

وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR-ANNABA UNIVERSITY
UNIVERSITÉ BADJI MOKHTAR-ANNABA



جامعة باجي مختار - عنابة

Année 2011

Faculté des Sciences

Département de Chimie

MEMOIRE

Présenté en vue de l'obtention du diplôme de **MAGISTER**

ECOLE DOCTORALE

Option : Synthèse de molécules bioactives et matériaux organiques

Thème :

**Etude théorique et expérimentale de l'effet du
solvant sur le dédoublement enzymatique du
menthol**

Par :

M. BOUGHANI Lazhar

Directeur de mémoire : M. DJEGHABA Zeineddine Professeur Université de Annaba

Devant un jury:

Président : M. REGAINIA Zine Professeur Université de Annaba

Examineurs : M. DJEROUROU Abdelhafid Professeur Université de Annaba

M. REBANI Nacer Professeur Université de Annaba

Membre invitée : Mme. BENAMIA Fatiha M.C Université de Annaba

Résumé :

Ce travail concerne l'étude de l'influence des solvants organiques sur la transestérification énantiosélective du (\pm)-menthol par l'acétate de vinyle en présence de la lipase de *Candida antarctica* B (CAL B). L'influence de la nature du solvant organique sur le facteur d'énantiosélectivité E a été étudiée par modélisation statistique. L'analyse et le traitement des données expérimentales ainsi que le calcul du modèle mathématique reliant le facteur E , variable à expliquer, aux descripteurs des solvants, variables explicatives, ont été effectués par régression linéaire multiple.

Le modèle calculé, montre que l'énantiosélectivité de la CAL B est favorisée par des solvants hydrophobes de faible volume moléculaire.

Mots clés : Transestérification, Alcool secondaire, Lipase, régression Linéaire multiple.

Abstract:

This work relates to the study of the influence of organic solvents on the énantiosélective transesterification of (\pm)-menthol with the vinyl acetate in the presence of the lipase of *Candida antarctica* B (CAL B).

The influence of the nature of organic solvent on the factor of enantioselectivity E was studied by statistical modeling. The experimental data analysis and processing as well as the calculation of the mathematical model connecting the factor E , variable to be explained, with the descriptors of solvents, explanatory variables, were carried out by multiple linear regression.

The calculated model, shows that the enantioselectivity of the CAL B is supported by hydrophobic solvents of low molecular volume.

Key words: transesterification, secondary Alcohol, Lipase, multiple linear regression.

Remerciements

Je tiens à exprimer ma profonde gratitude et mes remerciements à mon encadreur le professeur Zeineddine DJEGHABA de m'avoir accueilli dans son équipe avec gentillesse et bienveillance, ainsi que pour ses qualités pédagogiques et l'enthousiasme communicatif dont il sait faire preuve. C'est à la fois un privilège et une expérience exceptionnelle d'avoir pu bénéficier de ses conseils.

J'ai une attention toute particulière pour Monsieur Fouad FERKOUS, Professeur à l'Université de Badji Mokhtar Annaba, pour ses conseils fréquents et avisés et pour la confiance dont il m'honore.

J'adresse mes chaleureux remerciements à monsieur Zine REGAINIA, Professeur à l'université d'Annaba pour m'avoir fait l'honneur de présider le jury, mais également pour ses excellents cours dispensés en année théorique.

J'exprime ma vive reconnaissance à Monsieur Abdelhafid DJEROUROU, Professeur à l'Université d'Annaba et directeur de laboratoire de synthèse et biocatalyse organique, pour m'avoir fait l'honneur d'accepter d'examiner et de faire partie du jury de ce travail.

Je remercie également Monsieur Nacer REBANI, Professeur à l'Université d'Annaba, pour avoir accepté de faire partie de mon jury, ses critiques apporteront, j'en suis sur, un plus à ce travail.

Je remercie et j'exprime mon profond respects à madame F.BENAMIA maître de conférences à l'université Annaba, pour son soutien et pour le temps qu'elle a pu consacrer pour m'aider dans les analyses GC qui a fait l'objet de se travail. Merci pour sa patience, sa collaboration et son aide à la réalisation de ce manuscrit

Je voudrais tout particulièrement remercier Monsieur Khair-Eddine KRAIM et Monsieur Yousef SAIHI tout d'abord pour leurs soutien tout au long de ce travail, pour leurs conseils judicieux, ensuite pour toutes leurs remarques pertinentes, pour nos longues discussions mais surtout pour la bonne ambiance qui régnait au sein de l'équipe.

Je n'oublierai pas de remercier madame BANI pour son soutien, sa gentillesse et la sympathie qu'elle m'a manifestée

Un grand merci pour mes collègues de laboratoire et amis qui ont contribué par leur soutien et amitié, chacun à sa façon à la progression de mon travail dans une ambiance toujours amicale et stimulante.

Table de matières

Résumé.....	I
Remerciements.....	II
Table de matières.....	IV
Liste des tableaux.....	VII
Liste des schémas.....	VIII
Liste des abréviations.....	IX
Introduction générale.....	1

Première partie

Chapitre 1: Généralités sur les enzymes

1. Introduction	3
2. Structure des enzymes	5
3. Classification des enzymes	6
4. les hydrolases - cas des lipases	7
4.1. Mécanisme d'action des lipases	8
4.2. Aspect cinétique	9
4.2.1. Cinétique de Michaelis-Menten	9
4.2.2. Modélisation de l'énantiosélectivité	12

Chapitre 2 : Applications des lipases en synthèse organique

1. Introduction	15
2. Reactions en milieu aqueux	15
2.1. Régiosélectivité	15

2.1. Stéréosélectivité.....	17
3. Reactions en milieu organique	18
3.1. Régiosélectivité.....	18
3.2. Stéréosélectivité.....	20
4. Effet du solvants.....	22

Deuxième partie

Chapitre 1 : Méthodes d'analyses statistiques utilisées

1. Analyse statistique univariee.....	24
2. Analyse statistique bivariee.....	24
3. Analyse statistique multivariee.....	25
3.1. Base théorique de la régression linéaire multiple.....	25
3.2. Méthodes de sélection des variables explicatives.....	28
3.3. Calcul des modèles.....	29
3.4. Validation des modèles calculés.....	32

Chapitre 2 : Influence des solvants organiques sur l'enantioselectivité de CAL B dans la résolution du menthol

1. Collecte des données.....	39
2. Analyse et resultats statistique.....	43
2.1. Analyse statistique univariée.....	43
2.2. Analyse statistique bivariée	43
2.3. Analyse statistique multivariée.....	44

MATERIELS ET METHODES

1. Matériels.....	48
--------------------------	-----------

1.1. Chromatographie en phase gazeuse (CPG).....	48
1.2. Produits chimiques et enzymes.....	48
2. Méthodes.....	48
3. Techniques générales.....	48
3.1. Analyse qualitative par chromatographie sur couche mince.....	48
3.2. Séparation et analyse quantitative par chromatographie gazeuse.....	49
4. détermination des excès énantiomériques et de l'énantiosélectivité	50
4.1. Détermination des excès énantiomériques	50
4.2. Détermination de l'énantiosélectivité.....	51
CONCLUSION GENERALE.....	52
REFERENCES BIBLIOGRAPHIQUES.....	53

Liste des tableaux

tableau	Titre	Page
1	Classification des enzymes	6
2	Hydrolyse regioselective du 1,2,3,5-tetra- <i>O</i> -acetyl- β -D-ribofuranose <u>1</u> , catalysée par des lipases immobilisées sur un support d'octyl agarosa	16
3	Hydrolyse Enzymatique du 1,2,3,5-tetra- <i>O</i> -acetyl- β -D-ribofuranose <u>1</u> , catalysée par la CRL et PFL immobilisées sur Octyl agarosa et Glutaraldehyde	17
4	Estérification de la pyridoxine dans diverses conditions	20
5	Influence du solvant sur les performances de <i>la lipase de Carica papaya</i> (CPL), dans l'estérification du de l'acide (RS)-2-(4-chlorophenoxy) propionique <u>1</u>	21
6	Tableau d'analyse de la variance d'une régression linéaire multiple avec un terme indépendant	27
7	Résolution cinétique du (\pm)-menthol par transestérification par la CAL-B. Valeurs du facteur d'énantiosélectivité <i>E</i> , des excès énantiomériques et du taux de conversion	40
8	Type et symbole des descripteurs moléculaires	41
9	Variables explicatives des solvants utilisés	42
10	Paramètres statistiques de base déterminés pour les variables utilisés	43
11	Matrice de corrélations entre les 8 variables deux à deux pour les dix-neuf solvants	44
12	Modèles calculés par la stepwise pour la réponse <i>E</i> (<i>y</i>)	45
13	Modèles calculés par la validation l'effectif d'apprentissage, pour la réponse <i>E</i> (<i>y</i>)	46
14	Validation du modèle n°2 calculé pour l'effectif d'apprentissage, afin de prédire la réponse <i>E</i>	46
15	Modèle calculé et validé pour la réponse <i>E</i> (<i>y</i>)	47

Liste des schémas

schémas	Titre	Page
1	Avantage de la catalyse enzymatique par rapport à la catalyse chimique	3
2	Stabilisation des différents états de transition intermédiaires lors de l'hydrolyse d'une liaison peptidique par l' α -chymotrypsine.	4
3	Structure tertiaire d'une protéine	5
4	Mécanisme d'intervention de la triade (Ser-Hist-Asp) et formation de l'acyl-enzyme.	8
5	Représentation hyperbolique de Michaelis-Menten	11
6	Représentation de Lineweaver et Burke	11
7	Courbe de l'excès énantiomérique du substrat et celui du produit en fonction du taux de conversion	13
8	Courbe des énergies libres d'activation (énergie de Gibbs) dans une réaction de dédoublement enzymatique qui favorise l'énantiomère <i>R</i>	14
9	Hydrolyse régioselective du 1,2,3,5-tetra- <i>O</i> -acetyl- β -D-ribofuranose	16
10	L'hydrolyse enantioselective du (R,S)-2-acetoxy-4-phenyl-but-3-ene 1 , catalysée par une série de lipases en présence du toluène	17
11	L'hydrolyse enzymatique du DL-menthyl acétate 1 en présence de la lipase de <i>Burkholderia cepacia</i> , en présence de co-solvants organiques	18
12	La deacétylation du composé 1,3-di- <i>O</i> -propanoylresorcinols 1 en présence de la lipase de <i>candida antartica</i> B	19
13	Synthèse du mono ester de la pyridoxine 1 , en présence de La lipase de <i>Candida antartica</i> B	19
14	Estérfication de l'acide (RS)-2-(4-chlorophenoxy) propionique 1 , catalysée par la lipase de <i>Carica papaya</i> (CPL) en présence de solvants organiques	21
15	Résolution du (\pm)-menthol avec l'anhydride propionique, en testant la réaction dans différents solvants organique	22
16	différence d'énergie libre des états de transition des deux complexes	37
17	transestérfication du menthol par la CAL B en présence de solvants organiques	38
18	chromatogramme GC de la réaction	50

Liste des abréviations

R, S	Enantiomères <i>Rectus et Sinister</i>
E.C	Enzyme Classification
Ser	Serine
Hist	Histidine
Asp	Aspartique
ee	Excès énantiomérique
C	Conversion
<i>E</i>	Facteur de sélectivité
T°	Température
CALB	<i>Candida antractica</i> lipase B
CALA	<i>Candida antartica</i> lipase A
CRL	<i>Candida rugosa</i> lipase
LPP	Lipase <i>pancréatique du porc</i>
RML	<i>Rhizopus Miehi</i> lipase
PFL	<i>Pseudomonas fluorescens</i> lipase
HLL	<i>Humicola lanuginosa</i> lipase
MML	<i>Mucor miehei</i> lipase
AOL	<i>Aspergillus oryzae</i> lipase
DMF	Diméthylformamide
DMSO	Diméthylsulfoxyde
iPr ₂ O	Disopropylether
[Bmim]PF ₆	1-Butyl-3-méthylimidazolium hexafluorophosphate
3D MoRSE	3D Molecule Representation of Structures Based on Electron Diffraction
dl	Degré de liberté
SC	Sommes des carrées
MC	Moyenne des carrées
MSE	L'erreur quadratique moyenne (Mean Square Error)

Introduction générale

Les lipases (EC 3.1.1.3) sont des biocatalyseurs très utilisées dans des réactions d'hydrolyses, estérification et transestérification, etc. Ce sont des catalyseurs très largement utilisés en synthèse organique en raison de leur stabilité et de leur activité en milieu solvant. De plus, elles présentent une grande chimio, régio et stéréosélectivité¹, ce qui en fait des outils très sollicités dans diverses industries chimiques et surtout pharmaceutiques où c'est l'énantiosélectivité des enzymes qui est recherchée. A ce titre, la recherche de moyens permettant d'améliorer la sélectivité des enzymes est d'actualité. Parmi ces moyens, les solvants peuvent jouer un rôle non négligeable dans ce sens. En effet, il est maintenant connu que la nature du solvant peut avoir une influence sur le fonctionnement de l'enzyme et particulièrement sur l'énantiosélectivité du biocatalyseur.

C'est dans ce cadre que s'inscrit notre présent travail qui est consacré à l'étude de l'influence des solvants organiques sur la transestérification énantiosélective du (\pm)-menthol par l'acétate de vinyle en présence de la lipase de *Candida antarctica* B (CAL B).

L'action d'un solvant dans un milieu réactionnel est le résultat de différents types d'interactions entre ce dernier et les entités présentes dans le milieu réactionnel, à fortiori dans les réactions biocatalysées où ces interactions concernent en même temps le substrat et le biocatalyseur. Etant donné la complexité de ces interactions, nous avons décidé d'aborder cette problématique par une approche multiparamétrique quant à la nature des solvants. Pour ce faire, nous avons fait appel aux méthodes statistiques d'analyse des données et aux moyens informatiques de calcul afin de modéliser la relation entre l'énantiosélectivité à travers le facteur d'énantiosélectivité E et la nature du solvant à travers leurs descripteurs physico-chimiques et moléculaires.

Après avoir rappelé quelques propriétés générales des enzymes, les principales utilisations des lipases en synthèse organique et l'intérêt que peut avoir la présence d'un solvant dans les réactions biocatalysées, nous décrirons les méthodes statistiques de traitement des données utilisées et exposerons, enfin, les résultats que nous avons obtenus.

Le plan de notre travail sera le suivant :

Première partie

Chapitre 1: Généralités sur les enzymes

Chapitre 2: Applications des lipases en synthèse organique

Deuxième partie

Chapitre 1: Méthodes d'analyses statistiques utilisées

Chapitre 2: Influence des solvants organiques sur l'enantiosélectivité de la lipase de
Candida antarctica B dans la résolution du menthol

Conclusion générale

PREMIERE PARTIE

➤ ***Chapitre 1: Généralités sur les enzymes***

➤ ***Chapitre 2: Applications des lipases en synthèse organique***

1. INTRODUCTION

Les enzymes sont des catalyseurs biologiques de nature protéique remarquablement efficaces et douées d'une grande spécificité fonctionnelle. In vivo, elles interviennent dans tous les processus métaboliques de synthèse, de régulation, de reproduction et de dégradation et permettent à toutes ces réactions de se dérouler dans des conditions de température, de pression, de concentration, de pH et de milieu, remarquablement « douces ». Comme les catalyseurs chimiques, les enzymes n'interviennent que dans les réactions métaboliques thermodynamiquement possibles qu'elles accélèrent en diminuant l'énergie d'activation de la réaction, c'est-à-dire en stabilisant l'état de transition par rapport à l'état fondamental (schéma 1).

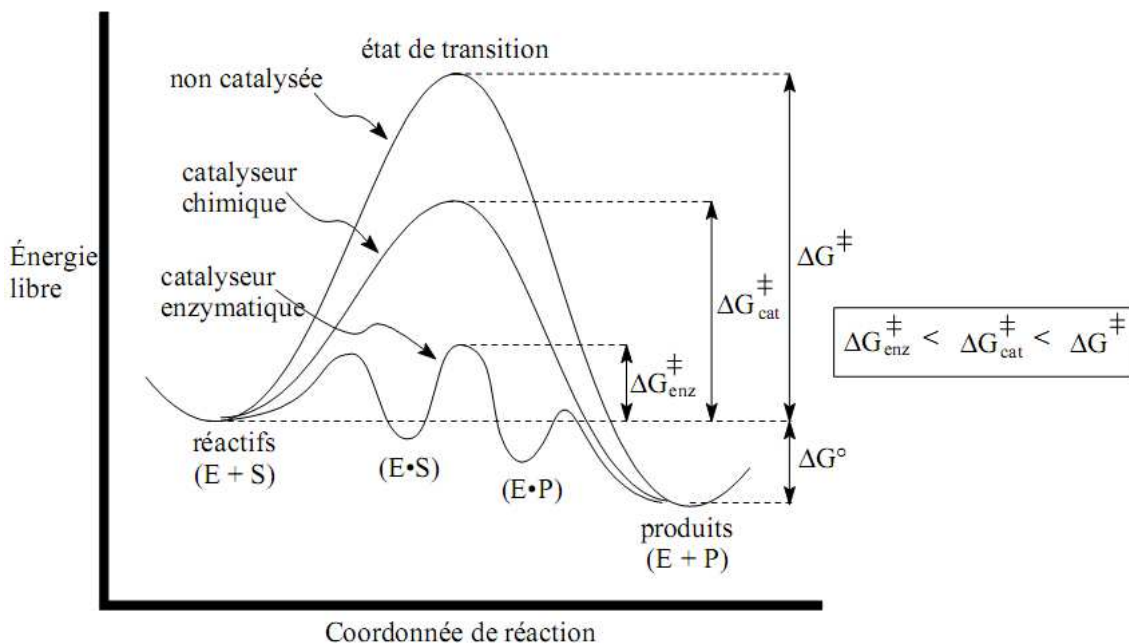


Schéma 1 : Avantage de la catalyse enzymatique par rapport à la catalyse chimique.

Différents types d'interactions sont utilisés par les enzymes afin d'augmenter leur affinité pour le substrat et stabiliser l'état de transition :

- Liaison covalente (stabilisation de 40 à 110 kcal/mole)
- Interactions électrostatiques (stabilisation d'environ 5 kcal/mole)
- Interactions ion-dipôle et dipôle-dipôle (stabilisation de 1 kcal/mole)
- Liaisons hydrogènes (stabilisation de 3-10 kcal/mole)
- Complexes de transfert de charge (implique les électrons π des anneaux aromatiques) (stabilisation < 3 kcal/mole)

- Interactions hydrophobes (stabilisation de 0,5 kcal/mole)

- Forces de Van der Waals (stabilisation de 0,5 kcal/mole)

L'exemple ci-dessous montre l'intervention de ces différentes interactions pour la stabilisation des différents états de transitions intermédiaires menant à la formation du produit.

Ⓒ et Ⓓ indiquent les extrémités C- et N-terminales du peptide.

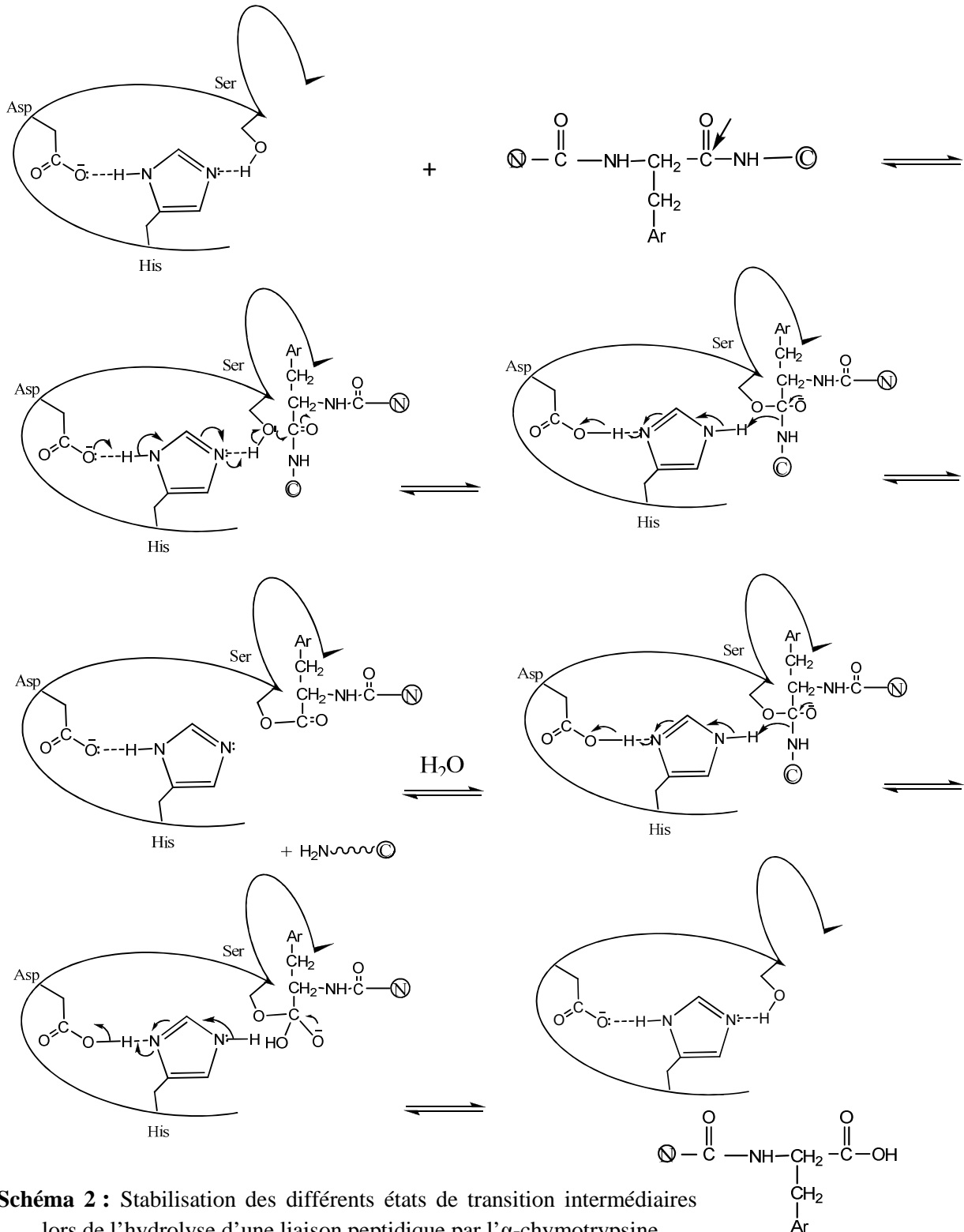


Schéma 2 : Stabilisation des différents états de transition intermédiaires lors de l'hydrolyse d'une liaison peptidique par l'α-chymotrypsine.

Cependant, le caractère fondamental des enzymes est leur spécificité pour une réaction ou un type de réaction et pour un ou un nombre limité de substrats et surtout, elles peuvent être chimio-, régio-, diastéréo- et énantiosélectives¹.

2. STRUCTURE DES ENZYMES

Comme toutes les protéines, les enzymes sont constituées d'un enchaînement d'acides aminés, appelé séquence ou structure primaire de la protéine. Cette chaîne se replie et s'enroule sur elle-même pour finalement adopter une structure spatiale compacte qu'on appelle structure tertiaire. Diverses forces provenant toutes des chaînes latérales des acides aminés constituant la protéine, interviennent pour stabiliser cette grosse molécule : ponts disulfures, liaisons hydrogènes, attractions et répulsions électrostatiques :

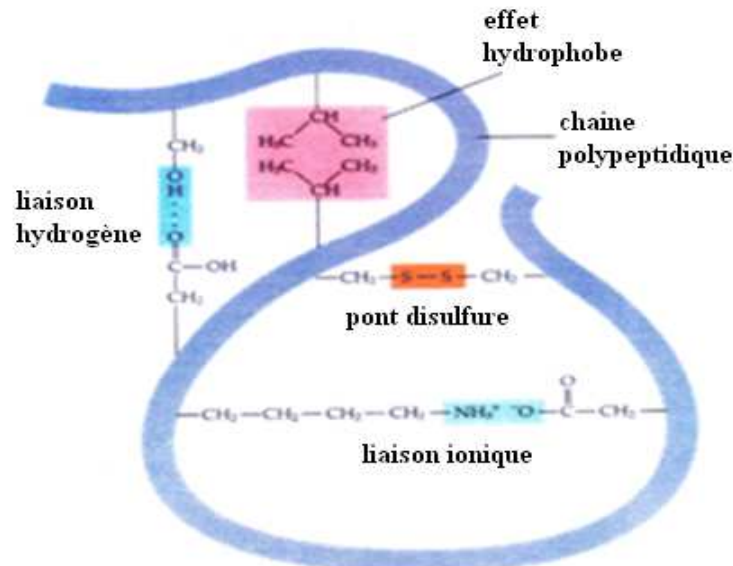


Schéma 3 : structure tertiaire d'une protéine.

Cette structure tertiaire est importante car elle permet l'aménagement d'une poche, appelée site actif de l'enzyme, où a lieu la complexation du substrat et l'activation de la réaction. C'est également la géométrie de ce site qui détermine la sélectivité des enzymes. En effet, pour que la catalyse ait lieu, la géométrie du substrat et celle du site actif, doit être compatibles. Il en résulte la formation d'un complexe « enzyme-substrat » qui a lieu, soit par simple complémentarité stérique, soit par ajustement induit, facilité par la plasticité dont est dotée fondamentalement la protéine.

3. CLASSIFICATION DES ENZYMES

Une enzyme est en général spécifique à une réaction et à un substrat donné. Une classification a été adoptée et chaque enzyme complètement caractérisée reçoit un numéro de code de quatre chiffres précédé de E.C (Enzyme Classification) :

- 1^{er} chiffre : type de réaction catalysée (il existe 6 classes ou groupes).
- 2^{ème} chiffre : (sous classe) type de groupement chimique ou de liaisons.
- 3^{ème} chiffre : nature précise des groupes et les mécanismes réactionnels en jeu.
- 4^{ème} chiffre : numéro d'ordre dans le groupe et dans le sous-groupe.

Tableau 1: Classification des enzymes.

	Classes d'enzymes	Exemple d'enzyme	N° Code
EC 1.	OXYDO-REDUCTASES		
	sur aldéhyde ou oxo	Formiate déshydrogénase	1.2.1.2
	sur liaison C-C	Fumarate réductase	1.3.1.6
	sur liaison amine	Glutamate déshydrogénase	1.4.1.2
	réductases diverses	Nitrate réductase (plantes)	1.6.6.2
	donneurs azotés	Nitrate réductase (<i>E. Coli</i>)	1.7.99.4
	oxygénases	Catéchol 1,2-dioxygénase	1.13.11.1
EC 2.	TRANSFERASES		
	de groupe monocarboné	Sérine hydroxyméthyl transférase	2.1.2.1
	d'aldéhyde ou de cétone	Transaldolase	2.2.1.2
	d'acyle	Choline acétyltransférase	2.3.1.6
	de glycosyle	Lévane sucrase	2.4.1.10
	de groupe azoté	Aspartate aminotransférase	2.6.1.1
EC 3.	HYDROLASES		
	Hydrolyse d'esters Glycosylases	Acétylcholine estérase	3.1.1.7
	Hydrolyse de peptide C-N autre que peptide	Alpha-amylase	3.2.1.1
	Hydrolyse d'anhydride	Chymotripsine	3.4.4.5

		Uréase	3.5.1.5
		Pyrophosphatase	3.6.1.1
EC 4.	LYASES		
	sur liaison C-C	Glutamate décarboxylase	4.1.1.15
	sur liaison C-O	Citrate déshydrogénase	4.2.1.4
	sur liaison C-N	Aspartate ammoniac lyase	4.3.1.1
	sur liaison P-O	Adénylate cyclase	4.6.1.1
EC 5.	ISOMERASES		
	Racémases, épimérasés	UDP-glucose 4-épimérase	5.1.3.2
EC 6.	LIGASES		
	(Enzymes effectuant des synthèses avec une source d'énergie)		
	forment C-O	Alanyl-ARNt-synthétase	6.1.1.7
	forment C-S	Acétyl-coenzyme A synthétase	6.2.1.1
	forment C-N	Glutamine synthétase	6.3.1.2
	forment C-C	Pyruvate carboxylase	6.4.1.3

4. LES HYDROLASES - CAS DES LIPASES

Parmi les enzymes les plus utilisées par les chimistes, dans le domaine de la recherche ou celui de l'industrie, les hydrolases sont les plus importantes et ce, en raison de leur faible coût et de facilité de manipulation. Elles peuvent être utilisées sous forme de poudre lyophilisée ou immobilisée sur un support.

Dans cette classe d'enzymes, on trouve les lipases (EC 3.1.1.3) qui sont les plus employées par les chimistes organiciens^{2,3}. Comme pour tous les biocatalyseurs auxquels fait appel le chimiste, les lipases sont sollicitées pour leur sélectivité et particulièrement leur énantiosélectivité qui permet l'accès à des molécules chirales dont l'intérêt, en pharmacie par exemple, n'est plus à démontrer.

4.1. Mécanisme d'action des lipases

Les lipases diffèrent entre elles par le nombre d'acides aminés constituant leur structure primaire. Par exemple, la lipase *Candida rugosa* contient plus de 500 acides aminés tandis que la lipase de *Candida antarctica* B en contient 317. Cependant, la caractéristique commune de toutes les lipases est que leurs sites actifs sont constitués de trois acides aminés, appelés la triade sérine, aspartate ou glutamate et histidine³. Cette triade d'acides aminés, ajoutée à quelques autres résidus, constitue le site actif de l'enzyme. Celui-ci est protégé par un « couvercle » (en fait un segment hélicoïdal hydrophobe) qui en bloque l'accès et ne s'ouvre qu'au contact d'une phase hydrophobe pour permettre la complexation du substrat, lipide hydrophobe, et faciliter ainsi la catalyse de la réaction. Cette activation interfaciale du biocatalyseur constitue la caractéristique principale des lipases⁴.

Le mécanisme d'intervention de la triade (Ser, Hist et Asp) lors de l'hydrolyse d'un ester est donné par le schéma suivant³ :

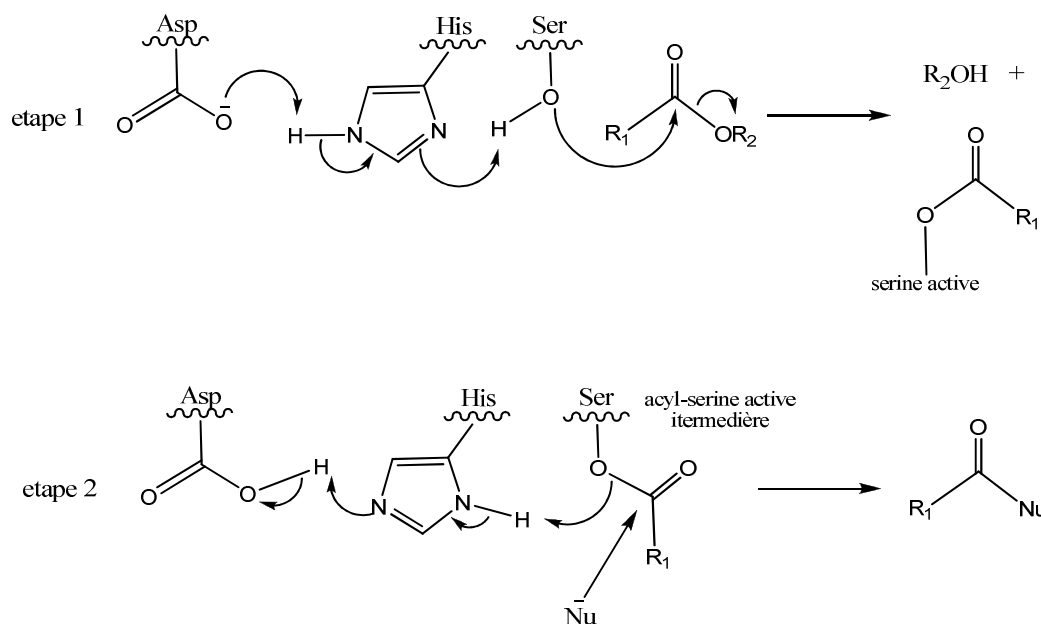
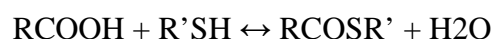
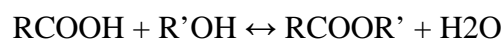


Schéma 4 : Mécanisme d'intervention de la triade (Ser-Hist-Asp) et formation de l'acyl-enzyme.

Comme le montre le schéma ci-dessus, la disposition particulière des 3 résidus formant la triade permet, dans une première étape, l'attaque nucléophile du substrat $R_1CO_2R_2$ pour donner un complexe appelé acyl-enzyme avec libération de l'alcool R_2OH . Dans une seconde étape, cet intermédiaire acyl-enzyme subit une attaque nucléophile par l'eau présente dans le milieu, régénérant ainsi l'enzyme par expulsion de l'acide R_1CO_2H , produit de la réaction.

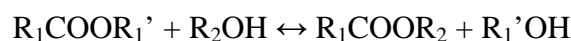
L'idée de l'utilisation des lipases en synthèse organique était basée sur l'hypothèse que d'autres nucléophiles que l'eau pourrait attaquer l'acyl-enzyme, en particulier dans un milieu pauvre en eau comme le milieu organique. C'est ainsi que depuis le début des années quatre vingt et les travaux de Zacks et A.M. Klibanov⁵, s'est développé toute une chimie autour de l'utilisation des lipases en synthèse organique qu'on peut résumer ci-dessous:

- Estérification et thioestérification :

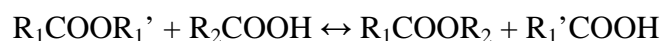


- Transestérification

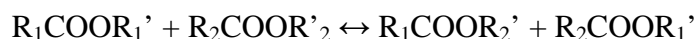
- Alcoolyse :



- Acidolyse :



- Interestérification :



4.2. Aspect cinétique

4.2.1. Cinétique de Michaelis-Menten

La vitesse de réaction enzymatique est mesurée à partir de la quantité de produit formée en un temps donné. L'affinité de l'enzyme pour son substrat est donnée par son K_m ou constante de Michaelis. Celle-ci est définie comme la concentration de substrat pour laquelle la vitesse de réaction enzymatique est la moitié de la vitesse de réaction maximale, elle a la même unité : mol.l^{-1} . De nombreux facteurs peuvent modifier la vitesse de réaction enzymatique tel que:

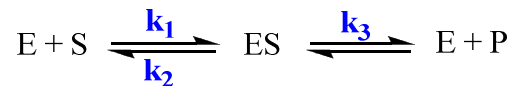
- Les concentrations en enzyme et en substrat
- Les caractéristiques physico-chimiques du milieu de réaction (température, pH, ...)
- La présence d'inhibiteurs de la réaction enzymatique. Un inhibiteur compétitif augmente la constante de Michaelis.

La constante de Michaelis est spécifique à chaque enzyme. C'est une constante cinétique caractérisant une réaction enzymatique. Elle reflète l'affinité de l'enzyme pour son substrat :

- Plus sa valeur est élevée, plus l'affinité du substrat pour l'enzyme est faible ;
- Plus elle est faible, plus l'affinité du substrat pour l'enzyme est grande.

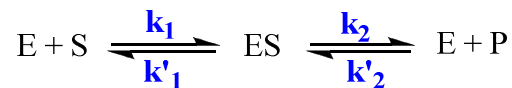
L'équation de Michaelis-Menten permet de calculer les caractéristiques d'une enzyme michaelienne, enzyme ne fixant qu'une molécule de substrat par molécule d'enzyme.

Le mécanisme de la réaction pour les enzymes à un seul substrat, ainsi que l'équation de la vitesse sont:



$$v = \frac{d[p]}{dt} = -d[S]/dt$$

a) Constante de Michaelis-Menten



K_m représente l'inverse de l'affinité de l'enzyme pour son substrat.

b) Équation de Michaelis-Menten

Cette équation est établie comme suit :

$$v = \frac{v_{max} [S]}{K_m + [S]} \quad (1)$$

Avec :

v : vitesse initiale (c'est à dire en absence de produit) de la réaction enzymatique pour une concentration de substrat $[S]$ (en $\mu\text{mol}/\text{min}$) ;

v_{max} : Vitesse initiale maximale mesurée pour une concentration saturante de substrat $[S]$ (en $\mu\text{mol}/\text{min}$) ;

$[S]$: Concentration en substrat (en mol/l) ; K_m : Constante de Michaelis spécifique de l'enzyme, c'est la concentration en substrat laquelle la vitesse initiale de la réaction est égale à $\frac{1}{2} v_{max}$ (en mol/l). Elle correspond à l'inverse de la constante d'affinité du substrat pour l'enzyme. Graphiquement, l'équation de Michaelis est une branche d'hyperbole.

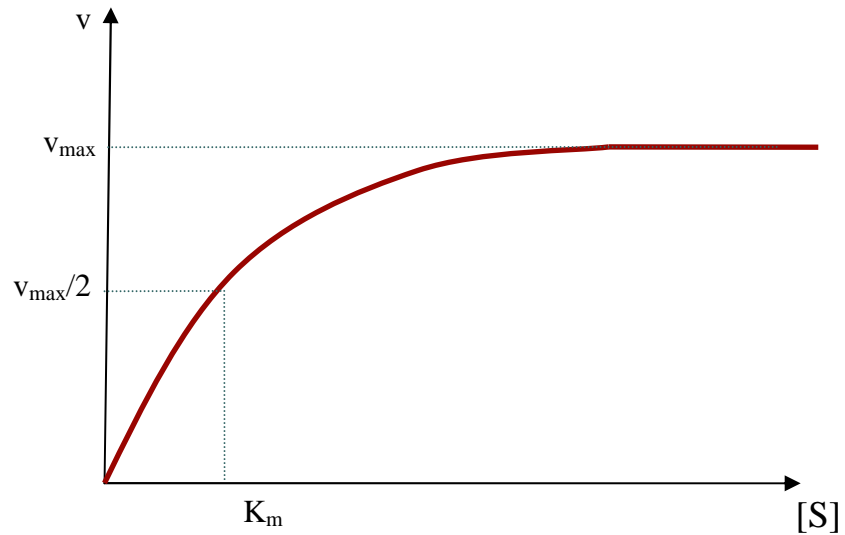


Schéma 5 : Représentation hyperbolique de Michaelis-Menten.

En pratique, on détermine les constantes de l'enzyme K_m et v_{\max} par la représentation des inverses (représentation de Lineweaver et Burke) qui est une droite d'équation.

$$\frac{1}{v_i} = \frac{K_m}{v_{\max} + [S]} + \frac{1}{v_{\max}} \quad (2)$$

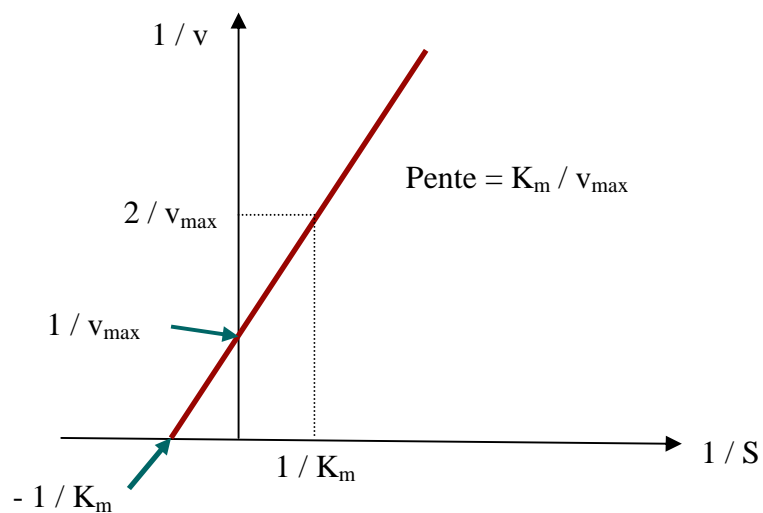
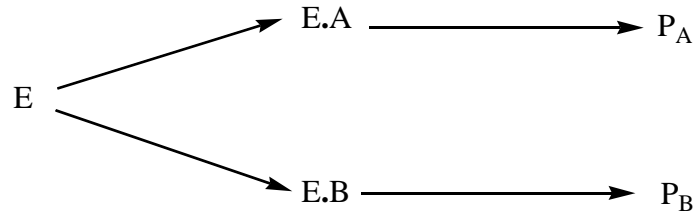


Schéma 6 : Représentation de Lineweaver et Burke.

4.2.2. Modélisation de l'énantiosélectivité

Dans le cas du dédoublement enzymatique d'un mélange racémique, il a été proposé un mécanisme réactionnel s'appliquant aux réactions non réversibles⁶.



Où,

E : l'enzyme

A : l'énantiomère R

B : l'énantiomère S

On peut donc exprimer les vitesses de disparition de A et B :

$$v_A = -\frac{d[A]}{dt} = [K_{cat}/k_m]_A [E][A] \quad (3)$$

$$v_B = -\frac{d[B]}{dt} = [K_{cat}/k_m]_B [E][B] \quad (4)$$

$$v_A/v_B = \frac{[K_{cat}/K_m]_A [A]}{[K_{cat}/K_m]_B [B]} \quad (5)$$

Le rapport v_A/v_B exprime la préférence que peut montrer l'enzyme vis-à-vis de l'un des substrats. Par intégration de ce rapport, Sih et col⁷, établissent une nouvelle équation qui définit un autre paramètre représentant la sélectivité de la réaction, le facteur d'énantiosélectivité E . Sa valeur est définie comme le rapport des constantes de spécificité pour les deux énantiomères (Eq.6)

$$E_{AB} = \frac{[K_{cat}/K_m]_A}{[K_{cat}/K_m]_B} \quad (6)$$

Il a été montré par Chen et col⁶, que ce facteur peut être exprimé par le taux d'avancement (C) de la réaction et respectivement (ee_S) l'excès énantiomérique du substrat n'ayant pas réagi et (ee_P) l'excès énantiomérique du produit formé (Eq.7, 8, 9)

$$E = \frac{\ln [(1-C)(1-ee_s)]}{\ln [(1-C)(1+ee_p)]} = \frac{\ln [(1-C)(1+ee_p)]}{\ln [(1-C)(1-ee_s)]} \quad (7)$$

Où,

$$C = \frac{ee_s}{ee_s + ee_p} \quad (8)$$

$$ee_s = \frac{[B]-[A]}{[B]+[A]} \quad ee_p = \frac{[P]-[Q]}{[P]+[Q]} \quad (9)$$

Lorsque les deux énantiomères A et B sont en compétition, il faut que le taux de conversion de la réaction ne dépassera pas 50%, afin d'éviter une décroissance rapide de la pureté optique du produit. On note qu'il est difficile de connaître avec exactitude le taux d'avancement de la réaction, indépendamment des méthodes expérimentales.

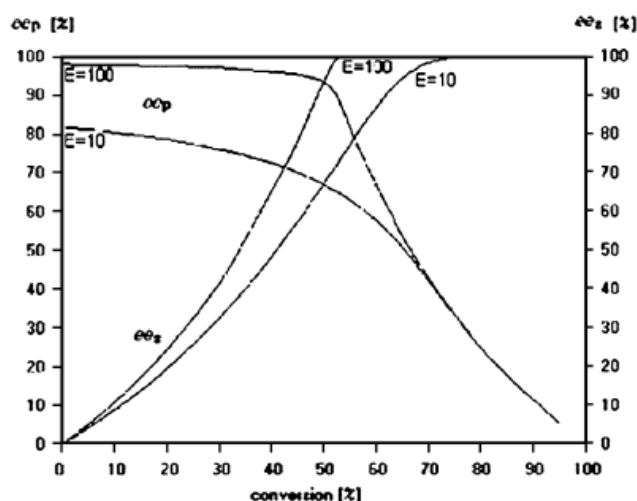


Schéma 7 : Courbe de l'excès énantiomérique du substrat et celui du produit en fonction du taux de conversion³.

L'énantioselectivité est le résultat d'une différence d'énergie libre d'activation entre les deux énantiomères (R et S), notée $\Delta_{R-S}\Delta G^\ddagger$, qui est relié au facteur E par la relation donnée par l'équation 10 et le schéma 8 ci-dessous³ :

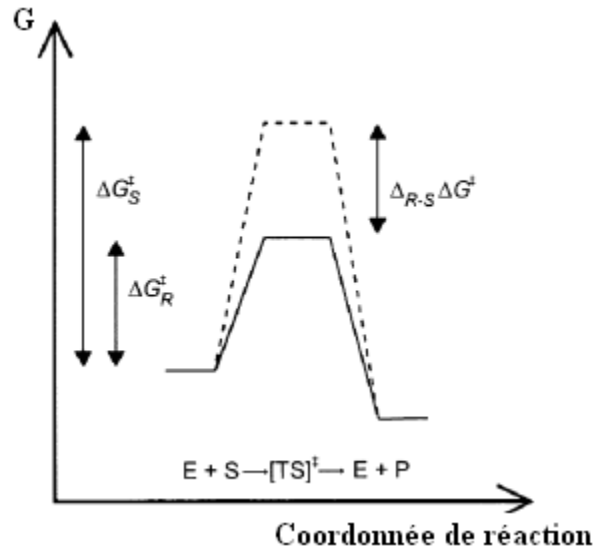


Schéma 8 : Courbe des énergies libres d'activation (énergie de Gibbs) $\Delta_{R-S}\Delta G^\ddagger$, dans une réaction de dédoublement enzymatique qui favorise l'énantiomère R .

$$\Delta_{R-S}\Delta G^\ddagger = \Delta G_R^\ddagger - \Delta G_S^\ddagger = -TR \ln E \quad (10)$$

Où,

$$E_{R,S} = \frac{(K_{cat}/K_m)_R}{(K_{cat}/K_m)_S} \quad (11)$$

Ceci permet de déduire que lorsque la vitesse initiale v augmente, le facteur d'énantioselectivité E augmente également, entraînant ainsi une baisse au niveau de l'énergie libre d'activation $\Delta_{R-S}\Delta G^\ddagger$.

1. INTRODUCTION

Les enzymes d'une manière générale et les lipases en particulier, sont devenues des outils de choix pour le chimiste organicien. En raison de leur grande sélectivité et des conditions réactionnelles douces dans lesquelles elles opèrent, les enzymes sont maintenant incontournables, en particulier dans la synthèse de composés chiraux qui sont très demandés par l'industrie pharmaceutique. Par ailleurs, en raison principalement des restrictions de plus en plus fortes en matière de pollution et de protection de l'environnement, d'autres industries intègrent les biotechnologies dans leurs processus de production. Parmi les biocatalyseurs utilisés, les lipases constituent le groupe le plus important. Elles sont non seulement capables d'assurer leur fonction naturelle d'hydrolyse des esters, mais peuvent également catalyser des réactions de transestérification, d'estérification, d'interestérification et d'aminolyse, en milieu non aqueux. Ces propriétés, ajoutées au label « naturel » qu'elles confèrent aux produits issus des processus biotechnologiques, permettent de retrouver les lipases impliquées dans la synthèse de nombreux produits aussi variés que les médicaments, les biopolymères, le biodiesel, les arômes, etc⁹.

Afin de mieux décrire notre travail par rapport à la littérature, nous rapportons dans ce chapitre les résultats obtenus dans les réactions catalysées par les lipases, en milieu aqueux et non aqueux.

2. REACTIONS EN MILIEU AQUEUX

2.1. Régiosélectivité :

Parmi les nombreux travaux montrant la régiosélectivité des lipases¹⁰⁻¹⁴, nous citerons ceux de Fernandez-Lorente et col¹⁵, décrivant l'hydrolyse de sucres peracétylés en présence de différentes lipases. Ces auteurs ont montré que parmi les différentes lipases immobilisées sur un support d'octyl agarosa, seule la lipase de *Candida rugosa* (CRL) et la lipase de *Pseudomonas fluorescens* (PFL) ont pu montrer, par hydrolyse régiosélective du 1,2,3,5-tetra-*O*-acetyl- β -D-ribofuranose **1**, des activités notables (Schéma 9, Tableau 2).

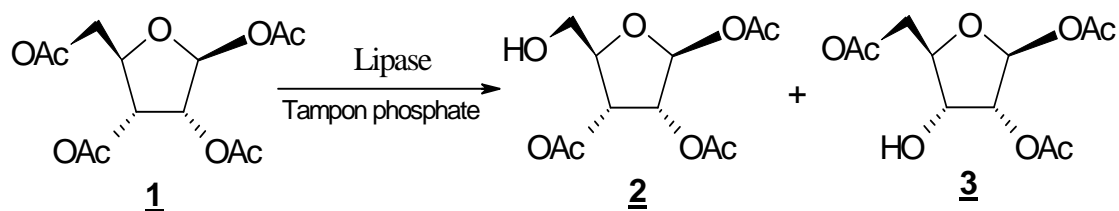


Schéma 9 : Hydrolyse régiosélective du 1,2,3,5-tetra-*O*-acetyl-β-D-ribofuranose **1**, en présence de tampon phosphate et catalysée par les lipases immobilisées sur un support d'octyl agarosa.

Tableau 2 : Hydrolyse régiosélective du 1,2,3,5-tetra-*O*-acetyl-β-D-ribofuranose **1**, catalysée par des lipases immobilisées sur un support d'octyl agarosa .

Lipase	Activité enzymatique (μmol/ml)	Rendement %
CRL	3.07	47
PFL	0.1	40
MML	0.0075	33
HLL	0.0073	25
AOL	0.008	20
CAL A	0.012	5
CAL B	0.03	2

NB: CRL: lipase de *Candida rugosa* ; PFL: lipase de *Pseudomonas fluorescens* ; MML: lipase de *Mucor miehei* ; HLL: lipase de *Humicola lanuginosa* ; AOL: lipase de *Aspergillus oryzae* ; CAL B: lipase de *Candida antarctica B* ; CAL A : lipase de *Candida antarctica A*.

Dans le cas de la lipase de *Pseudomonas fluorescens* (PFL), cette dernière, hydrolyse principalement l'ester sur la position 3 en donnant un mélange de deux produits triacetate **2** et **3** avec un rendement de 40%. La lipase de *Candida rugosa* (CRL) quant à elle, hydrolyse sélectivement l'ester sur la position 5 en donnant le mono triacetate **2** avec un rendement de 47%. L'immobilisation de ces deux lipases sur le glutaraldehyde agarosa, montre que l'activité de la lipase de *Candida rugosa* (CRL), dans les mêmes conditions réactionnelles a baissé de plus de 100 fois par rapport à son immobilisation sur octyl agarosa. Le même comportement est également observé dans le cas de la lipase de *Pseudomonas fluorescens* (PFL), (tableau 3) :

Tableau 3 : Hydrolyse Enzymatique du 1,2,3,5-tetra-*O*-acetyl- β -D-ribofuranose **1**, catalysée par la CRL et PFL immobilisées sur Octyl agarosa et Glutaraldehyde.

Lipase	Support d'immobilisation	Activité enzymatique ($\mu\text{mol/mL}$)	Rendement %
CRL	Octyl agarosa	3.07	47
PFL	Glutaraldehyde agarosa	0.03	33
CRL	Octyl agarosa	0.1	40
PFL	Glutaraldehyde agarosa	0.007	30

2.2. Stéréosélectivité

La stéréosélectivité des lipases a été utilisée soit pour effectuer des dédoublements de mélanges racémiques par résolution cinétique essentiellement, soit pour accéder à des composés optiquement actifs à partir de composés prochiraux^{3,16}.

Ghanem et col¹⁷, ont étudié l'hydrolyse énantiosélective du (R,S)-2-acetoxy-4-phenyl-but-3-ene **1**, catalysée par une série de lipases en présence de toluène.(schéma 10)

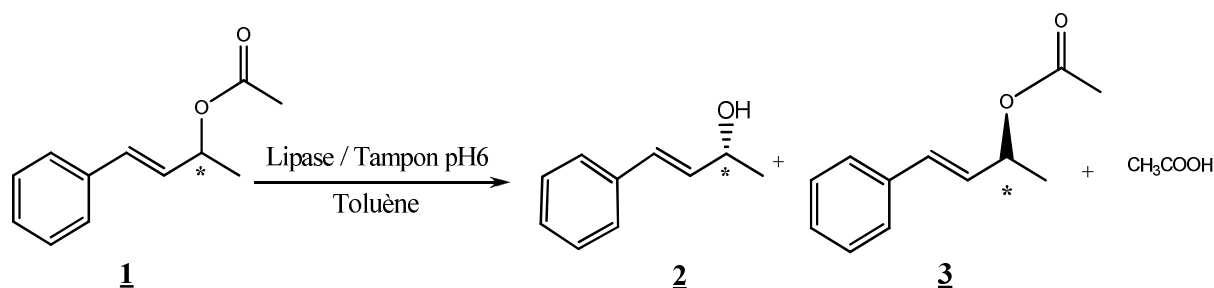


Schéma 10 : L'hydrolyse énantiosélective du (R,S)-2-acetoxy-4-phenyl-but-3-ene **1**, catalysée par une série de lipases en présence du toluène.

La présence d'un solvant organique peut également favoriser la sélectivité de l'enzyme. Lors de l'hydrolyse enzymatique de l'acétate de menthyle racémique **1**, Lijuan Yu et col¹⁸, ont montré que la présence de co-solvants organiques permet de multiplier par trois le facteur d'énantiosélectivité par rapport à la réaction sans co-solvant (Schéma 11). Les solvants utilisés dans cette réaction sont : DMF; DMSO; Acétonitrile et Acétone.

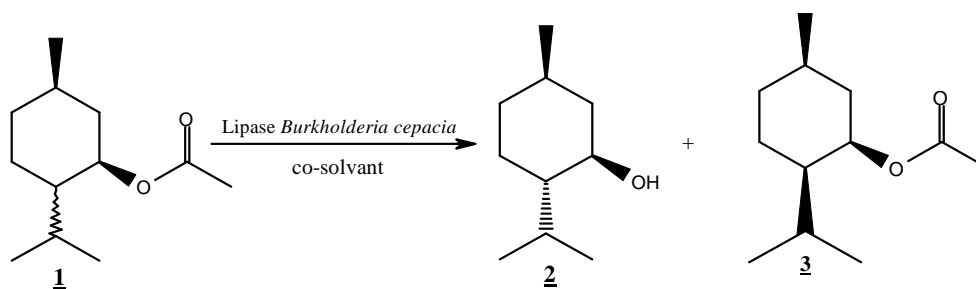


Schéma 11 : L'hydrolyse enzymatique du (\pm)-menthyl acétate **1** en présence de la lipase de *Burkholderia cepacia*, en présence de co-solvants organiques.

3. REACTIONS EN MILIEU ORGANIQUE

Initialement les lipases ont été employées pour hydrolyser des esters en milieu aqueux, l'eau étant l'environnement naturel des enzymes. Cependant, l'utilisation de l'eau limitait le champ d'action des lipases en synthèse organique. En plus des problèmes de solubilités des substrats, elle doit être éliminée pour que l'équilibre thermodynamique soit en faveur de la formation des produits. Toutefois, une quantité minimale d'eau est nécessaire pour le maintien de la conformation active de l'enzyme.

Parmi les avantages du travail en milieu organique, citons la récupération de l'enzyme pour de nouvelles utilisations et l'isolation aisée des produits réactionnels par simple filtration.

Notons enfin que la nature du solvant peut également influencer autant sur l'activité que sur la sélectivité du biocatalyseur¹⁹⁻²⁴.

3.1. Régiosélectivité :

Parmi les hydrolases, la lipase de *Candida antarctica* B (CAL B) possède une grande régiosélectivité dans les réactions de biotransformations. Elle a été mise à profit pour désacyler des dérivés du 1,3-di-O-propanoylresorcinols substitué en position 4 (**1**)²⁵. L'alcoolyse du composé **1**, donne régiospecifiquement le produit **2**. La deuxième fonction ester n'est pas touchée en raison de certainement de l'encombrement stérique créé par le groupement R adjacent. (Schéma 12).

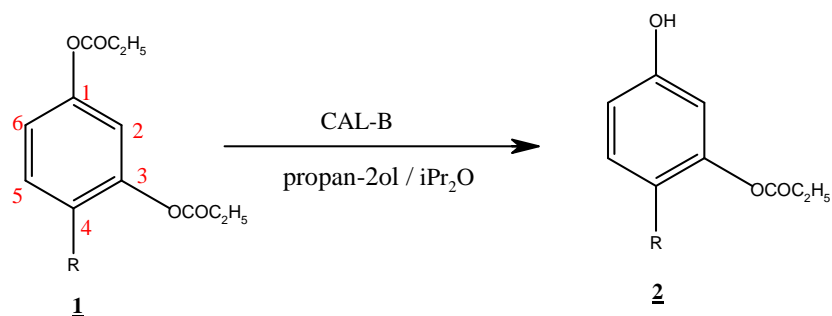


Schéma 12 : La deacétylation du composé 1,3-di-O-propanoylresorcinols **1** en présence de la lipase de *Candida antarctica B* (CAL B). Avec R= Et ; (CH₃)₃CCH₂C (CH₃)₂ ; Bn ; Cl ; Br.

Comme nous l'avons souligné plus haut, et à travers les exemples que nous allons citer, le milieu réactionnel a une influence non seulement sur l'élargissement de la palette des substrats, mais également, et surtout, sur la régio- et l'énantioselectivité des lipases.

Ainsi, Zhang et col¹⁰, ont montré lors de la synthèse de la pyridoxine par transestérification du 4,5-bis(hydroxyméthyl)-2-méthylpyridin-3-ol **1** (schéma 13), catalysée par la lipase CAL B, le comportement de l'enzyme est influencé par la nature du solvant employé, l'acétonitrile étant celui qui donne les meilleurs résultats en terme d'activité et de régiosélectivité (Tableau 4).

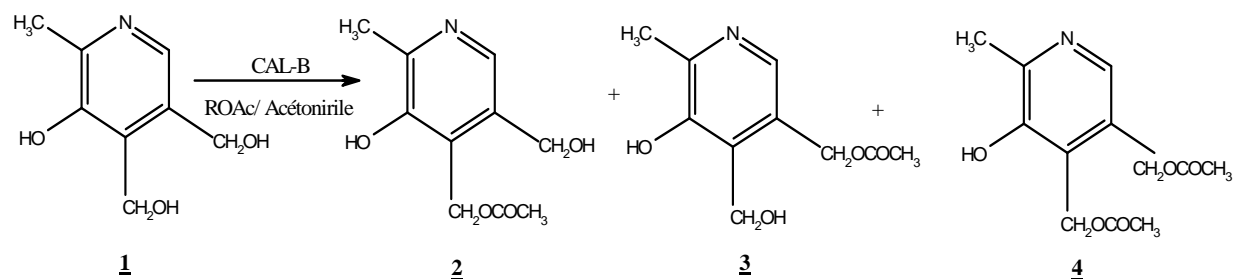


Schéma 13 : Synthèse du mono ester de la pyridoxine **1**, en présence de La lipase de *Candida antarctica B* (CAL B).

Tableau 4 : Estérfication de la pyridoxine dans diverses conditions.

Lipase	solvant	Donneur d'acyle	Temps de réaction (h)	Conversion (%)	<u>2</u> : <u>3</u> : <u>4</u>
CAL B	MeCN	Anhydride acétique	2	98,9	0 : 64 : 36
	MeCN	Acétate de vinyle	2	99,1	0 : 96 : 4
	MeCN	Acétate d'éthyle	6	26,9	0 : 96 : 4
	[Bmim]PF ₆	Anhydride acétique	2	63,5	82 : 4 : 14
	[Bmim]PF ₆	Acétate de vinyle	6	36,4	4 : 83 : 13
	[Bmim]PF ₆	Acétate d'éthyle	6	27,5	14 : 74 : 12

NB : CAL B : La lipase de *Candida antarctica B* ; MeCN : acétonitrile ; 2 : 4-*O*-acetylpyridoxine (4-AcPN) ; 3 : 5-*O*-acetylpyridoxine (5-AcPN) ; 4 : 4,5-di-*O*-acetylpyridoxine.

3.2. Stéréosélectivité

Une des plus importantes caractéristiques des biocatalyseurs est leur aptitude à différencier des énantiomères. Cette propriété des enzymes les place parmi les outils de choix tant en synthèse organique que dans le domaine de la biotechnologie^{9, 26-30}. Dans ce contexte, de nombreux auteurs, ont montré que le comportement des enzymes, en termes d'activité et de sélectivité, pouvait être modifié par de nombreux facteurs^{31,32}. Parmi ceux-ci, le milieu réactionnel joue un rôle très important, comme nous allons le voir à travers les exemples ci-dessous.

Il est très important de noter que la variation des solvants dans le cas de la résolution des racémiques, en présence des lipases peut avoir non seulement une influence sur la vitesse de la réaction mais aussi sur la sélectivité de l'enzyme³³⁻⁴¹.

C'est dans ce contexte, que Cheng et col⁴², ont décrit la sélectivité des lipases lors de l'estérfication de l'acide (RS)-2-(4-chlorophenoxy) propionique 1, catalysée par la lipase de *Carica papaya* (CPL) en présence de solvants organiques. Cette réaction a été choisie comme modèle afin d'étudier l'influence de l'alcool et les solvants organiques sur la performance de la lipase (Schéma 14).

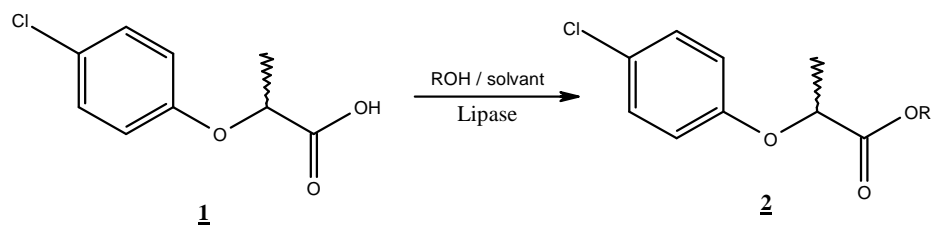


Schéma 14 : Estérification de l'acide (RS)-2-(4-chlorophenoxy) propionique 1, catalysée par la lipase de *Carica papaya* (CPL) en présence de solvants organiques.

Il a été montré que seule la nature de l'alcool et du solvant organique possèdent une influence sur l'activité enzymatique et l'énantiosélectivité de la lipase de *Carica papaya* (CPL). Pour les alcools R-OH, l'activité et la sélectivité de l'enzyme évoluent positivement avec l'encombrement stérique du groupement R de l'alcool. La nature du solvant, exprimée à travers le coefficient d'hydrophobicité $\text{Log}P$, a une influence sur le facteur d'énantiosélectivité E qui passe de 57, dans le cas du cyclohexane, à 87 lorsque c'est l'hexane ou l'isooctane qui sont employés comme solvant (Tableau 5). Cependant, aucune corrélation n'a été trouvée entre $\text{Log}P$ et le facteur d'énantiosélectivité E .

Tableau 5 : Influence du solvant sur les performances de la lipase de *Carica papaya* (CPL), dans l'estérification du de l'acide (RS)-2-(4-chlorophenoxy) propionique 1.

Solvant	$\text{Log}P$	ee _p %	V_R	V_S	E
Tétrachlorure de carbone	3	94.5	15.7	2.1	74
Cyclohexane	3.2	94.4	17.5	3.4	57
Hexane	3.5	89.1	10.3	1.2	87
Isooctane	4.5	77.5	2.6	0.3	87
Décane	5.6	91	4.3	0.7	63

NB : $\text{Log}P$: coefficient d'hydrophobicité du solvant ; ee_p% : l'excès énantiomérique du produit ; V_R et V_S : les vitesses initiales des substrats énantiomères ($\times 10^2 \text{mM h}^{-1}$) ; E : facteur d'enantioselectivité.

4. EFFETS DES SOLVANTS

A travers les exemples cités précédemment, nous avons vu que de nombreux facteurs pouvaient influencer sur le comportement des lipases, que ce soit dans les réactions d'hydrolyse ou de transestérification. Parmi ces facteurs, les solvants organiques, présents dans les réactions d'hydrolyse afin de solubiliser les substrats ou dans les réactions de transestérifications pour les avantages déjà cités, peuvent modifier autant l'activité que la sélectivité du biocatalyseur. De ce fait, la recherche de moyens pouvant aider au choix du solvant approprié pour une réaction donnée, est devenue une des préoccupations des chercheurs dans le domaine de la biocatalyse^{3, 43-47}.

C'est dans ce contexte que Shu Bai et col⁴⁸, ont évoqué le principe de l'effet de solvant sur les lipases en faisant la résolution du (\pm)-menthol avec l'anhydride propionique en présence de la lipase de *Candida rugosa*. La réaction a été réalisée dans différents solvants organiques (Schéma 15).

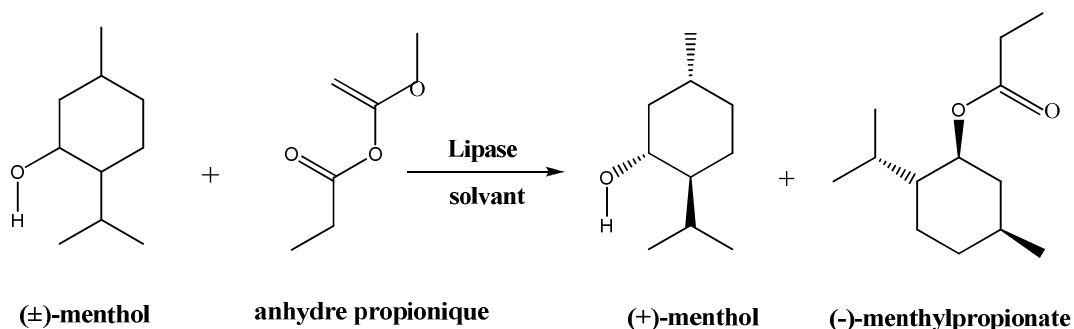


Schéma 15 : Résolution du (\pm)-menthol avec l'anhydride propionique, en testant la réaction dans différents solvants organique.

La nature des milieux organiques non seulement influe sur l'activité de l'enzyme, mais aussi l'énantiosélectivité, qui a été vérifiée par les travaux de plusieurs groupes⁴⁹⁻⁵¹. $LogP$ (logarithme du coefficient de partition d'un solvant donné entre le n-octanol et l'eau) est maintenant largement utilisé pour exprimer la polarité ou l'hydrophobicité d'un solvant. Dans cette étude, les effets des solvants sur les conversions et énantiosélectivité de la CRL, ont montré à travers les résultats obtenus dans ce travail, que les meilleures conversions sont dans le cas des solvants qui possèdent un $LogP$ élevé c'est-à-dire les solvants apolaires. De même il a été constaté que les meilleures énantiosélectivité sont dans la région des solvants apolaires.

L'étude concernant l'influence du solvant organique sur le comportement des lipases a fait déjà l'objet d'un travail de thèse de doctorat soutenue dans le même laboratoire, seulement qu'au cours de cette étude l'équipe de recherche au laboratoire a choisi comme descripteurs explicatives des paramètres physico-chimique seulement. Les résultats de cette étude obtenus nous ont conduits à conclure que la polarité du point de vue physico-chimique joue un rôle très important dans ce domaine.

De ce fait, la recherche de moyens pouvant aider au choix du milieu réactionnel approprié pour les réactions biocatalysées, est devenue une des préoccupations des chercheurs dans le domaine de la biocatalyse^{52,53} et pour mieux cerner notre étude sur les effets de solvants et le comportement de l'enzyme, nous avons orienté notre objectif de recherche sur le choix d'autres variables explicatives et cette fois ci nous avons essayé de modéliser le facteur d'enantiosélectivité E en fonction des descripteurs moléculaires afin de mieux expliquer le phénomène relation substrat-enzyme-solvant.

DEUXIEME PARTIE

- ***Chapitre 1: Méthodes d'analyses statistiques utilisées***
- ***Chapitre 2: Influence des solvants organiques sur l'enantiosélectivité de la lipase de Candida antarctica B dans la résolution du menthol***

Le but de la statistique est de dégager les significations de données, numériques ou non, obtenues au cours de l'étude d'un phénomène. Pour cela il faut distinguer les données statistiques qui sont les résultats d'observations recueillies lors de l'étude d'un phénomène, et la méthode statistique qui a pour objet l'étude rationnelle des données. La méthode statistique comporte plusieurs étapes.

1. ANALYSE STATISTIQUE UNIVARIEE

Le but de cette analyse est de caractériser les solvants utilisés en décrivant les différentes caractéristiques ou variables mesurées à l'aide de certains paramètres statistiques de base tels que la moyenne arithmétique (\bar{x} ou \bar{y}), qui est un paramètre de position et de tendance centrale, l'écart-type (s) qui mesure la dispersion des données autour de la moyenne, les valeurs minimales (x_{\min}) et maximales (x_{\max}) qui donnent toutes les deux une idée sur l'étendue des données et enfin, l'effectif (n) qui nous renseigne sur l'importance des données traitées⁵⁴.

Ces paramètres ont été calculés à l'aide du logiciel d'analyse et de traitement statistique des données⁵⁵, sur les valeurs des caractéristiques physico-chimiques des solvants et sur les valeurs des activités enzymatiques aussi bien pour les réactions d'hydrolyse enzymatique que pour celles de transestérification. Ces calculs ont été également effectués sur les valeurs du facteur d'énantiosélectivité E et sur celles l'énergie libre de Gibbs de transition $\Delta\Delta G^\ddagger$ obtenues dans le cas de l'étude de l'influence des différents solvants sur l'énantiosélectivité.

2. ANALYSE STATISTIQUE BIVARIEE

Cette analyse nous permet de voir le degré de la liaison entre les variables prises deux à deux. Nous avons calculé le coefficient de corrélation de « Bravais Pearson » qui est obtenu à l'aide de la formule suivante⁵⁴.

$$r = \frac{cov(x, y)}{s_x s_y} \quad (1)$$

Où,

$Cov(x,y)$: covariance des variables x et y .

s_x et s_y : les écart-types de x et de y .

Ce coefficient est compris entre -1 et +1. Il est en valeur absolue d'autant plus proche de 1 que la liaison entre les deux séries d'observations est nette, pour autant que cette relation soit linéaire ou approximativement linéaire. Au contraire, si le coefficient de corrélation est nul ou presque nul, les deux variables ne sont pas alors corrélées entre elles. D'autre part, le signe du coefficient de corrélation indique si la relation est croissante ou décroissante. En effet, lorsque le coefficient de corrélation est positif, les valeurs élevées d'une variable correspondent dans l'ensemble aux valeurs élevées de l'autre variable, et les valeurs faibles d'une variable correspondent aux valeurs faibles de l'autre variable. Par contre, lorsque la corrélation est négative, les valeurs élevées d'une variable correspondent dans l'ensemble aux valeurs faibles de l'autre variable et vis versa. Les corrélations ont été calculées pour les réactions d'hydrolyse, pour les réactions de transestérification et pour l'énantiosélectivité avec le logiciel MINITAB (X, 2000).

3. ANALYSE STATISTIQUE MULTIVARIEE

3.1. Base théorique de la régression linéaire multiple

L'analyse statistique multivariée a trait essentiellement au calcul de modèles à l'aide de la régression linéaire multiple. C'est une méthode statistique dont les principes sont bien connus et se retrouvent exposés dans la plus part des ouvrages classiques relatifs à la régression^{56,57} ou à la chimiométrie^{58,59}. Aussi nous nous contentons d'une représentation relativement succincte de la méthode afin de faciliter sa compréhension.

La régression linéaire repose sur un modèle qui suppose que les données de la variable à expliquer sont en fonction linéaire des données des variables explicatives :

$$y = b_0 \mathbf{1} + Xb + e \quad (2)$$

où, y est le vecteur colonne ($n \times 1$) des données de la variables à expliquer avec n correspondant au nombre d'observations (nombre de solvants), b_0 est l'ordonnée à l'origine de l'hyperplan de régression, $\mathbf{1}$ est un vecteur colonne ($n \times 1$) des valeurs unité, X est la matrice ($n \times p$) des données des variables explicatives avec p correspondant au nombre de variables explicatives prises en considération, b est le vecteur colonne ($p \times 1$) des coefficients de régression et e est le vecteur colonne ($n \times 1$) des résidus ou erreurs. Une simplification de notation couramment adoptée consiste à inclure l'ordonnée à l'origine b_0 en première position du vecteur b et à augmenter la matrice X à gauche d'une colonne de 1. Le modèle s'écrit alors :

$$y = Xb + e \quad (3)$$

Où, X est désormais une matrice de dimension $[n \times (p+1)]$ et b un vecteur colonne de dimension $[(p+1) \times 1]$. L'estimation \hat{b} du vecteur des paramètres b s'obtient par la méthode des moindres carrés. Si l'on désigne par \hat{y} le vecteur des valeurs estimées de la variable dépendante (ou variable à expliquer) à l'aide du modèle, on a :

$$\hat{y} = X\hat{b} \quad (4)$$

et le vecteur colonne \hat{e} des résidus observés s'obtient par :

$$\hat{e} = y - \hat{y} = y - X\hat{b} \quad (5)$$

L'estimation de b au sens des moindres carrés consiste à calculer les éléments du vecteur \hat{b} de façon à minimiser la somme des carrés des résidus,

$$\hat{e}'\hat{e} = \sum_{i=1}^n \hat{e}_i^2 \quad (6)$$

Où, \hat{e}' est la transposée de \hat{e} .

La solution est donnée par :

$$\hat{b} = (X'X)^{-1} X'y \quad (7)$$

Pour autant que l'inverse de la matrice $X'X$ existe⁶⁰. La transposée de la matrice X est notée X' et son inverse X^{-1} . Sous la condition d'existence citée précédemment, la matrice de variances-covariances du vecteur \hat{b} vaut⁶¹:

$$\sum \hat{b} = \hat{\sigma}^2 (X'X)^{-1} \quad (8)$$

où la variance résiduelle estimée appelée également erreur standard, $\hat{\sigma}^2$ se calcule par :

$$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n - p - 1} \quad (9)$$

La somme des carrés des écarts des résidus, $\hat{e}'\hat{e}$, peut encore s'écrire :

$$\hat{e}'\hat{e} = y'y - \hat{b}X'y \quad (10)$$

Soit la différence entre la somme des carrés des observations de la variable dépendante et la somme des carrés expliqués par la régression. Cette dernière somme des carrés est habituellement décomposée en une somme des carrés liée à la présence du terme indépendant dans le modèle,

$$\frac{\sum_{i=1}^n y_i^2}{n} \quad (11)$$

et en une somme des carrés liée à la présence des autres paramètres,

$$\frac{\hat{b}X'y \sum_{i=1}^n y_i^2}{n} \quad (12)$$

On peut ainsi construire un tableau d'analyse de la variance (tableau 6).

Tableau 6 : Tableau d'analyse de la variance d'une régression linéaire multiple avec un terme indépendant.

Source de variation	dl	SC	MC
Expliquée par la régression	p	$\frac{\hat{b}X'y \sum_{i=1}^n y_i^2}{n}$	$MC_{\text{régr}}$
Résidus	$n - p$	$\hat{e}'\hat{e}$	$\hat{\sigma}^2$
Variation totale	$n - 1$	$\frac{y'y \sum_{i=1}^n y_i^2}{n}$	-

Cette décomposition de la somme des carrés totale permet d'aboutir à la définition du coefficient de détermination R^2 , qui représente la part de la variance de la variable dépendante expliquée par la régression :

$$R^2 = \frac{\hat{b}_X' y - \sum_{i=1}^n y_i^2 / n}{y' y - \sum_{i=1}^n y_i^2 / n} = 1 - \frac{\hat{e}' \hat{e}}{y' y - \sum_{i=1}^n y_i^2 / n} \quad (13)$$

Ce coefficient de détermination R^2 est une estimation biaisée du coefficient théorique correspondant. Une estimation moins biaisée est formé par le coefficient de détermination ajusté, R^2_{adj} , dans la définition du quel on tient compte des nombres de degrés de liberté associés aux différentes sommes des carrés des écarts :

$$R^2_{adj} = 1 - \frac{\hat{e}' \hat{e} / n - p}{\left[y' y - \sum_{i=1}^n y_i^2 / n \right] / (n-1)} = 1 - (1 - R^2) \left(\frac{n-1}{n-p} \right) \quad (14)$$

La recherche de modèle exprimant l'activité enzymatique en fonction des paramètres physico-chimiques des solvants organiques a été effectuée dans le cas des réactions d'hydrolyse enzymatique et dans le cas des réactions de transestérification. Tandis que pour l'énantiosélectivité le calcul des modèles a été fait en exprimant le facteur d'énantiosélectivité E et l'énergie libre de Gibbs de transition $\Delta\Delta G^\ddagger$, en fonction des paramètres physico-chimiques des solvants organiques. Tous les calculs ont été réalisés avec le logiciel MINITAB (version 14.31).

3.2 Méthodes de sélection des variables explicatives

Différente stratégies visant à identifier les variables explicatives à conserver ou à éliminer du modèle ont été proposées afin de ne conserver comme prédicteurs que les variables peu ou pas corrélées entre elles mais corrélées à la variable dépendante. Ces méthodes sont les suivantes :

1- La méthode de toutes les combinaisons possibles, qui consiste à calculer toutes les équations de régression possibles, puis à choisir parmi elles celles qui sont les plus significatives. Cette méthode n'est utilisable, en pratique, que pour un nombre relativement faible de variables explicatives. Pour un grand nombre de variables explicatives, cette méthode nécessite, en effet, un temps de calcul trop important⁵⁴.

Lorsque le nombre de variables explicatives est très élevé, on utilise généralement

l'une des trois techniques suivantes : la sélection progressive, la sélection régressive et la sélection mixte ou régression mixte ou pas à pas « stepwise »⁵⁴.

2- La sélection progressive des variables explicatives consiste à choisir les variables de proche en proche, en assurant chaque fois le minimum de la variance résiduelle jusqu'au moment où l'introduction d'une nouvelle variable ne provoque plus de réduction significative de la variance résiduelle.

3- La sélection régressive des variables explicatives a pour principe de partir de l'équation de régression la plus complexe, dans laquelle figurent toutes les variables explicatives, et de procéder à des éliminations successives jusqu'au moment où tous les coefficients de régression partiels sont significatifs pour un niveau de signification α préalablement fixé.

4- La sélection mixte ou régression pas à pas « stepwise », qui est considérée comme une combinaison des deux précédentes techniques, procède par introductions successives des variables explicatives, mais de telle sorte, qu'avant l'introduction d'une variable supplémentaire, la signification des variables explicatives déjà présentes dans l'équation soit testée. Dans le cas échéant, les variables qui n'apporteraient pas une contribution significative à la régression sont éliminées. C'est la méthode que nous avons utilisée dans le calcul des modèles, (avec $\alpha = 5\%$, tant à l'entrée qu'à la sortie des variables du modèle), du fait que c'est un bon compromis entre les deux techniques précédentes.

3.3 Calcul des modèles

Nous avons vu précédemment que la construction des modèles qui constitue la phase ajustement de la régression linéaire est effectuée par la méthode des moindres carrés. Cependant, cette méthode donne parfois pour la même variable dépendante et pour les mêmes observations plusieurs modèles plus au moins équivalents et ceci pour la même valeur du niveau de signification α .

Aussi pour, d'une part, sélectionner et d'autre part, comparer les modèles entre eux, il faut calculer pour chaque équation de régression, des paramètres statistiques. Ces paramètres sont, en majorité, donnés automatiquement par le logiciel MINITAB et d'autres sont facilement calculés. Ils sont subdivisés en deux groupes. Le premier groupe est relatif aux paramètres concernant la qualité de l'ajustement du modèle et sera étudié dans cette partie alors que le deuxième groupe a trait aux paramètres concernant la qualité de la validation du modèle par la méthode du jackknife, et sera étudié plus en détails, par la suite dans la phase

validation. Les paramètres caractérisant l'ajustement du modèle sont :

- le coefficient de détermination R^2 ,
- le coefficient de détermination ajusté R^2_{adj} ,
- la valeur de la variable F_{obs} de FISHER,
- la valeur de la probabilité P ,
- la valeur de l'écart -type résiduel $s_{y.x}$.

a) Le coefficient de détermination R^2

Afin de se faire une idée sur la qualité de l'ajustement réalisé par la méthode des moindres carrés, on utilise le coefficient de détermination R^2 , cité précédemment. Ce paramètre correspond au carré du coefficient de corrélation. Il est compris entre 0 et 1 et s'exprime toujours en %.

Si la valeur de R^2 est proche de 1 ou 100%, l'ajustement est alors d'excellente qualité. Par contre, si la valeur de R^2 est faible et tend, plus particulièrement vers 0 ou 0%, l'ajustement est mauvais.

b) Le coefficient de détermination ajusté R^2_{adj}

Le R^2_{adj} est utilisé pour comparer des modèles comportant des nombres de termes différents. C'est la valeur de R^2 ajusté en fonction du nombre de degrés de liberté. Si une variable est ajoutée à l'équation, R^2 augmente même si cette variable n'a pas de valeur réelle. Pour compenser cet effet, MINITAB imprime également R^2 ajusté, qui constitue une estimation approximative sans biais de R^2 pour la population. Ce paramètre qui a été défini précédemment est donné par la formule suivante:

$$R^2_{adj} = 1 - \frac{n-1}{n-p} (1 - R^2) \quad (15)$$

Dans cette formule, p est le nombre de coefficients ajustés dans l'équation de régression et n représente le nombre d'observations (ou solvants) utilisées⁶²⁻⁶⁴.

c) La valeur de F_{obs} de FISHER

La valeur de F_{obs} est donnée par le test de signification des coefficients de régression du modèle. Ce test est réalisé par l'analyse de la variance citée ci-dessus. La valeur de F_{obs} est

comparée avec une valeur théorique $F_{1-\alpha}$ tirée à partir des tables de FISHER pour un niveau de signification α donné et pour k_1 et k_2 degrés de liberté. Si la valeur de $F_{\text{obs}} \geq F_{1-\alpha}$, alors les coefficients de régression sont significativement différents de zéro. Dans le cas contraire, si $F_{\text{obs}} < F_{1-\alpha}$, alors le modèle n'est pas significatif et par conséquent il est mauvais.

d) La probabilité P

Une autre façon de tester la signification du modèle est de calculer la probabilité P de mettre en évidence des différences significatives entre les valeurs des coefficients de régression partiels et la valeur zéro. Ensuite, on compare la valeur de cette probabilité P avec le niveau de signification $\alpha = 0,05$. Si la valeur de $P \leq \alpha = 0,05$, alors on dit qu'il existe des différences significatives entre zéro et les valeurs des coefficients de régression, et par conséquent le modèle calculé est significatif. Dans le cas contraire, si $P > \alpha = 0,05$, alors les différences entre zéro et les valeurs des coefficients de régression partiels ne sont pas significatives, et le modèle n'est pas significatif (modèle de mauvaise qualité).

e) L'écart type résiduel ($s_{y.x_i}$)

C'est la racine carrée de la variance résiduelle. Ce paramètre est également appelé **l'erreur d'estimation** ou **erreur de prédiction**, c'est l'erreur que l'on commettrait en utilisant l'équation de régression pour faire des estimations ou des prévisions ou des prédictions.

Ce paramètre constitue une mesure de la dispersion des points observés autour de la droite de régression de y en x . Il correspond à la part de la variance totale de y qui n'est pas "expliquée" ou "justifiée" par la régression. Pour « p » variables explicatives, l'écart type estimé s'écrit :

$$s_{y.x_i} = \sqrt{\frac{1}{n-p-1} \sum_{i=1}^n (y_i - \hat{y}_{(x_i)})^2} \quad (16)$$

Où : - n nombre total d'individus ou de composés

- p nombre de variables explicatives intervenant dans le calcul de $\hat{y}_{(x_i)}$.

Si l'écart résiduel est très faible et tend vers zéro, il s'agit d'une bonne équation de régression qui prédit des valeurs qui se rapprochent des vraies valeurs (réelles). Si, par contre, la valeur de l'écart type résiduel est grande, nous avons alors une très mauvaise équation de

régression qui prédit des valeurs fortes différentes des vraies valeurs observées.

3.4 Validation des modèles calculés

a) Introduction

Après avoir calculé les paramètres d'une équation de régression linéaire simple ou multiple, il est indispensable de vérifier que la solution obtenue répond au problème posé, c'est-à-dire permet de prédire de façon suffisamment précise la valeur prise par la variable dépendante y pour des valeurs données des variables explicatives x . A cette fin, on recourt à diverses techniques de validation. Ces techniques aboutissent au calcul de paramètres statistiques qui mesurent l'exactitude des estimations obtenues à l'aide de l'équation. Ces mêmes paramètres peuvent servir de critères de choix de plusieurs équations ont été calculées pour résoudre un problème donné.

b) Les sources d'erreur de prédiction

La qualité des prédictions réalisées à l'aide d'une équation de régression est influencée par de nombreux phénomènes, qui donnent lieu à plusieurs sources d'erreur. Dans un monde idéal, une équation de régression correctement développée devrait permettre d'estimer sans erreur les valeurs de la variable dépendante y en fonction des valeurs prises par les variables explicatives x et ce, pour l'ensemble de la population concernée.

En pratique, cependant, plusieurs sources d'erreur entrent en jeu et donnent naissance à des écarts entre les valeurs observées et les valeurs prédites de la variable dépendante pour les mêmes objets. Ces sources d'erreur se répartissent en quatre catégories : les erreurs liées à l'équation de régression, les erreurs liées au manque de représentativité du jeu de données de l'ajustement, le bruit aléatoire entachant ces mêmes données et le bruit aléatoire entachant les données du jeu de validation.

c) Les paramètres de mesure de l'erreur de prédiction

Lorsqu'on utilise une équation de régression linéaire pour prédire la valeur de la variable dépendante y , on souhaite que la valeur prédite \hat{y} s'écarte le moins possible de la valeur réelle y pour l'objet en question; c'est-à-dire que la différence $y - \hat{y}$ soit proche de zéro. Cette différence constitue une mesure naturelle de l'aptitude à la prédiction de l'équation de régression pour l'objet considéré.

En pratique, on souhaite chiffrer l'exactitude des prédictions obtenues à l'aide d'une équation de régression pour l'ensemble de la population des objets à laquelle elle est susceptible d'être appliquée.

Le paramètre le plus couramment employé à cette fin est **l'erreur quadratique moyenne** (Mean Square Error, MSE), ou espérance mathématique des carrés des écarts entre les valeurs réelles de y_i et les valeurs prédites \hat{y}_i correspondantes pour l'ensemble des objets.

$$MSE = E(y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (17)$$

Afin de disposer d'un paramètre exprimé dans les mêmes unités que la variable dépendante, on considère souvent sa racine carrée, notée RMSE (Root Mean Square Error).

Faute de pouvoir mesurer la valeur de la variable dépendante pour l'ensemble des individus de la population, ce qui rendrait d'ailleurs inutile l'établissement d'une équation de régression, l'erreur quadratique moyenne doit être estimée. Dans les situations les plus simples, il est possible de calculer une situation de ce paramètre à partir des formules issues de la théorie des modèles linéaires et d'informations concernant la distribution théorique des erreurs⁵⁶.

Cependant, dans la plupart des cas il n'est pas toujours possible de suivre cette approche et on estime alors, l'erreur quadratique moyenne à partir d'un nombre limité d'objets appartenant à un jeu de données de validation réservées spécialement à cet usage (validation externe) ou au jeu de données d'ajustement lui-même (validation interne).

La validation externe

La validation externe permet d'estimer l'erreur quadratique moyenne à partir d'un échantillon d'objets prélevés dans la population à laquelle l'équation de régression s'applique mais différents des objets faisant partie de l'échantillon d'ajustement. Soit n_v , l'effectif de l'échantillon de validation.

L'erreur quadratique moyenne calculée à partir de cet échantillon porte le nom **d'erreur quadratique moyenne de prédiction** (Mean Square Error of Prediction : MSEP) et se calcule simplement par:

$$MSEP = \frac{1}{n_v} \sum_{i=1}^{n_v} (Y_i - \hat{Y}_i)^2 \quad (18)$$

Sa racine carrée se note RMSEP ou parfois RMSP (Root Mean Square Error of Prediction). A partir des mêmes individus, on peut estimer la variance de prédiction pour toute la population :

$$\hat{\sigma}_p^2 = \frac{1}{n_v - 1} \sum_{i=1}^{n_v} (\hat{y}_i - \bar{y})^2 \quad (19)$$

Où ;
$$\bar{y} = \frac{1}{n_v} \sum_{i=1}^{n_v} \hat{y}_i$$

Sa racine carrée est **l'écart-type de prédiction** (SEP), appelée aussi fréquemment **écart-type de validation (externe)** (SEV).

Pour autant qu'elle s'appuie sur un échantillon de validation d'effectif élevé et constitué d'objets représentatifs de l'ensemble de la population, la validation externe fournit une estimation fiable de l'erreur quadratique moyenne.

La validation interne

Par opposition à la validation externe, la validation interne emploie les données du jeu d'ajustement pour procéder à la validation.

Une première façon de procéder consiste à estimer l'erreur quadratique moyenne à l'aide de l'ensemble des données du jeu d'ajustement. On définit ainsi **l'erreur quadratique moyenne d'ajustement**, aussi appelée **erreur quadratique moyenne d'estimation** par :

$$MSEE = \frac{1}{n_a - dl} \sum_{i=1}^{n_a} (\hat{y}_i - \bar{y})^2 \quad (20)$$

Dans cette expression, n_a représente l'effectif de l'échantillon d'ajustement et dl le nombre de degrés de liberté du modèle. Ce nombre de degrés de liberté est tout simplement le nombre de coefficients estimés lors du calcul de l'équation de régression.

Dans le cas de la régression linéaire multiple cette erreur quadratique moyenne d'ajustement est un estimateur de la variance résiduelle du modèle. Cependant, l'erreur quadratique moyenne d'ajustement sous-estime l'erreur quadratique moyenne du modèle,

puisqu'elle ne prend en compte que la variance résiduelle, négligeant complètement la variance des coefficients de régression. Dès lors, il n'est pas recommandable d'employer ce paramètre ou sa racine carrée pour la validation d'un modèle.

Une seconde façon de valider une équation de régression linéaire à partir des données du jeu d'ajustement repose sur le mécanisme de la validation croisée⁶⁵. Cette méthode de validation permet de valider le modèle d'ajustement à l'aide d'un jeu de données indépendant de celui ayant servi à estimer les coefficients du modèle, tout comme la validation externe, mais contrairement à celle-ci, elle évite de devoir sacrifier une part non négligeable des données disponibles à la seule validation.

Cette méthode appelée **validation croisée complète** (full cross-validation) ou méthode du jakknife consiste à estimer les coefficients de l'équation d'ajustement à partir du jeu de données d'ajustement dont on enlève l'observation i et à calculer, pour cette observation, le résidu de prédiction $y_i - \hat{y}_i$. Ce résidu n'est rien d'autre que la différence observée entre la valeur de la variable dépendante pour l'observation i (y_i) et la valeur prédite, pour cette même observation, à l'aide de l'équation d'ajustement établie en enlevant l'observation i du jeu de données d'ajustement (\hat{y}_i). On répète cette démarche pour chaque observation du jeu de données d'ajustement et on calcule ensuite **la somme des carrés des erreurs de prédiction** (Predicted Residual Error Sum of Squares, PRESS)⁶⁶,

$$PRESS = \sum_{i=1}^{n_a} (y_i - \hat{y}_i)^2 \quad (21)$$

en divisant cette quantité par l'effectif n_a du jeu de données d'ajustement, on aboutit à **l'erreur quadratique moyenne de validation croisée** (Mean Square Error of cross-validation, MSECv) :

$$MSECv = \frac{1}{n_a} PRESS \quad (22)$$

qui est une estimation de l'erreur quadratique moyenne théorique définie précédemment. On considère souvent sa racine carrée afin de disposer d'un paramètre de validation exprimé dans les mêmes unités que la variable dépendante, et qu'on note :

$$RMSECV = \sqrt{\frac{PRESS}{n_a}} = \left(\frac{PRESS}{n_a}\right)^{1/2} \quad (23)$$

Enfin, la validation de l'équation de régression peut aussi reposer sur la technique du **bootstrap**⁶⁷. Cette technique, consiste à simuler m (généralement supérieur à 30) échantillons de même taille n que l'échantillon initial. Ils sont obtenus par tirage au hasard avec remise parmi les n individus observés au départ, ceux-ci ayant tous la même probabilité $1/n$ d'être choisis. Certains individus auront de ce fait un poids élevé (2,3,...) alors que d'autres seront absents (poids nul).

Le bootstrap n'est rien d'autre qu'une technique de simulation particulière, fondée sur la distribution empirique de l'échantillon de base.

Cette méthode est employée pour analyser la variabilité de paramètres statistiques simples en produisant des intervalles de confiance de ces paramètres.

Elle donne dans la plupart des cas une bonne image de la précision statistique de l'estimation de l'échantillon⁶⁸.

En ce qui nous concerne, nous avons pris en considération comme paramètres statistiques pour juger de la qualité de la validation interne des modèles :

- le PRESS statistique,
- l'erreur quadratique moyenne de la validation croisée (méthode jackknife) (RMSECV),
- et le coefficient de détermination de prédiction $R^2_{\text{préd}}$ obtenu simplement par :

$$R^2_{\text{préd}} = 1 - \frac{PRESS}{\sum_{i=1}^{n_a} (y_i - \bar{y})^2} \quad (24)$$

Les valeurs de ces paramètres sont fournies automatiquement par le logiciel MINITAB⁵⁵.

Ce chapitre est consacré à l'étude de l'influence du solvant sur l'énantiosélectivité lors de réaction de transestérification catalysée par une lipase.

Un substrat racémique S_{RS} est constitué de deux énantiomères S_R et S_S en quantité égale. La résolution cinétique enzymatique est un processus dans lequel l'un des énantiomères de ce mélange racémique réagit plus rapidement avec l'entité chirale, que constitue l'enzyme Ez, que l'autre énantiomère.

La différence dans les vitesses de réaction des deux énantiomères S_R et S_S , découle de la différence d'énergie libre des états de transition ($\Delta\Delta G^\ddagger$) des deux complexes enzyme-substrat, $[Ez-S_R]^*$ et $[Ez-S_S]^*$, comme le montre le schéma suivant :

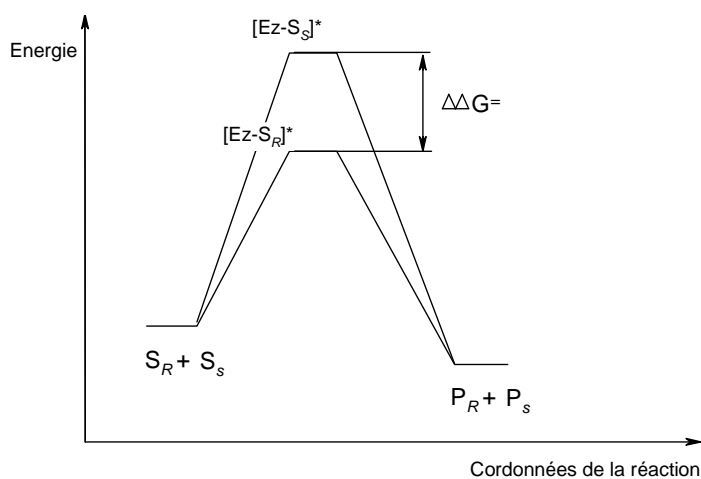


Schéma 16 : différence d'énergie libre des états de transition des deux complexes.

Pour rappel, l'énantiosélectivité est quantifiée par le facteur d'énantiosélectivité E qui est le rapport des constantes de spécificité de chacun des deux énantiomères. Il est également lié à $\Delta\Delta G^\ddagger$ par la relation : $\Delta\Delta G^\ddagger = -RT \cdot \ln E$.

De nombreux travaux, consacrés à l'étude de l'effet de solvant sur l'énantiosélectivité des lipases, ont montré que celle-ci est affectée par le choix du solvant utilisé^{3,69,70}.

Parmi les hypothèses émises pour expliquer les mécanismes par lesquels les solvants influent sur l'énantiosélectivité de l'enzyme, citons : (i) la modification des interactions enzyme-substrat lorsque le solvant pénètre dans le site actif, (ii) la modification de la conformation de la protéine et de sa flexibilité, changeant ainsi le processus de reconnaissance moléculaire et (iii) orientation des substrats dans le site catalytique de l'enzyme⁷¹.

A travers la littérature citée précédemment, il ressort que les recherches de corrélations entre l'énantiosélectivité et la nature du milieu réactionnel, ne prennent en compte qu'un seul paramètre physico-chimique du solvant à la fois, en général $LogP$, alors que l'association d'au moins un autre indicateur physico-chimique pourrait, dans certains cas, aboutir à une meilleure corrélation, comme l'ont montré Valivety et col⁷², dans leur étude sur les effets des solvants dans la biocatalyse en milieu organique. En associant $LogP$ et un autre indicateur de polarité des solvants, ils aboutissent à une meilleure corrélation entre l'énantiosélectivité de l'enzyme et la nature du solvant. Cette corrélation a été limitée à regrouper seulement des descripteurs physico-chimique et d'autre part $LogP$ n'a jamais été corrélé aux descripteurs moléculaires qui peuvent contribuer à l'explication de la relation « solvant-énantiosélectivité ».

C'est dans ce contexte que nous nous sommes proposés d'étudier l'influence des solvants organiques sur l'énantiosélectivité, en cherchant à modéliser la relation « nature du solvant-énantiosélectivité » par une analyse statistique multivariée.

Pour ce faire, nous avons choisi d'étudier la résolution cinétique du (\pm)-menthol par transestérification enzymatique.

Cette réaction, catalysée par la lipase de *Candida antarctica* B, est réalisée dans 19 solvants organiques différents :

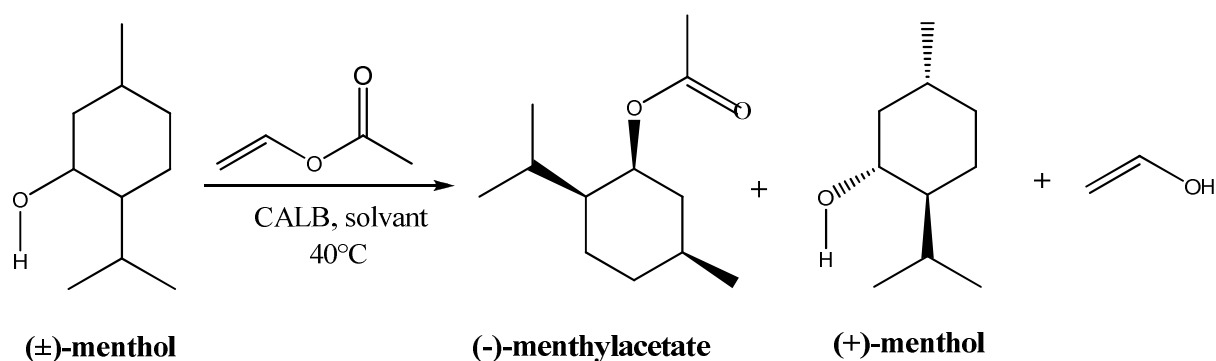


Schéma 17 : transestérification du menthol par la CAL B en présence de solvants organiques.

Dans tous les solvants utilisés où un facteur d'énantiosélectivité a été décelé, c'est le (-)-menthol qui est préférentiellement transestérifié par la CAL B pour former le (-)-menthylacétate.

1. COLLECTE DES DONNEES

La collecte des données a consisté en la détermination des excès énantiomériques (ee) et le calcul du facteur d'énantiosélectivité E et du taux de conversion C , à partir des formules suivantes (voir chapitre I) :

$$E = \frac{\ln [(1 - C)(1 - ee_s)]}{\ln [(1 - C)(1 + ee_s)]}$$

Avec :

$$C = \frac{ee_s}{ee_s + ee_p}$$

Les excès énantiomériques du substrat (ee_s) (menthol racémique) et du produit, ((-)-menthylacetate) (ee_p), sont déterminés par chromatographie en phase gazeuse utilisant une colonne chirale.

Le tableau 7 ci-dessous, résume l'ensemble de ces données pour chaque solvant utilisé :

Tableau 7 : Résolution cinétique du (\pm)-menthol par transestérification par la CAL B. Valeurs du facteur d'énantiosélectivité E , des excès énantiomériques et du taux de conversion.

N°	Solvant	E	ee _s % (+)-menthol	ee _p % (-)-menthylacetate	Conversion %
1	DMF	00,00	00	00	00
2	Acétonitrile	03,86	43	24	64
3	1,4-dioxane	09,00	62	32	66
4	Acétone	05,93	25	17	60
5	2-butanone	08,90	52	33	61
6	THF	06,08	61	33	65
7	Diethylether	03,95	65	26	71
8	Dichloromethane	07,72	27	43	38
9	Chloroforme	12,43	67	23	74
10	Benzene	11,60	38	30	56
11	1-chlorobutane	10,82	60	32	65
12	Toluene	09,20	34	36	49
13	Tetrachloromethane	11,06	60	62	49
14	Chlorobenzene	11,14	61	35	63
15	Dibutylether	12,92	06	36	15
16	Cyclohexane	15,62	76	45	63
17	Heptane	08,22	74	36	67
18	Octane	06,18	45	32	58
19	Nonane	14,08	67	25	73

La récolte des descripteurs est réalisée à partir d'une base de données en ligne utilisant le serveur E-DRAGON1.0, c'est un Laboratoire virtuel de chimie computationnelle⁷³.

E-DRAGON est la version électronique à distance du logiciel bien connu dragon, qui est une application pour le calcul de descripteurs moléculaires développés par le groupe (Milano Chemometrics and QSAR Research). Ces descripteurs peuvent être utilisés pour évaluer la structure moléculaire-activité ou structure-propriété.

DRAGON fournit plus de 1600 descripteurs moléculaires. L'utilisateur peut calculer non seulement le type d'atome le plus simple, d'un groupe fonctionnel et du fragment, mais aussi plusieurs descripteurs topologiques et géométriques. Ce logiciel nous a permis de calculer pour chaque solvant plus de 1400 descripteurs ces derniers ont été triés et filtrés en utilisant l'analyse statistique descriptive de base qui donne la matrice de corrélation. Celle-ci nous a permis de choisir les descripteurs les mieux corrélés (facteur de corrélation r proche de 1). Les descripteurs employés pour prédire l'énantiosélectivité E ainsi que leur type de catégorie, sont résumés dans le tableau 8.

Tableau 8 : Type et symbole des descripteurs moléculaires.

Symbole	Type des descripteurs moléculaires
LogP	Propriété moléculaire
XMOD	Connectivité
Mor01m	3D-MoRSE
ww	Topologique
S0K	Topologique
QXXm	Géométrique
QYYm	Géométrique

Les sept descripteurs qui ont été choisis à partir de la matrice de corrélation pour l'analyse et la modélisation de la réponse E , sont rassemblés avec leurs valeurs dans le tableau 9.

Tableau 9 : Variables explicatives des solvants utilisés.

N°	Solvant	Variables explicatives						
		LogP	XMOD	Mor01m	ww	S0K	QXXm	QYYm
1	DMF	-01,01	15,11	15,44	28,00	09,61	06,82	09,46
2	Acetonitrile	-00,34	8,84	02,16	2,33	05,00	06,71	06,71
3	1,4-dioxane	-00,27	20,0	23,11	168,00	05,51	08,90	09,30
4	Acetone	-00,24	10,97	04,03	04,00	00,75	09,91	02,52
5	2-butanone	00,29	14,20	05,05	05,67	01,36	16,19	02,23
6	THF	00,46	16,00	15,11	80,00	07,61	06,04	06,60
7	Diethylether	00,89	15,49	16,13	35,00	07,61	02,34	19,32
8	Dichloromethane	01,25	16,26	15,78	05,00	02,76	04,08	12,93
9	Chloroforme	01,97	19,92	35,82	12,00	03,25	14,38	14,39
10	Benzene	02,13	18,00	18,13	168,00	00,00	07,39	14,78
11	1-chlorobutane	02,64	18,37	4,29	5,43	16,00	44,54	08,70
12	Toluene	02,73	20,36	25,91	251,00	12,90	07,53	16,68
13	Tetrachloromethane	02,83	23,00	64,08	22,00	03,61	25,59	25,59
14	Chlorobenzene	02,84	23,54	36,55	251,00	15,65	07,40	34,64
15	Dibutylether	03,21	27,49	53,87	330,00	20,53	04,53	111,86
16	Cyclohexane	03,44	18,00	21,52	168,00	00,00	09,73	09,73
17	Heptane	04,57	20,49	31,27	126,00	13,65	03,62	55,60
18	Octane	05,18	23,49	41,19	210,00	16,00	04,07	82,40
19	Nonane	05,24	26,49	52,48	330,00	20,53	04,67	116,41

NB: **Log P** : coefficient de partage eau/octanol, **XMOD**: indice de connectivité de RANDIC modifié, **Mor01m**: 3D-MoRSE - signal 01, **ww**: hyper-detour index , **S0K**: indice de symétrie de Kier, **QXXm** : valeur de Qxx COMMA2 / pesée par la masse atomique , **QYYm**: valeur de Qyy COMMA2 value / pesée par la masse atomique.

2. ANALYSE ET RESULTATS STATISTIQUES

2.1. Analyse statistique univariée

Les résultats de cette analyse, obtenus dans le cas de la résolution de (\pm)-menthol par transestérification, en utilisant E comme variable à expliquer et les sept variables explicatives (tableau 7 et 9), sont donnés par le tableau 10.

Tableau 10 : Paramètres statistiques de base déterminés pour les variables utilisés.

Variables	Effectifs (n)	Moyennes (\bar{x} ou \bar{y})	Ecart type (s)	$x_{\min} \sim x_{\max}$
Log $P(x_1)$	19	01,99	01,90	-01,01 ~ 05,24
XMOD (x_2)	19	18,74	04,87	08,84~ 27,48
Mor01m (x_3)	19	25,36	18,06	02,15~ 64,08
ww (x_4)	19	115,9	116,1	02,30~ 330,00
SOK (x_5)	19	08,54	06,92	00,00 ~ 20,53
QXXm (x_6)	19	10,23	09,94	02,34 ~ 44,53
QYYm (x_7)	19	29,46	35,74	02,23 ~ 116,41
$E(y)$	19	08,90	03,92	00,00 ~ 15,62

2.2. Analyse statistique bivariée

A partir des matrices de données de dimensions 19 x 8 (19 solvants organiques testés et 8 variables quantitatives) dont 7 sont des paramètres physico-chimiques des solvants et une variable à expliquer qui est : $E(y)$ obtenu pour l'alcool racémique, nous avons calculé les corrélations deux à deux entre ces variables. Les résultats obtenus figurent dans le tableau 11.

Tableau 11 : Matrice de corrélations entre les 8 variables deux à deux pour les dix-neuf solvants.

	LogP (x_1)	XMOD (x_2)	Mor01m (x_3)	ww (x_4)	S0K (x_5)	QXXm (x_6)	QYYm (x_7)
XMOD (x_2)	0,773 (0,000)***						
Mor01m (x_3)	0,672 (0,002)**	0,882 (0,000)***					
ww (x_4)	0,632 (0,004)**	0,769 (0,000)***	0,590 (0,008)**				
S0K (x_5)	0,572 (0,011)*	0,670 (0,002)**	0,467 (0,044)*	0,643 (0,003)**			
QXXm (x_6)	-0,001 (0,996)	-0,021 0,933	-0,113 0,644	-0,393 0,096	-0,033 0,894		
QYYm (x_7)	0,729 (0,000)***	0,778 (0,000)***	0,727 (0,000)***	0,746 (0,000)***	0,770 (0,000)***	-0,300 0,213	
E (y)	0,612 (0,005)**	0,612 (0,005)**	0,488 (0,034)*	0,499 (0,030)*	0,109 0,656	0,245 0,341	0,311 0,154

Les valeurs entre parenthèses représentent les probabilités du test de signification des coefficients de corrélation. * : corrélation significative au niveau $\alpha = 0,05$ ($P \leq 0,05$). **: corrélation hautement significative au niveau $\alpha = 0,01$ ($P \leq 0,01$). *** : corrélation très hautement significative au niveau $\alpha = 0,001$ ($P \leq 0,001$).

Le tableau 11 montre que E (y) est corrélé positivement avec $LogP$, XMOD, Mor01 et ww et que les trois autres variables explicatives ne présentent aucune corrélation avec la réponse. On constate aussi que $LogP$ (x_1) est corrélé avec la plupart des paramètres explicatives.

2.3. Analyse statistique multivariée

La stepwise a sélectionné deux modèles significatifs pour E (y) (tableau 12), après suppression de trois valeurs aberrantes qui sont : octane, heptane et 1,4-dioxane.

Tableau 12 : Modèles calculés par la stepwise pour la réponse $E(y)$.

N°	Effectifs des solvants (n)	Modèle	Paramètres statistiques							
			De l'ajustement du modèle					validation interne du modèle		
			R ² %	R ² _{adj} %	F _{obs}	P	s _{y,x}	R ² _{pred} %	PRESS	$\bar{\epsilon}_q$ (MSE)
1	16	$E(y) = 5,2 + 2,2 \text{ Log}P$	77	75	47	0,000***	2,1	67,04	88,4	4,41
2	16	$E(y) = 6 + 2,7 \text{ Log}P - 0,2 \text{ SOK}$	86	84	40	0,000***	1,7	78,34	58,1	2,88

NB : * : modèle significatif au niveau $\alpha = 0,05 (P \leq 0,05)$. ** : modèle hautement significatif au niveau $\alpha = 0,01 (P \leq 0,01)$. *** : modèle très hautement significatif au niveau $\alpha = 0,001 (P \leq 0,001)$.

Comme le montre ce tableau, des deux modèles sélectionnés, celui dont l'expression associe les deux indicateurs $\text{Log}P$ et l'indice de symétrie de *Kier* SOK, (modèle n°2) est meilleur en termes de qualité d'ajustement et de prédiction, que celui exprimé uniquement avec $\text{Log}P$ (modèle n°1). A partir de ce résultat nous avons effectué la validation de ce modèle.

• Validation du modèle n°2

La validation du meilleur modèle calculé sur l'effectif total des solvants, à savoir le modèle n°2 ($E(y) = 6 + 2,7 \text{ Log}P - 0,2 \text{ SOK}$), a consisté à partager aléatoirement les seize données utilisées pour l'ajustement en deux groupes, de dix solvants qui constitueront l'effectif testé et six solvants formeront l'effectif de validation. Les résultats de cette étude sont rassemblés dans les tableaux 13 et 14.

L'application de la stepwise sur l'effectif testé, a donné deux modèles significatifs. Ces modèles sont résumés avec leurs paramètres statistiques dans le tableau 13.

Tableau 13 : Modèles calculés par la validation de l'effectif d'apprentissage, pour la réponse $E(y)$.

N°	Effectif des solvants testés (n)	Modèle	Paramètres statistiques							
			De l'ajustement du modèle					validation interne du modèle		
			R ² %	R ² _{adj} %	F _{obs}	P	s _{y,x}	R ² _{pred} %	PRESS	\bar{e}_q
1	10	$E(y) = 5,25 + 2,65 \text{ Log}P$	86	85	50,6	0,000***	1,87	78	45	3,5
2	10	$E(y) = 6,5 + 3,1 \text{ Log}P - 0,22 \text{ SOK}$	95	94	72	0,000**	1,15	85	30	1,3

NB : * : modèle significatif au niveau $\alpha = 0,05(P \leq 0,05)$. ** : modèle hautement significatif au niveau $\alpha = 0,01(P \leq 0,01)$. *** : modèle très hautement significatif au niveau $\alpha = 0,001(P \leq 0,001)$.

L'examen du tableau 13, montre que les deux modèles calculés pour l'effectif des dix solvants, présentent une meilleure qualité d'ajustement. Seulement on remarque que la qualité de prédiction est nettement meilleure dans le cas du modèle n°2.

Comme on peut le constater, ce modèle s'exprime avec les mêmes indicateurs de solvants que le meilleur modèle calculé avec l'effectif total des solvants (modèle n°2 du tableau 12).

Nous avons par la suite, procédé à la validation de ce modèle sur la série des six solvants. Le tableau 14, ci-dessous, résume les paramètres statistiques de cette validation

Tableau 14 : Validation du modèle n°2 calculé pour l'effectif d'apprentissage, afin de prédire la réponse E .

Modèle	Equation	R ² %	Moyennes arithmétiques de y_{est} et y_{obs}		Moyennes arithmétiques (\bar{e}_a) et quadratiques (\bar{e}_q) des résidus	
			\bar{y}	\bar{y}_{est}	\bar{e}_a	\bar{e}_q
M calculé sur 10 et validé sur 6	$y_M = - 1,93 + 0,93 \hat{y}_M$	93	10	12,4	2,8	2,9

Les résultats de cette validation confirment que le modèle n°2 du tableau 13, calculé sur l'effectif de dix solvants, est le meilleur du point de vue qualité d'ajustement et de prédiction.

L'équation exprimant ce modèle ainsi que ses paramètres statistiques sont résumés dans le tableau 15 ci-dessous :

Tableau 15 : Modèle calculé et validé pour la réponse E (y).

Modèle	Paramètres statistiques							
	De l'ajustement du modèle					validation interne du modèle		
	$R^2\%$	$R^2_{adj}\%$	F_{obs}	P	$s_{y,x}$	$R^2_{pred}\%$	PRESS	\bar{e}_q
$E(y) = 6,5 + 3,1 \text{ Log}P - 0,22 \text{ SOK}$	95	94	72	0,000**	1,15	85	30	1,3

NB : * : modèle significatif au niveau $\alpha = 0,05$ ($P \leq 0,05$). ** : modèle hautement significatif au niveau $\alpha = 0,01$ ($P \leq 0,01$). *** : modèle très hautement significatif au niveau $\alpha = 0,001$ ($P \leq 0,001$).

Comme on peut le constater, ce modèle possède d'excellents paramètres statistiques tant en ajustement qu'en prédiction, et s'exprime à travers deux paramètres : un paramètre physico-chimique des solvants, $\text{Log}P$ et c'est aussi un descripteur moléculaire expérimentale qui caractérise la polarité et l'hydrophobicité du solvant et un descripteur moléculaire théorique, SOK , qui est un indice topologique de symétrie obtenu à partir de la structure 2D de la molécule, et donnent des informations sur sa taille et sa forme globale. Cet indice a été décrit dans la littérature, comme descripteur stérique^{74,75}. Ce facteur qui a été décrit comme paramètre influant l'état de transition du complexe enzyme-substrat dans la résolution cinétique des énantiomères en présence de solvants organiques⁷⁶.

Ce modèle montre que l'énantiosélectivité de la CAL B dans la résolution du menthol, est effectivement influencée par la nature des solvants organiques. Le facteur d'énantiosélectivité E serait favorisé par des solvants de valeurs de $\text{Log}P$ croissantes et des volumes moléculaires décroissants. Cependant, étant donnée les valeurs respectives des coefficients de régression (3,1 pour $\text{Log}P$ et 0,22 SOK), l'hydrophobicité du solvant est la plus influente.

1. Matériels :

1.1. Chromatographie en phase gazeuse (CPG)

Nous avons employé cette technique d'analyse afin d'identifier les produits, vérifier leur pureté, suivre la cinétique des réactions, et calculer les paramètres d'énantiosélectivité des composés chiraux. L'appareil utilisé est un SHIMADZU GC- 17. Colonne capillaire chirale Beta-dex 325 (30m x 0.25mm x 0.25 μ m) ; détecteur FID ; température maximale 300°C). Injecteur en mode Split.

1. 2. Produits chimiques et enzymes

Les composés racémiques ainsi que les autres produits et solvants utilisées dans l'étude sont d'origine commerciale (FLUKA, SIGMA ALDRICH). La Novozym 435[®] est une préparation industrielle produite par Novozyme Corporation et commercialisée par Sigma, résultant de l'immobilisation de la lipase immobilisée de *Candida antarctica* B sur une résine acrylique macroporeuse. Elle est fournie avec une activité de 10000 U/g.

2. Méthodes

Dans un flacon fermant hermétiquement, une solution d'acétate de vinyle (20mM) et (\pm)-menthol (10mM) est agitée dans 10 ml de solvant anhydre en présence de CAL B (0.5g) et le tamis moléculaire. Le mélange réactionnel est ensuite porté en agitation à 40°C. La cinétique de la réaction est suivie par chromatographie en phase gazeuse (SHIMADZU GC-17A à Colonne capillaire chirale β -Dex 325), par injection à intervalle régulier de prélèvements de 1 μ l du mélange réactionnel.

3. Techniques générales

a) Analyse qualitative par chromatographie sur couche mince

Toutes les réactions mises en œuvre dans cette étude ont tout d'abord été analysées qualitativement par chromatographie sur couche mince. Les plaques de silice utilisées sont de dimensions 20 x 20 cm (Silica gel 60, Merck). Des dépôts de quelque μ L de prélèvements réactionnels sont effectués. Le type d'éluant utilisé est un mélange de l'éther de pétrole/Ether diéthylique (90/10).

En fonction de leur structure chimique, les composés sont révélés selon l'une des procédures suivantes

- pulvérisation d'une solution aqueuse de permanganate de potassium réagissant de façon non spécifique avec tous les composés réducteurs et conduisant à l'apparition d'une coloration jaune.
- exposition à des vapeurs saturées en iode (révélateur universel).

b) Séparation et analyse quantitative par chromatographie gazeuse

Les conditions d'analyse GC utilisées pour identifier le produit (\pm)-Acétate de menthyle et séparer ces énantiomères (-) et (+) ainsi que ceux des deux énantiomères du menthol sont :

- température initiale : 120°C
- temps initiale : 12 min
- température finale : 200°C
- temps finale : 30 min
- vitesse : 6°C/min

Les temps de rétention représenté dans le chromatogramme ci-dessous pour les deux énantiomères de l'ester racémique (\pm)-Acétate de menthyle et ces énantiomères (-) et (+) ainsi que ceux des deux énantiomères du menthol ont été déduites par rapport à des référence optiquement pures :

(+)-menthol, $t_R = 12.92$ min. // (-)-menthol, $t_R = 13.03$ min.

(-)-acétate de menthyle, $t_R = 14.50$ min. // (+)-acétate de menthyle, $t_R = 14.64$ min.

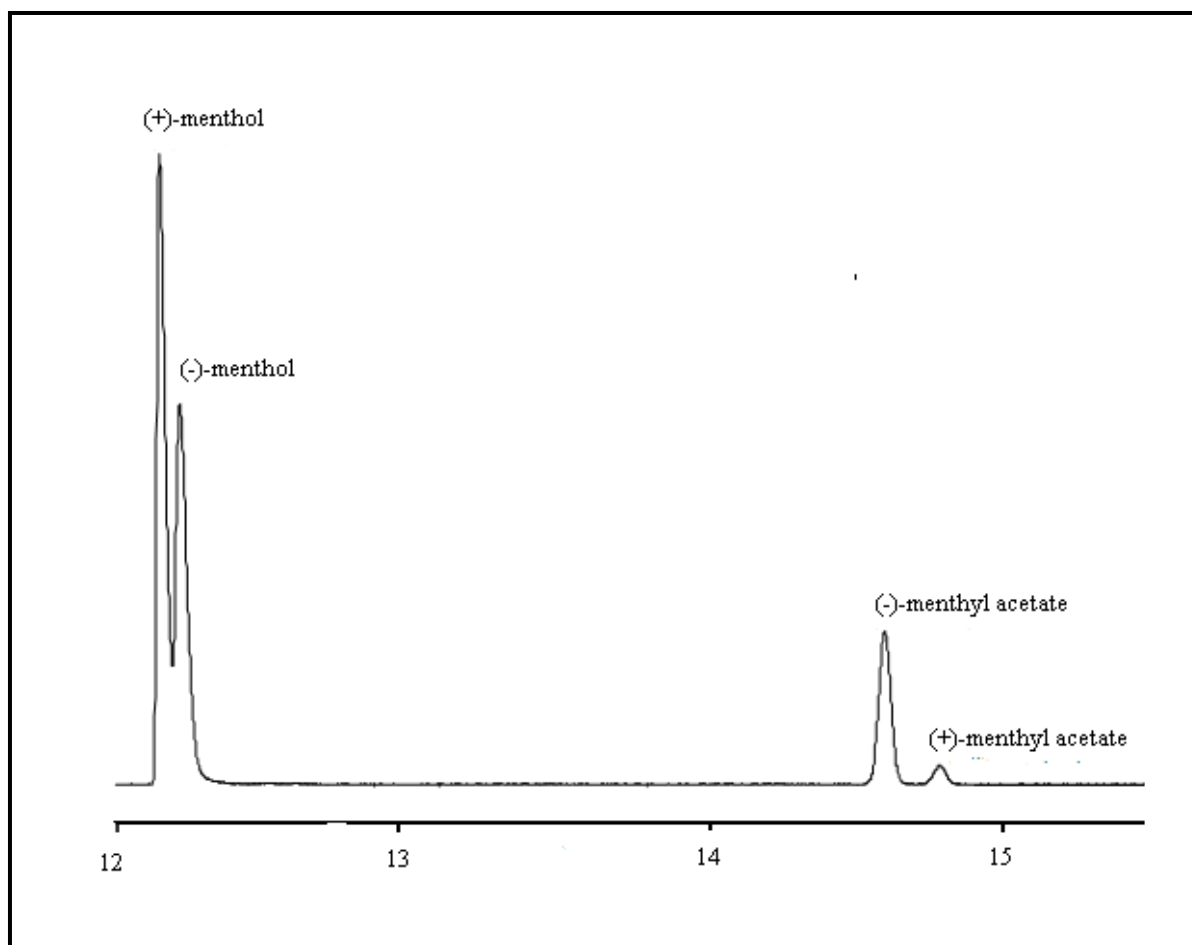


Schéma 17 : chromatogramme GC de la réaction.

4. Détermination des excès énantiomériques et de l'énantiosélectivité

4.1 Détermination des excès énantiomériques

A partir des résultats de la CG sur colonne chirale, les excès énantiomériques du substrat (ee_s) et les excès énantiomériques du produit (ee_p) sont calculés à partir des équations (1) et (2) ci-dessous. Le taux de conversion (C) de la réaction est calculé à partir de l'équation (3).

$$\% ee_s = \frac{(R-S)}{(R+S)} \times 100 \quad (1)$$

$$\% ee_p = \frac{(S-R)}{(S+R)} \times 100 \quad (2)$$

$$\% C = \frac{ee_s}{ee_s + ee_p} \quad (3)$$

4.2 Détermination de l'énantiosélectivité E

L'énantiosélectivité représente la capacité de l'enzyme à différencier les énantiomères (R) et (S). Dans cette étude, L'énantiosélectivité (E) a été calculé selon la méthode des excès énantiomériques.

$$E = \frac{\ln[(1 - C)(1 - ee_s)]}{\ln[(1 - C)(1 + ee_s)]}$$

Ou

$$E = \frac{\ln [(1 - C)(1 + ee_p)]}{\ln [(1 - C)(1 - ee_p)]}$$

Conclusion générale

Au cours de ce travail, nous avons cherché à connaître l'effet de solvants sur l'énantiosélectivité des enzymes. Etant donné la complexité du phénomène, à savoir les nombreuses interactions entre les différents acteurs du milieu réactionnel, nous avons fait appel à une approche multiparamétrique.

Nous avons pris comme réaction modèle, la transestérification énantiosélective du menthol par l'acétate de vinyle en présence de la lipase de *Candida antarctica* B (CAL B), que nous avons réalisé dans différents solvants et mesuré le facteur d'énantiosélectivité E pour chacun d'eux.

Le traitement des résultats obtenus a été effectué par régression linéaire multiple en utilisant un logiciel spécifique en prenant le facteur E comme variable à expliquer (réponse) et les paramètres physico-chimiques et moléculaires des solvants employés comme variables explicatives.

Le modèle obtenu montre que l'énantiosélectivité de la CAL B dans la résolution du menthol est effectivement influencée par la nature des solvants organiques. Le facteur d'énantiosélectivité E serait favorisé par des solvants dont des valeurs de $\text{Log}P$ croissantes et des volumes moléculaires décroissants. En d'autres termes, les solvants hydrophobes de faible volume moléculaire favoriseraient l'énantiosélectivité de la CAL B lors de la résolution cinétique du menthol. L'explication pouvant corroborer ces observations, pourrait provenir des interactions hydrophobes que réaliserait le solvant avec la partie hydrophobe de l'enzyme proche du site actif. L'interaction favorable du solvant avec la poche enzymatique conjuguée à un faible volume du solvant favoriserait l'énantiosélectivité de la protéine.

Pour conforter notre conclusion, l'étude d'autres lipases et d'autres substrats est nécessaire.

- [1] Faber.K. *Biotransformations in Organic Chemistry*, 4th Edition, Ed.; Springer-Verlag Berlin New York, **2000**, 5 p.
- [2] V. Gotor-Fernandez, R. Brieva, V. Gotor. *Molecular Catalysis B: Enzymatic*. **2006**, 40 111-120.
- [3] A. Ghanem. *Tetrahedron*. **2007**, 63, 1721-1754.
- [4] R.D. Schmid, R. Verger. *Angew. Chem. Int. Ed.*, 37, **1998**, 1608-1633.
- [5] A. Zacks, A.M. Klivanov. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 82, 3192-3196.
- [6] C. S. Chen, Y. Fujimoto, G. Girdaukas, C. J. Sih. *J. Amer. Chem. Soc.* **1982**, 104, 7294-7299.
- [7] C. J. Sih, S.Wu, Y. Fujimoto. *J. Amer. Chem. Soc.* **1987**, 128, 43.
- [8] J. Ottosson, J.C. Rotticci-Mulder, D. Rotticci, K. Hult. *Protein Sci.* **2001**, 10, 1769-1774.
- [9] F. Hasan, A. A. Shah, A. Hameed. *Enz. Microb. Technol.* **2006**, 39, 235-251.
- [10] D. H. Zhang, S. Bai, Y. Sun. *Food Chem.* **2007**, 102, 1012-1019.
- [11] M. Terreni, R. Salvetti, L. Linati, R. Fernandez-Lafuente, G. Fernandez-Lorente, A. Bastida, J. M. Guisan. *Carbohydrate Research*. **2002**, 337, 1615-1621.
- [12] T. Miyazawa, M. Hamada, R. Morimoto, T. Murashima, T. Yamada. *Tetrahedron Letters*. **2008**, 49, 175-178.
- [13] I. Perez-Victoria and J. C. Morales. *Tetrahedron*. **2006**, 62 878-886.
- [14] X. Zhao, X. Tao, D. Wei, Q. Song. *Eur. J. Med. Chem.* **2006**, 41, 1352-1358.
- [15] G. Fernandez-Lorente, J. M. Palomo, J. Cocca, C. Mateo, P. Moro, M. Terreni, R. Fernandez-Lafuente and J. M. Guisan. *Tetrahedron*. **2003**, 59, 5705-5711.
- [16] A. Ghanem, H.Y. Aboul-Enein. *Tetrahedron Asymmetry*. **2004**, 15, 3331-3351.
- [17] A. Ghanem, V. Shurig. *Tetrahedron Asymmetry*. **2003**, 14, 57-62.
- [18] L. Yu, Y. Xu, X. Wang, X.Yu. *J. Mol. Catal B: Enzymatic*. **2007**, 47, 149-154.
- [19] P. Xue, X. H. Yan, Z.Wang. *Chin. Chem. Lett.* **2007**, 18, 929 – 932.
- [20] P. Ciuffreda, L. Alessandrini, G. Terraneo and E. Santaniello. *Tetrahedron Asymmetry*. **2003**, 14, 3197-3201.
- [21] L. M. Levy, G. De Gonzalo and V. Gotor. *Tetrahedron Asymmetry*. **2004**, 15, 2051-2056.
- [22] A. Mezzetti, C. Keith and R. J. Kazlauskas. *Tetrahedron Asymmetry*. **2003**, 14, 3917-3924.

- [23] F. Secundo and G. Carrea. *Chem. Eur. J.* **2003**, 9, 3194-3199.
- [24] L.S. Chua and M.R. Sarmidi. *Enzyme. Microb. Technol.* **2006**, 38, 551-556.
- [25] Toshifumi Miyazawa, Manabu Hamada, Ryohei Morimoto, Takashi Murashima and Takashi Yamada. *Tetrahedron Letters.* **2007**, 48, 8334 -8337.
- [26] A. Zaks. *Curr. Opin. Chem. Biol.* 2001, 5, 130-136.
- [27] P. Villeneuve, J. M. Muderhwa, J. Graille, M. J. Haas. *J. Mol. Catal., B Enzymatic.* **2000**, 9, 113-148.
- [28] S. Benjamin, A. Pandey. *Yeast.* **1998**, 14, 1069-1087.
- [29] V. M. Balcao, A. L. Paiva, F. X. Malcata. *Enzyme. Microb. Technol.* **1996**, 18, 392-416.
- [30] G. Yadav, K. M. Devi. *Biochem. Eng. Journal.* **2004**, 17, 57-63.
- [31] S. K. Khare, M. Nakajima. *Food Chemistry.* **2000**, 68, 153-157.
- [32] P. D. Desai, A. M. Dave, and S. Devi. *J. Mol. Cat B: Enzymatic.* 2004, 31, 143-150.
- [33] T. Sakurai, A. L. Margolin, A. J. Russell, A. M. Klivanov. *J. Am. Chem. Soc.* **1988**, 110, 7236.
- [34] Y.-F. Wang, J. J. Lalonde, M. Momongan, D.E. Bergbreiter, C. H. Wong. *J. Am. Chem. Soc.* **1988**, 110, 7200.
- [35] K. Nakamura, Y. Takebe, T. Kitayama, A. Ohno. *Tetrahedron Letters.* **1991**, 32, 4941.
- [36] F. Secundo, S. Riva, G. Carrea. *Tetrahedron Asymmetry.* **1992**, 3, 267.
- [37] S. Parida, J. S. Dordick. *J. Am. Chem. Soc.* **1991**, 113, 2253.
- [38] T. Anthonsen, J. A. Jongejan. *Methods Enzymol.* **1997**, 286, 473.
- [39] T. Anthonsen, B. H. Ho. *Chem. Phys. Lipids.* **1998**, 93, 199.
- [40] L. Kvittingen, *Tetrahedron.* **1994**, 50, 8253.
- [41] E. Rubio, F. M. Alfonso, A. M. Klivanov. *J. Am. Chem. Soc.* **1991**, 113 695.
- [42] Y-C. Cheng, S-W. Tsai. *Tetrahedron Asymmetry.* **2004**, 15, 2917-2920.
- [43] A. Zaks, A. M. Klivanov. *J. Am. Chem. Soc.* **1986**, 108, 2767.
- [44] E. Rubio, M. A. Fernandez, A. M. Klivanov. *J. Am. Chem. Soc.* **1991**, 108, 2767.
- [45] D. E. Stevenson, A. C. Storer. *Biotech. Bioeng.* **1991**, 37, 519.
- [46] A. Zaks, A.M. Klivanov. *enzymatic reactions in organic media.* **1996**, Eds. Blackie, London.

- [47] A. Fishman and U. Cogan. *J. Mol. Catal B : Enzymatic*. **2003**, 22, 193-202.
- [48] Shu Bai, Zheng Guo, Wei Liu, Yan Sun. *Food Chemistry*. **2006**, 96, 1-7.
- [49] A. Schmid, J.S.Dordick, B. Hauer, A.Kiener, M. Wubbolts. *Nature*. **2001**,409, 258.
- [50] S. Hazarika, P. Goswami, N. N. Dutta. *Chem. Eng. Journal*. **2003**, 94, 1-10.
- [51] G. Carrea, G. Ottolina, Sergio Riva. *Trends in Biotechnology*. **1995**, 13, 63-70.
- [52] A. M. Klivanov. *A. M. Nature*. **2001**, 409, 241-246.
- [53] Persson. M, Costes. D, Wehtje. E, Adlercreutz. P. *Enzym. Microb. Technol*. **2002**, 31, 88-93.
- [54] P. Dagnélie. *Statistique Théorique et Appliquée*, Tome 2, Edition : Université LARCIER et de BOECK. **1999**, 650 p.
- [55] *Minitab reference manual*. Release 14.31 for windows. PA State College, Minitab, 1047 p.
- [56] N. R. Draper and H. Smith. *Applied regression analysis*. Chichester, Wiley and sons, 3rd edition. **1998**, 706 p.
- [57] S. Weisberg. *Applied linear regression*. New York, Wiley. **1985**, 324 p.
- [58] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman. *Chemometrics: a textbook*. Amsterdam, Elsevier. **1988**, 488 p.
- [59] M. Otto. *Chemometrics*. Wiley-VCH, Weinheim, **1999**.
- [60] R. S. Searle. *Matrix algebra for the biological science*. New York, Wiley. **1966**, 256 p.
- [61] R. Palm. *Les critères de validation des équations de régression linéaire*. Note de statistique et d'informatique. Faculté des Sciences Agronomiques (Gembloux). **1988**, 27 p.
- [62] R. Tomassone, E. Lesquoy, et C. Miller. *La régression : nouveau regards sur une ancienne méthode statistique*, Masson, INRA. **1983**, 180 p.
- [63] P. Dagnélie. *Analyse statistique à plusieurs variables*. GEMBLoux. Presses Agronomiques. **1982**, 362 p.
- [64] Ph. Besse. *Pratique de la modélisation statistique*. Publications du laboratoire de statistique et probabilités, Université Paul Sabatier, France. **2001**.
- [65] M. Stone. *Journal of the Royal Statistical Society*. **1974**, 36, 111-147.
- [66] D. M. Allen. *Technometrics*. **1974**,16, 125-127.
- [67] B. Efron and G. Gong. *American Statistician*. **1983**, 37, 36-48.

- [68] R. Welsch, Leverage, S. Kotz, N. L Johnson. *Encyclopedia of Statistical Sciences*. Wiley, New York, **1983**.
- [69] G. Carrea, S. Riva. *Angew. Chem. Int. Ed.* **2000**, 39, 2226-2254.
- [70] Y. Wang, Q. Li, Z. Zhang, J. Ma, Y. Fen. *J. Mol. Catal.B : Enzym.* **2008**, 10, 1016.
- [71] Y. Zhou, *Computational Study of Enzyme Enantioselectivity*, Doctoral Thesis, Delf University of Technology, **2006**.
- [72] R.H. Valivety, G.A. Johnston, C.J. Suckling, P.J. Halling. *Biotechnol. Bioeng.* **1991**, 38, 1137-1143.
- [73] <http://www.vcclab.org/lab/edragon>.
- [74] Roberto Todeschini and Viviana Consonni. *Handbook of Molecular Descriptors, Methods and Principles in Medicinal Chemistry*, Edited by R. Mannhold; H. Kubinyi; H. Timmerman, Volume11, 412-434 p.
- [75] G. Ding, J. Chen, X. Qiao, L. Huang, J. Lin, X. Chen. *Chemosphere.* **2006**, 62, 1057–1063.
- [76] J. Ottosson, L. Fransson, J. W. King b, K. Hult. *Biochimica et Biophysica Acta.* **2002**, 1594, 325-334.