

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE

UNIVERSITE MOHAMED KHIDER BISKRA  
FACULTE DES SCIENCES EXACTES ET  
DES SCIENCES DE LA NATURE ET DE LA VIE

Département des Sciences de la matière

**THESE**

Présentée par

**ZERROUG Enfale**

En vue de l'obtention du diplôme de :

**Doctorat en chimie**

**Option :**

Chimie Moléculaire

*Intitulé:*

**Analyse in silico de plusieurs séries de  
pharmacomolécules hétérocycliques appliquée à la  
conception de médicaments**

Soutenue le : **24 septembre 2020**

Devant la commission d'Examen :

M.Ammar DIBI	Prof	Université Hadj Lakhder-Batna	Président
M.Salah BELAIDI	Prof	Université Mohamed Khider- Biskra	Directeur de thèse
M.Mahmoud OMARI	Prof	Université Mohamed Khider- Biskra	Examineur
M.Terrissa LABIB	Prof	Université Mohamed Khider- Biskra	Examineur
M.Samir CHTITA	Prof	Université Hassan II –Casablanca-Maroc	Invité

"وَاللَّهُ خَلَقَكُمْ ثُمَّ يَتَوَفَّاكُمْ وَمِنْكُمْ مَنْ يُرَدُّ إِلَىٰ أَرْذَلِ الْعُمُرِ لِكَيْ لَا يَعْلَمَ بَعْدَ عِلْمٍ شَيْئًا إِنَّ اللَّهَ عَلِيمٌ قَدِيرٌ"

سورة النحل الآية 70

**"Allah vous a créés! Puis Il vous fera mourir. Tel parmi vous sera reconduit jusqu'à l'âge le plus vil, de sorte qu'après avoir su, il arrive à ne plus rien savoir. Allah est, certes, Omniscient et Omnipotent "** Quran, Sourat AN-NAHL (LES ABEILLES) le verset 70

*A mes très chers parents*

*A mes frères*

*A ma sœur Ibtihal*

*A tous ceux qui me sont chers*

### REMERCIEMENTS

*Au nom d'"Allah", le plus bienfaisant, le plus miséricordieux,*

*Je voudrais exprimer mes remerciements et ma gratitude envers mon superviseur Monsieur **Salah BELAIDI** chef d'équipe de chimie informatique et pharmaceutique du laboratoire de chimie moléculaire et environnement (LMCE) Université Mohamed Kfider de Biskra, au sein duquel ce travail a été réalisé. Je vous remercie pour avoir facilité mon intégration au sein de votre équipe, Vos conseils, votre disponibilité, votre rigueur scientifique, ainsi que la confiance que vous m'avez accordée m'ont permis de mener à bien cette thèse.*

*Je voudrais adresser mes sincères remerciements à Monsieur le professeur **Ammar DIBI**, pour avoir accepté de présider le jury et d'avoir consacré son temps à la lecture et à l'évaluation de ce travail.*

*J'adresse mes respectueux remerciements au professeur **Mahmoud OMARI** directeur de notre laboratoire (LMCE) et le président du comité de formation de doctorant, qui m'a fait l'honneur d'accepter de participer au jury de cette thèse et pour son temps consacré à évaluer mon travail.*

*Je suis honorée que le Professeur **Terrissa LABIB** ait accepté de faire partie des examinateurs de ce travail. Je vous remercie m'avoir accueilli dans votre laboratoire et de m'avoir initiée aux bases de l'intelligence artificielle « ANN ». Ainsi que pour vos encouragements, vos conseils et vos commentaires constructifs qui m'ont été précieux,*

*Je tenais à exprimer ma plus profonde gratitude au professeur **Samir CHITTA**, pour sa collaboration sur toute la période de mon doctorat, durant laquelle il a toujours su se rendre disponible pour évaluer mes travaux et pour répondre à mes nombreuses questions. Je vous remercie pour vos conseils, sans oublier votre modestie qui imposent le respect et sachez que vous êtes un modèle pour tous. C'est un honneur pour moi de vous compter parmi les membres du jury de ma thèse.*

## REMERCIEMENTS

*Mes vifs et sincères remerciements vont également à mon enseignant le docteur **Nadjib MELKEMI**, pour ses encouragements, ces précieux conseils, son soutien ainsi que sa gentillesse.*

*J'adresse mes remerciements à tous les membres de notre équipe de chimie informatique et pharmaceutique du laboratoire de recherche LMCE, tout particulièrement, le docteur **Imane BENBERAHIM** pour sa disponibilité, ses conseils et son aide au début de mon parcours en doctorat.*

*Je tenais aussi à adresser mes remerciements au professeur **Nouredine ZERHOUNI** pour sa gentillesse, sa bonne humeur et sa convivialité depuis mon arrivée au laboratoire. Sa passion communicative de la science et son ouverture d'esprit, ses conseils et surtout la pression qu'il m'a mis, ainsi que les moments d'échanges et de discussions avec les membres d'équipe (Abir, Safa, et Fawzi) qui m'ont donné l'envie et l'inspiration d'aller encore plus loin dans l'application de l'intelligence artificielle dans mon domaine.*

*Je voudrais exprimer mes sincères remerciements au professeur **Mutasem TAHA**, directeur du laboratoire chimie médicinale de l'université de Jordanie, Amman, au sein duquel a été réalisée une partie de cette thèse et à l'ingénieurs **Ghada Tufaha**. Je vous remercie infiniment pour votre aide, votre disponibilité. Le travail avec vous a été une grande et une agréable expérience d'apprentissage.*

*Je voudrais saluer la patience, le soutien, la serviabilité et les encouragements dont a fait part de mon amie et ma sœur **Fatima MEBARKI** dans les moments où j'en avais le plus besoin. Puisse Dieu t'apporter le bonheur dans ta vie.*

*À ma sœur **Ibtihal**, son aide, son soutien et ses remarques ont été pour moi une source de courage et de confiance. J'implore Dieu pour qu'il lui apporte le bonheur, et l'aide à réaliser tous ses vœux.*

*À mes chers parents, pour qui tous les mots du monde ne sauraient exprimer l'immense amour que je leur porte ni la profonde gratitude que je leur témoigne. C'est à travers leurs soutiens, leurs sacrifices, leur patience infinie et à la confiance qu'ils m'ont accordé pendant toutes*

## REMERCIEMENTS

*mes années d'études en particulier ces quatre dernières années, que j'ai opté pour cette noble profession. Ce document vous est dédié, rien de ce qui s'y trouve n'y serait arrivé sans vos duaaas, vos prières, vos conseils et les mots justes que mon cher père a su trouver pour m'aider à me dépasser. J'espère avoir répondu aux espoirs que vous avez fondés en moi et vous avoir rendue fière de moi. Que Dieu tout puissant vous garde et vous procure santé, bonheur et longue vie pour que vous demeuriez le flambeau illuminant mon chemin.*

*Enfin, je tiens à préciser que cette aventure qui a été enrichissante en tout point de vue et qui m'a beaucoup apporté, aussi bien sur le plan de la recherche scientifique que sur celui du **savoir-être scientifique**, m'a appris qu'aucun de nous, en agissant seul, ne peut atteindre le succès et qu'on ne va jamais y arriver avec un état d'esprit égoïste, méprisant et individualiste. Une bonne recherche, un bon article c'est un vrai travail d'équipe. Il faut apprendre à s'encourager, à partager et à s'entraider avec bienveillance.*

REMERCIEMENTS	
SOMMAIRE	
LISTE DES FIGURES & SCHÉMAS	
LISTE DES TABLEAUX	
ABRÉVIATIONS	
INTRODUCTION GÉNÉRALE.....	2

## CHAPITER I : Maladie d'Alzheimer & Cibles thérapeutiques

PARTIE 1 : MALADIE D'ALZHEIMER .....	7
1.1. INTRODUCTION.....	7
1.2. NEUROPATHOLOGIE DE LA MALADIE D'ALZHEIMER .....	8
1.2.1. Physiopathologie.....	9
1.2.2. Les symptômes .....	9
1.2.3. Étiologie .....	10
1.2.4. Facteurs de risque et de protection.....	13
1.3. APPROCHES THERAPEUTIQUES.....	14
PARTIE 2: CIBLES THÉRAPEUTIQUES.....	17
2.1. ACETYLCHOLINESTERASE.....	17
2.1.1. Neurotransmission .....	17
2.1.2. Structure tridimensionnelle d'acétylcholinestérase .....	18
2.1.3. Implication de l'AChE dans la maladie d'Alzheimer .....	20
2.2. KINASES DEPENDANTES DES CYCLINES 5 .....	21
2.2.1. Protéines kinases.....	21
2.2.2. Kinases dépendantes des cyclines.....	22
2.2.3. Structure de Kinases dépendantes des cyclines 5.....	22
2.2.4. Rôles de Cdk5/P25 dans la maladie d'Alzheimer .....	24
REFERENCES BIBLIOGRAPHIQUE .....	26

## CHAPITER II : Conception et découverte de médicaments assistés par ordinateur

1. INTRODUCTION.....	34
1.1. REVOLUTION CHRONOLOGIQUE DE MEDICAMENTS.....	34
1.2. CONCEPTION DE MEDICAMENTS ASSISTEE PAR ORDINATEUR.....	34
2. CRIBLAGE VIRTUEL «IN SILICO » .....	36
2.1. CRIBLAGE VIRTUEL BASE SUR LE LIGAND « LIGAND-BASED » .....	37
2.1.1. Recherche de similarité .....	38
2.1.2. Modèles pharmacophoriques « Ligand-Based ».....	38
2.1.3. Relation structure-activité quantitative « ligand-based ».....	40
2.2. CRIBLAGE VIRTUEL BASE SUR LA STRUCTURE « STRUCTURE-BASED » .....	43
2.2.1. Modèles pharmacophoriques « structure-based » .....	44
2.2.2. Conception de novo « de novo design » .....	46
2.2.3. Relation quantitative structure-activité dépendant du récepteur (RD-QSAR).....	47
2.3. ÉVALUATION DES METHODES DE CRIBLAGE VIRTUEL.....	55
REFERENCES BIBLIOGRAPHIQUE.....	59

## CHAPITER III : Travaux réalisés

Partie 1 : Étude de drug-likeness et de la relation structure-activité/propriété qualitative d'inhibiteurs d'AChE des dérivés de la pyridazine.....	69
-----------------------------------------------------------------------------------------------------------------------------------------------------	----

<b>1. MATERIELS ET METHODES</b> .....	70
1.1. METHODES DE CALCUL UTILISEES .....	70
1.2. METHODES D'OPTIMISATION MULTIPARAMETRES DU SYSTEME NERVEUX CENTRAL (MPO DU SNC) .....	71
1.3. PROPRIETES DRUG- LIKENESS ET LES INDICES DE LA LIPOPHILIE .....	72
<b>2. RESULTATS ET DISCUSSION</b> .....	74
2.1. SCORE DE DESIRABILITE D'OPTIMISATION MULTIPARAMETRES DU SNC .....	74
2.2. ÉTUDE SAR/SPR D'UNE SERIE DE DERIVE DE LA PYRIDAZINE .....	75
<b>3. CONCLUSION</b> .....	80

## Partie 2 : Étude QSAR basée sur un réseau de neurones artificiels pour le criblage Virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase .....

<b>1. MATERIELS ET METHODE</b> .....	83
1.1. LOGICIELS .....	83
1.2. ENSEMBLE DE DONNÉES .....	83
1.3. METHODES STATISTIQUES .....	87
1.3.1. Classification des données .....	87
1.3.2. Sélection des descripteurs .....	87
1.3.3. Les outils statistiques utilisés pour le développement du modèle QSAR .....	88
1.3.4. Estimation de la capacité prédictive du modèle QSAR .....	89
<b>2. Résultats et discussion</b> .....	92
2.1. METHODE DU PARTITIONNEMENT K-MOYENNES « K-MEANS » .....	92
2.2. ANALYSE EN COMPOSANTES PRINCIPALES (ACP) .....	92
2.3. REGRESSION LINEAIRE MULTIPLE (MLR) .....	94
2.4. CONCEPTION ET FORMATION DU MODELE DE RESEAUX DE NEURONES ARTIFICIELS (ANN) .....	96
2.4.1. Évaluation de la capacité prédictive du réseau de neurones .....	98
2.4.2. Domaine d'applicabilité .....	99
2.4.3. CONTRIBUTION DES DESCRIPTEURS SELECTIONNES .....	100
<b>3. CONCLUSION</b> .....	105

## Partie 3 : Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants de CDK5 .....

<b>1. MATERIAUX ET METHODES</b> .....	108
1.1. LOGICIELS .....	108
1.2. ENSEMBLE DE DONNEES .....	109
1.3. PREPARATION DE LA STRUCTURE CRISTALLINE CDK5 .....	113
1.4. PARAMETRES DE DOCKING .....	117
1.5. ANALYSE COMPARATIVE DES CONTACTS INTERMOLECULAIRES BASEE SUR LE DOCKING (DB-CICA) .....	114
1.6. GENERER LES PHARMACOPHORES CORRESPONDANT AUX MODELES REUSSIS DE DB-CICA .....	116
1.7. CRIBLAGE IN SILICO DE NOUVEAUX INHIBITEURS DE CDK5 .....	117
<b>2. RESULTATS ET DISCUSSION</b> .....	118
2.1. MODELISATION DB-CICA DES INHIBITEURS DE CDK5 .....	118
2.2. LES MODELES PHARMACOPHORES OPTIMAUX BASES SUR DBCICA .....	119
2.3. VALIDATION DE L'HYPOTHESE PHARMACOPHORE PAR LA COURBE ROC .....	120
2.4. Raffinement du modèle pharmacophor « volumes d'exclusion » .....	122
2.5. CRIBLAGE IN SILICO DE LA BASE DE DONNEES NCI POUR SELECTIONNER DES NOUVEAUX INHIBITEURS DE CDK5 PUISSANTS .....	123
<b>3. CONCLUSION</b> .....	124
<b>REFERENCES BIBIOGRAPHIQUE</b> .....	125
<b>CONCLUSION GÉNÉRALE</b> .....	134
<b>RÉSUMÉ</b>	



### Chapitre I

- Fig.1.** Croquis dessiné par Alzheimer représentant dégénérescences neurofibrillaires du cerveau d'Auguste D.
- Fig.2.** Hémisphère d'un cerveau sain (à G) et atteint d'Alzheimer (à D).
- Fig.3.** Les lésions associées à la maladie Alzheimer au plan anatomique.
- Fig.4.** Les voies impliquées dans le processus de métabolisation de la protéine précurseur de l'amyloïde (APP).
- Fig.5.** La structure de la mémantine (Ebixa®).
- Fig.6.** La structure des inhibiteurs de l'acétylcholinestérase.
- Fig.7.** Structure de monomère de la TcAChE (1EVE:PDB) : structure 3D avec représentation des hélices  $\alpha$ , feuillets  $\beta$  et résidus du site actif. La coloration en arc-en-ciel va du domaine C-terminal en rouge vers domaine N-terminal en bleu.
- Fig.8.** Schématisation des mécanismes de phosphorylation/déphosphorylation.
- Fig.9.** Structure de la kinase CDK5 lié à p25.
- Fig.10.** Dégénérescence neurofibrillaire induite par l'hyperphosphorylation de protéine Tau.
- Schéma 1.** Rôle de l'AChE au niveau des synapses.
- Schéma 2.** Vue schématique du site actif de TcAChE.

### Chapitre II

- Fig.1.** Pharmacophore généré à l'aide du logiciel discovery studio.
- Fig.2.** Représentation schématique des étapes principales d'une conception de novo.
- Fig.3.** La méthode de docking.
- Fig.4.** Les courbes de ROC représentent l'évolution, pour chaque fraction de la chimiothèque.
- Schéma 1.** Les différentes étapes de développement du médicament.
- Schéma 2.** Classification des méthodes de criblage virtuel (avec QSAR : relation quantitative structure-activité, RD-QSAR : relation quantitative structure-activité dépendant du récepteur).
- Schéma 3.** Les différentes étapes d'une étude QSAR.
- Schéma 4.** Identification d'un site de liaison sur une cible thérapeutique.
- Schéma 5.** Étapes de la méthodologie de conception de médicaments de novo.
- Schéma 6.** Représentation de la sélection de 'n' molécules dans une chimiothèque 'N'.

### Chapitre III

- Fig.1.** Structures 3D des dérivés de la pyridazine.
- Fig.2.** Les graphes représentent les six fonctions des propriétés physico-chimiques utilisées pour générer le score MPO du SNC.
- Fig.3.** Ensemble de molécules avec leurs valeurs de log D et MW. Les molécules avec une bonne perméabilité (cercles bleus) sont situées dans le Triangle d'or.
- Fig.4.** Sites donneurs et accepteurs du composé 14.
- Fig.5.** a) Projection des composés dans les deux axes, F1 et F2; les points bleus et rouges sont respectivement les ensembles de formation et de test. b) Cercle de corrélation de la variable dans les deux axes, F1 et F2.
- Fig.6.** L'architecture du réseau d'ANN sur Matlab.
- Fig.7.** a) La courbe de la régression pour le modèle ANN. b) Courbes de performances durant l'apprentissage, validation et le test du modèle ANN.
- Fig.8.** Comparaison des capacités prédictives des modèles MLR et ANN.
- Fig.9.** Diagramme de Williams du résidu normalisé par rapport au levier pour le modèle ANN.
- Fig.10.** Contribution des descripteurs dans le modèle ANN.
- Fig.11.** Représentation schématique du résumé de travail réaliser.
- Fig. 12.** Présentation du schématique de préparation des ligands pour les étapes de docking.
- Fig.13.** Représentation schématique de différentes étapes de la méthode db-CICA.
- Fig.14.** Les étapes pour la génération manuelle de l'hypothèse de liaison Hypo-I guidée par le modèle I de db-CICA.
- Fig.15.** Les courbes ROC du pharmacophore à base de db-CIC: **A**: Hypo-I, **B** : Hypo-I raffiné.
- Fig.16.** Les points pharmacophoriques d'Hypo-I. Le modèle Hypo-I raffiné avec 58 volumes d'exclusion.
- Schéma 1.** Représentation schématique du flux du travail suivi dans cette étude.
- Schéma 2.** Architecture du réseau ANN et le principe de l'algorithme de rétropropagation.

## LISTE DES TABLEAUX

- Tableau 1 :** Scores de MPO du SNC et les scores individuels transformés (T0) des composés anti-AChE.
- Tableau 2 :** Paramètres physico-chimiques des dérivés de la pyridazine.
- Tableau 3 :** Paramètres d'aptitude au médicament et les indices de lipophilie des dérivés de la pyridazine.
- Tableau 4 :** Le résumé des résultats.
- Tableau 5 :** Structures chimiques et l'activité anti-AChE des composés étudiés.
- Tableau 6 :** Descripteurs calculés et leur type.
- Tableau 7 :** Les résultats de partitionnement k-moyennes.
- Tableau 8 :** Corrélation (Pearson (r)) entre l'activité et les variables indépendantes.
- Tableau 9 :** Multi-colinéarité du modèle MLR.
- Tableau 10 :** Paramètres statistiques obtenus par le modèle MLR et leurs valeurs de seuil.
- Tableau 11 :** Résultats statistiques du modèle ANN-QSAR.
- Tableau 12 :** Activité prédictive ( $pIC_{50}$ ) et l'effet de levier (hi) de 17 *hits* identifiés.
- Tableau 13 :** Composés impliqués dans la modélisation basée sur la structure CDK5.
- Tableau 14 :** Modèles db-CICA les mieux classés, leurs paramètres correspondants et leurs critères statistiques.
- Tableau 15 :** Atomes de contact du site de liaison critique proposés par le modèle dbCICA
- Tableau 16 :** Performances du pharmacophore sélectionné par le modèle dbCICA-I.
- Tableau 17 :** Les *hits* les plus actifs captés par le pharmacophore raffiné Hypo-I et leur activité CDK5 prédictive in silico.

## ABRÉVIATIONS

<b>AChEI:</b>	Acetyl-Cholinesterase-Inhibitor.
<b>ACP :</b>	Principal Component Analysis
<b>AD :</b>	Applicability Domain
<b>ALA:</b>	Alanine.
<b>ANN :</b>	Artificial Neural Networks.
<b>APP :</b>	Amyloid Precursor Protein.
<b>ASN:</b>	Asparagine
<b>ASP :</b>	Aspartic acid
<b>ATP :</b>	Adénosine Triphosphate.
<b>A<math>\beta</math> :</b>	Amyloid $\beta$ -protein.
<b>B3LYP :</b>	Becke 3-Parameter Lee-Yang-Parr.
<b>CDK5 :</b>	Cyclin-dependent Kinase 5
<b>CHE</b>	Centre Hospitalo-Universitaire
<b>ChEMBL :</b>	Chemical European Molecular Biology Laboratory
<b>dbCICA :</b>	Docking-Based Comparative Intermolecular Contacts Analysis
<b>DFT :</b>	Density Functional Theory.
<b>DNF :</b>	Dégénérescence Neuro-Fibrillaires.
<b>EF :</b>	Enrichment Factor.
<b>GA:</b>	Genetic Algorithm
<b>GLN:</b>	Glutamine
<b>GLU:</b>	Glutamate.
<b>HBA:</b>	Hydrogene Bond Acceptor.
<b>HBD:</b>	Hydrogène Bond Donor.
<b>HIS:</b>	Histidine.
<b>HOH:</b>	Water
<b>HTS :</b>	High Throughput Screening.
<b>Hypo:</b>	Hypothèse.
<b>IC<sub>50</sub> :</b>	The half maximal inhibitory concentration
<b>ILE:</b>	Isoleucine
<b>LEU:</b>	Leucine
<b>LOO:</b>	Leave-One-Out.
<b>LYS:</b>	Lysine
<b>MA :</b>	Maladie d'Alzheimer.
<b>MLR :</b>	Multiple Linear Regression.
<b>MM :</b>	Mécanique Moléculaire.

## | ABRÉVIATIONS

<b>NCI :</b>	National Cancer Institute.
<b>nM :</b>	Nanomolar.
<b>OCDE :</b>	Organisation de Coopération et de Développement Economique
<b>PAS :</b>	Périphérique Anionique Site.
<b>PDB :</b>	Protein Data Bank.
<b>PHE :</b>	Phenylalanine
<b>PHF :</b>	Paired helical filament.
<b>RD-QSAR :</b>	Receptor Dependent -Quantitative Structure/Activity Relationships.
<b>RMSD</b>	Root-mean-square deviation
<b>ROC:</b>	Receiver Operating Characteristic
<b>SAR :</b>	Structure/Activity Relationships.
<b>Se:</b>	Sensibilité.
<b>SER:</b>	Sérine.
<b>Sp :</b>	Spécificité.
<b>SPR :</b>	Structure/Property Relationships.
<b>THEL CV :</b>	Transformed human liver epithelial cell viability
<b>TRP :</b>	Tryptophan
<b>TYR:</b>	Tyrosine
<b>VAL:</b>	Valine

# Introduction Générale

## Introduction Générale

Modélisation moléculaire, docking moléculaire, criblage virtuel, QSAR, pharmacophore et *de novo design*; tous ces termes appartiennent à des disciplines de recherche de la chimio-informatique et la bioinformatique structurale. Ces dernières sont bien établies avec de nombreuses applications dans le domaine des sciences du vivant et de la santé et qui sont apparues au cours de ces dernières années avec le développement important et accéléré des techniques computationnelles utilisées dans le domaine de la conception de médicaments assistés par ordinateur (*Computer-Aided Drug Design*)<sup>1-3</sup>. Les technologies de la CADD ont amélioré l'efficacité de la découverte de médicaments en faisant gagner du temps et de l'argent principalement dans les premières phases de la recherche<sup>4-7</sup>. Pour ces raisons, la CADD est devenue actuellement l'axe de recherche privilégié de développement thérapeutique qui intéresse de plus en plus les chercheurs spécialisés en découverte de médicaments.

La CADD met en œuvre une puissance de calcul dans l'espace chimique et biologique combiné, afin de rationaliser la conception, le développement et l'optimisation de médicaments<sup>8</sup>. Elle est un ensemble de protocoles *in silico* et des algorithmes qui aident à la conception de petites molécules chimiques « *drug-like* » les plus susceptibles de se lier à une cible médicamenteuse<sup>8-10</sup>.

Dans notre étude, nous nous basons sur un ensemble de méthodes de la CADD, pour la prédiction et le développement de nouveaux inhibiteurs puissants de l'acétylcholinestérase (AChEI) et des kinases dépendantes des cyclines 5 (CDK5). Ces deux cibles thérapeutiques sont parmi les cibles désignées dans le traitement de la maladie d'Alzheimer.

Avec le vieillissement de la population, la maladie d'Alzheimer est devenue un problème majeur de santé publique. La maladie d'Alzheimer est le type de démence le plus répandu, représentant au moins les deux tiers des cas de démence chez les personnes âgées de plus de 65 ans<sup>11</sup>. En 2019, ADI (Alzheimer's Disease International) estime à plus de 50 millions le nombre de personnes atteintes de démence dans le monde, ce chiffre devrait passer à 152 millions d'ici 2050<sup>11</sup>. La maladie d'Alzheimer est une maladie neurodégénérative progressive causée par la mort des cellules neuronales. La déficience de la mémoire est suivie de problèmes de la résolution, du jugement, du manque de motivation, de la désorganisation et du trouble du langage<sup>12-13</sup>.

## Introduction Générale

De nos jours, les médicaments disponibles sur le marché ne constituent que des traitements palliatifs qui soulagent temporairement les patients. En réponse à ce constat, un nombre croissant de recherches tentent de mieux comprendre les mécanismes biologiques liés à la maladie d'Alzheimer dans l'espoir de développer de nouveaux composés thérapeutiques capables de stopper, voire d'inverser, le déclin cognitif et la perte de mémoire. Parmi les voies de recherche, l'une d'elles s'attaque à la dérégulation de protéine kinase dépendante du cycline 5 (Cdk5) qui est considérée comme l'un des événements majeurs dans le processus neurodégénératif de la maladie d'Alzheimer<sup>14</sup>. Une autre se repose sur l'hypothèse cholinergique, considérée comme l'une des lésions principales engendrées lors de la maladie d'Alzheimer. Le déficit en acétylcholine peut atteindre jusqu'à 90% dans les stades sévères de la maladie, les médecins font habituellement appel aux inhibiteurs de l'acétylcholinestérase<sup>15</sup>.

Dans le cadre de la recherche de ma thèse, nous nous sommes particulièrement intéressés aux méthodes de criblage virtuel. Dans un premier temps, nous nous sommes focalisés sur les méthodes à base de ligands qui ont une activité anti-AChE, en faisant intervenir la méthode QSAR (Relation Quantitative Structure-Activité) en appliquant un réseau de neurones artificiels (ANN). L'objectif de ce travail est d'explorer ces stratégies d'analyse et de des protocoles de criblage virtuel, afin de concevoir et de développer de nouvelles entités chimiques ayant des propriétés inhibitrices de l'AChE.

Dans le deuxième axe de notre recherche, nous nous sommes focalisés sur les méthodes basées sur la structure de la protéine CDK5. L'enrichissement toujours croissant en structures tridimensionnelles dont la qualité en terme de résolution est en nette amélioration, nous a guidés vers une méthode de 3D-QSAR. Notamment, l'analyse comparative des contacts intermoléculaires basée sur le docking (db-CICA), pour extraire des modèles pharmacophoriques optimaux et hautement liés à l'activité inhibitrice de CDK5. Les pharmacophores ont ensuite été utilisés comme requêtes de recherche pour extraire de nouveaux inhibiteurs de l'enzyme CDK5 depuis la base de données de l'Institut National du Cancer (*National Cancer Institute* ou NCI) américain.

Dans ce contexte, cette thèse est présentée sous forme de trois chapitres brièvement décrit comme suit:



### Chapitre I : Maladie d'Alzheimer & Cibles thérapeutiques

Ce chapitre est exposé en deux parties: La première représente une étude bibliographique de la maladie d'Alzheimer. Et dans la deuxième, nous allons nous attarder sur des généralités concernant les deux cibles thérapeutiques AChE et CDK5 puisqu'elles sont considérées comme des cibles impliquées dans le traitement de la maladie d'Alzheimer.

### Chapitre II: Conception et découverte de médicaments assistées par ordinateur (CADD)

Dans le second chapitre, nous avons présenté de façon brève l'évolution du domaine de la conception de médicaments. Par la suite, nous nous sommes focalisés sur des méthodes de criblage virtuel et leurs techniques qui sont devenues indispensables pour la conception de nouvelles substances médicamenteuses.

### Chapitre III: Travaux réalisés

Les travaux de recherche effectués au cours de cette thèse nous ont permis d'aboutir à des résultats qui seront ensuite exposés et discutés en trois parties dans ce chapitre. Tout au long de ces travaux.

- **Partie 1:** Étude de drug-likeness et de la relation structure-activité/propriété qualitative d'inhibiteurs d'AChE des dérivés de la pyridazine
- **Partie 2 :** Étude QSAR basés sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase.
- **Partie 3 :** Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants de CDK5

Nous concluons par un bilan général présentant les résultats obtenus, indiquant l'importance du domaine de la chimio-informatique et de la bio-informatique dans la conception des médicaments.

### Références bibliographiques :

1. Zhang, W. (2016). Computer-Aided Drug Discovery. *Methods in Pharmacology and Toxicology*. (1Ed.), Humana Press.305.
2. Yu, W., & MacKerell, A. D. (2017). Computer-aided drug design methods. *Methods .Mol .Biol*. 1520, 85–106.
3. Hoque, I., Chatterjee, A., Bhattacharya, S., & Biswas, R. (2017). An approach of Computer-Aided Drug Design (CADD) tools for in silico pharmaceutical drug design and development. *Int. J. Adv. Res. Biol. Sci*, 4(2), 60-71
4. Cohen, N. C. (1996). *Guidebook on molecular modeling in drug design*. (Ed). Elsevier. 361.
5. Siju, E. N., Rajalakshmi, G. R., Paulose, A. P., Fathima, T., Dhanya, P. P., Hariraj, N., & Rahul, K. (2017). CADD: PHARMACOLOGICAL APPROACHES IN DRUG DESIGN AND DRUG DISCOVERY.892–908.
6. Leelananda, S. P., & Lindert, S. (2016). Computational methods in drug discovery. *Beilstein journal of organic chemistry*, 12(1), 2694-2718.
7. Gao, Q. Yang, L. and Zhu, Y. (2010), Pharmacophore based drug design approach as a practical process in drug discovery. *Current Computer-Aided Drug Design*, 6 (1), 37–49.
8. Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach. *Chemico-biological interactions*, 171(2), 165-176.
9. Abuhammad, A., & Taha, M. (2016). Innovative computer-aided methods for the discovery of new kinase ligands. *Future medicinal chemistry*, 8(5), 509-526.
10. Singh, J., Chuaqui, C. E., Boriack-Sjodin, P. A., Lee, W. C., Pontz, T., Corbley, M. J., ... & Papadatos, J. L. (2003). Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGF $\beta$  receptor kinase (T $\beta$ RI). *Bioorganic & medicinal chemistry letters*, 13(24), 4355-4359.
11. Alzheimer's Disease International. 2019. *Rapport Mondial Alzheimer 2019 : Attitudes à l'égard de la démence*. Londres : Alzheimer's Disease International.
12. Kumar, A., & Tsao, J. W. (2018). Alzheimer Disease. In *StatPearls [Internet]*. StatPearls Publishing.
13. Tang, Y., Lutz, M. W., & Xing, Y. (2019). A systems-based model of Alzheimer's disease. *Alzheimer's & Dementia*, 15(1), 168-171.
14. Dhavan, R., & Tsai, L. H. (2001). A decade of CDK5. *Nature reviews Molecular cell biology*, 2(10), 749.
15. Zhou, Y., Wang, S., & Zhang, Y. (2010). Catalytic reaction mechanism of acetylcholinesterase determined by Born– Oppenheimer ab initio QM/MM molecular dynamics simulations. *The journal of physical chemistry B*, 114(26), 8817-8825

# **CHAPITER I**

## **Maladie d'Alzheimer & Cibles thérapeutiques**

## Partie 1 : Maladie d'Alzheimer

### 1.1. Introduction

#### 1.1.1. Auguste Dieter, premier cas diagnostiqué de la Maladie d'Alzheimer

L'histoire a commencé en 1901, quand Auguste atteinte d'un état de sénile jamais répertorié auparavant. Elle présentait une dégradation progressive de ses capacités cognitives incluant des troubles mnésiques et comportementaux. Ce diagnostic fut établi par un neuropsychiatre allemand, le Dr Alois Alzheimer. Par la suite, cette maladie portera son nom<sup>1</sup>.

Le 8 avril 1906, Auguste décède. Le Dr Alzheimer réalisa une autopsie de son cerveau et décrivit les deux principales lésions neuropathologiques de la maladie d'Alzheimer: l'apparition des plaques séniles et les dégénérescences neurofibrillaires (Figure 1), il conclut ce diagnostic par une description "maladie particulière du cortex cérébral"<sup>2</sup>. Ce n'est que dans les années 1970-1980 que la recherche sur la maladie d'Alzheimer est apparue, grâce aux progrès de la médecine et en raison de l'augmentation des cas en lien avec le vieillissement de la population. Ils ont alors identifié le peptide  $\beta$ -amyloïde ( $A\beta$ ) et la protéine *Tau* hyperphosphorylée qui constituent respectivement les plaques séniles et les dégénérescences neurofibrillaires<sup>3</sup>. Ils découvrent ainsi qu'un déficit en acétylcholine pouvait accélérer le déclin cognitif<sup>4</sup>. Ces découvertes sont toujours déterminantes actuellement dans la recherche d'un traitement pharmacologique contre la maladie d'Alzheimer. Enfin, dans les années 1990, les progrès en génétique ont permis d'identifier plusieurs gènes liés à l'apparition de la maladie.



**Fig.1.** Croquis dessiné par Alzheimer représentant la dégénérescence neurofibrillaire du cerveau d'Auguste D<sup>2</sup>.

### 1.1.2. Épidémiologie

La maladie d'Alzheimer (MA), forme de démence la plus répandue, représente 60 à 70% des cas de démence. 62% des professionnels de la santé du monde entier pensent toujours que la démence fait partie du vieillissement normal, sachant que 2 personnes sur 3 pensent qu'une démence est causée par le vieillissement normal<sup>5</sup>. La MA est désormais en passe de devenir un véritable fléau qui ronge de plus en plus la santé des personnes âgées aux quatre coins du monde<sup>6</sup>.

Actuellement, selon le rapport d'Alzheimer mondiale 2019, toutes les trois secondes, quelqu'un dans le monde développe une démence et près de 50 millions de personnes sont atteintes la MA dans le monde et ce chiffre devrait passer à 152 millions d'ici 2050, si des stratégies de réduction des risques ne sont pas mises en place à travers le monde<sup>5</sup>. En conséquence, le coût social et économique de la maladie va également s'accroître, à moins que des mesures curatives ou de prévention ne soient rapidement établies. Le coût total estimé de la démence dans le monde en 2019 est de 1 billion de dollars américains, ce chiffre atteindra 2 000 milliards de dollars d'ici 2030, ce coût est dépassé la valeur marchande des plus grandes entreprises du monde, y compris Apple et Microsoft<sup>5,7</sup>. La plus grande augmentation des cas se fera dans les pays faiblement ou moyennement industrialisés. Cela est dû en partie à une augmentation de l'espérance de vie, ainsi qu'à un manque d'accessibilité du diagnostic et de politique de prévention. Les taux d'incidence de la maladie d'Alzheimer sont légèrement plus élevés chez les femmes<sup>8</sup>. Dans les pays occidentaux où la population est vieillissante, la MA est la première cause de démence et la quatrième cause de décès, précédée par les maladies cardiaques, le cancer et les accidents vasculaires cérébraux.

En Algérie, près de 200.000 cas d'Alzheimer ont été recensés à l'échelle nationale entre 2017 et 2018, a révélé par le Pr Souhila Amalou, neurologue spécialisée dans la maladie d'Alzheimer au Centre hospitalo-universitaire (CHU) Franz Fanon de Blida<sup>9</sup>. Elle a recensé une moyenne de 25 nouveaux cas d'Alzheimer/mois, dans la seule wilaya de Blida. La statistique n'est évidemment pas exhaustive tant que la pathologie reste sous-diagnostiquée, pour diverses raisons. Pourtant, huit centres de mémoires sont opérationnels en Algérie.

## 1.2. Neuropathologie de la maladie d'Alzheimer

La maladie d'Alzheimer est un trouble neurodégénératif progressif du cerveau, qui représente une détérioration au fil du temps de différentes fonctions cérébrales et un dysfonctionnement des connexions entre les neurones qui provoquent une déficience

## Maladie d'Alzheimer & Cibles thérapeutiques

progressive et invalidante des fonctions cognitives et perte de la mémoire, ce qui conduit à la démence<sup>10,11</sup>. La démence est un nom collectif pour les syndromes d'atteinte progressive du cerveau qui causent un déclin de la capacité cognitive suffisamment grave pour gêner les activités de la vie quotidienne.



**Fig.2.** Hémisphère d'un cerveau sain (à G) et atteint d'Alzheimer (à D). Cliché du CHU de l'université de Chêne-Bourg, près de Genève, Suisse. REUTERS/Denis Balibouse.

### 1.2.1. Physiopathologie

Au cours du vieillissement chez des sujets sains l'atrophie corticale pourra être observée, avec une perte d'environ de 2% de son poids tous les 10 ans. Alors que chez les malades d'Alzheimer, cette atrophie est observée, avec une perte d'environ de 8% de son poids et sera surtout localisée au niveau de l'hippocampe (une structure cérébrale essentielle pour la mémoire) et le cortex entorhinal (impliquée dans les mécanismes de l'olfaction et de la mémoire) (Figure3)<sup>12</sup>. Puis ces lésions se propageront aux autres aires du cerveau comme celle du langage, de la réalisation des gestes et à reconnaissance des visages et d'autres fonctions vitales pour le patient. Les différentes lésions seront distinguées grâce à la technique de l'IRM ou du scanner dans l'étape de diagnostic.

### 1.2.2. Les symptômes

#### ➤ *D'un point de vue physiologique :*

La MA se caractérise par la présence entre les neurones de dépôts de plaque amyloïde extracellulaire et par la présence d'enchevêtrements neurofibrillaires intracellulaires dans le neurone, entraînent progressivement le dysfonctionnement des neurones et leur mort fonctionnelle<sup>16</sup>.

Plusieurs signes distinctifs de la maladie ont été rapportés tels que le stress oxydatif et faibles niveaux d'acétylcholine<sup>17, 18</sup>.

## Maladie d'Alzheimer & Cibles thérapeutiques

### ➤ D'un point de vue clinique :

Les médecins divisent les symptômes en symptômes « cognitifs » affectent la mémoire, le langage, le jugement et les processus de pensée et les symptômes « comportementaux » modifient les actions et les émotions du patient. Au début de la pathologie est caractérisée par des difficultés d'acquisition de nouvelles compétences et de recherche de mots en raison de troubles de la mémoire <sup>13,14</sup>. Plus la maladie progresse, plus les changements comportementaux deviennent manifestes, les fonctions cognitives de l'individu (langage, raisonnement, apprentissage, résolution de problèmes, prise de décision, manque de motivation et de la désorganisation, les émotions, l'humeur...). Ces troubles plongent progressivement le malade dans la dépression jusqu'à le stade sévère de la maladie. Il devient alors totalement dépendant des autres pour vivre et éprouve de grandes difficultés à communiquer. Il perd graduellement la mobilité de ses membres et reste généralement alité jusqu'à son décès <sup>15</sup>

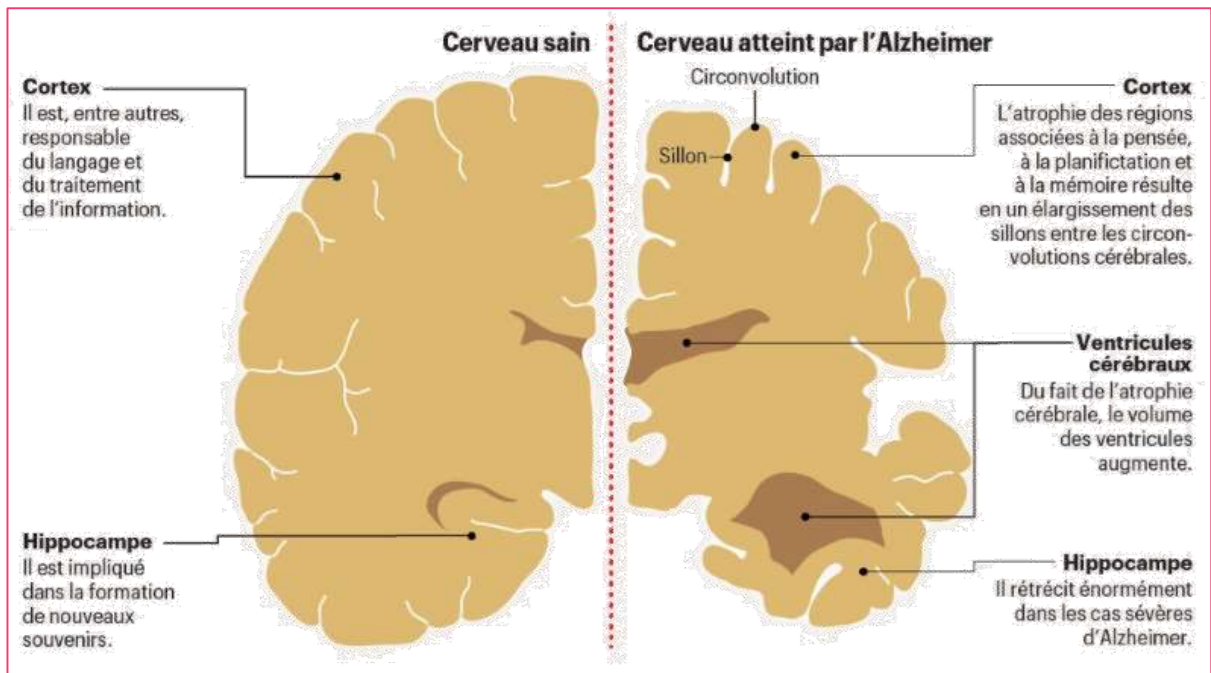


Fig.3. Les lésions associées à la maladie Alzheimer au plant anatomique.

### 1.2.3. Étiologie

L'étiologie de la maladie d'Alzheimer n'est pas complètement connue, cependant plusieurs hypothèses étiologiques ont été proposées et font la lumière sur une partie des mécanismes de la maladie, telles que l'hypothèse cholinergique, l'hypothèse de l'amyloïde et l'hypothèse de la protéine *Tau*. Les neurones sont affectés par deux lésions principales: le dépôt extracellulaire de plaque  $\beta$ -amyloïde ( $A\beta$ ) et la dégénérescence neurofibrillaire <sup>3,18,19</sup>. Ces deux lésions conduisant à une profonde perte de neurones et de récepteurs cholinergiques qui

provoque une diminution de l'activité cholinergique. D'après l'hypothèse cholinergique, une perte de la fonction cholinergique dans le système nerveux central contribue de manière significative au déclin cognitif associé à l'âge avancé et à la maladie d'Alzheimer<sup>4,20,21</sup>.

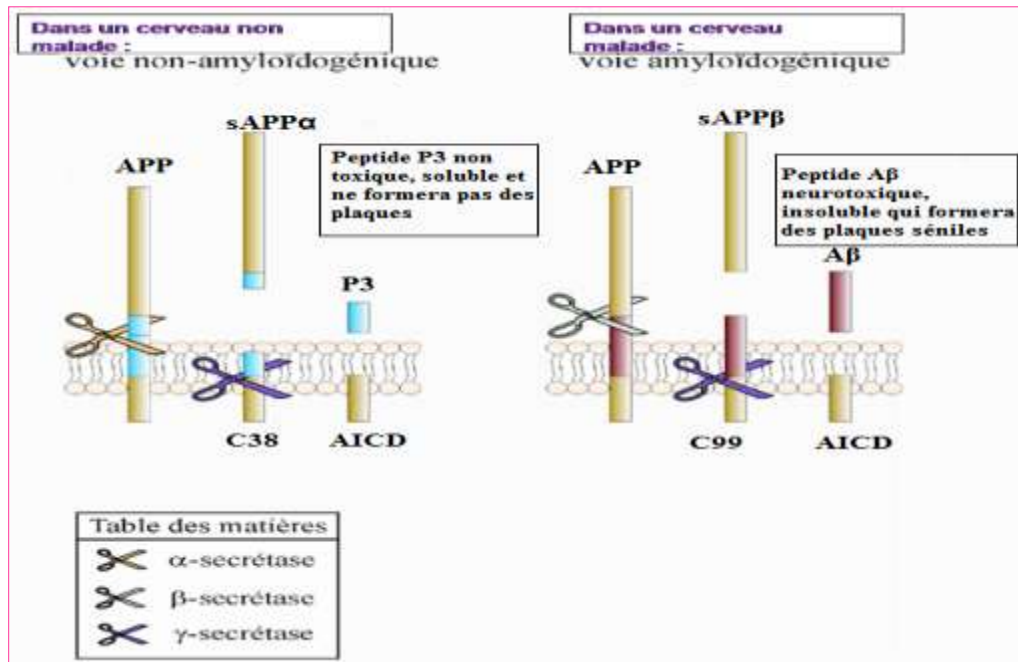
### 1.2.3.1. Dépôts extracellulaires du peptide $\beta$ -amyloïde

Les peptides  $A\beta$  sont issus du clivage de la protéine transmembranaire précurseur de la protéine amyloïde (APP) qui joue un rôle dans la neurotransmission, par sa présence en quantité importante au niveau des synapses<sup>22</sup>. Habituellement, l'APP est caractérisé par une voie non amyloïdogénique où ce précurseur est clivé par l' $\alpha$ -sécrétase et  $\gamma$ -sécrétase ce que produit le fragment sAPP $\alpha$  et p3 ont des effets neurotrophiques bénéfiques<sup>23</sup>. Chez les gens atteints la maladie d'Alzheimer, le l'APP est métabolisé en  $\beta$ -amyloïde ( $A\beta$ ). Cette voie amyloïdogène résulte d'une mutation de l'APP dont leur gène qui est localisé sur le chromosome 21. Cette voie amyloïdogène implique la libération du fragment sAPP $\beta$  (fragment de 99 acides aminés) et à son tour est rapidement clivé par la  $\gamma$ -sécrétase pour générer fragment  $\beta$ -amyloïde, qui ne sera pas dégradé par les lysosomes, ce qui produit une accumulation de la protéine  $\beta$ -amyloïde d'environ 40 à 42 acides aminés (Figure 4). L'élévation des niveaux de peptides  $A\beta$  de 42 acides aminés conduit à une agrégation de l'amyloïde qui provoque une toxicité neuronale<sup>22,24</sup> (Figure 4). Ces dépôts extracellulaires se passent entre les neurones, commencent par le cortex pour se propager ensuite dans l'hippocampe. Ces dépôts empêchent le fonctionnement des neurones et entraînent une mauvaise transmission des messages nerveux et provoquant un effet toxique pour le cerveau dû à une entrée massive de calcium, ce qui va augmenter la production de radicaux libres responsables du stress oxydant dans le cortex cérébral<sup>17,25</sup>.

### 1.2.3.2. Dégénérescences neurofibrillaires

La protéine *Tau* est une protéine d'association aux microtubules (MAP), elle se localise surtout dans les neurones du système nerveux central. La dégénérescence neurofibrillaire (DFN) correspond à une accumulation anormale de filaments hélicoïdaux intraneuronaux de protéine *Tau* hyperphosphorylée<sup>26, 27</sup>. La DFN survient d'abord dans l'hippocampe et peut ensuite être observée dans tout le cortex cérébral. Par ailleurs, il a été prouvé que l'hyperphosphorylation est induite par l'exposition de peptides  $A\beta$  provoquait la dysrégulation et l'activation de certaines protéines kinases (GSK-3b, CDK5, ERK2) qui entraînent une hyperphosphorylation de *Tau*, donc il existe un lien étiologique entre la cascade amyloïde et les dégénérescences neurofibrillaires<sup>29-31</sup>.





**Fig.4.** Les voies impliquées dans les processus de métabolisation de la protéine précurseur de l'amyloïde (APP).

### 1.2.3.3. Déficit cholinergique

Dans le cas des neurones cholinergiques, l'acétylcholine (ACh) joue le rôle d'un neurotransmetteur. Une carence en ACh est une caractéristique importante de la maladie Alzheimer. Le déficit en acétylcholine chez les malades est aggravé par l'action d'une enzyme l'acétylcholinestérase (AChE), qui hydrolyse l'acétylcholine<sup>20,21</sup>. Le vieillissement normal provoque une légère diminution de la concentration en acétylcholine, provoquant des oublis périodiques. Cependant, dans la MA, la concentration de l'ACh peut être réduite jusqu'à 90%, ce qui entraîne un déclin significatif de la mémoire et du comportement<sup>32,33</sup>. L'hypothèse cholinergique affirme que la perte de la fonction cholinergique dans le système nerveux central est associée aux symptômes cognitifs du dysfonctionnement de la mémoire.

Les origines de dégénérescences neurofibrillaires et le déficit cholinergique, ainsi que leurs causes et leurs importances qu'ils occupent dans le traitement de la maladie Alzheimer seront discutées en détail dans la partie deux de ce chapitre.

### 1.2.4. Facteurs de risque et de protection

La maladie d'Alzheimer est une pathologie multifactorielle. On distingue deux formes de maladie d'Alzheimer : la forme familiale (ou héréditaire) qui touche un petit pourcentage des personnes, puisqu'elle est beaucoup plus rare, dépendamment de la génétique des parents. La deuxième forme la plus répandue est la forme sporadique (ou non héréditaire). Elle concerne 99 % des cas, est due à un ensemble de facteurs de risque touchant à la génétique, à notre environnement et à notre style de vie<sup>34</sup>.

Les facteurs de risque génétique augmentant le risque de développer une MA, comme dans transmission héréditaire sont causées par des mutations sur trois gènes (le gène Amyloid Protein Precursor (APP) sur le chromosome 21, le gène préséniline 1 (PS1) sur le chromosome 14 et le gène Préséniline 2 (PS2) sur le chromosome 1). À côté de ces gènes, un 4ème gène est impliqué dans la pathogénie de la maladie d'Alzheimer dans la forme sporadique. Il s'agit du gène Apolipoprotéine E (ApoE4)<sup>35</sup>. Ce gène étant un facteur de risque, il n'est ni nécessaire et ni suffisant pour développer la maladie. Cela signifie qu'une personne qui n'est pas porteuse du gène ApoE4 peut quand même développer la maladie d'Alzheimer et qu'un individu porteur du gène ApoE4 n'en sera pas nécessairement atteint.

Le principal facteur de risque de la maladie d'Alzheimer est l'âge avancé et la prévalence augmente avec l'âge: 1% à 65 ans et jusqu'à 50% à 85 ans<sup>36</sup>. L'hypertension, l'hypercholestérolémie, les accidents cardiovasculaires et cérébraux et le diabète augmentent également le risque de développement d'Alzheimer, ainsi que le régime alimentaire et le style de vie contribuent à la progression de la maladie, tels que le tabagisme et l'obésité. Une dépression avancée peut provoquer une atrophie cérébrale, une réduction des niveaux des neurotrophines et l'augmentation des réponses inflammatoires dans le cerveau<sup>37, 38</sup>.

En revanche, le traitement des troubles dépressifs réduira également le risque de développer la MA, tout comme le traitement des troubles cardiovasculaires et du diabète<sup>37</sup>. Un régime alimentaire composé en grande partie d'aliments à base de plantes, d'huile d'olive et de poisson, riches en Oméga3 et en acides gras réduire le risque de développer la MA<sup>39</sup>. Une activité physique accrue est également bénéfique et la quantité d'activité semble être plus importante pour la prévention de la MA que le type d'activité. Une étude longitudinale de 8 ans a montré qu'un exercice d'au moins 30 minutes réduisait le risque de déclin cognitif<sup>37, 40</sup>.

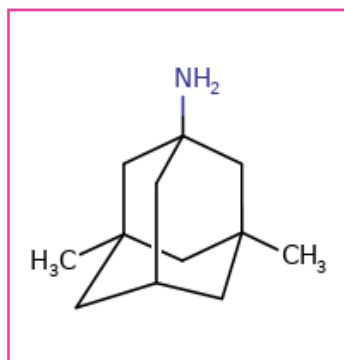
### 1.3. Approches thérapeutiques

#### 1.3.1. Thérapies actuellement disponibles

Actuellement, il n'existe toujours pas de traitement curatif ni de traitement permettant d'enrayer efficacement la progression de la maladie<sup>36</sup>. Cependant les médicaments qui sont utilisés spécifiquement pour cette maladie vont avoir pour objectif d'améliorer, de retarder ou de stabiliser la progression du déclin cognitif, ainsi que de diminuer au maximum les conséquences sur la vie quotidienne. De plus, la présence de la barrière hémato-encéphalique (BHE) devient un obstacle à l'entrée des drogues dans le cerveau et à l'atteinte des lésions. Ces difficultés mènent généralement à l'échec du traitement de la MA. Deux types de médicaments à cette fin sont approuvés par la FDA (*Food and Drug Administration*) à cette fin : les anti-acétylcholinestérasés et la mémantine.

##### 1.3.1.1. Les inhibiteurs des récepteurs glutamatergiques NMDA

La mémantine (Ebixa®) est un antagoniste non compétitif des récepteurs glutamatergiques *N*-Méthyl-D-Aspartate NMDA (Figure 5)<sup>41,42</sup>. Une forte concentration du neurotransmetteur glutamate induit un effet neurotoxique, appelé excitotoxicité. Cela se traduit par l'activation excessive de son récepteur *N*-Méthyl-D-Aspartate (NMDA). Ceux-ci deviennent très perméables et permettent l'influx massif intracellulaire des ions de  $\text{Ca}^{2+}$  induit par les dépôts amyloïdes. L'augmentation de concentration intracellulaire de  $\text{Ca}^{2+}$  provoque la production de radicaux libres et d'autres enzymes qui contribuent à la mort des cellules neuronales<sup>43</sup>. La mémantine, par son action antagoniste sur le récepteur du glutamate, bloque l'entrée massive d'ions  $\text{Ca}^{2+}$  ce qui retarde le processus de neurodégénérescence. Il est approuvé par la FDA pour le traitement de la maladie d'Alzheimer modérée à grave. Les vertiges, les courbatures, les maux de tête et la constipation sont des effets secondaires courants<sup>42</sup>.

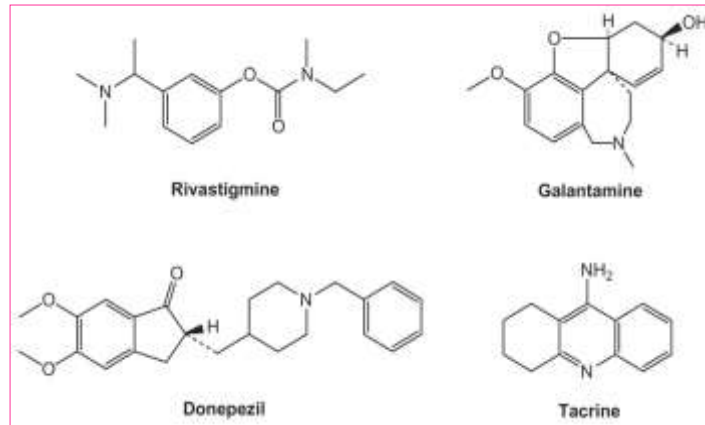


**Fig.5.** La structure de la mémantine (Ebixa®).

### 1.3.1.2. Les inhibiteurs de l'acétylcholinestérase

Les inhibiteurs de l'acétylcholinestérase agissent en augmentant le taux d'acétylcholine, c'est un neurotransmetteur utilisé par les cellules nerveuses pour communiquer entre eux et qui est important pour l'apprentissage, la mémoire et les fonctions cognitives. Le déficit neurotransmission cholinergique constitue actuellement un des axes principaux pour développer de nouveaux outils thérapeutiques pour le traitement de la maladie d'Alzheimer. Les inhibiteurs de l'acétylcholinestérase (AChEI) ont fait l'objet des recherches pharmacologiques au cours du siècle dernier et ont été développés en tant que produits neuropharmaceutiques<sup>44</sup>.

Quatre inhibiteurs d'acétylcholinestérase ont été développés et mis sur le marché pour le traitement symptomatique de la maladie d'Alzheimer. Il s'agit de la tacrine (Cognex®), le donépézil (Aricept®), la galanthamine (Reminyl®) et la rivastigmine (Exelon®)<sup>45</sup> (Figure 6). Ces médicaments sont approuvés que pour le soulagement symptomatique. La tacrine ayant été développée en tout premier, puis retirée du marché en raison de l'importance de ses effets secondaires. Les autres molécules ont également des effets secondaires pouvant largement retentir sur le confort du malade, tel que : des nausées, des vomissements, des douleurs abdominales, de douleurs musculaires, de troubles du sommeil. Le donépézil peut être utilisé à tous les stades de la maladie d'Alzheimer. La galantamine et la rivastigmine ne sont approuvées que pour le traitement de la maladie d'Alzheimer légère à modérée. Le donépézil et la galantamine sont des inhibiteurs rapides et réversibles de l'acétylcholinestérase. La rivastigmine est un inhibiteur lent et réversible de l'acétylcholinestérase et de la butyrylcholinestérase. La recherche de meilleurs inhibiteurs de l'AChE se poursuit pour le but de régler l'activité cholinergique centrale et diminuer les effets secondaires. Dans notre travail, nous nous intéressons à cette famille de médicaments (AChEI) pour criblage virtuel de nouveaux anti-acétylcholinestérases.



**Fig.6.** La structure des inhibiteurs de l'acétylcholinestérase.

### 1.3.2. Les futurs traitements

#### 1.3.2.1. Les perspectives thérapeutiques agissant sur la voie amyloïde

Le peptide  $A\beta_{42}$  est libéré après clivage par la sécrétase  $\beta$  puis  $\gamma$  à partir du précurseur de l'amyloïde. La  $\gamma$ -sécrétase est considérée comme une cible clé pour le développement de traitements modificateurs de la maladie. Les études ont montré que les inhibiteurs de la  $\gamma$ -sécrétase (GSI), administrés par voie orale, sont capables de réduire les concentrations d' $A\beta$  dans le cerveau. Mais ces inhibiteurs peuvent provoquer aggravation inattendue des déficits cognitifs et des effets indésirables<sup>46-48</sup>. Raven *et al.* ont développé une classe de modulateurs solubles de la 2-aminothiazole  $\gamma$ -sécrétase (SGSM) qui diminue de manière préférentielle les taux d' $A\beta_{42}$  dans le cerveau chez les souris d'une manière distincte des GSI développées antérieurement<sup>49</sup>. Le potentiel thérapeutique des inhibiteurs de  $\beta$ -sécrétase (BACE1) réduisent efficacement les taux d' $A\beta$  dans le cerveau, les essais cliniques ne démontrent toujours pas d'amélioration de la fonction cognitive chez les patients atteints de MA<sup>49,50</sup>. Il existe de nombreuses preuves montrent que les inhibiteurs de BACE1 peuvent agir en tant qu'agents modificateurs de la maladie et que leur utilisation devrait être poursuivie pour que les patients atteints de MA aient des effets bénéfiques<sup>51</sup>.

#### 1.3.2.2. Perspectives thérapeutiques ciblées dégénérescences neurofibrillaires

L'hyperphosphorylation de la protéine *Tau* perturbe la fonction biologique normale de ces protéines. Cette hyperphosphorylation est le résultat de la dérégulation de plusieurs protéines kinases telles que la kinase dépendant du cycline 5 (CDK5), Caséine kinase 1 (CK1) et glycogène synthase kinase 3 (GSK3)<sup>52,53</sup>. Ces kinases constituent une classe prometteuse de cibles potentielles pour le traitement de la MA. Parmi celles-ci, les plus étudiées sont  $3\beta$  (GSK 3 $\beta$ ) et CDK5 et sont en cours des essais cliniques.

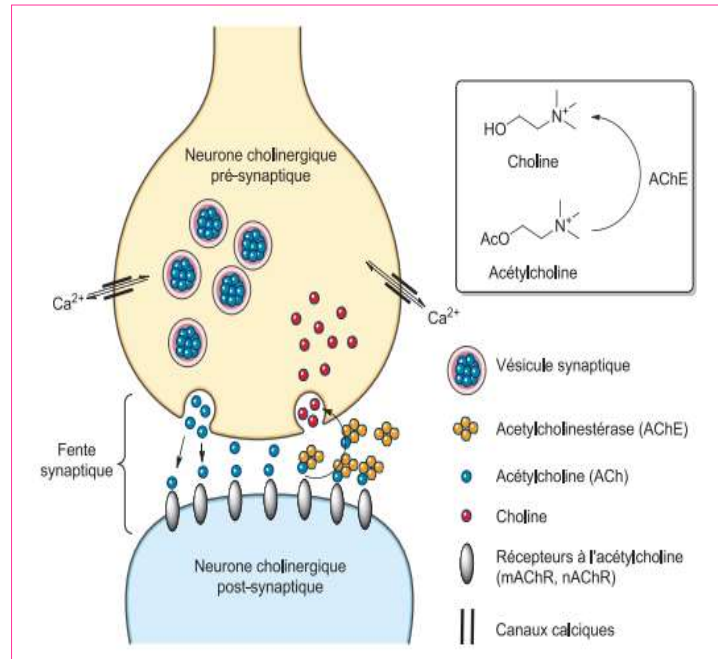
### Partie 2 : Cibles thérapeutiques

Dans cette partie, nous détaillerons le rôle et les caractéristiques structurales de l'enzyme d'acétylcholinestérase (AChE) et de kinase dépendante du cycline 5 (cyclin-dependent kinase 5 ou CDK5), ces deux cibles thérapeutiques sont parmi les cibles désignées dans le traitement de la maladie d'Alzheimer. Nous montrerons également la place importante qu'elles occupent au sein des préoccupations actuelles dans le traitement de la maladie Alzheimer. Nous avons basé dans notre étude sur ces deux cibles pour la prédiction et le développement de nouveaux inhibiteurs puissants d'AChE et de CDK5.

#### 2.1. Acétylcholinestérase

##### 2.1.1. Neurotransmission

La neurotransmission est un processus par lequel un influx nerveux est délivré entre un neurone présynaptique à un neurone post-synaptique, via un intervalle appelé fente synapses. L'arrivée d'un signal nerveux provoque l'ouverture de canaux calciques, de sorte que ces ions de calcium activent la libération des entités chimiques « acétycholine » par les vésicules qui les stockent au niveau de synapse, ensuite ils se lient à des récepteurs post-synaptiques « récepteur cholinergique », le signal nerveux est ainsi transmis au neurone post-synaptique 'initiant une réponse' (Schéma 1). Ces entités chimiques responsables de transmettre cette information nerveuse sont appelées neurotransmetteurs<sup>54</sup>. Après la transmission du signal, l'ACh doit être dégradée très rapidement, en quelques millisecondes, afin d'éviter la surexcitation des neurones cholinergiques et revenir à leur état de repos, ainsi permettre la transmission de l'influx nerveux suivant<sup>55</sup>. Cette fonction essentielle est assurée par une réaction d'hydrolyse due par l'acétylcholinestérase (AChE), ce dernier est considéré parmi les enzymes les plus rapides de la nature (en moyenne de 10.000 molécules de substrat hydrolysées par seconde)<sup>56</sup>. Dans le système nerveux central, l'acétylcholine intervient notamment dans l'apprentissage et la mémoire. C'est un messenger chimique clé utilisé par les neurones à transmettre des signaux à l'autre qui souscrivent à notre traitement cognitif ainsi que nos pensées de base.



**Schéma 1.** Rôle de l'AChE au niveau des synapses.

### 2.1.2. Structure tridimensionnelle d'acétylcholinestérase

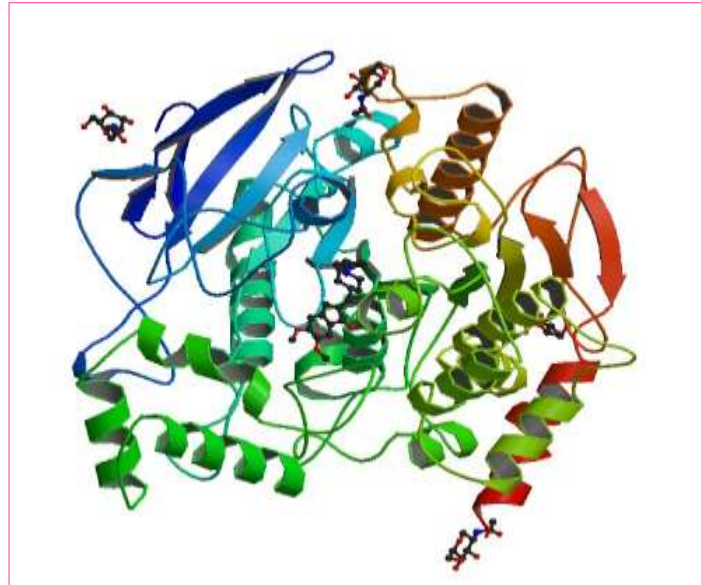
L'acétylcholinestérase (EC 3.1.1.7) est une protéase à sérine, regroupant 537 acides aminés, appartient à la famille des cholinestérases, ils sont capables d'hydrolyser les cholinesters dont fait partie l'acétylcholine (ACh), par laquelle il se produit les cholines et les esters au niveau des synapses dites cholinergiques. Les synapses sont retrouvées au niveau des jonctions neuromusculaires, ainsi que dans les zones du cortex en charge des fonctions cognitives (mémoire, orientation, jugement...etc.)<sup>20</sup>.

L'AChE a été identifiée pour la première fois en 1986 à partir du *Torpedo californica* (rayon électrique du Pacifique)<sup>58,59</sup>. Les structures primaires de AChE sont très proches : plus de 60% des résidus sont identiques entre l'AChE de *torpedo californica* et celle de mammifères, et en particulier au niveau du site actif<sup>60,61</sup>. La structure de l'AChE c'est un dimère entier, chaque monomère représente un repliement de type  $\alpha/\beta$  hydrolases (constitué de 11 feuilletts  $\beta$  et de 15 hélices  $\alpha$ ), ainsi qu'extrémité N-terminale présente courte portion de feuillet  $\beta$  qui n'interagit pas avec le reste de la structure. Deux hélices  $\alpha$  de chaque monomère, dont l'hélice  $\alpha$  C-terminale, s'associent pour former le dimère. Le paquet de 4 hélices  $\alpha$  assurent la cohésion des deux monomères, donc ils assurent la stabilité de ce dimère (Figure 7).

La résolution de la structure de la *TcAChE* a permis de révéler l'existence de deux zones essentielles pour la catalyse enzymatique : le site actif et le site périphérique, ce dernier est situé

## Maladie d'Alzheimer & Cibles thérapeutiques

à l'entrée de la gorge hydrophobe profonde (env.20Å) et étroite (env.5Å) bordée de résidus aromatiques conservés qui menant au site actif au fond<sup>60, 62</sup>(Schéma 2). Ces deux sites seront ainsi deux cibles potentielles en vue de l'inhibition de l'AChE par des composés thérapeutiques.



**Fig.7.** Structure de monomère de la *TcAChE* (1EVE:PDB) : structure 3D avec représentation des hélices  $\alpha$ , feuillets  $\beta$  et résidus du site actif. La coloration en arc-en-ciel va du domaine C-terminal en rouge vers domaine N-terminal en bleu.

### 2.1.2.1. Site actif

Dans le site actif se trouve la machinerie catalytique de l'enzyme, se subdivise en deux sous sites : le site anionique et le site estérasique<sup>63</sup> (Schéma 2).

➤ **Le site estérasique** : Ce site correspond au locus au niveau duquel la portion acétyl de la molécule d'ACh se fixe et forme l'intermédiaire tétraédrique ; celle-ci comprend :

- **Triade catalytique** : possède trois résidus d'acide aminés qui interviennent ensemble dans de certaines hydrolases sont responsable de l'hydrolyse de l'ACh (Nucléophile : SER200, Base : HIS440, Acide : GLU327).
- **Trou oxyanion** : contient les résidus GLY118, GLY119 et ALA201, ce site favorise la formation de l'intermédiaire tétraédrique en accueillant l'oxygène négativement chargé de son carbonyle de l'ACh et le stabilisée.
- **Poche acyle** : forme résidus hydrophobes TRP233, PHE288, PHE290 et Phe331, ici où groupement méthyle de la portion acétate ce stabilise. Cette poche assure un rôle essentiel dans la sélectivité d'AChE pour ces substrats.



## Maladie d'Alzheimer & Cibles thérapeutiques

➤ **Le site anionique** : Se située en face de site estérasique, au fond de la gorge, c'est un ensemble de résidus aromatique chargé négativement (TRP84, GLU199 et PHE330), contribue à la stabilisation de la charge positive portée par la portion choline de l'ACh, stabilisée par l'attraction électrostatique du résidu Glu199, mais le sera beaucoup plus effectivement par son interaction avec les électrons  $\pi$  des noyaux aromatiques des résidus TRP84 et PHE330.

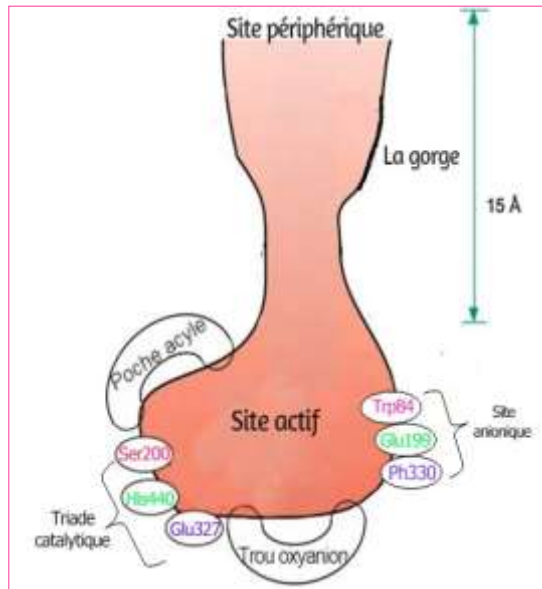


Schéma 2. Vue schématique du site actif de TcAChE.

### 2.1.2.2. Site périphérique

Le PAS est situé à environ 15 Å au-dessus du site actif. Il est caractérisé par ces résidus aromatiques (TYR70, ASP72, TYR121, TRP279 et TYR334) constituent ce qu'on appelle le site périphérique anionique (PAS) se localisant au bord de la cavité (gorge) de l'enzyme, il se lie aux substrats cationiques, et notamment aux ligands trop gros pour pénétrer dans la gorge. Il fournit un site de liaison pour les inhibiteurs allostériques<sup>63,64</sup>

### 2.1.3. Implication de l'AChE dans la maladie d'Alzheimer

L'hypothèse cholinergique fut introduite pour expliquer les pertes de mémoire et de cognition<sup>21</sup>. Elle suggère qu'un déficit en acétylcholine dans le système nerveux central contribue aux défaillances cognitives et mnésiques. Comme on a vu précédemment, que la maladie d'Alzheimer est l'atrophie du cortex cérébral. Celle-ci est notamment due à une perte massive de neurones cholinergiques et une diminution assez forte du taux de l'ACh dans le cerveau, notamment l'hippocampe et le cortex cérébral.

En effet, la présence de plaques amyloïdes neurotoxiques est connue pour influencer négativement les sites récepteurs de la choline, perturbant de fait son recyclage et inhibant la libération de l'ACh dans la fente synaptique<sup>65</sup>. Certaines études ont montré que l'hypothèse de l'activation des récepteurs cholinergiques et en particulier ceux de type muscariniques peut jouer un rôle dans la protéolyse de l'APP par l' $\alpha$ -sécrétase (voie non amyloïdogénique) et donc la synthèse du peptide  $\beta$ -amyloïde sera diminuée<sup>66,67</sup>. Les dégénérescences neurofibrillaires détériorent également les fonctions cholinergiques<sup>65, 20,21</sup>. D'autres études ont montré que l'AChE peut jouer un rôle médiateur dans d'autres processus liés à son site anionique périphérique (PAS), notamment l'accélération de l'agrégation d'amyloïde et il forme des complexes stables avec les plaques amyloïdes, ces complexes augmentent la neurotoxicité dépendante de  $\beta$ -amyloïde<sup>68</sup>. Le site anionique périphérique (PAS) est une cible à privilégier dans la recherche de traitements contre la MA. La plupart des médicaments anti-Alzheimer sont actuellement des inhibiteurs de l'AChE qui agissent au niveau de son site actif en vue de restaurer la fonction cholinergique.

En conclusion, pour ces diverses raisons et ainsi avec la disponibilité croissante de structures cristallines AChE, avec ou sans ligands liés, la recherche de nouveaux inhibiteurs AChEI prennent en compte le rôle du PAS dans l'apparition des plaques séniles, se focalise actuellement sur la production des inhibiteurs AChEI réversibles spécifiques qui agissent à la fois au niveau du site actif et du PAS. Devraient non seulement restaurer la fonction cholinergique, mais également empêcher l'assemblage de l'A $\beta$ , ce qui retardera le processus neurodégénératif. Ces rapports ont suscité un nouvel intérêt pour les AChEI.

## 2.2. Kinases dépendantes des cyclines 5

### 2.2.1. Protéines kinases

Les protéines kinases sont un groupe d'enzymes qui catalysent la phosphorylation des protéines cellulaires en transférant un groupe phosphate d'adénosine triphosphate (ATP) et en le liant de manière covalente à des groupes hydroxyle d'acides aminés, par exemple la tyrosine, la sérine ou la thréonine<sup>69-71</sup> (Figure 8). La phosphorylation des protéines ciblées entraîne des changements de conformation et altère la fonction des protéines ciblées en modifiant son activité enzymatique soit pour l'activer (phosphorylation positive), soit pour l'inhiber (phosphorylation négative).

Par conséquent, les kinases agissent comme des régulateurs importants des voies cellulaires, telles que la transduction du signal, la croissance cellulaire et la survie<sup>72,73</sup>. Certaines protéines

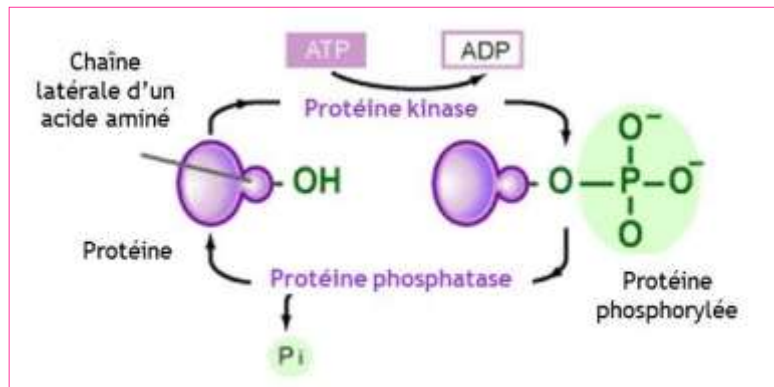
## Maladie d'Alzheimer & Cibles thérapeutiques

kinases ont besoin de l'activation par l'association avec une sous-unité régulatrice, la cycline, pour être fonctionnelles, on les appelle kinases dépendantes des cyclines (CDK).

### 2.2.2. Kinases dépendantes des cyclines

Les kinases dépendantes des cyclines (en anglais, cyclin-dependent kinase ou CDKs). Contrôlent une multitude d'événements tout au long du cycle cellulaire <sup>74,75</sup>. Les CDKs appartiennent à la famille des sérines/thréonine kinases, c'est-à-dire qu'elles catalysent le transfert d'un groupement phosphate sur les résidus sérine et thréonine de leurs substrats. Comme toutes les protéines kinases.

Dans notre travail nous nous sommes intéressés par les kinases dépendantes des cyclines



**Fig.8.** Schématisation des mécanismes de phosphorylation/déphosphorylation. 5 (CDK5). Le CDK5 est également une CDK particulière, car elle est activée par la fixation des protéines appelées p35 ou p39, ne partageant pas d'identités de séquences avec les cyclines, sauf au niveau de 17 acides aminés de p35 qui présentent une similarité avec une région conservée des cyclines <sup>76</sup>.

### 2.2.3. Structure de Kinases dépendantes des cyclines 5

Le cyclin-dependent-kinase 5 (CDK5) possède 58% et 60% d'homologies de séquence d'acides aminés avec CDK1 et CDK2 respectivement, cette similitude se situe surtout au niveau des domaines caractéristiques des kinases <sup>77,78</sup>. C'est un membre unique de la famille puisqu'il nécessite une association avec l'activateur p35 pour l'activation au lieu des cyclines <sup>75</sup>. L'activité de la CDK5 est limitée aux neurones, car son activateur p35 est spécifique pour les neurones.

La kinase CDK5 est une protéine de 292 acides aminés, d'un poids moléculaire de 33 kDa <sup>79,80</sup>. Tous comme les autres CDKs, CDK5 est constituée d'un petit lobe N-terminal et d'un

grand lobe C-terminal, le lobe N-terminal se compose d'un feuillet  $\beta$  et d'une hélice, le grand lobe C est composé d'un feuillet  $\alpha$ . C'est entre ces deux structures que l'on retrouve le site de liaison à l'ATP, et aux substrats, ainsi qu'une structure en boucle terminale « T-loop » dont l'orientation permet la reconnaissance du substrat (Figure 9).

### 2.2.3.1. Activation de CDK5

Malgré les similitudes structurales observées entre CDK 5 et les autres CDKs, l'activation de CDK5 diffère sur certains points que nous abordons ci-dessous. Contrairement aux autres CDKs du cycle cellulaire qui sont dépendantes des cyclines pour leur activation, CDK5 est activée par trois partenaires activateurs p39, p35 et p25.

#### ➤ Les protéines p35 et p39

La protéine p35, également appelé Nck5a (Neuronal Cdk5 Activator), il a été démontré que p35 était capable de s'associer et d'activer CDK5 in vivo et in vitro<sup>81</sup>. Le système Cdk5/p35 participe au développement normal des cellules neuronales et à la croissance axonale<sup>82</sup>. L'isoforme p39 ou Nck5ai (Neuronal CDk5 Activator Isoform) a été identifié par homologie de séquence avec p35<sup>83</sup>. Néanmoins, ces petites molécules adoptent une structure tridimensionnelle similaire au domaine globulaire des cyclines, permettant l'interaction avec les CDKs<sup>84</sup>. Les protéines p39 et p35 sont majoritairement exprimées au niveau des cellules post-mitotiques du système nerveux central.

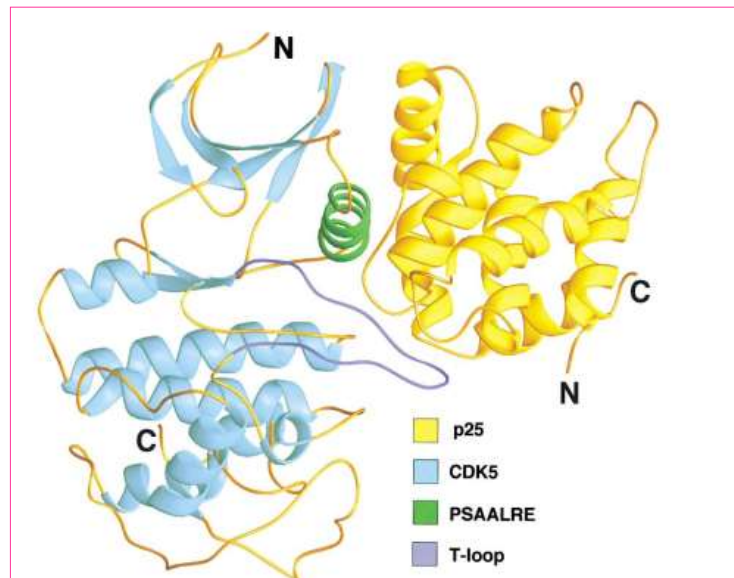


Fig.9. Structure de la kinase CDK5 lié à p25.

#### ➤ La protéine p25

La forme majoritaire activatrice de CDK5 dans le cerveau était p35. Le calpaïne

## Maladie d'Alzheimer & Cibles thérapeutiques

(Protéase activée par le calcium) clive la protéine p35 du côté N-terminal en deux protéines : p10 et p25, cette dégradation est induite par diverses conditions neurotoxiques telles que le stress oxydatif, l'exposition au peptide  $\beta$ -amyloïde et l'augmentation de la concentration du calcium<sup>85,86</sup>.

Cette dégradation n'augmente pas que l'activité kinase de CDK5, mais augmente aussi la stabilité du complexe, puisque le temps de demi-vie de l'activateur p25 est trois fois plus élevé que celui de p35 qui n'est que de 20 à 30 minutes<sup>87</sup>. Cette forme stable se relocalise anormalement dans le cytoplasme, et induit des phosphorylations associées à l'apoptose des neurones. L'accumulation de la forme CDK5/p25 est également associée à des dégénérescences anormales telles que celles caractérisant la maladie d'Alzheimer<sup>86</sup>.

### 2.2.4. Rôles de Cdk5/P25 dans la maladie d'Alzheimer

#### 2.2.4.1. Dégénérescence neurofibrillaire (DFN)

La protéine *Tau* (Tubulin associated unit) de la famille des MAPs (Microtubule Associated Protein) participe à l'architecture des neurones, nécessaire pour la stabilité du cytosquelette. Comme son nom l'indique, la protéine *Tau* interagit avec la tubuline des microtubules<sup>26,88</sup>.

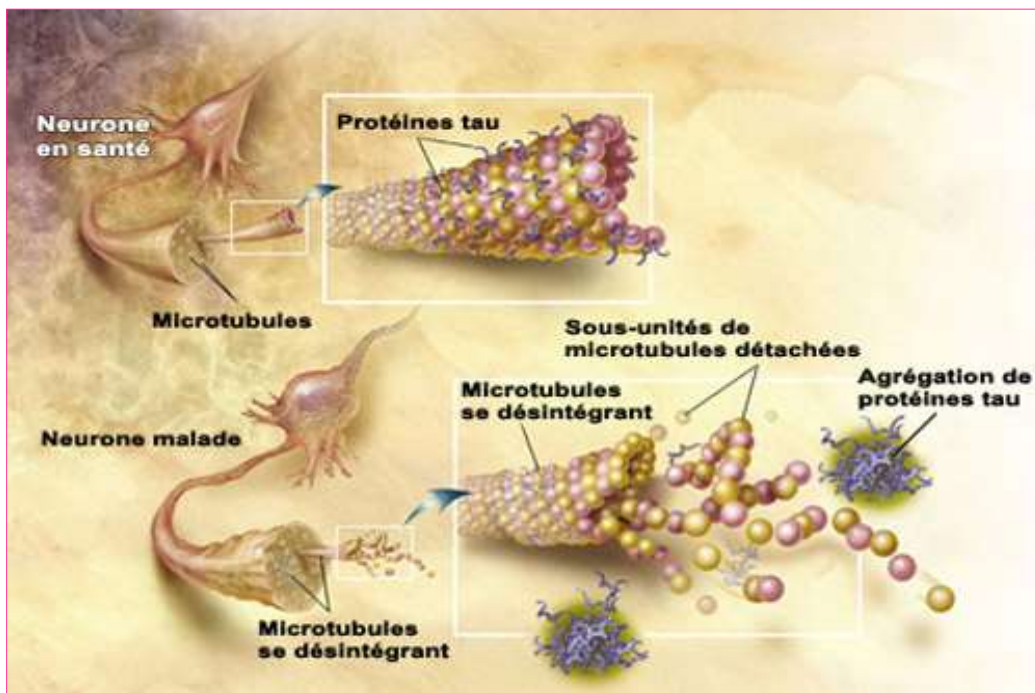
Dans le cerveau normal, l'équilibre entre la phosphorylation et la déphosphorylation de la protéine *Tau* module la stabilité du cytosquelette (la protéine *Tau* est phosphorylée sur 2-3 sites associés aux microtubules). Par conséquent, la rupture de cet équilibre provoque hyperphosphorylation anormale sur plus de 20 sites (l'ajout de groupements phosphates) par la CDK5, ces groupements supplémentaires gênent la polymérisation des microtubules ce qui conduit à leur déstructure et altération structurelle de la protéine *Tau*. La formation des agrégats des protéines *Tau*, des amas de protéines *Tau* se forment (paired helical filament PHF) constituant les dégénérescences neurofibrillaires DFN (enchevêtrements neurofibrillaires) observées dans les cerveaux de patients atteints de la maladie d'Alzheimer (Figure 10)<sup>26-28</sup>.

Les biologistes cellulaires et les neuroscientifiques ont manifesté un vif intérêt en découvrant que dérégulation du complexe CDK5/p35 déclenché par le peptide  $\beta$ -amyloïde ( $A\beta$ ), est impliquée dans les hyperphosphorylations de la protéine *Tau*<sup>85,89,90</sup>.

## Maladie d'Alzheimer & Cibles thérapeutiques

La dérégulation de la CDK5 est un événement majeur dans le processus neurodégénératif de la maladie d'Alzheimer. Chez les souris injectées avec  $A\beta_{1-40}$ <sup>92</sup>, la dérégulation de cette sérine-thréonine kinase dirigée via l'activation du calpaïne, un groupe de protéases cytosoliques dépendantes du calcium<sup>91</sup>. L'activation du calpaïne, déclenchée par l'augmentation des taux de calcium intracellulaire induite par l'exposition à l' $A\beta$ <sup>93-94</sup>, entraînera le clivage de l'activateur de CDK5/p35 à p25 (sous-unité pathogène), s'associe à la CDK5 formant un complexe CDK5/p25 hyperactif.

Globalement, ces données démontrent que la sur-activation de CDK5 est associée à une hyperphosphorylation de *Tau*, qui aboutit à la diminution de l'affinité entre la protéine *Tau* et les microtubules, ce qui peut entraîner leur déstabilisation et perturbe le transport axonal et finit par mort des neurones « dégénérescence neurofibrillaire DNF ». La perte neuronale est très fortement corrélée à la gravité des déficits cognitifs chez les patients Alzheimer. Ils surviennent d'abord dans l'hippocampe et peuvent ensuite être observés dans tout le cortex cérébral<sup>29-31</sup>. C'est pourquoi les kinases de type CDK5 constituent une classe prometteuse de cibles potentielles pour le traitement de la maladie d'Alzheimer. Ce sujet constitue un défi majeur pour les futures études.



**Fig.10.** Dégénérescence neurofibrillaire induite par l'hyperphosphorylation de protéine Tau.

### Références bibliographique :

1. Alzheimer, A. (1906) Über Einen Eigenartigen, Schweren Erkrankungsprozeß Der Hirnrinde. *Neurol. Cent*, 23, 1129–1136.
2. Alzheimer, A. (1911). Concerning unusual medical cases in old age. *Zeitschrift Fur Die Gesamte Neurologie Und Psychiatrie*, 4(1), 356-385.
3. Whalley, L. J. (1982). The dementia of Down's syndrome and its relevance to aetiological studies of Alzheimer's disease. *Annals of the New York Academy of Sciences*, 396(1), 39-53.
4. Atack, J. R., Perry, E. K., Bonham, J. R., Perry, R. H., Tomlinson, B. E., Blessed, G., & Fairbairn, A. (1983). Molecular forms of acetylcholinesterase in senile dementia of Alzheimer type: selective loss of the intermediate (10S) form. *Neuroscience letters*, 40(2), 199-204.
5. Alzheimer's Disease International. 2019. *Rapport Mondial Alzheimer 2019 : Attitudes à l'égard de la démence*. Londres.
6. Fratiglioni, L., Launer, L. J., Andersen, K., Breteler, M. M., Copeland, J. R., Dartigues, J. F., ... & Hofman, A. (2000). Incidence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology*, 54(11), 10-5.
7. Alzheimer's Disease International. 2018. *Rapport Mondial Alzheimer 2018 : Attitudes à l'égard de la démence*. Londres : Alzheimer's Disease International.
8. Vina, J., & Lloret, A. (2010). Why women have more Alzheimer's disease than men: gender and mitochondrial toxicity of amyloid- $\beta$  peptide. *Journal of Alzheimer's disease*, 20(2), 527-533.
9. La liberte (2010). Les risques d'alzheimer augmentent. <https://www.liberte-algerie.com/actualite/les-risques-dalzheimer-augmentent-77329>. Consulte-le : [15/03/2018].
10. Kumar, A., & Tsao, J. W. (2018). Alzheimer Disease. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing.
11. Heneka, M. T., Golenbock, D. T., & Latz, E. (2015). Innate immunity in Alzheimer's disease. *Nature immunology*, 16(3), 229-236.
12. Gosche, K. M., Mortimer, J. A., Smith, C. D., Markesbery, W. R., & Snowdon, D. A. (2002). Hippocampal volume as an index of Alzheimer neuropathology: findings from the Nun Study. *Neurology*, 58(10), 1476-1482.
13. Cummings, J. L., & Back, C. (1998). The cholinergic hypothesis of neuropsychiatric symptoms in Alzheimer's disease. *The American Journal of Geriatric Psychiatry*, 6(2), S64-S78.
14. Lyketsos, C. G., Carrillo, M. C., Ryan, J. M., Khachaturian, A. S., Trzepacz, P., Amatniek, J., ... Miller, D. S. (2011). Neuropsychiatric symptoms in Alzheimer's disease. *Alzheimer's & Dementia*, 7(5), 532–539.
15. Audrain, M. (2016). Thèse de doctorat. *Modélisation des phases précoces de la maladie d'Alzheimer par transfert de gènes*, Sorbonne Paris.

16. Knobloch, M., & Mansuy, I. M. (2008). Dendritic spine loss and synaptic alterations in Alzheimer's disease. *Molecular neurobiology*, 37(1), 73-82.
17. Chen, Z., & Zhong, C. (2014). Oxidative stress in Alzheimer's disease. *Neuroscience bulletin*, 30(2), 271-281.
18. Tublin, J. M., Adelstein, J. M., del Monte, F., Combs, C. K., & Wold, L. E. (2019). Getting to the heart of Alzheimer disease. *Circulation research*, 124(1), 142-149.
19. Hellström-Lindahl, E. (2000). Modulation of  $\beta$ -amyloid precursor protein processing and tau phosphorylation by acetylcholine receptors. *European journal of pharmacology*, 393(1-3), 255-263.
20. Terry, A. V., & Buccafusco, J. J. (2003). The cholinergic hypothesis of age and Alzheimer's disease-related cognitive deficits: recent challenges and their implications for novel drug development. *Journal of Pharmacology and Experimental Therapeutics*, 306(3), 821-827.
21. Mesulam, M. (2004). The cholinergic lesion of Alzheimer's disease: pivotal factor or side show?. *Learning & memory*, 11(1), 43-49.
22. Murphy, M. P., & LeVine III, H. (2010). Alzheimer's disease and the amyloid- $\beta$  peptide. *Journal of Alzheimer's Disease*, 19(1), 311-323.
23. Bell, K. F., Zheng, L., Fahrenholz, F., & Cuello, A. C. (2008). ADAM-10 over-expression increases cortical synaptogenesis. *Neurobiology of aging*, 29(4), 554-565.
24. Scheuner, D., Eckman, C., Jensen, M., Song, X., Citron, M., Suzuki, N., & Larson, E. (1996). Secreted amyloid  $\beta$ -protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nature medicine*, 2(8), 864-870.
25. Selkoe, D. J. (2001). Alzheimer's disease: genes, proteins, and therapy. *Physiological reviews*, 81(2), 741-766.
26. Grundke-Iqbal, I., Iqbal, K., Tung, Y. C., Quinlan, M., Wisniewski, H. M., & Binder, L. I. (1986). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proceedings of the National Academy of Sciences*, 83(13), 4913-4917.
27. Noble, W., Hanger, D. P., Miller, C. C., & Lovestone, S. (2013). The importance of tau phosphorylation for neurodegenerative diseases. *Frontiers in neurology*, 4(83),1-10.
28. Kolarova, M., García-Sierra, F., Bartos, A., Rícný, J., & Ripova, D. (2012). Structure and pathology of tau protein in Alzheimer disease. *International journal of Alzheimer's disease*, 2012,1-13.
29. Cedazo-Minguez, A., Popescu, B. O., Blanco-Millan, J. M., Akterin, S., Pei, J. J., Winblad, B., & Cowburn, R. F. (2003). Apolipoprotein E and  $\beta$ -amyloid (1-42) regulation of glycogen synthase kinase-3 $\beta$ . *Journal of neurochemistry*, 87(5), 1152-1164.
30. Lopes, J. P., Oliveira, C. R., & Agostinho, P. (2010). Neurodegeneration in an A $\beta$ -induced model of Alzheimer's disease: the role of Cdk5. *Aging cell*, 9(1), 64-77.



## Maladie d'Alzheimer & Cibles thérapeutiques

31. Takashima, A., Noguchi, K., Sato, K., Hoshino, T., & Imahori, K. (1993). Tau protein kinase I is essential for amyloid beta-protein-induced neurotoxicity. *Proceedings of the National Academy of Sciences*, 90(16), 7789-7793.
32. Giacobini, E. (2003). Cholinesterases: new roles in brain function and in Alzheimer's disease. *Neurochemical research*, 28(3-4), 515-522.
33. Inestrosa, N. C., Dinamarca, M. C., & Alvarez, A. (2008). Amyloid–cholinesterase interactions. *The FEBS journal*, 275(4), 625-632.
34. Société Alzheimer du Canada, maladie d'Alzheimer *De quoi s'agit-il?* Québec : Éditions de la Société d'Alzheimer, 2010.
35. Raber, J., Huang, Y., & Ashford, J. W. (2004). ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiology of aging*, 25(5), 641-650.
36. Hara, Y., McKeehan, N., & Fillit, H. M. (2019). Translating the biology of aging into novel therapeutics for Alzheimer disease. *Neurology*, 92(2), 84-93.
37. Rakesh, G., Szabo, S. T., Alexopoulos, G. S., & Zannas, A. S. (2017). Strategies for dementia prevention: latest evidence and implications. *Therapeutic advances in chronic disease*, 8(8-9), 121-136.
38. Reitz, C., & Mayeux, R. (2014). Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochemical pharmacology*, 88(4), 640-651.
39. Barnard, N. D., Bush, A. I., Ceccarelli, A., Cooper, J., de Jager, C. A., Erickson, K. I., ... & Morris, M. C. (2014). Dietary and lifestyle guidelines for the prevention of Alzheimer's disease. *Neurobiology of aging*, 35, S74-S78.
40. Silvestrelli, G., Lanari, A., Parnetti, L., Tomassoni, D., & Amenta, F. (2006). Treatment of Alzheimer's disease: from pharmacology to a better understanding of disease pathophysiology. *Mechanisms of ageing and development*, 127(2), 148-157.
41. Reisberg, B., R. Doody, et al. (2003). "Memantine in moderate-to-severe Alzheimer's disease." *New England Journal of Medicine*, 348(14), 1333-41.
42. Parsons, C. G., Stöffler, A., & Danysz, W. (2007). Memantine: a NMDA receptor antagonist that improves memory by restoration of homeostasis in the glutamatergic system-too little activation is bad, too much is even worse. *Neuropharmacology*, 53(6), 699-723.
43. Hynd, M. R., Scott, H. L., & Dodd, P. R. (2004). Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer's disease. *Neurochemistry international*, 45(5), 583-595.
44. Čolović, M. B., Krstić, D. Z., Lazarević-Paši, T. D., Bondžić, A. M., and Vasić, V. M. (2013). Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Curr. Neuropharmacol.* 11(3), 315- 335.
45. Craig, L. A., Hong, N. S., & McDonald, R. J. (2011). Revisiting the cholinergic hypothesis in the development of Alzheimer's disease. *Neuroscience & Biobehavioral Reviews*, 35(6), 1397-1409.

## Maladie d'Alzheimer & Cibles thérapeutiques

46. Wolf, M. S. (2002).  $\gamma$ -Secretase as a Target for Alzheimer's Disease. *Current topics in medicinal chemistry*, 2(4), 371-383.
47. Kumar, D., Ganeshpurkar, A., Kumar, D., Modi, G., Gupta, S. K., & Singh, S. K. (2018). Secretase inhibitors for the treatment of Alzheimer's disease: Long road ahead. *European journal of medicinal chemistry*, 148, 436-452.
48. P Imbimbo, B., & AM Giardina, G. (2011).  $\gamma$ -secretase inhibitors and modulators for the treatment of Alzheimer's disease: disappointments and hopes. *Current topics in medicinal chemistry*, 11(12), 1555-1570.
49. Raven, F., Ward, J. F., Zoltowska, K. M., Wan, Y., Bylykbashi, E., Miller, S. J., ... & Wagner, S. L. (2017). Soluble Gamma-secretase Modulators Attenuate Alzheimer's  $\beta$ -amyloid Pathology and Induce Conformational Changes in Presenilin 1. *EBioMedicine*, 24, 93-101.
50. Moussa, C. E. (2017). Beta-secretase inhibitors in phase I and phase II clinical trials for Alzheimer's disease. *Expert opinion on investigational drugs*, 26(10), 1131-1136.
51. Coimbra, J. R., Marques, D. F., Baptista, S. J., Pereira, C. M., Moreira, P. I., Dinis, T. C., ... & Salvador, J. A. (2018). Highlights in BACE1 inhibitors for Alzheimer's disease treatment. *Frontiers in chemistry*, 6, 1-10
52. Jayapalan, S., & Natarajan, J. (2013). The role of CDK5 and GSK3B kinases in hyperphosphorylation of microtubule associated protein tau (MAPT) in Alzheimer's disease. *Bioinformation*, 9(20), 10231030.
53. Jayapalan, S., Subramanian, D., & Natarajan, J. (2016). Computational identification and analysis of neurodegenerative disease associated protein kinases in hominid genomes. *Genes & Diseases*, 3(3), 228-237.
54. Graham, L. P., & Patrick, D. P. (2003). Chimie pharmaceutique. *Paris: Deboek*, 380-386.
55. Batool, Z., Sadir, S., Liaquat, L., Tabassum, S., Madiha, S., Rafiq, S & Perveen, T. (2016). Repeated administration of almonds increases brain acetylcholine levels and enhances memory function in healthy rats while attenuates memory deficits in animal model of amnesia. *Brain research bulletin*, 120, 63-74.
56. Quinn, D. M. (1987). Acetylcholinesterase: enzyme structure, reaction dynamics, and virtual transition states. *Chemical Reviews*, 87(5), 955-979.
57. Zhou, Y., Wang, S., & Zhang, Y. (2010). Catalytic reaction mechanism of acetylcholinesterase determined by Born– Oppenheimer ab initio QM/MM molecular dynamics simulations. *The journal of physical chemistry B*, 114(26), 8817-8825.
58. Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S. S., & Taylor, P. (1986). Primary structure of *Torpedo californica* acetylcholinesterase deduced from its cDNA sequence. *Nature*, 319(6052), 407-409.
59. Soreq, H., Ben-Aziz, R., Prody, C. A., Seidman, S., Gnatt, A., Neville, L., & Lipidot-Lifson, Y. (1990). Molecular cloning and construction of the coding region for human acetylcholinesterase

- reveals a G+ C-rich attenuating structure. *Proceedings of the National Academy of Sciences*, 87(24), 9688-9692.
60. Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toket, L., & Silman, I. (1991). Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science*, 253(5022), 872-879
  61. Delfino, R. T., Ribeiro, T. S., & Figueroa-Villar, J. D. (2009). Organophosphorus compounds as chemical warfare agents: a review. *Journal of the Brazilian Chemical Society*, 20(3), 407-428.
  62. Sussman, J. L., Harel, M., Frolow, F., Varon, L., Toket, L., Futerman, A. H., & Silman, I. (1988). Purification and crystallization of a dimeric form of acetylcholinesterase from *Torpedo californica* subsequent to solubilization with phosphatidylinositol-specific phospholipase C. *Journal of molecular biology*, 203(3), 821-823.
  63. Sanson, B. (2009). Thèse de Doctorat, La dynamique structurale de l'acétylcholinestérase: étude réalisée par cristallographie aux rayons X et une méthode spectroscopique complémentaire. Université Joseph-Fourier-Grenoble I, France.
  64. Ripoll, D. R., Faerman, C. H., Axelsen, P. H., Silman, I., & Sussman, J. L. (1993). An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. *Proceedings of the National Academy of Sciences*, 90(11), 5128-5132.
  65. Hellström-Lindahl, E. (2000). Modulation of  $\beta$ -amyloid precursor protein processing and tau phosphorylation by acetylcholine receptors. *European journal of pharmacology*, 393(1-3), 255-263.
  66. Hock, C., Maddalena, A., Raschig, A., Müller-Spahn, F., Eschweiler, G., Huger, K., ... & Wienrich, M. (2003). Treatment with the selective muscarinic ml agonist talsaclidine decreases cerebrospinal fluid levels of A $\beta$ 42 in patients with Alzheimer's disease. *Amyloid*, 10(1), 1-6.
  67. Kojro, E., & Fahrenholz, F. (2005). The non-amyloidogenic pathway: structure and function of  $\alpha$ secretases. In *Alzheimer's Disease* (pp. 105-127).
  68. Inestrosa, N. C., Alvarez, A., Perez, C. A., Moreno, R. D., Vicente, M., Linker, C., ... & Garrido, J. (1996). Acetylcholinesterase accelerates assembly of amyloid- $\beta$ -peptides into Alzheimer's fibrils: possible role of the peripheral site of the enzyme. *Neuron*, 16(4), 881-891.
  69. Edelman, A. M., Blumenthal, D. K., and Krebs, E. G. (1987), Protein serine/threonine kinases. *Annual review of biochemistry*, 56(1), 567-613.
  70. Hunter, T. (1995). Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell*, 80(2), 225-236.
  71. Ventura, J. J., & Nebreda, Á. R. (2006). Protein kinases and phosphatases as therapeutic targets in cancer. *Clinical and Translational Oncology*, 8(3), 153-160.
  72. Taylor, S. S., & Kornev, A. P. (2011). Protein kinases: evolution of dynamic regulatory proteins. *Trends in biochemical sciences*, 36(2), 65-77.
  73. Nussinov, R., & Tsai, C. J. (2013). Allostery in disease and in drug discovery. *Cell*, 153(2), 293-305.

74. Maccioni, R. B., Otth, C., Concha, I. I., & Muñoz, J. P. (2001). The protein kinase Cdk5. *European Journal of Biochemistry*, 268(6), 1518-1527.
75. Morgan, D. O. (1997). Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annual review of cell and developmental biology*, 13(1), 261-291.
76. Lew, J., Huang, Q. Q., Qi, Z., Winkfein, R. J., Aebersold, R., Hunt, T., & Wang, J. H. (1994). A brain-specific activator of cyclin-dependent kinase 5. *Nature*, 371(6496), 423.
77. Hellmich, M. R., Pant, H. C., Wada, E., & Battey, J. F. (1992). Neuronal cdc2-like kinase: a cdc2-related protein kinase with predominantly neuronal expression. *Proceedings of the National Academy of Sciences*, 89(22), 10867-10871.
78. Kobayashi, S., Ishiguro, K., Omori, A., Takamatsu, M., Arioka, M., Imahori, K., & Uchida, T. (1993). A cdc2-related kinase PSSALRE/cdk5 is homologous with the 30 kDa subunit of tau protein kinase II, a proline-directed protein kinase associated with microtubule. *FEBS letters*, 335(2), 171-175.
79. Tsai, L. H., Delalle, I., Caviness Jr, V. S., Chae, T., & Harlow, E. (1994). p35 is a neural-specific regulatory subunit of cyclin-dependent kinase 5. *Nature*, 371(6496), 419.
80. Tarricone, C., Dhavan, R., Peng, J., Areces, L. B., Tsai, L. H., & Musacchio, A. (2001). Structure and regulation of the CDK5-p25nck5a complex. *Molecular cell*, 8(3), 657-669.
81. Paglini, G., Pigino, G., Kunda, P., Morfini, G., Maccioni, R., Quiroga, S., ... & Cáceres, A. (1998). Evidence for the participation of the neuron-specific CDK5 activator P35 during laminin-enhanced axonal growth. *Journal of Neuroscience*, 18(23), 9858-9869.
82. Tang, D., Yeung, J., Lee, K. Y., Matsushita, M., Matsui, H., Tomizawa, K., ... & Wang, J. H. (1995). An isoform of the neuronal cyclin-dependent kinase 5 (Cdk5) activator. *Journal of Biological Chemistry*, 270(45), 26897-26903.
83. Tang, D., Chun, A. C., Zhang, M., & Wang, J. H. (1997). Cyclin-dependent Kinase 5 (Cdk5) Activation Domain of Neuronal Cdk5 Activator EVIDENCE OF THE EXISTENCE OF CYCLIN FOLD IN NEURONAL Cdk5a ACTIVATOR. *Journal of Biological Chemistry*, 272(19), 12318-12327.
84. Lee, M. S., Kwon, Y. T., Li, M., Peng, J., Friedlander, R. M., & Tsai, L. H. (2000). Neurotoxicity induces cleavage of p35 to p25 by calpain. *Nature*, 405(6784), 360.
85. Kusakawa, G. I., Saito, T., Onuki, R., Ishiguro, K., Kishimoto, T., & Hisanaga, S. I. (2000). Calpain-dependent proteolytic cleavage of the p35 cyclin-dependent kinase 5 activator to p25. *Journal of Biological Chemistry*, 275(22), 17166-17172.
86. Tarricone, C., Dhavan, R., Peng, J., Areces, L. B., Tsai, L. H., & Musacchio, A. (2001). Structure and regulation of the CDK5-p25nck5a complex. *Molecular cell*, 8(3), 657-669.
87. Ahn, K. W., Joo, Y., Choi, Y., Kim, M., Lee, S. H., Cha, S. H., ... & Kim, H. S. (2008). Swedish amyloid precursor protein mutation increases cell cycle-related proteins in vitro and in vivo. *Journal of neuroscience research*, 86(11), 2476-2487.

88. Mandell, J. W., & Banker, G. A. (1996). A spatial gradient of tau protein phosphorylation in nascent axons. *Journal of Neuroscience*, 16(18), 5727-5740
89. Alvarez, A., Toro, R., Cáceres, A., & Maccioni, R. B. (1999). Inhibition of tau phosphorylating protein kinase cdk5 prevents  $\beta$ -amyloid-induced neuronal death. *FEBS letters*, 459(3), 421-426.
90. Lopes, J. P., Oliveira, C. R., & Agostinho, P. (2010). Neurodegeneration in an A $\beta$ -induced model of Alzheimer's disease: the role of Cdk5. *Aging cell*, 9(1), 64-77.
91. Lopes, J. P., Oliveira, C. R., & Agostinho, P. (2007). Role of cyclin-dependent kinase 5 in the neurodegenerative process triggered by amyloid-Beta and prion peptides: implications for Alzheimer's disease and prion-related encephalopathies. *Cellular and molecular neurobiology*, 27(7), 943-957.
92. Ferreiro, E., Resende, R., Costa, R., Oliveira, C. R., & Pereira, C. M. (2006). An endoplasmic-reticulum-specific apoptotic pathway is involved in prion and amyloid-beta peptides neurotoxicity. *Neurobiology of disease*, 23(3), 669-678.
93. Lee, M. S., & Tsai, L. H. (2003). Cdk5: one of the links between senile plaques and neurofibrillary tangles?. *Journal of Alzheimer's Disease*, 5(2), 127-137.
94. Eckermann, K., Mocanu, M. M., Khlistunova, I., Biernat, J., Nissen, A., Hofmann, A., & Zhou, L. (2007). The  $\beta$ -propensity of Tau determines aggregation and synaptic loss in inducible mouse models of tauopathy. *Journal of Biological Chemistry*, 282(43), 31755-

## **CHAPITER II**

# **Conception et découverte de médicaments assistées par ordinateur**

### 1.Introduction

#### 1.1. Révolution chronologique de médicaments

Si on peut considérer que les maladies et les « essais thérapeutiques » remontent à l'existence de l'homme sur Terre, de la préhistoire jusqu'au début du XXe siècle, la grande majorité des médicaments provenaient de l'identification de principes actifs extraits de substances naturelles. Dans les premiers temps, il n'y avait aucune possibilité de comprendre la nature de la maladie<sup>1</sup>. Cependant, le début du XIXe siècle marque un tournant historique dans la recherche de nouveaux médicaments grâce à l'isolement de principes actifs de substances naturelles précédemment utilisées dans la médecine traditionnelle. En 1803, Friedrich Wilhelm isole la morphine à partir du pavot, qui a été commercialisée par les laboratoires Merck à partir des années 1827<sup>2</sup>. Cependant, la nécessité de développer d'autres sources de médicaments est rapidement apparue et a permis de nombreuses avancées tout au long du XXe siècle et jusqu'à aujourd'hui.

Plusieurs grandes disciplines ont permis la révolution de la recherche de nouveaux médicaments et ont abouti aux processus de R&D (Recherche et développement) mis en œuvre actuellement (Schéma 1)<sup>3</sup>. La pharmacologie, les techniques de biologie moléculaire et la robotique ont permis le développement de techniques de criblage à haut débit<sup>4,5</sup> utilisées lors de la phase préliminaire de R&D, qui permet l'identification rapide de composés prometteurs ayant des activités potentielles sur des cibles d'intérêt thérapeutiques. Le faible taux d'identification de *hits* par rapport aux coûts de la méthode a conduit les compagnies pharmaceutiques au développement de nouvelles approches telles que le criblage virtuel basé sur la bioinformatique, couplée à la pharmacologie et à la chimie, permet la conception de nouvelles molécules « *drug-like* » à l'aide des outils informatiques<sup>6,7</sup>.

#### 1.2. Conception de médicaments assistée par ordinateur

Le domaine de la conception de médicaments « *drug design* » peut être défini par un ensemble de processus inventifs capables de trouver de nouveaux médicaments basés sur les connaissances de cibles biologiques. La conception de médicaments repose fréquemment sur les techniques de modélisation sur l'ordinateur (*Computer-Aided Drug Design* ou **CADD**), en se basant sur des outils informatiques « *in silico* »<sup>8-12</sup>.

## Conception et découverte de médicaments assistée par ordinateur

En effet, l'outil informatique aide la conception de médicaments à des étapes spécifiques du processus de la conception<sup>13,14</sup> :

- Dans l'identification des composés potentiellement thérapeutiques, en utilisant le criblage virtuel, aussi dans la conception basée sur la structure de la cible et sur les propriétés des ligands.
- Dans le processus d'optimisation de l'affinité et de la sélectivité des molécules potentielles vers les têtes de série « lead ».
- Dans le processus d'optimisation de lead par rapport aux propriétés pharmacologiques recherchées tout en maintenant une bonne affinité de cette molécule<sup>15</sup>.



**Schéma 1.** Les différentes étapes de développement du médicament.

Un certain nombre de médicaments mis sur le marché proviennent d'une conception rationnelle basée sur des stratégies de criblage virtuel par exemple:

- Gefitinib (Iressa®) : médicament anticancéreux, au cours du processus de R&D, une approche pharmacophorique 3D suivie d'une approche pharmacophorique 2D a permis l'identification d'un *hit*, dont l'optimisation a donné la naissance au Gefitinib<sup>16</sup>.
- Pazopanib (Votrient®) : médicament indiqué pour le traitement du cancer du rein. Étude de Relations Structure-Activité (RSA) à partir de modèle par homologie de VEGFR2<sup>17</sup>.
- Boceprevir (Victrelis®) : indiqué dans le traitement de l'hépatite C chronique (cible : HCV protease). Étude RSA à partir de la structure cristallisée<sup>18</sup>.



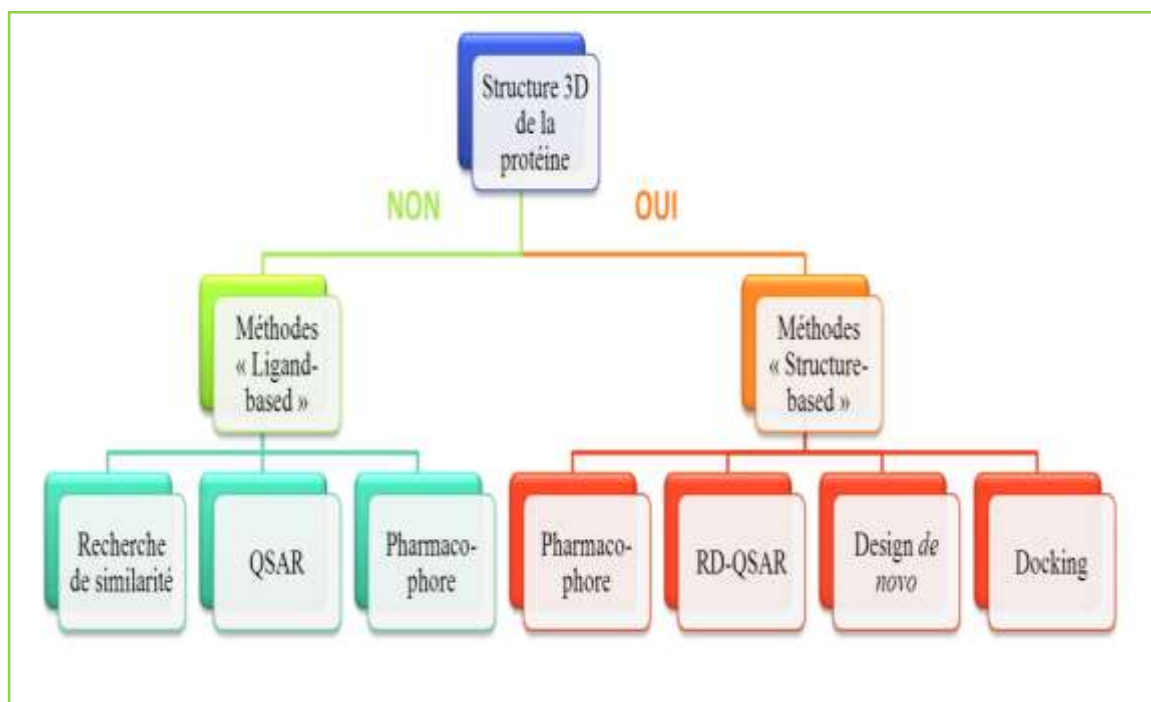
### 2. Criblage virtuel «in silico »

Le terme criblage virtuel regroupe un ensemble de techniques computationnelles ayant pour l'objectif d'exploration de bases de composés à la recherche de molécules d'intérêt thérapeutique<sup>19</sup>. Ce qui lui permet de prendre une importance croissante dans les processus de conception de médicaments, largement appuyé par les progrès en informatique et en bioinformatique<sup>20</sup>. Les méthodes de criblages *in silico* ont été développées pour diminuer le coût et le temps considérable nécessaire au processus de R&D en permettant de réduire le nombre considérable de molécules à tester expérimentalement et pour guider la sélection des composés les plus prometteurs, que ce soit lors des phases d'identification des *hits* ou d'optimisation des leads<sup>21</sup>.

Les méthodes de criblage consistant à prédire l'affinité d'un très grand nombre de molécules (collectées en chimiothèques) pour le site actif ciblé afin d'aboutir à une liste très réduite de molécules présentant l'activité souhaitée. La diversité moléculaire des chimiothèques criblées expérimentalement ou virtuellement est un critère important pour mieux couvrir l'espace chimique et donc pour augmenter les chances d'identifier de nouvelles touches « *hits* ». Les molécules doivent également avoir les propriétés physico-chimiques comme la biodisponibilité et les propriétés ADME (Absorption, Distribution, Métabolisme et Excrétion)<sup>8</sup> caractérisant un médicament. Pour cela, des règles élémentaires permettent de filtrer rapidement des chimiothèques et éliminer toutes les molécules qui ne répondant pas à ces règles ; les plus connues sont celles de Lipinski<sup>22</sup>, de Veber<sup>23</sup> et de triangle d'or<sup>24</sup>.

Le criblage virtuel nécessite des informations sur les propriétés physico-chimiques et/ou géométriques des composés actifs et sur la structure 3D de la cible. Selon la disponibilité ou non de ces informations, différentes méthodes de criblage virtuel peuvent être utilisées. Les méthodes basées sur les ligands (*Ligand-Based Drug Design LBDD*) utilisent les ligands connus comme guides pour définir les propriétés responsables de leur affinité<sup>25</sup>. Les méthodes basées sur la structure (*Structure-Based Drug Design SBDD*) sont utilisées lorsque les données sur la structure 3D de la cible biologique sont disponibles. Elles permettent d'évaluer le

potentiel d'interaction entre les composés criblés et le site d'interaction de la cible<sup>26</sup>. Au sein de ces deux grandes familles, différentes méthodes ont été développées (Schéma 2).



**Schéma 2.** Classification des méthodes de criblage virtuel (avec QSAR : relation quantitative structure-activité ; RD-QSAR : relation quantitative structure-activité dépendant du récepteur).

### 2.1. Criblage virtuel basé sur le ligand « Ligand-Based »

Les méthodes de criblage « ligand-based » reposent sur la connaissance préalable de ligands ayant une activité sur la cible thérapeutique connue. Il sera ainsi possible d'utiliser ces ligands comme une première base de « *hits* » afin d'identifier d'autres composés similaires, présentant des caractéristiques d'activité commune<sup>27,28</sup>. L'hypothèse que les molécules similaires vont avoir tendance à présenter des profils d'activités très proches à similaires est le principe de base de toutes les méthodes « ligand-based ». Différents types de descripteurs moléculaires pourront être calculés pour quantifier la similitude entre les composés. Ces méthodes présentent l'avantage d'être un protocole de découverte de médicament viable, même en l'absence de structure expérimentale tridimensionnelle (3D) du récepteur ciblé<sup>10,29</sup>. Suivant le nombre de ligands connus de la cible thérapeutique, plusieurs méthodes peuvent être employées : la recherche de similarité, le criblage de pharmacophores et les approches QSAR. Ces deux dernières méthodes sont les outils les plus importants et largement utilisés dans la conception de médicaments à base de ligands<sup>30</sup>.

Les modèles réussis à base de ligand permettent l'accès à de nouveaux ligands « *scaffold hopping* »<sup>31-33</sup>, le « *scaffold hopping* » fait la référence à la recherche assistée par ordinateur des composés ayant une activité biologique similaire, mais contenant des structures de noyau différent<sup>34</sup>. L'objectif fondamental de l'approche « *scaffold hopping* » est de découvrir des composés structurellement nouveaux à partir des composés actifs connus en modifiant la structure centrale de la molécule<sup>34,35</sup>. Les nouveaux composés offrent l'occasion de modifier les affinités et les sélectivités des récepteurs (profils de toxicité), ainsi que d'améliorer les profils pharmacocinétiques (propétés ADME) et d'obtenir des analogues brevetables<sup>35</sup>.

### 2.1.1. Recherche de similarité

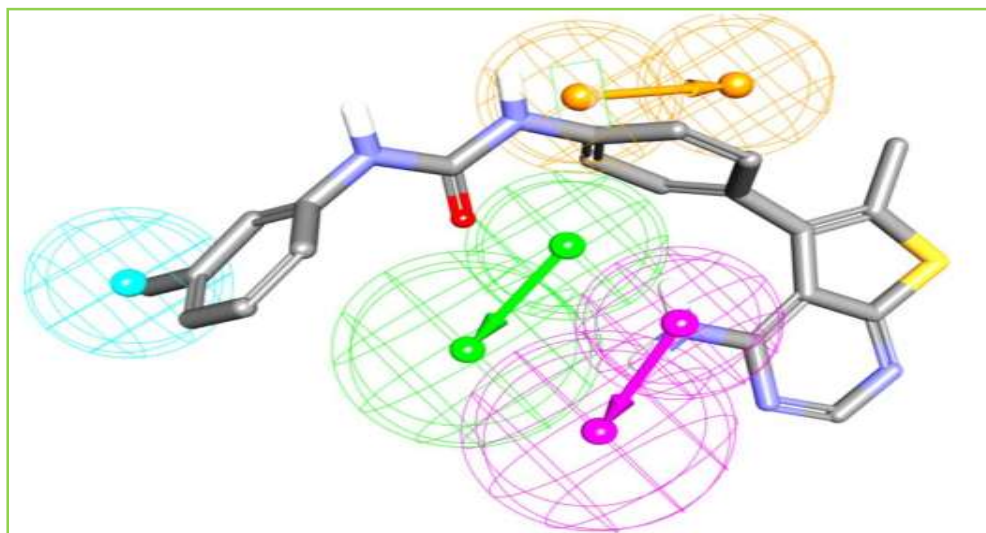
La recherche de similarité est la méthode à employer lorsque très peu de ligands ont été rapportés pour la cible biologique choisie. Cette méthode repose sur l'utilisation de descripteurs et de métriques de similarité permettant de comparer des molécules à cribler à un ou plusieurs ligands de référence pour prédire leur profil d'activité<sup>36</sup>. Les descripteurs de similarité permettent de définir au sein d'une base de données quels ligands sont les plus ressemblants aux ligands actifs connus. Ces descripteurs peuvent être obtenus mathématiquement ou expérimentalement pour transformer l'information chimique en un nombre utile et sont classés suivant leur dimension, qui représentent des caractéristiques structurales clés de composés. L'obtention des vecteurs de descripteurs pour chaque molécule de la chimiothèque, ainsi que les molécules actives de référence permettent ensuite de comparer ces molécules par métriques de similarité<sup>37</sup>. Ils sont traduits en coefficients dont la valeur est comprise dans l'intervalle [0;1], par exemple le cas du coefficient de Tanimoto «Tc» qui est le plus populaire et le plus couramment utilisé. Les valeurs de coefficients Tc les plus élevées sont associées aux molécules les plus similaires<sup>38</sup>.

### 2.1.2. Modèles pharmacophoriques « Ligand-Based »

Le concept de pharmacophore a été introduit en 1909 par Ehrlich<sup>39</sup>, qui a défini le pharmacophore comme des groupes ou des fonctions chimiques sont responsables de l'effet biologique d'une molécule. La définition officielle donnée par l'IUPAC (International Union of Pure and Applied Chemistry) indique que le pharmacophore possède l'ensemble des propriétés stériques et électroniques d'une molécule nécessaires pour assurer l'établissement d'interactions supramoléculaires optimales avec une cible biologique spécifique pour engendrer ou bloquer une réponse biologique<sup>40</sup>. Suivant cette définition les molécules qui partagent le

même pharmacophore pour une cible donnée devraient donc se lier de manière identique à ce récepteur et présenter des profils d'activité similaires. Les interactions sont définies comme des points pharmacophoriques : les donneurs et accepteurs de liaisons hydrogène, les groupements cationiques et anioniques réalisant des interactions ioniques, les groupements aromatiques conduisant à du  $\pi$ -stacking et les groupements hydrophobes donnant des interactions hydrophobes<sup>41</sup> (Figure 1).

Un modèle pharmacophorique est dit basé sur le ligand « ligand-based » lorsque on fait la superposition des molécules actives et en extrayant les caractéristiques physico-chimiques communes essentielles à leur activité biologique, sans connaître ou sans prendre en compte la structure du récepteur. Une fois le pharmacophore est généré, il peut ensuite être utilisé dans le processus de R&D pour cribler une chimiothèque à la recherche de nouveaux *hits* « pharmacophore searching »<sup>42</sup>.



**Fig.1.** Pharmacophore généré à l'aide du logiciel discovery studio<sup>43</sup>. Les différents types de points pharmacophoriques sont illustrés par des sphères colorées : le bleu indique les points hydrophobes; l'orange indique les groupements aromatiques; le vert pour les liaisons hydrogène accepteurs (HBA) et le violet pour les liaisons hydrogène donneurs (HBD).

### 2.1.2.1. Élucidation du pharmacophore :

L'élucidation d'un pharmacophore est un processus complexe divisé en plusieurs étapes<sup>44</sup>.

1. **Sélection des ligands de référence** : Les ligands de référence sont des composés actifs dont l'activité sur la cible biologique étudiée est renseignée et comparable pour tous.

**2. Recherche conformationnelle** : Les ligands utilisés pour créer le pharmacophore doivent être dans leurs conformations bioactives, c'est-à-dire la conformation avec laquelle ils se lient au récepteur. Cependant, lorsque celle-ci n'a pas été identifiée, une recherche conformationnelle doit être menée pour pouvoir inclure toutes les conformations à basse énergie des ligands dans l'étude.

**3. Détermination et représentation des points pharmacophoriques de chaque ligand** : Si les points pharmacophoriques d'un modèle sont basés sur les atomes ou sur les groupes topologiques (par exemple un oxygène ou un carbonyle), seules les molécules possédant exactement ces atomes ou ces groupes topologiques peuvent être identifiées en tant que *hits* lors d'un criblage. A l'opposé, si les points pharmacophoriques décrivent des propriétés chimiques, le nombre de molécules pouvant satisfaire les critères du pharmacophore augmente puisque différents éléments peuvent représenter une même fonction chimique (par exemple un azote et un oxygène peuvent être tous les deux accepteurs de liaisons hydrogène).

**4. Détermination de modèles de pharmacophore** : Il faut déterminer les points pharmacophoriques communs entre les différents ligands. Pour cela, les ligands doivent être alignés (superposé) selon leurs atomes, fragments ou caractéristiques chimiques. Il est possible de représenter les molécules sous forme de graphes, permettant la comparaison des structures moléculaires pour l'identification la sous-structure commune maximale (Maximum Common Substructure MCS) <sup>45,46</sup>. Pour déterminer quels sont les meilleurs modèles pharmacophoriques parmi tous ceux générés, un score est attribué à chaque modèle, ce qui permet de les classer <sup>47-50</sup>.

### 2.1.3. Relation structure-activité quantitative « ligand-based »

En raison de l'absence de structures expérimentales, les ligands connus qui se lient à la cible du médicament sont étudiés pour comprendre les propriétés structurales et physico-chimiques des ligands qui sont en corrélation avec l'activité pharmacologique souhaitée <sup>51</sup>. Pour cela, les méthodes QSAR « Quantitative Structure-Activity Relationship » ont pour objectif d'établir des relations mathématiques à l'aide de méthodes d'analyse de données, reliant des propriétés moléculaires appelées descripteurs à l'activité biologique mesurée dans les mêmes conditions expérimentales pour une série de composés, ce qui permet d'identifier et d'évaluer l'impact des propriétés influençant l'activité biologique <sup>52,53</sup>.

Lors d'une étude QSAR, il faut calculer à partir de la structure moléculaire tous les descripteurs moléculaires possibles qui portent 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. 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 utilisées dans la prédiction des activités biologiques de nouvelles molécules obtenues lors de phases de criblage ou d'optimisation des candidats-médicaments, ainsi que dans la conception de nouvelles structures<sup>54-56</sup>.

### 2.1.3.1. Principe

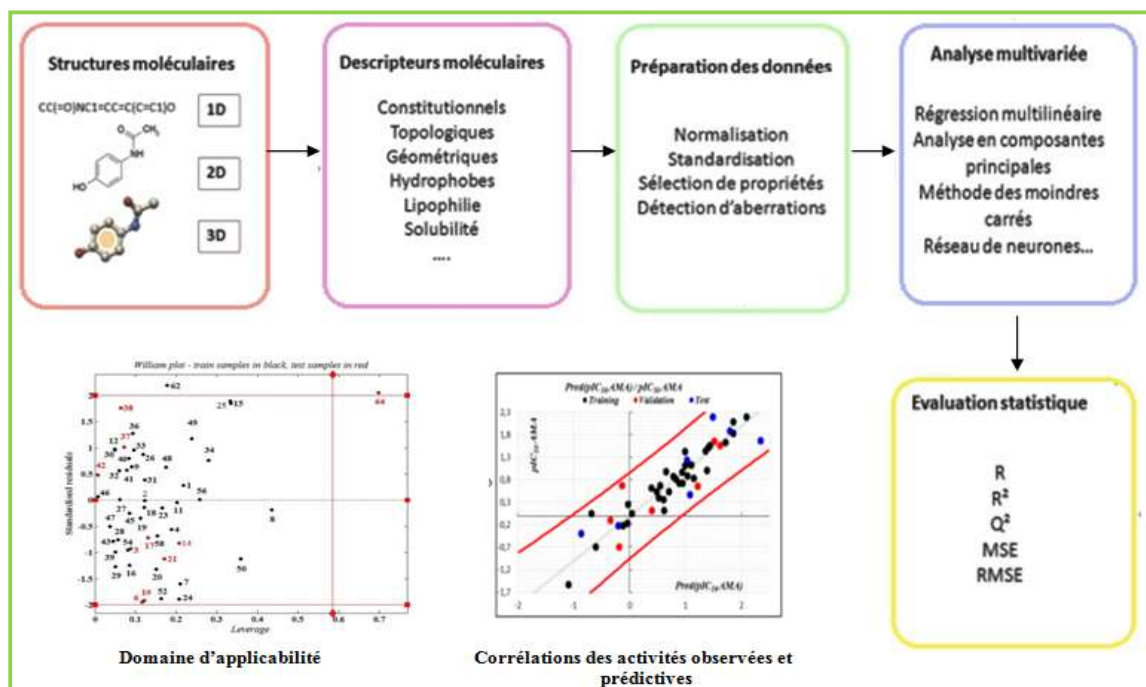
En pratique, le développement d'un modèle commence par la collecte de données expérimentales fiables et en nombre important, ensuite il est nécessaire de générer un nombre de descripteurs, caractérisant les structures moléculaires, utilisées dans le développement du modèle QSAR<sup>57</sup>. Une fois le modèle construit, l'influence des composés du jeu d'entraînement sur le modèle (robustesse du modèle) est estimée par des méthodes de validation interne. Pour estimer le pouvoir prédictif du modèle, il est nécessaire de disposer de données expérimentales supplémentaires, jeu de données de validation externe afin de déterminer la capacité du modèle à prédire ces valeurs<sup>58,59</sup>. Le schéma 3 montre les étapes générales mises en œuvre dans la modélisation QSAR. Des règles précises ont également été mises en place récemment par l'OCDE (Organisation de Coopération et de Développement Économique) pour la validation des modèles QSAR<sup>60,61</sup>. La validation des modèles repose alors sur cinq grands principes<sup>62</sup>. L'évaluation de chacun des cinq principes est une condition importante afin de proposer des modèles applicables dans le plan expérimental :

- **Un effet défini** : la base de données (l'activité/propriété ciblée) doit être fiable et définie avec un protocole expérimental identifié.
- **Un algorithme non ambigu** : l'algorithme sur lequel repose le modèle doit garantir la transparence du calcul. Il convient notamment d'être prudent lorsque des méthodes non transparentes et difficilement reproductibles ont été utilisées pour élaborer le modèle QSAR.
- **Un domaine d'applicabilité défini** : le domaine d'applicabilité et les limitations du modèle doivent être décrits pour permettre l'évaluation de l'espace chimique dans lequel on peut faire les prévisions avec confiance. Les prévisions issues d'un modèle ne contenant aucune information sur le domaine d'applicabilité ne peuvent pas être acceptées.

## Conception et découverte de médicaments assistée par ordinateur

- **Des mesures appropriées du degré d'ajustement, de la robustesse et de la prévisibilité** : ce principe traduit la nécessité d'une validation statistique du modèle. Des statistiques relatives à la validation interne (degré d'ajustement et robustesse) et la validation externe (prévisibilité) doivent être disponibles. Par exemple, les statistiques portant sur le modèle de régression peuvent être consignées en utilisant le coefficient de corrélation, le coefficient de corrélation de la validation croisée, l'erreur quadratique moyenne du modèle.

- **Si possible, une interprétation des mécanismes sous-jacents** : Il est reconnu qu'il n'est pas toujours possible, d'un point de vue scientifique, de fournir une interprétation mécaniste d'un modèle QSAR. L'intention du principe 5 n'est pas de rejeter les modèles qui n'ont pas de base mécaniste apparente, mais de faire en sorte que soit envisagée la possibilité d'une association mécaniste entre les descripteurs utilisés dans un modèle et le résultat final.



**Schéma 3.** Les différentes étapes d'une étude QSAR<sup>55</sup>.

### 2.1.3.2. Outils statistiques pour le développement et la validation de modèles

Le succès de tout modèle QSAR dépend grandement du choix des descripteurs moléculaires et de la capacité de générer la relation mathématique appropriée entre les descripteurs et l'activité biologique d'intérêt. Depuis les premiers jours de QSAR, il était clair que la définition des descripteurs moléculaires est la partie cruciale de la méthode<sup>63, 64</sup>. Les descripteurs moléculaires sont généralement classés en descripteurs physico-chimiques, topologiques électroniques et quantiques. Les logiciels développés récemment permettent maintenant de générer un grand nombre de descripteurs moléculaires pouvant être utilisés pour

les méthodes QSAR<sup>65, 66</sup>. Lorsqu'une grande quantité de descripteurs est introduite, certains d'entre eux peuvent contenir des informations redondantes, entraînant un problème de colinéarité. La méthode Analyse en composantes principales est parmi les méthodes statistiques qui nous aident à sélectionner les descripteurs en évitant ce problème.

Pour élaborer un modèle QSAR, nous avons besoin d'une méthode d'analyse de données, cette méthode permet de quantifier la relation qui existe entre l'activité et la structure (descripteurs). Il existe plusieurs méthodes pour construire un modèle et analyser les données statistiques de ce dernier, certaines sont linéaires telles que la régression linéaire multiple (MLR)<sup>67</sup>, la régression des moindres carrés (PLS)<sup>68</sup> et non linéaires comme les réseaux de neurones artificiels (ANN)<sup>69</sup>. Il existe un grand nombre d'exemples de la littérature suggérant que les modèles ANN peuvent offrir des performances prédictives nettement meilleures que les approches statistiques traditionnelles telles que la MLR dans la modélisation QSAR<sup>70-72</sup>.

Une fois le modèle QSAR développé, doit être interprétés en analysant tous les paramètres statistiques de ce modèle, sa qualité doit être aussi étudiée et vérifiée par ce que l'on appelle la validation. Sa robustesse c'est-à-dire l'influence des composés de la série d'apprentissage sur le modèle est estimée par des méthodes de validation interne. Afin d'estimer son pouvoir prédictif, une validation externe est appliquée sur l'ensemble des données supplémentaires. Cette partie est bien détaillé dans les études réalisées du notre groupe<sup>73-77</sup>.

### 2.2. Criblage virtuel basé sur la structure « structure-based »

Conception de médicaments basée sur la structure utilise les interactions perçues au sein des complexes ligands-récepteurs pour générer des modèles de liaison et les utiliser dans la découverte et l'optimisation de nouveaux ligands bioactifs. Le criblage virtuel « structure-based » est considéré comme un équivalent *in silico* d'un test expérimental étudiant la liaison ligand-cible biomoléculaire<sup>77</sup>. Cependant, il se repose essentiellement sur la disponibilité de la structure 3D de la cible thérapeutique qui est obtenue par des méthodes expérimentales telles que la cristallographie par rayons X et la spectroscopie par Résonance Magnétique Nucléaire (RMN) dans des bases de données (tel que PDB: Protein Data Bank)<sup>78</sup>. Par contre, pour toutes les structures non déterminées expérimentalement, des méthodes de prédiction de la structure 3D par homologie de séquence peuvent être mises en œuvre<sup>79-81</sup>, basée sur la structure expérimentale d'un protéine proche (même classe voire même famille).



## Conception et découverte de médicaments assistée par ordinateur

Comme leur nom l'indique, ces méthodes utilisent la structure de la cible pour découvrir de nouveaux composés actifs. Pour cela, plusieurs approches peuvent être employées: le docking moléculaire, les approches pharmacophoriques, la méthode QSAR (RD-QSAR) et le design de novo<sup>79,82,83</sup>. Chacune de ces méthodes requiert la détermination du site de liaison sur la cible thérapeutique. Pour cela, deux solutions sont possibles : obtenir une structure 3D de la protéine co-cristallisée avec un ligand, ou utiliser des outils de prédiction<sup>84</sup> (Schéma 4). Cela nous donne l'information nécessaire pour identifier les groupements d'interactions du ligand dans la poche de liaison du récepteur.

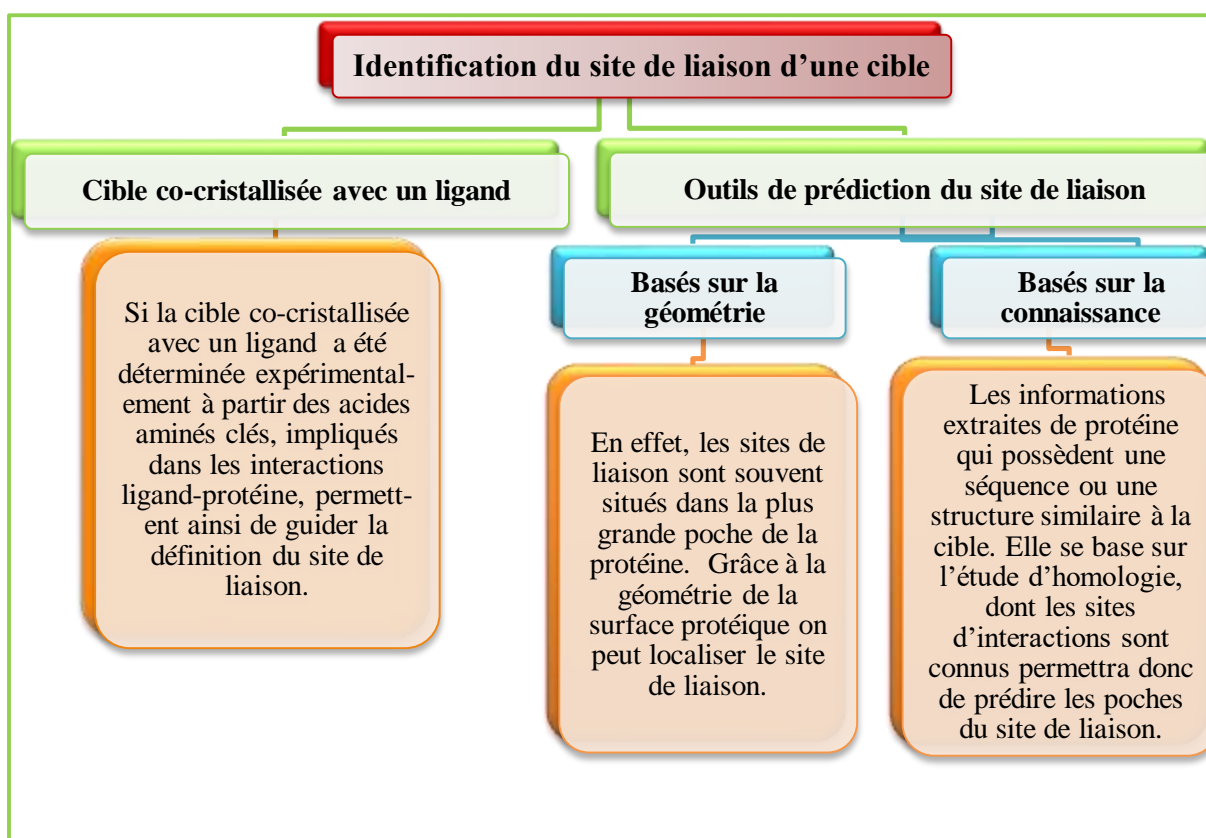


Schéma 4. Identification d'un site de liaison sur une cible thérapeutique.

### 2.2.1. Modèles pharmacophoriques « structure-based »

Les pharmacophores 3D décrivent l'arrangement spatial des propriétés chimiques nécessaires pour l'activité biologique à partir d'un ensemble de ligands actifs de référence<sup>85</sup>. Cependant, lorsqu'elle est disponible, prendre en compte la structure 3D du site de liaison permet d'obtenir de nouveaux modèles pharmacophoriques incorporant des informations cruciales ce que rendent les méthodes pharmacophoriques plus puissantes et inexploitées dans les modèles basés sur les ligands. Les pharmacophores 3D sont obtenus également à partir de la structure 3D de la protéine seule « receptor-based approach » où

Co-cristallisée avec un ligand « complexe-based approach »<sup>44</sup>. Dans notre étude on a construit un modèle pharmacophorique, en se basant sur la méthode « receptor-based approach », pour cribler la chimiothèque de NCI à la recherche de molécules potentiellement actives.

### 2.2.1.1. Approche basée sur le récepteur

L'approche basée sur le récepteur nécessite comme point de départ la structure 3D du site de liaison de la cible d'intérêt pour décrire ses propriétés et leurs relations spatiales, cette structure 3D ne contient pas un ligand co-cristallisé dans le site de liaison. Pour générer un modèle pharmacophorique, la démarche est la suivante <sup>44</sup> :

➤ Une fois le site de liaison identifié, une image négative du récepteur est créée qui comporte généralement un grand nombre de points pharmacophoriques. (Cette carte regroupe les propriétés complémentaires de celles identifiées dans le site de liaison. Par exemple, une propriété de type donneur de liaison a été identifiée dans le site de liaison sera représentée par une propriété de type accepteur de liaison sur la carte d'interaction.).

➤ Les points pharmacophoriques de même type et spatialement proches peuvent être regroupés et représentés par un point unique, ou les points pharmacophoriques d'intérêt peuvent être sélectionnés manuellement. Lorsque des ligands de la cible sont connus, ils peuvent également être utilisés pour guider la sélection des points pharmacophoriques <sup>86,87</sup>.

➤ Les pharmacophores générés, comportant chacun un plus petit nombre de points pharmacophoriques, afin de représenter l'ensemble de leurs combinaisons potentielles. Ces modèles sont ensuite testés sur un ensemble de composés actifs de référence et l'évaluation de leurs performances permet de choisir le modèle de pharmacophore à utiliser pour cribler la chimiothèque d'intérêt <sup>86,87</sup>.

➤ Enfin, les pharmacophores peuvent être complétés et raffinés par l'ajout de volumes d'exclusion (les contraintes stériques), qui représentent les régions stériquement interdites aux molécules à l'intérieur de la poche de la liaison. Les volumes d'exclusion peuvent être définis de manière explicite à partir des données structurales du récepteur ou positionnés manuellement par l'utilisateur.

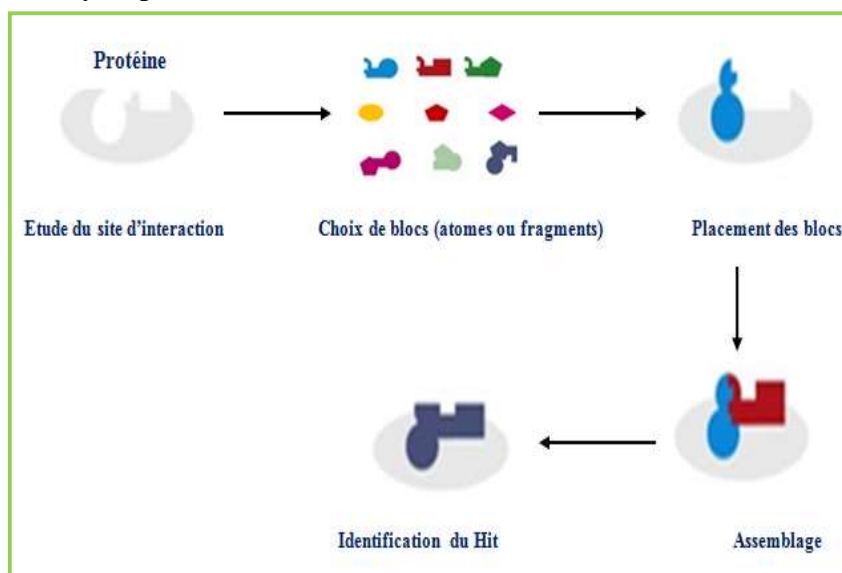
### 2.2.1.2. Approche basée sur le complexe ligand-récepteur

Lorsque des données de co-cristallisation protéine-ligand sont disponibles pour la cible étudiée, l'analyse des complexes ligand-récepteur permet d'obtenir des informations précises sur le mode d'interactions s'établissant entre le ligand et son récepteur <sup>88-90</sup>. Un modèle pharmacophorique unique peut être directement généré avec la structure de complexe.

Cependant, un seul modèle pharmacophorique généré avec une seule structure n'est pas assez fiable, il est donc recommandé d'utiliser les informations d'un ensemble de complexes différents dans le cas où plusieurs co-cristallisations de la cible sont disponibles. À partir de l'alignement de ces différents complexes, les groupes fonctionnels qui se retrouvent bien alignés permettent l'élaboration d'un modèle pharmacophorique commun à toutes ces structures<sup>88-91</sup>.

### 2.2.2. Conception de novo « de novo design »

Les méthodes de construction de novo design se basent sur la connaissance du site actif de la cible étudiée pour construire de nouveaux composés bioactifs qui seront spécifiquement adaptés à cette cible<sup>92</sup>. La première étape consiste à extraire toutes les informations concernant les interactions ligand-récepteur pour déterminer les sites d'interactions nécessaires à l'assemblage de divers blocs dans le site d'interaction. Ces blocs peuvent être soit un atome, soit un fragment de molécule. La mise en œuvre des approches de novo nécessite, dans un premier temps, l'identification des points d'interaction potentiels du site de liaison de la cible. Les interactions recherchées sont généralement de type de liaisons hydrogène, interactions électrostatiques et hydrophobes<sup>92,93</sup>.



**Fig.2.** Représentation schématique des étapes principales d'une conception de novo.

L'utilisation des atomes comme blocs permettent d'explorer une plus de diversité structurale et conduit à un nombre très important de solutions possibles ce qui rend l'extraction des composés les plus prometteurs très compliqués dont il peut être difficile d'extraire les « hits »<sup>92,93</sup>. En revanche, l'approche basée sur les fragments permet de réduire la taille de l'espace

## Conception et découverte de médicaments assistée par ordinateur

chimique à étudier. En conséquence, les approches basées sur les fragments et les algorithmes stochastiques (ces algorithmes permettant une optimisation locale à chaque étape de la construction des molécules dans le site de liaison de la cible) sont les plus utilisées et les fragments obtenus sont ensuite assemblés pour créer de nouveaux composés (Schéma 5).

Une solution de conception de novo doit être efficace en trois principaux : (i) l'assemblage des différents fragments moléculaires se fait par croissance « fragment-growth » ou bien par liaison « fragment-linking » ; (ii) l'échantillonnage de l'espace chimique accessible repose sur deux approches, la première appelée « depth-first search » (ou recherche en profondeur d'abord) ne conserve qu'une seule solution parmi toutes celles possibles à chaque niveau du processus de construction des ligands. À l'inverse, l'approche « breadth-first search » (ou recherche en largeur d'abord) permet à chaque niveau à toutes les solutions possibles de passer au niveau suivant jusqu'à la fin de la construction ; (iii) l'évaluation de la qualité des assemblages <sup>92</sup>.

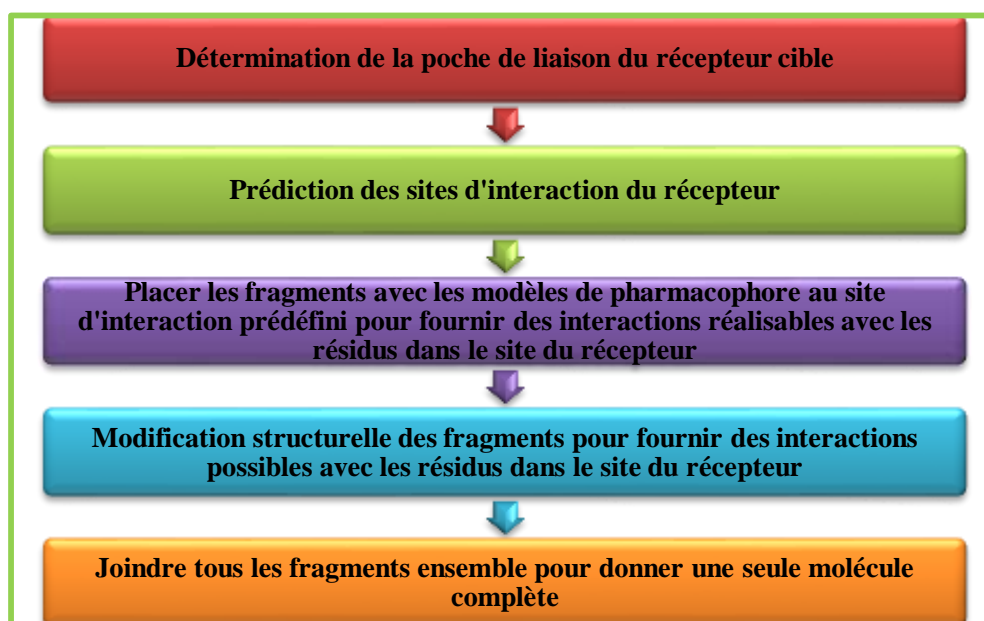


Schéma 5. Étapes de la méthodologie de conception de médicaments de novo

### 2.2.3. Relation quantitative structure-activité dépendant du récepteur (RD-QSAR)

Les interactions médicament-récepteur revêtent une importance fondamentale pour déterminer les propriétés du médicament représentées par son efficacité ou son activité. Toutefois, en raison de la complexité des systèmes médicament-récepteur, la méthode QSAR indépendante du récepteur (RI) a ignoré cet effet. Cette approche décrivant indirectement les

récepteurs hypothétiques par la série de médicaments<sup>95</sup>. L'incertitude liée à cette approche ne peut être éliminée que par l'utilisation de la structure 3D du récepteur de ligand. L'augmentation constante des données de structure et de la puissance de calcul disponibles a amélioré la capacité de modélisation en chimio-informatique et de nouvelles méthodes QSAR qui dépendent du récepteur (RD) ont été publiées.

Les méthodes RD-QSAR sont utilisées pour rassembler les énergies de liaison et d'interaction, en tant que descripteurs, de l'interaction entre le ligand et le récepteur. L'utilisation de QSAR dépendante du récepteur (RD) s'ajoute au modèle QSAR par l'inclusion des interactions ligand – récepteur<sup>95-97</sup>.

Le principe des méthodes 3D-QSAR dépendantes du récepteur consiste à établir une corrélation entre l'activité biologique représentée par l'énergie libre de liaison et des interactions ligand-récepteur déduites de la structure 3D d'un complexe correspondant<sup>97</sup>. Le formalisme des méthodes RD 4D-QSAR a été proposé en 2003 par Pan *et al.*<sup>98</sup>. L'approche proposée repose sur l'utilisation de simulations de dynamique moléculaire pour la modélisation des complexes récepteur-ligand<sup>98</sup>. RD 4D-QSAR est caractérisé par une grille d'occupation permet de détecter les caractéristiques physico-chimiques favorables ou défavorables à l'interaction des ligands. L'utilisation de simulations de dynamique moléculaire requiert la définition d'un référentiel stable pour le positionnement de la grille d'occupation. L'utilisateur doit donc définir plusieurs points d'interaction fixes entre le récepteur et chacun des ligands, sur la base de données expérimentales obtenues préalablement<sup>96</sup>.

### 2.2.4. Docking

Le docking est une technique informatique qui permet de prédire les interactions probables entre des ligands (substrat, activateur ou inhibiteur) et des acides aminés composant la structure d'une protéine (Figure 3), en se basant pour cela, sur une simulation de la conformation et de l'orientation du ligand qu'il adopte lors de sa liaison au récepteur<sup>99</sup>. À cet effet, les logiciels de docking combinent l'utilisation d'un algorithme de recherche conformationnelle qui permettant de générer des modes de liaisons putatives (les poses) du ligand dans le récepteur, et d'une fonction de score qui peut être considérées comme un indice d'affinité molécule-récepteur, employée pour classer les différentes poses selon un score qui prédit d'affinité<sup>100</sup>.

Les méthodes de docking ont deux objectifs principaux : (i) identifier les molécules qui sont des ligands véritables du récepteur parmi toutes celles étudiées et (ii) déterminer les modes de liaison adoptée par les ligands dans le récepteur<sup>101,102</sup>.

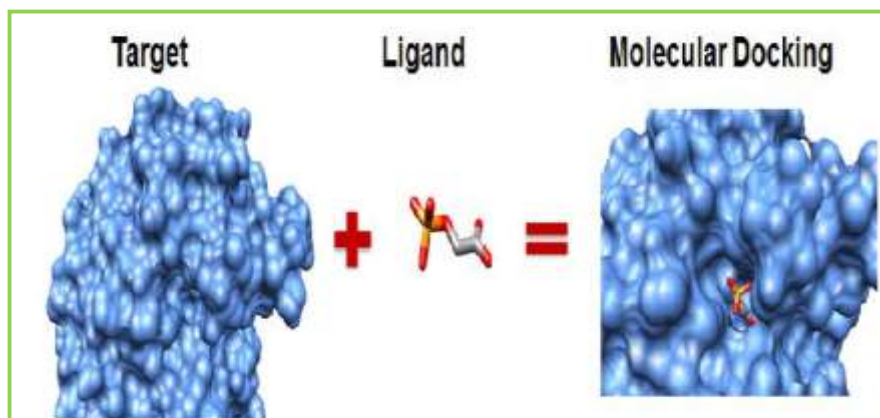


Fig. 3. La méthode de docking.

### 2.2.4.1. Classement des méthodes du docking

Les méthodes de docking peuvent être classées en applications de docking rigide et docking flexible selon le degré auquel elles considèrent la flexibilité du ligand et de la protéine pendant le processus d'amarrage 'Docking'<sup>103,104</sup>.

#### ➤ Docking avec ligand rigide «docking rigid-body »

Les premiers outils de docking obéissent au principe dit « lock-and-key » (principe « clef-serrure ») selon lequel le ligand qui représente la clef est complémentaire au niveau géométrique du site actif du récepteur, qui représente la serrure<sup>105,106</sup>. La recherche de la pose optimale se limite au positionnement des ligands dans le site de liaison, qui se fait par translation et rotation. L'intérêt des méthodes de docking rigide réside dans leur rapidité d'exécution qui permet de réaliser un premier filtrage de très larges chimiothèques, permettant de ne pas retenir des molécules aberrantes (trop grande, mauvaise complémentarité avec le site de liaison) avant de procéder, par exemple, à la mise en œuvre de méthodes de docking flexible<sup>107,80</sup>.

#### ➤ Docking avec ligand flexible

La présentation du système biologique comme un système dynamique, dont la liaison du ligand à un site d'interaction se fait en réalité dans la majorité des cas par un processus de sélection de conformation, qui se termine par une étape d'adaptation du ligand et du récepteur

l'un à l'autre<sup>108</sup>. La flexibilité du ligand, permettant d'explorer les conformations adoptées par celui-ci lors de la fixation au site de liaison dans le but de déterminer les bonnes poses du ligand qui correspondent à une énergie libre de liaison basse. Les liaisons dites flexibles sont toutes celles dont la rotation autour de l'axe est chimiquement valide et provoque le déplacement d'une « branche », c'est à dire d'un groupe d'atomes<sup>109</sup>. Aujourd'hui, de nombreux programmes de docking permettent d'ajouter un niveau de complexité au problème en incluant également une certaine flexibilité sur le récepteur<sup>105,110</sup>. En réalité, l'augmentation du nombre de degrés de liberté que cela représente est synonyme d'un décuplement du temps de calcul. Donc le docking de ligand flexible et la protéine rigide reste la méthode la plus souvent employée, car elle permet une évaluation rapide de bases de milliers, voire de millions de molécules<sup>109</sup>.

### 2.2.4.2. Algorithmes de recherche

En principe, un programme de docking doit être capable de générer les modes de liaison attendus pour des ligands dont la position adoptée au sein du site actif est connue (en général, via la structure cristallographique de la protéine co-cristallisée avec le ligand) dans un temps raisonnable. Tout algorithme de docking requiert une recherche conformationnelle la plus exhaustive possible de ce complexe, consiste à générer tous les mouvements de translation et de rotations systématiques de toutes les liaisons rotatives possibles du ligand dans un site de liaison putatif et prédéfini. Plus le degré de liberté de l'entité chimique augmente, plus l'espace conformationnel s'accroît et plus le temps de la recherche exhaustive serait trop long. L'utilisation d'un algorithme de recherche efficace permettant d'atteindre un bon compromis entre l'espace conformationnel couvert et le temps de calcul nécessaire est indispensable. À cette fin, différents algorithmes d'optimisation ont été utilisés au fil des années, tels que les algorithmes de recherche stochastique et déterministe<sup>99,110</sup>.

#### 2.2.4.2.1. Recherche aléatoire ou stochastique

Les algorithmes de recherche aléatoire ou algorithmes stochastiques procèdent à des changements aléatoires des variables de translations, rotations et torsions permettant de générer de nouvelles conformations de ligands ou une population de ligands, lesquels sont acceptés ou rejetés sur la base d'une fonction de probabilité (« fitness »)<sup>99</sup>. Les méthodes stochastiques ne garantissent pas l'obtention du minimum global de l'énergie, mais elles permettent d'obtenir un ensemble de conformères de faible énergie constituant un échantillon supposé être

représentatif. Les méthodes de Monte-Carlo<sup>111,112</sup>, les algorithmes génétiques comptent parmi les approches les plus utilisées<sup>113,114</sup>.

### 2.2.4.3.2. Les algorithmes de recherche déterministe ou de simulation

Deux méthodes principales les caractérisent : les méthodes de dynamique moléculaire et les méthodes de minimisation d'énergie.

➤ **Les méthodes de dynamique moléculaire** se basent sur les équations de mouvements de Newton pour simuler l'évolution d'un système récepteur-ligand au cours du temps à l'échelle atomique<sup>99,112</sup>. La simulation est initialisée par l'attribution de vitesses à chaque atome du système selon une distribution caractérisant une température donnée. Pour cela, la force de chaque atome est calculée à partir d'un changement dans l'énergie potentielle entre une nouvelle position et l'actuelle. La position de chaque atome est ensuite recalculée à très courts intervalles de temps en fonction des forces qui lui sont appliquées, dérivées des énergies potentielles et cinétiques des atomes environnants. Cette technique permet donc d'observer la trajectoire d'une molécule dans le site de liaison de la protéine cible et la stabilité des interactions établies.

Malheureusement, avec un temps de calcul raisonnables, les méthodes de dynamique moléculaire ne sont généralement pas capables de franchir les barrières de hautes énergies et en conséquence, l'identification du minimum global d'énergie se révèle être très difficile, voire quasi-impossible avec ces méthodes et les conformations des ligands obtenues sont généralement piégées dans des minimums locaux d'énergie. Malgré tout, de nombreuses approches ont été développées pour pallier à ce problème<sup>17,99,112,115</sup>. Une des méthodes consiste à réaliser des simulations de différentes parties du complexe ligand-récepteur en variant la température<sup>115</sup> tandis que d'autres proposent de lancer la simulation de dynamique moléculaire à partir de différentes positions initiales du ligand afin d'augmenter les chances de trouver un minimum local d'énergie. Malgré cela, les temps de simulation restent encore trop élevés pour pouvoir utiliser ces méthodes à large échelle lors d'un criblage virtuel. Cependant, les approches de dynamique moléculaire peuvent être employées sur un petit nombre de molécules à la fin d'un criblage virtuel pour optimiser la sélection des meilleurs composés.

➤ **Les méthodes de minimisation d'énergie** : reposent sur l'usage de champs de force pour optimiser, dans le cas des méthodes de docking, les positions atomiques d'un



complexe récepteur-ligand selon leurs énergies à la fois intra et intermoléculaires. Les temps d'exécution associés sont très courts et permettent d'obtenir des poses plus proches de la réalité biologique, notamment en affinant les positions des chaînes latérales des acides aminés impliqués dans l'interaction. Cette approche identifie donc des minimums énergétiques globaux et est majoritairement utilisée pour affiner les poses moléculaires. Elles ne sont donc pas utilisées seules, mais plutôt en complément d'autres méthodes de recherche.

### 2.2.4.4. Fonctions de score

Après avoir obtenu les conformations du ligand (générées) acceptées selon les différentes approches, leurs affinités pour le récepteur doivent être évaluées, les méthodes de docking utilisent des fonctions de score permettant de calculer cette affinité. Le rôle principal des fonctions de score est donc d'évaluer et de classer les poses<sup>101</sup> pour identifier le « vrai » mode de liaison des composés et les ligands les plus prometteurs<sup>99</sup>. Les fonctions de score peuvent être classées selon trois catégories : les fonctions de score basées sur les champs de force, sur les connaissances et les fonctions de score empiriques<sup>99</sup>.

➤ **Fonctions basées sur les champs de force** : Les champs de force dans leur forme standard évaluent la somme de deux énergies : l'énergie entre les atomes liés au sein d'une molécule donnée (énergie interne) et l'énergie entre atomes non-liés (énergie externe)<sup>99</sup>. Ce deuxième terme correspond au terme principal de l'énergie d'interaction protéine-ligand dans le cas d'une application à un problème de docking. Cependant, dans la plupart des fonctions de score basées sur les champs de force implémentée dans les logiciels de docking, une seule conformation de la protéine est prise en compte, ce qui permet de négliger l'énergie libre de la protéine et ainsi de simplifier les calculs de score<sup>116</sup>. Ces fonctions de score se basent sur les champs de force de mécanique moléculaire qui permettent de calculer l'énergie potentielle du système (EMM) qui est exprimée par une somme de quatre termes (Équation 1): les trois premiers termes estiment l'énergie interne du ligand. Le quatrième terme reflète l'affinité d'un ligand donné pour le site actif, mesuré en tant qu'énergie d'interaction, correspond à la contribution d'enthalpie estimée par les potentiels de Lennard Jones et de Coulomb. Le potentiel de Lennard-Jones évalue les interactions de Van der Waals et le potentiel de Coulomb détermine les interactions électrostatiques<sup>117</sup>. L'avantage de l'utilisation des fonctions de score réside en leur rapidité de calcul, mais leur grand inconvénient est qu'elles ne prennent pas en compte les effets entropiques et de solvation<sup>17,99,110</sup> ce qui a été amélioré dans différents logiciels (GOLD, AutoDock<sup>99,110</sup>, et les versions de DOCK récentes (5 et 6)<sup>99</sup>.

$$E_{MM} = \sum_{bond} K_r (r - r_{ref})^2 + \sum_{angles} K_\vartheta (\vartheta - r\vartheta_{ref})^2 + \sum_{dihedrals} K_\varphi [1 + \cos(n\varphi - \varphi_0)]^2 + \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

**Équation.1.** L'énergie potentielle d'un système (EMM) par les champs de force de la mécanique moléculaire. ( $K_r$ ,  $K_\vartheta$  et  $K_\varphi$  : facteurs de pénalité pour les liaisons, les angles et les angles dièdres ;  $r$  et  $r_{ref}$  : longueurs des liaisons mesurées et de référence ;  $\vartheta$  et  $r\vartheta_{ref}$  : valeurs des angles mesurées et de référence ;  $\varphi$  et  $\varphi_{ref}$  : valeurs des angles dièdres mesurées et de référence ;  $A_{ij}$  et  $B_{ij}$  : constantes attractives et répulsives mesurées expérimentalement ;  $R_{ij}$  : distance entre les atomes  $i$  et  $j$  ;  $q_i$  et  $q_j$  : charges des atomes  $i$  et  $j$  ;  $\epsilon$ : constante diélectrique)<sup>177</sup>

### ➤ Fonctions de score empiriques :

L'idée fondamentale des fonctions de score empiriques est que l'énergie d'interaction d'un complexe récepteur-ligand peut être interprétée comme une sommation d'interactions chimiques localisées. En d'autres termes, les fonctions de score empiriques ont pour but de reproduire les données expérimentales d'énergie de liaison et ils utilisent une somme de plusieurs termes individuels décrivant les différents types d'interactions moléculaires pondérés par des coefficients<sup>99,110,112</sup>. Les coefficients des différents termes sont calculés par des analyses de régression utilisant les données expérimentales d'affinité et les données structurales de cristallographie données de complexes protéine-ligand co-cristallisés<sup>110</sup>. L'avantage d'utiliser les fonctions de score empiriques est que leur forme est plus simple que celles basées sur les champs de force, leurs termes sont facilement à déterminer<sup>99</sup> et le temps de calcul sont relativement rapides<sup>110</sup>.

### ➤ Les fonctions de scores basées sur les connaissances ou « knowledge-based

Ces fonctions de score sont construites à partir de règles fondées sur une analyse statistique des complexes protéine-ligand résolus expérimentalement. Elles partent du principe que les distances interatomiques les plus représentées statistiquement dans les complexes constituent des contacts énergétiques favorables et qu'à l'inverse, les plus rares représentent des interactions moins stables<sup>99,118,119</sup>. Ainsi, leur paramétrisation dépend de la quantité d'informations expérimentales disponibles et leur apparition à la profusion de données structurales accessibles dans des bases de données telles que la Protein Data Bank. Les exemples populaires de ces fonctions de scores sont PMF<sup>118</sup> et Drug Score<sup>120</sup>.

### 2.2.4.5. Problématique liée aux méthodes de docking

➤ **Importance des structures de protéines** : L'un des éléments essentiels à considérer avant de réaliser un docking est la disponibilité et la qualité de la structure des protéines étudiées. La base de données de structures la plus populaire est la Protein Data Bank (PDB). Cependant, lorsque de nombreuses structures expérimentales ont été résolues pour une même protéine et que l'on veut n'en retenir qu'une seule pour le criblage virtuel, il n'est pas aisé de décider laquelle choisir. La qualité de la structure obtenue doit donc être évaluée à l'aide de différents paramètres. La résolution mesure la précision avec laquelle le modèle cristallographique reproduit les données expérimentales. En général, une structure cristallographique présentant une résolution inférieure à 2Å est considérée comme une source de donnée fiable<sup>112</sup>. Le facteur de température ou « B-factor » mesure le déplacement des atomes<sup>121</sup>, et donne des indications sur la mobilité locale de la structure de protéine<sup>122</sup>.

➤ **Rôle du solvant** La grande majorité des procédures de docking font intervenir une protéine et un ligand sans prendre en compte le solvant, que ce soit de façon implicite ou explicite. Cependant, cette approche est assez éloignée de la réalité si on veut simuler le milieu biologique dans lequel la protéine et le ligand vont entrer en contact. Les molécules d'eau peuvent notamment servir de pont (liaisons hydrogène) entre la protéine et le ligand permettant de stabiliser le complexe. Dans la grande majorité des criblages virtuels basés sur la structure, le solvant n'est pas pris en compte, ni de façon implicite ni de façon explicite. Le logiciel SLIDE<sup>123</sup> résout partiellement ce problème en déterminer quelles molécules d'eau sont nécessaires à l'établissement des interactions ligand-protéine et devraient être retenues et lesquelles devraient être déplacées.

➤ **Flexibilité de la protéine** La structure d'un site de liaison est directement affectée par la flexibilité d'une protéine. Les changements conformationnels observés sont généralement limités aux mouvements des chaînes latérales des acides aminés du site de liaison. Une très grande majorité des logiciels considèrent la protéine et donc le site de liaison comme rigide au cours du processus de docking, ces logiciels négligent donc le phénomène de flexibilité de la protéine lors de la liaison. Différentes approches permettent de prendre en compte la flexibilité d'une protéine ou de son site de liaison. Les protéines très flexibles constituent des cibles difficiles à étudier avec les méthodes de docking.

### 2.3. Évaluation des méthodes de criblage virtuel

L'intérêt porté aux méthodes de criblage virtuel réside dans le gain potentiel de temps et d'argent que représente leur utilisation pour la découverte des nouveaux *hits*. Puisque les résultats obtenus par criblage virtuel restent des prédictions et que leur fiabilité peut parfois être mise en doute, l'évaluation des méthodes de criblage virtuel est donc essentielle pour s'assurer de leurs performances et ainsi de valider ces résultats.

Il est également possible d'utiliser des résultats expérimentaux précédents (HTS ou autres données d'activité quantifiée) pour évaluer la capacité d'un logiciel à prédire des scores d'affinité proches des données recueillies expérimentalement. Les données de co-cristallographie protéine-ligand permettent également d'évaluer la capacité d'une méthode à repositionner un ligand selon son mode de liaison biologique. Ainsi, il est possible de comprendre pourquoi certaines prédictions des criblages virtuels peuvent être biaisées, soit par un mauvais positionnement des molécules dans le récepteur ciblé, soit par un calcul des scores d'affinité qui a estimé mal les énergies d'interaction réelles<sup>125-127</sup>. Les différents logiciels de criblage sont évalués principalement sur les données de banques d'évaluation grâce à différentes métriques de performance.

#### 2.3.1. Précision du positionnement des composés

##### 2.3.1.1. Écart quadratique-moyenne-racine (RMSD)

Cette mesure géométrique est basée sur la distance entre les positions atomiques d'une pose prédite et d'une pose co-cristallisée d'un ligand (Équation 2)<sup>129</sup>. Plus la prédiction du positionnement est précise, plus les différences entre les deux structures sont faibles, plus la valeur du RMSD est basse

$$RMSD(A, B) = \sqrt{\frac{1}{n} \sum_{i=1}^n ((A_{ix} - Y_{ix})^2 + (A_{iy} - Y_{iy})^2 + (A_{iz} - Y_{iz})^2)}$$

**Équation.2.** Formule du RMSD entre deux composés A et B (x,y et z : coordonnées cartésiennes, n : nombre d'atomes du composé).

### 2.3.1.2. L'espace réel du facteur RSR

Les valeurs de RSR<sup>128</sup> sont calculées à partir des cartes de densité électronique théoriques du modèle (le ligand de référence ou la pose prédite) et de la densité électronique expérimentale (Équation 3). Le modèle est validé si la valeur de RSR correspondante est faible, et vice-versa.

$$RSR = \frac{\sum |\rho_{obs} - \rho_{calc}|}{\sum |\rho_{obs} + \rho_{calc}|}$$

**Équation.3.** Calcul du RSR à l'aide de la densité électronique observée expérimentalement ( $\rho_{obs}$ ) et de la densité électronique calculée à partir des coordonnées atomiques du modèle : le ligand co-cristallisé ou la pose prédite ( $\rho_{calc}$ ).

### 2.3.2. Enrichissement d'une chimiothèque

Les méthodes de criblage virtuel doivent permettre de différencier les composés capables de se lier à la cible thérapeutique, dits « actifs », du reste des composés de la chimiothèque dits « inactifs ». Dans un scénario idéal, après avoir prédit le mode d'interaction correct d'un ligand dans le site de liaison, le calcul du score d'affinité entre le ligand et son récepteur indique une interaction favorable, et vice-versa<sup>129</sup>.

#### 2.3.2.1. Les métriques de performance

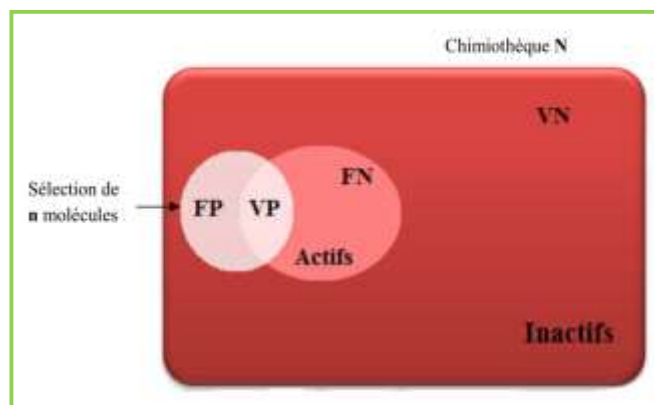
Pour évaluer les méthodes de criblage virtuel, des métriques de performance sont utilisées, au sein de la fraction étudiée de la chimiothèque **N** (Schéma 6), le taux d'actifs aussi appelés vrais positifs (VP) illustrant la sélectivité (Équation 4), et de decoys (somme des vrais négatifs VN et des faux positifs FP) témoin de la spécificité (Équation 5)<sup>129</sup>.

$$Se = \frac{VP}{VP+FN}$$

**Équation.4.** La sélectivité (Se) est définie comme le ratio du nombre d'actifs reconnus par la méthode dans une fraction donnée de la chimiothèque 'n' (VP : vrais positifs) sur le nombre d'actifs total de la chimiothèque 'N' ( faux négatifs (FN) avec VP), FN c'est les actifs non reconnus par la méthode de criblage virtuel.

$$Sp = \frac{VN}{VN+FP}$$

**Équation 5.** La spécificité (Sp) représente le ratio du nombre de decoys non reconnus dans la fraction de la chimiothèque 'n' (VN : vrais négatifs) sur le nombre de decoys total 'N' (FP : faux positifs avec VN), FP c'est les inactifs sélectionnés par le modèle.



**Schéma 6.** Représentation de la sélection de 'n' molécules dans une chimiothèque 'N'.

À partir de ces valeurs, des métriques de performances. On distingue généralement les métriques dites « classiques », telles que les facteurs d'enrichissement et les courbes de ROC (Receiver Operating Characteristic).

### ➤ Facteurs d'enrichissement

Le facteur d'enrichissement (Enrichment Factor EF) est une métrique de performance couramment utilisée. Il est calculé pour des fractions choisies de la chimiothèque, pour valider la capacité d'une méthode de criblage virtuel à retrouver des actifs dans une fraction donnée de la chimiothèque par rapport à une sélection aléatoire (Équation 5) <sup>129</sup>.

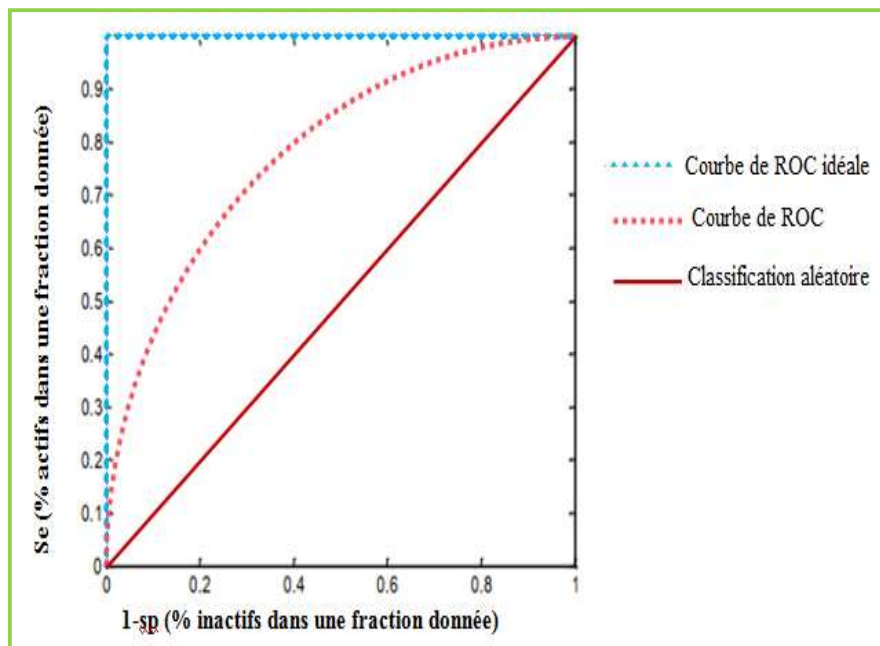
$$EF = \frac{VP/n}{(VP + FN)/N}$$

**Équation.5.** Calcul du facteur d'enrichissement (EF) de la chimiothèque (avec n: nombre de composés dans la fraction de la chimiothèque étudiée, N : nombre total de composés dans la chimiothèque, VP : vrais positifs, et FN : faux négatifs).

### ➤ Courbe de fraction d'efficacité du récepteur (ROC)

La courbe ROC (« Receiver Operating Characteristic curve ») trace l'évolution du taux de vrais positifs (la prédiction dit vrai et elle est c'est correct) qui représente sensibilité en fonction du taux de faux positifs (la prédiction dit vrai, mais c'est incorrect) qui représente spécificité ( $1-Sp$ ). Les courbes de ROC permettent de visualiser les fractions de vrais positifs et de faux positifs retrouvées à chaque rang d'une chimiothèque classée <sup>130</sup>.

L'aire sous la courbe de ROC (ou *Area Under the Curve* **AUC**) résume la capacité de discrimination globale d'une méthode en un unique indicateur chiffré qui peut être utilisé pour comparer rapidement les performances de plusieurs méthodes de criblage virtuel. Une classification aléatoire produit une AUC égale à 0,5 tandis qu'une discrimination parfaite est caractérisée par une AUC égale à 1. Plus une AUC est proche de 1, plus la classification réalisée est discriminante (Figure 4).



**Fig.4.** Les courbes de ROC, une classification aléatoire des composés de la chimiothèque est représentée par une diagonale allant du point (0,0) au point (1,1) du graphique. Si une méthode permet de discriminer les actifs et les decoys de manière plus efficace que le hasard. Pour une distribution idéale, c'est-à-dire que les actifs sont tous classés en premier, la courbe monte en ligne droite jusqu'au point (0,1) ( $Se=Sp=1$  pour tous les actifs) puis viennent ensuite les inactifs ( $Se=1, Sp=0$  pour tous les inactifs) représentés par une ligne horizontale entre les points (0,1) et (1,1) du graphique <sup>130</sup>.

### Références bibliographique :

1. Sneader, W. (2005). *Drug discovery: a history*. John Wiley & Sons.
2. Schmitz, R. (1985). Friedrich Wilhelm Sertürner and the discovery of morphine. *Pharmacy in history*, 27(2), 61-74.
3. Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature reviews Drug discovery*, 9(3), 203-214.
4. Pereira, D. A., & Williams, J. A. (2007). Origin and evolution of high throughput screening. *British journal of pharmacology*, 152(1), 53-61.
5. Feng, B. Y., Shelat, A., Doman, T. N., Guy, R. K., & Shoichet, B. K. (2005). High-throughput assays for promiscuous inhibitors. *Nature chemical biology*, 1(3), 146-148.
6. Hughes, JP, Rees, S., Kalindjian, SB et Philpott, KL (2011). Principes de découverte précoce de médicaments. *Revue britannique de pharmacologie*, 162 (6), 1239-1249.
7. Kerns, E.H., Li Di., 2008. Drug-Like Properties: Concepts, Structure Design and Methods from ADME to Toxicity Optimization, first ed. Elsevier, United States.
8. Zhang, S. (2011). Computer-aided drug discovery and development. In *Drug Design and Discovery*, 23-38.
9. Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chemico-biological interactions*, 171(2), 165-176.
10. Gao, Q., Yang, L., & Zhu, Y. (2010). Pharmacophore based drug design approach as a practical process in drug discovery. *Current computer-aided drug design*, 6(1), 37-49.
11. Leelananda, S. P., & Lindert, S. (2016). Computational methods in drug discovery. *Beilstein journal of organic chemistry*, 12(1), 2694-2718.
12. Macalino, S. J. Y., Gosu, V., Hong, S., & Choi, S. (2015). Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research*, 38(9), 1686-1701.
13. Ooms, F. (2000). Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Current medicinal chemistry*, 7(2), 141-158.
14. Tanrikulu, Y., Krüger, B., & Proschak, E. (2013). The holistic integration of virtual screening in drug discovery. *Drug Discovery Today*, 18(7-8), 358-364.
15. Duffy, B. C., Zhu, L., Decornez, H., & Kitchen, D. B. (2012). Early phase drug discovery: cheminformatics and computational techniques in identifying lead series. *Bioorganic & medicinal chemistry*, 20(18), 5324-5342.
16. Herbst, R. S., Fukuoka, M., & Baselga, J. (2004). Gefitinib—a novel targeted approach to treating cancer. *Nature Reviews Cancer*, 4(12), 956-965.
17. Harris, P. A., & Stafford, J. A. (2009). Discovery of pazopanib: a pan vascular endothelial growth factor kinase inhibitor. *Kinase Inhibitor Drugs*, 57-77.



18. Njoroge, F. G., Chen, K. X., Shih, N. Y., & Piwinski, J. J. (2008). Challenges in modern drug discovery: a case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection. *Accounts of chemical research*, 41(1), 50-59.
19. Walters, W. P., Stahl, M. T., & Murcko, M. A. (1998). Virtual screening an overview. *Drug discovery today*, 3(4), 160-178.
20. Leach, AR, Shoichet, BK et Peishoff, CE (2006). Prédiction des interactions protéine-ligand. Amarrage et notation: réussites et lacunes. *Journal de chimie médicinale* , 49 (20), 5851-5855.
21. Enyedy, I.J., & Egan, W.J. (2008). Can we use docking and scoring for hit-to-lead optimization?. *Journal of computer-aided molecular design*, 22(3-4), 161-168.
22. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23(1-3), 3-25.
23. Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, 45(12), 2615-2623.
24. Johnson, T. W., Dress, K. R., & Edwards, M. (2009). Using the Golden Triangle to optimize clearance and oral absorption. *Bioorganic & medicinal chemistry letters*, 19(19), 5560-5564.
25. Speck-Planche, A., Luan, F., & NDS Cordeiro, M. (2012). Role of ligand-based drug design methodologies toward the discovery of new anti-Alzheimer agents: futures perspectives in Fragment-Based Ligand Design. *Current medicinal chemistry*, 19(11), 1635-1645.
26. Aparoy, P., Kumar Reddy, K., & Reddanna, P. (2012). Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors. *Current medicinal chemistry*, 19(22), 3763-3778.
27. Johnson, M. A., & Maggiora, G. M. (1990). Concepts and applications of molecular similarity. *Berichte Der Bunsengesellschaft Für Physikalische Chemie*, 96(8), 1087-1087.
28. Hawkins, P. C., Skillman, A. G., & Nicholls, A. (2007). Comparison of shape-matching and docking as virtual screening tools. *Journal of medicinal chemistry*, 50(1), 74-82.
29. Mason, J. S., Good, A. C., & Martin, E. J. (2001). 3-D pharmacophores in drug discovery. *Current pharmaceutical design*, 7(7), 567-597.
30. Acharya, C., Coop, A., E Polli, J., & D MacKerell, A. (2011). Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Current computer-aided drug design*, 7(1), 10-22.
31. Kurogi, Y., & Guner, O. F. (2001). Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Current medicinal chemistry*, 8(9), 1035-1055.
32. Koehn, F. E., & Carter, G. T. (2005). The evolving role of natural products in drug discovery. *Nature reviews Drug discovery*, 4(3), 206-220.

33. Hu, Y., Stumpfe, D., & Bajorath, J. (2016). Recent advances in scaffold hopping: miniperspective. *Journal of medicinal chemistry*, 60(4), 1238-1246.
34. Bajorath, J. (2017). Computational scaffold hopping: cornerstone for the future of drug design? *Future Medicinal Chemistry*, 9(7), 629–631.
35. Jasial, S., Hu, Y., & Bajorath, J. (2016). Assessing the growth of bioactive compounds and scaffolds over time: implications for lead discovery and scaffold hopping. *Journal of chemical information and modeling*, 56(2), 300-307.
36. Koeppen, H., Kriegl, J., Lessel, U., Tautermann, C. S., & Wellenzohn, B. (2011). Ligand-Based Virtual Screening. *Virtual Screening*, 61–85
37. Willett, P., Barnard, J. M., & Downs, G. M. (1998). Chemical similarity searching. *Journal of chemical information and computer sciences*, 38(6), 983-996.
38. Willett, P. (2010). Similarity Searching Using 2D Structural Fingerprints. *Cheminformatics and Computational Chemical Biology, Methods Mol Biol*, 133–158
39. Ehrlich, P. (1909). Über den jetzigen Stand der Chemotherapie. *Berichte der deutschen chemischen Gesellschaft*, 42(1), 17-47.
40. Wermuth, C. G., Ganellin, C. R., Lindberg, P., & Mitscher, L. A. (1998). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry*, 70(5), 1129-1143.
41. Qing, X., Lee, X. Y., De Raeymaecker, J., Tame, J. R., Zhang, K. Y., De Maeyer, M., & Voet, A. (2014). Pharmacophore modeling: advances, limitations, and current utility in drug discovery. *Journal of Receptor, Ligand and Channel Research*, 7, 81-92.
42. Finn, P. W. (1996). Computer-based screening of compound databases for the identification of novel leads. *Drug Discovery Today*, 1(9), 363-370.
43. Ala`a Aziz,A.(2019). Doctoral thesis. Computer-aided drug design and discovery of new KDR (vascular endothelial growth factor receptor2) inhibitors via structure-based and ligand-based modeling followed by in vitro assay. School of Graduates Studies The University of Jordan.
44. Dror, O., Shulman-Peleg, A., Nussinov, R., & Wolfson, H. J. (2004). Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Current medicinal chemistry*, 11(1), 71-90.
45. Cone, M. M., Venkataraghavan, R., & McLafferty, F. W. (1977). Computer-aided interpretation of mass spectra. 20. Molecular structure comparison program for the identification of maximal common substructures. *Journal of the American Chemical Society*, 99(23), 7668-7671.
46. Schulz, K.-P. (1992). M. Johnson, G. Maggiora (Eds.): Concepts and Applications of Molecular Similarity, Wiley Interscience,. *Berichte Der Bunsengesellschaft Für Physikalische Chemie*, 96(8), 1087–1087.
47. Gillet, V. J. (2012). Pharmacophore Models in Drug Design. *Physico-Chemical and Computational Approaches to Drug Discovery*, 53, 151-170.

48. Kirchmair, J., Markt, P., Distinto, S., Wolber, G., & Langer, T. (2008). Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes?. *Journal of computer-aided molecular design*, 22(3-4), 213-228.
49. Clement, O. O., & Mehl, A. T. (2000). HipHop: pharmacophores based on multiple common-feature alignments. *Pharmacophore perception, development, and use in drug design*, 69-84.
50. Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., & Friesner, R. A. (2006). PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *Journal of computer-aided molecular design*, 20(10-11), 647-671.
51. Dearden, J. C., Cronin, M. T., & Kaiser, K. L. (2009). How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, 20(3-4), 241-266.
52. Akamatsu, M. (2002). Current state and perspectives of 3D-QSAR. *Current topics in medicinal chemistry*, 2(12), 1381-1394.
53. Verma, R. P., & Hansch, C. (2008). Camptothecins: a SAR/QSAR study. *Chemical reviews*, 109(1), 213-235.
54. Verma, J., Khedkar, V. M., & Coutinho, E. C. (2010). 3D-QSAR in drug design-a review. *Current topics in medicinal chemistry*, 10(1), 95-115.
55. Damale, M., N Harke, S., A Kalam Khan, F., B Shinde, D., & N Sangshetti, J. (2014). Recent advances in multidimensional QSAR (4D-6D): a critical review. *Mini reviews in medicinal chemistry*, 14(1), 35-55.
56. Ghose, A., & Viswanadhan, V. (2001). *Combinatorial Library Design and Evaluation: Principles, Software, Tools, and Applications in Drug Discovery*. CRC Press.
57. Zhao, M., Li, Z., Peng, L., Tang, Y. R., Wang, C., Zhang, Z., & Peng, S. (2007). Novel 1-oxyl-2-substitutedphenyl-4, 4, 5, 5-tetramethylimidazolines: Synthesis, selectively analgesic action, and QSAR analysis. *Bioorganic & medicinal chemistry*, 15(8), 2815-2826.
58. Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, 22(1), 69-77.
59. Chtita, S., Ghamali, M., Hmamouchi, R., Elidrissi, B., Bourass, M., Larif, M., ... & Lakhlifi, T. (2016). Investigation of antileishmanial activities of acridines derivatives against promastigotes and amastigotes form of parasites using quantitative structure activity relationship analysis. *Advances in Physical Chemistry*, 2016,1-6.
60. Degen, B. (2010). Organisation de coopération et de développement économiques (OCDE).

61. OECD, D. (2007). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q) SAR] Models. *Organisation for Economic Co-operation and Development, Paris, France.*
62. OECD, O. (2004). Principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationship models. In *37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology, Organisation for Economic Co-Operation and Development, Paris, France.*
63. Akamatsu, M. (2002). Current state and perspectives of 3D-QSAR. *Current topics in medicinal chemistry*, 2(12), 1381-1394.
64. Winkler, D. A. (2005). Overview of Quantitative Structure-Activity Relationships (QSAR). *Molecular Analysis and Genome Discovery*, 347–367.
65. MarvinSketch, displaying and characterizing chemical structures, substructures, and reactions, Marvin 17.1.2, 2017, ChemAxon (<http://www.chemaxon.com>)
66. Dragon 5.4, <http://www.disat.unimib.it>
67. Gelman, A., & Hill, J. (n.d.). Linear regression: the basics. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 31–52.
68. Statistical Basis of Regression and Partial Least-Squares Analysis. (2010). *Quantitative Drug Design*, 131–155.
69. Chauhan, H., Bernick, J., Prasad, D., & Masand, V. (2016). The Role of Artificial Neural Networks on Target Validation in Drug Discovery and Development. In *Artificial Neural Network for Drug Design, Delivery and Disposition* (pp. 15-27). Academic Press.
70. Chtita, S., Larif, M., Ghamali, M., Bouachrine, M., & Lakhlifi, T. (2015). Quantitative structure-activity relationship studies of dibenzo [a, d] cycloalkenimine derivatives for non-competitive antagonists of N-methyl-d-aspartate based on density functional theory with electronic and topological descriptors. *Journal of Taibah University for Science*, 9(2), 143-154.
71. Hemmateenejad, B., Safarpour, M. A., Miri, R., & Taghavi, F. (2004). Application of ab initio theory to QSAR study of 1, 4-dihydropyridine-based calcium channel blockers using GA-MLR and PC-GA-ANN procedures. *Journal of computational chemistry*, 25(12), 1495-1503.
72. Ghamali, M., Chtita, S., Ousaa, A., Elidrissi, B., Bouachrine, M., & Lakhlifi, T. (2017). QSAR analysis of the toxicity of phenols and thiophenols using MLR and ANN. *Journal of Taibah University for Science*, 11(1), 1-10.
73. Benbrahim, I. (2017). Étude par des méthodes quantiques et empiriques des relations SAR et QSAR dans des hétérocycles à intérêt médicamenteux. Doctoral thesis, Université Mohamed Khider - Biskra.
74. Salah, T. (2017). Étude par la chimie computationnelle des corrélations 2D-QSAR et 3D-QSAR de quelques composés bioactifs. Thèse de doctorat, Université Mohamed Khider - Biskra.
75. Kerassa, A. (2015). Étude par la modélisation moléculaire des relations structures-propriétés de quelques séries hétérocycliques bioactives. Thèse de doctorat, Université Mohamed Khider - Biskra.

76. Melkemi., N. (2013). Étude des corrélations entre les paramètres structuraux et les propriétés bioactives de quelque composés organiques. Thèse de doctorat, Université Mohamed Khider - Biskra.
77. Cummings, B. S. (2007). Phospholipase A2 as targets for anti-cancer drugs. *Biochemical pharmacology*, 74(7), 949-959.
78. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data Bank nucleic acids research, 28(1), 235-242.
79. Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Current medicinal chemistry*, 20(23), 2839-2860.
80. Tuccinardi, T. (2009). Docking-based virtual screening: recent developments. *Combinatorial chemistry & high throughput screening*, 12(3), 303-314.
81. Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintha, C. (2012). Homology modeling a fast tool for drug discovery: current perspectives. *Indian journal of pharmaceutical sciences*, 74(1), 1.
82. Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., ... & Vajda, S. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nature protocols*, 10(5), 733-755.
83. Butkiewicz, M., Lowe, E., Mueller, R., Mendenhall, J., Teixeira, P., Weaver, C., & Meiler, J. (2013). Benchmarking ligand-based virtual high-throughput screening with the PubChem database. *Molecules*, 18(1), 735-756.
84. Laurie, A. T., & Jackson, R. M. (2006). Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. *Current Protein and Peptide Science*, 7(5), 395-406.
85. Hessler, G., & Baringhaus, K. H. (2010). The scaffold hopping potential of pharmacophores. *Drug Discovery Today: Technologies*, 7(4), e263-e269.
86. Thangavelu, K., Chong, Q. Y., Low, B. C., & Sivaraman, J. (2014). Structural basis for the active site inhibition mechanism of human kidney-type glutaminase (KGA). *Scientific reports*, 4(1), 1-7.
87. Pascarella, S., Angelaccio, S., Contestabile, R., Fratte, S. D., Salvo, M. D., & Bossa, F. (1998). The structure of serine hydroxymethyltransferase as modeled by homology and validated by site-directed mutagenesis. *Protein Science*, 7(9), 1976-1982.
88. Khedkar, S. A., Malde, A. K., Coutinho, E. C., & Srivastava, S. (2007). Pharmacophore modeling in drug discovery and development: an overview. *Medicinal Chemistry*, 3(2), 187-197.
89. Kumar, S. P., & Jha, P. C. (2016). Multi-level structure-based pharmacophore modelling of caspase-3-non-peptide complexes: extracting essential pharmacophore features and its application to virtual screening. *Chemico-biological interactions*, 254, 207-220.
90. Kumar, R. B., Suresh, M. X., & Priya, B. S. (2015). Pharmacophore modeling, in silico screening, molecular docking and molecular dynamics approaches for potential alpha-delta bungarotoxin-4 inhibitors discovery. *Pharmacognosy magazine*, 11(12),19-28.

91. Vuorinen, A., & Schuster, D. (2015). Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods*, 71, 113-134.
92. Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery*, 4(8), 649-663.
93. Hartenfeller, M., & Schneider, G. (2010). De Novo Drug Design. *Chemoinformatics and Computational Chemical Biology*, 299–323.
94. Grover, M., Singh, B., Bakshi, M., & Singh, S. (2000). Quantitative structure–property relationships in pharmaceutical research–Part 2. *Pharmaceutical science & technology today*, 3(2), 50-57.
95. Polanski, J. (2009). Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. *Current medicinal chemistry*, 16(25), 3243-3257.
96. Schneider, G. (2005). Chemoinformatics. Concepts, Methods, and Tools for Drug Discovery. Herausgegeben von Jürgen Bajorath. *Angewandte Chemie*, 117(4), 515–515.
97. Ortiz, A. R., Pisabarro, M. T., Gago, F., & Wade, R. C. (1995). Prediction of drug binding affinities by comparative binding energy analysis. *Journal of medicinal chemistry*, 38(14), 2681-2691.
98. Pan, D., Tseng, Y., & Hopfinger, A. J. (2003). Quantitative structure-based design: formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. *Journal of chemical information and computer sciences*, 43(5), 1591-1607.
99. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery*, 3(11), 935-949.
100. Barril, X., & Soliva, R. (2006). Molecular Modelling. *Molecular BioSystems*, 2(12), 660-681.
101. Sottriffer, C., Klebe, G., Stahl, M., & Böhm, H.-J. (2003). Docking and Scoring Functions/Virtual Screening. *Burger's Medicinal Chemistry and Drug Discovery*, 281–331.
102. Andrusier, N., Mashiach, E., Nussinov, R., & Wolfson, H. J. (2008). Principles of flexible protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, 73(2), 271-289.
103. Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function, and Bioinformatics*, 47(4), 409-443.
104. Dias, R., & de Azevedo Jr., W. (2008). Molecular Docking Algorithms. *Current Drug Targets*, 9(12), 1040–1047.
105. Yuriev, E., Agostino, M., & Ramsland, P. A. (2011). Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition*, 24(2), 149-164.
106. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*, 161(2), 269-288.
107. Lyne, P. D. (2002). Structure-based virtual screening: an overview. *Drug discovery today*, 7(20), 1047-1055.

108. Vogt, A. D., & Di Cera, E. (2013). Conformational selection is a dominant mechanism of ligand binding. *Biochemistry*, 52(34), 5723-5729.
109. Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., & DesJarlais, R. L. (2005). Docking: successes and challenges. *Current pharmaceutical design*, 11(3), 323-333.
110. Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein–ligand docking: current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, 65(1), 15-26.
111. Metropolis, N., & Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, 44(247), 335-341.
112. Spyrakis, F., Cozzini, P., & Kellogg, G. E. (2003). Docking and scoring in drug discovery. *Burger's Medicinal Chemistry and Drug Discovery*, 601-684.
113. Lameijer, E. W., Bäck, T., Kok, J. N., & Ijzerman, A. P. (2005). Evolutionary algorithms in drug design. *Natural Computing*, 4(3), 177-243.
114. Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, 267(3), 727-748.
115. Di Nola, A., Roccatano, D., & Berendsen, H. J. (1994). Molecular dynamics simulation of the docking of substrates to proteins. *Proteins: Structure, Function, and Bioinformatics*, 19(3), 174-182.
116. Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., & Corbeil, A. C. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British journal of pharmacology*, 153(S1), S7-S26.
117. González, M. A. (2011). Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, 12, 169-200.
118. Muegge, I. (2006). PMF Scoring Revisited. *Journal of Medicinal Chemistry*, 49(20), 5895–5902.
119. Yang, C. Y., Wang, R., & Wang, S. (2006). M-score: a knowledge-based potential scoring function accounting for protein atom mobility. *Journal of medicinal chemistry*, 49(20), 5903-5911.
120. Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *Journal of molecular biology*, 295(2), 337-356.
121. Hawkins, P. C., Warren, G. L., Skillman, A. G., & Nicholls, A. (2008). How to do an evaluation: pitfalls and traps. *Journal of computer-aided molecular design*, 22(3-4), 179-190.
122. Davis, A. M., Teague, S. J., & Kleywegt, G. J. (2003). Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angewandte Chemie International Edition*, 42(24), 2718-2736.
123. Raymer, M. L., Sanschagrin, P. C., Punch, W. F., Venkataraman, S., Goodman, E. D., & Kuhn, L. A. (1997). Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *Journal of molecular biology*, 265(4), 445-464.
124. Carlson, H. A. (2002). Protein flexibility and drug design: how to hit a moving target. *Current opinion in chemical biology*, 6(4), 447-452.

125. McGaughey, G. B., Sheridan, R. P., Bayly, C. I., Culberson, J. C., Kretsoulas, C., Lindsley, S., ... & Cornell, W. D. (2007). Comparison of topological, shape, and docking methods in virtual screening. *Journal of chemical information and modeling*, 47(4), 1504-1519.
126. Mysinger, M. M., Carchia, M., Irwin, J. J., & Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, 55(14), 6582-6594.
127. Huang, N., Shoichet, B. K., & Irwin, J. J. (2006). Benchmarking sets for molecular docking. *Journal of medicinal chemistry*, 49(23), 6789-6801.
128. Schnecke, V., & Kuhn, L. A. (2000). Virtual screening with solvation and ligand-induced complementarity. In *Virtual Screening: An Alternative or Complement to High Throughput Screening?* (pp. 171-190). Springer, Dordrecht.
129. Kirchmair, J., Markt, P., Distinto, S., Wolber, G., & Langer, T. (2008). Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes?. *Journal of computer-aided molecular design*, 22(3-4), 213-228.
130. Triballeau, N., Acher, F., Brabet, I., Pin, J. P., & Bertrand, H. O. (2005). Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *Journal of medicinal chemistry*, 48(7), 2534-2547.



## **CHAPITER III**

### **Travaux réalisés**

## **Partie 1**

**Étude de drug-likeness et de la relation structure-activité/propriétés qualitative d'inhibiteurs d'AChE des dérivés de la pyridazine**

En chimie médicinale, le remplacement bio-isostérique est l'une des techniques les plus efficaces, utilisée dans la conception des médicaments<sup>1</sup>. Ces techniques sont utilisées dans le processus d'optimisation de ces derniers pour améliorer leurs propriétés. Environ 50% de toutes les molécules médicamenteuses utilisées dans le domaine médical contiennent des cycles phényle qui peuvent être substitués ou non<sup>2</sup>. Le remplacement bio-isostérique des cycles phényles par des cycles de pyridazine (1,2-diazine) ouvre l'accès à plusieurs milliers d'échantillons d'analogues de la diazine avec plus de possibilités d'interaction. Récemment, GlaxoSmithKline a examiné la possibilité d'administrer de la pyridazine, étant l'un des hétérocycles les plus utilisés dans la conception de médicaments<sup>3</sup>. Les composés contenant le cycle pyridazine s'avèrent être un ligand utile pour différentes cibles. Par conséquent, les dérivés de la pyridazine ont démontré des activités biologiques polyvalentes comme les antibactériens<sup>4</sup>, les anti-inflammatoires<sup>5</sup>, les antituberculeuses<sup>6</sup> et les antihypertenseurs<sup>7</sup>. Des dérivés des pyridazines ont été proposés comme « structure privilégiée » pour la découverte de médicaments neurodégénératifs<sup>8</sup>. Dans notre étude, nous nous sommes basés sur une classe particulière d'inhibiteurs de l'acétylcholinestérase (ACHéI) des dérivés de la pyridazine. L'acétylcholinestérase est considérée comme l'une des cibles visées pour le traitement de la maladie neurodégénérative « Alzheimer ».

Différentes techniques pour un criblage virtuel à base de ligand ont été mise en œuvre dans cette partie, telles que: l'optimisation multiparamètres du système nerveux central (MPO du SNC), la règle du triangle d'or, la relation qualitative entre la structure et l'activité/propriété (SAR/SPR), les propriétés *drug-likeness* et les indices de lipophilicité. L'objectif de ces techniques est de guider la conception, la sélection et l'identification rapide d'un composé « *drug-like* » parmi une classe d'inhibiteurs de l'acétylcholinestérase.

## 1. Matériels et méthodes

### 1.1. Méthodes de calcul utilisées

À l'aide du logiciel HyperChem (ver. 8.08)<sup>9</sup> nous avons pu déterminer l'optimisation géométrique de dix-sept composés de dérivés de la pyridazine (Figure 1) en utilisant la Mécanique Moléculaire (MM+). Les structures obtenues ont été ré-optimisées en utilisant la méthode semi-empirique PM3 [3]. Les propriétés physico-chimiques ont été calculées grâce au logiciel MarvinSketch (ver. 17.13.0)<sup>10</sup> et le module QSAR (ver.8.0.6) intégré au logiciel HyperChem.

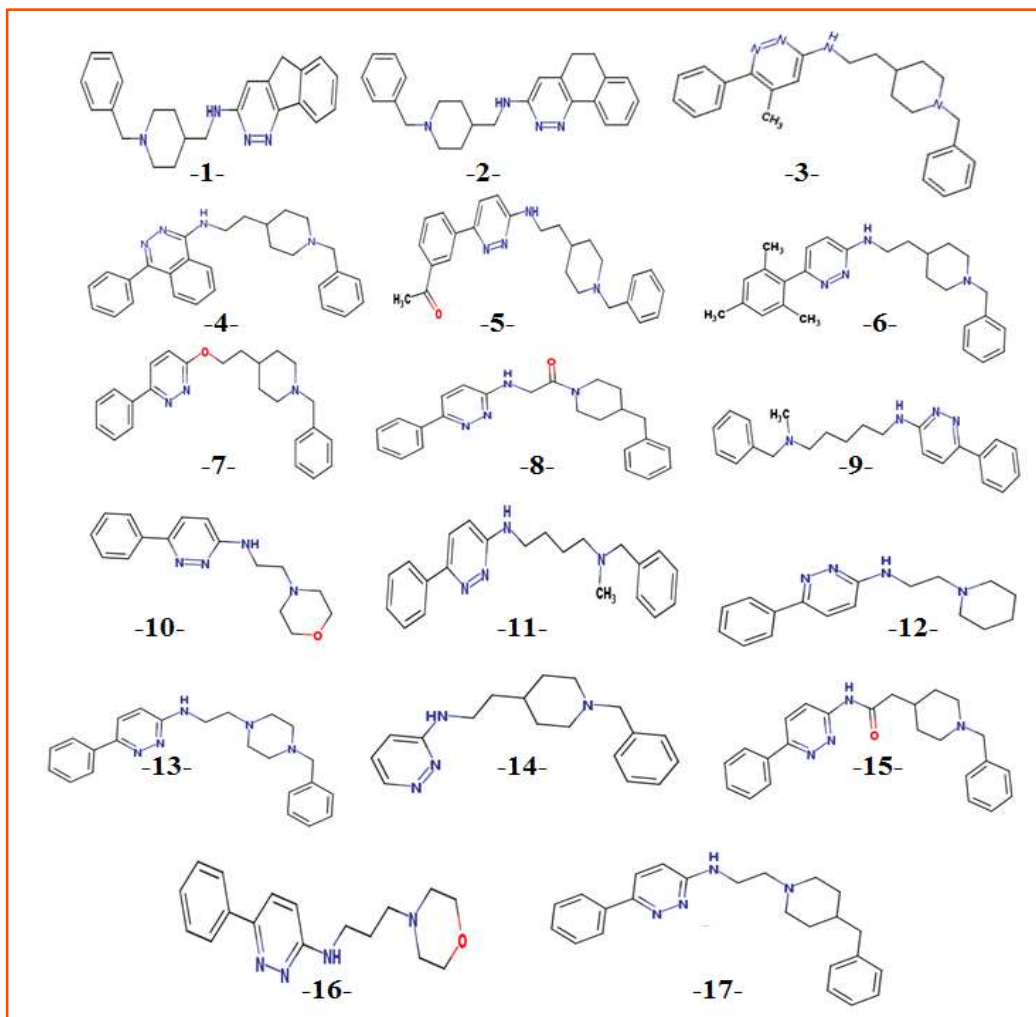


Fig.1. Structures 3D des dérivés de la pyridazine.

## 1.2. Méthodes d'optimisation multiparamètres du système nerveux central (MPO du SNC)

La conception des médicaments qui réagiront dans le système nerveux central (SNC) est confrontée à des défis particuliers. Ces médicaments en question doivent avoir des propriétés pharmacocinétiques, être sans danger pour l'organisme et avoir une grande chance de pénétrer la barrière hémato-encéphalique (BHE) <sup>11</sup>.

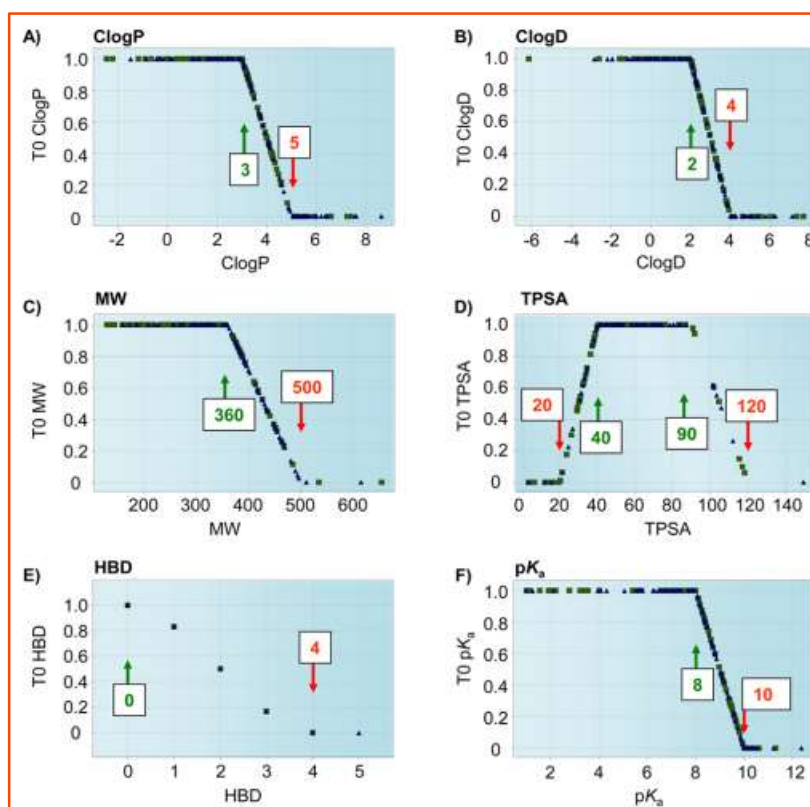
Un nouvel algorithme d'optimisation des paramètres du SNC (score de désirabilité MPO du SNC) a été développé pour améliorer la conception de ce type de molécules avec une meilleure pénétration de la BHE et des propriétés médicamenteuses accrues <sup>12,13</sup>. Ce nouvel algorithme MPO du CNS (équation 1) se base sur un ensemble de six paramètres physico-chimiques fondamentaux : ((1) coefficient de partage octanol/eau **LogP**; (2) coefficient de distribution à pH = 7,4 **logD** ; (3) poids moléculaire **MW**; (4) surface polaire topologique **TPSA** ; (5) nombre de liaisons hydrogène donneurs **HBD** ; (6) centre le plus basique **pKa**).

## Etude de drug-likeness et de la relation entre la structure et l'activité/propriété d'inhibiteurs d'AChE des dérivés de la pyridazine

Le score de désirabilité de ces propriétés physico-chimiques est de 0 à 1 pour chaque propriété. Allant des plus souhaitables ( $T0 = 1$ ) aux moins souhaitables ( $T0 = 0$ ) et sont respectivement représentés par des flèches vertes et rouges, dans la Figure 2. Le score total d'un composé varie de 0 à 6. Les composés ayant un score MPO du SNC  $\geq 4$  sont considérés comme des molécules ayant un bon profil *drug-like*<sup>12</sup>.

$$\text{Score MPO} = \sum \text{score}(\text{LogP} + \text{LogD} + \text{MW} + \text{TPSA} + \text{HBD} + \text{pKa}) \quad \text{Equation 1}$$

Un score de MPO du SNC élevé augmente les chances d'identifier un composé *drug-like* avec des propriétés ADME optimales et une bonne capacité de pénétration de la BHE, tout en maintenant un espace de sécurité *in vitro* favorable<sup>12-14</sup>.



**Fig.2.** Les graphes représentent les six propriétés physico-chimiques utilisées pour générer le score MPO du SNC.<sup>12</sup>

### 1.3. Propriétés drug-likeness et indices de la lipophilie

Le concept de *drug-likeness* permet d'optimiser les propriétés pharmacocinétiques de la structure moléculaire avant même que la substance ne soit synthétisée et testée. Les paramètres pharmacocinétiques *in vivo* sont fortement influencés par les propriétés physico-chimiques d'un médicament. Ces propriétés, principalement l'hydrophobicité, la distribution électronique, les caractéristiques de liaison hydrogène, la taille et la flexibilité de la molécule et bien sûr la

## Etude de drug-likeness et de la relation entre la structure et l'activité/propriété d'inhibiteurs d'AChE des dérivés de la pyridazine

présence de diverses caractéristiques pharmacophoriques, influencent le comportement de la molécule dans un organisme vivant, y compris la biodisponibilité, les propriétés ADME-Tox (Absorption, Distribution, Métabolisme, Élimination et Toxicité). Les propriétés *drug-likeness* comprennent les règles de Lipinski, de Veber<sup>15,16</sup> et du triangle d'or<sup>17</sup>.

La première étude approfondie de l'analyse des propriétés ADME a été effectuée par Lipinski et a abouti à la fameuse « règle des 5 », qui affirme qu'une bonne absorption intestinale est plus probable lorsque : le poids moléculaire (MW) <500u (u.m.a), que le nombre de liaisons hydrogène donneurs HBD<5 (en comptant la somme de tous les groupes NH et OH), que le coefficient de partition octanol/eau LogP < 5 et que le nombre de liaisons accepteurs d'hydrogène HBA<10 (en comptant tous les atomes d'azote et d'oxygène)<sup>15</sup>. Si un composé possède plus d'une violation de la règle des 5 (qui a un score situé entre 0 et 4), il sera susceptible de posséder une mauvaise biodisponibilité orale. Deux autres descripteurs ont été identifiés par Veber *et al.*: le nombre de bandes rotatives (NBR) <10 et la surface polaire (PSA) <140 Å<sup>2</sup><sup>16</sup>.

La lipophilie (caractéristique déterminant l'affinité chimique pour les graisses) d'une molécule est un critère extrêmement important. Elle est un paramètre qui permet de savoir si la molécule a plus tendance à être hydrophile (ayant une affinité pour l'eau), ou lipophile (donc hydrophobe). Les molécules trop lipophiles ont beaucoup plus de chances d'être instables métaboliquement, d'avoir des problèmes de toxicité et de faible solubilité aqueuse.

À l'inverse, une molécule trop hydrophile sera sujette à d'autres problèmes, notamment celui d'avoir une faible perméabilité des membranes. De ce fait, le chimiste médicinal doit comprendre, pour un type de structure donnée, quel est l'espace de lipophilie qui lui permettrait de trouver le meilleur compromis pour conserver des propriétés adéquates pour une molécule active.

L'efficacité du ligand (LE) est interprétée comme une mesure de la qualité de l'interaction entre un composé donné et la protéine cible<sup>17</sup>. L'efficacité du ligand est utile pour optimiser l'affinité par rapport à la taille moléculaire : **LE = 1,4pIC50 / NH**.

Pour faciliter l'optimisation de l'affinité vis-à-vis de la lipophilie, Leeson et Springthorpe<sup>18</sup> ont défini l'efficacité de la lipophilie du ligand LipE (LLE); ce paramètre combine à la fois la puissance et la lipophilie dans une tentative d'estimation de *drug-like*.

LipE tente d'améliorer la puissance tout en maintenant une lipophilie faible, ce qui rend l'interaction avec le récepteur souhaité<sup>17,19</sup> :  $\text{LipE} = \text{pIC50} - \text{LogP}$ .

Pour obtenir des propriétés ADME optimales, la taille moléculaire et la lipophilie sont des facteurs importants à prendre en compte. Keseru et Makara<sup>20</sup> proposent un indice de lipophilie dépendant de l'efficacité du ligand (LELP) qui met en corrélation trois facteurs : la puissance, la taille et la lipophilie. La limite inférieure de 'LE' est de 0,3 et la plage de lipophilie est de  $-3 < \log P < 3$ , ce qui définit une plage de scores LELP optimaux de  $-10 < \text{LELP} < 10$ <sup>20</sup> :

$$\text{LELP} = \text{LogP} / \text{LE}$$

## 2. Résultats et discussion

### 2.1. Score de désirabilité d'optimisation multiparamètres du SNC

Dans notre étude, l'algorithme MPO du SNC a été appliqué à un ensemble d'inhibiteurs de l'acétylcholinestérase (AChEI) qui sont des dérivés de la pyridazine, dont leur synthèse est caractérisée par Contreras *et al.*<sup>21,22</sup>.

Les valeurs transformées (T0) des six propriétés ont été déterminées pour chaque composé ainsi que les scores de désirabilité MPO du SNC sont présentés dans le Tableau 1. L'algorithme MPO du SNC montre que 12 composés (1, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17) affichaient un score MPO du SNC élevé  $\geq 4$ . Ce score montre que ces composés ont un bon profil *drug-like*, avec une capacité de pénétrer la BHE.

L'affinité d'un composé avec des propriétés ADME souhaitable *in vitro* augmente avec l'augmentation du score de désirabilité du SNC. Les 77% de médicaments ont un score MPO du SNC  $\geq 5$ . Ces composés ont montré un alignement complet des trois attributs ADME (une haute perméabilité apparente passive (Papp), un faible passif d'efflux de la glycoprotéine P (P-gp) et un faible clairance métabolique (Clint,u) « une valeur CLint,u plus élevée suggère un métabolisme plus rapide de médicament »). Les composés 10, 12, 13, 14 et 16 ont un score élevé de MPO du SNC  $\geq 5$ , ce qui montre que ces composés ont des propriétés ADME et des attributs de sécurité souhaitables ( faible toxicité cellulaire « THLE Cv > 100µM haute viabilité cellulaire » et faible inhibition de la liaison des dofétildes) avec une bonne capacité de pénétration de la barrière hématoencéphalique.

## Etude de drug-likeness et de la relation entre la structure et l'activité/propriété d'inhibiteurs d'ACHé des dérivés de la pyridazine

**Tableau 1.** Scores de MPO du SNC et scores individuels transformés (T0) des composés anti-AChE.

Comps	LogP		LogD		TPSA		MW (u.m.a)		NHD		Pka		MPO du SNC
	Valeur	T0	Valeur	T0	Valeur	T0	Valeur	T0	Valeur	T0	Valeur	T0	
1	4.21	0.39	2.41	0.79	41.05	1.00	370.5	0.99	1	0.83	9.2	0.40	4.4
2	4.65	0.18	2.86	0.57	41.05	1.00	384.52	0.89	1	0.83	9.20	0.40	3.8
3	4.97	0.02	3.09	0.46	41.05	1.00	386.54	0.87	1	0.83	9.28	0.36	3.5
4	5.44	0.00	3.57	0.22	41.05	1.00	422.57	0.62	1	0.83	9.28	0.36	3.0
5	4.01	0.49	2.13	0.94	58.12	1.00	414.55	0.68	1	0.83	9.28	0.36	4.3
6	5.99	0.00	4.12	0.00	41.05	1.00	414.59	0.68	1	0.83	9.28	0.36	2.7
7	4.82	0.09	3.05	0.48	38.25	0.91	373.50	0.97	0	1.00	9.17	0.42	3.9
8	3.71	0.65	3.71	0.15	58.12	1.00	386.50	0.87	1	0.83	3.70	1.00	4.5
9	4.64	0.18	2.62	0.69	41.05	1.00	360.50	1.09	1	0.83	9.43	0.29	4.0
10	1.68	1.00	1.64	1.00	50.28	1.00	284.36	1.00	1	0.83	6.35	1.00	5.8
11	4.20	0.40	2.18	0.91	41.05	1.00	346.48	1.00	1	0.83	9.43	0.28	4.4
12	2.75	1.00	1.58	1.00	41.05	1.00	282.39	1.00	1	0.83	8.54	0.73	5.6
13	3.47	0.77	2.93	0.54	44.29	1.00	373.50	0.97	1	0.83	7.78	1.00	5.1
14	2.42	1.00	0.54	1.00	41.05	1.00	296.42	1.00	1	0.83	9.28	0.36	5.2
15	4.04	0.48	2.67	0.67	58.12	1.00	388.51	0.87	1	0.83	8.75	0.63	4.5
16	1.74	1.00	1.59	1.001	50.28	1.00	298.39	1.00	1	0.83	7.00	1.00	5.8
17	4.61	0.19	3.33	0.34	41.05	1.00	372.51	0.97	1	0.83	8.66	0.67	4.0

### 2.2. Étude SAR/SPR d'une série de dérivé de la pyridazine

La modification de la structure chimique des molécules bioactives dans le but d'améliorer l'efficacité de l'activité pharmacologique et des propriétés pharmacocinétiques, tels que la solubilité, la stabilité et la perméabilité, est une étape cruciale lors de la découverte des médicaments. L'étude SAR/SPR consiste à déterminer la relation qualitative entre la structure et l'activité/les propriétés<sup>23,24</sup>. Dans notre étude, nous avons appliqué l'étude qualitative SAR/SPR sur une série de composés anti-acétylcholinestérase (Figure 1).

Les paramètres physico-chimiques impliqués sont : la surface moléculaire (SAG), le volume molaire (MV), l'énergie d'hydratation (HE), le poids moléculaire (MW), le coefficient de partage octanol/eau (LogP), la surface polaire (PSA), le nombre de liaisons rotatives (NRB) et les liaisons hydrogène donateurs/accepteurs de (HBD, HBA) (Tableaux 2 et 3).

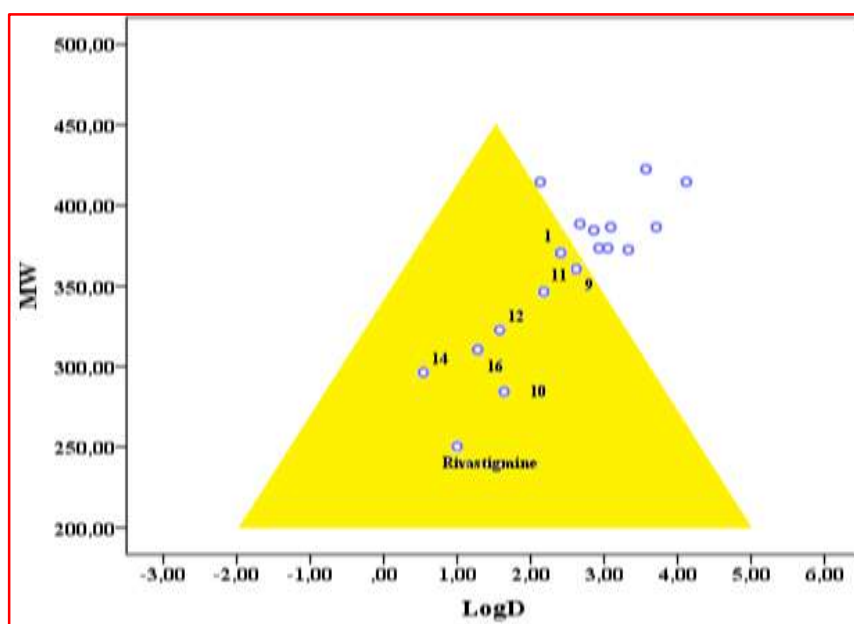
Le poids moléculaire (MW) est l'un des facteurs qui déterminent la perméabilité des médicaments. Les médicaments du SNC ont un poids moléculaire significativement réduit par rapport à d'autres composés thérapeutiques. Des molécules médicamenteuses avec  $MW \leq 450$  u(u.m.a) suggèrent qu'elles ont une meilleure pénétration à travers la BHE et une bonne absorption orale<sup>25,26</sup>. Ce qui est le cas pour notre série de composés (Tableau 1).



## Etude de drug-likeness et de la relation entre la structure et l'activité/propriété d'inhibiteurs d'ACHé des dérivés de la pyridazine

La règle de Waring<sup>27</sup> montre que le coefficient de distribution à pH 7,4 ( $\log D_{7,4}$ ) et le poids moléculaire (MW) ont un impact important sur les propriétés ADME. Selon la règle de Waring, il y a 74 % de chances d'atteindre les objectifs de haute perméabilité pour les composés avec  $MW < 414$  u et  $\log D_{7,4} > 1.3$ <sup>28</sup>.

Le triangle d'or<sup>29</sup> est un outil de visualisation qui aide les chimistes médicinaux à déterminer la perméabilité et la clairance *in vitro* de composés avant de les synthétiser. En traçant le poids moléculaire par rapport au coefficient de distribution à pH 7,4 ( $\log D_{7,4}$ ) pour une série de molécules. La probabilité de succès dans la maximisation de la puissance, de la stabilité et de la perméabilité des molécules bioactive est réalisée par déplacement des propriétés de conception dans une zone avec une ligne de base de  $\log D_{7,4} = -2,0$  à  $5,0$  et  $MW = 200$  u et une pointe entre  $\log D_{7,4} = 1,0$  et  $2,0$  et  $MW = 450$  u. Ces limites donnent une forme triangulaire, appelée triangle d'or (*Golden Triangle*). Pour notre série de molécules, les composés 1, 9, 10, 12, 14 et 16 sont situés à l'intérieur du triangle d'or (Figure 3) et ils se trouvent aussi dans les limites suggérées par Waring<sup>27</sup> cité précédemment (Tableau 1,2). Par conséquent, ces cinq composés ont une bonne perméabilité et une faible clairance.



**Fig. 3.** Ensemble de molécules avec leurs valeurs log D et MW. Les molécules (Cercles bleus) situées dans le triangle d'or ont une bonne perméabilité.

Le calcul de l'énergie d'hydratation (HE) dépend du type d'atome des groupes moléculaires, qui peuvent être donneurs ou accepteurs des liaisons hydrogène. La valeur de l'énergie d'hydratation la plus élevée a été observée pour le composé 14 (11,320 kcal/mol).

## Etude de drug-likeness et de la relation entre la structure et l'activité/propriété d'inhibiteurs d'AChE des dérivés de la pyridazine

Le composé 6 a la valeur de l'énergie d'hydratation la plus faible (6.310 kcal/mol) (Tableau 2). Le composé 14 (Figure 4) possède une liaison hydrogène donneur HBD (NH) et quatre liaisons hydrogène accepteurs HBA (trois accepteurs d'azote cyclique, NH). Le nombre des liaisons hydrogène est en corrélation avec le nombre d'hétéroatomes. Pajouhesh *et al.*<sup>30</sup> ont démontré que les médicaments du SNC ont 5 hétéroatomes ou moins (2 à 5 atomes d'azote et 0 à 2 atomes d'oxygène)<sup>30</sup> Ce qui suggère que l'augmentation de la liaison hydrogène diminue la pénétration de la BHE<sup>15</sup>.

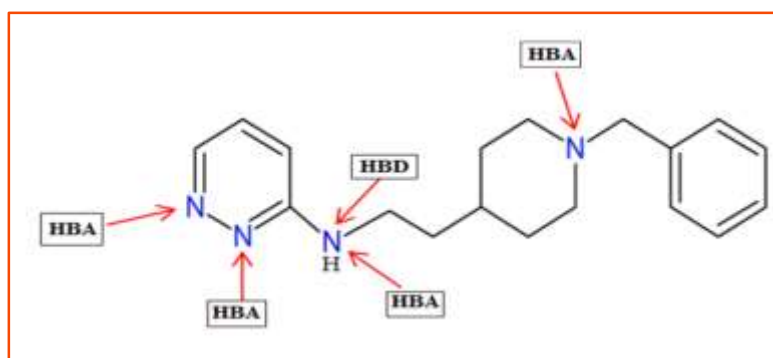


Fig.4. Sites donneurs et accepteurs du composé 14.

La lipophilie (LogP) est un facteur important dans le processus de la solubilité, des propriétés ADME et de l'activité pharmacologique. Une bonne biodisponibilité (perméabilité et solubilité) orale doit avoir un LogP modéré compris entre 0 et 3. En effet, pour un LogP trop élevé, le médicament a une faible solubilité aqueuse. Dans le cas contraire, le médicament a des difficultés à pénétrer les bicouches lipidiques des membranes cellulaires<sup>31</sup>. Pour plusieurs classes de substances actives du SNC, Hansch et Leo ont découvert que la pénétration de la barrière hématoencéphalique est optimale lorsque les valeurs LogP se situent entre 1,5 et 2,7<sup>32</sup>. Les composés 10, 12, 14 et 16 ont des valeurs LogP (respectivement 1,68, 2,75, 2,42 et 1,3) (Tableau 2) qui montre qu'ils sont considérés comme des composés médicamenteux du SNC en raison de leur meilleure pénétration cérébrale<sup>33</sup>. Ils peuvent être administrés par voie orale avec une bonne perméabilité intestinale<sup>34</sup>.

La surface polaire (PSA) d'une molécule est actuellement définie comme la somme des surfaces occupées par les atomes polaires (l'oxygène et l'azote). La limite supérieure de la PSA pour qu'une molécule puisse pénétrer le cerveau est d'environ 90Å<sup>30,35</sup>. La PSA a été utilisée pour calculer le pourcentage d'absorption (%ABS) en fonction de la formule d'équation suivante : (%ABS = 109 ± 0,345TPSA). Les valeurs TPSA ont été trouvées dans la gamme de 58,12-38,25 (Tableau 2). Nous pouvons évidemment observer que tous les composés

## Etude de drug-likeness et de la relation entre la structure et l'activité/propriété d'inhibiteurs d'AChE des dérivés de la pyridazine

présentaient un grand pourcentage d'absorption %ABS allant de 88,949 à 94,838%. Ce qui indique qu'ils devraient avoir une bonne perméabilité de la membrane plasmique cellulaire.

Le nombre de liaisons rotatives (NRB) exprime la flexibilité d'une molécule<sup>16</sup>. La flexibilité moléculaire est liée à la facilité avec laquelle la molécule traverse la membrane. Le nombre de liaisons rotatives que doit avoir un médicament du SNC pour être efficace sont inférieur à 8 NBR<sup>30</sup>. Tous les composés de notre série d'inhibiteurs de l'AChE ont un NBR < 8, à l'exception des composés 9 et 11 qui ont respectivement un NBR de 10 et de 9.

**Tableau 2.** Paramètres physico-chimiques des dérivés de la pyridazine.

Composés	MV (Å <sup>3</sup> )	SAG (Å <sup>2</sup> )	HE (kcal/mol)	LogD	ABS%	pIC <sub>50</sub>
1	1125.910	654.050	-7.890	2.41	94.838	8
2	1156.590	668.590	-8.640	2.86	94.838	7.41
3	1199.420	699.040	-8.100	3.09	94.838	7.68
4	1269.240	730.050	-8.900	3.57	94.838	6.36
5	1256.520	733.920	-9.140	2.13	88.949	7.27
6	1295.910	752.850	-6.310	4.12	94.838	5.52
7	1149.010	673.330	-8.610	3.05	95.804	6.85
8	1153.150	679.560	-10.360	3.71	88.946	3.92
9	1133.210	645.440	-9.880	2.62	94.838	6.13
10	874.10	528.890	-10.930	1.64	91.653	3.10
11	1091.770	626.210	-9.680	2.18	94.838	4.96
12	897.280	538.840	-8.380	1.58	94.838	4.15
13	1147.60	670.790	-10.030	2.93	93.719	5.82
14	940.820	557.820	-11.320	0.54	94.838	6.52
15	1157.060	644.330	-8.720	2.67	88.949	5.38
16	928.170	562.930	-10.600	1.59	91.653	3.39
17	1131.680	655.170	-9.820	3.33	94.838	4.62

### 2.2.1. Drug-likeness et indices de la lipophilie

L'objectif principal de cette étude est d'évaluer la biodisponibilité orale de dix-sept dérivés de la pyridazine (Figure 1). Une biodisponibilité orale élevée est souvent une considération non-négligeable pour le développement des molécules bioactives<sup>36,37</sup>. D'après le Tableau 3, tous les composés sont conformes aux règles de Lipinski et de Veber, ce qui suggère qu'ils sont associés avec une solubilité aqueuse et une perméabilité intestinale acceptables. Ces caractères pharmacocinétiques sont les premières étapes de la biodisponibilité orale.

## Etude de drug-likeness et de la relation entre la structure et l'activité/propriété d'inhibiteurs d'AChE des dérivés de la pyridazine

L'indice de l'efficacité de ligand lipophile « **LipE** » et celui de la lipophilie dépendant de l'efficacité du ligand « **LELP** » ont une relation avec l'espace de sécurité *in vitro*<sup>14</sup>. Si la lipophilie d'un ligand est trop élevée, il y a une plus grande probabilité qu'il se lie à des cibles différentes de celles souhaitées et par conséquent, le risque de toxicité est plus grand<sup>19</sup>. L'efficacité de la lipophilie des ligands LipE « **LLE** » permet de sélectionner des composés puissants ayant une faible lipophilie, ce qui rend l'interaction du ligand, avec le récepteur souhaité possible. 90 % des médicaments ont une valeur LipE supérieur à 3,3<sup>12</sup>. Les résultats du Tableau 3 montrent que les composés 1, 4 et 14 ont la plus haute valeur de LipE et ont donc été considérés comme les composés optimaux.

Le score LELP optimal est compris entre -10 et 10. Les composés 10, 12, 14 et 16 ont des valeurs de LELP situés dans cet intervalle. D'autre part, les composés ayant une valeur de LELP inférieur à 16,5 sont conformes aux règles de Lipinski (score ROF = 4)<sup>21</sup>. À l'exception des composés 4 et 6, qui ont respectivement un LELP de 19,706 et de 24,028 et qui sont donc en accord avec leur faible score ROF < 4.

**Tableau 3.** Paramètres d'aptitude au médicament et indices de la lipophilie des dérivés de la pyridazine.

Comp.	Règle de Lipinski					Règle de Veber		Indices de la lipophilicité				
	MW <500	LogP <5	NHD <5	NHA <10	ROF-score	NBR <10	TPSA <140	LE -	LipE -	LELP	<i>pIC</i> <sub>50</sub>	NH
1	370.500	4.210	1	4	4	5	41.050	0.400	3.790	10.525	8	28
2	384.520	4.650	1	4	4	5	41.050	0.358	2.760	12.999	7.41	29
3	386.540	4.970	1	4	4	7	41.050	0.371	2.710	13.405	7.68	29
4	422.570	5.440	1	4	3	7	41.050	0.276	0.870	19.706	6.31	32
5	414.550	4.010	1	5	4	8	58.120	0.328	3.260	12.214	7.27	31
6	414.590	5.990	1	4	3	7	41.050	0.249	-0.470	24.028	5.52	31
7	373.500	4.820	0	4	4	7	38.250	0.343	2.030	14.073	6.85	28
8	386.500	3.710	1	4	4	6	58.120	0.189	0.210	19.605	3.92	29
9	360.500	4.640	1	4	4	10	41.050	0.330	1.490	14.057	6.13	26
10	284.360	1.680	1	5	4	5	50.280	0.207	1.420	8.129	3.1	21
11	346.480	4.200	1	4	4	9	41.050	0.267	0.760	15.726	4.96	26
12	382.390	2.750	1	4	4	5	41.050	0.277	1.400	9.939	4.15	21
13	373.500	3.470	1	5	4	7	44.290	0.291	2.350	11.924	5.82	28
14	296.420	2.420	1	4	4	6	41.050	0.507	4.100	4.772	6.52	18
15	386.500	4.040	1	4	4	6	58.120	0.259	1.340	15.555	5.38	29
16	298.390	1.300	1	5	4	6	50.280	0.216	2.090	6.026	3.39	22
17	372.510	4.610	1	4	4	7	41.050	0.231	0.010	19.957	4.62	28

*pIC*<sub>50</sub> = Log(1/*IC*<sub>50</sub>) in nM<sup>22,23</sup>

### 3. Conclusion

Les inhibiteurs de l'AChE sont indiqués pour le traitement des troubles légers à modérés de la maladie d'Alzheimer. Pour identifier des composés « *drug-like* » ayant de bonnes chances de devenir des agents du SNC commercialisés, les propriétés physico-chimiques d'une série de 17 inhibiteurs de l'AChE ont été analysées et comparées avec le médicament cholinergique « Rivastigmine »<sup>37</sup>. L'utilisation du score de MPO du SNC, la règle de Lipinski, la règle de Veber, le triangle d'or, les indices de la lipophilie et les approches SAR/SPR (Tableau 4) ont montré que les composés 10, 12, 14 et 16 ont une meilleure pénétration à travers la barrière hématoencéphalique, une bonne perméabilité intestinale et biodisponibilité orale, un profil ADME optimal et un espace de sécurité favorable. Le médicament « Rivastigmine » répond également à ces règles. Les composés précédemment sélectionnés sont susceptibles d'obtenir des résultats *in vitro*. Ces résultats aident les chercheurs à concevoir des médicaments anti-acétylcholinestérase efficaces et puissants.

Tableau 4. Le résumé des résultats obtenus.

Comps	CNS MPO	SAR/SPR							Drug-likeness		lipophilicité
	- ≥5	Triangle d'or -	Règle de Waring -	LogP 1.5-2.7	HBD ≤2	Hétéro-atom ≤5	TPSA ≤76	NRB <8	Score-ROF 4	Règle de Veber -	LELP -10< et <10
1		X	X		X	X	X	X	X	X	
2			X		X	X	X	X	X	X	
3			X		X	X	X	X	X	X	
4					X	X	X	X		X	
5					X	X	X	X	X	X	
6			X		X	X	X	X		X	
7			X		X	X	X	X	X	X	
8			X		X	X	X	X	X	X	
9		X	X		X	X	X	X	X	X	
10	X	X	X	X	X	X	X	X	X	X	X
11		X	X		X	X	X	X	X	X	
12	X	X	X	X	X	X	X	X	X	X	X
13	X		X		X	X	X	X	X	X	
14	X	X	X	X	X	X	X	X	X	X	X
15			X		X	X	X	X	X	X	
16	X	X	X	X	X	X	X	X	X	X	X
17			X		X	X	X	X	X	X	
Rivastigmine	X	X	X	X	X	X	X	X	X	X	-

X : Le composé satisfait à la règle. Colonne vert : les composés sont susceptibles d'obtenir des résultats *in vitro*.

N.B : Il n'est pas possible de calculer le LEPE du médicament « Rivastigmine », car leur activité IC<sub>50</sub> doit être mesurée par le même test que les autres composés, avec des conditions expérimentales identiques

## **Partie 2**

**Étude QSAR basée sur un réseau de neurones  
artificiels pour le criblage virtuel de nouveaux  
inhibiteurs puissants de l'acétylcholinestéra**

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

Dans cette étude, nous avons mis en œuvre un flux de travail informatique en se basant sur une série d'inhibiteurs de l'acétylcholinestérase pour trouver les caractéristiques structurales nécessaires qui influencent leur activité biologique et pour la recherche de nouveaux composés, dit « *hits* », qui ont une activité inhibitrice de l'AChE.

Les relations structure-activité quantitatives (QSAR) visent la recherche d'une relation entre les molécules de grandes bases de données et leurs paramètres biologiques pour développer un modèle statistique. L'étude QSAR se repose sur un principe qui affirme que la différence de propriétés structurales est responsable de variations observées dans la réponse biologique des molécules<sup>38</sup>. La qualité d'un modèle est déterminée par son aptitude à l'ajustement et à la prédiction<sup>39,40</sup>. L'un des objectifs importants de la modélisation QSAR est la prédiction de l'activité biologique de nouveaux composés chimiques qui entrent dans le domaine d'applicabilité du modèle développé<sup>41</sup>. L'approche QSAR la plus populaire est basée sur la régression linéaire multiple (MLR). Cependant, dans de nombreux cas, cette méthode présente certaines limites et donne des résultats statistiques médiocres<sup>42</sup>. Il existe un grand nombre d'études suggérant que les méthodes non linéaires peuvent offrir des performances prédictives importantes dans la modélisation. L'application des méthodes non linéaires ont montré un intérêt considérable important dans l'étude QSAR. La méthode non linéaire du réseau de neurones artificiels (ANN) prend en compte toutes les relations non linéaires entre les descripteurs physico-chimiques et l'activité biologique correspondante. En effet, l'un des attributs les plus notables de l'ANN est sa capacité à faire des prédictions sur les nouvelles données avec un degré de précision similaire à celui obtenu avec la base d'apprentissage<sup>43</sup>. Les descripteurs apparaissant dans le modèle QSAR sont destinés au criblage virtuel de nouvelles molécules qui ont une activité puissante et améliorée<sup>44,45</sup>.

Le but principal de notre recherche est de construire un modèle QSAR-ANN capable d'expliquer la relation entre l'activité inhibitrice de l'AChE des dérivés de la pyridazine et leurs descripteurs structurels. Initialement, une analyse de régression linéaire multiple (MLR) a été utilisée pour sélectionner les descripteurs les plus pertinents avec l'activité anti-AChE. Par la suite, ces descripteurs ont été utilisés pour la création d'un modèle QSAR-ANN. Puis, le modèle QSAR développé a été utilisé pour sélectionner de nouveaux inhibiteurs d'AChE en fonction des caractéristiques chimiques considérées comme statistiquement significatives pour l'activité biologique.

# Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

## 1. Matériels et méthodes

L'objectif principal de notre étude est de développer un modèle QSAR suffisamment robuste sur le plan statistique pour prédire de manière précise et fiable l'activité biologique de nouveaux inhibiteurs de l'AChE. Pour atteindre cet objectif et pour obtenir un modèle QSAR applicable dans le plan expérimental, nous nous sommes concentrés sur les cinq principes adoptés par l'OCDE (Organisation de coopération et de développement économique): (i) un point final défini; (ii) un algorithme non ambigu; (iii) un domaine d'applicabilité défini; (iv) des mesures appropriées de la qualité d'ajustement, de la prédictivité et (v) une interprétation mécaniste (Schéma 1)<sup>46</sup>.

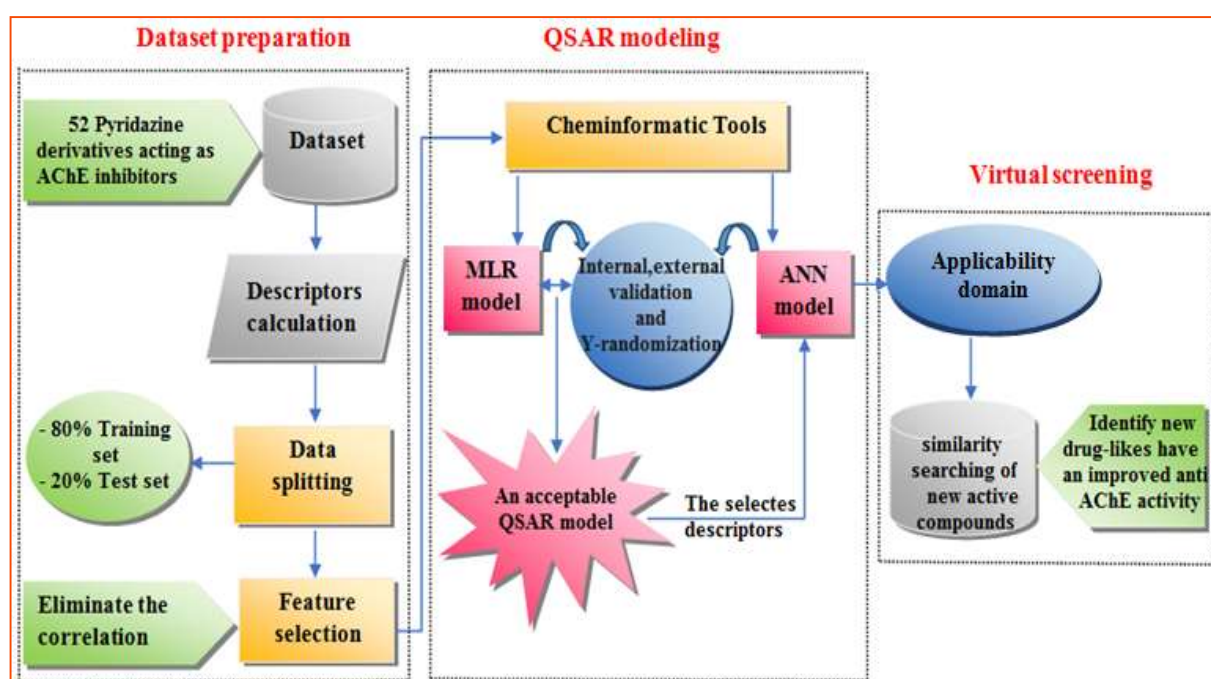


Schéma 1. Représentation schématique du flux de travail suivi dans cette étude.

### 1.1. Logiciels

- Chimique: MarvinSketch (17.27.0), ChemOffice Ultra 2004, HyperChem 7.0, Gaussien 09W.
- Statistique : Matlab (2017), XLSAT2017, MLR Y-Randomization Test 1.2., Applicability domain toolbox Matlab,1.0.

### 1.2. Ensemble de données

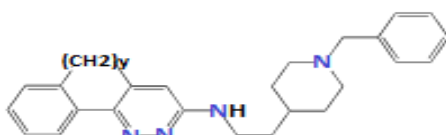
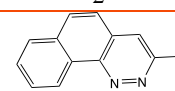
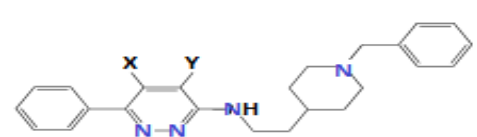
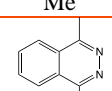
Les inhibiteurs de l'AChE des dérivés de la pyridazine sont synthétisés et caractérisés par la méthode décrite par Contreras *et al.*<sup>22,21</sup> (Tableau 5). La bioactivité de ces inhibiteurs a été exprimée sous forme de valeurs de IC<sub>50</sub> en nM, c'est la concentration du composé nécessaire pour inhiber à moitié l'activité de l'AChE (*The half-maximal inhibitory concentration IC50*).



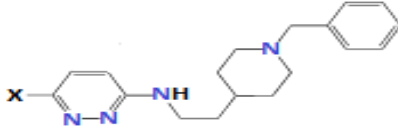
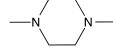
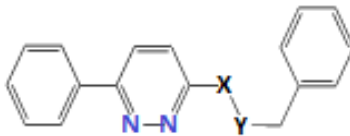

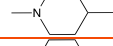
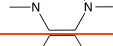
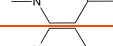
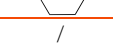

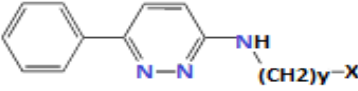
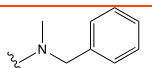
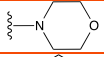
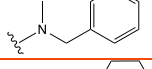
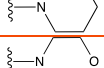

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

Les descripteurs moléculaires utilisés dans notre analyse QSAR sont présentés dans le Tableau 6. Pour calculer ces descripteurs théoriques, nous avons utilisé le logiciel HyperChem (ver.8.08)<sup>9</sup> pour faire une pré-optimisation géométrique pour les structures chimiques des molécules à l'aide de champs de force mécanique moléculaire (MM+). Après cela, les structures minimisées obtenues précédemment ont été optimisées et affinées en utilisant la méthode semi-empirique PM3 par le même logiciel. Par la suite, les conformères à basse énergie ont été utilisés pour calculer les descripteurs moléculaires à l'aide de MarvinSketch 17.13.0<sup>10</sup>, ChemOffice<sup>47</sup> et des propriétés du module QSAR (ver. 8.0.6) intégrés au logiciel HyperChem. Les composés ont été exportés vers le logiciel Gaussian 09W<sup>48</sup> pour les optimiser par la méthode de la théorie de la densité fonctionnelle (B3LYP / 6-31G (d) DFT)<sup>49,50</sup> afin de calculer les descripteurs quantiques. Par conséquent, 26 descripteurs théoriques ont été calculés pour chaque composé anti-AChE de l'ensemble des données de 52 composés.

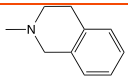
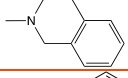
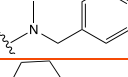
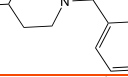
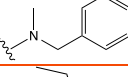
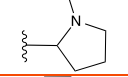
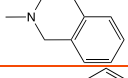
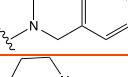
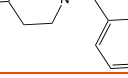

**Tableau 5.** Structures chimiques et l'activité anti-AChE des composés étudiés.

Composés	X	Y	pIC <sub>50</sub> (Exp.)*	pIC <sub>50</sub> (Pred.)	
				MLR	ANN
					
1 <sup>t</sup>		1	8.00	6.85	7.88
2		2	7.41	7.05	7.40
42			7.66	7.04	7.65
43		3	7.66	7.27	7.64
					
3 <sup>t</sup>	Me	H	7.68	6.58	6.77
4 <sup>t</sup>		/	6.36	7.29	7.23
22	i-Pr	H	7.21	6.76	6.82
25	Et	H	7.57	6.51	7.30
29	H	i-Pr	6.37	6.78	6.37
38	H	Me	6.49	6.58	6.77
51	H	H	6.92	6.05	6.91

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

					
5 <sup>t</sup>	3-Ac-Ph	/	7.27	7.21	7.45
6 <sup>t</sup>	2,4,6-(Me) <sub>3</sub> -Ph	/	5.52	6.95	6.20
14	H	/	6.52	5.44	6.53
20 <sup>t</sup>	4-CN-Ph	/	7.00	7.02	7.00
23	2-Cl-Ph	/	7.10	7.04	7.09
24	2-Me-Ph	/	7.05	6.58	6.77
26	4-(NMe <sub>2</sub> )-Ph	/	6.68	7.09	6.68
27	2-MeO-Ph	/	6.96	6.85	6.96
28	4-F-Ph	/	6.46	6.76	6.82
37	3,4-OCH <sub>2</sub> O-Ph		7.14	6.03	7.14
39	3-pyridinyl	/	7.24	6.75	7.24
41 <sup>t</sup>	MeO	/	6.66	5.74	6.91
44 <sup>t</sup>	3,5-(CF <sub>3</sub> ) <sub>2</sub> -Ph	/	7.25	8.21	6.94
45	2-naphthyl	/	7.25	7.29	7.23
46	Cl	/	7.14	7.53	7.14
47	2-Et-Ph	/	7.06	6.51	7.30
51	3-AcNH-Ph	/	7.24	7.09	7.25
52	2-thiophenyl	/	7.01	7.65	7.00
					
7	OCH <sub>2</sub> CH <sub>2</sub>		6.85	7.08	6.84
8	NHCH <sub>2</sub> CO		3.92	3.97	3.93
13	NHCH <sub>2</sub> CH <sub>2</sub>		5.82	4.37	5.83
17	NHCH <sub>2</sub> CH <sub>2</sub>		4.62	5.32	4.64
32	NHCOCH		4.74	4.08	4.75
36	NHCOCH <sub>2</sub>	/	4.77	5.90	4.81
48	SCH <sub>2</sub> CH		7.20	6.82	7.18
					
9 <sup>t</sup>		5	6.13	5.43	6.13
10 <sup>t</sup>		2	3.10	2.48	3.25
11		4	4.96	5.51	4.96
12 <sup>t</sup>		2	4.15	4.09	3.30
16		3	3.39	3.22	3.40

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

18		3	4.46	4.84	4.46
19		5	5.00	5.78	5.01
21		2	3.77	4.23	3.78
30		1	5.25	6.31	5.24
31		4	4.82	5.72	4.82
33		1	4.21	4.75	4.22
34		2	4.08	4.18	4.08
35		3	4.89	4.99	4.90
40		0	4.19	5.31	4.19
49	NHCH <sub>2</sub> CO		4.82	4.56	4.82

\*pIC<sub>50</sub>Exp en  $\mu\text{M}^{22,23}$ ; †: test.

**Tableau 6.** Descripteurs calculés et leurs types.

Descripteurs	Type
Polarisabilité ( $\alpha$ ), Surface topologique polaire (TPSA), Grille de surface (SAG) Indice de Wiener ( $W$ ), Indice de Balaban ( $J$ ), Connectivité totale ( $TC$ ).	<b>Topologique</b>
Coefficient de partage ( $LogP$ ), coefficient de distribution ( $LogD$ ), Coefficient de Hansch ( $\pi$ ), Réfractivité molaire ( $MR$ ), Énergie d'hydratation ( $HE$ ), Centre de base ( $Pka$ ), Nombre d'accepteurs de liaisons H (HBA), Nombre de donneurs de liaisons H (HBA), Énergie de solvabilité (E-sol).	<b>Physico-chimique</b>
Volume molaire ( $MV$ ), Poids moléculaire ( $MW$ ), Nombre de liaisons rotatives ( $NRB$ ), Volume de Van Der Waals ( $VDWV$ ).	<b>Géométrique</b>
La plus haute énergie orbitale moléculaire occupée ( $EHOMO$ ), La plus basse énergie orbitale molaire inoccupée ( $ELUMO$ ), Moment dipolaire ( $\mu$ ), Charge nette des atomes les plus négatifs ( $Qmin$ ), Charge nette des atomes les plus positifs ( $Qmax$ )	<b>Quantique</b>
Chaleur de formation ( $H^\circ$ ), Point de fusion ( $MP$ )	<b>Thermodynamique</b>

### 1.3. Méthodes statistiques & apprentissage automatique

#### 1.3.1. Classification des données

Les méthodes de classification, appelées aussi partition des données sont appliquées à l'analyse de bases de données et à la classification des composés. Ces méthodes permettent d'assembler des objets (observations ou individus) dans des classes (clusters) de manière à ce que les objets appartenant au même cluster soient plus similaires entre eux que les objets appartenant aux autres clusters. La base de données a été divisée en deux ensembles: un ensemble d'apprentissages (*training set*) utilisé pour construire le modèle et un ensemble de test (*test set*) utilisé pour évaluer la capacité prédictive du modèle construit. Le processus d'assemblage a été effectué de manière à ce que l'ensemble d'apprentissages (de formation) capture toutes les caractéristiques de l'ensemble de données. Pour assurer une excellente prédiction, les molécules de l'ensemble de test doivent être structurellement très similaires aux molécules de l'ensemble d'apprentissages<sup>51</sup>. Dans cette étude, la base de données actuelle a été scindée à l'aide de la technique du partitionnement k-moyennes « k-means »<sup>52</sup>, dans laquelle les composés ont été classés en un nombre défini de clusters (K). De sorte que chaque cluster contienne un ensemble de composés qui sont les plus proches voisins dans l'espace chimique déterminé à l'aide du logiciel XLSTAT<sup>53</sup>. 80% des composés ont été sélectionnés comme ensemble d'apprentissage et les 20% restant comme ensemble de test<sup>54</sup>.

#### 1.3.2. Sélection des descripteurs

L'élimination de la corrélation entre les descripteurs est une étape cruciale lors de l'étude QSAR, dans le but d'éviter la redondance d'informations dans le modèle et sélectionner initialement des descripteurs importants parmi un grand ensemble obtenu à partir de calculs (mathématique ou expérimentale). La méthode d'analyse de composantes principales (ACP), disponible dans le logiciel XLSTAT, est une méthode d'analyse de données quantitatives très efficace, pour réduire la dimension de l'espace de représentation des données et rendre l'information moins redondante<sup>55</sup>. L'objectif de la méthode ACP est de transformer un ensemble de variables corrélées entre eux en un nouvel ensemble de variables non corrélées, appelées composantes principales, tel qu'un maximum d'information soit conservé. Parmi les descripteurs corrélés, ceux qui présentaient une forte corrélation avec les valeurs pIC50 ont été retenus et les autres ont été retirés de la matrice de données. Cette méthode nous permet de faire une visualisation d'observations dans un espace de deux ou trois dimensions, afin d'identifier les groupes homogènes.

### 1.3.3. Les outils statistiques utilisés pour le développement du modèle QSAR

Tout au long de ces travaux, nous avons utilisé deux types différents de méthodes : la régression linéaire multiple « multiple linear regression (MLR) »<sup>56</sup> et les réseaux de neurones artificiels « artificial neural networks (ANN) »<sup>57</sup>. Ces deux méthodes sont disponibles dans les logiciels XLSTAT<sup>53</sup> et MATLAB (ver. 9.3)<sup>58</sup>.

#### 1.3.3.1. Régression linéaire multiple (MLR)

La méthode MLR a été utilisée pour étudier et modéliser la relation linéaire entre les variables indépendantes (descripteurs) et la variable dépendante (activité pIC<sub>50</sub>). L'avantage de cette méthode est qu'elle est transparente, et que la prédiction de l'activité de nouveaux composés peut être faite facilement. La méthode MLR se base sur l'hypothèse que la propriété *Y* dépend linéairement des différentes variables (les descripteurs) *X1*, *X2*..., *Xn*, selon la relation suivante<sup>56</sup> :

$$Y = B_0 + \sum B_n X_n,$$

avec : *Y* est la variable dépendante (à expliquer ou à prédire); *Xn* sont les variables indépendantes (explicatives) ; *B*<sub>0</sub> est la constante de l'équation du modèle ; *B*<sub>n</sub> représente les coefficients de descripteurs dans l'équation du modèle.

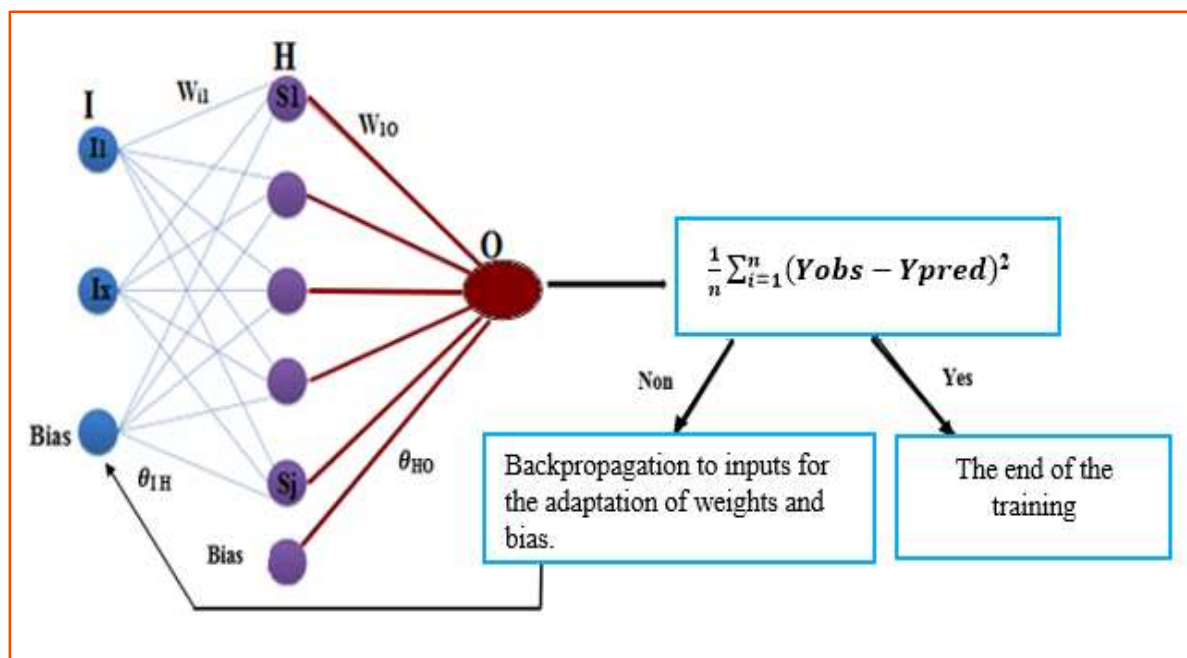
#### 1.3.3.2. Le réseau de neurones artificiels (ANN)

Un réseau de neurones artificiels est une méthode de prédiction d'inspiration biologique basées sur l'architecture d'un réseau de neurones humain<sup>59</sup>. Le schéma 2 montre une architecture typique d'un ANN. Celui-ci est formé de trois couches : une couche d'entrée (I), une couche cachée (H) et une couche de sortie (O). Chaque couche est composée d'un ou plusieurs neurones, représentés dans ce diagramme par de petits cercles. Les neurones artificiels ressemblent à des expressions mathématiques simples et sont le centre des opérations mathématiques du modèle ANN<sup>60,61</sup>. Le nombre de neurones de la couche d'entrée est égale au nombre de variables d'entrée (descripteurs) plus un (biais). Leur fonction est de recevoir les informations de la couche d'entrée et de les transférer à tous les neurones de la couche cachée. Les neurones de la couche cachée effectuent une combinaison linéaire des entrées reçues qui sont multipliées par des poids (les valeurs de *W<sub>K</sub>*), puis appliquent à cette valeur une fonction

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

d'activation « f » généralement non linéaire. La valeur obtenue Y (activité biologique) est la couche de sortie.

L'algorithme neuronal le plus couramment utilisé dans les recherches QSAR est celui de la rétropropagation « *back-propagation* » (BP)<sup>62</sup>. Le principe de cet algorithme consiste à adapter les poids  $W_K$  afin de minimiser l'erreur entre la sortie donnée par le réseau (activité calculée) et la sortie souhaitée (activité expérimentale). L'erreur obtenue par cette comparaison est ensuite rétropropagée vers la couche d'entrée en modifiant les poids des neurones. Ce processus est répété jusqu'à ce qu'une erreur négligeable soit trouvée<sup>63</sup>. La performance globale du modèle ANN final a été évaluée en termes d'erreur quadratique moyenne (MSE), sa qualité et sa robustesse ont été estimées à l'aide des paramètres statistiques.



**Schéma 2.** Architecture du réseau ANN et le principe de l'algorithme de rétropropagation, avec  $S_j = F_s \sum_{i=x}^{N_i} W_{Xj} + \theta_{IH}$   $Y_{pred} = F_1 \sum_{h=j}^{N_h} W_{jO} S_j + \theta_{HO}$ . **I**: couche d'entrée; **H**: couche cachée, **O**: couche de sortie; **S<sub>j</sub>**: entrée reçue par le neurone j de la couche cachée H;  $W_{Xj}$ : poids de la connexion entre le neurone x de la couche I et le neurone j de la couche H;  $\theta_{IH}$ : biais de la connexion de la couche I à travers la couche H;  $F_s$ : fonction d'activation non linéaire;  $Y_{pred}$ : valeur du neurone de la couche de sortie O;  $F_1$ : fonction d'activation linéaire.

### 1.3.4. Estimation de la capacité prédictive du modèle QSAR

Selon l'OCDE<sup>46</sup>, la validation est une caractéristique importante pour évaluer la fiabilité des modèles QSAR prédictifs. Pour illustrer cela, diverses mesures de validation ont été utilisées: validation interne, validation externe et test d'hasardisation (Y-randomisation)<sup>64-72</sup>.

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

Pour évaluer la qualité de l'ajustement d'un modèle QSAR, divers paramètres statistiques ont été utilisés, tels que le coefficient de corrélation ( $R$ ), le coefficient de détermination ( $R^2$ ), l'erreur standard d'estimation (SEE) et le rapport de variance ( $F$ ) (test de Fisher). Les expressions mathématiques et les valeurs de seuil de ces paramètres sont résumées dans le Tableau 10. Une valeur de  $R^2$  élevée (proche de 1) et une petite valeur de SEE impliquent qu'une corrélation plus significative a été obtenue avec un bon ajustement du modèle QSAR<sup>65</sup>.

### 1.3.4.1. Validation interne

La validation interne d'un modèle QSAR est effectuée à l'aide de l'ensemble d'apprentissage (*training test*) en utilisant une technique de validation croisée Leave-one-out (LOO)<sup>65</sup>, dans laquelle une valeur élevée du coefficient de corrélation ( $Q^2_{cv}$ ) est considérée comme une preuve de la capacité prédictive du modèle<sup>64</sup>. D'autre part, la valeur du  $Q^2_{cv}$  augmente avec l'augmentation de la valeur du dénominateur  $(\sum(Y_{obs} - \bar{Y}_{training})^2)$  (Tableau 10). Ainsi, même s'il y a une grande disparité dans les valeurs d'activité calculée et observée, des valeurs de  $Q^2_{cv}$  satisfaisantes peuvent être obtenues dans le cas où les molécules présentent une gamme considérablement large de données de réponse. Donc, une valeur acceptable de  $Q^2_{cv}$  ne garantit pas des valeurs d'activité prévues proches des valeurs observées<sup>64</sup>. Pour cela, Roy *et al.*<sup>66-68</sup> ont défini de nouvelles métriques de  $r_m^2$  pour éviter cette erreur et mieux mettre en évidence la prévisibilité du modèle, y compris la moyenne de  $r_m^2(LOO)$  (la valeur moyenne de  $r_m^2$  et  $r^2$ ) et de  $\Delta r_m^2(LOO)$  (la différence absolue entre  $r_m^2$  et  $r^2$ ) (Tableau 10). Pour une bonne prédiction du modèle QSAR,  $\Delta r_m^2(LOO)$  doit être inférieur de 0,2 et la valeur moyenne de  $r_m^2$  doit être supérieure de 0,5<sup>68</sup>.

#### 1.3.4.1.1. Validation par le test d'hasardisation (Y-randomisation test)

Le test de hasardisation (Y-randomisation test)<sup>69</sup> est une technique de validation largement utilisée pour vérifier la robustesse des modèles. Dans cette approche de validation, les valeurs de la variable cible sont redistribuées de manière aléatoire sur l'ensemble de l'apprentissage et un nouveau modèle est dérivé, l'opération étant répétée plusieurs fois. Un modèle est dit robuste si le coefficient de corrélation moyen au carré des modèles randomisés ( $R^2_r$ ) est inférieur à la moyenne du coefficient de corrélation au carré du modèle non randomisé ( $R^2$ ). Le degré de variation des valeurs de ( $R^2_r$ ) et de ( $R^2$ ) se reflète dans la valeur du paramètre  $R^2_p$ <sup>67,70</sup>, selon l'équation suivante:  $R^2_p = R^2(R^2 + R^2_r)$ . La valeur  $R^2_p$  doit être supérieure à 0,5

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

pour garantir que le modèle développé n'est pas obtenu par hasard. Les résultats de la randomisation ont été générés à l'aide du programme "MLR Y-Randomization Test 1.2"<sup>71</sup>.

### 1.3.4.2. Validation externe

La puissance prédictive réelle d'un modèle QSAR est celle de tester sa capacité à prédire l'activité des composés à partir d'un ensemble de test externe (composés non utilisés pour le développement du modèle). Le but d'un bon modèle QSAR est non seulement de prédire l'activité des composés d'ensemble d'apprentissages, mais aussi de prédire l'activité ( $pIC_{50}$ ) des molécules de test. La capacité prédictive du modèle est basée sur le coefficient de corrélation  $R^2_{pred}$  entre l'activité observée et l'activité prédite pour l'ensemble de tests. Les paramètres  $r^2_{m(test)}$  de l'ensemble de test sont similaires aux  $r^2_m$  (LOO) utilisés dans la validation interne.

Selon Golbraikh et Tropsha<sup>65</sup>, les critères suivants ont été utilisés spécifiquement pour la validation externe basée sur l'ensemble de validation

- (i)  $Q^2 > 0.5$  (ii)  $R^2_{pred} > 0.6$ .
- (ii) (iii)  $(r^2 - r_0^2 / r^2) < 0.1$  et  $0.85 \leq k \leq 1.15$  or  $(r^2 - r_0'^2 / r^2) < 0.1$  et  $0.85 \leq k' \leq 1.15$
- (iii) (iv)  $|r_m^2 - r_m'^2| < 0.3$

### 1.3.4.3. Domaine d'applicabilité

Un modèle idéal est celui qui est capable de prédire l'activité de n'importe quelle molécule imaginable. Malheureusement, cela est souvent loin d'être possible : la taille limitée du jeu d'entraînement rend l'espace chimique des modèles construits limité. Par conséquent, lorsqu'une molécule se situe en dehors de cet espace chimique, la prédiction sera plus faible. Pour prévenir ce type des problèmes, le domaine d'applicabilité (AD) permet de définir la zone dans laquelle un composé pourra être prédit avec confiance. Le domaine d'applicabilité correspond donc à la région de l'espace chimique incluant les composés du jeu d'apprentissage et les composés similaires proches dans ce même espace<sup>72</sup>. L'approche la plus courante pour déterminer le domaine d'applicabilité est décrite par Gramatica<sup>73,74</sup>. Le graphe qui représente le domaine d'applicabilité s'appelle le graphe de Williams. Il est obtenu en traçant les valeurs de la variation des résiduels standardisés de la variable indépendante R (axe de y) par rapport aux valeurs de levier h (axe de x) qui représente la distance entre les valeurs des descripteurs et leurs moyennes, calculées à l'aide du domaine d'applicabilité intégré au logiciel Matlab. Seuls les composés, dont leur valeur de levier ne dépasse pas le seuil  $h^* = 3p/n$  (où p est le nombre de



## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

descripteurs plus 1 et  $n$  est le nombre des composés), peuvent être prédit efficacement avec le modèle considéré.

### 2. Résultats et discussion

#### 2.1. Méthode du partitionnement k-moyennes « K-Means »

L'ensemble de données (52 composés) a été divisé en 11 clusters ( $K=11$ ) ayant des caractéristiques structurales homogènes (Tableau 7). À partir de chaque cluster obtenu, un composé a été sélectionné pour l'ensemble de test et le reste des composés représentent l'ensemble d'apprentissage. En conséquence, nous avons obtenu 41 composés d'ensemble d'entraînement et 11 composés d'ensemble de test. Cette méthode nous a permis de maintenir la similarité structurale entre ces deux ensembles pour construire un modèle QSAR prédictif.

**Tableau 7.** Les résultats du partitionnement k-moyennes.

Nombre de clusters	Nombre de composés dans le cluster	Partition des composés							
1	7	1	2	13	18	31	40	42	
2	6	3	23	24	25	38	47		
3	4	26	43	45	4				
4	9	8	7	5	15	27	29	32	46 50
5	3	28	22	6					
6	8	30	11	17	19	9	35	48	51
7	3	39	16	10					
8	6	33	14	21	12	34	37		
9	3	36	49	20					
10	2	52	41						
11	1	44							

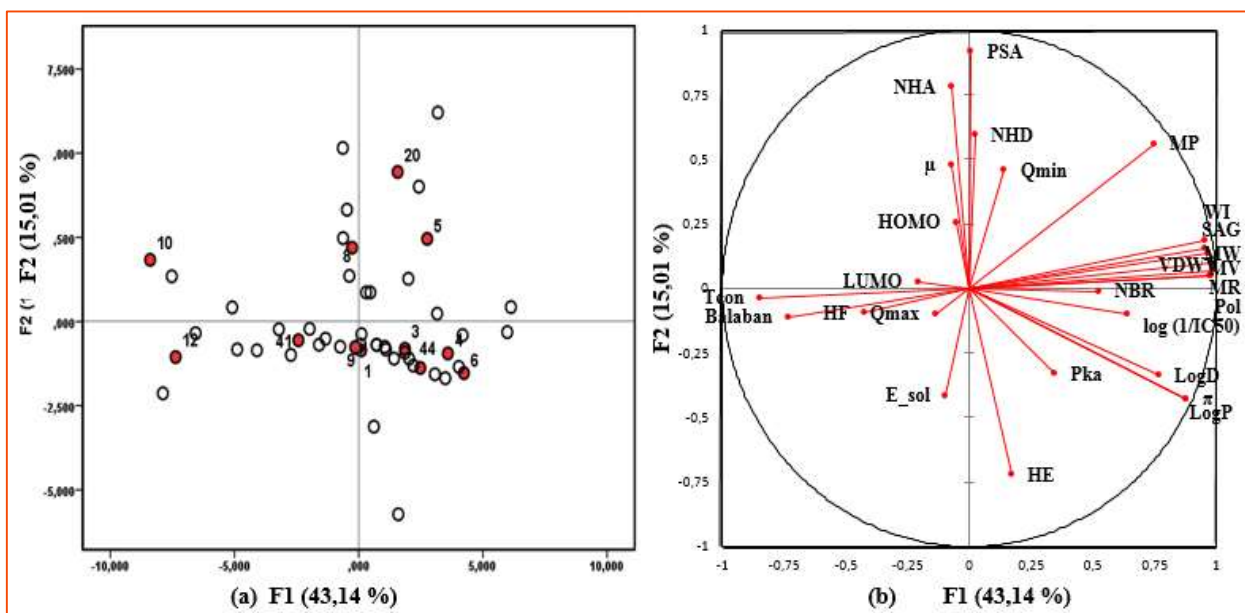
#### 2.2. Analyse en composantes principales (ACP)

Au cours de ces travaux, l'ACP a été utilisée pour déterminer la corrélation entre les descripteurs et déterminer ceux qui ont une forte corrélation avec l'activité, pour éviter le risque d'éliminer les variables dans l'espace d'entrée qui peuvent être liées à la variable de sortie du modèle QSAR. Les résultats de l'ACP ont montré que le nombre total de variables a été réduit en deux composantes principales (F1 et F2) pour rendre l'information moins redondante, avec une variance totale de 58,15 % qui montre le degré d'information représenté par ces composantes.

Le cercle de corrélation des variables (Figure 5b) montre que les descripteurs: MW, SAG,  $\alpha$ ,  $\pi$ , W, MR, MV et VDWW contiennent des informations similaires (informations redondantes), c'est-à-dire qu'il existe une corrélation entre eux. Pour éliminer cette corrélation,

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

les descripteurs SAG,  $\alpha$ ,  $\pi$ , WI, MR, MV et VDWV ont été éliminés, en dehors du poids moléculaire (MW) car il a une bonne corrélation avec l'activité ( $pIC_{50}$ ) comparé à celle des autres (Tableau 8). La projection des composés dans les deux axes F1 et F2 (Figure 5a) montre que chaque composé de l'ensemble de test est situé à proximité d'au moins un composé de l'ensemble d'entraînement dans l'espace 2D. Cette représentation confirme que les deux ensembles ont des caractéristiques structurales similaires.



**Fig. 5.** a) Projection des composés dans les deux axes, F1 et F2; les points noirs et rouges représentent l'ensemble d'apprentissage et de test, respectivement b) Cercle de corrélation des variables dans les deux axes, F1 et F2.

**Tableau 8.** Corrélation (Pearson (r)) entre l'activité et les variables indépendantes.

Variables	$pIC_{50}$	Variables	$pIC_{50}$
MW	0,607	NBR	0,310
LogP	0,583	VDWV	0,631
$\Pi$	0,583	Qmin	-0,036
PSA	-0,110	Qmax	0,116
LogD	0,229	HF	-0,201
MV	0,596	Balaban	-0,441
MR	0,632	Tcon	-0,489
Pol	0,614	MP	0,449
SAG	0,550	E_sol	0,026
HE	0,144	NBR	0,310
Pka	0,635	VDWV	0,631
HOMO	-0,082	Qmin	-0,036
LUMO	0,060	Qmax	0,116
$\mu$	-0,010	HF	-0,201
WI	0,572	Balaban	-0,441
NHD	-0,045	Tcon	-0,489
NHA	0,014	MP	0,449
		E_sol	0,026

### 2.3. Régression linéaire multiple (MLR)

La sélection de la combinaison de descripteurs ayant une bonne corrélation avec l'activité anti-AChE des composés étudiés est la prochaine étape vers la construction d'un modèle QSAR prédictif. Pour atteindre cette étape cruciale, nous avons soumis 19 descripteurs avec 42 composés d'ensembles d'entraînement. L'étude de la régression linéaire multiple descendante (MLR) a été utilisée pour le développement du modèle, qui consiste à éliminer les descripteurs un par un avec des coefficients de régression non significatifs, jusqu'à obtenir le meilleur ensemble des descripteurs. Dans notre étude, les cinq descripteurs les plus significatifs obtenus, selon la méthode MLR, sont logP, logD, MW, NHD et NBR.

La multicollinéarité augmente les erreurs types des coefficients. L'augmentation des erreurs types signifie que les coefficients de certaines variables indépendantes ne peuvent pas être significativement différents de zéro. Pour confirmer l'absence de multicollinéarité entre ces cinq descripteurs, un facteur d'inflation de la variance (**VIF**) a été calculé pour chaque descripteur de la régression<sup>64</sup>, comme suit:  $VIF=1/(1-r^2)$  où, r est le coefficient de corrélation d'une variable indépendante avec tous les autres descripteurs du modèle. Si la valeur du VIF est entre 1 et 5, le modèle correspondant est acceptable; si la valeur  $VIF > 5$ , le modèle correspondant est instable et une nouvelle vérification est nécessaire. Les valeurs VIF des cinq descripteurs sont présentées dans le Tableau 9. Ces résultats montrent que le modèle obtenu a une signification statistique.

**Tableau 9.** Multicollinéarité du modèle MLR.

Descripteurs	MW	LogP	LogD	NHD	NBR
<b>Tolérance*</b>	0.374	0.235	0.284	0.760	0.700
<b>VIF</b>	2.674	4.261	3.521	1.316	1.430

\* VIF est l'inverse de la tolérance

Le modèle de QSAR obtenu par la méthode MLR s'exprime comme suit:

$$pIC_{50} = -1.43 + 0.02MW + 1.46LogP - 1.73LogD - 0.95NHD - 0.29NBR.$$

Selon les significations statistiques indiquées dans le Tableau 10, l'équation ci-dessus montre une corrélation assez forte entre l'activité et les cinq variables explicatives, avec un coefficient de corrélation de 87 % (donc proche de 100 %). Les valeurs relativement élevées de  $R^2$  et  $R^2 > 0,6$  et une faible valeur de **SEE** indiquent que le modèle proposé est fiable et bien ajusté. L'équation de régression a une signification statistique, ce qui est confirmé par la valeur

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

élevée de **F** (les coefficients des variables sélectionnées dans l'équation ne sont pas égaux à zéro) (Tableau 10).

Concernant sa robustesse, le coefficient  $Q^2_{(cv)}$  montre que le modèle est robuste et stable (de par la stabilité de ses paramètres vis-à-vis des molécules du jeu d'entraînement) avec une bonne prévisibilité. Le test de hasardisation (Y-randomisation test) a montré que la moyenne des valeurs  $R^2_r$  et  $Q^2_{r(cv)}$ , après 10 essais aléatoires, étaient inférieure aux valeurs  $R^2$  et  $Q^2_{cv}$  du modèle original et  $R^2_p > 0,5$ , ce qui indique que le modèle développé est suffisamment robuste et que les résultats ne sont pas obtenus au hasard.

Une valeur acceptable de  $R^2_{perd} = 0,65$  signifie un bon potentiel prédictif externe de ce modèle QSAR obtenu. Toutes les métriques  $r_m^2$  de validation qualitative<sup>66-68</sup> et les critères de Golbraikh<sup>65</sup> calculés pour l'ensemble d'apprentissage et l'ensemble de test, sont conformes aux limites acceptables, ce qui reflète la fiabilité et la prédictivité du modèle MLR-QSAR développé (Tableau 10).

Après avoir confirmé la robustesse d'un modèle MLR-QSAR développé avec de bonnes qualités statistiques, nous avons utilisé les cinq descripteurs moléculaires sélectionnés dans ce modèle comme couches d'entrée pour construire un réseau de neurones artificiels (ANN) à trois couches, en utilisant une stratégie d'apprentissage BP.

**Tableau 10.** Paramètres statistiques obtenus par le modèle MLR et leurs valeurs de seuil.

Paramètres	Equation	Seuil	Score du modèle
<b>Critères d'adaptation</b>			
$R^2$	$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{calc})^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2}$	>0.6	0.75
<b>R</b>	$R = \sqrt{R^2}$	>0.6	0.87
<b>SEE</b>	$SEE = \sqrt{\frac{\sum(Y_{obs} - Y_{calc})^2}{N-p-1}}$	Une petite valeur	0.36
<b>F</b>	$F = \frac{\sum(Y_{calc} - \bar{Y}_{training})^2 / P}{\sum(Y_{calc} - \bar{Y}_{training})^2 / N-P-1}$	Une grande valeur	21.787
<b>Validation interne</b>			
$Q^2_{cv}$	$Q^2 = 1 - \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - \bar{Y}_{training})^2}$ PRESS = $\sum(Y_{obs} - Y_{pred})^2$	>0.5	0.66
$R_r$	La moyenne du coefficient de corrélation pour les données randomisées	<R	0.3
$R_r^2$	la moyenne du coefficient de détermination pour les données randomisées	< $R^2$	0.09

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

$Q^2_{r(cv)}$	La moyenne du coefficient de détermination du CV pour les données randomisées	$< Q^2_{cv}$	-0.23
$R_p^2$	Pénalise le modèle pour les petites différences entre les valeurs de $R^2$ et $R_r^2$	$> 0.5$	0.73
$\bar{r}_m^2(LOO)$	$(r_m^2 + r'_m{}^2)/2$	$> 0.5$	0.56
$\Delta r_m^2(LOO)$	$ r_m^2 - r'_m{}^2 $	$< 0.2$	0.17
<b>Validation externe</b>			
$R^2_{pred}$	$R^2_{pred} = 1 - \frac{\sum(Y_{obs(test)} - Y_{pred(test)})^2}{\sum(Y_{obs(test)} - \bar{Y}_{training})^2}$	$> 0.5$	0.65
$\bar{r}_m^2(test)$	Les valeurs moyennes de $r_m^2$ et $r'_m{}^2$ .	$> 0.5$	0.61
$\Delta r_m^2(test)$	La différence absolue entre $r_m^2$ et $r'_m{}^2$	$< 0.2$	0.15
$(r^2 - r_0^2 / r^2)$	Critères de Critolbraikh et Tropsha	$< 0.1$	0.09
$(r^2 - r_0'^2 / r^2)$	Critères de Golbraikh et Tropsha	$< 0.1$	0.003
$K$	$K = \frac{\sum(Y_{obs} \times Y_{pred})^2}{\sum(Y_{pred})^2}$	$0.85 \leq k \leq 1.15$	1
$K'$	$K' = \frac{\sum(Y_{obs} \times Y_{pred})^2}{\sum(Y_{obs})^2}$	$0.85 \leq k' \leq 1.15$	0.97

$Y_{obs}$  et  $Y_{calc}$  se référer aux valeurs observées et calculées, respectivement ;  $\bar{Y}_{obs}$  et  $\bar{Y}_{calc}$  Leurs moyennes, respectivement ;  $Y_{pred}$  : valeurs prédictives par la méthode LOO ;  $N$  : le nombre de composés ;  $P$  : le nombre de descripteurs.

$$r_m^2 = r^2 \times (1 - \sqrt{(r^2 + r_0^2)}); \quad r'_m{}^2 = r^2 \times (1 - \sqrt{(r^2 + r_0'^2)}); \quad r^2 = 1 - \frac{\sum(Y_{obs} - K \times Y_{pred})^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2}; \quad r_0'^2 = 1 - \frac{\sum(Y_{obs} - K' \times Y_{obs})^2}{\sum(Y_{obs} - \bar{Y}_{pred})^2}.$$

### 2.4. Conception et formation du modèle de réseaux de neurones artificiels (ANN)

Dans cette étude, l'activité anti-AChE a été estimée par un modèle ANN. En se basant sur l'algorithme de rétropropagation. L'ANN désirée doit avoir une généralisabilité élevée et une faible probabilité de sur-ajustement. Dans notre cas, un réseau d'ANN à quatre couches a été utilisé. Cinq neurones constituent la couche d'entrée et décrivent les cinq variables indépendantes. Un neurone, qui code l'activité anti-AChE ( $pIC_{50}$ ), constitue la couche de sortie (Figure 6). Malheureusement, il n'y a ni résultats théoriques disponibles ni règles empiriques satisfaisantes nous permettant de déterminer le nombre de couches cachées et le nombre de neurones constituant ces couches. Il suffit d'avoir une ou deux couches cachées dans le cas d'une petite base de données pour éviter le sur-ajustement des modèles. Avant le processus d'apprentissage, la base de données a été divisée en trois sous-ensembles: l'ensemble d'entraînement, qui était utilisé pour l'apprentissage des paramètres du modèle. Pendant le processus d'apprentissage, l'erreur de l'ensemble d'entraînement diminue. Afin d'éviter le sur-ajustement à chaque étape de l'apprentissage, le modèle est vérifié par l'ensemble de validation. Lorsque l'erreur de validation commence à augmenter, l'apprentissage est terminé, alors que l'ensemble de données de test était utilisé pour évaluer les performances prédictives du réseau sur de nouvelles données (Figure 6).

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

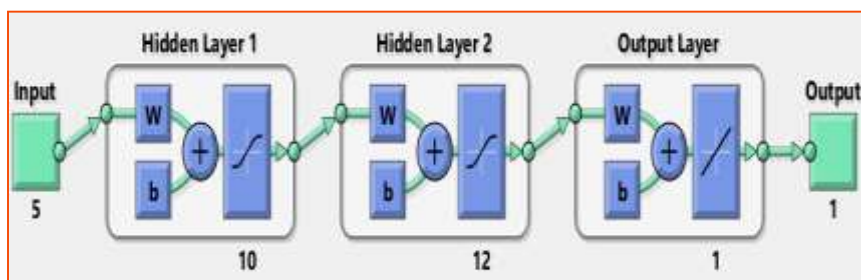


Fig.6. L'architecture du réseau d'ANN sur Matlab.

Pour déterminer la meilleure architecture ANN avec une capacité optimale de prédiction, différents modèles ANN ont été développés, formés et testés. Le modèle ANN choisi est celui qui a la plus petite valeur d'erreur quadratique moyenne (**MSE**) et une valeur élevée de coefficient de détermination **R<sup>2</sup>**. La meilleure architecture du réseau ANN trouvée après une large recherche était [5-10-12-1] avec deux couches cachées ayant 10 et 12 neurones (Figure 6). Ce réseau est caractérisé par une valeur **MSE** = 0,06 et une valeur **R<sup>2</sup>** = 0,96, ce qui indique que l'activité anti-AChE (**pIC<sub>50</sub>**) sont significativement corrélés avec les cinq variables adoptées dans ce travail. Ce réseau a une bonne capacité de reconnaissance avec des valeurs de performance faible 0,01 et 0,13 pour l'ensemble de l'entraînement et de test.

La courbe d'analyse de régression, illustrée dans la figure 7a montre la régression non linéaire entre l'activité observée et prédictive avec une valeur du coefficient de corrélation (**R**=0,98), ce qui indique que l'ajustement est raisonnablement bon pour tous les ensembles de données. La figure 7b montre la performance de l'ANN par rapport à la valeur de **MSE** pour l'ensemble d'entraînement, celui de test et celui de validation :

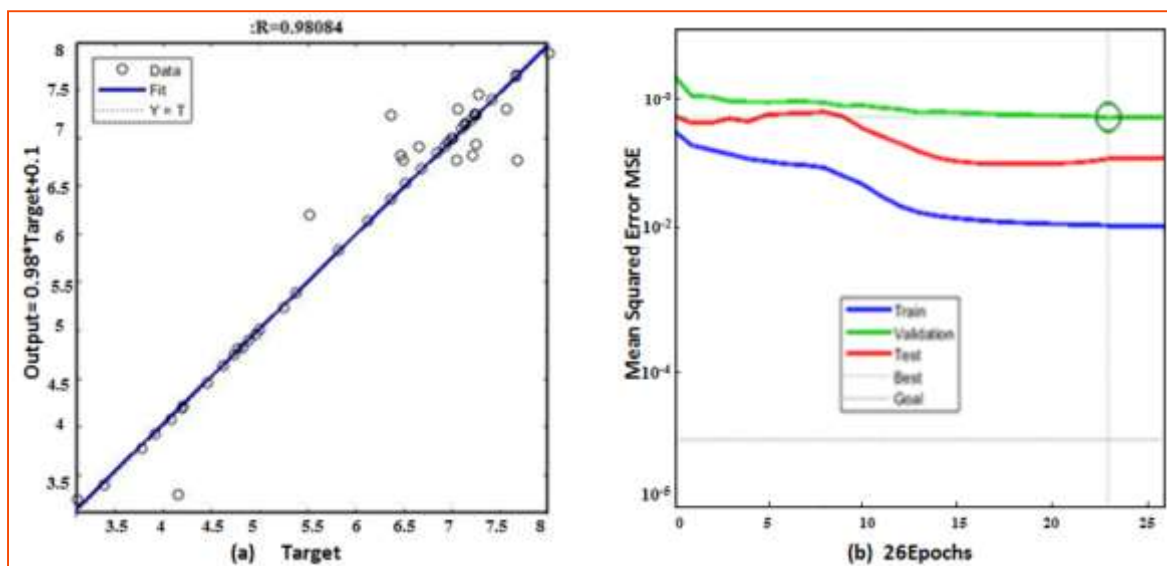
- Le graphique indique le nombre d'itérations (*epoch*) à laquelle la performance de validation a atteint le minimum. La figure n'indique aucun problème majeur avec l'apprentissage, puisqu'il y a une diminution de sa valeur MSE, et elle se stabilise lorsqu'elle atteint le point minimum (epoch 23).
- Les courbes de validation et de test ont des caractéristiques similaires. Si la courbe de test augmente considérablement avant l'augmentation de la courbe de validation, il est possible qu'il y ait eu sur-ajustement, ce qui n'est pas le cas dans notre modèle ANN.
- L'erreur quadratique moyenne a diminué en augmentant le nombre de l'*epoch*.

### 2.4.1. Évaluation de la capacité prédictive du réseau de neurones

La capacité du réseau à prédire l'activité anti-AChE a été estimée par une validation interne et externe, qui comprend : le chi-carré (**X<sup>2</sup>**) et la racine carrée de l'erreur quadratique

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

moyenne (RMSE). Ces paramètres sont utilisés pour décider si le modèle possède la qualité



**Fig.7.** (a) La courbe de la régression pour le modèle ANN. (b) Courbes de performances durant l'apprentissage, validation et le test du modèle ANN.

prédictive reflétée dans le  $R^2$ . Ils sont particulièrement utiles pour créer et valider des modèles pour les ensembles de données non linéaires<sup>40,75</sup>. Pour un bon modèle prédictif, les valeurs de chi ( $\chi^2$ ) et RMSE devraient être faibles (respectivement  $<0,5$  et  $<0,3$ ). D'après les résultats présentés dans le Tableau 11, l'architecture ANN [5-10-12-1] a été raisonnablement adéquate pour tous les ensembles de données avec une bonne capacité prédictive. Ainsi, le modèle ANN peut être utilisé de manière tout à fait satisfaisante pour le criblage d'une nouvelle série des inhibiteurs de l'AChE.

**Tableau 11.** Résultats statistiques du modèle ANN-QSAR.

Paramètres	Validation interne				Validation externe		
	$R^2$	RMSE	$\chi^2$	$Q^2_{cv}$	$R^2$	$\bar{r}_m^2$	$\Delta r_m^2$
ANN-QSAR	0.96	0.24	0.39	0.85	0.95	0.88	0.02

D'après la Figure 8, qui présente une comparaison entre la capacité prédictive des modèles MLR et ANN, la courbe d'activité prédictee par le modèle ANN est presque identique à l'activité observée, contrairement à celle du modèle MLR. Cela est confirmé par les paramètres statistiques calculés de ces modèles révélant la supériorité du modèle ANN par rapport au modèle MLR. Cette étude nous a permis de conclure que l'architecture ANN [5-10-

12-1] peut établir une relation non linéaire satisfaisante entre les caractéristiques structurales des dérivatives de la pyridazine et leur activité anti-AChE.

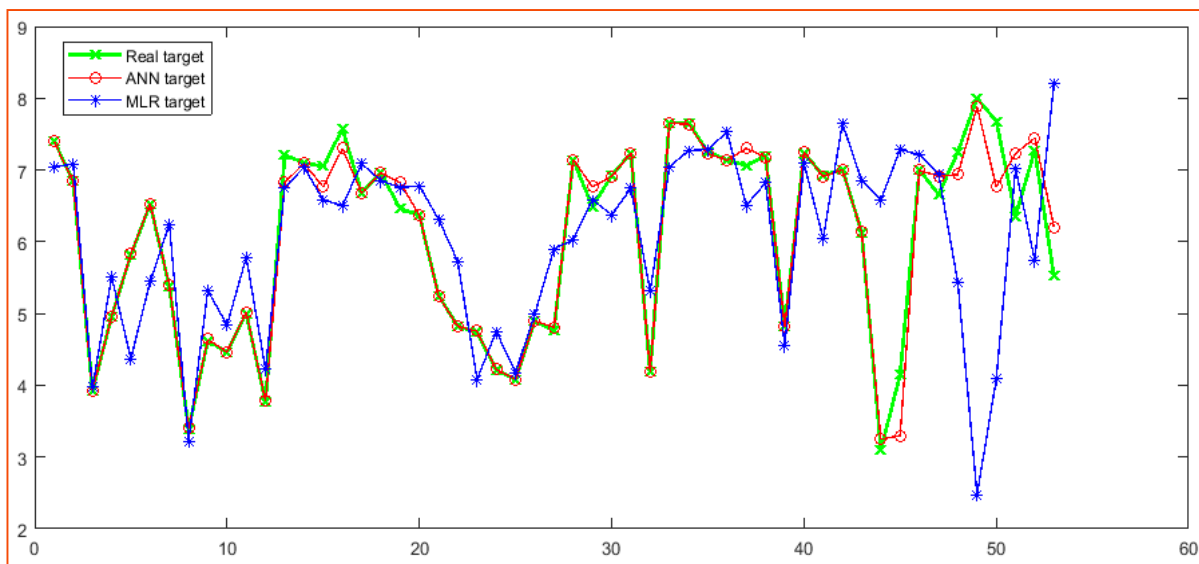


Fig.8. Comparaison de capacité prédictive des modèles MLR et ANN.

#### 2.4.2. Domaine d'applicabilité

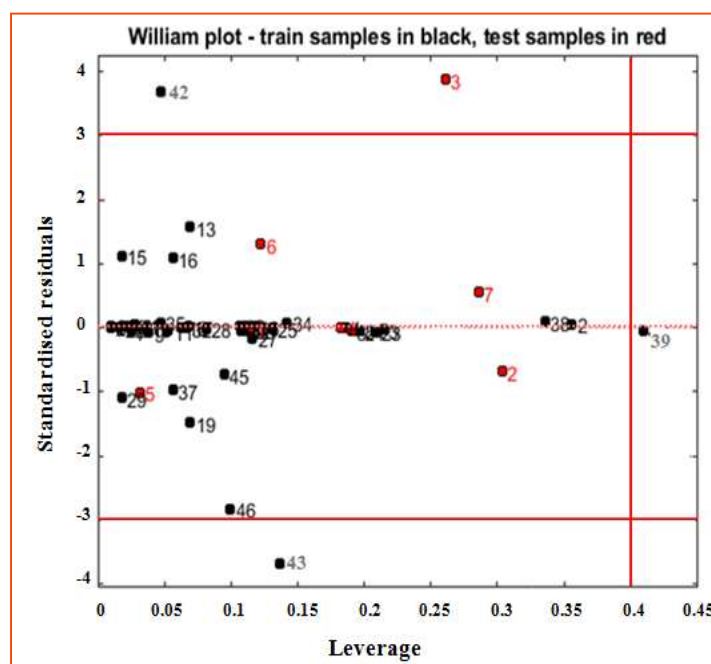
Même les modèles les plus exhaustifs, dignes de confiance et validés, ne peuvent pas prédire des propriétés de manière fiable pour l'intégralité des composés chimiques existants. Le domaine d'applicabilité (DA) permet de définir la zone dans laquelle un composé pourra être prédit avec confiance. La courbe de Williams du modèle QSAR est illustrée dans la Figure 9. La valeur de levier ( $h^*$ ) a été établie à 0,4 pour le modèle QSAR développé. D'après cette valeur, il n'y a qu'un seul composé se trouvant hors du domaine d'applicabilité défini (a une valeur  $h_i > 0,4$ ). Le DA permet de vérifier la fiabilité de la prédiction du modèle, en se basant sur l'écart type, où l'on peut trouver 3 composés aberrants. Les composés restants, qui relèvent du domaine d'applicabilité, peuvent être prédit efficacement avec le modèle ANN développé. Par conséquent, le modèle peut être appliqué avec un degré élevé de confiance dans ce domaine d'applicabilité, pour le criblage de nouveaux composés anti-AChE puissants.

#### 2.4.3. Contribution des descripteurs sélectionnés

L'un des objectifs de l'analyse QSAR est de comprendre les forces qui régissent l'activité d'une classe particulière de composés et d'aider à la conception des nouvelles molécules susceptibles d'être un médicament. Par conséquent, l'évaluation des descripteurs s'est avérée



## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase



**Fig.9.** Diagramme de Williams du résidu normalisé par rapport au levier pour le modèle ANN. Les ensembles d'entraînement et de test sont représentés par des points noirs et rouges, respectivement.

très intéressante et utile pour mieux comprendre l'activité anti-AChE. Nous avons donc décidé d'estimer leur contribution relative. Les contributions des descripteurs  $i$  ( $i=1-5$ ) ont été estimées à partir du réseau d'architecture formé [5-10-12-1]. Nous avons exclu le descripteur "i" et ses poids correspondants de l'ANN [5-10-12-1] et nous avons refait l'apprentissage de réseau avec la nouvelle architecture [4-10-12-1]. La moyenne des valeurs absolues des dérivations  $\Delta_{mi}$ , entre l'activité observée et celle estimée pour tous les composés, a été calculée. Ce processus a été répété pour chaque descripteur. La contribution  $C_i$ , du descripteur "i" est donné par Zheng

*et al.*<sup>76</sup> comme suit:

$$C_i \% = \frac{\Delta_{mi}}{\sum_{i=1}^5 \Delta_{mi}}$$

Selon la figure 10, les descripteurs relatifs au coefficient de partage (LogP) et le coefficient de distribution (LogD) sont très pertinents dans l'établissement de la relation quantitative structure-activité anti-AChE des dérivés de la pyridazine. La lipophilicité a été la première des descripteurs à être identifié comme un facteur important pour la pénétration

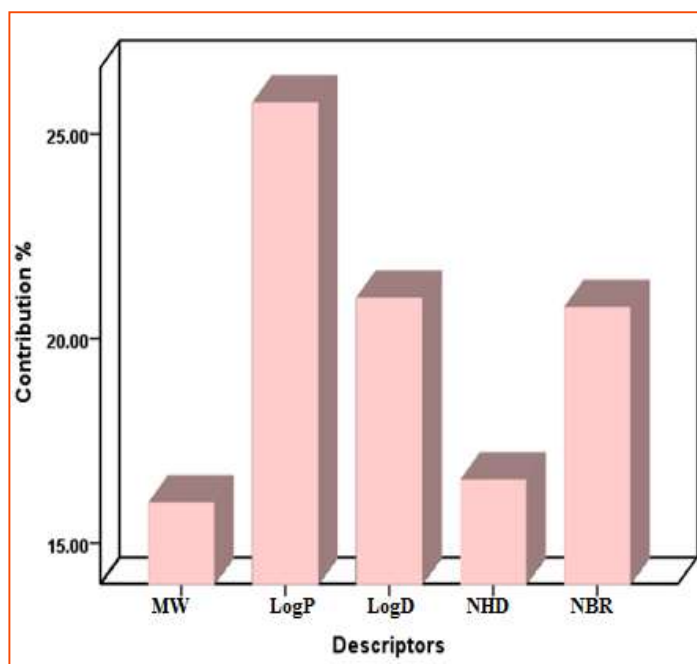
du SNC. Ces deux descripteurs (LogP et LogD) jouent un rôle important dans la détermination de la lipophilicité de composés qui ont une activité anti-AChE, pour identifier leurs capacités de pénétrer la barrière hémato-encéphalique (BHE). Pour plusieurs classes de substances actives du SNC, Hansch et Leo ont constaté que la pénétration de la BHE est optimale lorsque les valeurs LogP sont comprises entre 1,5 et 2,7<sup>30</sup>. Lorsque la lipophilie augmente, la probabilité

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

de se lier à d'autres cibles protéiques hydrophobes que celle souhaitée augmente ; par conséquent, le risque de toxicité est plus élevé. Une analyse de petites molécules de type médicament a suggéré que pour une meilleure pénétration cérébrale et pour une bonne perméabilité intestinale les valeurs LogD doivent être supérieures à 0 et inférieures à 3<sup>30</sup>.

La contribution des descripteurs montre l'importance de la flexibilité moléculaire et des liaisons d'hydrogène sur l'activité anti-AChE. La flexibilité moléculaire est liée à la facilité avec laquelle la molécule traverse la BHE. Elle a une corrélation avec le nombre de liaisons rotatives et le nombre d'anneaux dans la molécule. Les médicaments du SNC ont moins de liaisons rotatives que les autres classes de médicaments. Le NBR d'un médicament efficace pour le SNC est  $< 8$ <sup>30</sup>, ce qui suggère que les liens entre les systèmes aromatiques de l'inhibiteur et l'AChE sont essentiels pour produire une haute affinité. Kryger *et al.*<sup>77</sup> ont affirmé que le médicament Donepezil (E2020) nécessite au moins deux liaisons rotatives de chaque côté de la pipéridine, dans laquelle les deux fractions aromatiques de E2020 interagissent avec l'acides aminé tryptophan (Trp86 et Trp286 ; numérotation AChE humaine).

Le NHD peut influencer de manière significative le profil d'un médicament anti-AChE. Les liaisons d'hydrogène ont une corrélation avec le nombre d'hétéroatomes. La pénétration du SNC nécessite une somme de ces hétéroatomes  $\leq 5$ , ce qui suggère que l'augmentation de la liaison hydrogène diminue la pénétration la BHE. L'équation QSAR développée détermine l'importance du poids moléculaire (MW) sur l'activité des composés étudiés. Les médicaments destinés au SNC ont tendance à avoir un  $MW \leq 450$  pour faciliter la pénétration cérébrale et l'absorption orale de la molécule<sup>26</sup>.



2.5.

Fig.10. Contribution des descripteurs dans le modèle ANN.

### Criblage virtuel des inhibiteurs de l'AChE

La similarité moléculaire est l'une des méthodes du criblage virtuel (*ligand-based*). Elle se base sur le principe de la similarité des propriétés structurales d'une base de données dont leurs activités sont inconnues, mais leur structure est similaire à celle d'une molécule dont son activité biologique est connue (structure de référence). Les mesures de similarité moléculaire tiennent en compte la forme des molécules et leurs propriétés physico-chimiques.

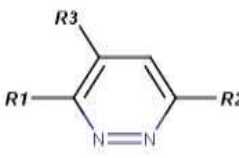
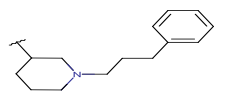
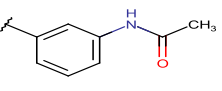
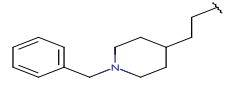
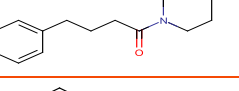
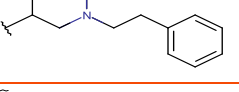
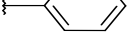
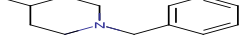
L'objectif principal du criblage virtuel dans cette étude est d'identifier de nouvelles structures ayant une activité anti-AChE améliorée et qui doit être dans le domaine d'applicabilité du modèle QSAR-ANN développé. À cette fin, un ensemble de composés a été recueilli à partir de la base de données ChEMBL<sup>78</sup>; ces composés sont similaires (d'environ 60 à 80 %) au composé 1 (Tableau 5). Ce composé s'avérera d'être l'un des plus puissants inhibiteurs de l'AChE et a un bon profil ADME<sup>79</sup>. Les composés de la base de données ChEMBL ont été filtrés selon les règles du triangle d'or, Lipinski, Veber<sup>15-17</sup> et le score MPO du SNC<sup>12</sup>. Nous avons sélectionné 19 composés ayant une bonne biodisponibilité avec des attributs ADME optimaux et une bonne capacité de pénétrer la barrière hémato-encéphalique (BHE).

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

Ensuite, l'activité anti-AChE prédictive (pIC<sub>50</sub>) de ces composés a été calculée à l'aide du modèle QSAR-ANN développé. La valeur de levier pour chaque composé a été calculée et comparée à l'effet de levier ( $h^* = 0,4$ ) (Tableau 12). Selon les résultats présentés dans le Tableau 12, deux composés 6 et 17 ont des valeurs de levier  $h_i > 0,4$  et doivent donc être exclus du domaine défini de validité. Les autres composés qui se situent dans le domaine d'applicabilité du modèle ANN (leurs  $h_i < 0,4$ ) sont donc considérés comme des *hits* ont une activité anti-inhibiteurs de l'AChE. Le composé 11 (pIC<sub>50</sub>=9,05) présente une bonne activité anti-AChE par rapport aux autres composés (Tableau 12). Contreras *et al.*<sup>21</sup>, ont démontré que pour une activité inhibitrice élevée de l'AChE, celle-ci nécessite la présence d'un cycle de la pyridazine centrale et d'une tête cationique, ce qui correspond au composé 11. Ce dernier a une activité plus puissante que celle du composé de référence (pIC<sub>50</sub>=8) et du médicament donépézil « Aricept® » (pIC<sub>50</sub>=7,32)<sup>22</sup>.

Le donépézil est le seul médicament de l'AChEI qui inhibe la double liaison de l'AChE. Il se fixe dans le site actif de l'enzyme (AChE) pour l'inhiber et il interagit avec le site périphérique (PSA) pour bloquer la formation de plaques d'Aβ<sup>80,81</sup>. Le donépézil est approuvé pour traiter les symptômes associés à la maladie d'Alzheimer<sup>82, 83</sup>.

**Tableau 12.** Activités prédictives (pIC<sub>50</sub>) et la valeur de levier (h<sub>i</sub>) de 19 *hits* identifiés.

Comp.	R1	R2	R3	pIC <sub>50</sub> (pred.)	h <sub>i</sub>
					
1	CH <sub>3</sub>		H	4.883	0.259
2			H	6.967	0.304
3	H		H	3.528	0.153
4	CH <sub>3</sub>		H	4.008	0.157
5			H	6.323	0.313

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

6			Me	6.659	0.288
7			Me	6.859	0.280
8			H	8.825	0.295
9	H		H	3.328	0.242
10	H		H	3.561	0.459
11			H	9.052	0.383
12			H	3.340	0.176
13			H	6.382	0.231
14			H	8.765	0.334
15			H	1.632	0.192
16			H	6.981	0.238
17			H	1.813	0.712
18			H	3.379	0.320
19			H	8.283	0.303

## Étude QSAR basée sur un réseau de neurones artificiels pour le criblage virtuel de nouveaux inhibiteurs puissants de l'acétylcholinestérase

### 3. Conclusion

Un réseau de neurones artificiels (ANN) a été utilisé comme méthode non linéaire pour créer un modèle QSAR et interpréter la relation non linéaire entre l'activité inhibitrice de l'AChE des dérivés de la pyridazine et leurs descripteurs structuraux sélectionnés par la méthode MLR. Les résultats confirment que le modèle non linéaire ANN a une bonne stabilité et une bonne capacité prédictive par rapport à celle du modèle linéaire (MLR). Le meilleur modèle de réseau neurone a une architecture [5-10-12-1] avec une petite valeur de MSE (0,06) et un coefficient de corrélation élevé ( $R^2 = 0,96$ ), ce qui indique que l'activité pIC<sub>50</sub> est en corrélation significative avec les 5 variables adoptées dans ces travaux. La précision et la prévisibilité de ce modèle, nous ont permis de l'utiliser d'une manière très satisfaisante pour le criblage d'une nouvelle série de inhibiteurs de l'AChE.

Enfin, nous avons combiné le domaine d'applicabilité du modèle QSAR-ANN avec le criblage virtuel de la base de données ChEMBL, en se basant sur la méthode de similarité moléculaire afin de découvrir de nouveaux inhibiteurs de l'AChE. Le criblage virtuel fondé sur le modèle QSAR-ANN mis au point a clairement atteint son objectif en identifiant les analogues potentiels de la pyridazine comme inhibiteurs de l'AChE. Par conséquent, le *hit* 11 s'est avéré être l'inhibiteur le plus puissant avec une valeur de pIC<sub>50</sub>=9,05 et il a montré une activité plus puissante que celle du donépézil (Aricept®).

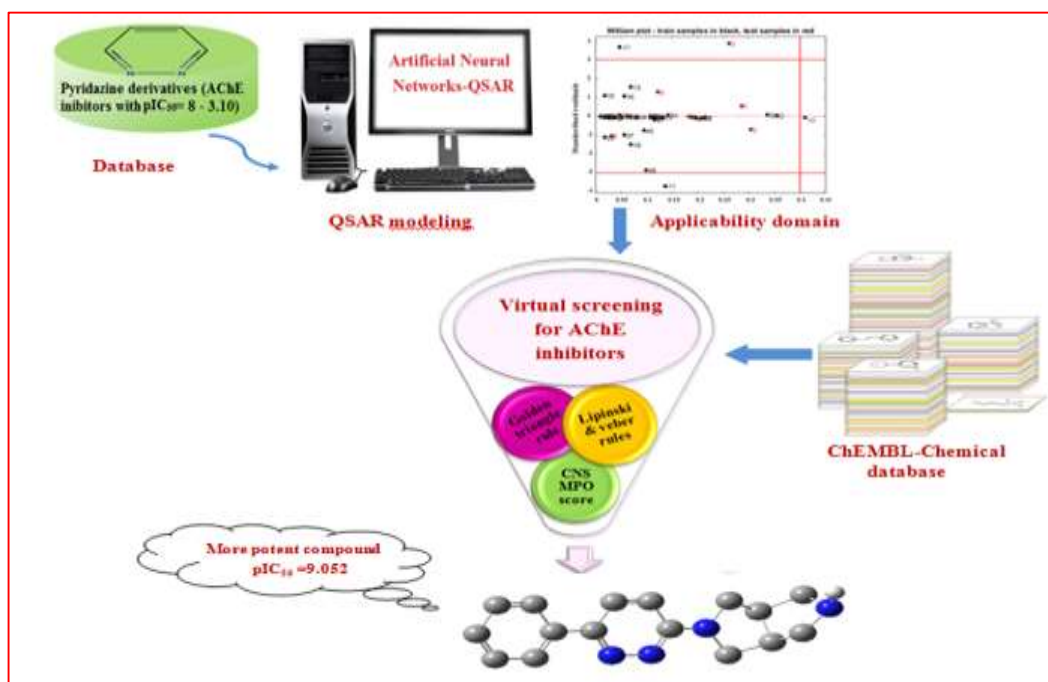


Fig.11.Représentation schématique du résumé de travail réaliser.

## **Partie 3**

**Analyse comparative des contacts intermoléculaires  
basés sur le docking (3D-QSAR) pour le criblage *in  
silico* de nouveaux inhibiteurs puissants de CDK5**

## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

La dégénérescence neurofibrillaire est l'une des deux lésions caractéristiques de la maladie d'Alzheimer. Nous avons vu dans la deuxième partie du chapitre 1, que cette lésion cérébrale est caractérisée par l'enchevêtrement neurofibrillaire dû à la suractivation de CDK5 « kinases dépendantes des cyclines 5 ».

La protéine CDK5 et son activateur p35 spécifique aux neurones sont des molécules essentielles à la migration neuronale. L'activation du calpaïne (protéase activée par le calcium) déclenchée par l'augmentation du taux de calcium intracellulaire induite par l'exposition à des plaques d'A $\beta$ <sup>84</sup>, conduit au clivage de la sous-unité régulatrice p35 à p25. Cette sous-unité pathogène s'associe à la kinase pour former un complexe CDK5/p25 hyperactif et mal localisé<sup>85</sup>. La suractivation de CDK5 est associée à une hyperphosphorylation de la protéine *Tau*, qui aboutit à la formation d'enchevêtrements neurofibrillaires<sup>86</sup>. Les inhibiteurs de CDK5 protègent les neurones contre les phosphorylations anormales de la protéine *Tau* et la mort neuronale.

Le docking moléculaire est l'une des méthodes les plus utilisées dans le criblage virtuel à base de structure (*structure-based*). Comme nous avons vu en détail dans le chapitre 2 (paragraphe 2.2.4), cette méthode implique l'ajustement virtuel de ligands dans le site de liaison correspondant. Elle se base sur des algorithmes pour trouver les interactions attractives et répulsives entre des groupes complémentaires au sein de complexes ligand-protéine<sup>87,88</sup>.

Le docking peut être complexe en raison du degré de flexibilité conformationnelle du ligand et de son état d'ionisation<sup>89,90</sup>, ainsi que de la complexité du comportement des molécules d'eau à la liaison du ligand au sein du site actif de la cible<sup>91-94</sup>. Les conformations finales dockées sont également sélectionnées en fonction de leurs scores. La précision de la fonction de scoring a un impact majeur sur la qualité des résultats du docking moléculaire<sup>90,95</sup>.

En conséquence, le modélisateur moléculaire doit trouver la combinaison optimale d'algorithmes de recherche et de fonction de score capables de classer correctement les poses dockées des ligands dans le site actif de la cible. De plus, il doit décider de conserver des molécules d'eau cristallographiquement explicites ou non dans le site de liaison, avant l'amarrage du ligand<sup>96</sup> et choisir aussi quel est l'état d'ionisation de ligands qu'on doit prendre en considération.



## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

L'intérêt porté au CDK5, ainsi que les problèmes liés au docking et le manque de méthodes de validation, nous a incités à utiliser l'analyse 3D-QSAR pour valider les configurations du docking. Cette méthode est l'analyse comparative des contacts intermoléculaires basée sur le docking<sup>97,98</sup>.

L'analyse comparative des contacts intermoléculaires basée sur le docking ou « *Docking-Based Comparative Intermolecular Contacts Analysis db-CICA* » est une nouvelle méthode 3D-QSAR pour valider les paramètres de docking et construire des pharmacophores à base de structure<sup>97-99</sup>. L'objectif de cette approche est d'identifier le groupe d'atomes du site de liaison (le site actif de la cible), qui entrent en contact avec les ligands actifs tout en évitant les ligands faiblement actifs. Si un tel ensemble d'atomes du contact est identifié pour un groupe de ligands dockés, alors on peut valider les paramètres du docking correspondant, cela voudrait dire que le docking a réussi à docker les molécules dans le site de liaison de manière à expliquer la variation de leur bioactivité<sup>97-100</sup>. Cette approche est basée sur le nombre et la qualité des contacts entre les ligands dockés et les résidus d'acides aminés du site actif de la cible.

La méthode db-CICA permet d'évaluer une configuration de docking en fonction de sa capacité à aligner un ensemble de ligands dans la poche de liaison correspondante, de manière à ce que les ligands puissants entrent en contact avec les points du site de liaison distincts de ceux approchés par les ligands de faible affinité. En d'autres termes, la méthode db-CICA évalue la corrélation entre l'affinité des ligands et leurs contacts avec les atomes du site de liaison<sup>97,98</sup>. Les modèles optimaux de la méthode db-CICA, qui représentent les contacts du site de liaison critique (ensemble d'atomes discriminant l'affinité dans le site actif de la cible), peuvent être traduits en modèles pharmacophores. Ensuite, ces derniers sont utilisés pour le criblage virtuel de la base de données structurale de l'Institut National du Cancer (NCI) à la recherche de nouveaux inhibiteurs puissants de CDK5.

## 1. Matériaux et méthodes

### 1.1. Logiciels

- MarvinView (version 5.1.4), ChemAxon Ltd., États-Unis
- DiscoveryStudio (DS 2.0), Biovia Inc., États-Unis
- MATLAB (version R2007a), The MathWorks Inc., États-Unis

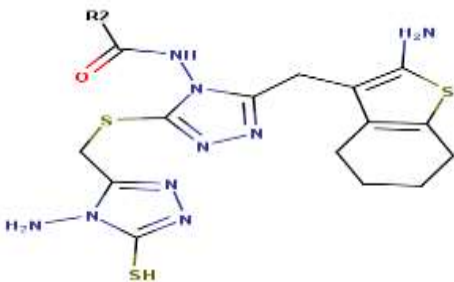
## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

### 1.2. Ensemble de données

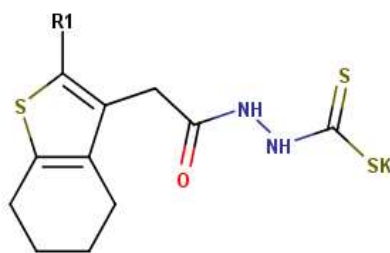
Quatre-vingts inhibiteurs de CDK5 (Tableau 13) ont été recueillis à partir de la littérature<sup>101-103</sup> depuis la base de données de l'Institut européen de bioinformatique (Chemical European Molecular Biology Laboratory « ChEMBL »). Les inhibiteurs de CDK5 recueillis ont tous été soumis à un dosage biologique en utilisant une méthode identique et leur bioactivité ont été exprimées sous forme de valeurs de IC<sub>50</sub> en nM, qui est la concentration du composé nécessaire pour inhiber à moitié l'activité de l'AChE (*The half-maximal inhibitory concentration IC50*). Les transformations logarithmiques de (1/ IC<sub>50</sub>) ont été utilisés en corrélation structure-activité afin de corrélérer de manière linéaire la bioactivité au changement d'énergie libre.

Les structures chimiques des inhibiteurs collectés ont été exportées de ChEMBL au format SMILES dans le logiciel Discovery Studio (version 2.5.5) et transformées automatiquement en structures 3D, puis enregistrées au format (SD). Chaque inhibiteur a été dockée dans deux états: ionisé et non ionisé selon les indications de MarvinView.

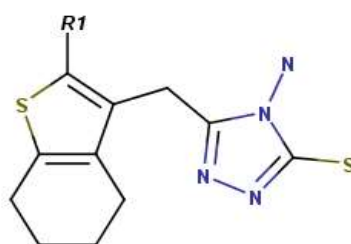
**Tableau 13.** Composés impliqués dans la modélisation basée sur la structure CDK5.

Composés	pIC <sub>50</sub>	R1	R2	R3	Ref
					
1	-3,81	NHCOCH <sub>2</sub> Cl	-4-C <sub>6</sub> H <sub>4</sub>	-	101
2	-3,79	NHCOCH <sub>2</sub> Cl	-C <sub>6</sub> H <sub>5</sub>	-	101
3	-3,81	NHCOCH <sub>3</sub>	-C <sub>6</sub> H <sub>5</sub>	-	101
4	-3,69	NHCH <sub>2</sub> CH <sub>2</sub> COOH	-C <sub>6</sub> H <sub>5</sub>	-	101
5	-3,61	NHCOCH <sub>2</sub> Cl	-CH <sub>3</sub>	-	101
6	-3,71	NHCOCH <sub>3</sub>	-CH <sub>3</sub>	-	101
7	-3,79	NHCOC <sub>6</sub> H <sub>5</sub>	-CH <sub>3</sub>	-	101
8	-3,79	NHCH <sub>2</sub> CH <sub>2</sub> COOH	-CH <sub>3</sub>	-	101

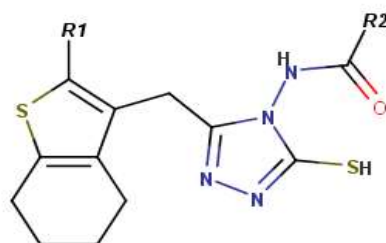
**Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants**



9	-3,79	NHCOCH2Cl	-	-	101
10	-2,39	NHCOCH3	-	-	101
11	-2,54	NHCOC6H5	-	-	101
12	-2,57	NHCH2CH2COOH	-	-	101

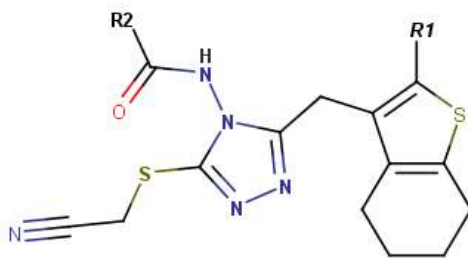


13	-1,63	NHCOCH2Cl	-	-	101
14	-2,59	NHCOCH3	-	-	101
15	-2,79	NHCOC6H5	-	-	101
16	-2,43	NHCH2CH2COOH	-	-	101

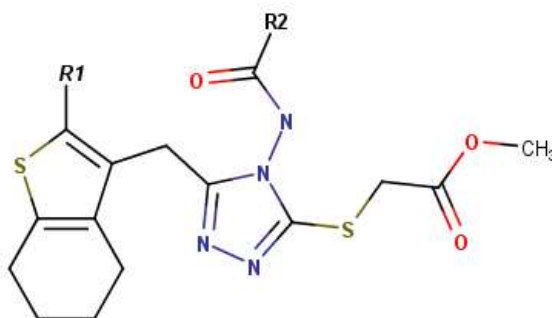


17	-2,54	NHCOCH2Cl	-4-ClC6H4	-	101
18	-2,66	NHCOCH3	-4-ClC6H4	-	101
19	-2,26	NHCOC6H5	-4-ClC6H4	-	101
20	-2,66	NHCH2CH2COOH	-4-ClC6H4	-	101
21	-2,86	NHCOCH2Cl	-C6H5	-	101
22	-2,76	NHCOCH3	-C6H5	-	101
23	-2,95	NHCOC6H5	-C6H5	-	101
24	-2,83	NHCH2CH2COOH	-C6H5	-	101
25	-2,84	NHCOCH2Cl	-CH3	-	101
26	-2,77	NHCOCH3	-CH3	-	101
27	-2,79	NHCOC6H5	-CH3	-	101
28	-2,64	NHCH2CH2COOH	-CH3	-	101

Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

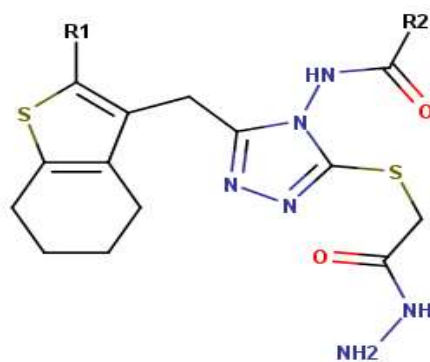


29	-1,76	NHCOCH2Cl	-4-ClC6H4	-	101
30	-2,62	NHCOCH3	-4-ClC6H4	-	101
31	-2,74	NHCOC6H5	-4-ClC6H4	-	101
32	-2,62	NHCH2CH2COOH	-4-ClC6H4	-	101
33	-2,82	NHCOCH2Cl	-C6H5	-	101
34	-2,59	NHCOCH3	-C6H5	-	101
35	-2,69	NHCOC6H5	-C6H5	-	101
36	-3,69	NHCH2CH2COOH	-C6H5	-	101
37	-1,78	NHCOCH2Cl	-CH3	-	101
38	-2,66	NHCOCH3	-CH3	-	101
39	-2,81	NHCH2CH2COOH	-CH3	-	101

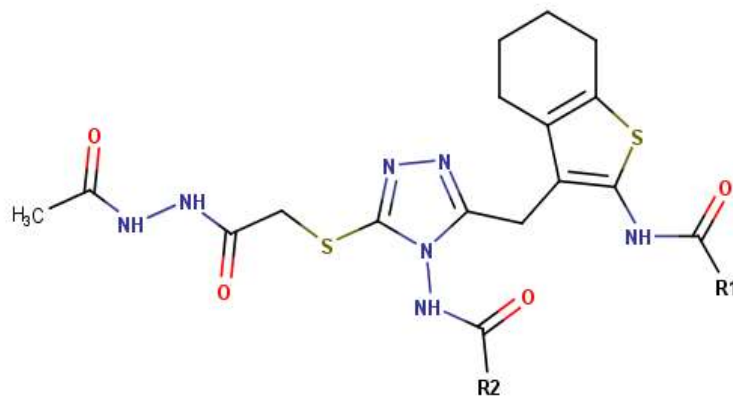


40	-2,65	NHCOCH2Cl	-4-ClC6H4	-	101
41	-2,58	NHCOCH3	-4-ClC6H4	-	101
42	-2,47	NHCOC6H5	-4-ClC6H4	-	101
43	-2,73	NHCH2CH2COOH	-4-ClC6H4	-	101
44	-2,83	NHCOCH2Cl	-C6H5	-	101
45	-2,76	NHCOCH3	-C6H5	-	101
46	-2,54	NHCOC6H5	-C6H5	-	101
47	-2,81	NHCH2CH2COOH	-C6H5	-	101
48	-2,84	NHCOCH2Cl	-CH3	-	101
49	-2,58	NHCOCH3	-CH3	-	101
50	-2,75	NHCOC6H5	-CH3	-	101
51	-2,86	NHCH2CH2COOH	-CH3	-	101

Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants



52	-1,59	NHCOCH2C1	-4-ClC6H4	-	101
53	-1,81	NHCOCH3	-4-ClC6H4	-	101
54	-2,37	NHCOC6H5	-4-ClC6H4	-	101
55	-2,30	NHCH2CH2COOH	-4-ClC6H4	-	101
56	-1,56	NHCOCH2C1	-C6H5	-	101
57	-2,28	NHCOCH3	-C6H5	-	101
58	-2,42	NHCOC6H5	-C6H5	-	101
59	-2,65	NHCH2CH2COOH	-C6H5	-	101
60	-1,78	NHCOCH2C1	-CH3	-	101
61	-2,39	NHCOCH3	-CH3	-	101
62	-2,25	NHCOC6H5	-CH3	-	101
63	-2,97	NHCH2CH2COOH	-CH3	-	101



64	-2,80	NHCOC6H5	-CH3	-CH3	101
65	-2,56	NHCH2CH2COOH	-CH3	-CH3	101
66	-2,56	NHCOCH2C1	-CH3	-C6H5	101
67	-2,64	NHCOCH3	-CH3	-C6H5	101
68	-2,68	NHCH2CH2COOH	-CH3	-C6H5	101
69	-2,65	NHCOCH2C1	-CH3	CH2C1	101
70	-2,67	NHCOCH3	-CH3	CH2C1	101
71	-2,42	NHCH2CH2COOH	-CH3	CH2C1	101
72	-2,54	NHCOCH2C1	-C6H5	-CH3	101
73	-2,74	NHCOCH3	-C6H5	-CH3	101
74	-2,81	NHCOCH3	-C6H5	CH2C1	101
75	-2,66	NHCOCH2C1	-CH3	-CH3	101

## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

76	-2,51	NHCOCH3	-CH3	-CH3	101
77	-2,53	NHCOCH2Cl	-CH3	-	101
78	-2,66	NHCOCH3	-CH3	-	101
79	-1,99	-	-	-	102
80	-2,62	-	-	-	103

### 1.3. Préparation de la structure cristalline CDK5

La structure cristalline de CDK5 a été obtenue depuis la banque de données des protéines (code PDB: 3O0G, résolution: 1,95 Å). Des atomes d'hydrogène ont été ajoutés à la protéine en utilisant DS 2.5. Des molécules d'eau explicites ont été conservées ou supprimées selon les besoins<sup>91,92</sup>, c'est-à-dire que les inhibiteurs ont été dockés en absence ou en présence de molécules d'hydratation.

## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

### 1.3. Paramètres de Docking

Les composés recueillis ont été dockés dans le site actif de CDK5 en utilisant trois algorithmes de recherche: LigandFit<sup>104</sup>, LibDock<sup>105</sup>, et CDocker<sup>106</sup>. Les poses les mieux classées ont été sélectionnées en utilisant 7 fonctions de score: Jain, LigScore1, LigScore2, PLP1, PLP2, PMF et PMF04<sup>105, 107-110</sup>.

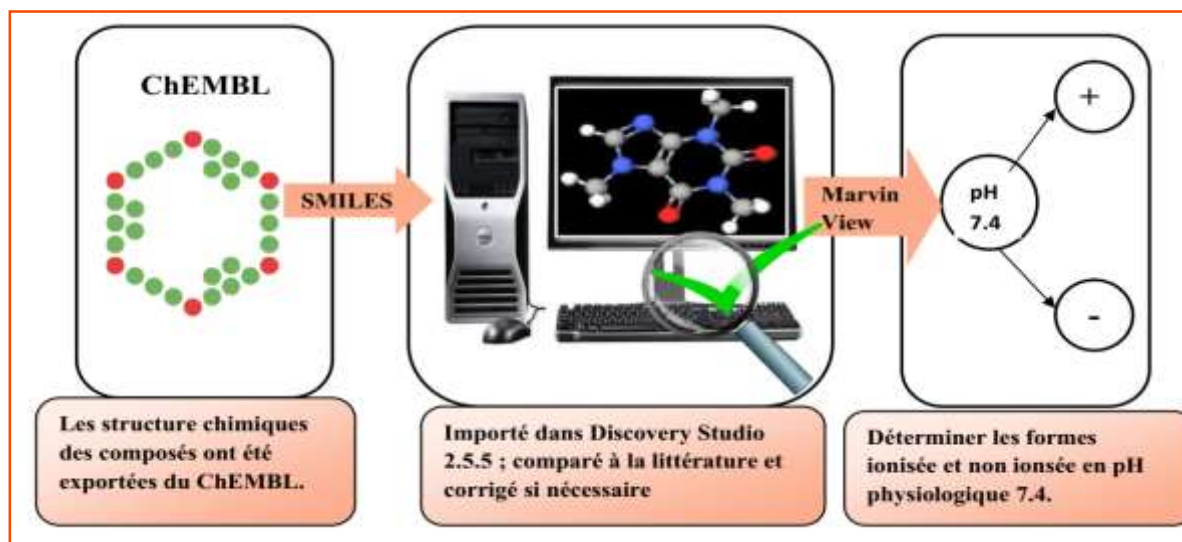


Fig. 12. Présentation schématique de la préparation des ligands pour les étapes de docking.

### 1.5. Analyse comparative des contacts intramoléculaires basée sur le docking (db-CICA)

1. Attribution d'un code binaire basé sur les contacts: Les poses dockées de chaque inhibiteur de CDK5 sont évaluées pour déterminer leurs atomes proches dans le site de liaison de la cible. Une valeur de contact intermoléculaire égale à « un » est affectée à des voisins atomiques (entre les ligands et le site actif de la cible) situés à une valeur inférieure ou égale à un seuil de distance prédéfini, sinon, une valeur de contact de « zéro » est attribuée. Par exemple, si la distance entre l'atome A du ligand et l'atome B de la poche de liaison était inférieure ou égale au seuil prédéfini, ce contact (B) se verra attribuer la valeur 1. Les interactions ligand-récepteurs, y compris les liaisons hydrogène et les forces de Van der Waals, illustrent des forces optimales à des distances comprises entre 2,0 et 3,5Å. Dans notre étude, nous avons choisi le seuil de distance 2,5Å. Finalement, une matrice 2D est générée, où les lignes correspondent à des ligands et les colonnes correspondent à différents atomes du site de liaison de la cible. La matrice est remplie du code binaire : des « zéros » et des « un ».

## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

2. Une matrice binaire a été construite pour chaque configuration de docking. En conséquence, un total de 84 fichiers de matrices binaires a été généré dans notre étude par le produit de : 3 algorithmes de recherche, 2 états d'ionisation de ligands, 2 états d'hydratation de la protéine, 7 fonctions du scoring et d'un seuil de distance.
3. Chaque colonne individuelle de la matrice est régressée par rapport aux bioactivités moléculaires correspondantes ( $pIC_{50}$ ). Les colonnes qui présentent une corrélation négative avec la bioactivité sont inversées, c'est-à-dire que les zéros sont convertis en *un* et vice-versa et sont exclus de l'étape suivante.
4. Après avoir exclu les colonnes inversées (contacts négatifs ou volumes d'exclusion), nous avons appliqué un algorithme génétique (GA) sur la matrice binaire résultante (contenant des colonnes de contact corrélées positivement avec la bioactivité) pour la recherche de la meilleure somme des contacts intermoléculaires ligand-récepteur capable d'expliquer la variation de la bioactivité (Tableau 13). Il a été demandé à la GA de balayer des combinaisons de 4 à 10 contacts intermoléculaires proportionnels (positifs) suivis de 10 contacts inversement proportionnels (négatifs). Les modèles db-CICA sont jugés sur la base de trois critères: coefficient de corrélation ( $R^2$ ), leave-one-out ( $R^2_{LOO}$ ) et  $R_{K-Fold}$ . Le meilleur modèle de la somme de colonne est sélectionné comme modèle représentatif db-CICA.
5. Il a été demandé à la GA de générer les meilleurs modèles db-CICA issus d'ensembles de quatre contacts positifs simultanés (colonnes de contacts sommés) et de répéter l'analyse pour identifier les meilleurs modèles de la somme 5, 6, 7, 8, 9 et 10 des contacts positifs. Chaque ensemble de contacts sommés est traité indépendamment pour identifier le modèle db-CICA correspondant dans chaque cas.
6. L'algorithme db-CICA permet d'identifier les contacts intermoléculaires ayant des contributions plus élevées dans les modèles dbCICA optimaux, à partir de leur poids: chaque colonne de contacts positifs peut apparaître jusqu'à 3 fois dans le modèle. Les poids sont distribués de manière aléatoire dans la première génération et soumis à une mutation uniquement (et non croisée) lors des itérations suivantes.
7. Par la suite, la GA est utilisée pour rechercher le meilleur modèle produit en combinant les colonnes inversées (colonnes d'exclusion, voir les étapes 4 et 5) avec les meilleurs modèles de la somme des contacts positifs. Dans le projet actuel, nous avons implémenté un seul paramètre d'exclusion : dix contacts négatifs ont été autorisés.



## **1.6. Générer les pharmacophores correspondant aux modèles réussis de db-CICA**

Les meilleurs modèles db-CICA ont été utilisés pour la construction des modèles pharmacophores, pour ensuite être utilisés en tant que requêtes de recherche 3D afin d'explorer de nouveaux inhibiteurs de CDK5. Les pharmacophores ont été construits en suivant les étapes suivantes:

1. Les paramètres de docking offrant les meilleurs modèles db-CICA ont été sélectionnés (Tableau 14 et 15 ) pour prédire la bioactivité d'inhibiteurs. Cela se fait en substituant les contacts des composés dockés dans l'équation de régression des modèles réussis de dbCICA. Les composés puissants (ceux ayant un écart entre les bioactivités ajustée et observée, très faible) ont été conservés dans la poche de liaison. Tandis que les autres molécules ont été exclues.

2. Les contacts positifs (les atomes du site de liaison de la cible) sélectionnés par les modèles db-CICA ont été annotés pour identifier les fractions des ligands les plus proches. Un consensus des composés puissants pour positionner des fractions de caractères physico-chimiques communes, proches de certains atomes de contact importants (sélectionné par db-CICA), nous a incité à placer un point pharmacophorique correspondant sur cette région. Par exemple, si des composés dockés puissants conviennent à placer des fragments hydrophobes adjacents à un point de contact significatif, une caractéristique hydrophobe est placée au-dessus des fragments hydrophobes des ligands. Les fonctions pharmacophoriques ont été introduites manuellement à l'aide de la bibliothèque de fonctions DS 2.5.

3. Des contacts corrélés négativement, c'est-à-dire avec la bioactivité des ligands (ceux inversés à l'étape 4 de la section précédente), sont utilisés pour démarquer les limites stériques du site de liaison de la cible. Les sphères d'exclusion ont été introduites manuellement sur ces contacts à l'aide de la bibliothèque de fonctions DS 2.5.

4. Les modèles pharmacophores ont été validés à l'aide d'une analyse de courbe ROC (*Receiver Operating Character*). Les détails de l'analyse ROC sont présentés dans le chapitre 2 (paragraphe 2.3.1.2).

5. Pour améliorer la capacité des modèles pharmacophores résultants à capter les composés actifs, il était nécessaire de les décorer avec des sphères d'exclusion supplémentaires à l'aide du logiciel HIPHOP-REFINE<sup>97,98</sup>.

## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

### 1.7. Criblage *in silico* de nouveaux inhibiteurs de CDK5

Les modèles pharmacophores efficaces (qui sont stériquement raffinés) ont été utilisés pour filtrer la base de données structurale NCI. Les *hits* capturés ont été filtrés selon les règles de Lipinski<sup>15</sup> et de Veber<sup>16</sup>. Les *hits* restants ont été ajustés aux pharmacophores db-CICA sélectionnés en utilisant l'option "meilleur ajustement" dans CATALYST de DS2.5..

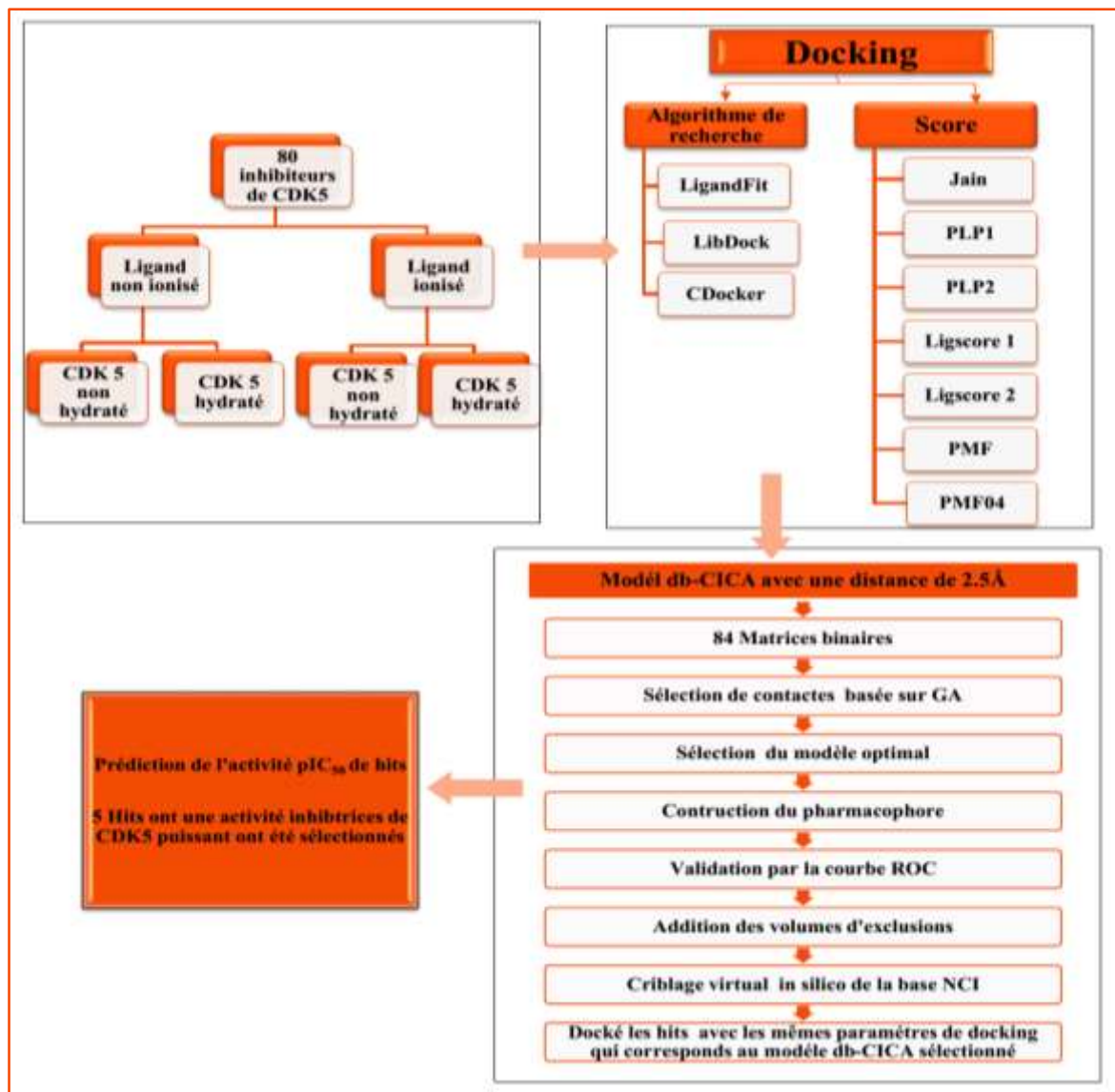


Fig.13. Représentation schématique des différentes étapes de la méthode db-CICA.

## 2. Résultats et discussion

### 2.1. Modélisation db-CICA des inhibiteurs de CDK5

La méthode db-CICA tente d'identifier les atomes du site actif de la cible qui entrent en contact avec les ligands puissants en évitant les inactifs. L'identification d'un tel ensemble d'atomes de contact valide les paramètres de docking, cela voudrait dire que le docking a réussi à docker les ligands de manière à expliquer la variation de la bioactivité. De plus, le nombre limité d'atomes de contact significatifs définis par le modèle db-CICA peut être converti en modèles pharmacophores pour permettre l'exploration de nouveaux ligands<sup>97,99,111</sup>.

Le Tableau 14 indique le nombre de contacts positifs, négatifs et les critères statistiques des modèles db-CICA optimaux basés sur des paramètres du docking différents. Ces modèles partagent le même état d'ionisation des ligands « ligands ionisés ». Dans les formes ionisées, les substituants d'acide carboxylique ( $pK_a \approx 4.5$ ) et les substituants d'amino, « amines tertiaires », ( $pK_a \approx 9.0-9.5$ ) ont respectivement des charges négatives et positives formelles. La présence d'un groupe fonctionnel ionisable (principalement cationique) favorise la pénétration de la BHE<sup>112</sup>. En outre, Les composés qui possèdent un atome d'azote tertiaire (une caractéristique de nombreux médicaments du système nerveux central) montrent un degré plus élevé de pénétration du cerveau<sup>30</sup>.

**Tableau 14.** Modèles dbCICA les mieux classés, leurs paramètres correspondants et leurs critères statistiques.

Modèle dbCICA	Algorithme De recherche	État d'ionisation	Présence d'eau <sup>a</sup>	Fonction de score	Nombre de contacts positifs <sup>b</sup>	Nombre de contacts négatifs <sup>c</sup>	R <sup>2</sup> <sup>d</sup>	R <sub>LOO</sub> <sup>e</sup>	R <sub>5-fold</sub> <sup>f</sup>
<b>I</b>	LibDock	Ionisé	Présent	Ligscore1	4	10	0.63	0.61	0.62
<b>II</b>	LibDock	Ionisé	Absent	PMF	4	10	0.62	0.60	0.58
<b>III</b>	CDocker	Ionisé	Présent	PMF04	4	10	0.59	0.59	0.56

**a)** L'état d'hydratation de la structure cristallographique de la cible ; **b)** Nombre optimal de combinaisons de contacts ligand/site de la cible qui améliorant la bioactivité ; **c)** Nombre optimal de combinaisons de contacts ligand/site de liaison qui sont défavorables à la bioactivité ; **d)** Coefficients de détermination ; **e)** Coefficients de corrélation de la validation croisée déterminés par la technique leave-one-out technique ; **f)** Coefficients de corrélation de validation croisée déterminés par k-fold.

## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

Le modèle dbCICA-I a les meilleurs critères statistiques parmi les modèles sélectionnés. Les molécules d'eau explicites semblent être importante, puisque le modèle dbCICA-I est caractérisé par un site de liaison hydraté. Les molécules d'eau peuvent notamment servir de pont (liaisons hydrogène) entre la protéine et le ligand, ce qui permet de stabiliser ce complexe.

Le Tableau 15 présente les atomes de contact critiques des acides aminés du site de liaison de CDK5 et leurs poids, proposés par le modèle dbCICA-I. Les atomes des contacts critiques du modèle représentent les interactions attractives concomitantes (principe de base de l'analyse dbCICA qui consiste à rechercher des contacts simultanés et non pas des contacts séparés), dont la somme de ces contacts est directement proportionnelle à la bioactivité. En revanche, les contacts négatifs (inversement proportionnels à la bioactivité) représentent les interactions répulsives (conflits stériques).

**Tableau 15.** Atomes de contact du site de liaison critique proposés par le modèle dbCICA-I

Modèle dbCICA	Atomes de contact favorisés (contacts positifs) <sup>a</sup>		Atomes de contact défavorisés (contacts négatifs) <sup>d</sup>
	Les acides aminés et identités des atomes correspondantes <sup>b</sup>	Poids <sup>c</sup>	
<b>I</b>	ASP84:C	2	ASN131:OD1 GLN130:CG GLN130:HG2
	ASP86:N	2	HOH311:O HOH332:O LYS128:HE2
	HOH327:H1	1	LYS89:HD1 B:PHE80:CG VAL18:HG12
	VAL64:HG22	3	VAL64:HG21

**a)** Contacts du site de liaison proportionnelle à la bioactivité ; **b)** Acides aminés du site de liaison et leurs contacts atomiques significatifs. Les codes atomiques sont tels que fournis par le format du fichier de PDB (par exemple, ILE111: CA code l'atome du carbone (A) de l'acide aminé isoleucine du numéro 111); **c)** Degré de signification (le poids) de l'atome de contact correspondant; **d)** Contacts du site de liaison inversement proportionnelle à la bioactivité.

### 2.2. Les modèles pharmacophores optimaux basés sur dbCICA

La figure 14 montre comment le modèle dbCICA-I (Tableau 15) a été traduit en modèle pharmacophore correspondant « Hypo-I ». Initialement, les poches de liaison ont été annotées en rendant les atomes significatifs de contacts sous des formes sphériques (Figure 14a).

Par la suite, nous avons sélectionné quelques composés puissants où  $IC_{50} \leq 100$  nM. Il faut prendre en considération que les composés puissants sont ceux qui ont un écart plus faible entre leur activité expérimentale et leur activité prédictive (déterminée par la régression

## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

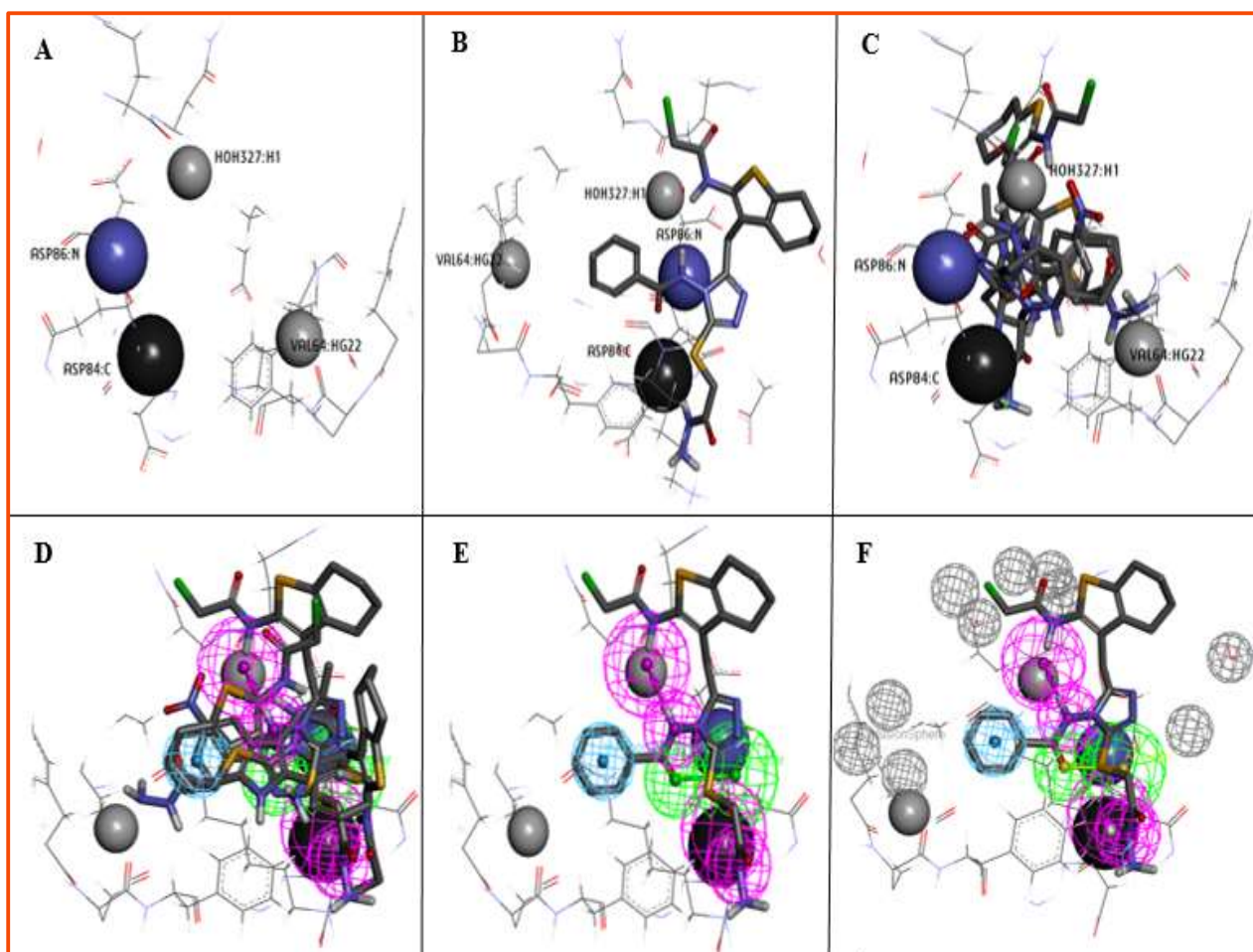
d'activité des inhibiteurs et de la somme de leurs contacts du modèle dbCICA correspondant). Ensuite, ces composés ont été alignés dans la poche du site de liaison. Des caractéristiques pharmacophoriques appropriées ont été placées sur les fonctions chimiques communes aux composés alignés, comme dans la Figure 14d.

La description suivante illustre, comment le modèle pharmacophore Hypo-I a été généré à partir du modèle optimal dbCICA-I correspondant (Figure14): l'émergence du contact positif ASP86:N (NH de l'acide aminé ASP86 a un caractère donneur) combinée au consensus des ligands puissants sur le placement de groupes carbonyles à proximité, nous a contraint à placer une liaison hydrogène accepteur à cette position. Ainsi, une interaction de type donneur entre les groupements d'amines NH (amines primaires et secondaires) des ligands et les deux acides aminés HOH327 /ASP84:C (carbone de la fonction carbonyle), nous a incité à placer deux points pharmacophoriques de type donneur autour de ces deux acides aminés. L'accord des composés puissants sur la disposition d'une fonction hydrophobe (cycles aromatiques) près de l'hydrogène (HG21) de VAL64 (apparu comme contact significatif dans le modèle dbCICA-I, comme indiqué dans le Tableau 15) nous a contraint à placer une caractéristique de pharmacophore hydrophobe aromatique, comme il est indiqué sur la figure 14d.

### 2.3. Validation de l'hypothèse pharmacophore par la courbe ROC

L'hypothèse pharmacophorique préparée manuellement (Hypo-I) a été validée en évaluant sa capacité à capter de manière sélective divers composés actifs de CDK5 à partir d'une vaste liste d'actifs et d'inactifs. Pour chaque composé actif, environ 33 composés inactifs ont été choisis au hasard dans la liste de ROC, elle même préparée comme décrite dans les travaux de Taha *et al.*<sup>97-99</sup>. La liste à tester a été construite en y incorporant 12 composés actifs connus et 229 composés inactifs. Les composés actifs et inactifs ont des caractéristiques similaires (structuellement homogène). Le Tableau 16 et la figure 15 montrent que le modèle pharmacophore Hypo-I a de bonnes performances globales avec des valeurs ROC-AUC supérieures à 80%.

## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants



**Fig.14.** Les étapes pour la génération manuelle de pharmacophore (**Hypo-I**) guidée par le modèle I de dbCICA (Tableaux 14 et 15) : **a)** les atomes de contact significatifs dans le site de liaison, c'est-à-dire, selon le modèle I de dbCICA, sont représentés par des sphères. **b)** La pose du composé 56 ( $pIC_{50pred} = -1,66$ ) le plus puissant dockée dans le site de liaison, **c)** Les poses dockées des composés puissants 13,37 et 79 dans site de liaison (leur  $pIC_{50pred}$  est : -1,93 ; -1,93 ; -2,74 respectivement) **d)** Caractéristiques pharmacophoriques placés manuellement sur les fractions chimiques communes aux ces composés, **e)** La pose du ligand 56 et son lien avec les caractéristiques pharmacophoriques proposées, **f)** Sphères d'exclusion adaptées aux atomes du site de liaison présentant des corrélations négatives avec la bioactivité (comme dans le modèle I de dbCICA).

**Tableau 16.** Performances du pharmacophore sélectionné par le modèle dbCICA-I.

Modèle de pharmacophore	AUC	ACC	SPC	TPR	FNR
Hypo-I	84	94	96	61	3
Hypo-I raffiné	96	92	95	25	1

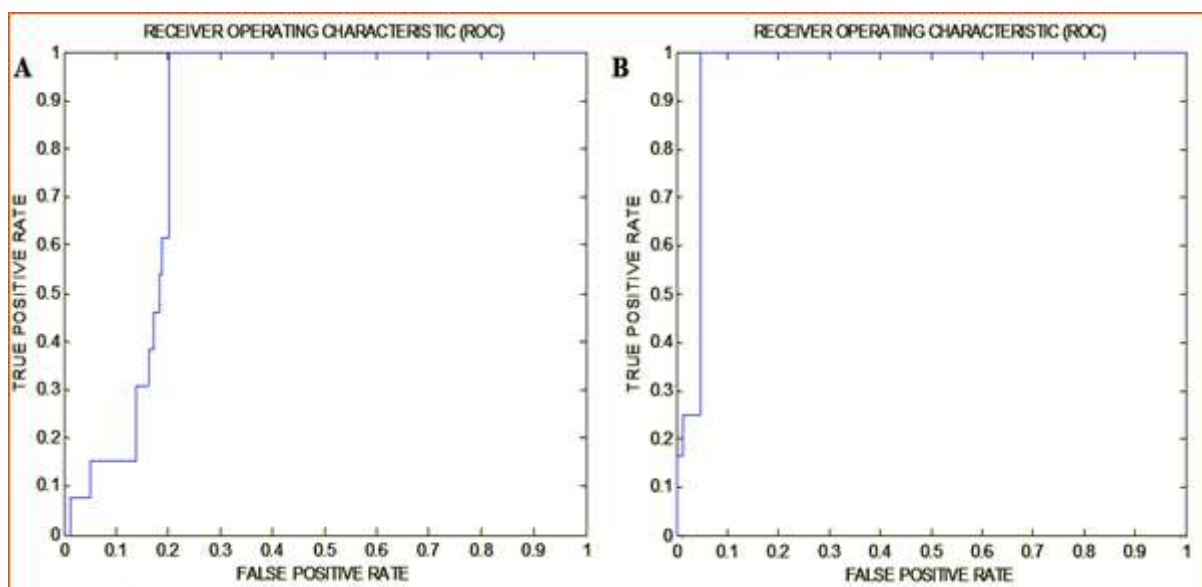
## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

**AUC:** aire sous la courbe (%); **ACC:** précision globale (%), **SPC:** spécificité globale (%) **TPR:** Taux global positif réel (%); **FNR:** Taux global de faux négatif (%).

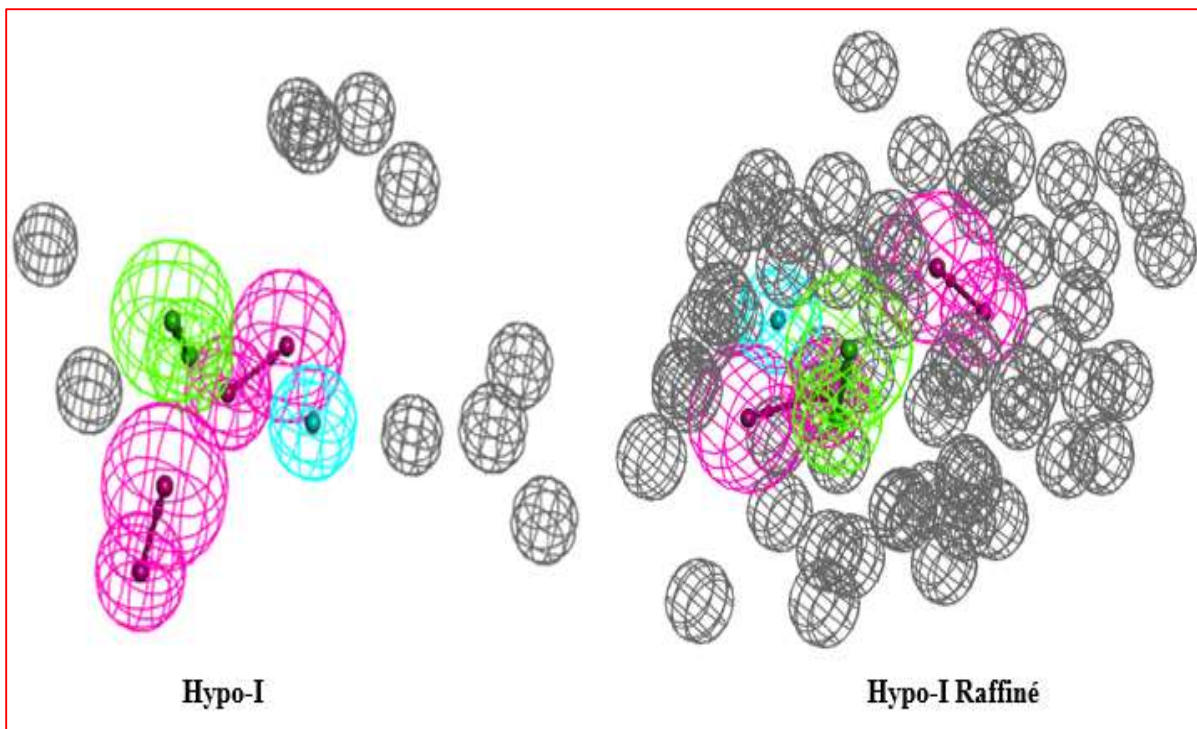
### 2.4. Raffinement du modèle pharmacophore « volumes d'exclusion »

Le modèle du pharmacophore Hypo-I a été décoré avec des volumes d'exclusion utilisant le module HipHop-Affiner. Ce module utilise des composés d'entraînement inactifs pour construire des volumes d'exclusion qui ressemblent aux contraintes stériques dans le site de liaison. Les volumes d'exclusion représentent les régions stériquement interdites à l'intérieur de la poche de liaison et sont occupés par les composés inactifs. Ces régions sont ensuite remplies avec des volumes d'exclusion <sup>111,113-115</sup>.

Le raffinement stérique a amélioré la capacité du pharmacophores Hypo-I à capter les actifs. La valeur ROC-AUC du modèle Hypo-I raffiné (96 %) est significativement plus élevée que son homologue non raffiné, comme il est illustré dans le Tableau 16. Il ressort clairement, d'après la Figure 15, que le modèle Hypo-I raffiné a une capacité élevée à capter les actifs, en conséquence, ce modèle a été sélectionné en tant que requêtes de recherche sur la base de données structurale NCI.



**Fig.15.** Les courbes ROC du pharmacophore à base du modèle dbCIC: **A:** Hypo-I, **B :** Hypo-I raffiné.



**Fig.16.** Les points pharmacophoriques d'Hypo-I: Les sphères bleu clair représentent les caractéristiques hydrophobes, les sphères vertes représentent les liaisons accepteurs d'hydrogène, les sphères violettes représentent les liaisons donneurs d'hydrogène et les sphères grises représentent les régions d'exclusion. Le modèle Hypo-I est raffiné avec 58 volumes d'exclusion.

## 2.5. Criblage *in silico* de la base de données NCI pour sélectionner des nouveaux inhibiteurs de CDK5 puissants


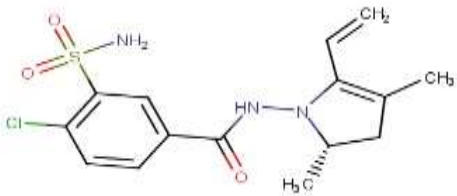

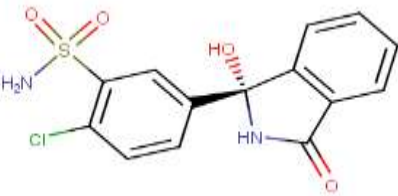
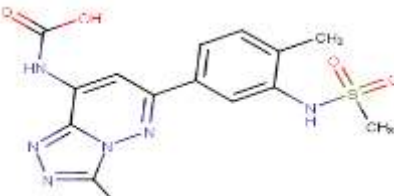
La version raffinée d'Hypo-I (Figure 16) a été utilisée comme requête de recherche 3D pour filtrer la base de données de l'institut national du Cancer NCI (comportant 238 819 composés), dans le but d'explorer de nouveaux inhibiteurs de CDK5.

Les *hits* capturés ont été filtrés sur la base des règles de Lipinski<sup>15</sup> et de Veber<sup>16</sup>. Ils ont ensuite été dockés dans la protéine CDK5 (code PDB: 3O0G) en utilisant les mêmes conditions de docking que celle du modèle dbCICA-I correspondant. Les poses résultantes ont été analysées pour la recherche des contacts critiques en fonction des modèles dbCICA-I, et pour prédire leur activité inhibitrice du CDK5. Le Tableau 17 présente les *hits* actifs sélectionnés à partir de cette requête 3D et leur bioactivité (pIC50).



## Analyse comparative des contacts intermoléculaires basés sur le docking (3D-QSAR) pour le criblage *in silico* de nouveaux inhibiteurs puissants

**Tableau 17.** Les *hits* les plus actifs captés par le pharmacophore raffiné Hypo-I et leur activité CDK5 prédictive *in silico*.

Composés	Code de NCI	pIC50	Composés	Code de NCI	pIC50
					
1	76243	7,308	4	757075	8,382
					
2	380092	7,577	5	69200	7,845
					
3	768126	8,114			

### 3. Conclusion

Les inhibiteurs de CDK5 sont considérés comme des traitements potentiels pour les troubles neurodégénératifs tels que la MA. Dans la recherche actuelle, nous avons mis en œuvre une vaste gamme de configurations de docking pour docker 80 inhibiteurs dans le site actif de la cible CDK5. Nous avons utilisé la méthode dbCICA pour identifier et valider les configurations de docking optimales. Le modèle dbCICA résultant a été utilisé pour construire le modèle pharmacophore correspondant, ensuite ce modèle a été validé par analyse ROC. Le meilleur modèle pharmacophore a été utilisé pour cribler la liste des composés NCI. La bioactivation inhibitrice de CDK5 a été significative et importante dans 5 *hits*. Ces derniers ont une activité inhibitrice améliorée par rapport aux composés initiaux (Tableau 13).

### Références bibliographique :

1. Patani, G. A., & LaVoie, E. J. (1996). Bioisosterism: a rational approach in drug design. *Chemical reviews*, 96(8), 3147-3176.
2. Wermuth, C. G. (2011). Are pyridazines privileged structures?. *MedChemComm*, 2(10), 935-941.
3. Ritchie, T. J., Macdonald, S. J., Peace, S., Pickett, S. D., & Luscombe, C. N. (2012). The developability of heteroaromatic and heteroaliphatic rings—do some have a better pedigree as potential drug molecules than others?. *Med Chem Comm*, 3(9), 1062-1069.
4. El-Sayed, H. A., Moustafa, A. H., Haikal, A. Z., & El-Ashry, E. S. H. (2009). Synthesis and antibacterial activity of some glucosyl- and ribosyl-pyridazin-3-ones. *Nucleosides, Nucleotides and Nucleic Acids*, 28(3), 184-192.
5. Refaat, H. M., Khalil, O. M., & Kadry, H. H. (2007). Synthesis and anti-inflammatory activity of certain piperazinylthienylpyridazine derivatives. *Archives of pharmacal research*, 30(7), 803-811.
6. Moldoveanu, C., Mangalagiu, G., Drochioiu, G., Caprosu, M., Petrovanu, M., & Mangalagiu, I. I. (2003). New antituberculosis compounds derived from diazine. *An. Stiint. Univ. "Al. I. Cuza" Iasi*, 11(1), 367-374.
7. Siddiqui, A. A., Mishra, R., Shaharyar, M., Husain, A., Rashid, M., & Pal, P. (2011). Triazole incorporated pyridazinones as a new class of antihypertensive agents: Design, synthesis and in vivo screening. *Bioorganic & medicinal chemistry letters*, 21(3), 1023-1026.
8. Gualtieri, F., Dei, S., Manetti, D., Romanelli, M. N., Scapecchi, S., & Teodori, E. (1995). The medicinal chemistry of Alzheimer's and Alzheimer-like diseases with emphasis on the cholinergic hypothesis. *Il Farmaco*, 50(8), 489-503.
9. HyperChem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA (2008).
10. MarvinSketch 17.1.2 (2017), Chemaxon (<http://www.chemaxon.com>).
11. Ghose, A.K., Herbertz, T., Hudkins, R.L., Dorsey, B.D., Mallamo, J.P., 2011. Knowledge-based, central nervous system (CNS) lead selection and lead optimization for CNS drug discovery, *ACS Chem. Neurosci.* 3(1), 50-68.
12. Wager, T. T., Hou, X., Verhoest, P. R., & Villalobos, A. (2010). Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS chemical neuroscience*, 1(6), 435-449.
13. Wager, T. T., Hou, X., Verhoest, P. R., & Villalobos, A. (2016). Central nervous system multiparameter optimization desirability: application in drug discovery. *ACS chemical neuroscience*, 7(6), 767-775.
14. Wager, T. T., Chandrasekaran, R. Y., Hou, X., Troutman, M. D., Verhoest, P. R., Villalobos, A., & Will, Y. (2010). Defining desirable central nervous system drug space through the alignment of molecular properties, in vitro ADME, and safety attributes. *ACS chemical neuroscience*, 1(6), 420-434.
15. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23(1-3), 3-25.

16. Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, 45(12), 2615-2623.
17. Hopkins, A. L., Groom, C. R., & Alex, A. (2004). Ligand efficiency: a useful metric for lead selection. *Drug discovery today*, 9(10), 430-431.
18. Leeson, P. D., & Springthorpe, B. (2007). The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews Drug Discovery*, 6(11), 881-890.
19. Schultes, S., de Graaf, C., Haaksma, E. E., de Esch, I. J., Leurs, R., & Krämer, O. (2010). Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discovery Today: Technologies*, 7(3), e157-e162.
20. Keserü, G.M., Makara, G.M., 2009. The influence of lead discovery strategies on the properties of drug candidates. *Nat. Rev. Drug Discov*, 8(3), 203-212.
21. Contreras, J. M., Rival, Y. M., Chayer, S., Bourguignon, J. J., & Wermuth, C. G. (1999). Aminopyridazines as acetylcholinesterase inhibitors. *Journal of medicinal chemistry*, 42(4), 730-741.
22. Contreras, J. M., Parrot, I., Sippl, W., Rival, Y. M., & Wermuth, C. G. (2001). Design, synthesis, and structure-activity relationships of a series of 3-[2-(1-benzylpiperidin-4-yl) ethylamino] pyridazine derivatives as acetylcholinesterase inhibitors. *Journal of medicinal chemistry*, 44(17), 2707-2718.
23. Salah, T., Belaidi, S., Melkemi, N., & Tchouar, N. (2015). Molecular Geometry, Electronic Properties, MPO Methods and Structure Activity/Property Relationship Studies of 1, 3, 4-Thiadiazole Derivatives by Theoretical Calculations. *Reviews in Theoretical Science*, 3(4), 355-364.
24. Melkemi, N., & Belaidi, S. (2014). Structure-property relationships and quantitative structure-activity relationship modeling of detoxication properties of some 1, 2-dithiole-3-thione derivatives. *Journal of Computational and Theoretical Nanoscience*, 11(3), 801-806.
25. Atkinson, F., Cole, S., Green, C., & Van de Waterbeemd, H. (2002). Lipophilicity and other parameters affecting brain penetration. *Current Medicinal Chemistry-Central Nervous System Agents*, 2(3), 229-240.
26. Van de Waterbeemd, H., Camenisch, G., Folkers, G., Chretien, J. R., & Raevsky, O. A. (1998). Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *Journal of drug targeting*, 6(2), 151-165.
27. Waring, M. J. (2009). Defining optimum lipophilicity and molecular weight ranges for drug candidates—molecular weight dependent lower logD limits based on permeability. *Bioorganic & medicinal chemistry letters*, 19(10), 2844-2851.
28. Martin, Y. C. (2005). A bioavailability score. *Journal of medicinal chemistry*, 48(9), 3164-3170.
29. Johnson, T. W., Dress, K. R., & Edwards, M. (2009). Using the Golden Triangle to optimize clearance and oral absorption. *Bioorganic & medicinal chemistry letters*, 19(19), 5560-5564.
30. Pajouhesh, H., & Lenz, G. R. (2005). Medicinal chemical properties of successful central nervous system drugs. *NeuroRx*, 2(4), 541-553.
31. Kerns, E.H., *Li Di.*, 2008. *Drug-Like Properties: Concepts, Structure Design and Methods from ADME to Toxicity Optimization*, first ed. Elsevier, United States.

32. Hansch, C., Leo, A., 1979. Substituent constants for correlation analysis in chemistry and biology, Third ed. Wiley, New York.
33. Bazooyar, F., Taherzadeh, M., Niklasson, C., Bolton, K., 2013. Molecular modeling of cellulose dissolution. *Journal of Computational and Theoretical Nanoscience*, 10(11), 2639-2646.
34. Fichert, T., Yazdaniyan, M., & Proudfoot, J. R. (2003). A structure–permeability study of small drug-like molecules. *Bioorganic & medicinal chemistry letters*, 13(4), 719-722.
35. Palm, K., Stenberg, P., Luthman, K., & Artursson, P. (1997). Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharmaceutical research*, 14(5), 568-571
36. Benbrahim, Imane & Belaidi, Salah & M Alafeefy, Ahmed. (2016). In Silico Modeling of Substitution-Induced Effect and Structure Property/Activity Relationship Profile of 1, 3, 4-Oxadiazole Derivatives. *REVIEWS IN THEORETICAL SCIENCE*.
37. Jann, M. W. (2000). Rivastigmine, a new-generation cholinesterase inhibitor for the treatment of Alzheimer's disease. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 20(1), 1-12
38. Roy, K., Kar, S., & Das, R. N. (2015). Statistical methods in QSAR/QSPR. In *A primer on QSAR/QSPR modeling* (pp. 37-59). Springer, Cham.
39. Roy, K., & Mitra, I. (2011). On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Combinatorial chemistry & high throughput screening*, 14(6), 450-474.
40. Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., & Agrawal, R. K. (2011). Validation of QSAR models-strategies and importance. *International Journal of Drug Design & Discovery*, 3, 511-519.
41. Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set in descriptor space: a review. *Alternatives to laboratory animals*, 33(5), 445-459.
42. Ghamali, M., Chtita, S., Ousaa, A., Elidrissi, B., Bouachrine, M., & Lakhlifi, T. (2017). QSAR analysis of the toxicity of phenols and thiophenols using MLR and ANN. *Journal of Taibah University for Science*, 11(1), 1-10.
43. Bishop, C. M. (1994). Neural networks and their applications. *Review of scientific instruments*, 65(6), 1803-1832.
44. Willett, P. (2006). Similarity-based virtual screening using 2D fingerprints. *Drug discovery today*, 11(23-24), 1046-1053.
45. Sheridan, R. P., & Kearsley, S. K. (2002). Why do we need so many chemical similarity search methods?. *Drug discovery today*, 7(17), 903-911.
46. OECD, D. (2007). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q) SAR] Models. *Organisation for Economic Co-operation and Development, Paris, France*.
47. ChemOffice, PerkinElmer Informatics, 2010.

48. Gaussian 09, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford CT, 2010.
49. Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6), 3098.
50. Lee, C., & Yang, W. R. Parr G, 1998. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 785-789.
51. Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M., & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environmental health perspectives*, 111(10), 1361-1375.
52. Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). An introduction to classification and clustering. *Cluster Analysis*, 5, 1-13.
53. XLSTAT, 2018 Add-in software, **XLSTAT Company**. [www.xlstat.com](http://www.xlstat.com).
54. Tropsha, A. (2010). Best practices for qsar model development, validation, and exploitation. *Molecular Informatics*.
55. Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis, *Chemometr. Intell. Lab.*, 2, 37–52.
56. Darlington, R. B., & Hayes, A. F. (2016). *Regression analysis and linear models: Concepts, applications, and implementation*. Guilford Publications.
57. Andrea, T. A., & Kalayeh, H. (1991). Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *Journal of Medicinal Chemistry*, 34(9), 2824-2836.
58. MATLAB 9.3.0 (R2017b) and Statistics toolbox release, The Math Works, Inc., Natick, Massachusetts, United States.
59. Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31-44.
60. Chauhan, H., Bernick, J., Prasad, D., & Masand, V. (2016). The Role of Artificial Neural Networks on Target Validation in Drug Discovery and Development. *Artificial Neural Network for Drug Design, Delivery and Disposition*, 15–27.
61. Boudergua, S., Alloui, M., Belaidi, S., Al Mogren, M. M., Ibrahim, U. A. E., & Hochlaf, M. (2019). QSAR Modeling and Drug-Likeness Screening for Antioxidant Activity of Benzofuran Derivatives. *Journal of Molecular Structure*, 1189, 307-314.

62. Salt, D. W., Yildiz, N., Livingstone, D. J., & Tinsley, C. J. (1992). The use of artificial neural networks in QSAR. *Pesticide science*, 36(2), 161-170.
63. Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design*, PWS Pub. Co., Boston, 3632.
64. Zaheer-ul-Haq, & Madura, J. D. (Eds.). (2015). *Frontiers in Computational Chemistry*.1(1Eds). Bentham Science Publishers.
65. Golbraikh, A., & Tropsha, A. (2002). Beware of  $q^2$ !. *Journal of molecular graphics and modelling*, 20(4), 269-276.
66. Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. *Journal of chemical information and modeling*, 52(2), 396-408.
67. Pratim Roy, P., Paul, S., Mitra, I., & Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules*, 14(5), 1660-1701.
68. Mitra, I., Roy, P. P., Kar, S., Ojha, P. K., & Roy, K. (2010). On further application of  $r$  as a metric for validation of QSAR models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 24(1), 22-33.
69. Van der Voet, H. (1994). Comparing the predictive accuracy of models using a simple randomization test. *Chemometrics and intelligent laboratory systems*, 25(2), 313-323.
70. Mitra, I., Saha, A., & Roy, K. (2010). Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Molecular Simulation*, 36(13), 1067-1079.
71. Ambure, P., Aher, R. B., Gajewicz, A., Puzyn, T., & Roy, K. (2015). “NanoBRIDGES” software: open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and Intelligent Laboratory Systems*, 147, 1-13. 2
72. Roy, K., Kar, S., & Ambure, P. (2015). On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems*, 145, 22-29.
73. Gramatica, P. (2007). Principles of QSAR models validation: internal and external. *QSAR & combinatorial science*, 26(5), 694-701.
74. Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, 22(1), 69-77.
75. Schneider, G., & Wrede, P. (1998). Artificial neural networks for computer-based molecular design. *Progress in biophysics and molecular biology*, 70(3), 175-222.
76. Zheng, F., Bayram, E., Sumithran, S. P., Ayers, J. T., Zhan, C. G., Schmitt, J. D., ... & Crooks, P. A. (2006). QSAR modeling of mono- and bis-quaternary ammonium salts that act as antagonists at neuronal nicotinic acetylcholine receptors mediating dopamine release. *Bioorganic & medicinal chemistry*, 14(9), 3017-3037.
77. Kryger, G., Silman, I., & Sussman, J. L. (1999). Structure of acetylcholinesterase complexed with E2020 (Aricept®): implications for the design of new anti-Alzheimer drugs. *Structure*, 7(3), 297-307.
78. Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Overington, J. P. (2011). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*,

79. Zerroug, A., Belaidi, S., BenBrahim, I., Sinha, L., & Chtita, S. (2019). Virtual screening in drug-likeness and structure/activity relationship of pyridazine derivatives as Anti-Alzheimer drugs. *Journal of King Saud University-Science*, 31(4), 595-601.
80. Murphy, M. P., & LeVine III, H. (2010). Alzheimer's disease and the amyloid- $\beta$  peptide. *Journal of Alzheimer's Disease*, 19(1), 311-323.
81. Seltzer, B. (2005). Donepezil: a review. *Expert opinion on drug metabolism & toxicology*, 1(3), 527-536.
82. Dunn, N. R., Pearce, G. L., & Shakir, S. A. W. (2000). Adverse effects associated with the use of donepezil in general practice in England. *Journal of Psychopharmacology*, 14(4), 406-408.
83. Ferreira, E., Resende, R., Costa, R., Oliveira, C. R., & Pereira, C. M. (2006). An endoplasmic-reticulum-specific apoptotic pathway is involved in prion and amyloid-beta peptides neurotoxicity. *Neurobiology of disease*, 23(3), 669-678.
84. Patrick, G. N., Zukerberg, L., Nikolic, M., de La Monte, S., Dikkes, P., & Tsai, L. H. (1999). Conversion of p35 to p25 deregulates Cdk5 activity and promotes neurodegeneration. *Nature*, 402(6762), 615.
85. Lee, M. S., & Tsai, L. H. (2003). Cdk5: one of the links between senile plaques and neurofibrillary tangles?. *Journal of Alzheimer's Disease*, 5(2), 127-137.
86. Hecht, D., & Fogel, G. B. (2009). Computational intelligence methods for docking scores. *Current Computer-Aided Drug Design*, 5(1), 56-68.
87. Menikarachchi, L. C., & Gascón, J. A. (2010). QM/MM approaches in medicinal chemistry research. *Current topics in medicinal chemistry*, 10(1), 46-54.
88. Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, 261(3), 470-489.
89. Englebienne, P., & Moitessier, N. (2009). Docking Ligands into Flexible and Solvated Macromolecules. 4. Are Popular Scoring Functions Accurate for this Class of Proteins? *Journal of Chemical Information and Modeling*, 49(6), 1568-1580
90. Homans, S. W. (2007). Water, water everywhere—except where it matters?. *Drug discovery today*, 12(13-14), 534-539.
91. Bodnarchuk, M. S. (2016). Water, water, everywhere... It's time to stop and think. *Drug Discovery Today*, 21(7), 1139-1146.
92. Hecht, D., & Fogel, G. B. (2009). Computational intelligence methods for docking scores. *Current Computer-Aided Drug Design*, 5(1), 56-68.
93. Bissantz, C., Folkers, G., & Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *Journal of medicinal chemistry*, 43(25), 4759-4767.
94. Wang, R., Liu, L., Lai, L., & Tang, Y. (1998). SCORE: A new empirical method for estimating the binding affinity of a protein-ligand complex. *Molecular modeling annual*, 4(12), 379-394.
95. Jain, A. N. (1996). Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *Journal of computer-aided molecular design*, 10(5), 427-440.

96. Taha, M. O., Habash, M., Al-Hadidi, Z., Al-Bakri, A., Younis, K., & Sisan, S. (2011). Docking-based comparative intermolecular contacts analysis as new 3-D QSAR concept for validating docking studies and in silico screening: NMT and GP inhibitors as case studies. *Journal of chemical information and modeling*, 51(3), 647-669.
97. Al-Sha'er, M. A., & Taha, M. O. (2012). Application of docking-based comparative intermolecular contacts analysis to validate Hsp90 $\alpha$  docking studies and subsequent in silico screening for inhibitors. *Journal of molecular modeling*, 18(11), 4843-4863.
98. Abuhammad, A., & Taha, M. (2016). Innovative computer-aided methods for the discovery of new kinase ligands. *Future medicinal chemistry*, 8(5), 509-526.
99. Jaradat, N. J., Khanfar, M. A., Habash, M., & Taha, M. O. (2015). Combining docking-based comparative intermolecular contacts analysis and k-nearest neighbor correlation for the discovery of new check point kinase 1 inhibitors. *Journal of computer-aided molecular design*, 29(6), 561-581.
100. Shiradkar, M., Thomas, J., Kanase, V., & Dighe, R. (2011). Studying synergism of methyl linked cyclohexyl thiophenes with triazole: Synthesis and their cdk5/p25 inhibition activity. *European journal of medicinal chemistry*, 46(6), 2066-2074.
101. Shrestha, S., Natarajan, S., Park, J. H., Lee, D. Y., Cho, J. G., Kim, G. S., ... & Baek, N. I. (2013). Potential neuroprotective flavonoid-based inhibitors of CDK5/p25 from *Rhus parviflora*. *Bioorganic & medicinal chemistry letters*, 23(18), 5150-5154.
102. Demange, L., Abdellah, F. N., Lozach, O., Ferandin, Y., Gresh, N., Meijer, L., & Galons, H. (2013). Potent inhibitors of CDK5 derived from roscovitine: Synthesis, biological evaluation and molecular modelling. *Bioorganic & medicinal chemistry letters*, 23(1), 125-131.
103. Venkatachalam, C. M., Jiang, X., Oldfield, T., & Waldman, M. (2003). LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *Journal of Molecular Graphics and Modelling*, 21(4), 289-307.
104. Rao, S. N., Head, M. S., Kulkarni, A., & LaLonde, J. M. (2007). Validation studies of the site-directed docking program LibDock. *Journal of chemical information and modeling*, 47(6), 2159-2171.
105. Wu, G., Robertson, D. H., Brooks, C. L. 3rd, and Vieth, M. (2003). Detailed analysis of gridbased molecular docking: A case study of CDOCKER-A CHARMM-based MD docking algorithm. *Journal of Computational Chemistry*, 24 (13), 1549–1562.
106. Jain, A. N. (2006). Scoring functions for protein-ligand docking. *Current Protein and Peptide Science*, 7 (5), 407–420
107. Gehlhaar, D. K., Verkhivker, G. M., Rejto, P. A., Sherman, C. J., Fogel, D. R., Fogel, L. J., & Freer, S. T. (1995). Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chemistry & biology*, 2(5), 317-324.
108. Muegge, I. (2000). A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. *Perspectives in Drug Discovery and Design*, 20(1), 99-114.
109. Muegge, I., & Martin, Y. C. (1999). A general and fast scoring function for protein– ligand interactions: a simplified potential approach. *Journal of medicinal chemistry*, 42(5), 791-804.



110. Tuffaha, G. O., Ma'mon, M. H., & Taha, M. O. (2019). Discovery of new JNK3 inhibitory chemotypes via QSAR-Guided selection of docking-based pharmacophores and comparison with other structure-based pharmacophore modeling methods. *Journal of Molecular Graphics and Modelling*, 91, 30-51.
111. Clark, D. E. (2003). In silico prediction of blood–brain barrier permeation. *Drug discovery today*, 8(20), 927-933.
112. Al-masri, I. M., Mohammad, M. K., & Taha, M. O. (2008). Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening. *ChemMedChem: Chemistry Enabling Drug Discovery*, 3(11), 1763-1779.
113. Taha, M. O., Qandil, A. M., Zaki, D. D., & AlDamen, M. A. (2005). Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling. *European journal of medicinal chemistry*, 40(7), 701-727.
114. Taha, M. O., Bustanji, Y., Al-Bakri, A. G., Yousef, A. M., Zalloum, W. A., Al-Masri, I. M., & Atallah, N. (2007). Discovery of new potent human protein tyrosine phosphatase inhibitors via pharmacophore and QSAR analysis followed by in silico screening. *Journal of Molecular Graphics and Modelling*, 25(6), 870-884.
115. Taha, M. O., Atallah, N., Al-Bakri, A. G., Paradis-Bleau, C., Zalloum, H., Younis, K. S., & Levesque, R. C. (2008). Discovery of new MurF inhibitors via pharmacophore modeling and QSAR analysis followed by in-silico screening. *Bioorganic & medicinal chemistry*, 16(3), 1218-1235.

# CONCLUSION GÉNÉRALE

## Conclusion Générale

L'objectif des travaux réalisés au cours de cette étude est l'application des méthodes de modélisations à la conception de nouveaux inhibiteurs puissants de l'acétylcholinestérase (AChEI) et de kinases dépendantes des cyclines 5 (CDK5). Nous nous sommes basées sur les méthodes de criblage virtuel à base de structure du récepteur et à base de ligands.

Dans la première partie, l'utilisation du score de désirabilité MPO, les approches SAR et les règles de *Drug likeness* nous ont offert la possibilité de sélectionner quatre composés potentiels parmi une série d'inhibiteurs de l'acétylcholinestérase (AChEI). Ces composés ont une meilleure pénétration à travers la barrière hémato-encéphalique (BHE) ainsi qu'une bonne perméabilité intestinale et biodisponibilité orale tout en présentant des caractéristiques d'ADME *in vitro* acceptables. Ces résultats nous ont aidés à sélectionner des composés « *Drug-Like* » efficaces pour le SNC.

Dans la deuxième partie, nous avons utilisé l'intelligence artificielle pour explorer de nouvelles molécules pouvant servir comme agent thérapeutique efficace contre la maladie d'Alzheimer. Un réseau de neurones artificiels (ANN) a été mis en place pour prédire l'activité anti-AChE d'une série de dérivés de la pyridazine. Leurs structures chimiques sont caractérisées par un ensemble approprié de descripteurs moléculaires. À l'aide de la méthode ACP, nous avons éliminé tous les descripteurs qui sont fortement corrélés pour rendre les informations moins redondantes. La méthode QSAR-ANN a été mise en application au moyen d'un réseau neuronal (NN) à trois couches en utilisant les descripteurs moléculaires qui ont une forte corrélation avec l'activité anti-AChE et qui ont été sélectionnés par la méthode de régression linéaire multiple (MLR). La relation quantitative non linéaire est plus satisfaisante entre les caractéristiques structurelles des dérivatives de la pyridazine et leur activité anti-AChE. Le meilleur réseau neuronal de calcul avait une architecture [5-10-12-1]. Tous les paramètres de validation ont montré que le modèle ANN peut être utilisé de manière très satisfaisante pour le criblage d'une nouvelle série de molécules ayant une activité anti-AChE. La technique de criblage virtuel basée sur la méthode de similarité moléculaire et le domaine d'applicabilité développés par QSAR-ANN ont permis la découverte d'un nouveau molécule « *hit* » à une activité anti-AChE, qui s'est avéré plus puissant que ceux du composé de référence et du donépézil (Aricept).

## Conclusion Générale

La protéine CDK5 joue un rôle très important dans la neurodégénérescence. Des perturbations dans sa fonction ont généralement entraîné des problèmes pathologiques tels que la maladie d'Alzheimer. Les inhibiteurs de CDK5 se sont révélés prometteurs dans le traitement des neuropathologies. Une approche de calcul basée sur la structure a été mise en œuvre, celle-ci est l'analyse comparative des contacts intermoléculaires basée sur le docking (dbCICA). Cette méthode nous a permis de sélectionner les paramètres de docking qui ont réussi à dockée les ligands de manière à expliquer la variation de leur bioactivité et identifier le groupe d'atomes du site actif de la cible qui entrent en contact avec les ligands actifs. Ensuite, les contacts du site de liaison critique identifiés (ensemble d'atomes discriminant l'affinité dans le site de liaison) sont traduits en modèle pharmacophore validé par la courbe ROC. Ce dernier est utilisé dans la recherche de nouveaux inhibiteurs de CDK5, à partir de la base de données structurale NCI. Ce criblage virtuel à base de structure nous a permis d'obtenir cinq *hits* prometteurs et puissants.

Pour conclure, ces méthodes, bien que différentes dans leur approche, proposent un intérêt important dans la découverte de nouvelles molécules thérapeutiques « *hits* » désignés dans le traitement de la maladie d'Alzheimer et sont susceptibles d'obtenir des résultats cliniques prometteurs.

Ce travail a été couronné par une publication d'un article dans une revue de renommée internationale classe A, intitulé « **Virtual Screening in Drug-Likeness and Structure/Activity Relationship of Pyridazine Derivatives as Anti-Alzheimer Drugs** » et par deux autres articles qui sont en cours de soumission.

# RÉSUMÉ

La maladie d'Alzheimer (MA), forme de démence la plus répandue, est caractérisé par le déclin de la capacité cognitive suffisamment grave pour gêner les activités de la vie quotidienne. En 2019, ADI (*Alzheimer's Disease International*) estime à plus de 50 millions le nombre de personnes atteintes de démence dans le monde, ce chiffre devrait passer à 152 millions d'ici 2050. Dans notre étude nous avons utilisé les deux méthodes de criblage virtuel à base de ligands (une étude qualitative SAR et une étude quantitative QSAR) et à base de structures (Docking et 3D-QSAR) pour la prédiction et le développement de nouveaux inhibiteurs puissants de l'acétylcholinestérase (AChEI) et de kinases dépendantes des cyclines 5 (CDK5). Ces deux cibles thérapeutiques sont parmi les cibles désignées dans le traitement de la maladie d'Alzheimer. D'après les résultats présentés et discutés dans cette étude, il nous est apparu que ces méthodes, bien que différentes dans leur approche, proposent un intérêt dans la découverte de nouveaux *hits* désignés dans le traitement de la maladie d'Alzheimer et qui sont susceptibles d'obtenir des résultats cliniques.

**Mots clés :** Alzheimer, AChE, CDK5, Criblage virtuel, SAR/SPR, AN-QSAR, 3D-QSAR, Docking.

## Résumé

La maladie d'Alzheimer (MA), forme de démence la plus répandue, est caractérisé par le déclin de la capacité cognitive suffisamment grave pour gêner les activités de la vie quotidienne. En 2019, ADI (*Alzheimer's Disease International*) estime à plus de 50 millions le nombre de personnes atteintes de démence dans le monde, ce chiffre devrait passer à 152 millions d'ici 2050. Dans notre étude nous avons utilisé les deux méthodes de criblage virtuel à base de ligands (une étude qualitative SAR et une étude quantitative QSAR) et à base de structures (Docking et 3D-QSAR) pour la prédiction et le développement de nouveaux inhibiteurs puissants de l'acétylcholinestérase (AChEI) et de kinases dépendantes des cyclines 5 (CDK5). Ces deux cibles thérapeutiques sont parmi les cibles désignées dans le traitement de la maladie d'Alzheimer. D'après les résultats présentés et discutés dans cette étude, il nous est apparu que ces méthodes, bien que différentes dans leur approche, proposent un intérêt dans la découverte de nouveaux *hits* désignés dans le traitement de la maladie d'Alzheimer et qui sont susceptibles d'obtenir des résultats cliniques.

**Mots clés :** Alzheimer, AChE, CDK5, Criblage virtuel, SAR/SPR, ANN-QSAR, db-CICA, Docking.

## Abstract

Alzheimer's disease (AD), the most common form of dementia, is characterized by a decline in cognitive ability severe enough to interfere with activities of daily living. In 2019, ADI (Alzheimer's Disease International) estimates that more than 50 million people worldwide are living with dementia. This number will expected to rise to 152 million by 2050. In our study we used the two methods of virtual screening based on ligand (a qualitative SAR / SPR study, a quantitative QSAR study) and based on structure (Docking, db-CICA method) for the prediction and the development of new inhibitors. Powerful acetylcholinesterase (AChEI) and cyclin-dependent kinase 5 (CDK5), these two therapeutic targets are one of the targets designated in the treatment of Alzheimer's disease. From the results presented and discussed in this present investigation, we have found that these methods, although different in their approach, offer a definite interest in the discovery of new *hits*, which are susceptible obtain clinical results to treat Alzheimer's disease.

**Keywords:** Alzheimer, AChE, CDK5, virtual screening, SAR/SPR, ANN-QSAR, db-CICA, Docking.

## ملخص

مرض الزهايمر هو أكثر أنواع الخرف شيوعًا، يتميز بتراجع القدرة المعرفية الشديدة والتي تتعلق بأنشطة الحياة اليومية. في عام 2019 المنظمة الدولية لمرض الزهايمر قدرت أن أكثر من 50 مليون شخص حول العالم يعانون من الخرف؛ هذا العدد من المتوقع ان يرتفع الى 152 مليون شخص بحلول عام 2050.

خلال هذه الدراسة قمنا بتطبيق مجموعة من الطرق والأساليب الحسابية المستخدمة في تصميم الأدوية وذلك بالاعتماد على البنية الكيميائية للمركبات اعتمادا على دراسة كمية و كيفية (SAR/SPR, QSAR) وعلى البنية البروتينية للمستقبل (Docking, db-CICA) وذلك من أجل التنبؤ والحصول على مثبطات جديدة ضد المستقبل أستيل كولين (AChE) و CDK5 حيث أن هذين المستقبلين يعتبران من بين المستقبلات المهمة في علاج مرض الزهايمر.

من خلال النتائج التي تحصلنا عليها والتي تم مناقشتها في هذا البحث، تبين لنا ان هذه الطرق على الرغم من اختلافها في مقارباتها إلا أننا من خلالها تحصلنا على مركبات جديدة وواعدة، لها فرصة في أن تنتقل إلى التجارب السريرية لعلاج المرض الزهايمر.

**الكلمات المفتاحية :** الزهايمر, AChE, CDK5, SAR/SPR, ANN-QSAR, Docking, db-CICA.