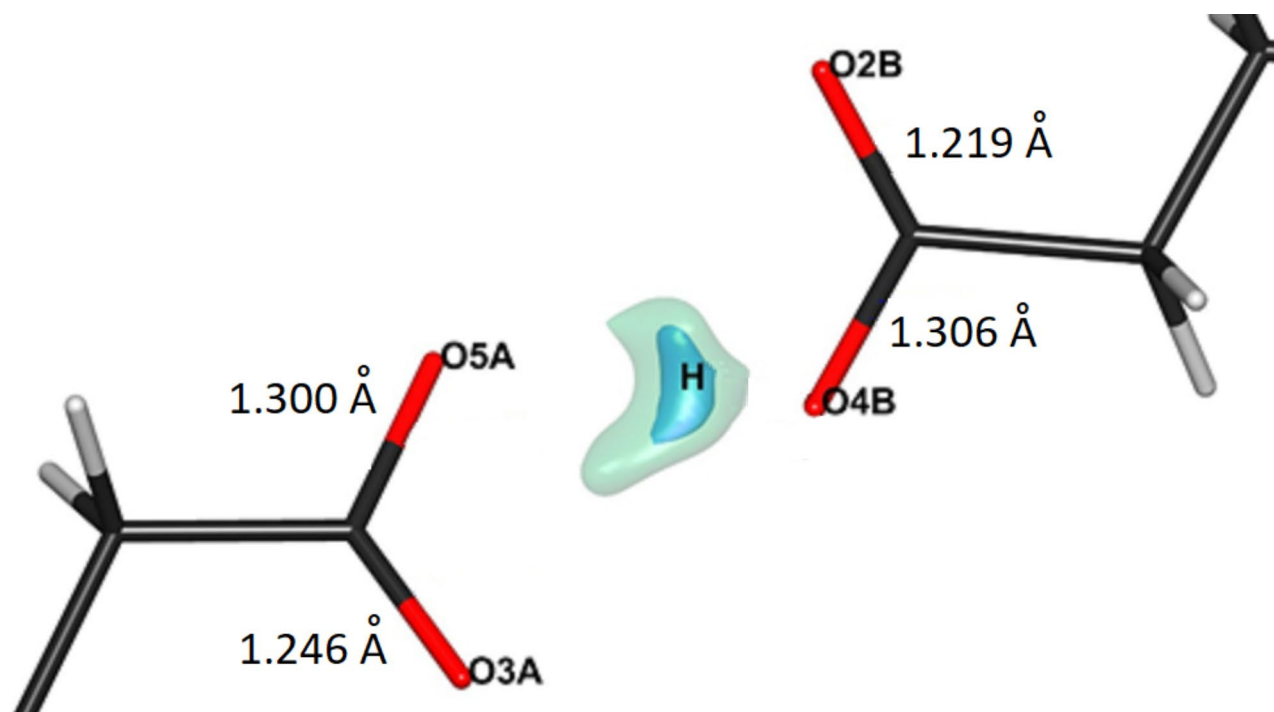
**Fig. 1** Crystallographic autostereogram of four symmetry homologues in the crystal structure of the title compound. Displacement ellipsoids are drawn at the 67% probability levels. The symmetry cards are from left to right:  $(x, -y + 1/2, z - 3/2)$ ,  $(x, y, z - 1)$ ,  $(x, -y + 1/2, z - 1/2)$ ,  $(x, y, z)$

The C-O bond lengths suggest a carboxylic acid character for both interacting moieties, with one short distance and one longer distance:  $d(C1B-O4B) = 1.306(3)$  Å,  $d(C1-O2B) = 1.219(3)$  Å on one side and  $d(C1A-O5A) = 1.300(4)$  Å,  $d(C1A-O3A) = 1.246(4)$  Å on the other molecule. Therefore, an H omit map was computed in the region around the O4B and O5A atoms which should bear a hydrogen atom (Fig. 2). The omit map shows an elongated density between the O4B and O5A atoms. To increase the quality of the signal, a Fourier residual map with application of a simplified noise reduction technique, inspired from Ursby & Bourgeois [35] was computed. The Fourier synthesis was performed on  $(F_{obs} - F_{calc}) * R$ , where  $R = |F_{obs} - F_{calc}| / [ |F_{obs} - F_{calc}| + \text{Sigma}(F_{obs}) * \text{gof} ]$

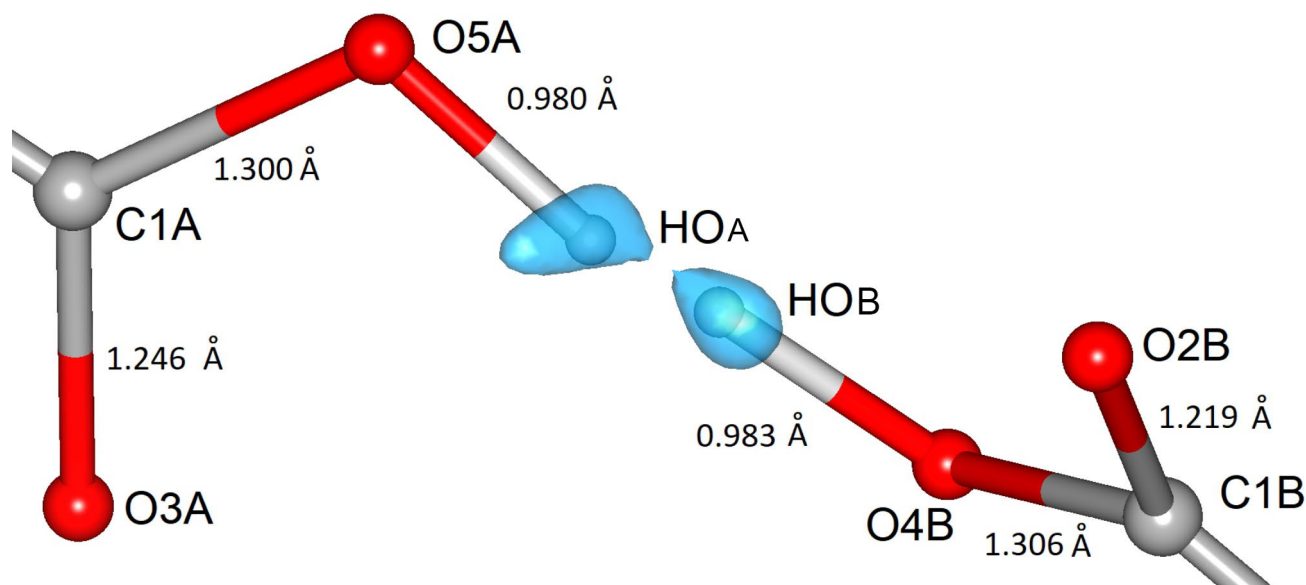
The  $R$  coefficient depends on the structure factor uncertainty  $\sigma(F_{obs})$  multiplied by the goodness of fit ( $\text{gof}$ ); it attenuates  $F_{obs} - F_{calc}$  differences which are small compared to the estimated error. The resulting omit map with noise reduction shows two residual peaks (Fig. 3). The main residual peak appears to be related to a hydrogen atom bound to the O4B atom which shows the longest C-O distance. The smallest peak corresponds to a hydrogen atom bound to O5A. The sum of occupancies of the two hydrogen atoms was constrained to unity and the occupancy values refined to  $q(\text{HOB}) = 0.66(8)$  and  $q(\text{HOA}) = 0.34(8)$ . The 0.006 Å difference between the two longer C-O bond lengths is however only around two sigmas of their uncertainty (ca. 0.003 Å).

This disordered hydrogen atom HOB/HOA suggests the presence of an asymmetric double-well low-barrier hydrogen bond (LBHB) between the two independent organic molecules in the crystal structure of Sodium Laurate / Lauric Acid.

Standard hydrogen bonds are longer (e.g. 2.8 Å for an  $O \cdots O$  H-bond), and the hydrogen atom is clearly bound



**Fig. 2** Fourier residual omit H map in the O4B...O5A region. Contour levels 0.4 (blue) and 0.5 (green)  $e.\text{\AA}^{-3}$



**Fig. 3** The omit H map in blue at contour level at  $0.27 e.\text{\AA}^{-3}$ . The Fourier residual map using terms  $(F_{\text{obs}} - F_{\text{calc}}) * R$  was computed by applying a noise reduction technique

to one of the oxygen atoms. When the pKa of the interacting moieties is closely matched, a LBHB becomes possible at short distances ( $\sim 2.55 \text{\AA}$ ). When the distance further decreases ( $< 2.29 \text{\AA}$ ) the bond is characterized as a single-well or short-strong hydrogen bond [36]. The present crystal structure, with an O...O distance of  $2.452(3) \text{\AA}$ , corresponds to an unsymmetrical double well, corroborated by

the interaction between two independent molecules in the crystal asymmetric unit.

The positive charge of the  $\text{Na}^+$  ion is compensated by a formal negative charge of  $-1e$  shared by the pair of carboxylate/carboxylic acid groups. The O4B-HOB...O3A and O4B...HOA-O3A interactions fall therefore within the

negative-charge assisted hydrogen bond ( $[\text{COO-H-COO}]^-$  CAHB) category.

The typical C-O<sup>-</sup> distance is 1.25(2) Å in carboxylate groups while in the carboxylic acid group, the bond lengths are  $d(\text{C}=\text{O})=1.21(1)$  Å and  $d(\text{C-O})=1.31(2)$  Å [37]. The C1B-O4B and C1A-O5A bonds, involving the oxygen atoms bound to the disordered hydrogen atom, have lengths very slightly smaller than the typical C-O distance in carboxylic acids. The two other C-O distances are intermediate between the C=O distance in carboxylic acids and the carboxylate C-O<sup>-</sup> distance. The C1B-O2B distance of 1.219(3) Å on the carboxylic group with two thirds protonation is close to C=O distance in carboxylic acids while the C1A-O3A distance 1.246(4) Å (on COOH group with one third protonation) is closer to the carboxylate C-O<sup>-</sup> bond length.

Disordered hydrogen atom within a hydrogen bond has been observed in the recent crystal structure of 4-[(morpholin-4-yl)-carbothioidyl]benzoic acid [38]. It reveals the case of a COOH...HOOC double hydrogen bond between two carboxylic acids related by an inversion centre, where the H-O hydrogen is shared between the oxygen atoms, the two C-O distances 1.266 and 1.268 Å being nearly similar.

In the crystal structure of potassium hydrogen acetylenedicarboxylate [39, 40] a hydrogen atom is found between two oxygen atom of two HOOC-C<sub>2</sub>H<sub>2</sub>-COO<sup>-</sup> moieties. This situation resembles that of the NaLLA crystal structure, except that it has single H atom which is located on the 2-fold symmetry axis relating the two oxygen atoms.

**Table 2** Selection of bond lengths (Å) and bond angles (°) for non-H atoms with uncertainty of values given in parentheses. Symmetry operators : (i)  $-x+1$  ;  $y+\frac{1}{2}$  ;  $-z+3/2$  (ii)  $-x+1$  ;  $-y+1$  ;  $-z+1$ .

atoms	distance	atoms	distance / angle
Na1 O4B (i)	2.322(2)	Na1 O2B (ii)	2.322(2)
Na1 O3A (i)	2.406(2)	Na1 O5A (ii)	2.386(2)
O3A C1A	1.246(4)	O2B C1B O4B	123.5(2)
O5A C1A	1.300(4)	O2B C1B C2B	121.9(2)
C2A C1A	1.514(4)	O4B C1B C2B	114.5(2)
O2B C1B	1.219(3)	O5A C1A O3A	122.9(2)
O4B C1B	1.306(3)	O5A C1A C2A	114.7(2)
C2B C1B	1.511(4)	O3A C1A C2A	122.3(2)

**Table 3** Geometric details of hydrogen bonds (Å, °) (D-donor; A-acceptor; H-hydrogen) of the title compound and symmetry code of acceptor atom

Interaction D-H...A	$d(\text{D}-\text{H})$	$d(\text{H}\cdots\text{A})$	$d(\text{D}\cdots\text{A})$	$\angle\text{D}-\text{H}\cdots\text{A}$	symmetry
O4B—HOB...O5A	1.02	1.48	2.452 (3)	169	$x, y, z$
O5A—HOA...O4B	1.02	1.48	2.452 (3)	169	$x, y, z$
C2B—H2D...O2B <sup>i</sup>	1.09	2.62	3.489 (3)	136	$x, -y+\frac{1}{2}, z-\frac{1}{2}$
C3A—H3B...O2B <sup>ii</sup>	1.09	2.50	3.394 (3)	138	$-x+1, y+\frac{1}{2}, -z+3/2$

## Crystal structure description

X-ray analysis of the crystal structure reveals that the title compound crystallizes in the monoclinic space group  $P2_1/c$  with  $Z=4$  (Table 1). Its molecular structure is shown in Fig. 1 and selected interatomic distances and angles are listed in Table 2. Table 3 lists the hydrogen bonds observed in the crystal structure. The asymmetric unit of this complex, shown in Fig. 1 consists of a sodium Na<sup>+</sup> cation and two interacting fatty acid molecules sharing a single negative formal charge.

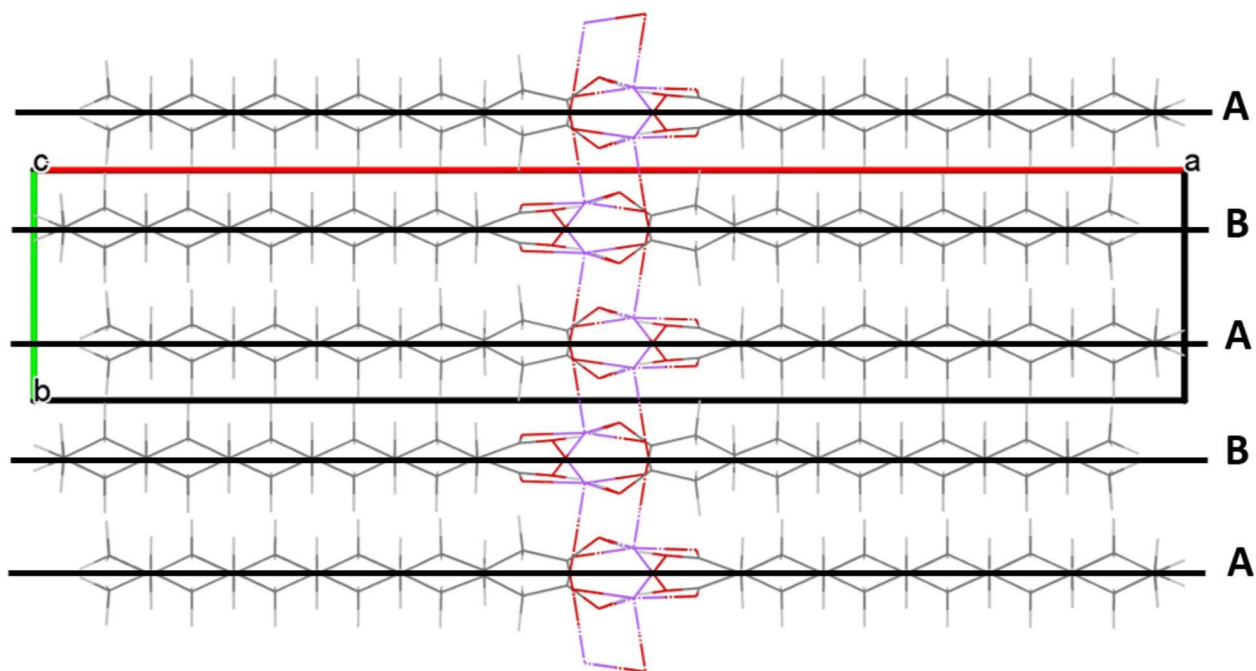
The packing of the NaLLA crystal structure displays van der Waals dispersive (hydrophobic) contacts between hydrogen atoms of almost perfectly parallel *n*-alkyl chains, as it is systematically observed in saturated fatty acids crystal structures. The asymmetric low barrier hydrogen bond involves the carboxylic heads of the two laurate / lauric acid molecules. The Na<sup>+</sup>...Oxygen electrostatic interactions and two weak C-H...O hydrogen bonds complete the interaction network.

Further comparisons of the packing observed in the present crystal structure and the ones typically seen in saturated fatty acids crystal structure reveal however striking differences.

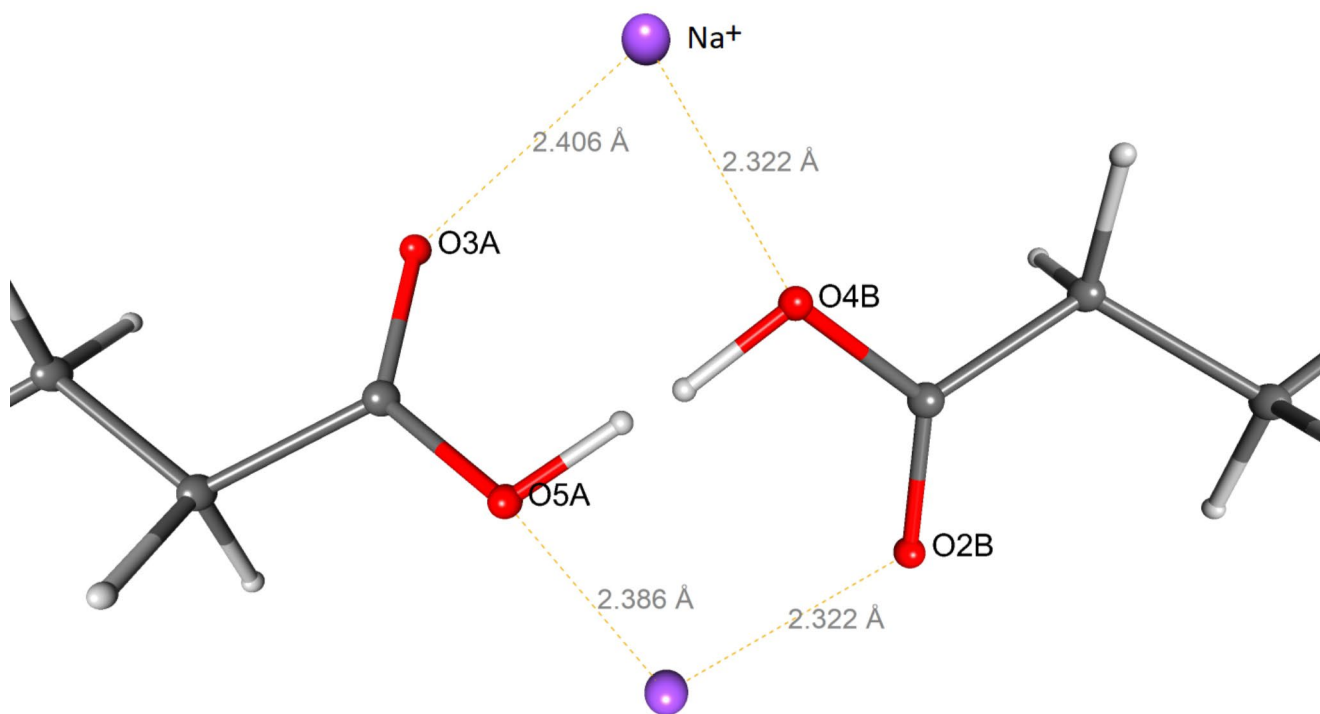
At first, in C<sub>n</sub>OOH, *n* odd- and even-numbered fatty acids crystal structures [11, 24], molecules are related by inversion centers found among symmetry elements of their space groups (very often monoclinic  $P2_1/a$  with  $Z=4$ ,  $A2/a$  with  $Z=8$  or  $P\bar{1}$  with  $Z=2$ ), with one molecule per asymmetric unit. Therefore, the conformations of interacting carboxylic groups in these centrosymmetric dimers are similar: they are either *cis* or *trans* with C-C-C=O torsion angle very close to 0° or 180°, respectively. Even-numbered saturated fatty acids (as is C12 lauric acid) noteworthy present systematic *cis* conformations with C-C-C=O torsion angles not exceeding 4°. In the present NaLLA crystal structure, this conformation is significantly distorted, with a *cis* conformation and C-C-C=O dihedral angles of 12.1° and 31.8° for the two molecules in the asymmetric unit. Consequently here, none of the two carboxyl groups are coplanar with the hydrocarbon chain of the molecule they belong to.

Furthermore, an interaction between two carboxylic groups where the two COOH groups related by inversion symmetry are facing each other in a  $R_2^2(8)$  pattern, with two identical C-O-H...O=C hydrogen bonds is often seen in





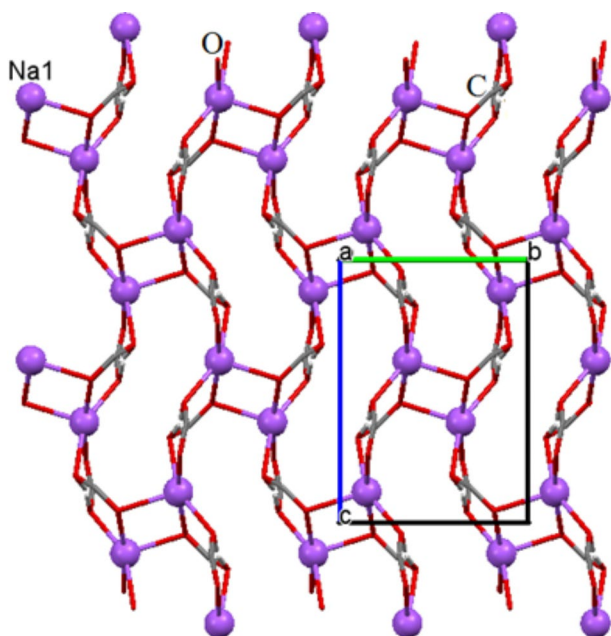
**Fig. 4** View along the *c* axis of the bidimensional supramolecular structure of the NaLLA crystal structure. Axes A and B show the parallelism of double chains



**Fig. 5** View of the two closest sodium cations interacting with the oxygen atoms of the title compound

fatty acids crystal structures. Such a COOH...COOH homomeron [41] is not observed here. Instead, the carboxylic groups are slightly rotated, forming a O-H...O hydrogen bond. These strong O-H...O hydrogen bonds, connecting opposite double chains, form layers spreading along [100] direction (Fig. 4; Table 3). Interestingly, the carboxylic oxygen atoms not involved in O-H...O hydrogen bonds, despite their strong hydrogen bond acceptor nature, only form weak C-H...O interactions with neighboring molecules (Table 3). The main interaction partners in the crystal packing of the non-protonated carboxylic oxygen atoms are the Na<sup>+</sup> cations. Indeed, there are four O...Na<sup>+</sup> interactions ( $2.320 \text{ \AA} < d_{O...Na^+} < 2.406 \text{ \AA}$ ) between the four oxygen atoms of the asymmetric unit and two Na<sup>+</sup> cations located on both sides of the interacting carboxyl groups (Fig. 6).

The distortion of the usually observed *cis* conformations of the C-C-C=O groups might be attributable to these Na<sup>+</sup> cations present in the crystal packing. The four Na<sup>+</sup> ions in the unit cell are arranged around a plane perpendicular to the [100] direction roughly located at coordinate  $x \sim 0.5$ . The Na<sup>+</sup> cations are surrounded by five negatively charged oxygen atoms, three of C-O functions O3A, O3A<sup>i</sup> ( $i = 1-x, 1-y, 1-z$ ) and O2B<sup>iii</sup> ( $iii = x, 1/2-y, -1/2+z$ ) and two corresponding to C-O-H groups, O5A<sup>iii</sup> ( $iii = x, 1/2-y, -1/2+z$ ) and O4B. The Na...O distances, ranging from 2.322(3) to 2.406(3) Å (Table 2; Fig. 5), are consistent with related penta-coordinated Na complexes [42, 43]. The Na1...Na1<sup>i</sup> and Na1...Na1<sup>ii</sup> separations are 3.341(3) and 5.183(3) Å respectively.



**Fig. 6** Projection of a 2-D layers of sodium cations in the crystal structure along the *a*-axis. The carbon chains of lauric acid and laurate molecules have been omitted for clarity

Also, it is worth noting that these distances are significantly shorter than the sum of the van der Waals radius of the oxygen atom ( $R_{vdw} \sim 1.52 \text{ \AA}$ ) and of the ionic radius of the Na<sup>+</sup> ion ( $R_{Na^+} \sim 1.16 \text{ \AA}$ ). Hence these strong Na<sup>+</sup>...O electrostatic attractions may induce the distorted conformations of the fatty acid heads.

Figure 4 shows clearly the respective disposition of the aliphatic hydrocarbons of lauric acid/laurate anion and of the inorganic parts in the title compound. The polymeric chains of sodium cations, which are connected by oxygen bridges around the  $x = n + 1/2$  planes (Fig. 5), form 2-D layers parallel to the *bc* plane.

The adjacent lauric acid/laurate molecules form face-to-face arrangements in the crystal packing (Fig. 1). These chains, which are almost parallel along the *a*-axis direction, form zig-zag planes of adjacent molecules which are perpendicular to one another (Fig. 7) [24].

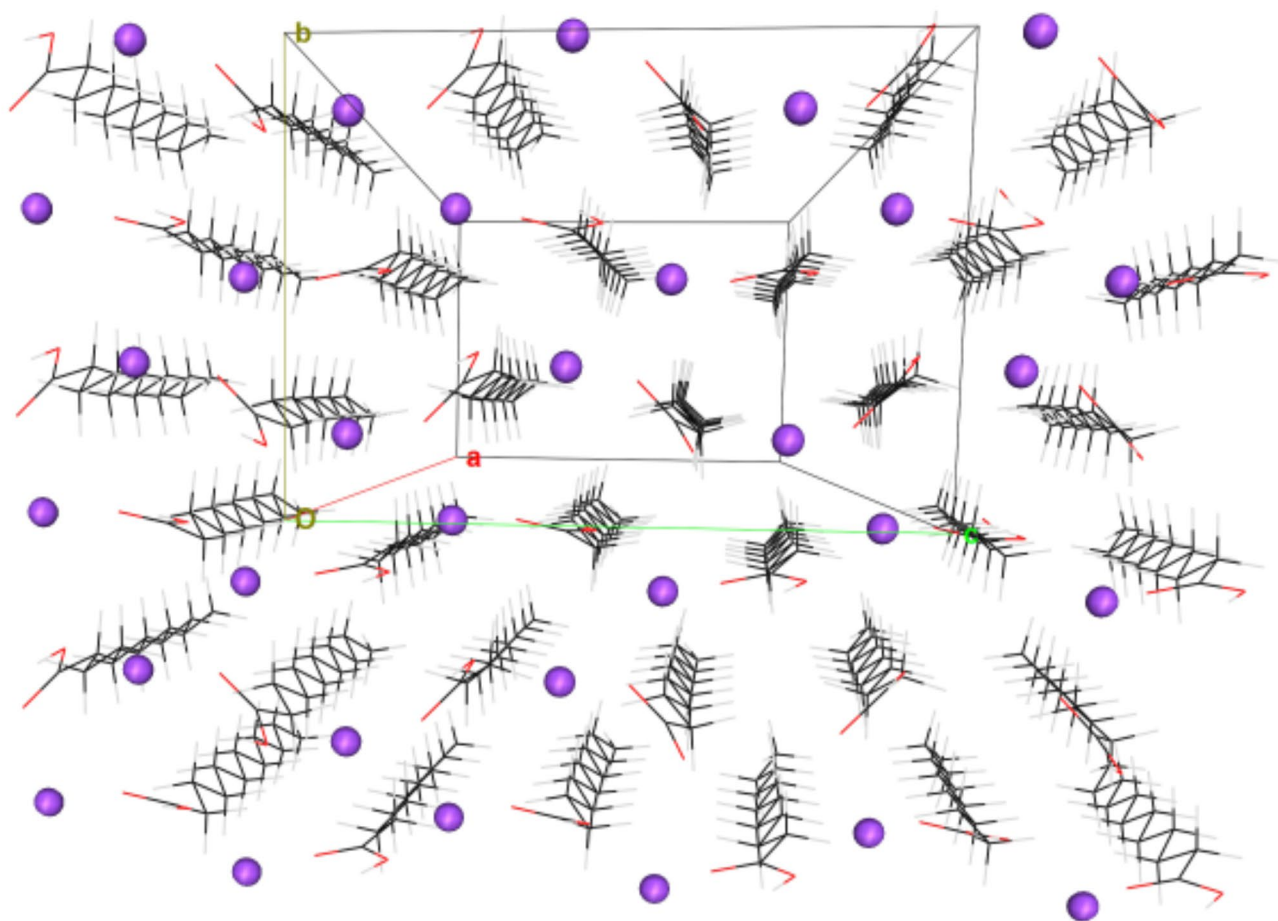
### Hirshfeld surface analysis

The Hirshfeld surface is representative of the region in space where molecules come into contact. In the Hirshfeld surface analysis, the disordered H atom was placed on the most occupied position. The Hirshfeld representation of the Na<sup>+</sup> cation, the neutral organic molecule (C<sub>12</sub>H<sub>24</sub>O<sub>2</sub>), and the organic anion (C<sub>12</sub>H<sub>23</sub>O<sub>2</sub>)<sup>-</sup> of the asymmetric unit highlights intermolecular interactions between the alkyl chains in the crystal structure (Fig. 8).

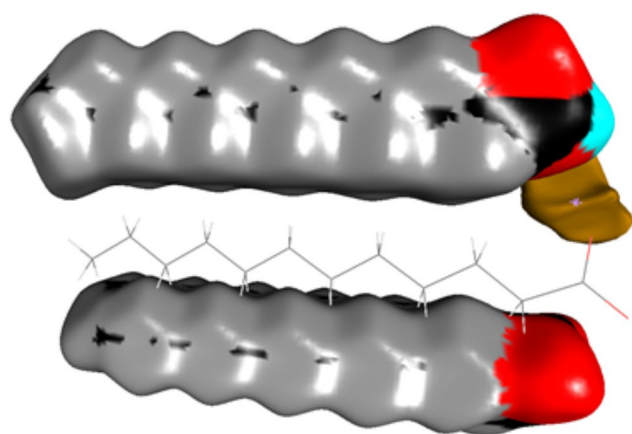
The 2D fingerprint plots of the Hirshfeld surface allows to highlight the atoms participating in contacts (Fig. 9). The graph shown in Fig. 9a. represents the H...O/O...H contacts. It is characterized by two symmetrical spikes at short distance  $d_i + d_e \sim 1.6 \text{ \AA}$ . These spikes are characteristic of the strong O-H...O hydrogen bond present in the crystal structure, but there are also C-H...O contacts at longer distance. The graph in Fig. 9e representing H...H contacts is characterized by an end pointing towards the origin along the diagonal corresponding to  $d_i = d_j \sim 1.2 \text{ \AA}$ . These are C-H...H-C contacts representing 70.8% of intermolecular contacts (Table 4), due to the parallel arrangement of the fatty acids alkyl chains, and the contact between terminal methyl groups of adjacent molecules along the [100] direction. The 2D fingerprint decomposition also shows other contacts: O...Na (10.8%) which is the second major contact and C...H-C which is in third position at 8.3%.

The contacts enrichment [46], derived from the Hirshfeld surface analysis, is computed from the ratio between the actual contact surface  $C_{xy}$  in the crystal structure and the  $R_{xy}$  (random) contact computed as if all types of contacts had the same probability to form. The equiprobable contact surfaces  $R_{xy}$  are computed from the product of probabilities  $S_x$  and  $S_y$  which correspond to the proportions of chemical species





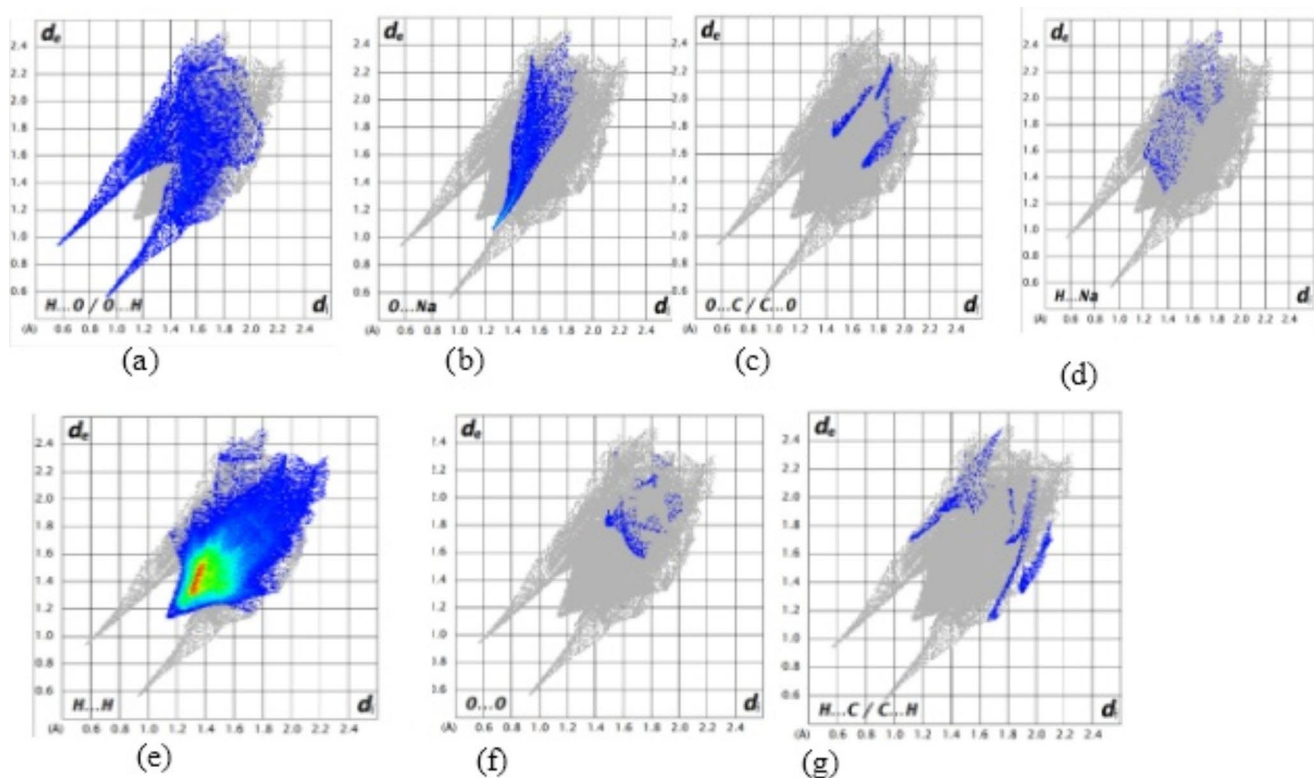
**Fig. 7** Crystallographic autostereogram of a layer of NaLLA molecules parallel to the  $(bc)$  plane carbon: black, hydrogen: grey, oxygen: red and sodium: purple. The view was made with MoProViewer [44]



**Fig. 8** Hirshfeld Surface around the lauric acid, the laurate and  $\text{Na}^+$  moieties of the asymmetric unit. The surface is colored according to the inner chemical type: hydrogen H-C: grey, hydrogen H-O: blue, carbon: black, oxygen: red)

X and Y on the Hirshfeld surface. The contact enrichment ratio ( $E_{xy}$ ) is a powerful tool to deduce which types of contacts are favoured or disfavoured in the crystal packing. An enrichment ratio  $E_{xy}$  larger than unity for a given pair of chemical species X...Y indicates that these contacts are over-represented. The contact types  $C_{xy}$  and their enrichment were computed with the MoProViewer program [45] and are shown in Table 4. For instance, the H-C hydrogen occupies 77.7% of the surface, yielding  $R_{\text{H-C,H-C}} = S_{\text{H-C}}^2 = 60.4\%$  and an enrichment  $E_{\text{H-C,H-C}} = C_{\text{H-C,H-C}} / R_{\text{H-C,H-C}} = 70.8 / 60.4 = 1.19$ .

More than three quarters of the surface area is generated by H atoms. The O...Na electrostatic interactions and the O...H-O hydrogen bonds are the two strongly enriched contacts ( $E_{\text{O,H-O}} > 8.3$ ) and constitute the two strong electrostatic interactions stabilizing the crystal packing. The large C-H...H-C contact surface between hydrophobic parts of lauric molecules is slightly over-represented at  $E_{\text{H-C,H-C}} = 1.19$ . Globally the hydrophilic and hydrophobic contacts are



**Fig. 9** Fingerprint plots in  $(d, d_e)$  of the main contacts on the Hirshfeld surface. The graphs were obtain with CrystalExplorer17 program [45]

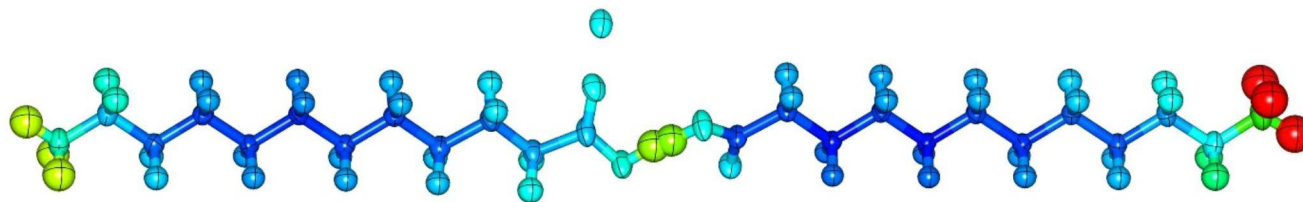
enriched by 5.26 and 1.13 coefficients, respectively. Conversely the cross contacts are strongly under-represented at  $E=0.28$ , in accordance with the presence of hydrophobic and hydrophilic layers in the crystal packing.

### Atomic thermal motion

Using transferred multipolar pseudoatoms in a crystal structure refinement at subatomic resolution allows better estimation of atomic anisotropic thermal displacement parameters [31]. In the present NaLLA crystal structure, equivalent atomic  $B$  factors give insight to the dynamic properties of lauric acid bilayers. The two molecules in the asymmetric unit display similar dynamic behavior, with average equivalent  $B$  factors for non-H atoms of  $1.6(4) \text{ \AA}^2$  and  $1.7(3) \text{ \AA}^2$ . Atomic thermal displacements within each molecule vary in similar ways along the fatty acid chain (Fig. 10). Highest  $B$  factors are found at the extremity of the fatty acids, for atoms of the terminal methyl groups:  $B(\text{C12A})=2.20 \text{ \AA}^2$  and  $B(\text{C12B})=2.86 \text{ \AA}^2$ . The lowest atomic thermal factors, ranging from  $1.36$  to  $1.52 \text{ \AA}^2$ , are found for C4 to C9 atoms located in the middle of the fatty acid molecules. These groups are involved in tight packing of parallel alkyl chains, where numerous H...H contacts can be found.

**Table 4** Nature of intermolecular contacts on the Hirshfeld surface by chemical types. The first row and first column indicate the chemical species involved in the contacts. The second row shows the atom surfaces related to the atom types that are indicated in the first row. The percentage of actual contact types  $C_{xy}$  between chemical species is then given followed by the enrichment ratios  $E_{xy}$ . The reciprocal contacts X...Y and Y...X have been merged. The major contacts as well as the most enriched ones are highlighted in bold. The hydrophobic H-C atoms bound to carbon were distinguished from the polar Ho atom of the COOH moiety. The computation was done as if the polar hydrogen atom H-O was bound to atom O4B. The computation was done on an ensemble of lauric acid, laurate and  $\text{Na}^+$  moieties which not in contact in order to obtain an integral surface around each moiety, as in Fig. 7. At the end of the table, the atoms were grouped in hydrophobic (H-C and C) and hydrophilic (H-O and O) atoms

atom	Na	H-O	O	H-C	C
surface $S_x$ %	5.5	1.3	9.1	77.7	6.4
Na	0.5		$C_{xy}$ Actual	Contacts	(%)
H-O	0.5	0.0			
O	<b>10.8</b>	1.5	0.0		
H-C	1.4	0.0	3.4	<b>70.8</b>	
C	1.5	0.2	1.0	<b>8.3</b>	0.2
Na	0.87		Enrichment	Ratio $E_{xy}$	
H-O	3.11	0			
O	<b>8.37</b>	<b>8.32</b>	0		
H-C	0.12	0	0.26	<b>1.19</b>	
C	1.68	1.28	1.09	0.94	0.74
%surface	Hydrophilic	Hydrophobic		Hydrophilic/Hydrophobic	
%contacts	15.9	84.1			
Enrichment	2.5	70.7			
	5.26	1.12			
					26.7
					0.28



**Fig. 10** NaLLA crystal structure asymmetric unit, with anisotropic thermal displacement parameters represented by ellipsoids at 75% probability. Atoms are colored according to their equivalent isotropic  $B$  factors: from dark blue for  $B < 1.5 \text{ \AA}^2$  to red for  $B > 4 \text{ \AA}^2$

## Conclusion

The asymmetric unit of the new crystal contains a sodium cation and two lauric/laurate molecules in a head-to-head configuration; the elongated hydrophobic chains are parallel to the long  $b$  axis. The cell neutrality is obtained by a single negative formal charge shared among the two interacting carboxylate/carboxylic heads of the fatty acids through an asymmetric low barrier hydrogen bond. A low-barrier hydrogen bond (LBHB) is a special type of hydrogen bond. LBHBs can occur when the  $pK_a$  of the two heteroatoms are close, which allows the hydrogen to be more equally shared between them. This hydrogen-sharing causes the formation of especially short, strong hydrogen bonds [47]. The present crystal structure of NaLLA contains formally a carboxylate group and a carboxylic acid. Therefore, the two interacting moieties ( $\text{COO}^- \dots \text{HOOC}$ ) have identical  $pK_a$ s and are prone to form a LBHB, which is the case in the present crystal structure. The LBHB is found unsymmetrical with occupancy factors equal ca. to one and two thirds, in accordance with the different C–O bonds lengths found within the  $\text{COO}^- \dots \text{HOOC}$  moiety and with the fact that the interaction occurs between two independent molecules of the asymmetric unit.

**Acknowledgements** The PMD2 X-ray diffraction facility (<http://crm2.univ-lorraine.fr/lab/fr/services/pmd2x>) of the Institut Jean Barriol, Université de Lorraine is acknowledged for the X-ray diffraction measurements, data processing and analysis, and for providing reports for publication. I.G. thanks the Government of Senegal for scholarships and the support of Nancy ROTARY Club.

**Data Availability** The CIF file and hkl reflections dataset generated and analysed during the current study are available at the Cambridge Structural Database repository (Deposition Number 2,143,409), [<https://www.ccdc.cam.ac.uk/>]

## Declarations

**Conflicts of interest/Competing interests:** None.

## References

- Metz J, Lassnre M (1996) *Ann N Y Acad Sci* 792:82–90
- Afig MA, Rahman RA, Man YC, Al-Kahtani HA, Mansor TST (2013) *Int. Food Res. J.*, 20:2035
- German BJ (2008) *Sci Aliments* 28:176–186
- Kavitha G, Vijayarohini P, Swamidoss CMA (2020) *Mater Today-Proc.*, 33:2782–2791
- Dayrit Fabian M (2015) *J Am Oil Chem Soc* 92:1–15
- Li YL, Wang S, Zhang X, Chen YM, Ning JN, Liu GF, Zhang GQ (2011) *Mater Sci Forum* 675:227–230
- Godquin-Giroud AM, Marchon JC, Guillon D, Skoulios A (1984) *J Phys Lett* 45:681–684
- Matveeva AG, Yurtov EV, Prokopova LA (2012) *Theor Found Chem Eng* 46:395–400
- Piper SH, Malkin T, Austin HE (1926) *J Chem Soc* 129:2310–2318
- Thiband J, Dupré LTF (1930) *C R Acad Sci Paris* 191:200
- Gbabode G, Negrier P, Mondieig D, Moreno Calvo E, Calvet T, Cuevas-Diarte M, À (2007) *Chem Eur J* 11:3150–3159
- Moreno E, Cordobilla R, Calvet T, Cuevas-Diarte MA, Gbabode G, Negrier P, Oonk HA (2007) *New J Chem* 31:947–957
- Gbabode G, Negrier P, Mondieig D, Moreno E, Calvet T, Cuevas-Diarte M (2009) *J Alloys Compd* 469:539–551
- Bernstein J, Davis RE, Shimoni L, Chang NL (1995) *Angew Chem Int Ed* 34:1555–1573
- Goto M, Asada E (1978) *Bull Chem Soc Jap* 51:70–74
- Morley WM, Vand V (1949) *Nature* 163:285–285
- Vand V, Aitken A, Campbell RK (1949) *Acta Crystallogr* 2:398–403
- Vand V, Morley WM, Lomer TR (1951) *Acta Crystallogr* 4:324–329
- Lomer TR (1955) *Nature* 176:653–654
- Lomer TR, Spanswick RM (1961) *Acta Crystallogr* 14:312–313
- Lomer TR (1963) *Acta Crystallogr* 16:984–988
- Groom CR, Bruno IJ, Lightfoot MP, Ward SC (2016) *Acta Crystallogr B: Struct Sci Cryst Eng* 72:171–179
- Von Sydow E (1956) *Arkiv Kemi* 9:231
- Moreno-Calvo E, Gbabode G, Cordobilla R, Calvet T, Cuevas-Diarte MA, Negrier P, Mondieig D (2009) *Chem Eur J* 15:13141–13149
- Martínez-Casado FJ, Ramos-Riesco M, Rodríguez-Cheda JA, Redondo-Yélamos MI, Garrido L, Fernández-Martínez A, Poulain A (2017) *Phys Chem Chem Phys* 19:17009–17018
- Gražulis S, Chateigner D, Downs RT, Yokochi AF, T, Quirós M, Lutterotti L, Le Bail A (2009) *J Appl Crystallogr* 42:726–729

27. Rigaku Oxford Diffraction. (2019) CrysAlis PRO Software System, version 1.171.40.67a; Rigaku Corporation: Oxford, UK
28. Sheldrick GM (2015) *Acta Crystallogr A* 71:3–8
29. Sheldrick GM (2015) *Acta Crystallogr C Struct chem* 71:3–8
30. Jelsch C, Guillot B, Lagoutte A, Lecomte C, (2005) *J Appl Crystallogr* 38:38–54
31. Domagała S, Fournier B, Liebschner D, Guillot B, Jelsch C (2012) *Acta Crystallogr* 68:337–351
32. Gilli P, Bertolasi V, Ferretti V (1993) *J Am Chem Soc* 116:909–915
33. Grabowski SJ (2021) *Crystals* 11:5
34. Madsen GKH, Iversen BB, Larsen FK, Kapon M, Reisner GM, Herbststein FH (1998) *J Am Chem Soc* 120:10040–10045
35. Ursby T, Bourgeois D (1997) *Acta Crystallogr A* 53:564–575
36. Schiøtt B, Iversen BB, Madsen GKH, Larsen FK, Bruice TC (1998) *Proc. Natl. Acad. Sci.* 95:12799–12802.
37. D'ascenzo L, Auffinger P (2015) *Acta Crystallogr B* 71:164–175
38. Agnimonhan FH, Bendeif EE, Akanni LA, Gbaguidi AF, Martin E, Wenger E, Lecomte C (2020) *Acta Crystallogr E: Crystallogr Commun* 76:581–584
39. Leban I, Golič L, Speakman JC (1973) *J Chem Soc Perkin Trans* 2:703–705
40. Kojić-Prodić B, Molčanov K (2008) *Acta Chim Slov* 55:692–708
41. Desiraju GR (1989) *J Am Chem Soc* 111:8725–8726
42. Zhang B, Xie CZ, Wang XQ, Shen GQ, Shen DZ (2004) *Acta Crystallogr Sect E: Struct Rep Online* 60:m1293–m1295
43. Ying-Qun Y, Shao-Hua Z, Si-Ping T, Li-Xia F (2013) *Chin J Struct Chem* 32:63–66
44. Guillot B, Enrique E, Huder L, Jelsch C (2014) MS19. 001. *Acta Crystallogr* 70:C279
45. Wolff SK, Grimwood DJ, McKinnon JJ, Turner MJ, Jayatilaka D, Spackman MA (2012) *Crystal Explorer 17.5.*, Université de Western Australia
46. Jelsch C, Ejsmont K, Huder L (2014) *IUCr J* 1:119–128
47. Gilli G, Gilli P (2000) *J Mol Struct* 552:1–15

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



---

---

*Références partie E*

---

---





- [1] M. Nespolo, G. Ferraris, S. Đurovič, et Y. Takéuchi, « Twins vs. modular crystal structures », *Z. für Krist.*, 219, 773-778, **2004**.
- [2] M. Nespolo et G. Ferraris, « Applied geminography – symmetry analysis of twinned crystals and definition of twinning by reticular polyholohedry », *Acta Crystallogr. A: Found.*, 60, 89-95, **2004**.
- [3] J. E. Padilla et T. O. Yeates, « A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning », *Acta Crystallogr. D Biol. Crystallogr.*, 59, 1124-1130, **2003**.
- [4] F. Jbilou, P. Dole, P. Degraeve, C. Ladavière, et C. Joly, « A green method for polybutylene succinate recycling: Depolymerization catalyzed by lipase B from *Candida antarctica* during reactive extrusion », *Eur. Polym. J.*, 68, 207-215, **2015**.
- [5] J. Uppenberg N. Oehrner, M. Norin, K. Hult, G. J. Kleywegt *et al.*, « Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols », *Biochem.*, 34, 16838-16851, **1995**.
- [6] L. Dettori, C. Jelsch, Y. Guiavarc-h, S. Delaunay *et al.*, « Molecular rules for selectivity in lipase-catalysed acylation of lysine », *Process Biochem.*, 74, 50-60, **2018**.
- [7] M. J. Jacques, F. Annie, J. A. Christian, « Les composés phénoliques des végétaux : un exemple de métabolites secondaires d'importance économique », 15<sup>e</sup>-ed., **2005**.
- [8] J. W. Wood et T. D. Fontaine, « Synthetic plant-growth regulators. 111. 2,4-dichloro-phenoxyacetyl derivatives of amino acids », *J. Org.Chem.*, 17, 891-896, **1952**.



---

*Conclusion Générale et Perspectives  
de recherche*

---



Cette thèse, qui présente un travail pluridisciplinaire, avait initialement pour objectif principal de réaliser des réactions de couplage pseudo-peptidiques entre l'acide bétulinique (BTA) et le 2-chlorocinnamoyl ou le chlorure de phénoxyacétyle, suivies de la cristallisation de ces dérivés. Ce couplage avait pour objectif d'augmenter la solubilité du BTA qui est un composé ayant un potentiel thérapeutique.

La spectroscopie infrarouge (IR) a été utilisée pour confirmer la présence de la fonction ester dans les dérivés obtenus avant de procéder à l'étape de cristallisation. Malgré plusieurs tentatives de cristallisation, il n'a pas été possible d'obtenir des cristaux de ces dérivés.

Les succès obtenus dans cette synthèse ouvrent la voie à de nouvelles opportunités de recherche et d'application, que ce soit dans le domaine de la santé, de l'environnement ou de la science fondamentale. Il serait cependant intéressant d'utiliser d'autres méthodes de purification comme la chromatographie à haute performance afin de purifier davantage les dérivés, pour espérer les cristalliser et les caractériser.

En revanche, les cristaux de laurate de sodium, acide laurique (NaLLA), ont été obtenus lors d'une tentative de cocrystallisation de la lysine avec le laurate de sodium dans le but de former un cocrystal lysine-laurate de sodium. Le cocrystal aurait ensuite été broyé afin d'éventuellement provoquer une réaction d'estérification. L'analyse des cristaux par la diffraction des rayons X n'a pas révélé la présence de lysine mais d'un mélange laurate de sodium/acide laurique. Une nouvelle forme cristalline a été obtenue dont le groupe d'espace est  $P2_1/c$ . L'unité asymétrique contient deux molécules indépendantes, une de laurate de sodium et l'autre d'acide laurique. L'approche de la surface de Hirshfeld ainsi que l'analyse des contacts d'enrichissement ont été utilisées pour mener une étude détaillée sur les interactions atomiques. Ces analyses ont permis de caractériser une interaction particulière entre le groupement carboxylique de l'acide laurique

et le groupement carboxylate du laurate de sodium. Ces deux groupes sont reliés par une liaison hydrogène OH...O asymétrique à double puits et à faible barrière d'énergie (LBHB). Ce type particulier de liaison hydrogène a fait l'objet d'une publication parue en 2023 dans le Journal of Chemical Crystallography.

L'orientation de la thèse s'est également tournée vers le domaine de la cristallographie des protéines. Au cours de cette réorientation, deux protéines ont fait l'objet d'une étude approfondie : Tout d'abord, une attention particulière a été portée à la lipase B de *Candida Antarctica* (CALB). Ensuite, la seconde protéine d'intérêt a été la Neuropiline 1 humaine (Nrp1).

En ce qui concerne l'enzyme CALB, celle-ci a été cristallisée avec de la lysine pour tenter de comprendre sa spécificité vis-à-vis de cette enzyme. CALB est une enzyme utilisée pour tenter d'acyler des acides aminés avec des acides gras. Deux sites d'acylation sont possibles avec la lysine car cet acide aminé contient deux groupements amine. Les tests biochimiques avaient montré que l'amine de la chaîne latérale était préférentiellement acylé. Un cocrystal CALB/Lysine a été obtenu et l'enregistrement des données a été réalisé au Synchrotron Soleil de Paris-Saclay. Le cristal obtenu présente une macule par pseudo-mériédrie. Il n'a pas été possible de finaliser l'affinement probablement parce que la macule n'a pas été correctement modélisée. Il en résulte des cartes de densité électronique bruitées qu'il est difficile d'interpréter avec confiance. Les cartes de densité électronique semblent en accord avec les tests biochimiques. En effet, d'une part la lysine semble se fixer dans le site actif de CALB et d'autre part c'est l'amine de la chaîne latérale qui se trouve à proximité des résidus catalytiques plutôt que l'amine de la chaîne principale.

Comme il est également possible d'acyler des molécules appartenant à la famille des polyphénols, une tentative de cristallisation de la protéine CALB avec un flavonoïde ; la rutine, a été entreprise. Bien que des cristaux aient été obtenus, leur pouvoir de diffraction s'est avéré insuffisant, voire inexistant. À l'avenir, il serait intéressant de trouver de nouvelles conditions de cristallisation pour améliorer la diffraction et obtenir des informations structurales précieuses sur l'interaction entre l'enzyme CALB et un substrat potentiel.

L'étude cristallographique du fragment b1 de la neuropiline-1 (Nrp1) représente une avancée significative dans la compréhension des mécanismes moléculaires sous-jacents au développement embryonnaire et à la croissance vasculaire. Les mutations introduites pour exposer le site de liaison du facteur de croissance endothélial vasculaire (VEGF) démontrent

l'intérêt de l'approche employée pour obtenir une structure cristalline permettant la fixation de ligands. De plus, la cocrystallisation avec le peptide KDKPPR a éclairé les interactions spécifiques entre Nrpl et ce ligand potentiel, fournissant ainsi des informations prometteuses pour le développement de thérapies ciblées contre certains cancers. Ces études ont abouti à la publication dans la revue scientifique *Molecules*. L'utilisation de méthodes analytiques telles que l'approche de la surface de Hirshfeld et l'analyse des contacts d'enrichissement a enrichi notre compréhension des interactions atomiques au sein de ces systèmes complexes. Cela nous a permis de déterminer les résidus d'acides aminés qui interagissent au niveau du site d'interaction du facteur de croissance VEGF. Ces approches complémentaires ont renforcé la robustesse et la fiabilité de nos résultats, ouvrant la voie à des interprétations plus précises et à de nouvelles perspectives dans le domaine de la biologie structurale.

Cette thèse démontre clairement la valeur d'une approche interdisciplinaire et de la combinaison de techniques expérimentales avancées pour aborder des problématiques complexes en chimie et en biologie. Les résultats obtenus ont des implications potentielles pour des applications biotechnologiques et médicales, et ils fournissent également des connaissances fondamentales qui contribueront à éclairer les développements scientifiques.

Une des perspectives intéressantes serait d'explorer des complexes enzymatiques impliquant non seulement la CALB, mais également d'autres enzymes, cofacteurs et substrats. L'analyse des interactions complexes au sein de systèmes biologiques plus réalistes pourrait révéler des mécanismes encore plus complexes et ouvrir de nouvelles cibles pour des applications biomédicales ou biotechnologiques.

La combinaison de la cristallographie aux rayons X avec d'autres techniques structurales avancées, comme la cryo-microscopie électronique (cryo-EM) et la résonance magnétique nucléaire (RMN), pourrait fournir une compréhension plus complète des interactions moléculaires et des changements conformationnels, en particulier pour les systèmes plus vastes et flexibles.





---

*ANNEXES*

---



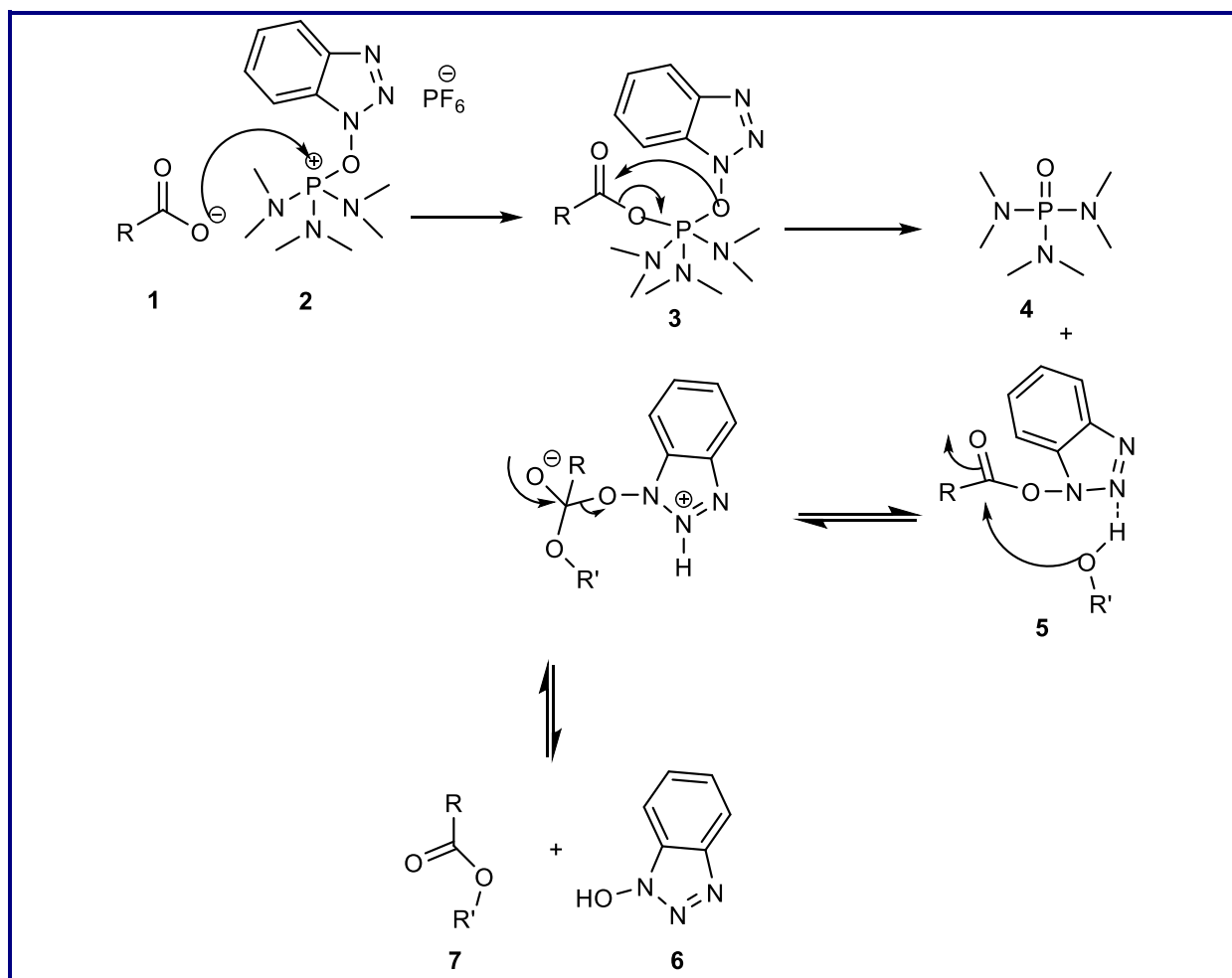
## Table des matières

ANNEXES .....	197
ANNEXE I : Généralités sur les mécanismes des réactions d'estérification.....	201
ANNEXE II : Structures topologiques des substrats et ligand utilisés au cours de ce projet .....	203
ANNEXE III : Liste des publications et communications .....	204



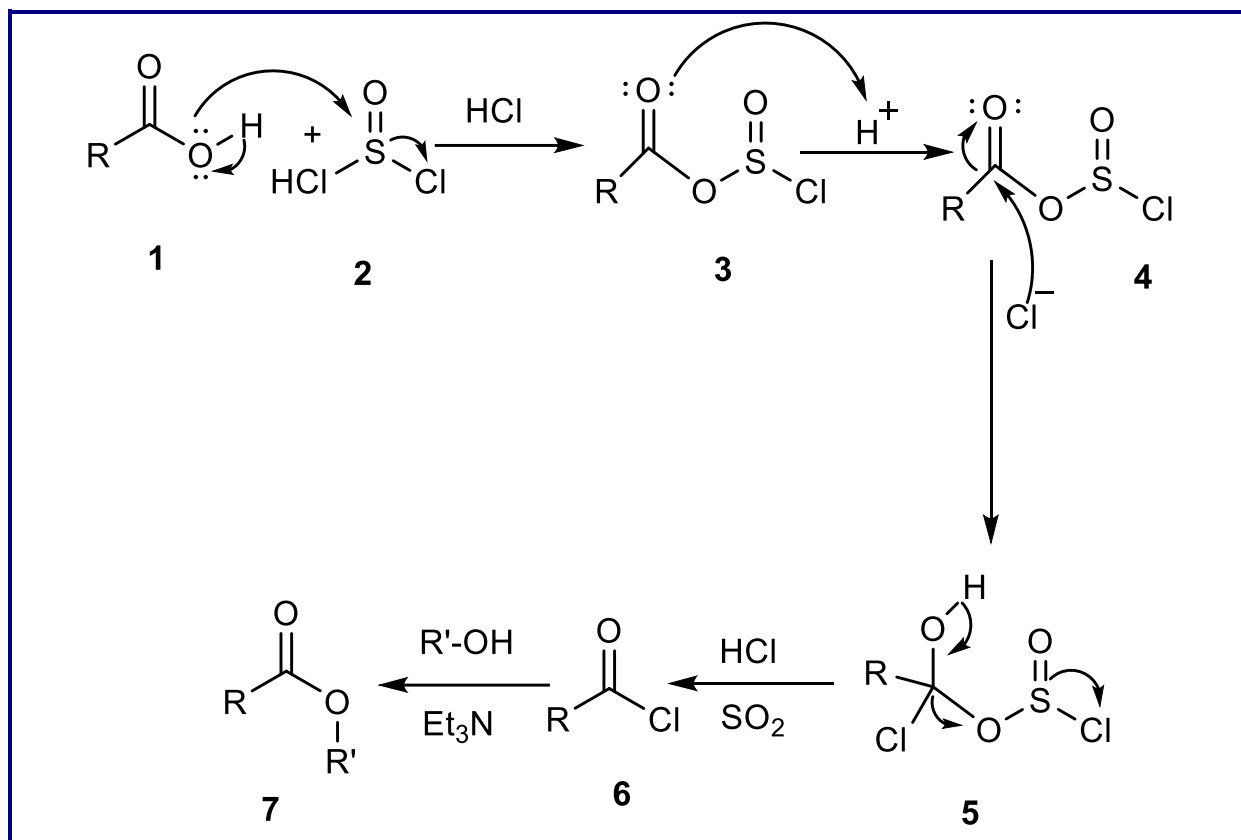
## ANNEXE I : GENERALITES SUR LES MECANISMES DES REACTIONS D'ESTERIFICATION

✚ Mécanisme de l'estérification avec HOBT / DMAP



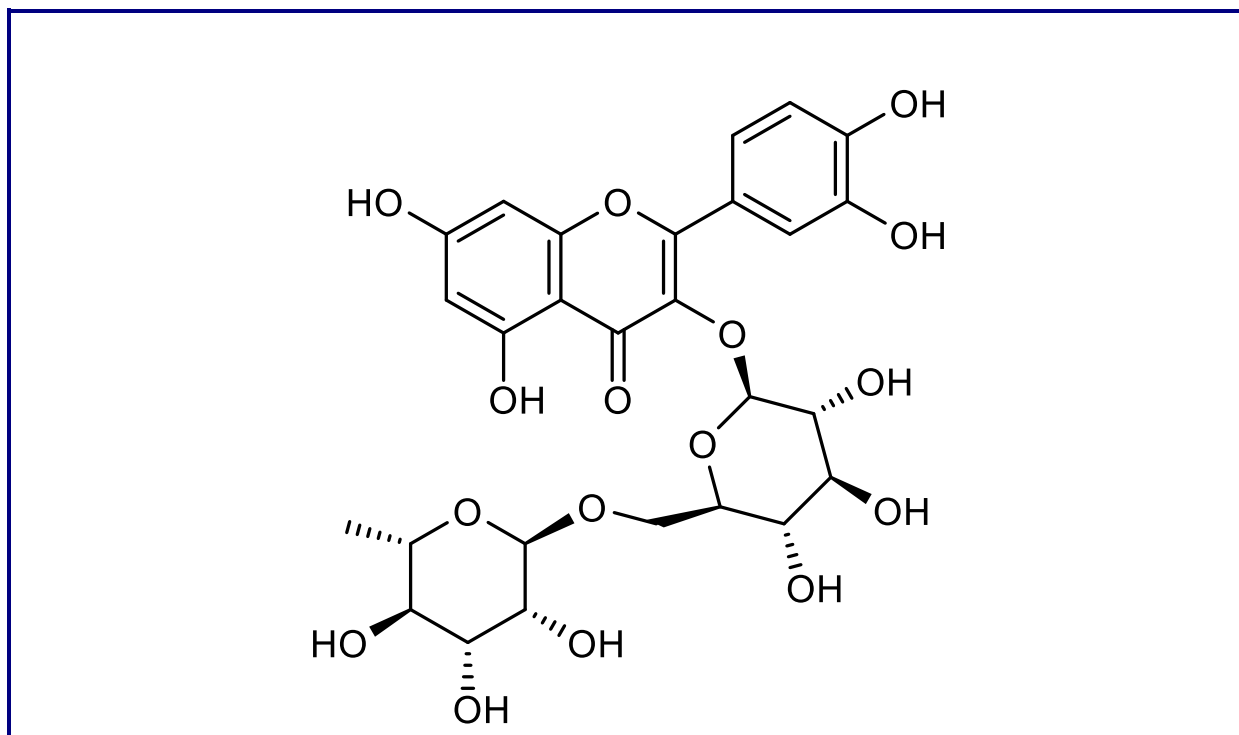
Mécanisme de l'estérification avec HOBT/DMAP

✚ Mécanisme de l'estérification avec  $\text{SOCl}_2$

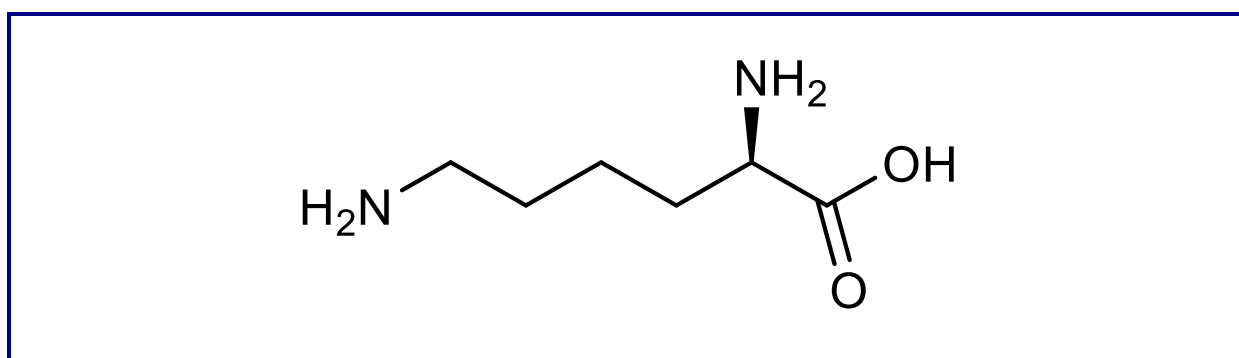


Mécanisme de l'estérification avec  $\text{SOCl}_2$

**ANNEXE II : STRUCTURES TOPOLOGIQUES DES SUBSTRATS ET LIGAND UTILISES AU COURS  
DE CE PROJET**



Structure chimique de La rutine, également connue sous le nom de quercétine-3-O-rutinoside, est un glycoside flavonoïde. La poudre de rutine est cristalline et jaune. Elle possède de puissantes propriétés antioxydantes et anti-inflammatoires. Elle agit en neutralisant les radicaux libres et en réduisant le stress oxydatif, qui peut endommager la peau et accélérer le vieillissement cutané.



Structure chimique de la lysine

### ANNEXE III : LISTE DES PUBLICATIONS ET COMMUNICATIONS

[C1] Modification chimique, Purification, encapsulation, co-cristallisation de triterpènes à activité antibactérienne et anticancéreuse en vue d'améliorer leur biodisponibilité, **I. Goudiaby**, M. Camara, A. Gassama, C. Jelsch. « *Salon International du Médicament et des Dispositifs Médicaux* », Juillet **2018**, Dakar, Sénégal.

[C2] Modification chimique, encapsulation, et co-cristallisation de triterpènes en vue d'améliorer leur biodisponibilité. **I. Goudiaby**, M. Camara, A. Gassama, C. Jelsch. « *Chimie et sécurité en Afrique de l'Ouest* », Juin **2019**, Ziguinchor, Sénégal.

[C3] Structures cristallographiques et fonctions des protéines Candida Antarctica Lipase B (CALB) et la Neuropiline-1 (Nrp-1). **I. Goudiaby**, C. Didierjean, A. Gassama, F. Favier, M. Camara, C. Jelsch. « *Recherche et Innovation dans un contexte de pandémie* », Novembre **2022**, Ziguinchor, Sénégal.

[C4] Etudes cristallographiques des protéines Candida Antarctica Lipase B (CALB) et la Neuropiline-1 (Nrp-1). **I. Goudiaby**, M. Camara, A. Gassama, C. Didierjean, F. Favier, C. Jelsch. « *La recherche scientifique au service de la transformation sociale* », Juin **2022**, Bambey, Sénégal.

[P1] Case of Charge-Assisted Hydrogen Bonding in the Crystal Structure of Sodium Laurate, Lauric Acid. **I. Goudiaby**, B. Guillot, E. Wenger, S. Soudani, C. ben Nasr, M. Camara and A. Gassama, C. Jelsch, *J. Chem. Crystallogr.* (**2022**), 53, 93-104.

[P2] New Crystal Form of Human Neuropilin-1 b1 Fragment with Six Electrostatic Mutations Complexed with KDKPPR Peptide Ligand. **I. Goudiaby**, T. E. Malliavin, E. Mocchetti, S. Mathiot, S. Acherar, C. Frochot, M. Barberi-Heyob, B. Guillot, F. Favier, C. Didierjean, and C. Jelsch. *Molecules* (**2023**), 28, 5603.

[C5] Mutation et cristallisation de la Neuropiline1- b1 (Nrp1-b1) en vue d'améliorer son mode d'interaction avec d'autres ligands. **I. Goudiaby**, M. Camara, A. Gassama, C. Didierjean, F. Favier, C. Jelsch. « *Chimie et Développement minier : Défis et Opportunités pour l'Afrique* », Août **2023**, Conakry, Guinée.



## Résumé

Cette thèse porte sur les Études Structurales et Fonctionnelles des Protéines Lipase B de *Candida Antarctica* (CALB), de la Neuropiline-1 (Nrp-1) et de l'Acide Laurique.

Un des premiers objectifs de cette thèse était d'améliorer la solubilité de l'acide bétulinique (BTA) dans l'eau en modifiant sa fonction alcoolique par une réaction de couplage pseudo-peptidique avec le 2-chlorocinnamoyl et le chlorure de phénoxyacétyle.

Parallèlement, cette thèse s'est penchée sur la recherche de cocristaux d'acide laurique et de lysine. Des cristaux de laurate de sodium, acide laurique (NaLLA) ont été obtenus et la structure a été résolue grâce à la diffraction des rayons X sur monocristal. Ces données ont permis d'étudier les interactions atomiques en utilisant l'approche de la surface de Hirshfeld et l'enrichissement des contacts.

Ensuite, cette thèse s'est réorientée vers le domaine de la cristallographie des protéines. Au cours de cette réorientation, deux protéines ont été soumises à une étude approfondie.

Dans un premier volet, une attention particulière a été portée à la lipase B de *Candida Antarctica* (CALB), une enzyme présentant des propriétés énantiométriques intéressantes pour la production de molécules par biocatalyse. Une cocrystallisation de la CALB en présence de la lysine a été entreprise pour mieux comprendre les mécanismes de fixation de cette dernière sur le site actif de la protéine. Les cristaux obtenus ont été caractérisés par la diffraction des rayons X et la structure a été résolue par la méthode du remplacement moléculaire. Il s'est avéré que les cristaux présentent une macule avec une loi de type  $(h,-k,-l)$ .

La seconde protéine d'intérêt était la Neuropiline-1, souvent abrégée Nrp-1, une protéine humaine transmembranaire. Elle est composée de plusieurs domaines, dont un domaine extracellulaire impliqué dans la liaison à certains ligands, et un domaine intracellulaire participant à la signalisation cellulaire. Nrp-1 interagit avec plusieurs facteurs de croissance et molécules de signalisation, tels que les sémaphorines de classe 3 (Sema 3, notamment Sema3A) et le facteur de croissance de l'endothélium vasculaire (VEGF), pour réguler la croissance des neurones. En plus de son rôle dans le système nerveux, la Neuropiline-1 est également impliquée dans la régulation de la croissance des vaisseaux sanguins (angiogénèse) et d'autres processus physiologiques. Le Nrp-1 est une cible thérapeutique potentielle car elle est supposée jouer un rôle dans le développement de certains cancers en facilitant la vascularisation tumorale.

L'étude cristallographique a été réalisée sur le fragment b1 de Nrp-1 car ce fragment est soluble dans l'eau et renferme le site de liaison du facteur de croissance endothélial vasculaire (VEGF). Dans les cristaux de la forme native de Nrp1-b1, ce site de liaison est obstrué par l'empilement cristallin. Il n'est donc pas possible d'utiliser les cristaux de la forme native pour tester des inhibiteurs ciblant le site de liaison du VEGF. Dans le but d'obtenir une autre forme cristalline, des mutations ont été introduites au niveau des acides aminés chargés situés à la surface de la protéine. Cela a permis la formation d'un nouvel empilement cristallin, laissant accessible le site de liaison du VEGF. Dans une perspective plus analytique, le variant a été co-cristallisé avec le peptide de séquence primaire KDKPPR. Les interactions atomiques ont été minutieusement analysées en utilisant de nouvelles approches basées sur des modèles de densité électronique multipolaire. Pour approfondir l'analyse, des simulations de dynamique moléculaire ont été mises en œuvre, permettant de comparer la dynamique du variant à celle de la forme native, ainsi qu'à celle des formes libres et en complexe avec le peptide KDKPPR.

**Mots clés :** Acide laurique, Lipase B de *Candida Antarctica*, Neuropiline-1, surface de Hirshfeld, Macle, Lysine, Facteur de Croissance de l'Endothélium Vasculaire (VEGF), ligand peptidique.

## Abstract

This thesis focuses on structural and functional studies of *Candida Antarctica* Lipase B protein (CALB), Neuropilin-1 (Nrp-1) and Lauric Acid.

One of the first objectives of this thesis was to improve the solubility of betulinic acid (BTA) in water by modifying its alcoholic function via a pseudo-peptide coupling reaction with 2-chlorocinnamoyl and phenoxyacetyl chloride.

In parallel, this thesis focused on the search for co-crystals of lauric acid and lysine. Crystals of sodium laurate, lauric acid (NaLLA) were obtained and the structure was resolved using single-crystal X-ray diffraction. These data were used to study atomic interactions using the Hirshfeld surface approach and contact enrichment.

This thesis was then reoriented towards the field of protein crystallography. During this reorientation, two proteins were subjected to in-depth study.

Firstly, particular attention was paid to *Candida Antarctica* lipase B (CALB), an enzyme with interesting enantioselective properties to produce molecules by biocatalysis. Co-crystallization of CALB in the presence of lysine was undertaken to better understand the mechanisms by which lysine binds to the protein's active site. The crystals obtained were characterized by X-ray diffraction and the structure was solved by the molecular replacement method. The crystals were found to be macular with a  $(h, -k, -l)$ -type distribution.

The second protein of interest was Neuropilin-1, often abbreviated to Nrp-1, a human transmembrane protein. It is composed of several domains, including an extracellular domain involved in binding to certain ligands, and an intracellular domain involved in cell signalling. Nrp-1 interacts with several growth factors and signaling molecules, such as class 3 semaphorins (Sema 3, notably Sema3A) and vascular endothelial growth factor (VEGF), to regulate neuronal growth. In addition to its role in the nervous system, Neuropilin-1 is also involved in the regulation of blood vessel growth (angiogenesis) and other physiological processes. Nrp-1 is a potential therapeutic target, as it is thought to play a role in the development of certain cancers by facilitating tumor vascularization.

The crystallographic study was carried out on the b1 fragment of Nrp-1, as this fragment is water-soluble and contains the binding site for vascular endothelial growth factor (VEGF). In crystals of the native form of Nrp1-b1, this binding site is blocked by crystal packing. It is therefore not possible to use crystals of the native form to test inhibitors targeting the VEGF

binding site. To obtain an alternative crystal form, mutations were introduced in the charged amino acids on the protein surface. This resulted in the formation of a new crystal stack, leaving the VEGF binding site accessible. From a more analytical perspective, the variant was co-crystallized with the primary sequence peptide KDKPPR. Atomic interactions were meticulously analyzed using new approaches based on multipolar electron density models. To extend the analysis, molecular dynamics simulations have been implemented, enabling comparison of the dynamics of the variant with those of the native form, as well as with those of the free forms and in complex with the KDKPPR peptide.

**Key words:** Lauric acid, *Candida Antarctica* Lipase B, Neuropilin-1, Hirshfeld surface, Macle, Lysine, Vascular Endothelial Growth Factor (VEGF), peptide ligand.