

وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR - ANNABA UNIVERSITY

UNIVERSITE BADJI MOKHTAR - ANNABA



جامعة باجي مختار - عنابة

Année 2015

Faculté des Sciences
Département de Chimie

THESE

Présentée pour obtenir le diplôme de DOCTORAT

Option : Chimie organique Appliquée

THEME

*Etude de la relation quantitative structure-activité
inhibitrice des enzymes hydrolytiques : cas des
alpha-glucosidases*

Présentée Par:

M. SAIHI Youcef

Directeurs de Thèse : FERKOUS Fouad Professeur à l'Université de Annaba

DJEGHABA Zeineddine Professeur à l'Université de Annaba

Devant le jury suivant:

Président : REGAINIA Zine Professeur à l'Université de Annaba

Examineurs :

KHATMI Djameleddine Professeur à l'Université de Guelma

ABDAOUI MOHAMED Professeur à l'Université de Guelma

MADI Fatiha Maitre de conférences à l'Université de Guelma

Abstract :

In this work of thesis we present two theoretical studies of the quantitative relation structure-inhibitory activity of the α -glucosidase. The first study was carried out starting from a data set composed of 57 molecules derived from xanthone and curcuminoide. The second study was carried out starting from a library of 37 molucles derived from N-(phenoxyalkyl) phthalimide.

In the first study, we have established a nonlinear model by employing the artificial neurons networks method (ANN) and the genetic algorithms (GA) as training and features selection methods respectively. The determination coefficients resulting from training, internal and external validation are: $R^2 = 94,1\%$, $Q^2_{cv} = 86,9\%$, $Q^2_{ext} = 86,7\%$. Compared with those obtained by the multiple-linear regression, these good results prove the great capacity of the ANN to model the inhibitory activity of the α -glucosidase and prove the power of its predictive capacity.

In the second study, the model obtained is consisted of three descriptors weighted by the volume of Van Der Walls, weight and electronegativity respectively. The reliability, the stability and the predictive capacity of the model are evaluated and proved by internal validation, external validation and the randomization test.

Key words: α -glucosidase, QSAR, ANN, MLR, xanthone, curcuminoide, N-(phénoxyalkyl) phthalimide, inhibitors.

Résumé :

Dans ce travail de thèse nous présentons deux études théoriques de la relation quantitative structure-activité inhibitrice de l' α -glucosidase. La première a été réalisée à partir d'une bibliothèque de 57 molécules dérivées de xanthone et de curcuminoïde. La deuxième a été réalisée à partir d'une bibliothèque de 37 molécules dérivées de N-(phénoxyalkyl)phtalimide.

Dans la première étude, nous avons établi un modèle non linéaire en employant la méthode des réseaux de neurones artificiels (RNA) et les algorithmes génétiques (AG) en tant que méthodes d'apprentissage et de sélection de variables respectivement. Les coefficients de détermination issus de l'apprentissage de la validation interne et de la validation externe sont : $R^2 = 94,1\%$, $Q^2_{cv} = 86,9\%$, $Q^2_{ext} = 86,7\%$. Comparés à ceux obtenus par la régression linéaire multiple (RLM), ces bons résultats prouvent la grande capacité des RNA à modéliser efficacement l'activité inhibitrice de l' α -glucosidase et la puissance de son pouvoir prédictif.

Dans la deuxième étude, le modèle obtenu est formé de trois termes qui représentent trois descripteurs pondérés par le volume de Van Der Waals, la masse et l'électronégativité respectivement. La fiabilité, la stabilité et le pouvoir prédictif du modèle sont évalués et prouvés par la validation interne, la validation externe et le test de randomisation.

Mots clés : α -glucosidase, QSAR, RNA, RLM, xanthone, curcuminoïde, N-(phénoxyalkyl)phtalimide, inhibiteurs.

Je dédie ce travail à

Ma mère

La mémoire de mon père

Ma petite famille : ma femme et mes filles

*Ma grande famille : mes frères, mes sœurs
et leurs enfants*

*Mes amis particulièrement
Khairo et Abdelkader*

Remerciements

Ce travail de recherche a été réalisé au sein du laboratoire de Chimie Organique Appliquée, groupe « Bioconversion et Synthèse Organique », sous la direction de Messieurs Fouad FERKOUS et Zeineddine DJEGHABA Professeurs à l'Université de Annaba.

Je tiens à remercier chaleureusement, en tout premier lieu, mes co-directeurs de thèse les Professeurs Fouad FERKOUS et Zeineddine DJEGHABA pour l'intérêt et la confiance qu'ils ont su m'accorder tout au long de ma thèse.

Mes meilleurs remerciements s'adressent à Monsieur Zine REGAINIA Professeur à l'université de Annaba pour m'avoir fait l'honneur de présider le jury de cette thèse.

Je remercie vivement Monsieur le Professeur Djameleddine KHATMI qui me fait l'honneur d'être présent dans ce jury pour examiner et évaluer ce travail.

J'adresse mes respectueux remerciements à Monsieur Mohamed ABDAOUI professeur à l'université de Guelma pour avoir accepté d'être membre de ce jury et d'avoir accepté d'examiner ce manuscrit.

J'adresse également mes remerciements à Madame Fatima MADI, Maître de conférences à l'université de Guelma, d'avoir accepté de participer à ce jury.

Que tous ceux qui m'ont accordé un soutien, une aide technique ou un conseil, trouvent ici l'expression de ma profonde reconnaissance.

Un grand merci à tous mes amis pour leur soutien, et particulièrement à mon ami Khairo. Heureusement que tu étais là.

De chaleureux remerciements à l'égard de ma famille, pour mes sœurs et mes frères, mais surtout pour mes parents. C'est grâce à eux, si je suis arrivé jusque là. Merci encore pour tout ce que vous avez fait pour moi. Merci Sabrina pour ton aide, ta compréhension, ton soutien et ta grande patience.

Et pour finir, merci aussi à tous ceux que j'ai pu oublier.

Liste des abréviations

Asp : Acide Aspartique.

CAZy : Carbohydrate-Active Enzymes.

Coef : Coefficients.

DFP : di-isopropylfluorophosphate

DQM : Déviation Quadratique Moyenne.

EC : Enzyme classification

Eq : Equation.

GH : Glycosyle Hydrolases

Glu : Acide Glutamique

IC₅₀ : la concentration d'inhibiteur pour laquelle la vitesse initiale de formation de produit diminue jusqu'à la moitié de sa valeur mesurée en absence d'inhibiteur.

IUB-MB : Union Internationale de Biochimie et de Biologie Moléculaire.

QSAR : la Relation Quantitative Structure-Activité (Quantitative Structure-Activity Relationship).

RLM : Régression Linéaire Multiple.

RNA : Réseaux de Neurones Artificiels.

SN2 : Substitution Nucléophile type 2.

Tyr : Tyrosine

Liste des figures

Chapitre 1 : Généralités sur les glycosidases et leurs inhibiteurs

- Figure 1** : les différents modèles de l'inhibition compétitive. 24
- Figure 2** : représentation graphique de la vitesse initiale en fonction de la concentration de substrat 25
- Figure 3** : représentation double inverse $\frac{1}{v_i} = f\left(\frac{1}{[S]}\right)$ selon *Lineweaver-Burk* 26
- Figure 4** : représentation graphique $\frac{[S]}{v_i} = f([S])$ selon *Hanes-Woolf*. 26
- Figure 5** : représentation graphique $K_M = f([I])$ selon *Lineweaver-Burk*. 26
- Figure 6** : détermination expérimentale d'IC₅₀ à partir de la courbe $\frac{v_i^I}{v_i} = f([I])$. 27
- Figure 7** : mécanisme d'inhibition non compétitive pure. 28
- Figure 8** : représentation graphique de $V_i^I = f([S])$ dans le cas d'inhibition non compétitive pure. 29
- Figure 9** : représentation graphique double inverse selon *Lineweaver-Burk* pour l'inhibition non compétitive pure. 30
- Figure 10** : tracé double inverse de *Lineweaver-Burk* dans le cas d'inhibition non compétitive mixte (pour $K_I < K_I'$). 31
- Figure 11** : tracé double inverse de *Lineweaver-Burk* dans le cas d'inhibition non compétitive mixte (pour $K_I > K_I'$). 32
- Figure 12** : modèles d'inhibition non compétitive mixte³⁶. 32
- Figure 13** : formation du complexe inactif ESI³⁶. 34
- Figure 14** : représentation graphique double inverse selon *Lineweaver-Burk* pour inhibition incompétitive. 35
- Figure 15** : structure chimique de l'acarbose. 36
- Figure 16** : structures chimiques de Valiénamine, Validamine, Valiolamine et de Voglibose. 41
- Figure 17** : structures chimiques de quelques iminosucres inhibiteurs de 43

glycosidases.

Figure 18 : structures chimiques des polyphénols inhibiteurs de glycosidases. 44

Figure 19 : structures chimiques de quelques inhibiteurs d'origine naturelle dérivés de flavanoides. 45

Figure 20 : structures chimiques de tannin hydrolysable et de tannin condensé. 47

Figure 21 : Quelques chalcones naturels ayant un pouvoir inhibiteur sur les glycosidases 47

Figure 22 : Structures chimiques des inhibiteurs de l' α -glucosidase dérivés des chalcones 48

Figure 23 : Structure de salacinol et de kotalanol 48

Chapitre 2 : QSAR : PRINCIPE ET METHODOLOGIE

Figure 1: présentation de la méthodologie de QSAR 56

Figure 2 : Réseau maillé. 57

Figure 3 : structure du toluène. 65

Figure 4 : représentation d'un neurone formel. 69

Figure 5 : architecture des réseaux de neurones. 70

Figure 6 : principe de la validation croisée leave-*n*-out. 75

Chapitre3 : Modélisation de la relation non linéaire quantitative structure activité inhibitrice de l' α -glucosidase par les dérivés de Xanthone et de Curcuminoïde.

Figure 1 : Structures développées des dérivés du xanthone et du curcuminoïde 88

Figure 2 : représentation schématique de l'algorithme de rétropropagation 95

Figure 3 : Influence du nombre de neurones de la couche cachée sur les valeurs des coefficients de corrélation. 99

Figure04 : représentation des résidus en fonction des valeurs expérimentales de pIC50. 101

Figure 5 : Représentation des valeurs des coefficients de détermination des sous-ensembles de validation et de test en fonction de celles de l'ensemble d'apprentissage. 101

Figure 6 : représentations graphiques des valeurs calculées de pIC50, obtenues par les modèles RLM et RNA, en fonction de celles expérimentales. 103

Figure 7 : Contribution des descripteurs du modèle ANN. 104

Liste des Schémas

Chapitre 1 : Généralités sur les glycosidases et leurs inhibiteurs

Schéma 1 : Hydrolyse simplifiée de la liaison glycosidique.	8
Schéma 2 : Mécanisme d'inversion de configuration anomérique de la réaction d'hydrolyse par les glycosidases.	9
Schéma 3 : Mécanisme de rétention de configuration anomérique de la réaction d'hydrolyse par les glycosidases.	10
Schéma 4 : état de transition lors de l'hydrolyse glycosidique.	11
Schéma 5 : conformations possibles lors de l'hydrolyse glycosidique (demi chaise « ${}^4H_3, {}^3H_4$ » ou bateau « ${}^{2,5}B, B_{2,5}$ »).	12
Schéma 6 : Similarités des mécanismes chez diverses familles de glycosidases.	13
Schéma 7 : Similarités dans l'orientation des liaisons dans des enzymes de mécanismes opposés.	13
Schéma 8 : Mécanisme de la réaction enzymatique selon Michaëlis-Menten.	19
Schéma 9 : mécanisme réactionnel de l'inhibition compétitive.	23
Schéma 10 : mécanisme d'inhibition incompétitive.	35
Schéma 11 : mécanisme d'inhibition irréversible.	39
Schéma 12 : mécanisme d'inhibition de la protéase par le di-isopropyl fluorophosphate.	40

Liste des tableaux

Chapitre 1 : Généralités sur les glycosidases et leurs inhibiteurs

Tableau 1 : Mécanismes réactionnels et acides aminés impliqués dans la catalyse enzymatique des familles des glycosidases classées en clans.	15
Tableau 2 : Division de la famille GH13 en sous familles.	17
Tableau 3 : résumé des modifications des paramètres cinétiques en présence d'inhibiteurs réversibles.	38

Chapitre 2 : QSAR : PRINCIPE ET METHODOLOGIE

Tableau.1 : quelques paramètres utilisés dans l'analyse	59
Tableau.2 : liste des blocs de descripteurs fournis par le logiciel Dragon.	60
Tableau.3 : liste des descripteurs constitutionnels générés par le logiciel Dragon.	61
Tableau.4 : présentation de quelques descripteurs moléculaires du bloc de groupements fonctionnels.	63
Tableau.5 : Quelques descripteurs topologiques fournis par logiciel Dragon.	66

Chapitre3 : Modélisation de la relation non linéaire quantitative structure activité inhibitrice de l' α -glucosidase par les dérivés de Xanthone et de Curcuminoïde.

Tableau 1: Valeurs expérimentales de pIC ₅₀ et celles calculées par le modèle RNA pour les 57 molécules listées dans la figure 1.	89
Tableau 2 : Glossaire des descripteurs sélectionnés par les algorithmes génétiques.	90
Tableau 3 : Valeurs de pIC ₅₀ expérimentales et valeurs calculées par le modèle linéaire RLM.	96
Tableau 4 : Les valeurs des coefficients de corrélation correspondant aux nombres de neurones de la couche cachée.	102
Tableau 5 : Coefficients de détermination et erreurs carrées moyennes obtenus par les modèles RLM et RNA.	100
Tableau 6 : La contribution C_i de chaque descripteur.	104

SOMMAIRE

Introduction générale	2
Chapitre 1 : Généralités sur les glycosidases et leurs inhibiteurs	
1. Introduction	8
2. Mécanismes d'action	8
2.1. Mécanisme avec inversion de configuration anomérique	8
2.2. Mécanisme avec rétention de configuration anomérique	9
2.3. Quelques cas particuliers	10
2.4. Etat de transition	11
3. Classifications des glycosidases	14
3.1. Classification traditionnelle	14
3.2. Classification structurale	14
3.2.1. Principe et classification	14
3.2.2. Répartition des familles des glycosides hydrolases en sous familles	17
4. Inhibition enzymatique	19
4.1. Bref rappel sur la cinétique enzymatique	19
4.1.1. Equation de <i>Michaëlis-Menten</i>	19
4.1.2. Détermination de V_{\max} et K_M	21
4.2. Types d'inhibition enzymatique	22
4.2.1. Inhibition réversible	22
4.2.1.1. Inhibition compétitive	22
4.2.1.1.1. Cinétique de l'inhibition compétitive	24
4.2.1.1.2. Activité fractionnaire et degré d'inhibition	28
4.2.1.1.3. Mesure du paramètre IC_{50}	28
4.2.1.2. Inhibition non compétitive	29
4.2.1.2.1. Inhibition non compétitive pure	29
4.2.1.2.1.1. Détermination du Degré d'inhibition et de l' IC_{50}	31
4.2.1.2.2. Inhibition non compétitive mixte	31
4.2.1.2.2.1. Détermination du degré d'inhibition (i) et de l' IC_{50}	33

4.2.1.2.2.2. Modèles d'inhibition non compétitive mixte	33
4.2.1.3. Inhibition incompétitive	34
4.2.1.3.1. Degré d'inhibition et IC ₅₀	36
4.2.2. Inhibition irréversible	39
4.3. Inhibiteurs des glycosidases	41
4.3.1. Aminocyclitols polyhydroxylés	41
4.3.2. Iminosucres polyhydroxylés	42
4.3.3. Polyphénols	43
4.3.3.1. Flavanoïdes	45
4.3.3.2. Tanins	46
4.3.4. Chalcones	47
4.4.5. Salacinol et katalanol	48
5. Références	49

Chapitre 2 : QSAR : PRINCIPE ET METHODOLOGIE

1. Introduction	55
2. Principe de QSAR	55
3.3. D QSAR	57
4. Application de QSAR	57
5. Outils et Méthodologie de QSAR	58
5.1. Paramètres biologiques	58
5.2. Utilisation des descripteurs moléculaires	59
5.2.1. Descripteurs générés par le logiciel DRAGON	59
5.2.1.1. Les descripteurs constitutionnels	60
5.2.1.2. Descripteurs de groupements fonctionnels	62
5.2.1.3. Les descripteurs topologiques	64
5.2.1.4. Les descripteurs géométriques	66
5.3. Sélection des descripteurs	67
5.3.1. Sélection objective	67
5.3.2. Sélection subjective	68
5.3.2.1. Introduction progressive	68
5.3.2.2. Elimination progressive	68
5.3.2.3. Sélection pas à pas	68

5.3.2.4. Sélection par les algorithmes génétiques	68
5.4. Méthodes de modélisation des données	69
5.4.1. Réseaux de neurones	69
5.4.2. La régression linéaire multiple	71
5.5. Validation du modèle	73
5.5.1. Coefficients et tests statistiques standards	73
5.5.2. Validation interne	75
5.5.2.1. La procédure leave- <i>n</i> -out	75
5.5.2.2. La procédure leave-one-out	76
5.5.2.3. Test de randomisation	76
5.5.3. Validation externe	77
5.6. Domaine d'applicabilité	77
6. Références	78

Chapitre3 : Modélisation de la relation non linéaire quantitative structure activité inhibitrice de l' α -glucosidase par les dérivés de Xanthone et de Curcuminoïde.

1. Introduction	84
2. Ensemble de molécules	84
3. Dessin et optimisation des structures	90
4. Génération des descripteurs moléculaires	90
5. Sélection des variables et formation du modèle	90
6. Utilisation des réseaux de neurones dans la construction du modèle non linéaire	91
7. Résultats et discussion	96
7.1 Comparaisons entre le modèle RNA et RLM	102
7.2 Analyse de la contribution des descripteurs dans le modèle RNA	104
8. Conclusion	105
9. Références	106
Conclusion générale	108

Introduction générale

Introduction générale

Les carbohydrates tels que le saccharose et l'amidon représentent la majorité des constituants dans le régime alimentaire de l'être humain. L'action de l'enzyme α -amylase, un des constituants de la salive, sur l'amidon libère de nouvelles grosses molécules non absorbées par le corps, connues sous le nom "oligosaccharides". Ces derniers réagissent au niveau de l'intestin grêle avec une autre enzyme appelée α -glucosidase (EC 3.2.1.20) en libérant du glucose, molécule de plus petite taille, capable d'être absorbée par le corps humain. L'inhibition de l' α -glucosidase bloque le processus de la dégradation des oligosaccharides, stoppant la digestion des carbohydrates, et permettant ainsi de contrôler le niveau du glucose dans le sang chez les diabétiques¹. Cet effet bénéfique, des inhibiteurs des glycosidases, sur la santé a attiré l'attention des chercheurs et des firmes pharmaceutiques pour développer et enrichir la classe thérapeutiques des médicaments anti diabète type II².

Une grande importance est donnée aux inhibiteurs de l' α -glucosidase dans l'objectif d'enrichir cette classe thérapeutique. Malheureusement, le manque d'information détaillée sur le mécanisme d'inhibition de l' α -glucosidase d'origine humaine, le coût et l'indisponibilité de cette enzyme sur le marché ont poussé les chercheurs à utiliser les outils de la modélisation statistique et moléculaire pour développer de nouvelles structures de base afin d'enrichir cette classe thérapeutique d'inhibiteurs et minimiser le coût financier des tests expérimentaux³⁻⁵.

La quantification de la relation existante entre la structure moléculaire et l'activité biologique (QSAR : Quantitative Structure Activity Relationship) est devenue, actuellement, un outil indispensable dans le domaine de la conception de médicaments, notamment dans le cas d'absence d'information sur le site actif de l'enzyme. Techniquement, cette approche se base principalement sur quatre piliers : le choix de la structure de base ; le choix des paramètres physico-chimiques (descripteurs moléculaires) ; la méthode de sélection des descripteurs et l'algorithme d'apprentissage⁶. La découverte d'une telle relation permet de prédire les propriétés physiques et chimiques ainsi que l'activité biologique des composés, de développer de nouvelles théories ou de comprendre les phénomènes observés. Elle permet également de modéliser d'analyser des familles de composés et de prévoir et de sélectionner ceux possédant une activité biologique potentielle⁷⁻⁹.

Introduction générale

Récemment, plusieurs travaux de recherche, traitant de la modélisation de l'inhibition de l' α -glucosidase par la méthode QSAR, ont été rapportés dans la littérature. Dans la plupart des modèles de QSAR construits, la régression linéaire multiple est très utilisée comme algorithme d'apprentissage. Cela est justifié par les avantages qu'elle présente (simple à construire, simple à interpréter la contribution de chaque variable)¹⁰. Toutefois son inconvénient majeur est que l'apprentissage effectué suit une fonction linéaire ce qui minimise l'exploration de l'espace de recherche^{11,12}. Parmi les méthodes alternatives à la régression linéaire multiple, la méthode des réseaux de neurones, potentiellement intéressante et efficace lorsque les effets des descripteurs sont non linéaires¹³.

L'utilisation de ces deux méthodes d'apprentissage, pour modéliser l'activité inhibitrice de l' α -glucosidase par certains ensembles de molécules, constitue le cœur de cette thèse qui vise à développer des modèles QSAR justes, stables, robustes, possédant un pouvoir prédictif puissant et capables d'être généralisés à d'autres structures moléculaires. Ceci leur permettra d'être utilisables dans le criblage virtuel et dans le développement de nouvelles molécules antidiabétiques¹³. En effet, nous présentons deux études de la relation quantitative structure-activité inhibitrice de l' α -glucosidase. La première concerne des dérivés de xanthone et de curcuminoïde et la deuxième des dérivés de N-(phénoxyalkyl)phtalimide.

Notre travail est divisé en trois chapitres.

Le premier chapitre présente un aperçu général sur les enzymes glycosidases, la structure de leur site actif, leur classification, leur mécanisme d'action, un bref rappel sur la cinétique enzymatique montrant les différents types d'inhibition et les différents composés inhibiteurs des glycosidases y est également présenté

Le deuxième chapitre intitulé «QSAR : principe et méthode » présente la méthode QSAR, son principe, son application et les différentes étapes à respecter pour la réaliser.

Le troisième chapitre est consacré à la méthodologie du travail réalisé dans cette thèse, aux résultats obtenus et à leurs discussions.

Enfin, nous terminerons ce manuscrit par une conclusion générale et des perspectives.

Références :

1. L. EL BLIDI, "Synthèse chimio-enzymatique et évaluation d'inhibiteurs potentiels de glycosidases, analogues de la valioline" Thèse de doctorat, Université BLAISE PASCAL, France, 2006.
2. A. K. Ghose, V. N. Viswanadhan, "Combinatorial Library Design and Evaluation: Principles, Software Tools, and Applications in Drug Discovery" Ed. *Marcel Dekker, Inc. Etats Unies*, 2001.
3. T. I. Oprea, "Cheminformatics in Drug Discovery" Ed. *WILEY-VCH Verlag. Allemagne*, 2005.
4. E. A. Rekká ; P. N. Kourounakis "Chemistry and Molecular Aspects of Drug Design and Action" Ed. *Taylor & Francis Group, LLC. Etats Unies*, 2008.
5. A. K. Ghose, V. N. Viswanadhan, "Combinatorial Library Design and Evaluation: Principles, Software Tools, and Applications in Drug Discovery" Ed. *Marcel Dekker, Inc. Etats Unies*, 2001.
6. F. Li, H. Wu, L. Li, X. Li, J. Zhao, W. J. G. M. Peijnenburg, *Ecotoxicol. Environm. Saf.* 80, 2012, 273–279.
7. T. Puzyn, J. Leszczynski, M. T. D. Cronin, « Recent advances in QSAR studies- Methods and applications » *Springer Dordrecht Heidelberg London New York*, 2010.
8. M. Tichý, *Science.Tot. Env.* 109, 1991, 407-410.
9. B. Peric, J. Sierra, E. Martí, R. Cruañas, M. A. Garau, *Ecotoxicol. Environm. Saf.* 115, 2015, 257-262.
10. H. Nocairi, M. Hanafi, E.M. Qannari. *Rev. Statistique Appliquee*, LIII(2), 2005, 29-41.
11. L. Douali, A. R. Schmitzer, D. Villemin, A. Jarid, D. Cherqaoui, *Phys. Chem. News.* 34, 2007, 131-144.
12. P.Desmet. Recherche et application en Marketing, 11(2), 1996, 17-28.
13. E. Tenorio-Borroto, C. G. Rivas, J. C. Vásquez Chagoyán, N. Castañedo, F. J. Prado-Prado, X. García-Mera, H. González-Díaz. *Bioorg. Med. Chem.* 20, 2012, 6181–6194.

Chapitre 1

Généralités sur les glycosidases et leurs inhibiteurs

1. Introduction :

Les glycosidases (glycoside hydrolases) sont des enzymes très répandues et impliquées dans un grand nombre de processus biologiques fondamentaux du métabolisme des sucres : la digestion des polysaccharides alimentaires dans la bouche et l'intestin grêle, la dégradation des polysaccharides au niveau du lysosome, ou encore la glycosylation des protéines au niveau du réticulum endoplasmique et de l'appareil de Golgi¹.

Les glycosidases catalysent la dégradation hydrolytique des polysaccharides en oligosaccharides et en hydrates de carbone monomères (Schéma 1). De toutes les enzymes présentes dans la nature, ce sont les plus abondantes, les plus évoluées et comptent parmi les plus efficaces. Elles arrivent à augmenter jusqu'à 10^{17} fois la vitesse de la réaction de coupure de la très stable liaison glycosidique C-O ($347 \text{ kJ}\cdot\text{mol}^{-1}$).²

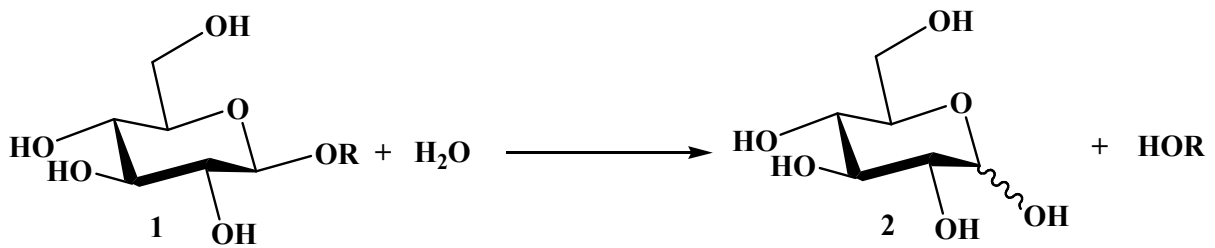


Schéma 1 : Hydrolyse simplifiée de la liaison glycosidique.

2. Mécanismes d'action :

Le mode d'action de l'hydrolyse enzymatique des liaisons glycosidiques le plus commun s'opère par une catalyse au cours de laquelle une paire d'acides aminés, provenant des résidus glutamate ou aspartate du site actif de l'enzyme, est mise en jeu. Dans ce cas, un acide aminé joue le rôle du donneur de proton acide/base et l'autre agit en tant que base/nucléophile. Ce mode d'action peut se dérouler selon deux mécanismes distincts³ qu'on retrouve majoritairement dans les réactions impliquant les glycosidases.

2.1. Mécanisme avec inversion de configuration anomérique :

Les enzymes qui libèrent un sucre de configuration opposée, agissent avec inversion de configuration anomérique (par simple substitution nucléophile). Elles utilisent un mécanisme catalytique non-covalent pour exploiter la catalyse acido-basique générale.

Dans ce cas, les deux résidus acide et basique sont suffisamment séparés, d'une distance moyenne de 10,5Å, ce qui permet l'insertion du substrat et la molécule d'eau dans le site actif. Le processus se fait en une étape : un des deux résidus carboxyliques permet le départ de l'aglycone par son rôle de catalyseur acide, le second acide carboxylique déprotoné active une molécule d'eau en générant un hydroxyle qui attaque le carbone anomérique. Selon ce mécanisme de type SN2, il y a inversion de la configuration anomérique⁴⁻⁶ (schéma 2). Cette réaction s'effectue en passant par un état de transition dont la structure est proche d'un cation oxocarbenium.

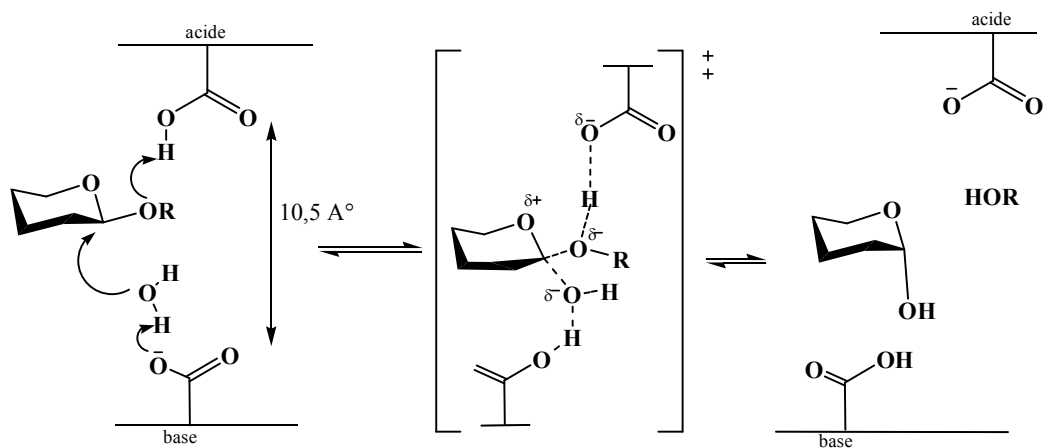


Schéma 2: Mécanisme d'inversion de configuration anomérique de la réaction d'hydrolyse par les glycosidases⁷.

2.2. Mécanisme avec rétention de configuration anomérique :

Les enzymes agissant avec rétention de configuration anomérique, libèrent un produit dont la stéréochimie du carbone anomérique est la même que celle du substrat (après deux inversions successives). Dans ce cas, les résidus carboxyliques sont moins éloignés (5,5Å) et l'hydrolyse a lieu en deux étapes. Le nucléophile effectuant la première attaque sur le carbone anomérique est l'acide carboxylique déprotoné de l'enzyme (Asp ou Glu), formant un intermédiaire covalent. Par conséquent, le substituant du glucoside est relâché à cette première étape. Par la suite, un deuxième résidu acide active une molécule d'eau qui, en attaquant le carbone anomérique, libère le glucide de l'enzyme. Deux étapes de type SN2 se suivent donc, il y a rétention de la configuration du carbone anomérique (schéma 3).

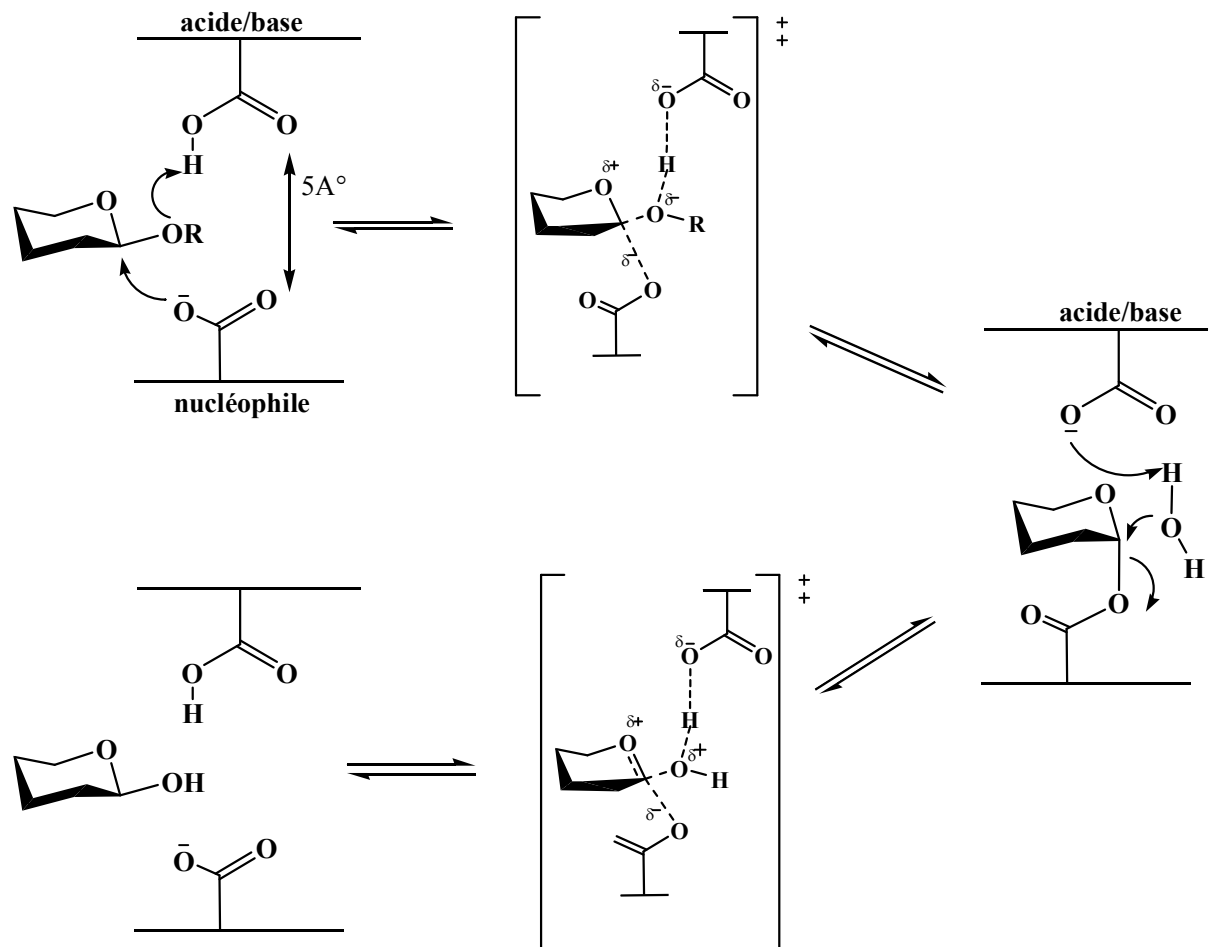


Schéma 3 : Mécanisme de rétention de configuration anomérique de la réaction d'hydrolyse par les glycosidases⁷.

2.3. Quelques cas particuliers :

Il existe néanmoins quelques exemples de glycosidases agissant selon des mécanismes différents de ceux mentionnés ci-dessus. Il s'agit des cas particuliers où l'acide catalytique peut être remplacé par un phosphate inorganique, comme, par exemple : la maltose phosphorylase (EC 2.4.1.8, famille GH65), la sucrose phosphorylase (EC 2.4.1.7, famille GH13), ou la cellobiose phosphorylase (EC 2.4.1.20, famille GH94)⁷⁻⁹. Certaines glycosidases comme les myrosinases, les chitinases et les endo N-acétylglucosaminidases présentent un mécanisme moléculaire d'hydrolyse n'impliquant l'intervention que d'un seul acide aminé catalytique¹⁰. D'autres impliquent des mécanismes d'oxydoréduction et/ou d'élimination^{11,12}.

2.4. Etat de transition :

Dans l'état de transition (schéma 4), les liaisons entre le nucléophile et l'oxygène glycosidique du groupement partant (ou l'eau dans l'étape de déglycosylation) sont partiellement formées et rompues. Pendant cette période très courte, la molécule possède donc un caractère d'ion oxocarbenium. Ceci implique une hybridation sp^2 du carbone anomérique et donc un caractère de double liaison entre celui-ci et l'oxygène endocyclique. Il apparaît également une charge délocalisée le long de cette liaison. Tout ceci conduit le pyranose à se distordre de manière à conduire le système à cet état de transition.

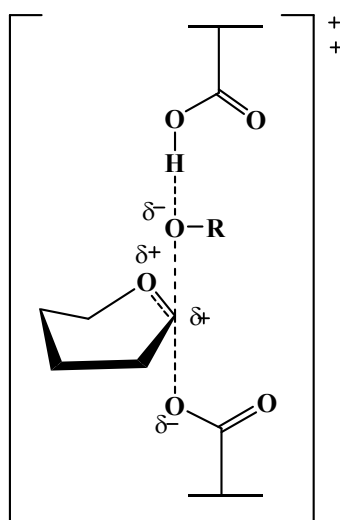


Schéma 4 : état de transition lors de l'hydrolyse glycosidique.

Les mécanismes d'inversion et de rétention de configuration anomérique font intervenir 4 états de transition possibles de type ion oxocarbenium⁷⁻⁹ (schéma 5). Parmi les β -glycosidases agissant selon un mécanisme avec rétention de configuration anomérique, les conformères 4H_3 , ${}^{2,5}B$ et $B_{2,5}$ ont été observés respectivement chez les familles GH5 et GH7^{13,14}, GH11¹⁵ et enfin GH26¹⁶. Pour les β -glycosidases agissant selon un mécanisme d'inversion de configuration anomérique, c'est la transition en conformère de type ${}^{2,5}B$ qui est observée chez les familles GH6, GH8 et GH48.¹⁷

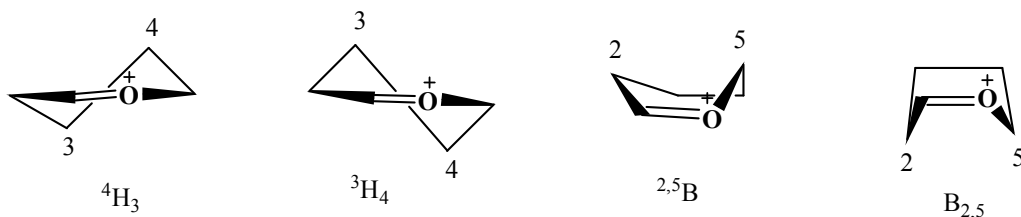


Schéma 5 : conformations possibles lors de l'hydrolyse glycosidique (demi chaise « 4H_3 , 3H_4 » ou bateau « ${}^{2,5}B$, $B_{2,5}$ »).

Chez les glycosidases agissant sur les liaisons β - anomériques, ces complexes de Michaelis (enzyme-substrat) correspondent à des distorsions de la forme cyclique nécessaire pour amener la liaison glycosidique dans une position pseudo axiale ¹⁸. Cette orientation favorise une attaque nucléophile directe. De telles distorsions du cycle accompagnant la formation du complexe de Michaelis sont inutiles pour les glucoamylases (enzymes agissant sur des liaisons α - anomériques) et leurs homologues de la famille GH15. En effet, la liaison glycosidique est déjà dans une position axiale facilitant l'attaque nucléophile. En fait, le mécanisme moléculaire d'inversion des glycosidases, agissant sur des liaisons α - anomériques, passe par une attaque nucléophile réalisée par une molécule d'eau. Cette molécule d'eau, préalablement activée par un résidu basique, va attaquer le carbone anomérique du glycoside non distordu, dans le même temps, avec rupture de la liaison glycosidique axiale (schéma 6) ¹⁹.

Des similarités, telles que celles observées précédemment, nécessitent une distorsion du substrat dans le site actif d'une β -glycosidase possédant un mécanisme de rétention de configuration anomérique. Cette distorsion place la liaison glucosidique dans une orientation axiale (schéma 7A) similaire à celle d'une liaison α - anomérique dans le site actif d'une glycosidase avec un mécanisme d'inversion de configuration anomérique (schéma 7B).

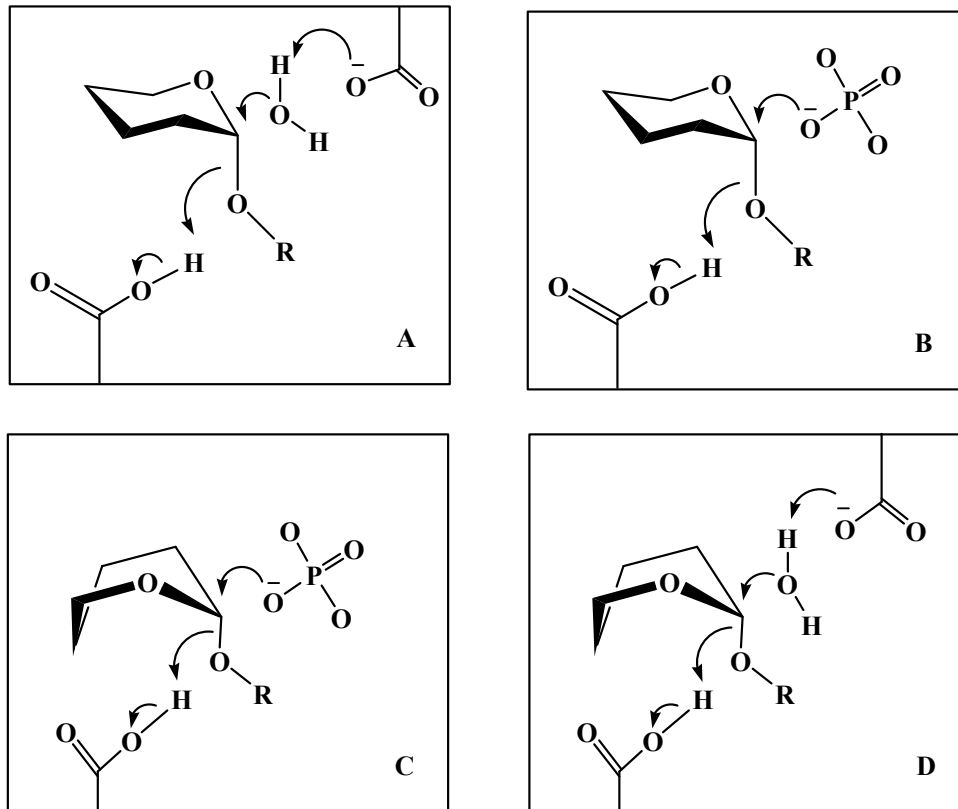


Schéma 6 : Similarités des mécanismes chez diverses familles de glycosidases.

- (A) Mécanisme des glycosidases hydrolysant des liaisons α chez les familles GH15, GH37, GH63, GH78, GH92 et GH95.
- (B) Mécanisme des phosphorylases de la famille GH65.
- (C) Mécanisme des phosphorylases de la famille GH94 hydrolysant des liaisons β .
- (D) Mécanisme des enzymes des familles GH8, GH9 et GH48 hydrolysant des liaisons β .

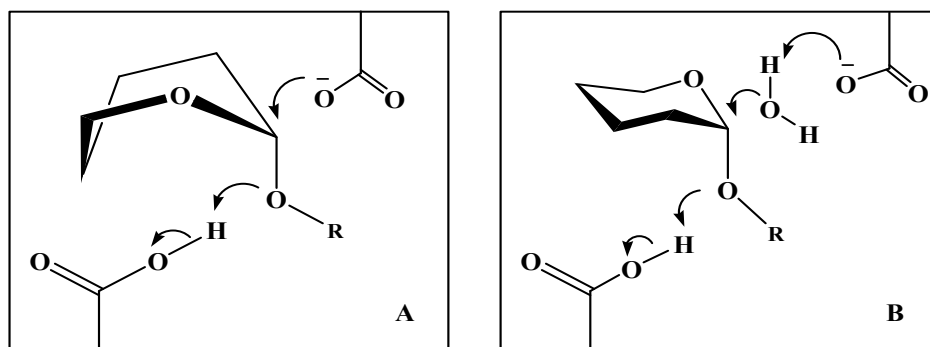


Schéma 7 : Similarités dans l'orientation des liaisons dans des enzymes de mécanismes opposés.

3. Classifications des glycosidases :

3.1. Classification traditionnelle :

Dans la nomenclature de l'Union Internationale de Biochimie et de Biologie Moléculaire (IUB-MB) basée sur le type de réaction catalysée et la spécificité de substrat, les glycosidases possèdent un numéro du type EC 3.2.1.x. Les trois premiers chiffres indiquent qu'elles hydrolysent des liaisons O-glycosidiques, le dernier (x) est variable et dépend du substrat transformé²⁰. Cette classification systématique permet de nommer précisément la spécificité de substrat d'une enzyme et possède en outre, l'avantage de pouvoir classer rapidement une enzyme. Cependant, ce système ne tient pas compte des similarités de séquence protéique et donc de la structure tridimensionnelle. De plus, cette classification n'est pas systématique quant au mécanisme d'action des enzymes. Elle paraît mal adaptée aux glycosidases qui peuvent agir sur différents substrats et elle ne reflète pas les aspects structuraux des enzymes. C'est notamment le cas de la β -glucosidase BglA, de *Pyrococcus furiosus* DSM 3638. Elle est annotée avec le numéro EC 3.2.1.21, on pourrait penser qu'elle n'a que cette activité, or elle montre des activités β -galactosidasique (EC 3.2.1.23) et β -xylosidasique (EC 3.2.1.37)²¹. Ou encore, il peut se trouver que deux numéros EC différents décrivent la même activité.

3.2. Classification structurale :

3.2.1. Principe et classification :

Avec l'accumulation de séquences des glycosidases (GHs), une nouvelle classification était indispensable car la diversité de propriétés et le comportement de certaines enzymes étaient inexplicables par rapport à la seule référence EC. Ces raisons ont amené *Bernard Henrissat*, à partir de 1989, à chercher une autre alternative pour classer ces enzymes selon leurs propriétés structurales. Il a donc créé une classification, basée sur la similarité de séquences en acides aminés.²² Elle est basée sur la détection de segments structuraux constituant le cœur hydrophobe des protéines globulaires. Grâce à cette approche, des similarités dans le repliement tridimensionnel peuvent être détectées entre les protéines. Les séquences ayant une forte similarité entre elles ont été classées dans les mêmes familles.²³ Cette classification diffère de la classification de l'IUB-MB par le fait que des enzymes de spécificité différente peuvent appartenir à la même famille. Elle permet de prendre en considération la structure tridimensionnelle ainsi que le mécanisme moléculaire

d'action des glycosidases. Elles sont ainsi regroupées en familles (notées GH) en fonction des similarités dans leur séquence d'acides aminés. Certaines glycosidases sont multifonctionnelles. Elles contiennent des domaines catalytiques qui appartiennent à différentes familles des glycosidases.

Puisqu'il existe une relation directe entre similarités de séquence en acides aminés et similarités de repliement ²⁴, il peut être admis que les membres d'une même famille possèdent des repliements similaires. Cette situation permet de prédire les structures tridimensionnelles générales des sites actifs des membres de chaque famille, si cette information est connue pour un ou plusieurs de ses représentants.

Le mécanisme d'action d'une enzyme étant donné par la structure du site actif et par la position des différents groupes fonctionnels, le mécanisme et la stéréochimie de la réaction seront conservés dans un clan (exception faite de la famille GH4). Toutes les glycosidases agissent généralement par un mécanisme qui implique deux résidus importants. Chez ces enzymes, les acides aminés catalytiques sont le plus souvent des aspartates et/ou des glutamates (Tableau 1).

Tableau 1 : Mécanismes réactionnels et acides aminés impliqués dans la catalyse enzymatique des familles des glycosidases classées en clans²¹.

Clan de glycosidases	Famille de glycosidases	Nombre de familles	Mécanisme réactionnel	Acides aminés impliqués dans la catalyse enzymatique ^a
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113, 128	19	réention	Glu Glu
GH-B	7, 16	2	réention	Glu Glu
GH-C	11, 12	2	réention	Glu Glu
GH-D	27, 31, 36	3	réention	Asp Asp
GH-E	33, 34, 83, 93	4	réention	Tyr+Glu

				non connu
GH-F	43, 62	2	inversion	non connu non connu
GH-G	37, 63	2	inversion	non connu non connu
GH-H	13, 70, 77	3	réention	Asp Glu
GH-I	24, 46, 80	3	inversion	non connu ou Asp non connu ou Glu
GH-J	32, 68	2	réention	Asp Glu
GH-K	18, 20, 85	3	réention	oxygène carbonyle Glu
GH-L	15, 65, 125	3	inversion	Glu ou phosphate Glu
GH-M	8, 48	2	inversion	Asp ou non connu Glu
GH-N	28, 49	2	inversion	Asp Asp

a = le premier groupement cité est le composé nucléophile/base, le second est celui du donneur de proton

Cet effort a abouti à la classification de 291 séquences en acides aminés en 35 familles de glycosidases différentes. Cette courte liste de familles a été, par la suite, mise à jour plusieurs fois ²⁵. A l'heure actuelle, 133 familles de glycosidases différentes sont répertoriées et peuvent être consultées sur internet.

Par ailleurs, beaucoup de ces familles, présentant des similarités dans le repliement tridimensionnel des enzymes qu'elles contiennent, ont été regroupées pour former des clans ou super-familles²⁶. On dénombre actuellement 14 clans (Clan GH : Clan des GlycosylHydrolases) appelés A, B, C, D, E, F, G, H, I, J, K, L, M et N (Tableau 1).

Réactualisée en permanence, la classification CAZy (Carbohydrate-Active enZymes) est maintenant étendue à d'autres classes d'enzymes actives sur les sucres : les glycosyltransférases (96 familles), les polysaccharides lyases (23 familles), et les carbohydrates estérases (16 familles).

3.2.2. Répartition des familles des glycosidases hydrolases en sous familles :

Les familles de protéines définies dans CAZy permettent d'avoir un certain pouvoir prédictif : le même repliement, le même mécanisme d'action, et les mêmes résidus catalytiques sont conservés à l'intérieur d'une même famille. Or depuis quelques années, l'équipe CAZy participe à des projets de séquençages de génomes, intervenant au niveau de l'identification et de l'annotation fonctionnelle des gènes de CAZymes.

CAZy est devenue une des principales références en ce qui concerne les CAZymes. Bien que l'inclusion d'une enzyme dans une famille particulière permet de prédire plusieurs caractéristiques structurales et mécanistiques, il n'était pas toujours possible de fournir une prédiction précise sur la fonction de cette enzyme.

Devant l'accumulation de séquences enzymatiques dans une même famille telle que la famille GH 13 (contenant 20902 séquences), *Stam et al.*, dans le but d'améliorer les procédures d'annotation de séquences, ont fourni un effort de division des familles de GHs en sous-familles suivant l'idée que des enzymes possédant des séquences très similaires doivent partager des propriétés biochimiques très proches^{27,28}. La famille GH 13, qui regroupe beaucoup de glycosidases a été subdivisée en 40 sous-familles (Tableau 2). Des α -amylases (EC 3.2.1.1) appartenant à cette même famille ont été rangées dans 13 sous-familles qui sont GH13_1, GH13_2, GH13_5, GH13_6, GH13_7, GH13_15, GH13_19, GH13_24, GH13_27, GH13_28, GH13_32, GH13_36 et GH13_39. Quant aux α -glucosidases (EC 3.2.1.20), elles se retrouvent dans des sous-familles différentes de celles des α -amylases. Il s'agit des sous-familles GH13_17, GH13_21, GH13_30 et GH13_40 (Tableau 2).

Tableau 2 : Division de la famille GH13 en sous familles²¹.

Sous-familles	Numéros EC	Activités enzymatiques
GH 13_1	3.2.1.1	α -amylase
GH 13_2	3.2.1.1	α -amylase
	2.4.1.19 3.2.1.133	cyclodextringlucanotransferase maltogenic α -amylase
GH 13_4	2.4.1.4 ; 3.2.1.-	Amylosucrase sucrose hydrolase
GH 13_5	3.2.1.1	α -amylase
GH 13_6	3.2.1.1	α -amylase
GH 13_7	3.2.1.1	α -amylase

GH 13_8	2.4.1.8	1,4- α -glucanbranching enzyme
GH 13_9	2.4.1.8	1,4- α -glucanbranching enzyme
GH 13_10	3.2.1.141	4- α -(1,4- α -glucano) trehalosetrehalohydrolase
GH 13_11	3.2.1.68	Isoamylase
GH 13_12	3.2.1.41	Pullulanase
GH 13_13	3.2.1.41	Pullulanase
GH 13_14	3.2.1.41	Pullulanase
GH 13_15	3.2.1.1	α -amylase
GH 13_16	5.4.99.16	malto α -glucosyltransférase
GH 13_17	3.2.1.20	α -glucosidase
GH 13_18	2.4.1.7	sucrose phosphorylase
GH 13_19	3.2.1.38	maltohexaose-forming α -amylase
GH 13_20	3.2.1.133	maltogenic α -amylase
GH 13_21	3.2.1.20	α -glucosidase
GH13_22	2.4.1.183	α -1,3-glucan synthase
GH13_23	nd	activité inconnue
GH13_24	3.2.1.1	α -amylase
GH13_25	3.2.1.33	amylo- α -1,6-glucosidase
GH13_26	5.4.99.15	(1,4)- α -glucan 1- α -glucosylmutase
GH13_27	3.2.1.1	α -amylase
GH13_28	3.2.1.1	α -amylase
GH13_29	3.2.1.93	α -phosphotrehalase
GH13_30	3.2.1.20	α -glucosidase
GH13_31	3.2.1.70 3.2.1.10	glucan1,6- α -glucosidase oligo-1.6-glucosidase
GH13_32	3.2.1.1	α -amylase
GH13_33	5.4.99.16	trehalose synthase
GH13_34		aminoacidtransproter
GH13_35		aminoacidtransproter β -amylase
GH13_36	3.2.1.1	α -amylase
GH13_37	nd	activité inconnue
GH13_38	nd	activité inconnue
GH13_39	3.2.1.1 ; 3.2.1.41	α -amylase pullulanase
GH13_40	3.2.1.20	α -glucosidase

4. Inhibition enzymatique :

4.1. Bref rappel sur la cinétique enzymatique :

Il existe plusieurs mécanismes enzymatiques différents pouvant décrire la réaction catalysée par les enzymes. En 1913, *Leonor Michaëlis* et *Maud Menten* ont proposé un modèle simple, dans lequel les enzymes, dites *Michaëlinnes*, fonctionnent selon le mode suivant : un substrat S se lie avec une enzyme E pour donner un intermédiaire ES, puis cet intermédiaire se dissocie pour donner un produit P avec régénération de l'enzyme E^{29,30}.

4.1.1. Equation de *Michaëlis-Menten* :

La modélisation mathématique de la cinétique d'une réaction selon *Michaëlis-Menten* se base sur deux termes : les concentrations des différentes molécules ou complexes de molécules et les constantes d'association et de dissociation de ces molécules entre elles (schéma 8)³¹.



Avec : k_1 = constante d'association de E + S

k_{-1} = constante de dissociation du complexe ES

k_2 = constante de réaction de ES en E + P

k_{-2} = constate d'association de E + P

Schéma 8 : Mécanisme de la réaction enzymatique selon *Michaëlis-Menten*.

L'équation de *Michaëlis-Menten* cherche à établir une expression de la vitesse initiale V_0 de la réaction en fonction de grandeurs connues (fixées par l'expérimentateur ou mesurées).

Pour cela, il faut se placer dans des conditions expérimentales particulières, à savoir : concentration en substrat [S] très largement supérieure à la concentration totale en enzyme $[E]_T$ et absence ou quasi-absence de produit P³². La première condition est obtenue en choisissant des quantités adaptées d'enzyme et de substrat à introduire dans le milieu réactionnel, la seconde en réalisant les mesures suffisamment rapides pour que la quantité de substrat transformé en produit soit faible³³. A partir du moment où ces conditions sont réunies, on peut effectuer les approximations suivantes :

1. Les mesures de cinétique seront toujours faites pour des concentrations de produit très faible : c'est la mesure de la vitesse initiale (v_i) pour $[P] \cong 0$ et $[S] \cong [S]_0$ où $[S]_0$ est la concentration du substrat à l'instant initial. Cela signifie que la vitesse d'apparition du complexe ES par association de E et de P (donnée par $V = K_{-2}[E][P]$) est négligeable.
2. La concentration totale du substrat $[S]_0$ est très grande par rapport à $[E]_T$, la concentration maximale en complexe $[ES]_{\max}$ est limitée par $[E]_T$ et sera donc toujours négligeable comparée à $[S]_0$, même à saturation de tous les sites actifs. Or $[S] = [S]_0 - [ES]$, donc si $[ES]$ est négligeable face à $[S]_0$, on a l'approximation suivante : $[S] = [S]_0$.
3. Dès l'addition de l'enzyme dans la solution de substrat, il s'établit un équilibre rapide entre les formes libres de l'enzyme $[E]$, du substrat $[S]$ et du complexe $[ES]$ (appelé complexe de *Michaëlis*), on parle d'hypothèse de l'état stationnaire. En terme cinétique cela se traduit par $\frac{d[ES]}{dt} = 0$

Après avoir posé ces hypothèses on peut calculer la vitesse d'apparition du produit en résolvant le système d'équations suivant :

$$\begin{cases} V = K[ES] & eq. 1 \\ K_M = \frac{[E][S]}{[ES]} & eq. 2 \\ [S]_0 = [S] + [ES] + [P] & eq. 3 \\ [E]_0 = [E] + [ES] & eq. 4 \end{cases}$$

Avec $K_M = \frac{K_2 + K_{-1}}{K_1}$: la constante de dissociation du complexe $[ES]$ (appelée constante de *Michaëlis-Menten*).

La vitesse d'apparition du produit est donnée par l'équation suivante :

$$V = \frac{K_2[E]_0[S]_0}{K_M + [S]_0} \quad eq. 5$$

Soit $V_{max} = K_2[E]_0$ (vitesse maximale pour $[S]_0 \gg K_M$), la vitesse s'écrit sous la forme :

$$V = \frac{V_{max}[S]_0}{K_M + [S]_0} \quad eq. 6 \quad (\text{Appelée équation de Michaëlis})$$

4.1.2. Détermination de V_{max} et K_M :

Les paramètres cinétiques de la réaction enzymatique V_{max} et K_M sont définis de la façon suivante :

- V_{max} représente la vitesse maximale de la réaction. Elle est atteinte pour une concentration élevée en substrat et traduit la saturation de l'enzyme par le substrat, c'est-à-dire lorsque l'enzyme est entièrement sous forme de complexe (ES)³⁴.
- K_M représente la concentration en substrat pour laquelle la vitesse est égale à la moitié de V_{max} ou pour laquelle la moitié des sites actifs de l'enzyme est occupée. Ainsi plus le K_M est faible et plus l'affinité de l'enzyme pour le substrat est importante³⁴.

Ces paramètres cinétiques peuvent facilement être déterminés :

- en traçant la courbe double inverse selon *Lineweaver* et *Burk* $\frac{1}{V_i}$ en fonction de $\frac{1}{[S]}$. L'équation de la vitesse initiale (eq.6) devient :

$$\frac{1}{V_i} = \frac{K_M + [S]}{V_{max}[S]} \quad eq. 7$$

- En portant sur un graphique les valeurs de $\frac{[S]}{V_i}$ obtenues pour différentes concentrations en substrat $[S]$. Cette représentation graphique est appelée (représentation de *Hanes-Woolf*)³². En effet, on peut transformer l'équation de *Michaelis-Menten* (eq.6) en l'équation suivante :

$$\frac{[S]}{V_i} = \frac{[S]}{V_{max}} + \frac{K_M}{V_{max}} \quad eq. 8$$

4.2. Types d'inhibition enzymatique :

Si la vitesse d'une réaction enzymatique diminue dans des conditions où l'enzyme n'est pas dénaturée, cela signifie que l'enzyme est inhibée³⁵.

De nombreuses substances modifient l'activité d'une enzyme en s'y combinant, ce qui altère la liaison du substrat et/ou sa constante catalytique. Les substances qui diminuent ainsi l'activité d'une enzyme sont appelées des inhibiteurs.

Les inhibiteurs sont généralement des molécules de structure voisine du substrat, qui ne donnent pas de réaction ou réagissent beaucoup plus lentement que le substrat. L'étude de l'effet d'inhibiteur est fréquemment utilisée pour déterminer le mécanisme catalytique d'une réaction enzymatique, de mieux connaître la spécificité d'une enzyme ainsi qu'obtenir des données physiques et chimiques concernant le site actif de l'enzyme³².

Les inhibiteurs d'enzymes peuvent être classés en deux types : inhibiteurs réversibles et inhibiteurs irréversibles.

Les inhibiteurs réversibles s'associent à l'enzyme de manière non covalente, alors que les inhibiteurs irréversibles forment des liaisons covalentes stables avec l'enzyme. L'effet net de l'inhibition correspond à une diminution de la concentration de l'enzyme active³⁶.

4.2.1. Inhibition réversible :

On distingue dans l'inhibition réversible trois catégories principales : les inhibiteurs compétitifs, non compétitifs et incompétitifs.

4.2.1.1. Inhibition compétitive :

Généralement un inhibiteur compétitif ressemble structurellement au substrat à un point tel qu'il se lie au site actif de l'enzyme mais sans donner de réaction. L'inhibiteur (I) entre en compétition directement avec le substrat (S) pour un site actif de l'enzyme (E). La fixation de l'inhibiteur empêche celle du substrat et réciproquement. Leurs fixations sont mutuellement exclusives³⁴. Il y a donc formation de deux complexes (EI) et (ES). La présence du complexe (EI) ralentit la réaction de formation de produit (P) en diminuant la concentration de l'enzyme libre (E)³⁷.

Le mécanisme réactionnel est représenté dans le schéma 9 :

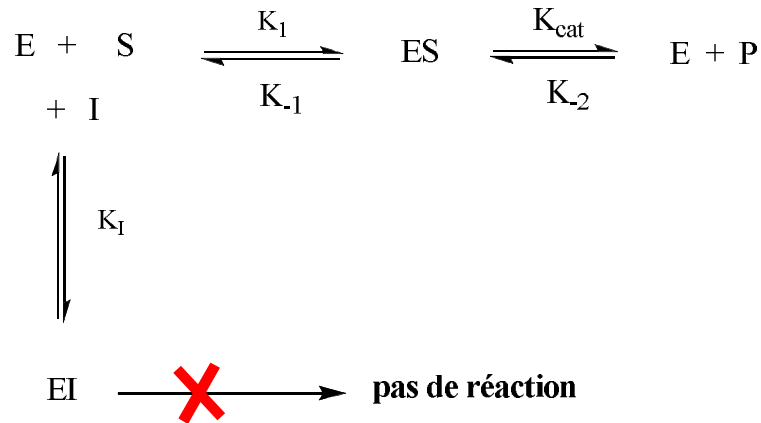


Schéma 9 : mécanisme réactionnel de l'inhibition compétitive.

Différents modèles rendent compte du mécanisme de l'inhibition compétitive.

- a. Le substrat et l'inhibiteur ont le même site de fixation (modèle classique).

Dans ce modèle il n'existe qu'un seul site de fixation pour les deux molécules (substrat et inhibiteur). La fixation exclusive résulte d'une analogie de structure entre le substrat et l'inhibiteur (figure.1 modèle1).

- b. Les sites de fixation du substrat et de l'inhibiteur sont distincts.

Il existe d'autres modèles où le substrat et l'inhibiteur se fixent sur des sites distincts. Malgré la possibilité de fixation de l'inhibiteur et du substrat sur l'enzyme dans des emplacements différents, le mécanisme de l'inhibition est considéré de type compétitif, pour diverses raisons structurales, parmi lesquelles :

- L'encombrement stérique de l'inhibiteur qui empêche la fixation du substrat sur son site actif (figure.1, modèle2),
- Le substrat et l'inhibiteur ont un groupe en commun qui se fixe à l'enzyme sur un troisième site de fixation (figure.1, modèle3),
- Les sites de fixation du substrat et de l'inhibiteur se recouvrent (figure.1, modèle4),
- La fixation de l'inhibiteur, généralement de structure différente de celle du substrat, induit un changement de conformation de l'enzyme qui déforme ou masque le site de fixation du substrat (et inversement) (figure.1, modèle5).

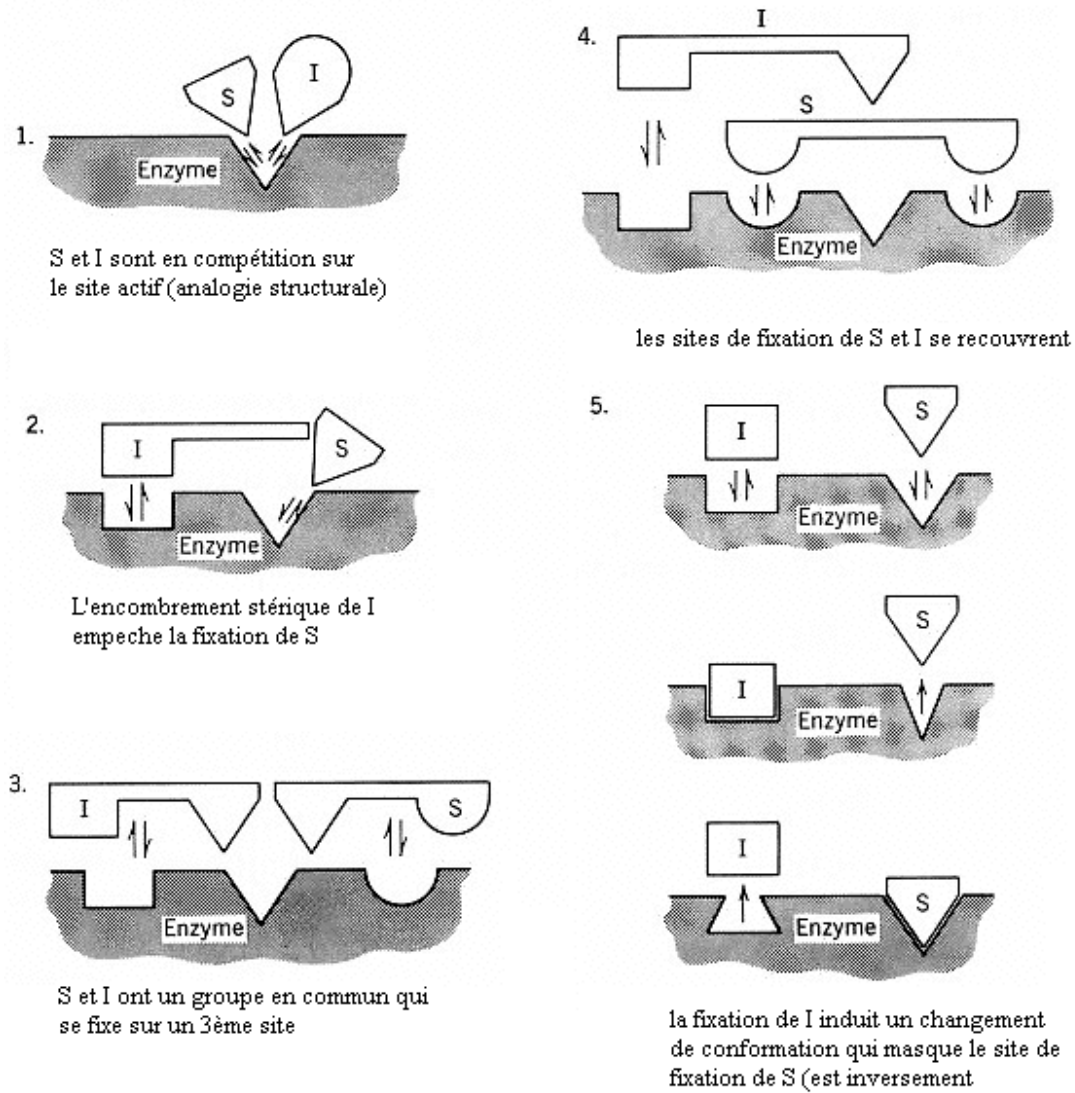


Figure 1 : les différents modèles de l'inhibition compétitive³⁶.

4.2.1.1.1. Cinétique de l'inhibition compétitive :

Si, à une concentration donnée de substrat [S], on appelle V_i^I et V_i , les vitesses initiales mesurées en présence et en absence d'inhibiteur sont respectivement :

$$V_i = \frac{V_{max} \cdot [S]}{K_M + [S]} \quad \text{eq.09}$$

$$V_i^I = \frac{V_{max} \cdot [S]}{K_M^{app} + [S]} \quad \text{eq.10}$$

Avec $K_M^{app} = \left(1 + \frac{[I]}{K_I}\right) \cdot K_M$ eq. 11

Avec :

- V_{max} : la vitesse maximale de réaction, $[I]$ la concentration de l'inhibiteur (I),
- K_I : constante d'inhibition, qui définit l'affinité de l'enzyme pour l'inhibiteur. Plus la valeur de cette constante est faible plus l'inhibiteur est efficace.
- K_M^{app} : constante de *MichaelisMentene* apparente. Cette constante augmente avec la concentration en inhibiteur ce qui ralentit la vitesse de formation de produit de réaction comme montré dans la figure.2 où la valeur du K_M (en absence d'inhibiteur) est accrue d'un facteur de $1 + \frac{[I]}{K_I}$ (présence d'inhibiteur). Quand $[I]$ est nulle, les deux équations (eq.9) et (eq.10) sont identiques.

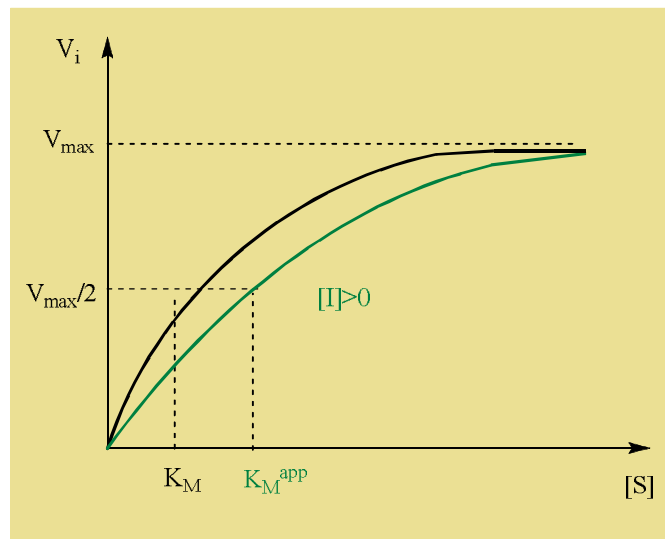


Figure 2 : représentation graphique de la vitesse initiale en fonction de la concentration de substrat.

On peut déterminer expérimentalement les valeurs de la vitesse maximale et la constante de *Michaelis* en traçant la courbe double inverse selon *Lineweaver-Burk* $\frac{1}{v_i} = f\left(\frac{1}{[S]}\right)$ (figure 3) ou selon la courbe de *Hanes-Woolf* $\frac{[S]}{v_i} = f([S])$ (figure 4).

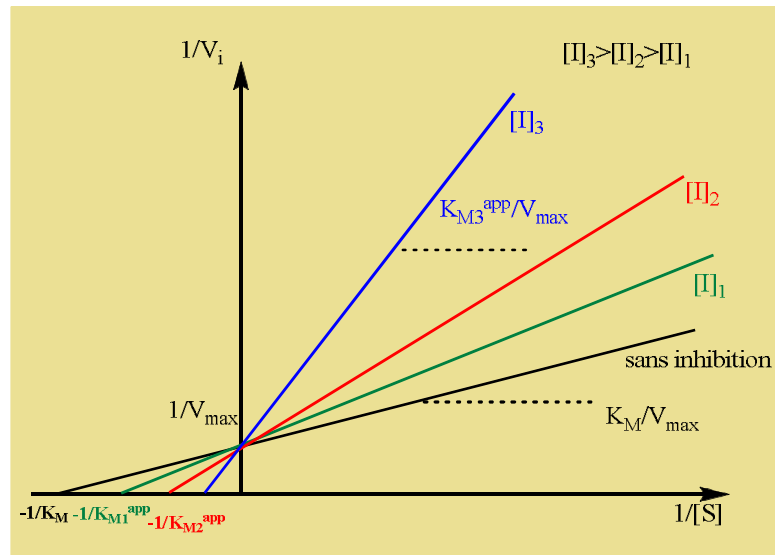


Figure 3 : représentation double inverse $\frac{1}{v_i} = f\left(\frac{1}{[S]}\right)$ selon *Lineweaver-Burk*

La représentation de *Hanes-Woolf* pour différentes concentration en inhibiteur (figure 4) donne plusieurs droites parallèles, qui permettent d'obtenir les différentes constantes de Michaelis K_M^{app} et la vitesse maximale V_{max} . Les droites de pente $\left(\frac{1}{V_{max}}\right)$ coupent l'axe des abscisses au $\left(\frac{1}{K_M^{app}}\right)$.

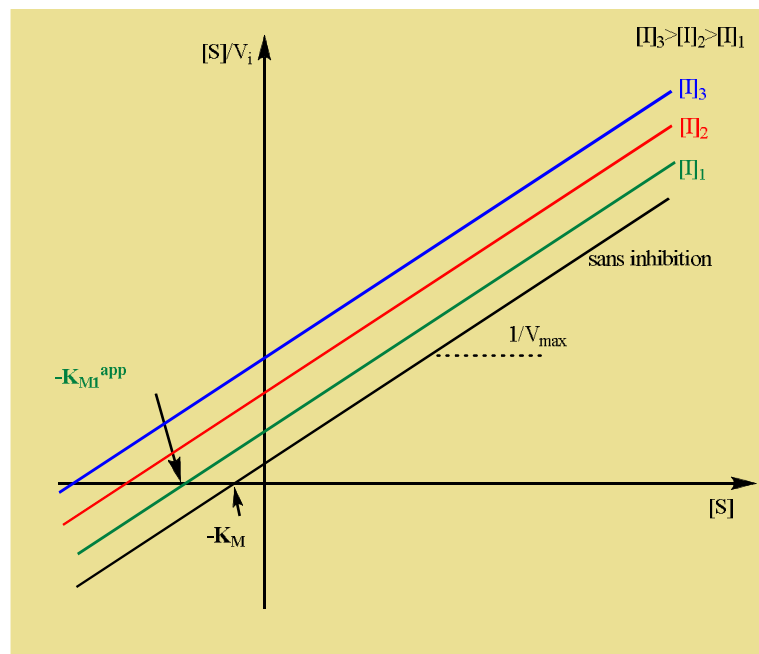


Figure 4 : représentation graphique $\frac{[S]}{v_i} = f([S])$ selon *Hanes-Woolf*.

La détermination expérimentale de la constante d'inhibition K_I est réalisée à partir de la représentation graphique de la constante de *Michaelis* pour différentes concentration d'inhibiteur (I).

$$K_M^{app} = K_M \left(1 + \frac{[I]}{K_I}\right) \quad eq. 12$$

On obtient une droite qui coupe l'axe des abscisses en $(-K_I)$ (figure 5).

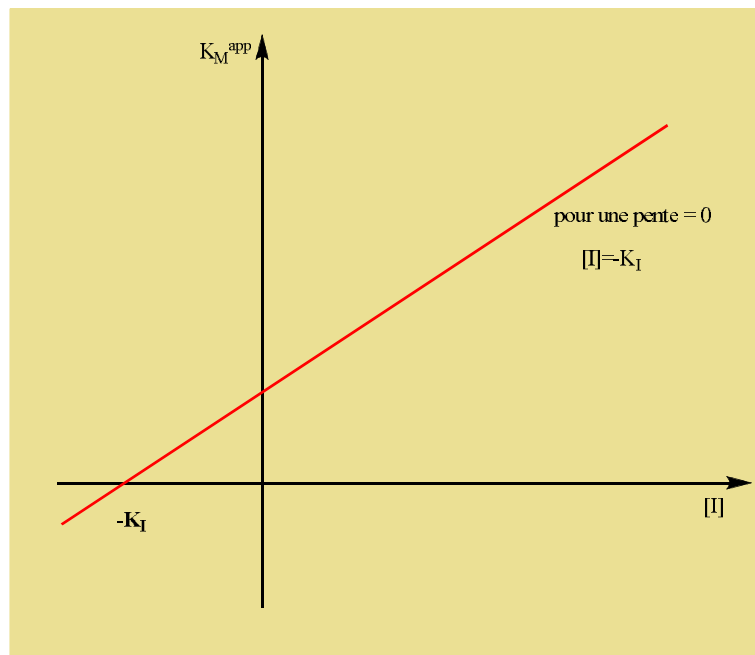


Figure 5 : représentation graphique $K_M = f([I])$ selon *Lineweaver-Burk*.

Il faut noter qu'en présence d'inhibiteur compétitif la constante de *Michaelis* K_M augmente avec l'augmentation de la concentration de l'inhibiteur (I). Le critère décisif de l'inhibition compétitive concerne la vitesse maximale de réaction V_{max} qui ne doit pas être affectée par la présence de l'inhibiteur I. Toutes les droites, dans le graphe double inverse de *Lineweaver-Burk* (figure 3), ont une intersection commune avec l'axe des ordonnées. Cela signifie que l'inhibiteur compétitif peut être libéré à concentration saturante de substrat.

4.2.1.1.2. Activité fractionnaire et degré d'inhibition :

L'activité fractionnaire (a) ou vitesse relative est définie par le rapport des vitesses initiales mesurées en présence et en absence d'inhibiteur³⁸ :

$$a = \frac{V_i^I}{V_i} = \frac{K_M + [S]}{K_M \left(1 + \frac{[I]}{K_I}\right) + [S]} \quad \text{eq. 13}$$

L'activité fractionnaire est souvent exprimée en terme de pourcentage d'inhibition ($i\%$) :

$$i = 1 - a = \frac{[I]}{[I] + K_I \left(1 + \frac{[S]}{K_M}\right)} \quad \text{eq. 14}$$

Le degré d'inhibition induit par un excès d'inhibiteur compétitif est maximal quand la concentration du substrat et la concentration de l'inhibiteur sont beaucoup plus importantes que K_M et K_I respectivement.

4.2.1.1.3. Mesure du paramètre IC₅₀ :

Le paramètre **IC₅₀** représente la concentration d'inhibiteur pour laquelle la vitesse initiale de formation de produit diminue jusqu'à la moitié de sa valeur mesurée en absence d'inhibiteur (eq. 15).

$$IC_{50} = [I] \text{ pour } V_i^I = \frac{V_i}{2} \quad \text{eq. 15}$$

La valeur expérimentale d'**IC₅₀** est obtenue à partir de la courbe $\frac{V_i^I}{V_i} = f([I])$ (figure.6).

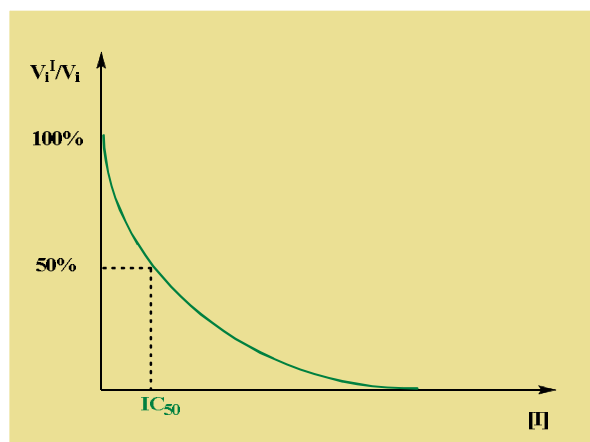


Figure 6 : détermination expérimentale d'**IC₅₀** à partir de la courbe $\frac{V_i^I}{V_i} = f([I])$.

Pour une inhibition compétitive, l'expression qui détermine la valeur d' IC_{50} est la suivante:

$$IC_{50} = \frac{K_I}{K_M} [S] + K_I \quad \text{eq. 16}$$

Pour $[S] = K_M$ $IC_{50} = 2K_I$

4.2.1.2. Inhibition non compétitive :

Dans le cas d'inhibition non compétitive, l'inhibiteur n'a aucune influence sur la fixation du substrat (et réciproquement), les sites de fixation du substrat et de l'inhibiteur sont distincts. En conséquence, l'inhibiteur se fixe de façon réversible à la fois sur l'enzyme (E) et sur le complexe enzyme-substrat (ES), de même le substrat se fixe à l'enzyme libre (E) et au complexe enzyme-inhibiteur (EI). Les inhibiteurs non compétitifs n'ont pas d'homologie structurale avec le substrat³⁶.

Il existe deux types d'inhibition non compétitive : l'inhibition non compétitive pure et l'inhibition non compétitive mixte.

4.2.1.2.1. Inhibition non compétitive pure :

Dans ce cas, la fixation de l'inhibiteur ne modifie pas la manière dont se fixe le substrat, mais empêche les ajustements conformationnels du site actif nécessaires pour la catalyse. Dans ce cas les valeurs de K_I et K_I' sont égales (figure 7)³⁹.

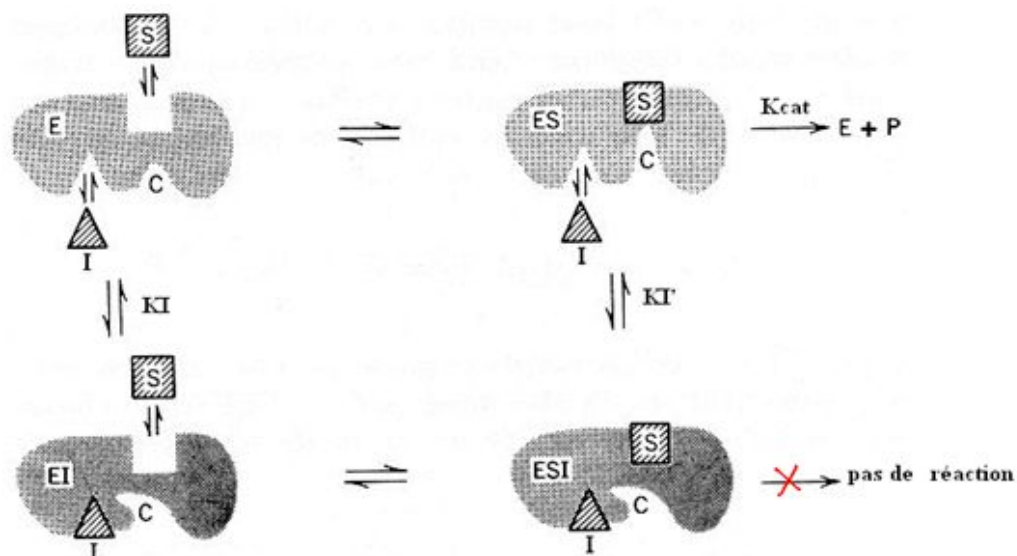


Figure 7 : mécanisme d'inhibition non compétitive pure³⁶.

En présence d'inhibiteur non compétitif, la représentation graphique de la vitesse initiale de réaction V_i^I (eq.09) en fonction de la concentration du substrat montre que la valeur de la constante de Michaelis K_M n'est pas modifiée. Cela peut signifier que l'affinité de l'enzyme libre (E) et celle du complexe (EI) pour le substrat (S) ne sont pas modifiées, par contre la vitesse maximale V_{max} a diminué (figure 8).

$$V_i^I = \frac{V_{max}}{1 + \frac{[I]}{K_I}} \cdot \frac{[S]}{K_M + [S]} \quad eq. 17$$

Soit
$$V_i^I = V_{max}^{app} \cdot \frac{[S]}{K_M + [S]} \quad eq. 18$$

Avec
$$V_{max}^{app} = \frac{V_{max}}{1 + \frac{[I]}{K_I}} \quad eq. 19$$

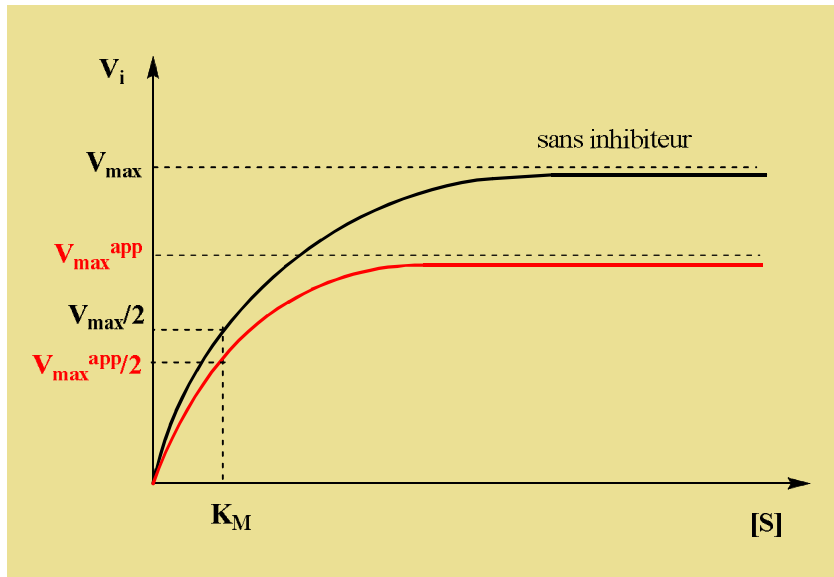


Figure 8 : représentation graphique de $V_i^I = f([S])$ dans le cas d'inhibition non compétitive pure.

Les paramètres cinétiques K_M et V_{max} peuvent être déterminés expérimentalement en traçant la courbe double inverse selon *Lineweaver-Burk* $\frac{1}{V_i} = f\left(\frac{1}{[S]}\right)$ (figure 9). En présence de différentes concentrations de l'inhibiteur (I), on obtient des droites qui coupent l'axe des abscisses sur le point $\left(-\frac{1}{K_M}\right)$ et coupent l'axe des ordonnées en différents points représentant l'inverse des vitesses maximales apparentes $\left(\frac{1}{V_{max}^{app}}\right)$. Plus la concentration de l'inhibiteur (I) augmente plus V_{max} diminue.

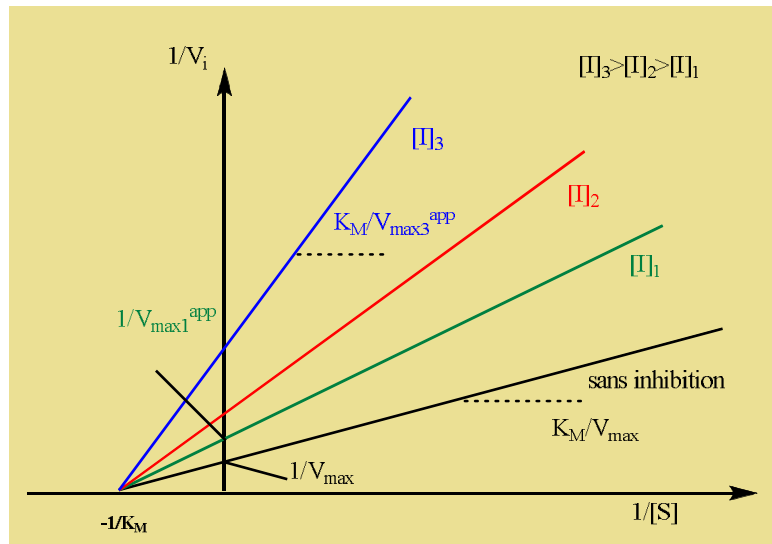


Figure 9 : représentation graphique double inverse selon *Lineweaver-Burk* pour l'inhibition non compétitive pure.

4.2.1.2.1.1. Détermination du Degré d'inhibition et de l'IC₅₀ :

Dans le cas de l'inhibition non compétitive pure, le degré d'inhibition (i) et le paramètre IC₅₀ sont exprimés par les équations (eq.20) et (eq.21) respectivement :

$$i = \frac{[I]}{K_I + [I]} \quad \text{eq. 20}$$

$$IC_{50} = K_I \quad \text{eq. 21}$$

4.2.1.2.2. Inhibition non compétitive mixte :

Dans ce cas, la liaison de l'inhibiteur (I) avec l'enzyme (E) influence la liaison du substrat (S) avec l'enzyme. Les valeurs de K_I et K_I' ne sont pas égales. K_M et V_{max} sont altérées par la présence de l'inhibiteur (I) et ne sont pas constantes³⁶ (figure 10 et figure 11).

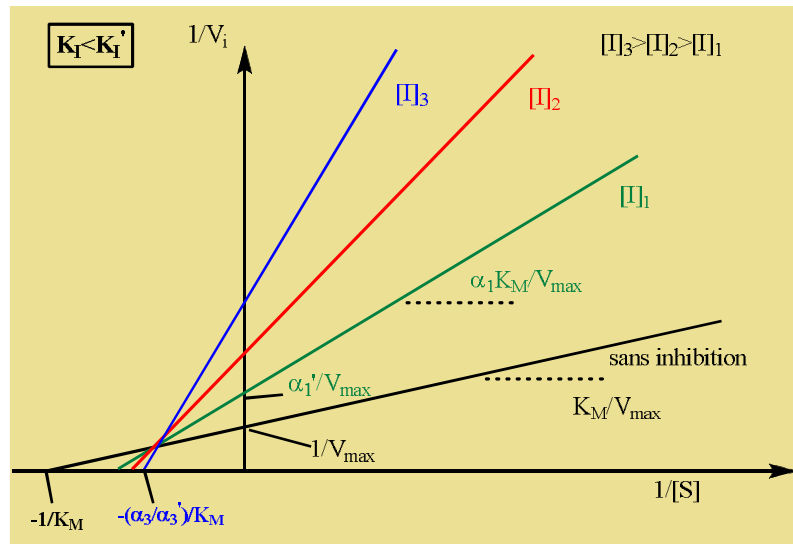


Figure 10 : tracé double inverse de *Lineweaver-Burk* dans le cas d'inhibition non compétitive mixte (pour $K_I < K_I'$).

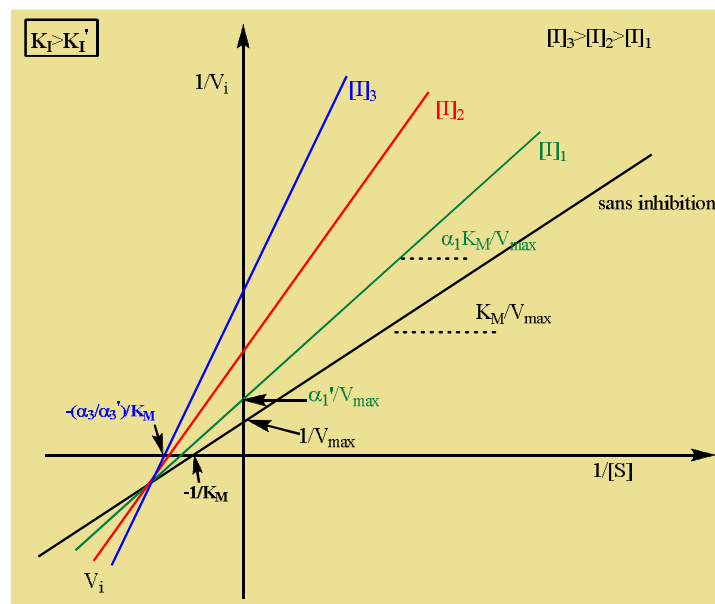


Figure 11 : tracé double inverse de *Lineweaver-Burk* dans le cas d'inhibition non compétitive mixte (pour $K_I > K_I'$).

Dans le cas d'une inhibition mixte la vitesse initiale est définie par l'équation suivante :

$$V_i^I = \frac{V_{max}}{\alpha'} \times \frac{[S]}{\frac{\alpha}{\alpha'} K_M + [S]} \quad eq.22$$

Avec $\alpha = 1 + \frac{[I]}{K_I}$ et $\alpha' = 1 + \frac{[I]}{K_I'}$

On peut écrire l'équation (eq.22) sous la forme :

$$V_i^I = V_{max}^{app} \cdot \frac{[S]}{K_M^{app} + [S]} \quad eq. 23$$

Avec : $V_{max}^{app} = \frac{V_{max}}{\alpha'}$ et $K_M^{app} = \frac{\alpha}{\alpha'} K_M$

Dans la représentation graphique double inverse de *Lineweaver-Burk* (figures 11 et 12), les points d'intersection des droites avec l'axe des ordonnées et avec l'axe des abscisses représentent les termes $\frac{\alpha'}{V_{max}}$ et $\frac{\alpha}{\alpha' K_M}$ respectivement.

4.2.1.2.2.1. Détermination du degré d'inhibition (*i*) et de l'IC₅₀ :

Le degré d'inhibition (*i*) et le paramètre IC₅₀ pour une inhibition non compétitive mixte sont définis ci-dessous par les équations (eq.24) et (eq.25) respectivement :

$$i = 1 - \frac{K_M + [S]}{\alpha K_M + \alpha' [S]} \quad eq. 24$$

$$IC_{50} = \frac{K_M + [S]}{\frac{K_M}{K_I} + \frac{[S]}{K_I'}} \quad eq. 25$$

4.2.1.2.2.2. Modèles d'inhibition non compétitive mixte :

Il existe deux modèles alternatifs d'inhibition non compétitive mixte (figure 12). Le modèle 1 suggère qu'il n'y a pas de passage conformationnel possible entre les deux complexes enzyme-substrat (ES) et enzyme-substrat-inhibiteur (ESI), c'est-à-dire que le complexe enzyme-inhibiteur (EI) peut fixer le substrat (S) pour donner le complexe inactif (ESI), mais le complexe (ES) ne peut pas fixer l'inhibiteur (I). Le modèle 2 suggère qu'il n'y a pas d'équilibre entre les deux complexes (EI) et (ESI), c'est-à-dire que le complexe (ES) peut fixer l'inhibiteur (I) pour donner le complexe inactif (ESI), en revanche le substrat (S) ne peut pas être fixé sur le complexe(EI)³².

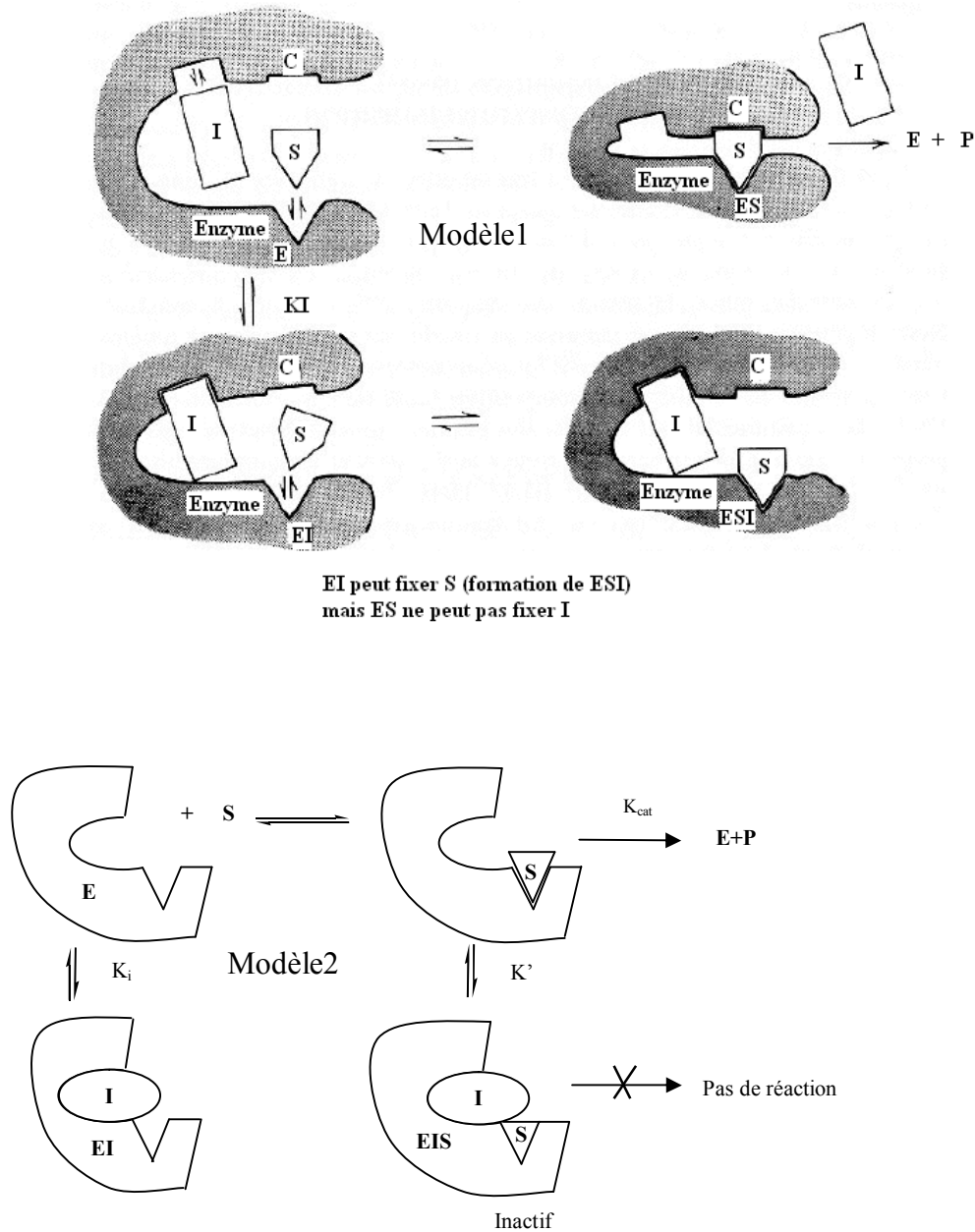


Figure 12 : modèles d'inhibition non compétitive mixte³⁶

4.2.1.3. Inhibition incompétitive :

Ce type d'inhibition est aussi appelé inhibition par blocage du complexe intermédiaire, l'enzyme et le substrat forment d'abord le complexe enzyme-substrat (complexe intermédiaire), puis l'inhibiteur se fixe à ce complexe. Il y a formation d'un complexe ternaire (ESI) inactif (schéma 10)⁴⁰.

$$V_i^I = V_{max}^{app} \times \frac{[S]}{K_M^{app} + [S]} \quad eq. 27$$

Avec : $V_{max}^{app} = \frac{V_{max}}{1 + \frac{[I]}{K_I}}$ et $K_M^{app} = \frac{K_M}{1 + \frac{[I]}{K_I}}$

Les différents paramètres cinétiques, vitesse maximale, constante de *Michaelis* et constante d'inhibition, sont déterminés à partir de la représentation graphique double inverse selon *Lineweaver-Burk* (figure 14).

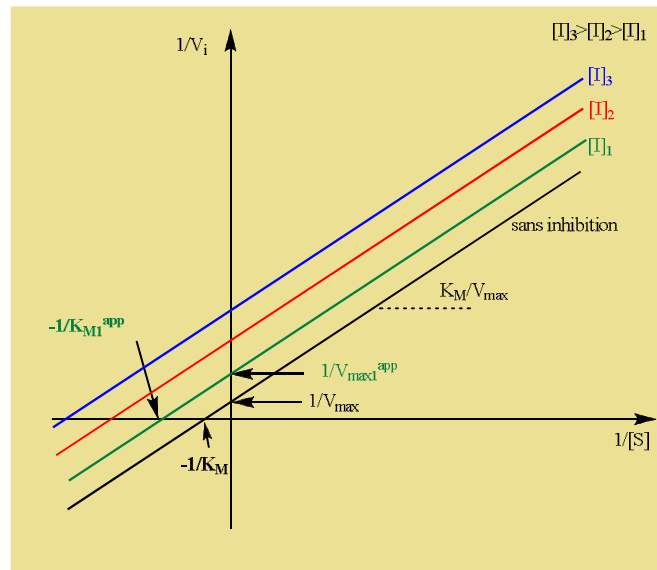


Figure 14 : représentation graphique double inverse selon *Lineweaver-Burk* pour inhibition incompétitive.

On obtient, selon le graphe, des droites parallèles de pente égale à $\left(\frac{K_M}{V_{max}}\right)$, qui coupent les axes des ordonnées et des abscisses aux points $\left(\frac{1 + \frac{[I]}{K_I}}{V_{max}}\right)$ et $\left(-\frac{1 + \frac{[I]}{K_I}}{K_M}\right)$ respectivement.

4.2.1.3.1. Degré d'inhibition et IC₅₀ :

Le degré d'inhibition (*i*) et le paramètre IC₅₀ pour une inhibition incompétitive sont donnés par les équations (eq.28) et (eq.29) respectivement :

$$i = \frac{[I][S]}{K_I K_M + (K_I + [I])[S]} \quad eq. 28$$

$$IC_{50} = \frac{K_I K_M}{[S]} + K_I \quad eq.29$$

$$(IC_{50} = 2K_I \quad \text{pour } [S] = K_M)$$

En présence de ce type d'inhibiteurs, les valeurs des deux paramètres V_{max} et K_M diminuent avec l'augmentation de la concentration de l'inhibiteur. Cela est expliqué par la diminution de concentration en enzyme active à cause de la formation du complexe inactif (ESI).

Cette inhibition ne peut être levée par un excès de substrat. En effet, plus il y a de substrat, plus il y a formation de complexe (ES) et plus l'équilibre de fixation de l'inhibiteur (I) est déplacé en faveur du complexe (ESI).

L'ensemble des modifications des paramètres cinétiques en présence d'inhibiteurs réversibles sont résumées dans le tableau 3.

Tableau 3 : résumé des modifications des paramètres cinétiques en présence d'inhibiteurs réversibles.

Type d'inhibition	Effet sur V_{max}	V_{max}^{app}	Effet sur K_M	K_M^{app}	Vitesse initiale V_i	Degré d'inhibition (i)	IC_{50}
Sans	-----	-----	-----	-----	$V_i = \frac{V_{max} \cdot [S]}{K_M + [S]}$	-----	-----
Compétitive	-----	V_{max}	augmentation	αK_M	$V_i^I = \frac{V_{max} \cdot [S]}{\alpha K_M + [S]}$	$i = \frac{[I]}{[I] + K_I(1 + \frac{[S]}{K_M})}$	$IC_{50} = \frac{K_I}{K_M}[S] + K_I$
Non compétitive pure $K_I=K_I'$	Diminution	$\frac{V_{max}}{\alpha'}$	Diminution	$\frac{K_M}{\alpha'}$	$V_i = \frac{\frac{V_{max}}{\alpha'} \cdot [S]}{\frac{K_M}{\alpha'} + [S]}$	$i = \frac{[I]}{K_I + [I]}$	$IC_{50} = K_I$
Non compétitive mixte avec $K_I > K_I'$	Diminution	$\frac{V_{max}}{\alpha}$	-----	K_M	$V_i = \frac{\frac{V_{max}}{\alpha} \cdot [S]}{K_M + [S]}$	$i = 1 - \frac{K_M + [S]}{\alpha K_M + \alpha' [S]}$	$IC_{50} = \frac{K_M + [S]}{\frac{K_M}{K_I} + \frac{[S]}{K_I'}}$
Non compétitive mixte avec $K_I < K_I'$	Diminution	$\frac{V_{max}}{\alpha'}$	Diminution	$\frac{\alpha K_M}{\alpha'}$	$V_i = \frac{\frac{V_{max}}{\alpha'} \cdot [S]}{\frac{\alpha K_M}{\alpha'} + [S]}$	$i = 1 - \frac{K_M + [S]}{\alpha K_M + \alpha' [S]}$	$IC_{50} = \frac{K_M + [S]}{\frac{K_M}{K_I} + \frac{[S]}{K_I'}}$
Incompétitive	Diminution	$\frac{V_{max}}{\alpha'}$	augmentation	$\frac{\alpha K_M}{\alpha'}$	$V_i = \frac{\frac{V_{max}}{\alpha'} \cdot [S]}{\frac{\alpha K_M}{\alpha'} + [S]}$	$i = \frac{[I][S]}{K_I K_M + (K_I + [I])[S]}$	$IC_{50} = \frac{K_I K_M}{[S]} + K_I$

$$\alpha = 1 + \frac{[I]}{K_I} \text{ et } \alpha' = 1 + \frac{[I]}{K_I'}$$

4.2.2. Inhibition irréversible :

L'action d'un inhibiteur est irréversible quand il se forme une liaison covalente entre l'enzyme et l'inhibiteur. Ce type d'inhibiteur est appelé inactivateur³⁷.

La cinétique de la réaction est similaire à celle d'une réaction en présence d'un inhibiteur non compétitif car il y a perte d'enzyme actif (schéma 11). Mais, contrairement à ce qui se passe dans l'inhibition réversible, la dilution ou la dialyse de la solution contenant le complexe enzyme-inhibiteur (EI) ne restaure pas l'activité enzymatique car le complexe ne peut pas se dissocier.

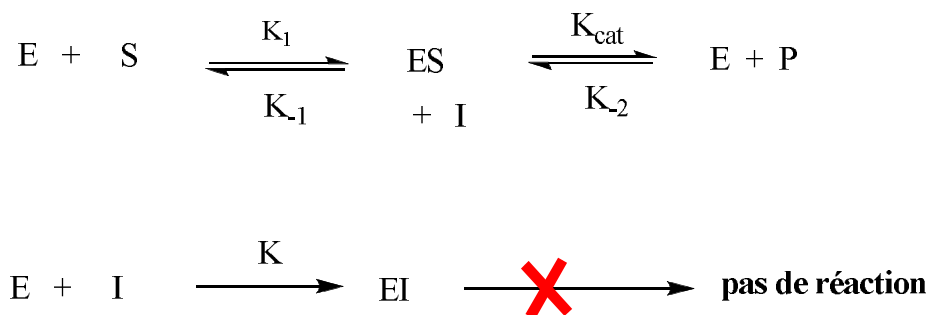


Schéma 11 : mécanisme d'inhibition irréversible.

L'étude de l'effet des inhibiteurs irréversibles est souvent utilisée pour déterminer les groupes actifs du site catalytique.

Un exemple d'inactivateur est le gaz neurotoxique, employé dans la deuxième guerre mondiale, le di-isopropylfluorophosphate (DFP).

L'étude de l'effet de ce composé sur l'activité des protéases à sérine a permis d'identifier la sérine 195 et l'histidine 57, les deux résidus impliqués dans la catalyse. La sérine 195 attaque le groupe fluorophosphate du DFP en formant une liaison covalente (phosphoester) qui rend cette inhibition irréversible (schéma 12)⁴¹.

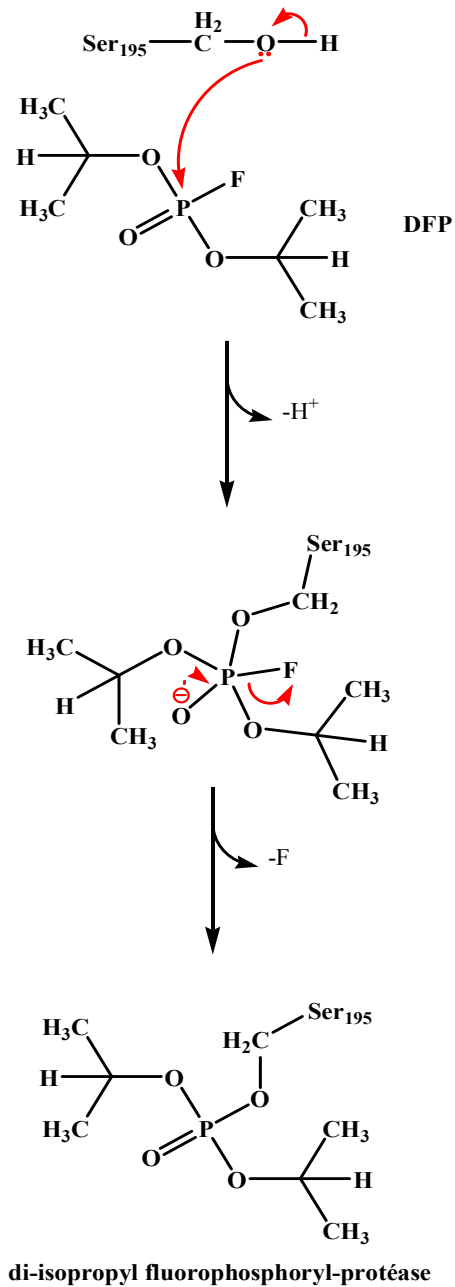


Schéma 12 : mécanisme d'inhibition de la protéase par le di-isopropylfluorophosphate.

4.3. Inhibiteurs des glycosidases :

La mise en œuvre d'inhibiteurs de glycosidases est actuellement un objectif de choix pour les chercheurs, notamment les chimistes, les biochimistes et les pharmaciens. Ces inhibiteurs, isolés ou synthétisés, ont souvent des structures mimant le substrat naturel de l'enzyme ou son état de transition. Ces molécules peuvent présenter plusieurs types d'activité pharmacologiques en tant qu'agents antidiabétiques, antiviraux, anti-malaria et anticancéreux.

Dans le reste de ce chapitre nous décrirons quelques inhibiteurs de glycosidases représentatifs de différentes familles.

4.3.1. Aminocyclitols polyhydroxylés :

Ces composés sont analogues du sucre, possédant un atome d'azote exocyclique protoné. Parmi les nombreux médicaments antidiabétiques, l'acarbose (figure 15) est un inhibiteur puissant d'enzymes digestives : l' α -amylase, α -glucosidase, la sucrase et la maltase ($IC_{50}=0,5\mu M$, saccharase)¹. Isolé dans les années 70 d'une souche bactérienne *Actinoplanes sp. EE50*⁴², l'acarbose est le plus largement utilisé pour le traitement du diabète type 2^{43,44}, commercialisé par la société Bayer sous le nom Glucor ou Glycobay.

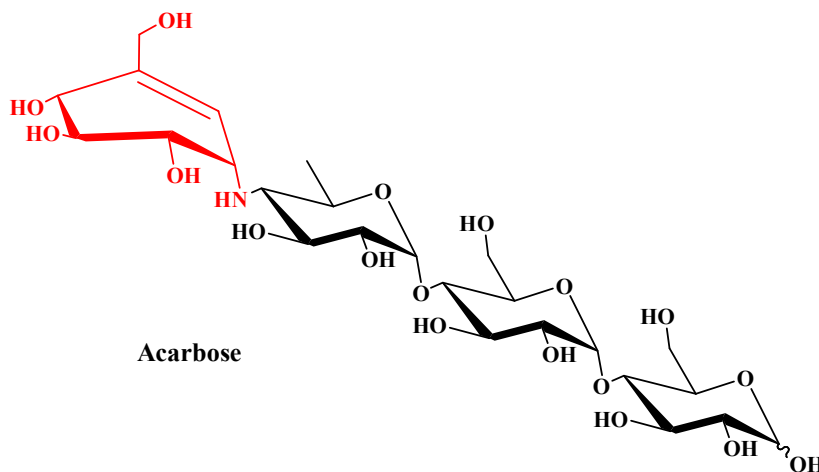


Figure 15 : structure chimique de l'acarbose.

La similarité structurale de l'acarbose à des oligosaccharides, en raison de ses résidus de glucose, est considérée comme étant responsable de la liaison de haute affinité pour les sites actifs des α -glucosidases. L'incapacité des enzymes digestives à hydrolyser l'acarbose est due à la présence d'un pont amine, qui est considéré le facteur déterminant de son effet inhibiteur.

L'unité (A) notée en rouge sur l'acarbose (figure 15), appelée valiénamine (figure 16), est également un bon inhibiteur d' α -glucosidases intestinales⁴⁵. Ce composé ainsi que la valioline (figure 16) ont été isolés en 1984 d'une souche de *Streptomyces lygrosopicus*⁴⁶. Le voglibose (figure 16) préparé à partir de la valioline (figure 16) est l'un des plus puissants inhibiteurs d' α -glucosidases (IC_{50} = 0,0158 μ M, maltase et IC_{50} = 0,0046 μ M, saccharase). Il est actuellement commercialisé par la société Japonaise Takeda sous le nom Basen pour le traitement du diabète type 2.

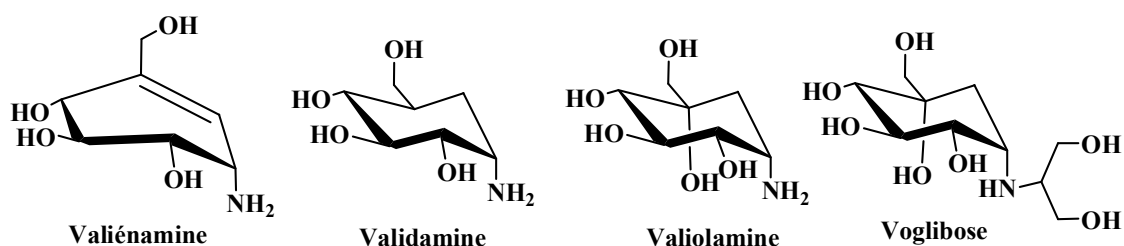


Figure 16 : structures chimiques de Valiénamine, Validamine, Valiolamine et de Voglibose.

4.3.2. Iminosucres polyhydroxylés :

Le remplacement du cycle oxygéné (pyranose, furanose) des structures monosaccharidiques par un cycle azoté (pipéridine, pyrrolidine, indolizidine) conduit à une classe « glycomimétiques » particulièrement intéressante : les iminosucres. Les composés de cette classe possèdent un effet inhibiteur puissant sur les glycosidases, ils miment le substrat de ces enzymes, du fait de la protonation de la fonction amine dans les conditions de pH voisin de neutralisation, en portant une charge positive qui entraîne une forte interaction avec les fonctions carboxylates des résidus du site actif. De plus, l'espèce chargée positivement est un mime de l'intermédiaire oxocarbenium impliqué dans le mécanisme d'hydrolyse⁴⁷.

Le premier azasucré naturel découvert a été la nojimyine (figure 17) qui mime le D-glucose. Il a été isolé en 1966 par l'équipe d'*Inoye*.⁴⁸ Cet azasucré naturel, isolé d'une souche bactérienne *Streptomyces*, en plus de ses propriétés antibiotiques, inhibe les α - et les β -glycosidases. Cependant ce composé est instable à cause de la présence du

groupement hydroxyle porté par le carbone C₁.⁴⁹ En 1967, l'équipe de Paulsen⁵⁰ a donc proposé la synthèse de désoxynojirimicine. Grâce à cette molécule, deux médicaments ont été mis sur le marché en tant qu'inhibiteurs de glycosidases : le Miglitol⁵¹ et le Zavesca/Miglustat⁵². Le Miglitol, qui possède une chaîne hydroxy éthyle sur l'atome d'azote, est prescrit pour le traitement du diabète type 2. L'avesca, analogue du miglitol, possède une chaîne butyle. Il permet de combattre la maladie de *Gaucher* en inhibant la glycosyltransférase.

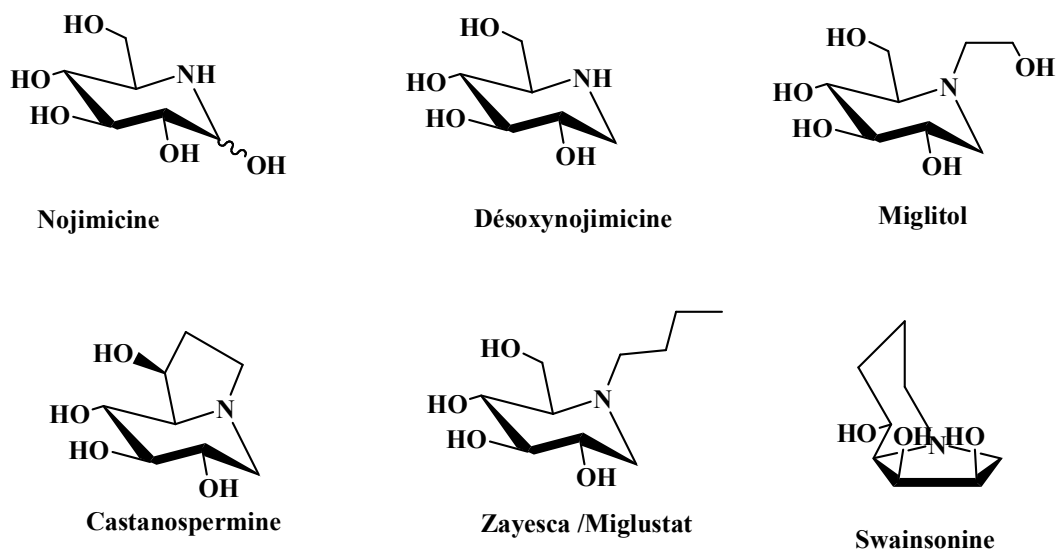


Figure 17 : structures chimiques de quelques iminosucres inhibiteurs de glycosidases.

La découverte puis la synthèse d'indolizidines a permis également de développer des inhibiteurs de glycosidases. Le swainsonine et le castanospermine sont deux exemples de molécules possédant un cycle à six chaînons et un cycle à cinq chaînons fusionnés. Le swainsonine, isolé de *Swainsona canescens*⁵³ en 1979 et synthétisée en 1984,⁵⁴ est un bon inhibiteur d' α -mannosidase. Le castanospermine, isolé de *Castanospermum australe*⁵⁵ en 1981 et synthétisée plus tard⁵⁶, est un inhibiteur d' α -et β -glucosidases. D'après des travaux publiés en 2005,⁵⁷ le castanospermine pourrait se révéler très intéressant contre la maladie de *dengue*. En effet, en inhibant une α -glucosidase de réticulum endoplasmique, il empêcherait la propagation du virus.

4.3.3. Polyphénols :

Les polyphénols constituent une famille de molécules organiques largement présente dans le règne végétal. Ils sont caractérisés, comme l'indique leur nom, par la présence de plusieurs groupements phénoliques associés en structures plus ou moins complexes,

généralement de haut poids moléculaire. Ces composés sont les produits du métabolisme secondaire des plantes.

Les polyphénols prennent une importance croissante, notamment grâce à leurs effets bénéfiques sur la santé⁵⁸. En effet, leur rôle d'antioxydants naturels suscite de plus en plus d'intérêt pour la prévention et le traitement du cancer⁵⁹, des maladies inflammatoires⁶⁰, des maladies cardiovasculaires^{61,62} et dans le traitement de troubles métaboliques comme l'obésité et le diabète⁶³. Ils sont également utilisés comme additifs pour l'industrie agroalimentaire, pharmaceutique et cosmétique⁶⁴.

Les polyphénols contiennent au moins un cycle aromatique portant un ou plusieurs groupes hydroxyles en plus d'autres substituants, ils peuvent être divisés en 15 grandes classes en fonction de leurs structures chimiques⁶⁵. Les plus importantes classes de polyphénols sont les tanins, les flavonoïdes, les lignanes, les stilbènes et les acides phénoliques tels que les acides hydroxybenzoïques, l'acide gallique, acide caféique et l'acide coumarique (Figure.18).

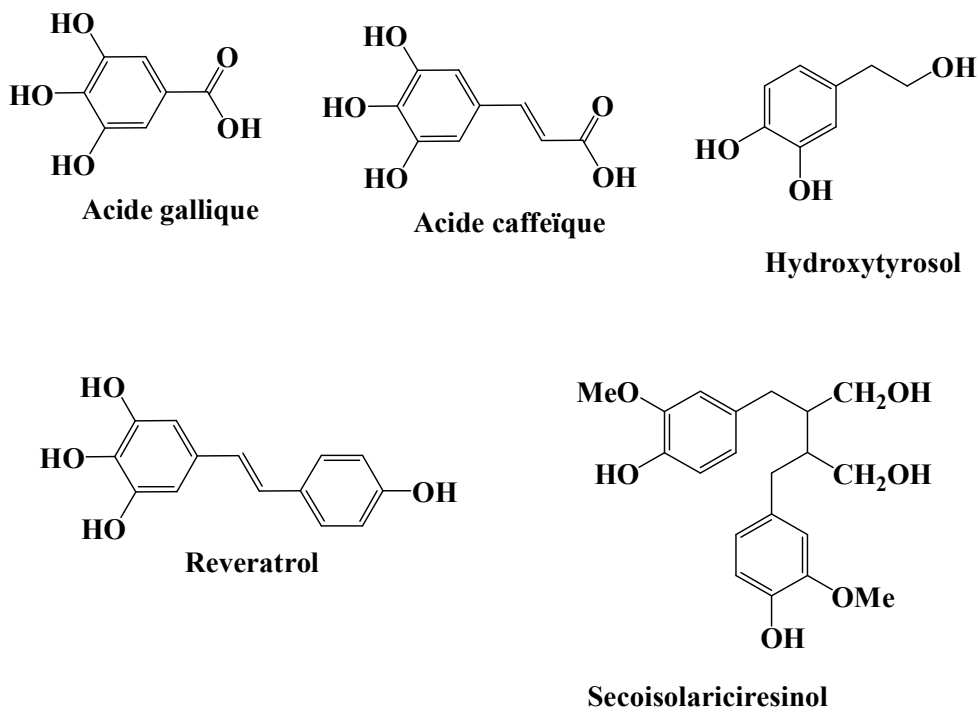


Figure 18 : structures chimiques des polyphénols inhibiteurs de glycosidases.

4.3.3.1. Flavanoïdes :

Les flavanoïdes sont les principaux polyphénols présents dans une grande variété de sources végétales⁶⁶⁻⁷¹. Leurs structures sont représentées par un noyau benzénique, condensé avec une pyranne hétérocyclique à six chaînons ou un noyau pyrone, qui en position 2 ou 3 porte un anneau de phényle comme produit de remplacement (figure 19). Plus de 10 000 flavanoïdes séparés et identifiés à partir de plantes sont divisées en sous-classes, tels que les anthocyanes, les flavonols, les flavanones, les flavones et isoflavones⁷². Les flavanoïdes présents dans les végétaux ont suscité un grand intérêt depuis les années 90 en raison de leur effet bénéfique sur la santé humaine notamment dans le traitement de troubles métaboliques comme l'obésité et le diabète⁷³. Plusieurs études ont été rapportées dans la littérature sur le pouvoir inhibiteur de ces composés et de leurs dérivés sur les glycosidase. Le vitexine et l'isovitexine, isolés à partir de l'extrait éthanolique du *haricot Azuki (vignaangularis)*⁷⁴ présentent un pouvoir inhibiteur sur les α -glucosidases intestinales. Ces deux composés portent des fragments glycosidiques, permettant une interaction avec les sites actifs d'enzymes. Le saponarine, isolé à partir des feuilles de *Tinosporacordifolia* a montré un effet inhibiteur sur l' α -glucosidase et sur la sucrase de différentes origines⁷⁵. Les valeurs d'IC₅₀ mesurées sur l' α -glucosidase de champignons, sur l'invertase de levure, sur la maltase intestinale de rat et sur la sucrase sont 55, 48, 48 et 35 mM respectivement. Un autre composé, le coumestrol, identifié dans les feuilles de soja, a montré un effet inhibiteur puissant sur l' α -glucosidase de levure avec une valeur d'IC₅₀ égale à 6 mM⁷⁶. Cependant, aucune donnée n'a été signalée prouvant son pouvoir inhibiteur sur les α -glucosidases de mammifères. Par contre, le dolichandroside et l'aloesaponrine, isolés à partir de *dolichanronefalcata* ont montré une activité inhibitrice à la fois sur l' α -glucosidase intestinale de rat et sur celle de levure⁷⁷. Le dolichandroside et l'aloesaponrine (figure 19) possèdent un pouvoir inhibiteur sur l' α -glucosidase de levure avec des valeurs d'IC₅₀ égales à 61 et 71,3 mM respectivement et sur l' α -glucosidase intestinale de rat avec des valeurs d'IC₅₀ égales à 28,7 et de 34,5 mM respectivement.

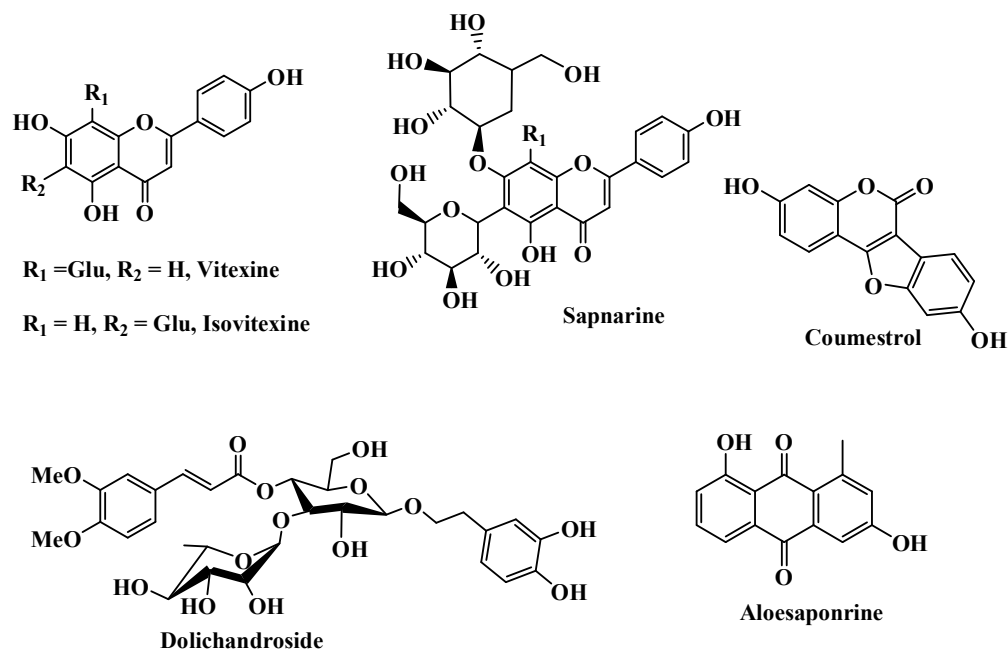


Figure 19 : structures chimiques de quelques inhibiteurs d'origine naturelle dérivés de flavanoides.

4.3.3.2. Tanins :

Les tanins sont considérés parmi les produits naturels les plus étudiés pour leur activité comme agents piègeurs de radicaux, agents anti-inflammatoires et inhibiteurs d'enzymes telles que l' α -glucosidase et l' α -amylase. Les tanins hydrolysables et condensés (figure 20), répartis en deux grandes classes de composés polyphénoliques, sont capables d'avoir des interactions avec les enzymes à travers des liaisons non spécifiques et donc souvent sélectionnés par le criblage virtuel des produits naturels visés pour les études des activités inhibitrices d'enzymes. Certains tanins hydrolysables présentent une activité inhibitrice puissante sur l' α -glucosidase. Il s'agit notamment des tanins des grains de *Eugenia Jambolana*⁷⁸ et rugosine D, rugosine A et Tellimagrandine II. Ces composés présentent des activités inhibitrices sur des α -glucosidases microbiennes plus élevée que celles sur les α -glucosidases intestinales de mammifères⁷⁹. Les tanins condensés ou proanthocyanidines oligomériques se sont avérés être des inhibiteurs puissants de l' α -glucosidase. Leur activité inhibitrice est plus élevée avec un haut degré de polymérisation, ce qui est cohérent avec l'idée que l'inhibition peut être due aux liaisons non spécifiques de ces composés avec les protéines⁸⁰.

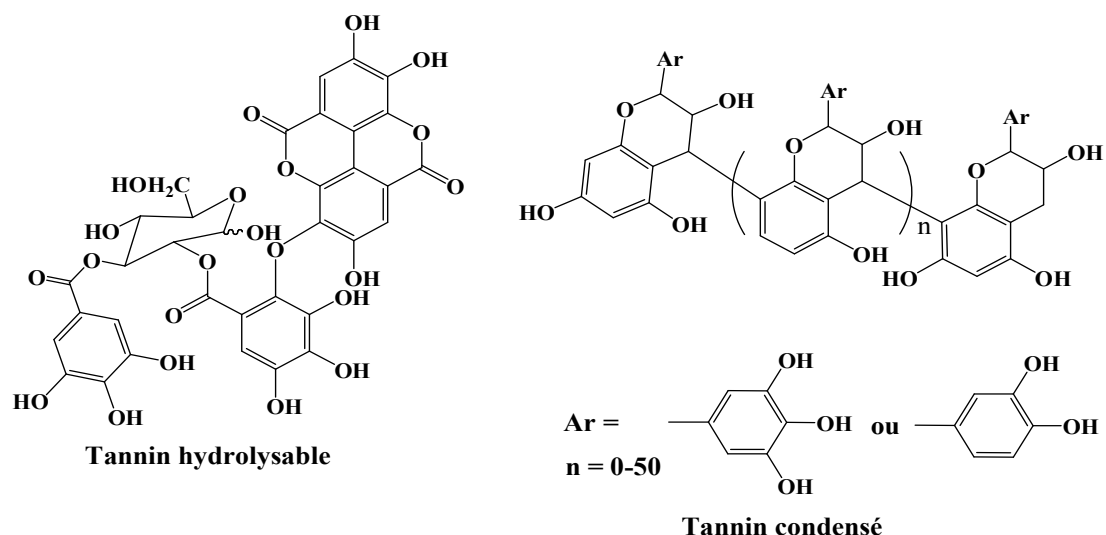


Figure 20 : structures chimiques de tannin hydrolysable et de tannin condensé.

4.3.4. Chalcones :

Les chalcones, considérés comme précurseurs de flavonoïdes, sont abondants dans les plantes⁸¹. Elles possèdent des activités pharmacologiques variées, telles que l'activité anti-inflammatoire^{82,83}, anticancéreuse⁸⁴ et l'anti-HIV⁸⁵. Trois composés naturels, isolés de la partie aérienne d'*Angelica keiskei*, le xanthoangelol, le isobavachalcone et le 4-hydroxyderricine ont montré une activité inhibitrice puissante sur la maltase de levure avec des valeurs d'IC₅₀ 20,32 ; 11,54 et 33,76 μM respectivement (figure 21)⁸⁶.

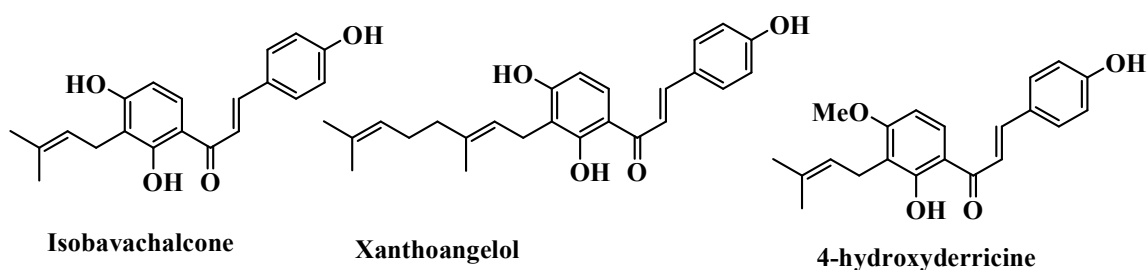


Figure 21 : Quelques chalcones naturels ayant un pouvoir inhibiteur sur les glycosidases

Récemment Woo Duck Seo *et al*⁸⁷ ont constaté que les chalcones aminées ont le potentiel d'agir comme une nouvelle classe d'inhibiteurs hautement spécifiques des α-glucosidases. Ils ont synthétisé une série de chalcones 1-20, et étudié leur activité inhibitrice vis-à-vis des glycosidases. Les composés non aminochalcones 1-12 (figure 22) n'avaient aucune activité inhibitrice, cependant, les aminochalcones 13-20 (figure 22) avaient d'excellentes activités inhibitrices contre les glucosidases (α-glucosidase, α-amylase et β-amylase). En

particulier, les sulfamides chalcones 17-20 avaient des activités inhibitrices de l' α -glucosidase plus puissantes que celles des chalcones aminées 13-16.

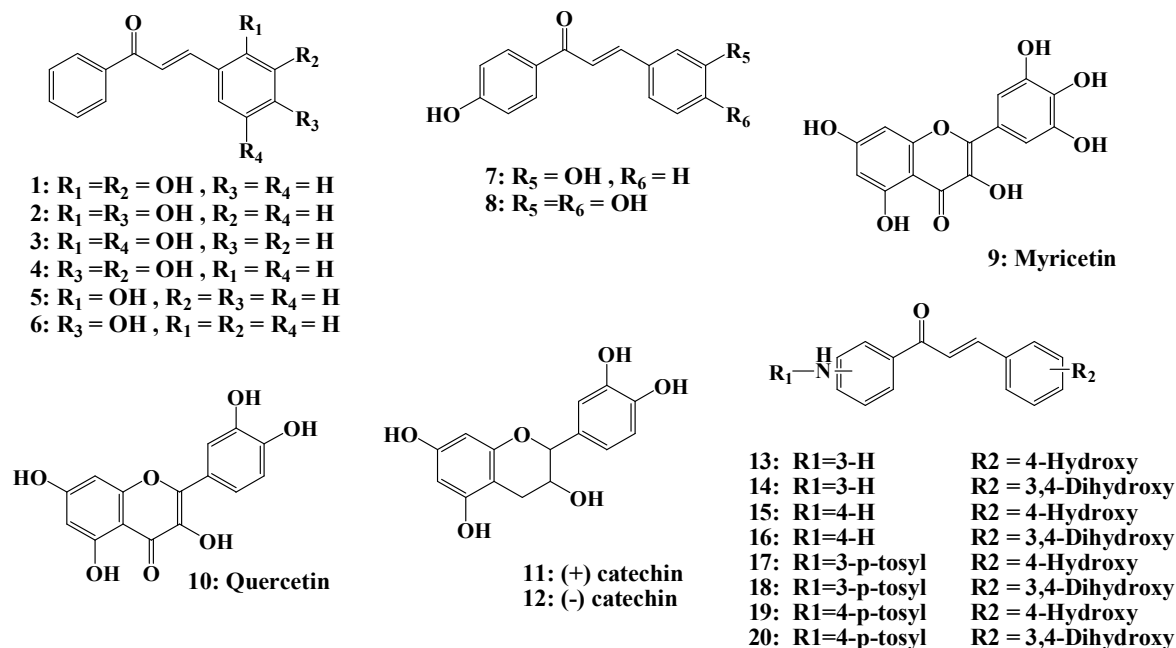


Figure 22 : Structures chimiques des inhibiteurs de l' α -glucosidase dérivés des chalcones

4.4.5. Salacinol et katalanol :

Récemment, une nouvelle classe d'inhibiteurs de glycosidases a été identifiée. Le salacinol et le kotalanol (Figure 23), ont été isolés des racines et des tiges d'une plante médicinale⁸⁸ : *Salaciareticulata*, ou "Kotalahimbutu" en Inde et au Sri Lanka et à partir d'autres variétés de *Salacia*: *Salaciaoblunga* en Inde⁸⁹ et *Salaciachinensis* en Thaïlande. Ces composés qui possèdent une structure zwitterionique comprenant un cation sulfonium stabilisé par un anion sulfate, sont considérés comme de bons inhibiteurs d' α -glucosidases ($K_i = 0,93 \mu M, K_i = 0,54 \mu M$; maltase d'intestin grêle de rat).

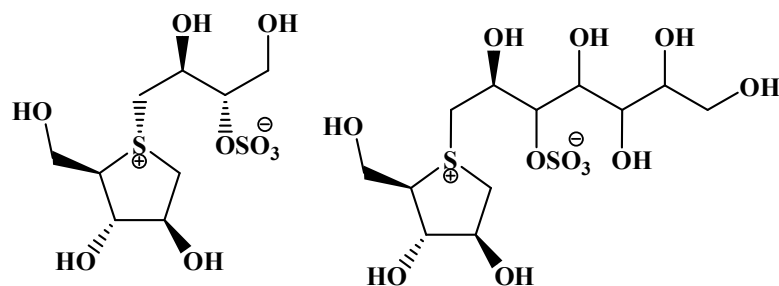


Figure 23 : Structure de salacinol et de kotalanol

5. Références :

1. L. EL BLIDI, "Synthèse chimio-enzymatique et évaluation d'inhibiteurs potentiels de glycosidases, analogues de la valioline" Thèse de doctorat, Université BLAISE PASCAL, France, 2006.
2. C. OUAIRY, « Synthèse de nouveaux glycoimidazoles, inhibiteurs potentiels de glycosidases préparation et utilisation en synthèse de glutaconaldéhydes oxydés en position 2 et de *n*-acylaminopentadiénals » Thèse de doctorat, Université Paris-sud, France, 2011.
3. DE. Koshland, *Biol.Rev.* () 90, 1953, 1171-1202.
4. E. MONTEL, "Synthèse chimio-enzymatique de dérivés de laminari-oligosaccharides, et leur utilisation biochimique" Thèse de doctorat, Université Joseph FOURIER - Grenoble 1, France, 2006.
5. J. McCarter, S. Withers, *CurrOpin.Struct.Biol.* 4, 1994, 885-892.
6. G. Davies, B. Henrissat, *Structure.* 3, 1995, 853 - 859.
7. F. CAMPS BRES, « Synthèse d'aminocyclitols, inhibiteurs potentiels de glycosidases lysosomales, via des aldolases » Thèse de Doctorat, Université BLAISE PASCAL, France, 2010.
8. A. Vasella, GJ. Davies, M. Bohm, *Curr.Opin. Chem. Bio.* 6, 2002, 619-629.
9. J. Davies, B. Evans, M. Jenney, J. Gregory, *J. Clin. Endocrinol.Metab.* 88, 2003, 6088-6097.
10. C. Rye, S. Withers, *Curr.Opin. Chem. Biol.* 4, 2000, 573-580.
11. B. Saha, *J. Ind. Microbiol. Biotechnol.* 30, 2003, 279-291.
12. V. Yip, A. Varrot, G. Davies, S. Rajan, X. Yang, J. Thompson, W. Anderson, S. Withers, *J. Am. Chem. Soc.* 126, 2004, 8354- 8355.
13. G. Sulzenbacher, H. Driguez, B. Henrissat, M. Schulein, G. Davies, *Biochemistry.* 35, 1996, 15280-15287.
14. A. Varrot, J. Macdonald, R. Stick , G. Pell, H. Gilbert, G. Davies, *Chem. Commun.* 47, 2003, 946-947.
15. E. Sabini, G. Sulzenbacher, M. Dauter, Z. Dauter., P. Jorgensen, M. Schulein, C. Dupont, G. Davies, K. Wilson, *Nat. Struct. Biol.* 3, 1996, 638-648.
16. V. M. A. Ducros, D. L. Zechel, G. N. Murshudou, H. J. Gilbert, L. Szabo, D. Stoll, S. G. Withers, G. J. Davies, *Angew. Chem. Int. Ed. Engl.* 41, 2002, 2824-2827.

17. D. M. Guerin, M. B. Lascombe, M. Costabel, H. Souchon, V. Lamzin, P. Beguin, B. G. Guimaraes, H. Souchon, B. L. Lytle, W. J. David, P. M. Alzari, *J.Mol.Biol.* 316, 2002, 1061-1069.
18. M. R. Stam, E. Blanc, P.M. Coutinho, B. Henrissat, *Carbohydr. Res.* 340, 2005, 2728-2734.
19. D. L. Zechel, S. G. Withers, *Acc. ChemRes.* 33, 2000, 11-8.
20. G. Placier, «Caractérisation d'une α -D-galactosidase acide d'Achatine strictement spécifique du seul résidu β -D-galactosyle. Applications potentielles». *Thèse doctorat. Université de Nantes, France, 1999.*
21. M. R. Stam, « Evolution et prédiction des activités des glycosides hydrolases » *Thèse doctorat de l'Université Aix-Marseille1, France, 2006.*
22. B. Henrissat, M. Claeysens, P. Tomme, L. Lemesle, J. P. Mornon, *Gene* 81, 1989, 83-95.
23. B. Henrissat, *Biochem. J.* 280, 1991, 309-316.
24. C. Chothia, A. M. Lesk, *EMBO J.* 5, 1986, 823-826.
25. B. Henrissat, A. Bairoch, *Biochem J.* 293, 1993, 781-788.
26. B. Henrissat, S. E. Heffron, M. D. Yoder, S. E. Lietzke, F. Journak, *Plant Physiol.* 107, 1995, 963-976.
27. D. Kim, B. H. Park, B. W. Jung, M. k. Kim, S. I. Hong, D. S. Lee, *Int. J. Mol. Sci.* 7, 2006, 571-589.
28. M. R. Stam, E. G. J. Danchin, C. Rancurel, P. M. Coutinho, B. Henrissat, (2006). *Prot. Engineering Design Selection.* 19, 2006, 555-562.
29. E. GALLIENNE, « Synthèse et évaluation d'inhibiteurs potentiels de glycosidases, analogues du salacinol » Thèse de doctorat, UNIVERSITE BLAISE PASCAL, France, 2005.
30. L. Michaelis, M. L. Menten, *Biochem.* 49, 1993, 333-369.
31. « Introduction à la cinétique enzymatique » cours de Biochimie licence science de la vie, (http://mon.ftp.a.moi.chez-alice.fr/Ecole/DEUG_SV2/SV2.htm), consulté le 20 mai 2014.
32. R. GARRETT, C. M. GRISHAM, « Biochimie » De Boeck, 2ème Edition américaine, Paris, 2000.
33. A. CORNISH-BOWDEN, M. JAMIN, V. SAKS, « Cinétique enzymatique » Grenoble Science, Portland press, Ltd. London, 2004.

34. A. RAISONNIER, « Enzymologie élémentaire » PCEM2, Université Pierre et Marie Curie, 2002.
35. G. Bjelaković, I. Stojanović, G. B. Bjelaković, D. Pavlović, G. Kocić, A. Daković, *Med. Bio.* 9(3), 2002, 201 – 206.
36. Cours des inhibitions des réactions enzymatiques (<http://biochimej.univ-angers.fr/Page2/COURS/4EnzymologieLicence/3CoursINHIBITEURS/1CoursInhibition.htm>), consulté le 20 mai 2014.
37. Cours de cinétique enzymatique, (<http://coursplbichat-lariboisiere.weebly.com>), consulté le 20 mai 2014.
38. M. DIXON, *Biochem. J.* 55(1), 1953, 170-171.
39. B. TOUSSAINT, « Les enzymes » cours de biochimie, Université Joseph Fourier Grenoble, 2010.
40. C. BLONSKI, « Enzymologie : Inhibition Enzymatique » notes de cours, université Paris-Sud11, Orsay, France.
41. M. B. Čolović, D. Z. Krstić, T. D. Lazarević-Pašti, A. M. Bondžić, V. M. Vasić, *Curr. Neuropharm.* 11(3), 2013, 315–335.
42. U.F. Wehmeier, *Biocatal. Biotransformation.* 21, 2003, 279–284.
43. L. Sim, “Structural and inhibition studies of human intestinal glucosidases” Thèse de doctorat, Université de Toronto, USA, 2010.
44. J.A. Balfour, D. McTavish, *Drugs.* 46(6), 1993, 1025–1054.
45. X. Chen, Y. Fan, Y. Zheng, Y. Shen, *Chem. Rev.* 103, 2003, 1955-1977.
46. Y. Kameda, N. Asano, M. Yoshikawa, M. Takeuchi, T. Yamaguchi, K. Matsui, S. Horii, H. J. Fukase, *Antibiot.* 37, 1984, 1301-1307.
47. T. D. Heightman, A. T. Vasella, *Angew. Chem. Int. Ed.* 38, 1999, 750-770.
48. S. Inouye, T. Tsuruoka, T. Niida, *J. Antibiot.* 19, 1966, 288-292.
49. S. Inouye, T. Tsuruoka, T. Ito, T. Niida, *Tetrahedron.* 24, 1968, 2125-2144.
50. H. Paulsen, I. Sangster, K. Heyns, *Chem. Ber.* 100, 1967, 802-811.
51. N. Katsilambros, P. Philippides, T. Toskas, J. Protopapas, D. Frangaki, M. Marangos, P. Siskoudis, K. Anastasopoulou, H. Xefteri, I. Hillebrand, *Arzneimittelforschung.* 36, 1986, 1136-1138.
52. F. M. Platt, G. R. Neises, R. A. Dwek, T. D. Butters, *J. Biol. Chem.* 269, 1994, 8362-8365.
53. S. M. Colegate, P. R. Dorling, C. R. Huxtable, *Aust. J. Chem.* 32, 1979, 2257-2264.
54. G. Fleet, M. J. Gough, P. W. Smith, *Tetrahedron Lett.* 25, 1984, 1853-1856.

55. L. D. Hohenschutz, E. A. Bell, P. J. Jewess, D. P. Leworthy, R. J. Pryce, E. Arnold, J. Clardy, *Phytochemistry*. 20, 1981, 811-814.
56. J. L. Reymond, A. A. Pinkerton, P. Vogel, *J. Org. Chem.* 56, 1991, 2128-2135.
57. K. Whitby, T. C. Pierson, B. Geiss, K. Lane, E. Engle, Y. Zhou, R. W. Doms, M. S. Diamond, *J. Virol.* 79, 2005, 8698-8706.
58. M. G. Stover, R. R. Watson, *Polyph.. Hum. Health. Dis.* 1, 2014, 3-7.
59. K.V.P. Chandra Mohan, P. Gunasekaran, E. Varalakshmi, Y. Hara, S. Nagini, *Cell. Biology. International.* 31(6), 2007, 599-608.
60. S. Vetal, S. L. Bodhankar, V. Mohan, P. A. Thakurdesai, *Food. Sci. HumanWell.*, 2(2), 2013, 59-67.
61. M. Quiñones, M. Miguel, A. Aleixandre, *Pharm. Res.* 68(1), 2013, 125-131.
62. A. Tresserra-Rimbau, E.B. Rimm, A. Medina-Remón, M.A. Martínez-González, R. de la Torre, D. Corella, J. Salas-Salvadó, E. Gómez-Gracia, J. Lapetra, F. Arós, M. Fiol, E. Ros, L. Serra-Majem, X. Pintó, G.T. Saez, J. Basora, J.V. Sorlí, J.A. Martínez, E. Vinyoles, V. Ruiz-Gutiérrez, R. Estruch, RM. Lamuela-Raventos, *Nutr. Metab. Cardiovasc. Dis.* 24(6), 2014, 639-647.
63. F. F. Anhê, Y. Desjardins, G. Pilon, S. Dudonné, M. InésGenovese, F. M. Lajolo, A. Marette, *PharmaNutrition.* 1(4), 2013, 105-114.
64. H. Boo, S. Hwang, C. SikBae, Su-H. Park, B. GuHeo, S. Gorinstein, *Indust. Crop. Prod.* 40, 2012, 129-135.
65. J. B. Xiao, G. Y. Kai, *Crit. Rev. Food Sci. Nutr.* 52, 2012, 85-101.
66. C. Blade, L.Arola, M. J. Salvado, *Mol. Nutr. Food Res.* 54, 2010, 37-59.
67. G. Morabito, D. Trombetta, K. S. Brajendra, K. P. Ashok, S. P. Virinder, C. Naccari, F. Mancari, A. Saija, M. Cristani, O. Firuzi, , L. Saso, *Biochimie.* 92, 2010, 1101-1107.
68. F. Natella, B. Lorrain, A. K. Prasad, V. S. Parmar, L. Saso, C. Scaccini, *Biochimie.* 92, 2010, 1147-1152.
69. M. H. Pan, , C. S. Lai, C. T. andHo, *Food. Funct.* 1, 2010, 15-31.
70. J. B. Xiao, H. Cao, Y. F. Wang, K. Yamamoto, X. L. Wei, *Mol. Nutr. Food Res.* 54, 2010, S253-S260.
71. J. B. Xiao, T. T. Chen, H. Cao, L. Chen, F. Yang, *Mol. Nutr. Food Res.* 55, 2011, 310-317.
72. W. Brand, B. Padilla, P. J. van Bladeren, G. Williamson, I. M. C. M. Rietjens, *Mol. Nutr. Food Res.* 54, 2010, 851-860.
73. S. I. Rizvi, M. Abu Zaid, *Clinic. Chim. Acta,* 354(1-2), 2005, 59-67.

74. Y. Yao, X. Z. Cheng, L. X. Wang, S. H. Wang, G. X. Ren, *Inter. J. Mol. Sci.* 12, 2011, 6445–6451.
75. S. Sengupta, A. Mukherjee, R. Goswami, S. Basu, *J. Enz. Inhib. Med. Chem.* 24, 2009, 684–690.
76. H. J. Yuk, Lee, M. J. Curtis-Long, J. W. Lee, Y. S. Kim, , H. W. Ryu, *Food Chem.* 126, 2011, 1057–1063.
77. P. Aparna, A. K. Tiwari, P. V. Srinivas, A. Z. Ali, V. Anuradha, , J. M. Rao, *Phytother. Res.* 23, 2009, 591–596.
78. R. Omar, L. Y. Li, T. Yuan, N. P. Seeram, *J. Nat. Prod.* 75, 2012, 1505–1509.
79. S. Ochir, M. Nishizawa, B. J. Park, K. Ishii, T. Kanazawa, M. Funaki, *J. Nat. Med.* 64, 2010, 275-280.
80. Y. A. Lee, E. J. Cho, T. Tanaka, T. Yokozawa, *J. Nut. Sci. Vit.* 53, 2007, 287–292.
81. J. F. Stevens, A. W. Taylor, G. B. Nickerson, Ivancic, J. Henning, A. Haunold, M. L. Deinzer, *Phytochemistry.* 53(7), 2000, 759-775.
82. K. V. Sashidhara, M. Kumar, R. K. Modukuri, R. Sonkar, G. Bhatia, A.K. Khanna, S. Rai, R. Shukla, *Bioorg. Med. Chem. Lett.* 21(15), 2011, 4480-4484.
83. S. H. Lee, G. S. Seo, J. Y. Kim, X. Y. Jin, H-D. Kim, D. H. Sohn, *Eur. J. Pharmaco.* 532(1-2), 2006, 178-186.
84. I. S. Aljančić, I. Vučković, M. Jadranin, M. Pešić, I. Đorđević, A. Podolski-Renić, S. Stojković, N. Menković, V. E. Vajs, S. M. Milosavljević, *Phytochemistry*, 98, 2014, 90-196.
85. M. Abdel-Aziz, S. Park, G. El-Din, A.A. Abuo-Rahma, M. A. Sayed, Y. Kwon, *Euro. J. Med. Chem.* 69, 2013, 427-438.
86. L. Luo, R. Wang, X. Wang, Z. Ma, N. Li, (2012). *Food Chem.* 131, 2012, 992–998.
87. W. D. Seo, J. H. Kim, J. E. Kang, H. W. Ryu, M. J. Curtis-Long, H. S. Lee, M. S. Yang, K. H. Park, *Bioorg. Med. Chem. Lett.* 15(24), 2005, 5514-5516.
88. M. Yoshikawa, T. Morakami, H. Shimada, H. Matsuda, J. Yamahara, G. Tanabe, O. Muraoka, *Tetrahedron Lett.* 38, 1997, 8367-7939.
89. H. Matsuda, T. Murakami, K. Yashiro, J. Yamahara, M. Yoshikawa, *Chem. Pharm. Bull.* 47, 1999, 1725-1729.

Chapitre 2

QSAR : principe et méthode

1. Introduction :

Les premiers essais de modélisation des relations structure-activité ont commencé à la fin du 19^{ème} siècle, lorsque *Crum-Brown* et *Frazer* ont postulé que l'activité biologique d'une molécule est une fonction de sa constitution chimique¹⁻³. Mais ce n'est qu'au début des années 60 que les travaux de *Corwin Hanch* ont proposé un modèle mathématique reliant l'activité biologique à la structure chimique⁴.

Aujourd'hui, l'utilisation de QSAR n'a cessé de progresser^{5,6}. Elle est devenue indispensable en chimie pharmaceutique et pour la conception de médicaments, notamment dans le cas où la disponibilité des échantillons est limitée ou les mesures expérimentales sont dangereuses, longues et chères^{7,8}.

Sans l'utilisation de grands instruments analytiques, les résultats des études QSAR peuvent fournir des informations utiles pour obtenir une meilleure connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent être alors utilisées dans la prédiction des activités biologiques de nouveaux composés ainsi que dans la conception de nouvelles structures⁹.

2. Principe de QSAR :

Le principe des méthodes QSAR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative la structure moléculaire, codée par des propriétés moléculaires appelées descripteurs, avec une activité en utilisant des méthodes d'analyse de données¹⁰. Par ces relations nous pouvons développer des modèles prédictifs de la forme générale suivante :

$$\text{Activité} = f(\text{descripteurs moléculaires})$$

L'objectif de ces méthodes est alors d'analyser les données structurales afin de déterminer les facteurs influençant l'activité mesurée¹¹.

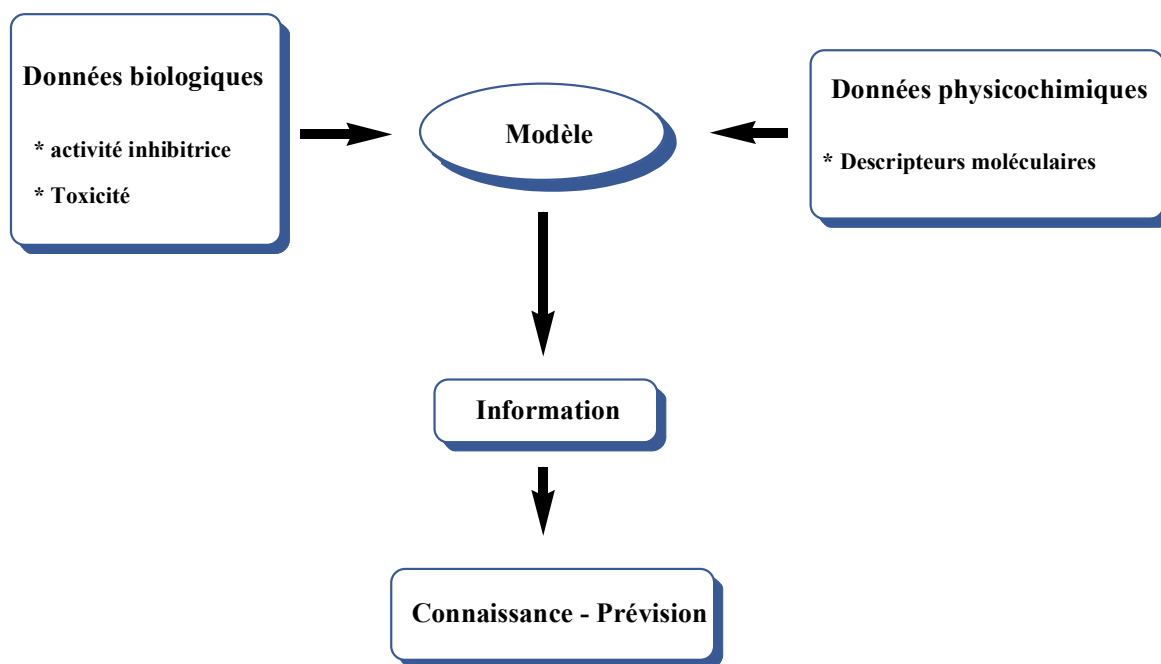


Figure 1: présentation de la méthodologie de QSAR.

En pratique, le développement d'un modèle commence par la collecte de données expérimentales fiables et en nombre important, ensuite il est nécessaire de générer un nombre de descripteurs, caractérisant les structures moléculaires, utilisés dans le développement du modèle QSAR¹². Une fois le modèle construit, il doit être évalué par des méthodes de validation (validations interne et externe) afin d'estimer sa robustesse et son pouvoir prédictif¹³⁻¹⁵. En fin, pour tout modèle, il est important de connaître quel type de molécules ce modèle est utilisable (connaître le domaine d'applicabilité)¹⁶.

Il y a plusieurs raisons pratiques qui justifient l'utilisation des études QSAR⁷.

- Pour prédire les propriétés de l'activité biologique et physico-chimiques par des moyens rationnels.
- Economie sur le coût de développement de produits (exemple de l'industrie pharmaceutique, des pesticides, des détergents,...etc.).
- Les prévisions pourraient réduire l'exigence de tests longs et coûteux pratiqués sur les animaux.

3. 3D QSAR :

Les études structure-activité en trois dimensions nous permettent d'obtenir des modèles qui établissent une relation quantitative entre une activité biologique et les paramètres structuraux calculés dans l'espace tridimensionnel pour un ensemble de molécules¹⁷⁻¹⁹.

Parmi les propriétés moléculaires utilisées dans ce genre d'études nous citerons :

- La surface moléculaire, accessible au solvant, ou surface de contact.
- Le potentiel électrostatique (position des groupements chargés).
- Les propriétés donneurs ou accepteurs de liaisons hydrogènes.
- Le potentiel de lipophilie moléculaire.

Le principe de cette méthode repose sur l'insertion d'une molécule dans un système de référence tridimensionnelle (réseau maillé ou Grid Box) comprenant un espace suffisamment large autour de la molécule, en calculant les champs de propriétés pour chaque molécule en chaque point de la grille²⁰ (figure.2).

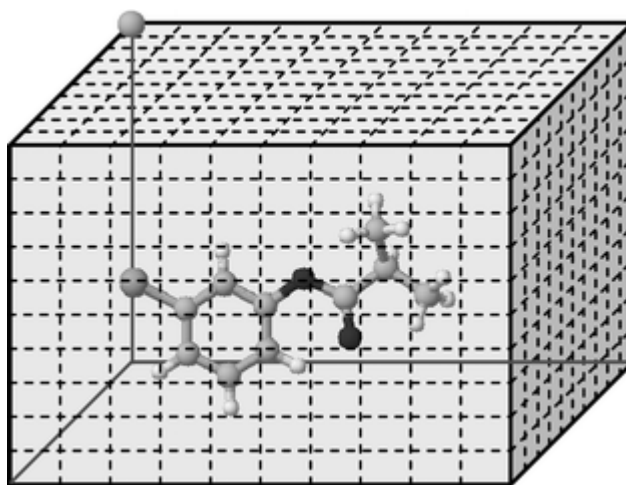


Figure 2 : Réseau maillé.

4. Application de QSAR :

Il existe un grand nombre d'applications des modèles QSAR en milieu industriel, dans le domaine de la recherche universitaire, en économie, en prévision météorologique... etc²¹. Nous rapportons ci-dessous quelques applications possibles des modèles QSAR²²⁻²⁴ :

- L'identification rationnelle de nouvelles pistes avec :
- l'optimisation de l'activité pharmacologique, biocide ou pesticide.

- la conception rationnelle de nombreux produits tels que des agents tensioactifs, des parfums, des colorants et des produits chimiques fins.
- l'identification des composés dangereux aux premiers stades de développement.
- la prédiction de la toxicité et les effets secondaires de nouveaux composés.
- la sélection des composés ayant des propriétés pharmacocinétiques optimales, que se soit la stabilité ou la disponibilité dans les systèmes biologiques.

5. Outils et Méthodologie de QSAR :

5.1. Paramètres biologiques :

Les modèles QSAR sont dépendants des données expérimentales utilisées pour leur construction. Le modélisateur doit tenir compte des données à modéliser. Le choix de la base de données est donc une étape très importante dans le développement des modèles QSAR.

Ces données devraient, idéalement, être de grande qualité, ce qui signifie qu'elles devraient être fiables et cohérentes. Il est donc important de les choisir parmi celles présentant des incertitudes faibles afin de limiter les barres d'erreurs expérimentales. De plus, le modélisateur doit s'assurer que les données expérimentales utilisées ont été obtenues selon le même protocole. En effet les conditions expérimentales ont, généralement, une forte influence sur les valeurs obtenues²⁵.

Il faut également que la distribution des données soit la plus homogène et normale que possible, car la plupart des méthodes statistiques sont basées sur ce type de distribution²⁶.

L'efficacité d'un modèle QSAR dépend également du type de molécules qui y sont incluses, plus le modèle présentera des composés de structures proches et similaires, plus il aura de chance d'être performant.

Les données biologiques sont généralement exprimées en logarithmes inverses ($\log 1/C$) afin d'obtenir des valeurs mathématiques plus élevées lorsque les structures sont biologiquement très efficaces^{27,28}. Quelques exemples de données biologiques, utilisées dans l'analyse QSAR, sont décrits dans le tableau.1.

Tableau 1 : quelques paramètres utilisés dans l'analyse QSAR²⁸.

Source d'activité	Paramètres biologiques
Constante de Michaelis-Menten	Log1/Km
Constante d'inhibition (récepteur isolé)	Log1/Ki
Constante d'inhibition (systèmes cellulaires)	Log1/IC50
Facteur de bioconcentration	logBCF
Vitesse de la réaction in vivo	logI (induction)
Vitesse pharmacodynamique	logT (clairance totale)
Mutation de gène	log TA98
Résistance croisée	logCR
Constante de vitesse	logK

5.2. Utilisation des descripteurs moléculaires :

L'activité biologique ne peut pas être reliée directement à la structure moléculaire ; cette dernière est codée par des grandeurs qui représentent d'une manière quantitative les informations contenues dans la structure moléculaire telles que les caractéristiques physico-chimiques et structurales. Ces grandeurs sont appelées descripteurs²⁹. Une fois ces derniers disponibles, il est possible d'établir des relations entre ceux-ci et une activité (ou propriété) à l'aide des outils de la modélisation³⁰. On en dénombre aujourd'hui des milliers, qui peuvent être calculés ou obtenus de manière empirique. De nombreux logiciels ont été développés pour calculer les différents descripteurs moléculaires tels que : OASIS^{31,32}, ADAPT^{33,34}, CODESSA³⁵⁻³⁷, DRAGON³⁸, Mol ConnZ³⁹, MOPAC⁴⁰, etc.

Dans ce travail de thèse nous avons utilisé le logiciel DRAGON.6.0 qui est capable de générer plus de 4800 descripteurs moléculaires.

5.2.1. Descripteurs générés par le logiciel DRAGON :

Dragon est une application utilisée pour le calcul de descripteurs moléculaires. Les descripteurs de Dragon peuvent être utilisés pour évaluer les relations structure-activité (ou structure-propriété moléculaire), ainsi que pour le criblage de bases de données.

Dragon fournit près de 5000 descripteurs moléculaires qui sont répartis en 29 blocs logiques (tableau.2). On peut calculer, avec Dragon, non seulement les types les plus

simples de descripteurs (constitutionnels), mais également plusieurs descripteurs topologiques et géométriques tels que : le volume et la surface moléculaires et d'autres propriétés comme l'alerte de *LIPINSKI*^{41,42} (appelée aussi : la règle de cinq) et certains indices utilisés dans la sélection de composés pour le criblage et /ou dans la conception de médicaments.

Nous décrivons dans ce chapitre ces différents blocs de descripteurs et nous nous limiterons surtout aux descripteurs les plus utilisés.

Tableau 2 : liste des blocs de descripteurs fournis par le logiciel Dragon.

N°	Nom du bloc	N°	Nom du bloc
1	Constitutional descriptors	16	RDF descriptors
2	Ring descriptors	17	3D-MoRSE descriptors
3	Topological indices	18	WHIM descriptors
4	Walk and path counts	19	GETAWAY descriptors
5	Connectivity indices	20	Randicmolecular profiles
6	Information indices	21	Functional group counts
7	2D matrix-based descriptors	22	Atom-centred fragments
8	2D autocorrelations	23	Atom-type E-state indices
9	Burden eigenvalues	24	CATS 2D
10	P_VSA-like descriptors	25	2D Atom Pairs
11	ETA indices	26	3D Atom Pairs
12	Edge adjacency indices	27	Charge descriptors
13	Geometrical descriptors	28	Molecular properties
14	3D matrix-based descriptors	29	Drug-like indices
15	3D autocorrelations		

5.2.1.1. Les descripteurs constitutionnels :

Les descripteurs constitutionnels sont les plus simples, d'un point de vue conceptuel et du coût calculatoire, et ils sont très couramment utilisés. Ils reflètent la composition chimique d'un composé sans aucune information sur sa géométrie moléculaire ni sur sa structure électronique.

Dragon calcule 43 descripteurs constitutionnels (**Tableau.3**), beaucoup d'entre eux sont bien expliqués par leur définition tels que : le poids moléculaire (MW), le nombre d'atomes (nAT), le nombre de liaison (nBT), etc.

Tableau 3 : liste des descripteurs constitutionnels générés par le logiciel Dragon.

N°	nom	Description
1	MW	molecular weight
2	AMW	average molecular weight
3	Sv	sum of atomic van der Waals volumes (scaled on Carbon atom)
4	Se	sum of atomic Sanderson electronegativities (scaled on Carbon atom)
5	Sp	sum of atomic polarizabilities (scaled on Carbon atom)
6	Si	sum of first ionization potentials (scaled on Carbon atom)
7	Mv	mean atomic van der Waals volume (scaled on Carbon atom)
8	Me	mean atomic Sanderson electronegativity (scaled on Carbon atom)
9	Mp	mean atomic polarizability (scaled on Carbon atom)
10	Mi	mean first ionization potential (scaled on Carbon atom)
11	nAT	number of atoms
12	nSK	number of non-H atoms
13	nBT	number of bonds
14	nBO	number of non-H bonds
15	nBM	number of multiple bonds
16	SCBO	sum of conventional bond orders (H-depleted)
17	RBN	number of rotatable bonds
18	RBF	rotatable bond fraction
19	nDB	number of double bonds
20	nTB	number of triple bonds
21	nAB	number of aromatic bonds
22	nH	number of Hydrogen atoms
23	nC	number of Carbon atoms
24	nN	number of Nitrogen atoms
25	nO	number of Oxygen atoms
26	nP	number of Phosphorous atoms
27	nS	number of Sulfur atoms
28	nF	number of Fluorine atoms
29	nCL	number of Chlorine atoms
30	nBR	number of Bromine atoms

31	nI	number of Iodine atoms
32	nB	number of Boron atoms
33	nHM	number of heavy atoms
34	nHet	number of heteroatoms
35	nX	number of halogen atoms
36	H%	percentage of H atoms
37	C%	percentage of C atoms
38	N%	percentage of N atoms
39	O%	percentage of O atoms
40	X%	percentage of halogen atoms
41	nCsp3	number of sp ³ hybridized Carbon atoms
42	nCsp2	number of sp ² hybridized Carbon atoms
43	nCsp	number of sp hybridized Carbon atoms

Ce type de descripteurs est employé pour obtenir les modèles QSAR les plus simples, mais dans certains cas ils sont défaillants car ils ne permettent pas la proposition de mécanismes d'interactions mises en jeu pour l'activité étudiée. Par exemple, ils ne permettent pas de distinguer les isomères. Or, pour la majorité des cas, la position d'un substituant modifie la valeur expérimentale de l'activité.

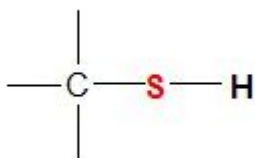
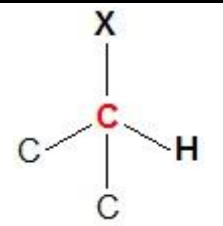
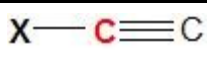
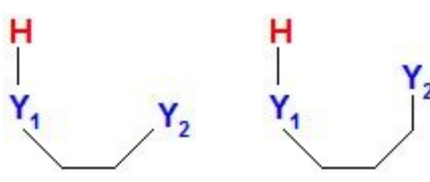
5.2.1.2. Descripteurs de groupements fonctionnels :

Ce sont des descripteurs moléculaires simples, définis comme étant le nombre de groupes fonctionnels spécifiques dans une molécule. Ils sont calculés sur la base de la composition moléculaire et des connectivités atomiques.

Dragon calcule 154 groupes fonctionnels. Le tableau.4 rassemble quelques descripteurs impliqués dans le bloc des groupements fonctionnels.

Tableau 4 : présentation de quelques descripteurs moléculaires du bloc de groupements fonctionnels.

N°	Symbole	Structure	Définition
01	nArOCN	$\text{Ar}-\text{O}-\text{C}\equiv\text{N}$	Nombre de cyanates (aromatique)
02	nArNCO	$\text{Ar}-\text{N}=\text{C}=\text{O}$	Nombre d'isocyanates (aromatique)
03	nArCONH ₂	$\begin{array}{c} \text{NH}_2 \\ \diagdown \\ \text{C}=\text{O} \\ \diagup \\ \text{Ar} \end{array}$	Nombre d'amides primaires (aromatique)
04	nRCSOH	$\begin{array}{cc} \text{OH} & \text{SH} \\ \diagdown & \diagdown \\ \text{C}=\text{S} & \text{C}=\text{O} \\ \diagup & \diagup \\ \text{Al} & \text{Al} \end{array}$	Nombre de thioacides (aliphatique)
05	nRCO	$\begin{array}{c} \text{Al} \\ \diagdown \\ \text{C}=\text{O} \\ \diagup \\ \text{Al} \end{array}$	Nombre de cétones (aliphatique)
06	nRNHO	$\begin{array}{c} \text{Al} \\ \diagdown \\ \text{N}-\text{O}-\text{Y} \\ \diagup \\ \text{Al} \end{array}$	Nombre d'hydroxylamines (aliphatique) Y = H, Al ou Ar
07	nRNO ₂	$\begin{array}{c} \text{O} \\ \diagdown \\ \text{Al}-\text{N} \\ \diagup \\ \text{O} \end{array}$	Nombre de groupes nitro (aliphatique) Al = H ou groupe aliphatique lié avec un carbone
08	nO(C=O) ₂	$\begin{array}{ccc} \text{Y} & & \text{Y} \\ \parallel & & \parallel \\ -\text{C}- & \text{O} & -\text{C}- \\ & & \end{array}$	Nombre d'anhydrides (thio-) Y = O ou S
09	nH ₂ O	$\text{H}-\text{O}-\text{H}$	Nombre de molécules d'eau

10	nSH		Nombre de thiols
11	nCHR2X		Nombre de CHR2X
12	nR#CX		Nombre de R#CX
13	nHBonds		Nombre de liaisons H intramoléculaires $Y_1 = \text{B, N, O, Al, P, S}$ $Y_2 = \text{N, O, F}$

Les lettres X, b, Ar, r, = et # représentent respectivement, dans les symboles des descripteurs moléculaires, un halogène, un benzène, un groupe aromatique lié par un carbone du groupement, tout cycle non aromatique, des doubles et triples liaisons.

5.2.1.3. Les descripteurs topologiques :

Les descripteurs topologiques sont inspirés de la théorie des graphes, et sont des indices obtenus à partir d'une structure 2D de la molécule, ils donnent des informations sur la taille globale du système, sa forme globale et ses ramifications.

Les descripteurs topologiques sont basés sur une représentation graphique de la molécule dont le système moléculaire est considéré comme un graphe $G[V,R]$ dans lequel les atomes sont des sommets V et les liaisons chimiques des arêtes R . Ils peuvent être sensibles à une ou plusieurs caractéristiques structurales de la molécule comme la taille, la forme, la symétrie, la ramification et la cyclicité, et peuvent également coder l'information chimique concernant les types d'atomes et la multiplicité des liaisons.

La distance topologique d entre deux atomes est définie comme le nombre de liaisons minimum reliant ces atomes. L'indice de *Wiener*^{43,44}, par exemple, permet de caractériser

le volume moléculaire et le taux de ramification des molécules. Il est calculé comme la somme de toutes les distances topologiques entre atomes au sein de la molécule.

$$\sum_{i,j} d_{ij}$$

Si l'on s'intéresse à la molécule de toluène, la figure.3 représente la structure telle qu'utilisée pour calculer cet indice. Les seules informations nécessaires sont les atomes et leur connectivité d'un point de vue purement topologique. Les distances topologiques au sein de la molécule sont sommées pour donner l'indice de *Wiener*, qui est égal à 40 pour le toluène.

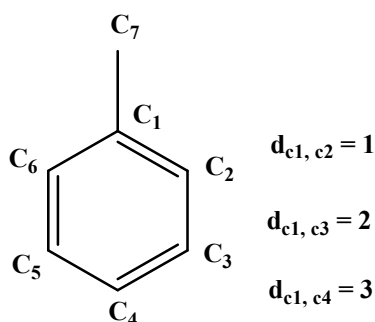


Figure 3 : structure du toluène.

Le tableau suivant rassemble quelques descripteurs topologiques issus du logiciel Dragon.

Tableau 5 : Quelques descripteurs topologiques fournis par logiciel Dragon.

N°	Nom	Description
1	Xt	total structure connectivity index
2	Ram	ramification index
3	Pol	polarity number
4	MSD	mean square distance index (Balaban)
5	MDDD	mean distance degree deviation
6	UNIP	unipolarity
7	SMTIV	Schultz Molecular
8	GMTI	Gutman Molecular
9	MAXDN	maximal electrotopological negative variation
10	MAXDP	maximal electrotopological positive variation
11	DELS	molecular electrotopological variation
12	TIE	E-state topological parameter
13	Psi_e_A	electrotopological state pseudoconnectivity index - type S average
14	Psi_e_0	electrotopological state pseudoconnectivity index - type 0
15	Psi_e_1	electrotopological state pseudoconnectivity index - type 1
16	Psi_e_tel	electrotopological state pseudoconnectivity index - type T
17	Psi_e_0d	electrotopological state pseudoconnectivity index - type 0d
18	Psi_e_1d	electrotopological state pseudoconnectivity index - type 1d
19	Psi_e_1s	electrotopological state pseudoconnectivity index - type 1s
20	BAC	Balaban centric index
21	LOC	lopping centric index

5.2.1.4. Les descripteurs géométriques :

Les descripteurs géométriques sont issus de la structure tridimensionnelle de la molécule, et sont évalués à partir des positions relatives des atomes d'une molécule dans l'espace, ainsi que des rayons et des masses atomiques.

En général, les descripteurs géométriques sont calculés soit à partir de la géométrie moléculaire optimisée par modélisation moléculaire soit à partir des coordonnées cristallographiques. Par conséquent, ils fournissent plus d'informations sur la structure

moléculaire par rapport aux descripteurs topologiques. Cependant, ils sont plus coûteux en calcul^{45,46}.

Les descripteurs géométriques calculés par Dragon sont divisés en sous-blocs tels que les indices de taille et les indices de forme. Parmi ces descripteurs on trouve le volume et la surface moléculaires, le moment d'inertie ou encore les distances et les angles dièdres entre atomes dans la molécule.

5.3. Sélection des descripteurs :

Lorsqu'une grande quantité de descripteurs est introduite, certains d'entre eux peuvent contenir des informations redondantes, entraînant un problème de colinéarité. De plus, les descripteurs calculés n'ont pas nécessairement une influence sur l'activité à modéliser.

Les descripteurs employés doivent être, autant que possible, porteurs de sens et facilement interprétables d'un point de vue chimique, aussi les modèles QSAR devraient être simples, transparents et compréhensibles d'un point de vue phénoménologique. Il est nécessaire donc d'éliminer les descripteurs dont l'influence est inférieure à celle de l'erreur, et de sélectionner uniquement les plus pertinents d'entre eux. De manière générale, pour qu'un descripteur soit retenu, il faut que son retrait entraîne une décroissance significative de performance du modèle. Il faut donc être attentif à ne pas perdre de l'information essentielle.

Finalement, le sens chimique des descripteurs utilisés doit être pris en considération puisque, plus les descripteurs sont reliés chimiquement au phénomène, plus la probabilité de faire intervenir des descripteurs par hasard est réduite.

La procédure de sélection et de réduction des variables (descripteurs) peut être effectuée en deux étapes⁴⁷ :

- Sélection objective.
- Sélection subjective.

5.3.1. Sélection objective :

Elle consiste en la sélection des variables en réduisant le nombre de descripteurs sans faire participer la variable dépendante (l'activité biologique). La première étape de cette procédure consiste à exclure tous les descripteurs ayant un pourcentage élevé de valeurs

identiques pour l'ensemble des composés (variance non significative). Cela permet de s'assurer que de tels descripteurs ne sont pas inclus par chance dans le modèle final. De même, lorsque deux descripteurs sont fortement corrélés et leur combinaison possède un coefficient de détermination supérieur au seuil requis ($R^2 > 0,90$), seul celui présentant la plus grande variance est retenu. Non seulement ces procédures évitent l'introduction, dans le modèle, de descripteurs inappropriés mais elles rendent la suite de l'analyse moins coûteuse en terme de temps de calcul, puisqu'elles réduisent le nombre de descripteurs restant à traiter.

5.3.2. Sélection subjective :

5.3.2.1. Introduction progressive :

Cette méthode consiste à incorporer, une à une, les variables au modèle en sélectionnant à chaque étape la variable dont la corrélation partielle avec la grandeur modélisée est la plus élevée⁴⁸.

5.3.2.2. Elimination progressive :

Cette méthode consiste en l'établissement du modèle avec l'ensemble des descripteurs pour ensuite ne garder que ceux qui permettent l'obtention d'un modèle ayant une bonne corrélation⁴⁸.

5.3.2.3. Sélection pas à pas :

C'est la combinaison des deux méthodes citées précédemment. Les variables sont incorporées une à une dans le modèle par sélection progressive. Cependant, à chaque étape, on vérifie que les corrélations partielles des variables précédemment introduites sont encore significatives⁴⁹.

5.3.2.4. Sélection par les algorithmes génétiques :

Les algorithmes génétiques (AG) sont inspirés du modèle des lois de la nature de *Darwin*, modèle basé sur le suivi des espèces les plus fortes^{50,51}. Dans une population donnée, les plus forts individus auront plus de chances de se reproduire et ainsi de transmettre leurs gènes pour produire de meilleurs descendants, augmentant la qualité globale de la population⁵²⁻⁵⁶.

Le processus génétique consiste en la reproduction des chromosomes (modèles) permettant d'obtenir de nouveaux descendants plus performants que leurs parents. Chaque chromosome est représenté par une chaîne de gènes (descripteurs) de longueur finie. Le classement des chromosomes dans la population se fait suivant une fonction d'évaluation. Cette fonction est d'origine statistique tels que : le coefficient de détermination R^2 , l'écart type S et le coefficient de détermination issu de la validation croisée, cette fonction est utilisée par les Algorithmes Génétiques, pour sélectionner les candidats aptes à la reproduction.

Les différentes opérations génétiques appliquées dans la production des populations de qualité à partir d'une population initiale sont : la reproduction, la sélection, le croisement, la mutation et la réparation⁵⁷⁻⁶².

5.4. Méthodes de modélisation des données :

5.4.1. Réseaux de neurones :

Les réseaux de neurones formels initiés dès 1943 par *Mc Culloch* et *Pitts*, sont analogues aux systèmes nerveux biologiques⁶³.

Les neurones formels (Figure.4) réalisent une combinaison linéaire des entrées reçues (descripteurs), puis appliquent à cette valeur une fonction d'activation f , généralement non linéaire (sigmoïdale, tangente, hyperbolique). La valeur obtenue Y (activité biologique) est la sortie du neurone⁶⁴.

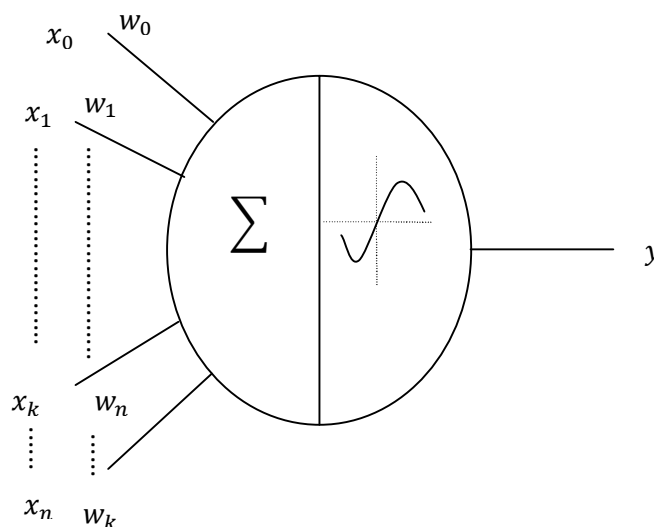


Figure 4 : représentation d'un neurone formel.

X_k (où $k = 1, 2, \dots, n$) sont les neurones de l'entrée (descripteurs) et W_k (où $k = 0, 1, 2, \dots, n$) sont les poids. W_0 est le poids associé à une entrée fixée à 1, appelée biais. L'équation du neurone est de la forme suivante :

$$y = f\left(W_0 + \sum_{k=1}^n W_k X_k\right) \quad eq.01$$

Les neurones seuls réalisent des fonctions assez simples, et ce sont leurs combinaisons, qu'on appelle réseaux de neurones, qui permettent de construire des fonctions particulièrement intéressantes (figure 5).

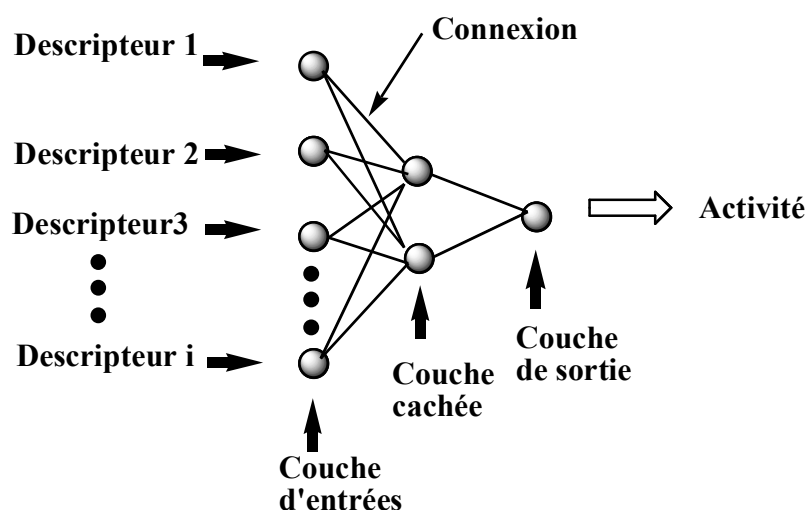


Figure 5 : architecture des réseaux de neurones.

L'architecture des réseaux de neurones est constituée de trois couches :

- Une couche d'entrée : constituée des neurones d'entrée, leur nombre est égal au nombre de variables d'entrée (descripteurs) plus un (biais). Chaque neurone est connecté aux neurones cachés.
- Une couche cachée : constituée d'un nombre variable de neurones. Pour chaque neurone caché, le réseau effectue une opération de somme pondérée avec les différents poids de chaque neurone d'entrée.
- Une couche de sortie : où le nombre de neurones de sortie est égal au nombre de propriétés (activités) à modéliser.

A chaque connexion entre les neurones des différentes couches on associe un poids. Le nombre de poids dépend alors du nombre de neurones de la couche cachée, il est donc possible de faire varier la complexité du réseau en augmentant ou en diminuant le nombre de neurones cachés. Pour ce faire, des cycles de présentation de données sont établis via un algorithme d'apprentissage. Celui utilisé dans notre démarche est un algorithme de rétro-propagation⁶⁵.

Le principe de l'algorithme de rétro-propagation consiste à minimiser la différence entre la sortie calculée (activité calculée) et les valeurs expérimentales de l'activité.

Défini en deux étapes, l'algorithme propage dans une première étape les entrées (descripteurs) vers l'avant jusqu'à obtenir une sortie (l'activité biologique) calculée par le réseau. Dans une seconde étape, la sortie calculée est comparée avec la valeur expérimentale. L'erreur obtenue par cette comparaison est ensuite rétro-propagée vers la couche d'entrée en modifiant les poids des neurones. Ce processus est réitéré jusqu'à l'obtention d'une erreur négligeable⁶⁷.

5.4.2. La régression linéaire multiple :

La régression linéaire multiple est la plus simple méthode statistique de modélisation et la plus appliquée dans les études de la relation structure-activité⁶⁸. La méthode a été popularisée par *Hansch* en reliant l'activité biologique aux propriétés expérimentales lipophiliques, électroniques et stériques pour des séries de composés⁶⁹.

La méthode RLM repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante Y (dans notre cas l'activité biologique) et une série de n variables indépendantes X_i (ici, les descripteurs moléculaires)⁷⁰. L'objectif est d'arriver à une équation mathématique de la forme :

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad eq. 02$$

Où a_i ($i = 0, 1, \dots, n$) sont les coefficients de la régression.

La détermination de l'équation 02 à partir d'un ensemble de données de p échantillons revient à résoudre un système de p équations.

$$y_1 = a_0 + a_1x_{1,1} + a_2x_{2,1} + \dots + a_nx_{n,1} + b_1 \quad eq. 03$$

$$y_2 = a_0 + a_1x_{1,2} + a_2x_{2,2} + \dots + a_nx_{n,2} + b_2 \quad eq. 04$$

⋮

$$y_p = a_0 + a_1x_{1,p} + a_2x_{2,p} + \dots + a_nx_{n,p} + b_p \quad eq. 05$$

Où :

Les résidus $b_i (i = 1, 2, \dots, p)$ représentent l'erreur du modèle,
 $y_i (i = 1, 2, \dots, p)$ représentent les variables dépendantes (activités),
 $x_i (i = 1, 2, \dots, p)$ représentent les variables indépendantes (descripteurs),
 $a_i (i = 1, 2, \dots, p)$ représentent les coefficients de la régression.

Cet ensemble d'équation peut être écrit sous la forme matricielle suivante :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & \dots & x_{n,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & \dots & x_{n,2} \\ \vdots & \vdots & & & & \\ \vdots & \vdots & & & & \\ 1 & x_{1,p} & x_{2,p} & \dots & \dots & x_{n,p} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} \quad eq. 06$$

Soit de manière condensée :

$$Y = XA + B \quad eq.07$$

Où Y , X , A et B représentent respectivement le vecteur de propriété, la matrice des attributs (descripteurs), la matrice des coefficients et la matrice des erreurs de régression.

La méthode RLM consiste alors à choisir les coefficients a_i de sorte à minimiser la somme des carrés des écarts entre les valeurs calculées de la propriété et celles expérimentales, l'équation du modèle devient donc :

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_{1,i} + \hat{a}_2 x_{2,i} + \dots + \hat{a}_n x_{n,i} \quad eq. 08$$

Ou sous forme matricielle :

$$X \cdot A = Y \quad eq. 09$$

Les coefficients peuvent être obtenus à partir de l'équation matricielle suivante :

$$A = (X^T \cdot X)^{-1} X^T \cdot Y \quad eq. 10$$

5.5. Validation du modèle :

La modélisation vise à fournir un modèle non seulement ajusté aux données expérimentales, mais pouvant également être généralisé à de nouveaux exemples⁷¹. Pour ce faire, plusieurs méthodes de validation, telles que la validation interne, la validation externe et le test de randomisation, sont employés pour estimer la fiabilité du modèle QSAR et pour déterminer sa pertinence pour une application donnée. Différents indicateurs statistiques sont également employés pour déterminer la qualité d'un modèle QSAR, nous citerons ci-dessous les plus répandus⁷¹.

5.5.1. Coefficients et tests statistiques standards :

La qualité du modèle est souvent visualisée sur un diagramme de dispersion, sur lequel sont portées les valeurs calculées de la propriété (activité biologique), en fonction de celles expérimentales. La qualité de la modélisation est meilleure quand les points de ce graphique sont proches de la droite d'ajustement⁷¹. L'ajustement des points à cette droite peut être évalué par le coefficient de détermination R^2 :

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad eq. 11$$

Où :

y : La valeur expérimentale de l'activité

\hat{y} : La valeur calculée de l'activité et

\bar{y} : La valeur moyenne des valeurs calculées de l'activité.

Plus la valeur de R^2 sera proche de 1 plus les valeurs calculées et observées sont corrélées. La déviation standard s est un autre indicateur statistique utilisé :

$$s = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n - p - 1}} \quad eq. 12$$

Où n est le nombre de données d'apprentissage et p le nombre de descripteurs utilisés dans la construction du modèle.

L'indice de *Fisher* F est également employé afin de mesurer le niveau de signification statistique du modèle, c'est-à-dire la qualité du choix des descripteurs constituant le modèle.

$$F = \frac{\sum((\hat{y}_i - \bar{y})^2(n - p - 1))}{(y_i - \hat{y}_i)^2 \cdot n} \quad eq. 13$$

La pertinence des descripteurs dans le modèle est évaluée par t-test de *student* en testant l'hypothèse nulle qui considère le descripteur comme non significatif. Pour une régression linéaire multiple, on suppose que le coefficient a_i , qui lui est associé, est nul.

Cette hypothèse est rejetée (avec un intervalle de confiance α) si le ratio t_i entre a_i et son erreur type $S(a_i)$ atteint la valeur du fractile d'ordre $(1-\alpha/2)$ de la loi de *student* à $(p-n-2)$ degrés de liberté.

$$|t_i| = \left| \frac{a_i}{S(a_i)} \right| > t_{1-\frac{\alpha}{2}}^{p-n-2} \quad eq. 14$$

5.5.2. Validation interne :

La validation interne est souvent la technique la plus employée, dans les études QSAR pour déterminer la stabilité du modèle et de tester l'influence de chaque échantillon de l'ensemble d'apprentissage sur le modèle final⁷². Pour ce faire, on emploie les techniques de la validation croisée (cross validation CV) leave-one-out et leave-*n*-out.

5.5.2.1. La procédure leave-*n*-out :

Cette procédure consiste à extraire chaque fois un certain nombre *n* de molécules de l'ensemble initial d'apprentissage à *k* molécules et à construire un nouveau modèle avec les *k-n* molécules restantes en utilisant les descripteurs choisis (les mêmes descripteurs utilisés dans la construction du modèle original)⁷³.

Le nouveau modèle est ensuite utilisé pour la prédiction sur les *n* molécules retirées. Ce processus est ensuite réitéré *p* fois pour retirer et prédire les valeurs de toutes les molécules de l'ensemble d'entraînement⁷⁴ (figure.6).

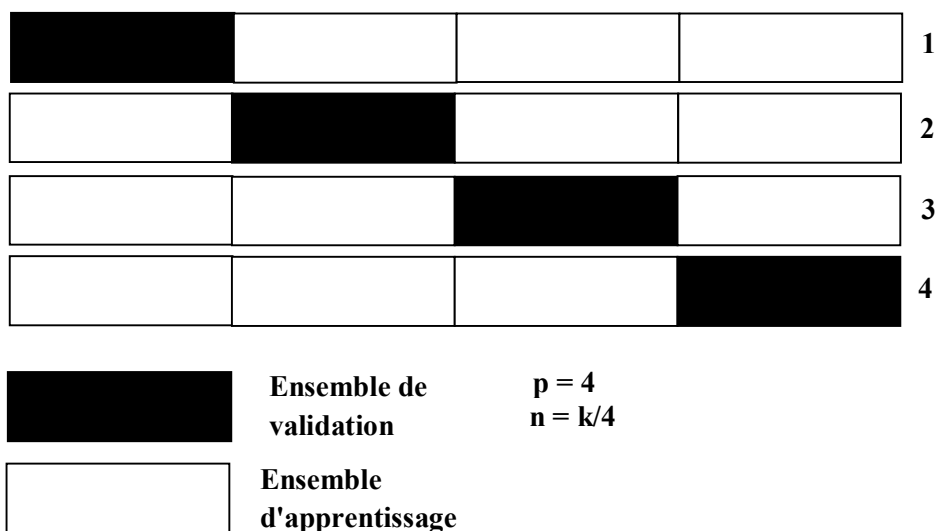


Figure 6 : principe de la validation croisée leave-*n*-out.

5.5.2.2. La procédure leave-one-out:

Pour un nombre N d'exemples d'apprentissage, on retire à chaque itération un exemple I de l'ensemble d'apprentissage initial. Une série d'apprentissage est réalisée pour les $N-1$ molécules restantes et la molécule retirée est prédite par le modèle formé⁷⁵.

La performance des modèles de régression est estimée avec les paramètres statistiques de la validation croisée décrits par les équations ci-dessous :

L'écart type de la validation croisée S_{CV-LOO} :

$$S_{CV-LOO} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{icv})^2} \quad eq. 15$$

Le coefficient de détermination de la validation croisée Q_{CV-LOO}^2 :

$$Q_{cv-loo}^2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^2 - \sum_{i=1}^N (y_i - y_{icv})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad eq. 16$$

Où

y_{icv} : Représente la valeur de l'activité de la molécule i calculée par la méthode de la validation croisée,

\bar{y} : La valeur moyenne de l'activité,

y_i : Les valeurs observées de l'activité.

5.5.2.3. Test de randomisation :

Le test de randomisation est utilisé par le modélisateur pour affirmer que les bonnes corrélations, entre les descripteurs et l'activité, présentées par le modèle QSAR ne sont pas dues à la chance. Pour se faire, les observations sont aléatoirement désorganisées, par exemple dix fois, en changeant aléatoirement la colonne des activités, mais les colonnes des descripteurs restent inchangées. On obtient donc dix modèles possédant des caractéristiques statistiques spécifiques. Si la randomisation des observations conduit à des modèles de prévisions faibles cela signifie que les capacités prédictives du modèle QSAR construit ne sont pas dues aux corrélations de chance^{76,77}.

5.5.3. Validation externe :

Afin de tester de manière fiable le pouvoir prédictif du modèle QSAR, il est nécessaire d'employer un ensemble de validation externe, non employé pour le développement du modèle. Pour se faire, soit un ensemble de molécules divisé en deux : un ensemble d'apprentissage sur lequel le modèle est développé et un ensemble de validation utilisé pour déterminer sa validité externe. Une fois l'ensemble de validation mis en place, il suffit alors d'appliquer le modèle QSAR aux molécules qui le composent et de déterminer la corrélation existant entre les activités calculées et celles expérimentales. Plus cette corrélation est importante, plus le modèle est capable de prédire les activités pour des molécules hors l'ensemble d'apprentissage⁷⁸⁻⁷⁹.

$$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^N (y_i - y_{iext})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad eq.17$$

Où y_{iext} : est les valeurs calculées des molécules composant l'ensemble de validation externe.

5.6. Domaine d'applicabilité :

Le domaine d'applicabilité est la région de l'espace chimique définie par les molécules de l'ensemble d'apprentissage du modèle⁸⁰. Les modèles QSAR ne peuvent pas prédire des propriétés de manière fiable pour l'intégralité des composés chimiques existants. En effet, un modèle QSAR n'est pas destiné à être employé en dehors de son domaine d'applicabilité, c'est-à-dire en dehors de l'espace chimique couvert par son ensemble d'apprentissage.

6. Références :

1. A. Crum Brown, T.R. Fraser, *Trans. Roy.Soc.* 25, 1868, 151-203.
2. Hansch, *Acc. Chem. Res.* 2, 1969, 232-239.
3. Hansch, T. Fujita, *J. Am. Chem. Soc.* 86, 1964, 1616-1626.
4. A. K. Debnath, *Mini Reviews in Medicinal Chemistry.* I, 2001,187-195.
5. M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today.* 3, 2000, 50-57.
6. M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today.* 3, 2000, 28-35.
7. T. I. Oprea, "Chemoinformatics in Drug Discovery" *Ed. WILEY-VCH Verlag. Allemagne*, 2005.
8. E. A. Rekká ; P. N. Kourounakis "Chemistry and Molecular Aspects of Drug Design and Action" *Ed. Taylor & Francis Group, LLC. Etats Unies*, 2008.
9. A. K. Ghose, V. N. Viswanadhan, "Combinatorial Library Design and Evaluation: Principles, Software Tools, and Applications in Drug Discovery" *Ed. Marcel Dekker, Inc. Etats Unies*, 2001.
10. F. Li, H. Wu, L. Li, X. Li, J. Zhao, W. J. G. M. Peijnenburg, *Ecotoxicol. Environm. Saf.* 80, 2012, 273–279.
11. A. Srinivas Reddy, S. Kumar, R. Garg, *J Mol Graph Model.* 28(8), 2010, 852–862.
12. M. Zhao, Z. Li, L. Peng, Y. R. Tang, C. Wang, Z. Zhang, S. Peng, *Bioorg. Med. Chem.* 15, 2007, 2815–2826
13. A. A. Toropov, A. P. Toropova, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, G. De Nucci, *Biochem. Biophys. Res. Comm.* 432(2), 2013, 214–225.
14. A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* 20, 2002, 269-276.
15. A. Tropsha, P. Gramatica, K.V. Gombar, *QSAR Comb. Sci.* 22, 2003, 69-77.
16. P. Gramatica, E. Giani, E. Papa, *J. Mol. Graph. Mod.* 25, 2007, 755-766.
17. J. Verma, VM. Khedkar, EC. Coutinho, *Curr Top Med Chem.*10(1), 2010, 95-115.
18. RC. Khunt, VM. Khedkar, RS. Chawda, NA. Chauhan, AR. Parikh, EC. Coutinho, *Bioorg. Med. Chem. Lett.* 22(1), 2012, 666-78.
19. V.V.S. Rajendra Prasad, G. Deepak Reddy, D. Appaji, G.J. Peters, Y.C. Mayur, *J. Mol. Graph. Model.* 40, 2013, 116-124
20. F. Ballante, R. Ragno, *J. Chem. Inf. Model.* 52 (6), 2012, 1674–85.
21. T. Puzyn, J. Leszczynski, M. T. D. Cronin, « Recent advances in QSAR studies-Methods and applications » *Springer Dordrecht Heidelberg London New York*, 2010.
22. M. Tichý, *Science.Tot. Env.* 109, 1991, 407-410.

23. B. Peric, J. Sierra, E. Martí, R. Cruañas, M. A. Garau, *Ecotoxicol. Environ. Saf.* 115, 2015, 257-262.
24. O. M. Bautista-Aguilera, G. Esteban, I. Bolea, K. Nikolic, D. Agbaba, I. Moraleda, I. Iriepa, A. Samadi, E. Soriano, M. Unzeta, J. Marco-Contelles, *Europ. J. Med. Chem.* 75, 2014, 82-95.
25. D. Fourches, E. Muratov, A. Tropsha, *J. Chem. Inf. Model.* 50, 2010, 1189–1204
26. A. Tropsha, *Mol Inform.* 29, 2010, 476–488
27. S. Chatterjee, A. S. Hadi, B. Price, « Regression analysis by example » *Wiley-Interscience, New York.* 2000
28. H. T. N. Phuong, « Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d’analogues Benzo[c]phénanthridiniques » *Thèse de doctorat, Université d’Angers. France,* 2007.
29. N. L. Mai, Y-M Koo, *Biochem. Engin. J.* 87, 2014, 33-40.
30. J. Wu, Y. Wang, Y. Shen, *Chemomet. Intel. Lab. Syst.* 131, 2014, 22-30.
31. O. Mekenyan, D. Bonchev, *Acta Pharm. Jugosl.* 36, 1986, 225-237.
32. <http://oasis-lmc.org>
33. R. Guha, J. R. Serra, P. C. Jurs, *J. Mol. Graph. Mod.* 23, 2004, 1–14.
34. <http://www.netsci.org/Resources/Software/Modeling/CADD/adapt.html>
35. A. R. Katritzky, V. S. Lobanov, M. Karelson, « Manuel du logiciel CODESSA » *Gainesville, FL. University of Florida.* 1994
36. www.semichem.com/codessa/codessa.html
37. AR. Katritzky, M. Karelson, R. Petrukhin, CODESSA PRO, University of Florida 2001-2005. <http://www.codessa-pro.com/>
38. R. Todeschini, Milano Chemometrics and QSPR Group, Talete srl, Milan, Italy. (http://www.talete.mi.it/products/dragon_description.htm).
39. <http://www.edusoft-lc.com/molconn/>
40. <http://openmopac.net/>
41. CA. Lipinski, *Drug. Discov. Today Techno.* 1(4), 2004, 337–341.
42. CA. Lipinski, *Drug. Discov. Today.* 8, 2003, 6-12.
43. H. S. Ramane, D. S. Revankar, A. B. Ganagi, *J. Indones. Math. Soc.* 18(1), 2012, 57-66.
44. J. Gasteiger, “Handbook of Chemoinformatics: From Data to Knowledge” *Ed. WILEY-VCH Verlag, Weinheim. Allemagne,* 2003.

45. M. Karelson “Molecular Descriptors in QSAR/QSPR” Ed. Wiley- Interscience. *Etats Unies*, 2000.
46. R. Todeschini; V. Consonni “Handbook of Molecular Descriptors: Methods and Principles in Medicinal Chemistry” Ed. Wiley – VCH, Weinheim. *Allemagne*, 2000.
47. K. KRAIM, « études de qsar sur des activités biologiques utilisant des produits d'origines naturels » *Thèse de doctorat, Université de Guelma-Algérie*, 2009.
48. É. Matzner-Løber, « Régression: Théorie et applications », *Springer Science & Business Media*, 2007.
49. Daniel Borcard, « Régression multiple - corrélation multiple et partielle », *Département de sciences biologiques Université de Montréal*, 2009
50. R Leardi, R Boggia, M Terrile, *J Chemom.* 6, 1992, 267–281.
51. H. Kubinyi, *J Chemom.* 10, 1996, 119–133.
52. P. Gramatica, P. Pilutti, E. Papa, *J. Chem. Inf. Comp. Sci.* 44, 2004, 1794–1802.
53. E. Papa, F. Villa, P. Gramatica, *J Chem Inf Model.* 45, 2005, 1256–1266
54. H. Liu, E. Papa, P. Gramatica, *Chem Res Toxicol.* 19, 2006, 1540–1548
55. P Gramatica, E Giani, E Papa, *J. Mol. Graph. Model.* 25, 2007, 755–766.
56. B. Bhatarai, P. Gramatica, *Chem Res Toxicol.* 23, 2010, 528–539.
57. M. Shahlaei, A. Madadkar-Sobhani, L. Saghaie, A. Fassihi, *Expert Systems with Applications*, 39(6), 2012, 6182-6191.
58. E. Pourbasheer, R. Aalizadeh, M. R. Ganjali, P. Norouzi, J. Shadmanesh, *J. Saud. Chem. Soc.*, 18(5), 2014, 681-688.
59. H. Gao, M. S. Lajiness, J. V. Drie, *J. Mol. Graph. Mod.*, 20(4), 2002, 259-268.
60. Chemomet. Intel. Lab. Syst, 139, 2014, 168-174.
61. M. O. Taha, A. M. Qandil, D. D. Zaki, M. A. AlDamen, M. Sarkhosh, N. Khorshidi, A. Niazi, R. Leardi, *Europ. J. Med. Chem.*, 40(7), 2005, 701-727.
62. X. Z. Wang, B. Perston, Y. Yang, T. Lin, J. A. Darr, *Chem. Engin. Res. Design*, 87(10), 2009, 1420-1429.
63. W. McCulloch, W. Pitts, *Bull. Math. Biophys.* 5, 1943, 115-133.
64. L. Douali, A. R. Schmitzer, D. Villemin, A. Jarid, D. Cherqaoui, *Phys. Chem. News.* 34, 2007, 131-144.
65. D. W. Salt, N. Yildiz, D. J. Livingstone, C. J. Tinsley, *Pestic. Sci.* 36, 1992, 161-170.
66. I. OTMANI, « l'analyse discriminante et le perceptron Multicouche », *Mémoire de Magister, Université Mentouri-Algérie*, 2011.

67. G. IDIOU, « Régression et modélisation par les réseaux de neurones », *Mémoire de Magister, Université Mentouri-Algérie*, 2009.
68. R. Rakotomalala, « Pratique de la Régression Linéaire Multiple : Diagnostic et sélection de variables », *version 2, Université Lumière Lyon 2*, 2011.
69. J. Confais, M. Le Guen, « premiers pas en regression lineaire avec SAS », *Revue MODULAD*, N° 35, 2006, 220-363.
70. K. Roy, P. P. Roy, *Europ. J. Med. Chem.*, 44(7), 2009, 2913-2922.
71. E. X. Esposito, A. J. Hopfinger, J. D. Madura, *Methods in Molecular Biology*. 275, 2004, 131-213
72. A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* 20, 2002, 269–276
73. DM. Hawkins, *J. Chem. Inf. Comput. Sci.* 44, 2004,1–12
74. A. Tropsha, P. Gramatica, VK. Gombar, *QSAR.Comb. Sci.* 22, 2003, 69–77.
75. A. Tropsha, P. Gramatica, VK. Gombar, *QSAR.Comb. Sci.* 22, 2003, 69–77.
76. L. Eriksson, J. Jaworska, A. Worth, *Environ. Health. Perspect.* 111, 2003,1361–1375
77. P. Gramatica, *QSAR. Comb. Sci.* 26, 2007, 694–701.
78. LM. Shi, H. Fang, W. Tong, *J. Chem. Inf. Comput. Sci.* 41, 2001, 186–195.
79. N. Nicola Chirico, P. Gramatica, *J. Chem. Inf. Model.* 51(9), 2011, 2320–2335.
80. T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. TD Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. JM. van de Sandt, W. Tong, G. Veith, C. Yang, *ATLA* 33, 2005, 155–173.

Chapitre 3

Modélisation de la relation quantitative non linéaire structure-activité inhibitrice de l' α -glucosidase par les dérivés de Xanthone et de Curcuminoïde.

1. Introduction :

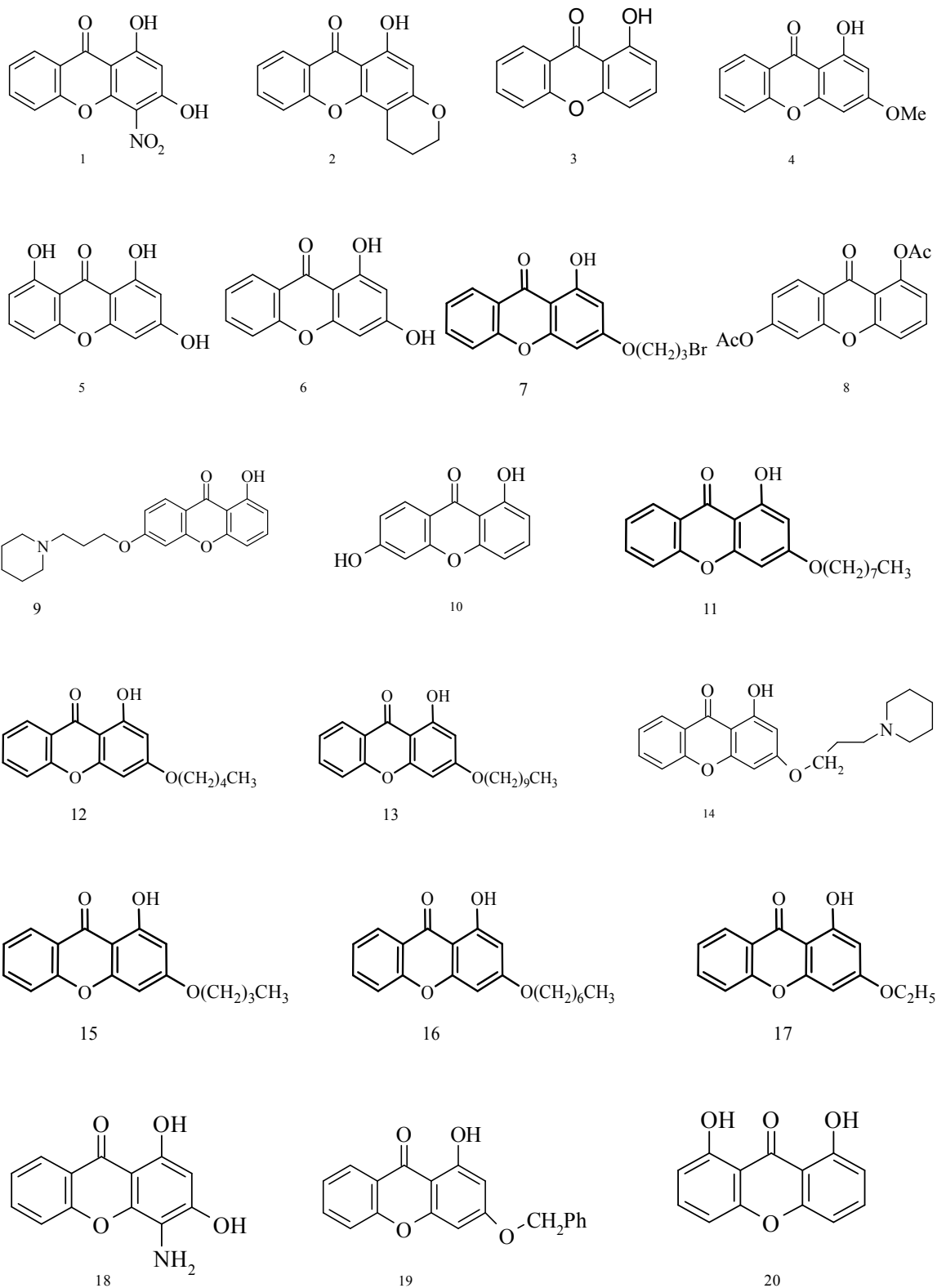
La régression linéaire multiple est parmi les méthodes statistiques la plus utilisée, du fait de la simplicité de ses modèles linéaires et de la facilité de l'interprétation de ses résultats¹. Cependant, cette méthode présente des limitations, car elle dépend des lois et des hypothèses statistiques, notamment dans le cas où les données ne sont pas distribuées normalement². Les réseaux de neurones ont été présentés comme une méthode alternative aux méthodes statistiques, potentiellement intéressants lorsque les effets des variables sont non linéaires³. Les réseaux de neurones présentent plusieurs avantages tels que la meilleure capacité de prédiction, d'adaptation et de généralisation au-delà de l'échantillon étudié, une meilleure stabilité des coefficients par une moindre sensibilité aux points aberrants et absence d'hypothèses sur la distribution des variables comme la normalité en régression linéaire par exemple².

Dans ce chapitre nous présentons une étude non linéaire de la relation quantitative structure-activité inhibitrice de l' α -glucosidase par les dérivés de xanthone et de curcuminoïde. Nous avons fait appel aux réseaux de neurones et aux algorithmes génétiques en tant que méthodes d'apprentissage et de sélection de variables respectivement. L'architecture du réseau de neurones utilisée dans cette étude est constituée de trois couches : couche d'entrée, couche cachée et couche de sortie⁴. Les résultats obtenus sont analysés, validés et comparés avec ceux obtenus par la régression linéaire multiple.

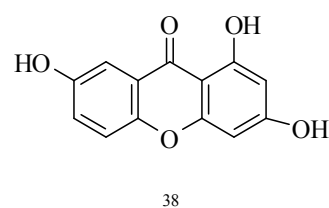
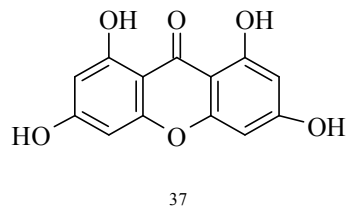
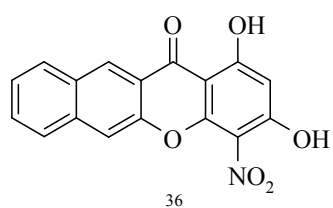
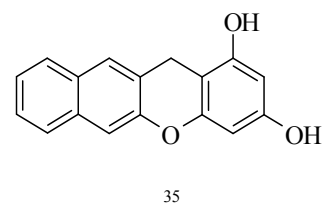
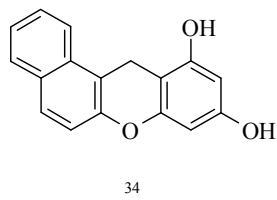
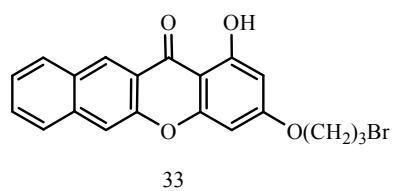
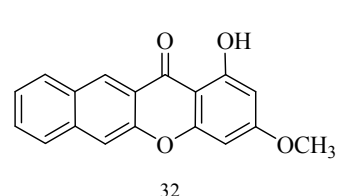
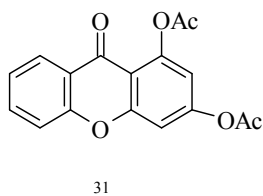
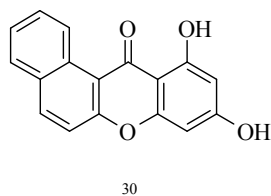
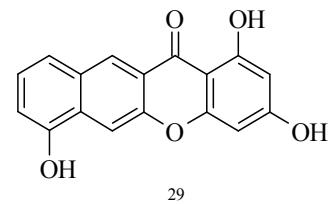
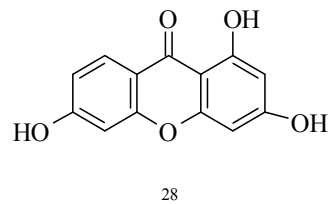
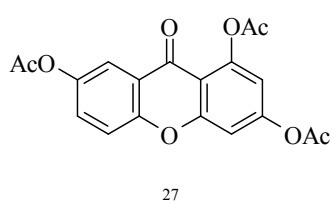
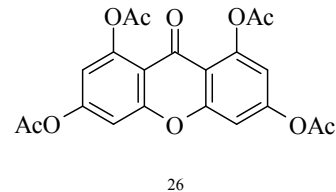
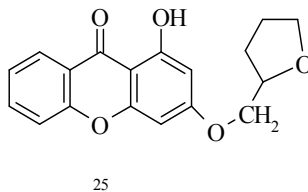
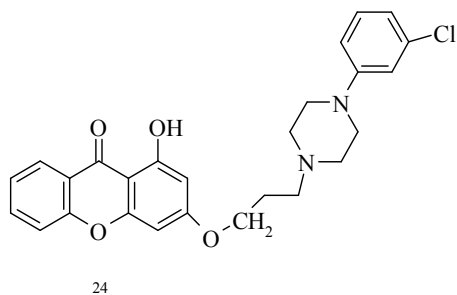
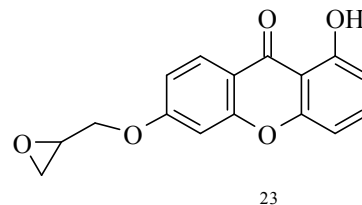
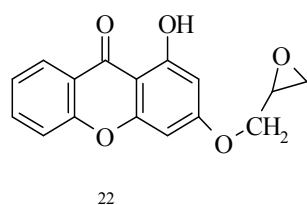
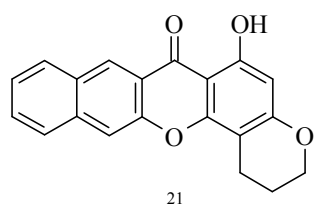
2. Ensemble de molécules :

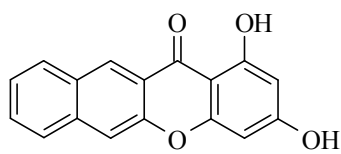
Dans cette étude, l'ensemble à modéliser est constitué de 57 molécules dérivées de xanthonnes et de curcuminoïdes, qui ont montré une activité inhibitrice vis-à-vis de l' α -glucosidase^{5,6}. Leurs structures moléculaires et leurs activités sont illustrées dans la figure 1 et le tableau 1 respectivement.

Dérivés du xanthone

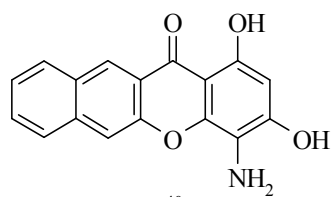


Chapitre 3: Modélisation de la relation quantitative non linéaire structure activité-inhibitrice de l' α -glucosidase par les dérivés de Xanthone et de Curcuminoïde.

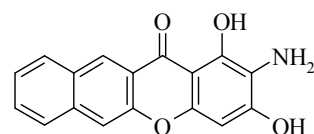




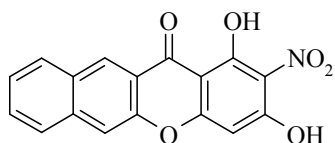
39



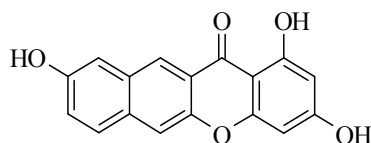
40



41

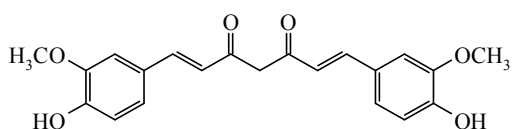


42

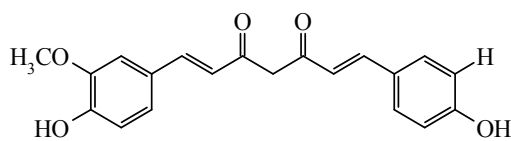


43

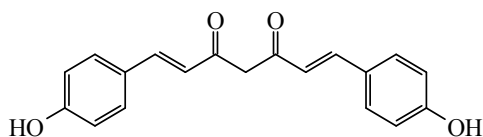
Dérivés du Curcuminoïde



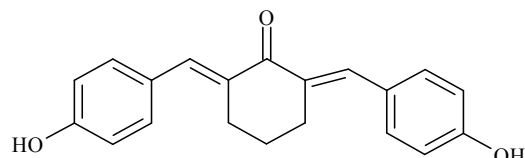
44



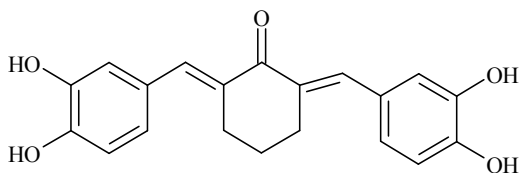
45



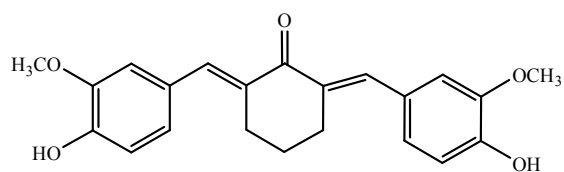
46



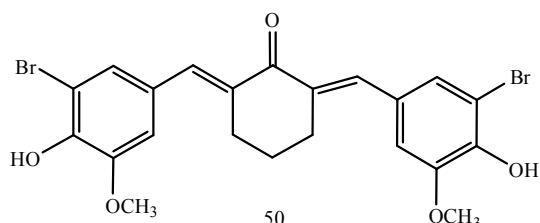
47



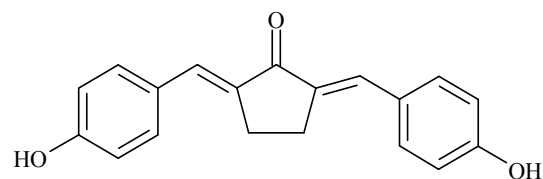
48



49



50



51

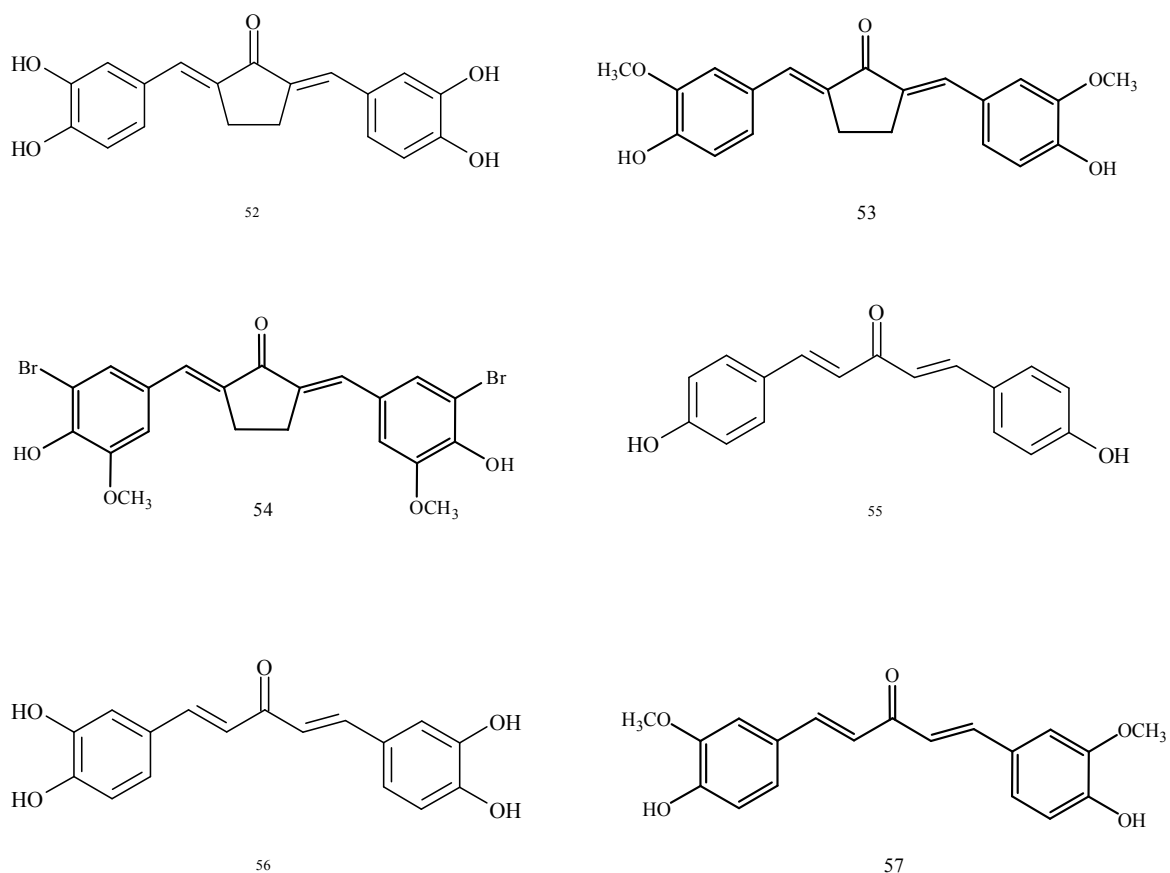


Figure 1 : Structures développées des dérivés du xanthone et du curcuminoïde

L'activité biologique inhibitrice est exprimée en terme d' IC_{50} : concentration micro molaire en inhibiteur, nécessaire pour inhiber 50% de l'activité enzymatique. Dans notre études nous avons exprimé l'activité inhibitrice par le rapport logarithmique pIC_{50} [$\text{Log}(1/IC_{50})$].

Tableau 1: Valeurs expérimentales de pIC_{50} et celles calculées par le modèle RNA pour les 57 molécules listées dans la figure 1.

molécule	pIC_{50} exp	pIC_{50} calc	ensemble	molécule	pIC_{50} exp	pIC_{50} cal	ensemble
1	-2,371	-2,134	Train	2	-2,297	-2,306	Val
3	-2,249	-2,059	Train	6	-2,166	-2,03	Val
5	-2,206	-2,065	Train	10	-2,114	-2,052	Val
7	-2,143	-1,941	Train	14	-2,062	-2,131	Val
9	-2,119	-2,072	Train	18	-1,992	-1,511	Val
11	-2,092	-2,089	Train	22	-1,823	-1,589	Val
13	-2,063	-2,073	Train	53	-1,723	-1,713	Val
15	-2,056	-2,156	Train	45	-1,63	-1,612	Val
17	-2,01	-2,155	Train	44	-1,571	-1,489	Val
19	-1,961	-2,057	Train	51	-1,512	-1,337	Val
21	-1,828	-1,827	Train	50	-1,467	-1,471	Val
23	-1,803	-1,694	Train	35	-1,303	-1,149	Val
25	-1,724	-1,658	Train	39	-0,919	-1,168	Val
26	-1,696	-1,578	Train	43	-0,763	-0,721	Val
27	-1,667	-1,556	Train				
28	-1,618	-1,73	Train				
47	-1,575	-1,553	Train				
57	-1,568	-1,482	Train				
54	-1,53	-1,506	Train				
31	-1,504	-1,799	Train				
33	-1,473	-1,714	Train				
34	-1,444	-1,276	Train				
55	-1,338	-1,271	Train				
36	-1,233	-1,246	Train				
38	-0,968	-1,184	Train				
40	-0,903	-0,867	Train				
42	-0,771	-0,785	Train				
48	-0,447	-0,469	Train				
56	-0,204	-0,224	Train				
				molécule	pIC_{50} exp	pIC_{50} cal	ensemble
				4	-2,238	-2,306	Test
				8	-2,123	-2,03	Test
				12	-2,082	-2,052	Test
				16	-2,045	-2,131	Test
				20	-1,913	-1,511	Test
				24	-1,791	-1,589	Test
				49	-1,672	-1,713	Test
				29	-1,601	-1,612	Test
				30	-1,543	-1,489	Test
				32	-1,496	-1,337	Test
				46	-1,362	-1,471	Test
				37	-1,167	-1,149	Test
				41	-0,799	-1,168	Test
				52	-0,415	-0,721	Test

L'ensemble des molécules est divisé en trois sous ensembles : un sous ensemble d'apprentissage constitué de 29 molécules (environ 50%), utilisé pour la construction du modèle QSAR et deux sous ensembles de 14 molécules chacun (environ 25%) réservés pour la validation interne et la validation externe.

3. Dessin et optimisation des structures :

Les structures des molécules ont été dessinées et optimisées à l'aide du logiciel Hyperchem⁷ en utilisant la méthode semi empirique AM1. Le programme de calcul est exécuté sur un PC portable dual core 2,4GHz.

4. Génération des descripteurs moléculaires :

Les descripteurs moléculaires utilisés sont générés à l'aide du logiciel Dragon 6, programme utilisée pour le calcul de descripteurs moléculaires, qui comprend 4887 descripteurs répartis en 29 blocs⁸.

5. Sélection des variables et formation du modèle :

Dans ce travail nous avons établi une relation QSAR non linéaire en utilisant la méthode des réseaux de neurones artificiels (RNA) couplée avec les algorithmes génétiques (AG) en tant que méthode de sélection de variables. Les cinq descripteurs moléculaires sélectionnés par les AG sont rassemblés dans le tableau 2.

Tableau 2 : Glossaire des descripteurs sélectionnés par les algorithmes génétiques.

Descripteur	Description
MATS7v	Moran autocorrelation of lag7/pondéré par le volume atomique de Van Der Waals
Mor15u	3D MorSE-signal15/non pondéré
H5u	H autocorrelation of lag5/non pondéré
R4e+	R maximal autocorrelation of lag4/ pondéré par l'électronégativité de Sanderson
nArOR	Nombre d'éthers aromatiques.

6. Utilisation des réseaux de neurones dans la construction du modèle non linéaire :

Dans ce travail nous présentons une étude non linéaire de la relation quantitative structure-activité inhibitrice de l' α -glucosidase par les dérivés de xanthone et de curcuminoïde. Pour se faire nous avons fait appel à la méthode des réseaux de neurones artificiels (RNA).

Le réseau utilisé dans cette étude est un réseau de neurones à multicouches, constitué de trois couches :

- Une couche d'entrée dont les neurones reçoivent l'information présentée au réseau. Cette couche est constituée de six neurones, cinq d'entre eux représentent les cinq descripteurs sélectionnés et le sixième neurone une valeur constante, appelée biais, toujours égale à 1.
- Une couche de sortie qui fournit les résultats de traitement réalisé par le réseau artificiel. Cette couche contient un seul neurone représentant l'activité biologique pIC_{50} .
- Une couche intermédiaire appelée couche cachée, elle contient un nombre variable de neurones.

Les neurones d'une même couche ne sont pas reliés entre eux, en revanche les neurones de chaque couche sont connectés à ceux de la couche suivante, chaque connexion d'un neurone j avec un autre i est associée à un coefficient de pondération synaptique W_{ij} appelé poids, qui détermine l'intensité de l'influence des entrées (descripteurs).

L'entrée d'un neurone est fonction des signaux provenant des autres neurones et des intensités des connexions. Cette fonction est la somme des entrées pondérées de leurs poids synaptiques.

Le fonctionnement du réseau de neurone utilisé dans cette étude, se déroule selon la méthode de rétropropagation. L'objectif de cette méthode est d'adapter les paramètres W_{ij} de façon à minimiser la fonction du coût, exprimée par l'erreur quadratique (MSE) entre la sortie donnée par le réseau (activité calculée) et la sortie désirée (activité expérimentale), en se basant sur la méthode des gradients itérative (eq01).

$$E = \frac{1}{2} \sum_{i=1}^N E_i^2 = \frac{1}{2} \sum_{i=1}^N [y_i(\text{exp}) - y(\text{calc})]^2 \quad \text{eq01}$$

Avec :

N : nombre de molécules dans le sous ensemble d'apprentissage.

E_i : représente l'erreur de la $i^{\text{ème}}$ donnée.

$y_i(\text{calc})$: représente la sortie (activité) calculée de la $i^{\text{ème}}$ donnée du sous ensemble d'apprentissage.

$y_i(\text{exp})$: représente la sortie (activité) expérimentale de la $i^{\text{ème}}$ donnée du sous ensemble d'apprentissage.

Le fonctionnement de l'algorithme du réseau de neurones utilisé se déroule selon deux étapes :

Dans la première étape, on commence l'entraînement du réseau par un choix aléatoire de poids synaptiques, le flux de l'information se propage de la couche d'entrée vers la couche de sortie, et chaque neurone de la couche cachée reçoit l'activation des neurones de la couche d'entrée qui lui sont connectés. La fonction d'activation pondérée (S_i) des neurones de la couche cachée est exprimée par l'équation suivante :

$$S_i = \sum_j^d W_{ij} X_j + \theta_i \quad \text{eq02}$$

Avec :

W_{ij} : représente le poids de la connexion entre le neurone j de la couche d'entrée et le neurone i de la couche cachée.

X_j : représente les entrées au neurone i (descripteurs).

θ_i : représente le poids de la connexion entre le neurone i et le biais de la couche d'entrée.

d : représente le nombre de neurones d'entrée.

Lorsque le niveau de l'activation pondérée du neurone i de la couche cachée dépasse un seuil d'activation (θ_i), une fonction d'activation (f) sigmoïdale détermine, à partir de l'activation pondérée (S_i) du neurone i , son activation réelle t_i , donnée par l'équation 3 :

$$t_i = f(S_i) = (1 + \exp - S_i)^{-1} \quad \text{eq03}$$

L'information continue à se propager jusqu'au neurone de la couche de sortie, qui produit la valeur calculée de l'activité \hat{y} exprimée par la combinaison linéaire suivante :

$$\hat{y} = \sum_{i=1}^A V_i t_i + \theta = \sum_{i=1}^A V_i f(S_i) + \theta \quad eq04$$

Avec :

V_i : représente le poids de la connexion du neurone de sortie avec le neurone i de la couche cachée.

θ : Poids de la concentration de la sortie avec le biais de la couche cachée.

Dans la deuxième étape, on calcule, pour toutes les molécules du sous ensemble d'apprentissage, l'erreur correspondante et les gradients de l'erreur par rapport à tous les poids synaptiques. Ensuite, l'algorithme se rétro-propage de la couche de sortie vers la couche d'entrée en ajustant les poids synaptiques. Ce processus est répété plusieurs fois jusqu'à ce que les sorties du réseau soient proches des sorties désirées. Une fois un minimum acceptable de la fonction du coût (MSE) est obtenu, les valeurs des poids sont fixées et utilisées pour la construction et la validation du modèle non linéaire.

L'adaptation des poids est basée sur la formule itérative suivante :

$$\Delta V_i (n + 1) = V_i (n + 1) - V_i(n) = -\alpha \frac{\partial E}{\partial V_i} \quad eq05$$

$$\Delta W_{ij}(n + 1) = W_{ij}(n + 1) - W_{ij}(n) = -\alpha \frac{\partial E}{\partial W_{ij}} \quad eq06$$

Où :

n et $n+1$: représentent le nombre d'itération dans la phase d'apprentissage.

α : représente le facteur ou pas d'apprentissage, détermine le taux d'apprentissage.

La même équation est utilisée pour l'ajustement des poids des biais θ et θ_i :

$$\Delta \theta_i (n + 1) = \theta_i (n + 1) - \theta_i(n) = -\beta \frac{\partial E}{\partial \theta_i} \quad eq07$$

$$\Delta\theta(n + 1) = \theta(n + 1) - \theta(n) = -\beta \frac{\partial E}{\partial \theta} \quad eq08$$

β : représente le pas d'apprentissage.

Les étapes réalisées par l'algorithme de rétropropagation sont résumées dans la figure 2 :

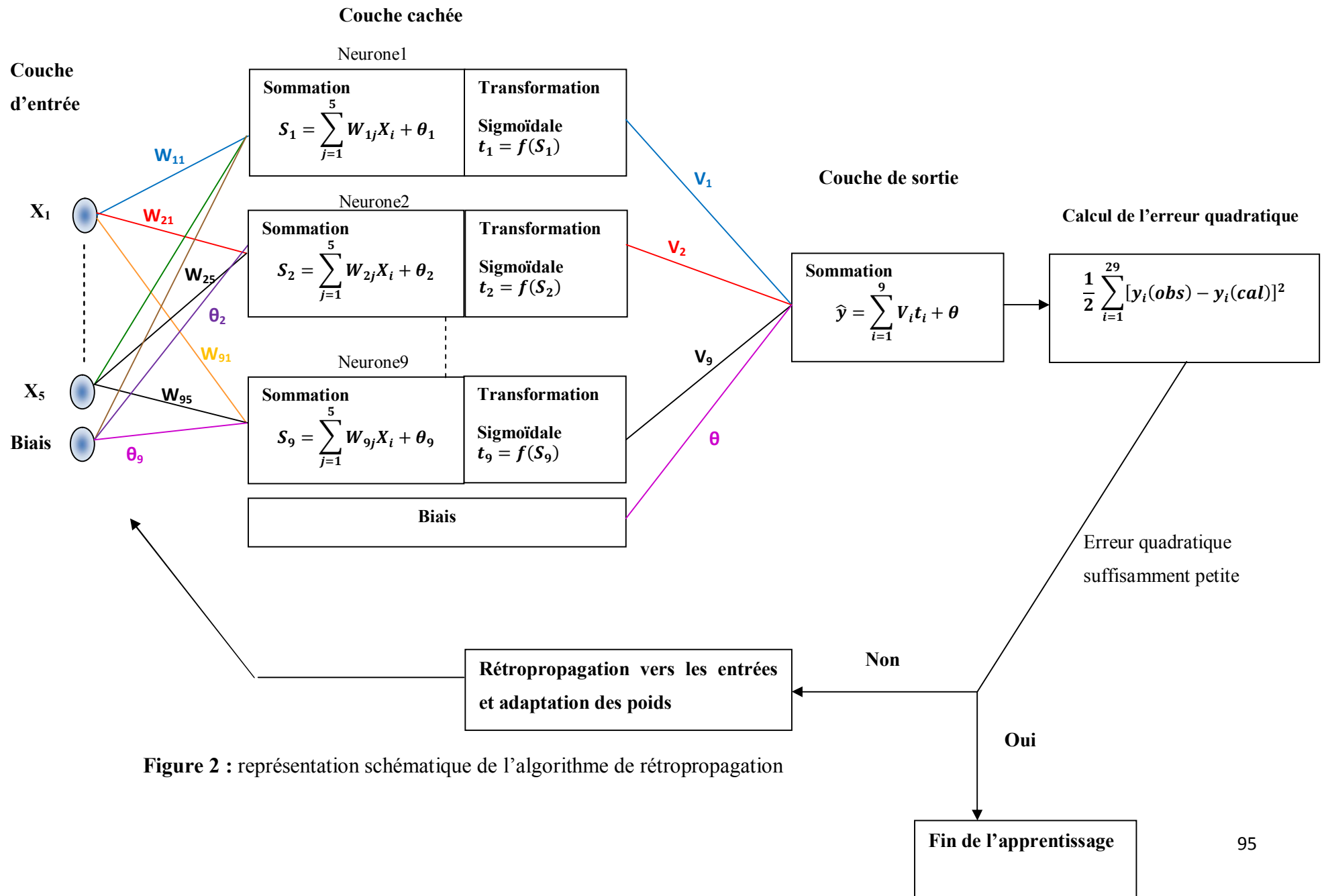


Figure 2 : représentation schématique de l'algorithme de rétropropagation

7. Résultats et discussion :

Lors d'une étude antérieure réalisée au sein de notre équipe de recherche⁹, nous avons établi une relation structure-activité inhibitrice de l' α -glucosidase avec 57 molécules dérivées de Xanthone et de curcuminoïde en utilisant la régression linéaire multiple (RLM) couplée avec les algorithmes génétiques en tant que méthode de sélection de variable. Cinq descripteurs ont été sélectionnés (cités dans le tableau 2) en tant que facteurs ayant une influence significative sur l'inhibition de l' α -glucosidase. Les résultats obtenus, à l'aide du logiciel MINITAB15.0, sont rassemblés dans le tableau 3.

Tableau 3 : Valeurs de pIC₅₀ expérimentales et valeurs calculées par le modèle linéaire RLM.

No de molécule	pIC ₅₀ exp.	pIC ₅₀ calc.	MATS7v	Mor15u	H5u	R4e+	nArOR
TSET							
1	-2,371	-2,071	-0,681	1,055	0,484	0,074	1,000
2	-2,297	-2,086	-0,205	0,717	0,363	0,097	2,000
4	-2,238	-2,162	-0,357	1,126	0,573	0,072	2,000
5	-2,206	-2,024	-0,538	1,093	0,377	0,064	1,000
7	-2,143	-2,026	-0,111	0,811	0,468	0,078	2,000
8	-2,123	-1,94	-0,248	0,764	0,528	0,038	1,000
9	-2,119	-2,06	-0,217	0,362	1,510	0,059	2,000
11	-2,092	-2,105	-0,226	0,322	1,623	0,050	2,000
12	-2,082	-2,139	-0,209	0,558	1,104	0,058	2,000
14	-2,062	-2,182	-0,256	0,204	1,488	0,064	2,000
16	-2,045	-2,13	-0,242	0,429	1,470	0,052	2,000
17	-2,010	-2,163	-0,240	0,915	0,548	0,071	2,000
18	-1,992	-1,609	-0,501	1,456	0,481	0,086	1,000
19	-1,961	-2,042	-0,236	1,080	0,692	0,066	2,000
20	-1,913	-2,035	-0,346	0,717	0,333	0,064	1,000
22	-1,823	-1,738	0,099	0,990	0,521	0,067	2,000
23	-1,803	-1,925	-0,044	1,005	0,552	0,065	2,000
24	-1,791	-1,724	-0,073	0,270	1,989	0,060	2,000
25	-1,724	-1,495	0,163	1,130	0,778	0,063	2,000
26	-1,696	-1,564	-0,262	1,176	1,238	0,020	1,000
28	-1,618	-2,026	-0,604	1,310	0,371	0,073	1,000
29	-1,601	-1,19	-0,037	1,391	0,655	0,061	1,000

Chapitre 3: Modélisation de la relation quantitative non linéaire structure activité-inhibitrice de l' α -glucosidase par les dérivés de Xanthone et de Curcuminoïde.

30	-1,543	-1,589	-0,206	1,300	0,542	0,048	1,000
31	-1,504	-1,758	-0,262	0,922	0,712	0,051	1,000
32	-1,496	-1,718	-0,185	1,529	0,876	0,064	2,000
33	-1,473	-1,744	-0,098	1,124	0,807	0,079	2,000
34	-1,444	-1,215	-0,244	1,370	0,631	0,102	1,000
35	-1,303	-1,009	-0,160	1,366	0,658	0,115	1,000
36	-1,233	-1,288	-0,308	1,463	0,481	0,109	1,000
38	-0,968	-1,442	-0,204	1,420	0,422	0,074	1,000
40	-0,903	-1,058	-0,213	1,879	0,752	0,079	1,000
41	-0,799	-0,873	-0,213	1,783	0,899	0,106	1,000
42	-0,771	-1,052	-0,308	1,377	0,696	0,158	1,000
43	-0,763	-0,723	0,208	1,720	0,695	0,065	1,000
45	-1,630	-1,231	-0,076	0,742	1,382	0,063	1,000
46	-1,362	-1,529	-0,274	0,328	0,931	0,048	0,000
48	-0,447	-0,694	0,071	0,985	0,974	0,066	0,000
49	-1,672	-1,764	0,105	0,161	1,701	0,052	2,000
50	-1,467	-1,434	-0,049	1,479	1,399	0,048	2,000
51	-1,512	-1,236	-0,228	0,845	0,585	0,058	0,000
52	-0,415	-0,556	0,032	1,473	0,687	0,073	0,000
53	-1,723	-1,735	0,105	0,184	1,700	0,053	2,000
54	-1,530	-1,479	-0,063	1,641	1,188	0,044	2,000
55	-1,338	-1,492	-0,160	0,452	0,680	0,039	0,000
PSET							
3	-2,249	-1,812	-0,179	0,637	0,287	0,076	1,000
6	-2,166	-1,974	-0,477	1,000	0,327	0,075	1,000
10	-2,114	-2,012	-0,477	0,943	0,331	0,073	1,000
13	-2,063	-2,131	-0,200	0,060	1,862	0,045	2,000
15	-2,056	-2,218	-0,259	0,627	0,879	0,064	2,000
21	-1,828	-1,707	-0,127	1,149	0,675	0,095	2,000
27	-1,667	-1,571	-0,047	0,859	0,978	0,025	1,000
37	-1,167	-1,991	-0,635	1,437	0,424	0,061	1,000
39	-0,919	-1,401	-0,176	1,350	0,596	0,065	1,000
44	-1,571	-1,371	0,038	1,232	1,420	0,061	2,000
47	-1,575	-1,618	-0,164	-0,243	0,873	0,056	0,000
56	-0,204	-0,53	0,208	1,131	0,797	0,059	0,000
57	-1,568	-1,384	0,238	1,131	1,320	0,035	2,000

L'équation décrivant le modèle QSAR par rapport à ces descripteurs, en utilisant la RLM-AG, est représentée avec ses paramètres statistiques comme suit :

$$pIC_{50} = - 2,01 + 1,17 \text{ MATS7v} + 0,485 \text{ Mor15u} + 0,414 \text{ H5u} + 6,05 \text{ R4e}^+ - 0,480 \text{ nArOR} \quad (\text{Equ.9})$$

$$N = 44; \quad R^2 = 85,70\%; \quad F = 45,53; \quad s = 0,197; \quad Q^2_{cv-100} = 0,815; \quad s_{cv-100} = 0,208; \quad Q^2_{Pset} = 0,66$$

Où :

N : représente le nombre de molécules constituant le sous ensemble d'apprentissage (TSET).

s : représente l'écart type de la régression.

R^2 : représente le coefficient de détermination de la régression.

F : représente le test de Fischer.

s_{cv-100} : représente l'écart type issue de la validation croisée.

Q^2_{cv-100} : représente le coefficient de détermination issue de la validation croisée.

Q^2_{Pset} : représente le coefficient de détermination issue de la validation externe.

Dans la présente étude, nous présentons dans un premier temps la construction d'un modèle QSAR non linéaire, décrivant la relation quantitative structure-activité inhibitrice de l' α -glucosidase par les 57 molécules dérivées du xanthone et du curcuminoïde, en utilisant les réseaux de neurones artificiels en tant que méthode d'apprentissage. Dans un second temps nous procéderons à une comparaison entre les résultats obtenus par ce modèle linéaire et ceux obtenus par le modèle linéaire utilisant la régression linéaire multiple (RLM). Les réseaux de neurones artificiels employés dans cette étude ont été générés en utilisant les cinq descripteurs apparus dans le modèle MLR. En effet, nous avons divisé, dans une première étape, l'ensemble total de molécules en trois sous ensembles: un sous ensemble d'apprentissage, un sous ensemble de validation et un sous ensemble de test. Le sous ensemble d'apprentissage, composé de 29 molécules, a été utilisé pour la génération du modèle RNA. Le sous ensemble de validation, composé de 14 molécules, a été utilisé pour la validation interne du modèle. Le sous ensemble de test, composé également de 14 molécules, a été utilisé pour évaluer et généraliser le modèle généré.

L'architecture RNA utilisée dans cette étude est composée de trois couches: la couche d'entrée comprenant cinq neurones (MATS7v, Mor15u, H5u, R4e +, et nArOR), la couche cachée et la couche de sortie (comprenant un neurone pIC_{50}). Les neurones de la couche

d'entrée envoient des signaux vers les neurones de la couche cachée et enfin à la couche de sortie selon l'algorithme de rétro-propagation. Cette procédure est suivie par un réajustement des poids de connexion conformément à l'erreur de prédiction. Les poids de connexion sont initialisés au hasard et modifiés au cours du processus d'apprentissage en utilisant l'algorithme de rétro-propagation (Levenberg-Marquardt)¹⁰⁻¹¹.

Un problème majeur dans les réseaux de neurones consiste en la façon de déterminer le nombre de neurones dans la couche cachée. Bien qu'il n'y ait pas de règle rigoureuse utilisée pour résoudre ce problème, certains auteurs ont proposé un paramètre ρ qui peut servir à déterminer le nombre d'unités cachées^{12,14} définies dans l'équation 10.

$$\rho = \frac{\text{Nombre de données dans l'ensemble d'apprentissage}}{\text{Nombre de connexion dans le réseau de neurones}} \quad \text{eq.10}$$

Une deuxième méthode pratique, a été utilisée dans ce travail pour déterminer le nombre de neurones dans la couche cachée, en utilisant la procédure essai-erreur. Cette procédure commence l'optimisation par un petit nombre de neurones que l'on peut augmenter si nécessaire jusqu'à l'obtention d'un nombre optimal de neurones. En effet, nous avons optimisé, dans cette étude, le nombre de neurones cachés en commençant par un seul neurone et arrivant jusqu'à 12 neurones. Les résultats obtenus pour chaque cas sont représentés sur la figure 3 et dans tableau 3 respectivement.

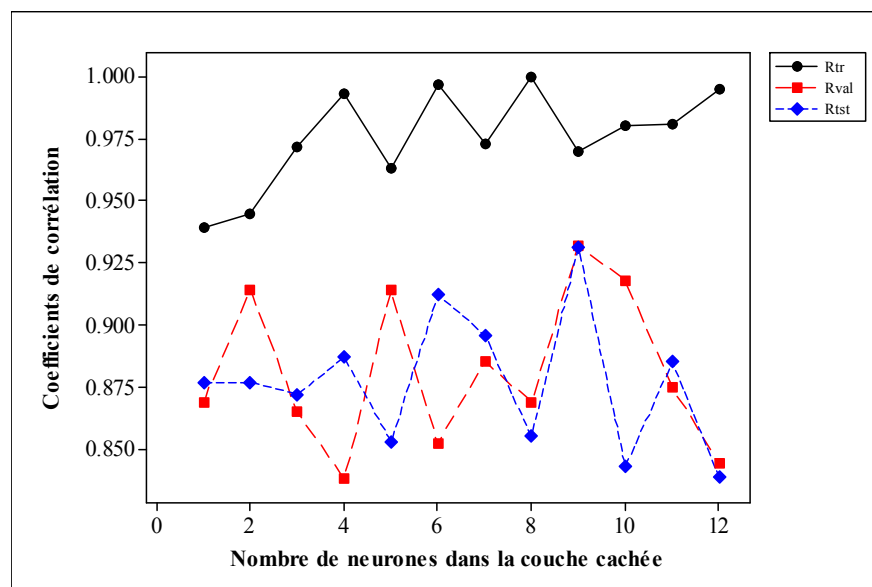


Figure 3 : Influence du nombre de neurones de la couche cachée sur les valeurs des coefficients de corrélation.

Pendant l'apprentissage de l'RNA, les poids du réseau sont ajustés de manière itérative pour minimiser la fonction de performance MSE (erreur quadratique moyenne) entre la valeur de la sortie du réseau et celle expérimentale (les résultats des calculs effectués par le réseau de neurones sont obtenus à l'aide du MATLAB version 10.0)¹⁵.

Tableau 4 : Les valeurs des coefficients de corrélation correspondant aux nombres de neurones de la couche cachée.

Nbr de Neurones	MSE	R _{tr}	R _{val}	R _{tst}
1	0,028	0,939	0,869	0,877
2	0,024	0,945	0,914	0,877
3	0,013	0,972	0,865	0,872
4	0,002	0,993	0,838	0,887
5	0,008	0,963	0,914	0,853
6	0,000	0,997	0,852	0,912
7	0,008	0,973	0,885	0,896
8	0,000	1,000	0,869	0,855
9	0,004	0,970	0,932	0,931
10	0,003	0,980	0,918	0,843
11	0,003	0,981	0,875	0,885
12	0,000	0,995	0,844	0,839

MSE : représente l'erreur quadratique moyenne

R_{tr}, R_{val}, R_{tst} : représentent les coefficients de corrélation pour les ensembles d'apprentissage, de validation et de test respectivement.

D'après le tableau 4, le modèle optimal correspond à 9 neurones dans la couche cachée (5-9-1), où les coefficients de corrélation des sous ensembles d'apprentissage, de validation et de test dépassent la valeur de 93% avec une valeur de MSE égale 0,004.

Les résidus des valeurs calculées de pIC₅₀ par l'RNA sont représentés en fonction des valeurs expérimentales de pIC₅₀ sur la figure 04. La distribution des résidus autour de la ligne zéro montre qu'aucune erreur systématique n'existe dans le modèle RNA construit.

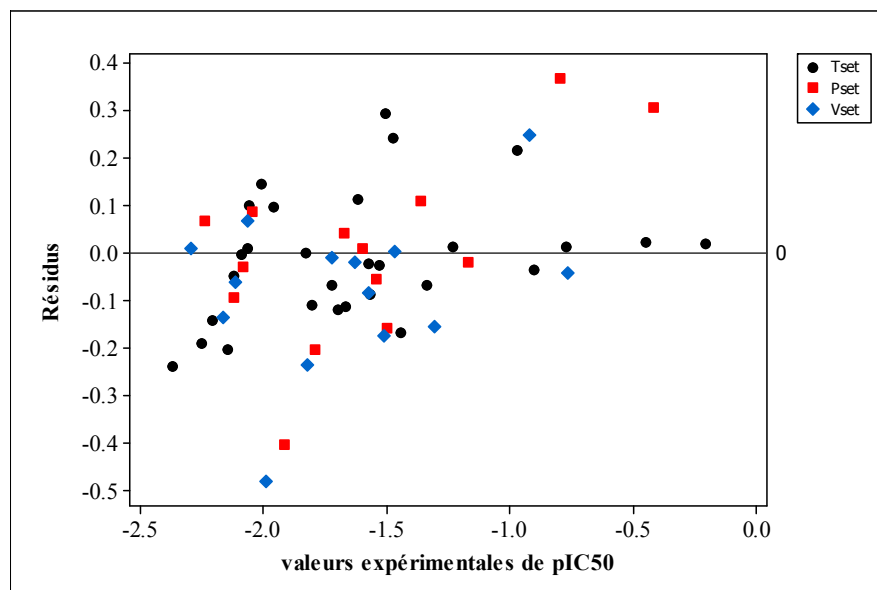


Figure04 : représentation des résidus en fonction des valeurs expérimentales de pIC50.

Pour montrer que les résultats du modèle RNA ne sont pas obtenus par hasard, un test de randomisation a été effectué. Les valeurs de pIC₅₀ des trois sous ensembles d'apprentissage, de validation et de test (Tset, Vset, et Pset) ont été changées de façon aléatoire en gardant les paramètres RNA (poids, matrice des descripteurs) inchangés. Cette opération est répétée dix fois, les résultats obtenus sont représentés sur la figure 5.

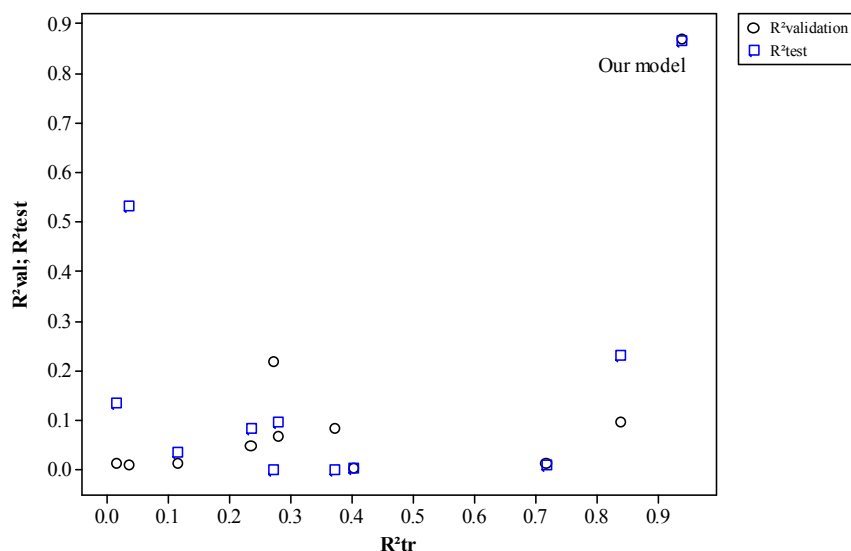


Figure 5 : Représentation des valeurs des coefficients de détermination des sous ensembles de validation et de test en fonction de celles de l'ensemble d'apprentissage.

La représentation graphique des valeurs des coefficients de détermination de la validation interne et de test en fonction du coefficient de détermination de l'apprentissage (Figure 5), montre que les valeurs des coefficients de détermination des trois sous ensembles calculés par les modèles générés sont faibles par rapport à celles de notre modèle. Ceci confirme que le modèle RNA n'est pas obtenu par hasard.

7.1 Comparaisons entre le modèle RNA et RLM :

Les paramètres statistiques obtenus par le modèle RLM et RNA sont listés dans le tableau 5.

Tableau 5 : Coefficients de détermination et erreurs carrées moyennes obtenus par les modèles RLM et RNA.

	RLM		RNA	
	DQM	R ²	DQM	R ²
Sous ensemble d'apprentissage	0,156	0,857	0,129	0,941
Sous ensemble de prédiction	0,194	0,660	0,188	0,867

Dans le but de comparer les résultats statistiques obtenus par le modèle RNA et ceux obtenus par le modèle RLM, nous avons tracé graphiquement les valeurs calculées de pIC₅₀, obtenus par les deux modèles RNA et MLR, en fonction de celles expérimentales (figure 6).

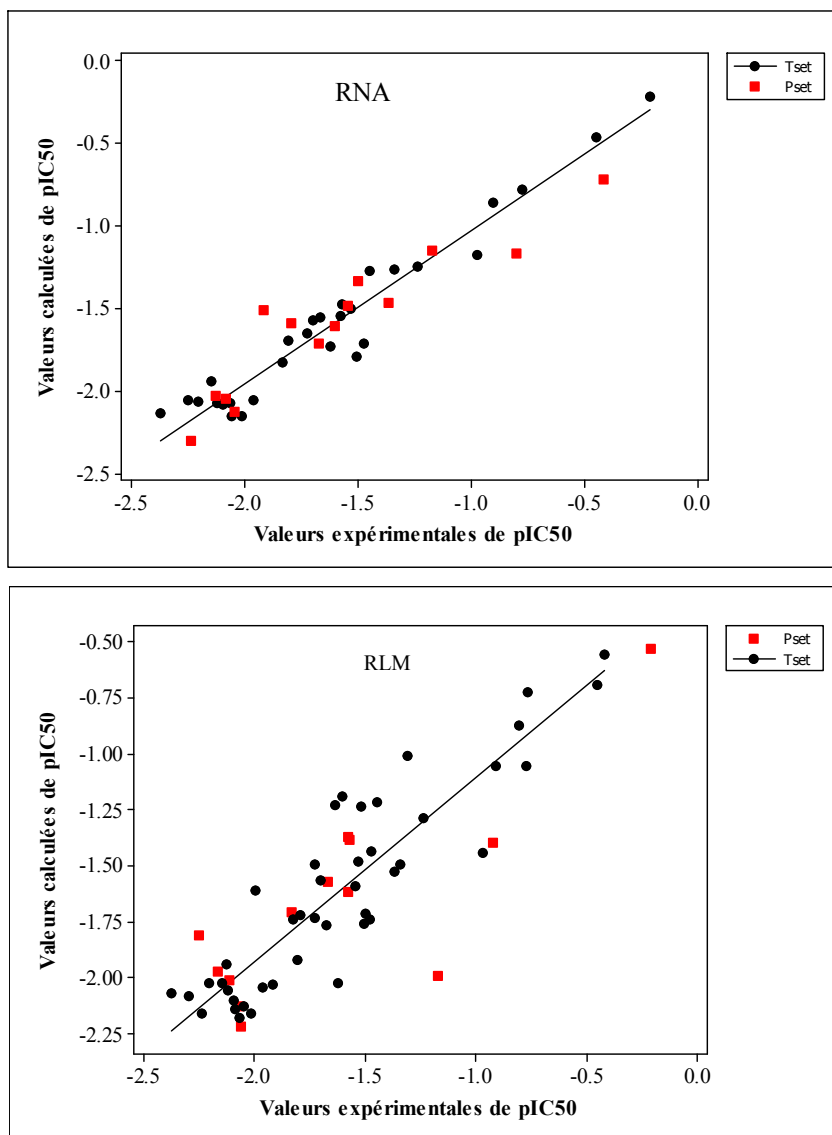


Figure 6 : représentations graphiques des valeurs calculées de pIC_{50} , obtenues par les modèles RLM et RNA, en fonction de celles expérimentales.

En se basant sur les résultats statistiques donnés par les deux modèles RNA et RLM, nous observons que le modèle RNA a une fiabilité et une capacité prédictive nettement supérieures à celles données par le modèle RLM. Les erreurs types de calcul sont plus faibles, les coefficients de détermination des ensembles d'apprentissage et de prédiction sont plus élevés avec le modèle RNA qu'avec le modèle RLM. Par conséquent, nous considérons que l'approche non linéaire est la plus convenable pour analyser la relation quantitative structure-activité inhibitrice de l' α -glucosidase par les dérivés de xanthone et de curcuminoïde.

7.2 Analyse de la contribution des descripteurs dans le modèle RNA :

Nous avons étudié l'effet des cinq descripteurs, formant le modèle RNA, sur le pouvoir inhibiteur de l' α -glucosidase par les dérivés de xanthone et de curcuminoïde.

La contribution de chaque descripteur est donnée par l'équation suivante^{16,18} :

$$C_i = \frac{100 * \Delta m_i}{\sum \Delta m_i} \quad eq11$$

C_i : la contribution du descripteur i .

Δm_i : la valeur moyenne de l'écart absolu entre les valeurs observées et les valeurs prédites de toutes les molécules.

Les résultats obtenus sont représentés dans le tableau 6 et la figure 7. D'après le tableau 6, les cinq descripteurs sont classés en fonction de leur contribution comme suit: MATS7v> Mor15u> R4e +> H5u> nArOR.

Tableau 6 : La contribution C_i de chaque descripteur.

	$\Delta pIC50$				
	MATS7v	Mor15u	H5u	R4e+	nArOR
Δm_i	0,299	0,238	0,194	0,235	0,172
C_i	26,271	20,904	17,046	20,664	15,114

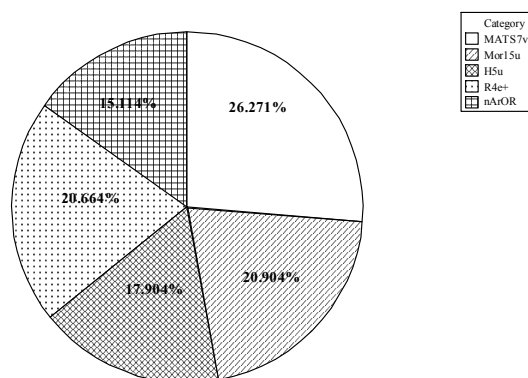


Figure 7 : Contribution des descripteurs du modèle ANN.

En conclusion, ces résultats montrent que le descripteur codé par le volume de Van der Waals exerce l'effet le plus important sur le pouvoir inhibiteur de l' α -glucosidase. Les deux descripteurs appartenant aux blocs de 3D MORS (Mor15u) et 2D autocorrelation (R4e +) sont classés en deuxième et troisième ordre respectivement.

8. Conclusion :

Au cours de ce travail nous avons appliqué la méthode des réseaux de neurones (RNA) dans le développement d'un modèle décrivant la relation quantitative structure-activité inhibitrice de l' α -glucosidase. Pour cela nous avons fait appel à la méthode des réseaux de neurones artificiels (RNA) à multicouches en tant que méthode d'apprentissage. Cinq descripteurs moléculaires sélectionnés par les algorithmes génétiques (AG) ont été utilisés dans l'apprentissage des RNA, qui sont : MATS7v, Mor15u, H5u, R4e + et nArOR. La comparaison des bons résultats obtenus par l'apprentissage des RNA avec ceux obtenus par la RLM montre que les RNA est une méthode adaptée à la modélisation de l'activité inhibitrice de l' α -glucosidase. Ils permettent non seulement d'obtenir une prédiction mais également de donner de meilleurs résultats que l'analyse par la régression linéaire multiple (RLM). Dans la RLM, l'activité inhibitrice est une combinaison linéaire des descripteurs. Par contre, dans les RNA leur pouvoir prédictif provient de leur capacité à produire des combinaisons non linéaires.

L'étude de la contribution des cinq descripteurs sélectionnés a montré que le descripteur moléculaire MATS7v, pondéré par le volume de Van Der Waals, exerce l'effet le plus important sur le pouvoir prédictif de l'activité inhibitrice de l' α -glucosidase.

9. Références :

1. H. Nocairi, M. Hanafi, E.M. Qannari. *Rev. Statistique Appliquee*, LIII(2), 2005, 29-41.
2. P.Desmet. Recherche et application en Marketing, 11(2), 1996, 17-28.
3. E. Tenorio-Borroto, C. G. Rivas, J. C. Vásquez Chagoyán, N. Castañedo, F. J. Prado-Prado, X. García-Mera, H. González-Díaz. *Bioorg. Med. Chem.* 20, 2012, 6181–6194.
4. L. Douali, A. R. Schmitzer, D. Villemin, A. Jarid, D. Cherqaoui. *Phys. Chem. News.* 34, 2007, 131-144.
5. Yan Liu, Zhuofeng Ke, Jianfang Cui, Wen-Hua Chen, Lin Ma, Bo Wang. *Bioorg. Med. Chem.* 16(15), 2008, 7185-7192.
6. Z. Du, R. Liu, W. Shao, X. Mao, L. Ma, L. Gu, Z. Huang, A. Chan. *European Journal of Medicinal Chemistry.* 41(2), 2006, 213-218.
7. www.hyper.com
8. <http://www.taletе.mi.it>
9. K. Kraim, D. Khatmi, Y. Saihi, F. Ferkous, M. Brahim. *Chemometrics Intelli. Lab. Syst.* 97(2), 2009, 118-126.
10. M. Jalali-Heravi, M. Asadollahi-Baboli, P. Shahbazikhah. *European Journal of Medicinal Chemistry*, 43, 2008, 548-556.
11. M. Arab Chamjangali, M. Beglari, G. Bagherian. *J. Mol. Graphics. Model.* 26, 2007, 360-367.
12. M. Zahouily, J. Rakik, M. Lazar, M. A. Bahlaoui, A. Rayadh, N. Komaha. *ARKIVOC.* xiv, 2007, 245-256.
13. A. Golbraikh, A. Tropsha. *J. Mol. Graph. Model.* 20(4),2002, 269-276.
14. T. A. Andrea, H. Kalayeh. *J. Med. Chem.* 34, 1991, 2824-2836.
15. <http://www.mathworks.com>
16. R. Guha, P. C. Jurs, *J. Chem. Inf. Model.* 45(3), 2005, 800-806.
17. F. Zheng, E. Bayram, S. P. Sumithran, J. T. Ayers, C. G. Zhan, J. D. Schmitt, L. P. Dwoskina, P. A. Crooks. *Bioorg. Med. Chem.* 14(9), 2006, 3017-3037.
18. D. Cherqaoui, M. Esseffer, D. Villemin, J. M. Cence, M. Chastrette, D. Zakarya. *New J. Chem.* 22, 1998, 839-8343.

Conclusion générale

Conclusion Générale

Notre travail de thèse a été consacré à la modélisation de la relation quantitative structure-activité inhibitrice de l' α -glucosidase, pour construire des modèles de QSAR fiables, robustes, stables et précis, capables de prédire efficacement cette activité.

Pour cela nous avons, dans un premier temps, mené notre étude à partir d'une bibliothèque de 57 molécules, d'origine naturelle, dérivées de xanthone et de curcuminoïde. Cette même étude a été ensuite appliquée à une bibliothèque de 37 molécules, d'origine synthétique, dérivées du N-(phénoxyalkyl) phthalimide respectivement.

Dans la première étude, nous avons fait appel aux réseaux de neurones artificiels en tant que méthode d'apprentissage non linéaire pour modéliser l'activité inhibitrice de l' α -glucosidase, exprimée par la grandeur pIC_{50} . Cinq descripteurs moléculaires ont été sélectionnés par les algorithmes génétiques afin de construire le modèle non linéaire RNA, exprimés par : MATS7v, Mor15u, H5u, R4e +, et nArOR. La comparaison des bons résultats du modèle non linéaire de RNA avec ceux obtenus par le modèle linéaire de RLM a montré que les RNA sont des méthodes capables de modéliser efficacement l'activité inhibitrice de l' α -glucosidase, et que leurs résultats sont meilleurs et plus fiables que ceux obtenus par la méthode RLM.

Dans la deuxième étude, nous avons utilisé la méthode de régression linéaire multiple et les algorithmes génétiques dans le développement du modèle QSAR en tant que méthodes d'apprentissage et de sélection respectivement.

La dimension du modèle développé est déterminée par la technique du point de virage et est constituée de trois descripteurs moléculaires pondérés par le volume de Van Der Waals, la masse et l'électronégativité respectivement. La fiabilité et le pouvoir prédictif du modèle sont testés et évalués par l'utilisation de la validation croisée, le test de randomisation et la validation externe.

Les résultats des deux études montrent que l'activité inhibitrice de l' α -glucosidase est influencée significativement par le volume de Van Der Waals et l'électronégativité. Cela peut être en relation avec la taille et la structure du site actif de l' α -glucosidase. Pour élucider en détail le mécanisme d'inhibition et connaître le type d'interaction de l' α -glucosidase avec ces molécules, il est intéressant d'utiliser, avec la méthode QSAR, d'autres méthodes de chimie-informatique tel que le docking moléculaire.

Conclusion Générale

Il serait également intéressant d'employer les deux modèles QSAR construits lors de nos études dans le criblage virtuel des banques de molécules disponibles afin de découvrir d'autres structures moléculaires capables d'être de bons inhibiteurs de l' α -glucosidase.