

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET
DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DE SAÏDA - Dr MOULAY TAHAR



FACULTE DES SCIENCES
Département de Chimie

MEMOIRE

Présenté par :
MOSTEFAI NORIA

En vue de l'obtention du

Diplôme de Master en Chimie

Spécialité : Chimie Théorique et Computationnelle

Thème

**Elaboration des modèles QSAR qui
prédisent une activité associée à la maladie
d'Alzheimer.**

Soutenu le 14-07-2021

devant le jury composé de:

Présidente

Pr. MOSTEFAI Asmaa

Université de Saida

Encadrant

Dr .GUENDOZI Abdelkrim

Université de Saida

Examineur

Dr. HADJI Djebbar

Université de Saida

Année universitaire 2020/2021

Remerciements



Tout d'abord, je remercie Dieu le tout puissant pour la volonté, la santé et la patience qu'il m'a donné durant toutes ces années d'études. Je tiens à exprimer mes vifs remerciements à mon encadreur Mr. Guendouzi Abdelkrim pour son aide, ses conseils.

Je tiens également à remercier les membres de jury de mon mémoire d'avoir accepté d'examiner et de juger ce travail.

Je remercie sincèrement tous mes enseignants durant les cinq années d'étude, leurs aides, leurs conseils, et leurs disponibilités ont été pour moi un point fort dans la réussite de mon travail.

DEDICACE

Je dédie ce modeste travail, Aux deux personnes les

plus

chères au monde pour moi, mon père et ma mère. Que

Dieu vous protège, Amén.

A mon frère Yahia

A ma seule sœur Bochra

A tous mes chères amies :Hadjira, Widad, Fatima,

Maroua, Souad, Halima

A ma famille Mostefai

pour leur encouragement

permanente et leur soutien moral.

NORIA

الملخص :

مرض الزهايمر هو اضطراب تنكسي في الدماغ يجمع بين اضطرابات الذاكرة السائدة ، والاضطرابات المعرفية أو الاضطرابات السلوكية التي تؤثر على الحياة اليومية للمرضى. السبب الجذري غير واضح ومشاركة الإنزيمات المختلفة في الحالات المرضية تؤكد مدى تعقيد المرض. تحظى تقنيات علاقة النشاط الكمي (QSAR) في تصميم الأدوية ضد مرض الزهايمر باهتمام كبير.

ABSTRACT

Alzheimer's disease (AD) is a degenerative disease of the brain that combines predominant memory disorders, cognitive disorders or behavioral disorders that affect the daily lives of patients.

The root cause is unclear, and the involvement of different enzymes in the pathological conditions underscores the complexity of the disease. Quantitative Activity Relation (QSAR) techniques are of great interest in the design of drugs against Alzheimer's disease, in this work used the method of quantum chemistry via DFT to calculate molecular descriptors for the development of QSAR model to predict the bioactivity (IC50-50% inhibition concentration) of Pyrimidine derivatives as an inhibitor of Alzheimer's disease. The QSAR model linked the molecular parameters of the compounds studied to their activity.

RESUME :

La maladie d'Alzheimer (MA) est une maladie dégénérative du cerveau qui associe des troubles prédominants de la mémoire, des troubles cognitifs ou des troubles du comportement qui affectent le quotidien des patients.

La cause profonde n'est pas claire et l'implication de différentes enzymes dans les conditions pathologiques souligne la complexité de la maladie. Les techniques de relation quantitative d'activité (QSAR) sont d'un grand intérêt dans la conception de médicaments contre la maladie d'Alzheimer, dans ce travail a utilisé la méthode de la chimie quantique via DFT pour calculer des descripteurs moléculaires pour le développement de modèle QSAR pour prédire la bioactivité (CI50-concentration d'inhibition 50 %) de dérivées de Pyrimidine comme inhibiteur de la maladie d'alzheimer. Le modèle QSAR a lié les paramètres moléculaires des composés étudiés à leur activité.

Table des matières

INTRODUCTION GENERALE	1
-----------------------------	---

BIBLIOGRAPHIE

CHAPITRE I

LES DERIVEES DE PYRIMIDINE ET LA MALADIE D'ALZHEIMER

I.1 INTRODUCTION	5
------------------------	---

I.2 GENERALITE SUR LES PYRIMIDINES.....	5
---	---

I.2.1 DEFINITION	5
------------------------	---

I.2.2 METHODES GENERALES DE SYNTHESE DES PYRIMIDINES	6
--	---

I.2.3 INTERET BIOLOGIQUE DES PYRIMIDINES.....	7
---	---

I.3 LA MALADIE D'ALZHEIMER.....	7
---------------------------------	---

I.3.1 HISTORIQUE	7
------------------------	---

I.3.2 DEFINITION	9
------------------------	---

I.3.3 PREVALENCE.....	9
-----------------------	---

I.3.3.1 En fonction de l'âge.....	9
-----------------------------------	---

I.3.3.2 En fonction des pays.....	10
-----------------------------------	----

I.3.3.2 En fonction du sexe.....	10
----------------------------------	----

I.3.4 LA PHYSIOPATHOLOGIE	11
---------------------------------	----

I.3.4.1 Les plaques séniles	11
-----------------------------------	----

I.3.4.2 Les dégénérescences neurofibrillaires (DNF).....	11
--	----

I.3.4.3 Les neurotransmetteurs.....	12
-------------------------------------	----

I.3.5 LES PHASES DE LA MALADIE D'ALZHEIMER	12
--	----

I.3.6 FACTEURS DE RISQUE ET FACTEURS PROTECTEURS.....	13
---	----

I.3.7 SYMPTOMES.....	15
----------------------	----

I.3.8 TRAITEMENT	19
------------------------	----

BIBLIOGRAPHIE

CHAPITRE II:

LES METHODES UTILISEES DANS LA MODELISATION MOLECULAIRE.

II.1 INTRODUCTION.....	25
------------------------	----

II.2 METHODES QUANTIQUES	25
--------------------------------	----

II.2.1 EQUATION DE SCHRÖDINGER	26
--------------------------------------	----

II.2.2 L'APPROXIMATION DE BORN-OPPENHEIMER	26
--	----

II.2.3 LA THEORIE DE LA FONCTIONNELLE DE LA DENSITE (DFT).....	27
--	----

II.2.4 LES BASES D'ORBITALES.....	27
-----------------------------------	----

II.3 QSPR/QSAR.....	28
---------------------	----

II.3.1 DEFINITION.....	28
------------------------	----

II.3.2 PRINCIPE.....	28
----------------------	----

II.3.3 METHODOLOGIE GENERALE D'UNE ETUDE QSAR/QSPR	28
--	----

II.3.4 METHODES STATISTIQUES.....	34
-----------------------------------	----

II.3.4 TEST DE LA SIGNIFICATION GLOBALE DE LA REGRESSION	37
--	----

II.3.5 VALIDATIONS DES MODELES	39
II.3.6 APPLICATION DE QSAR.....	40

BIBLIOGRAPHIE

CHAPITRE III :

ETUDE QUANTITATIVE DES

PROPRIETES QSAR D'UNE SERIE DE PYRIMIDINES.

III.1 INTRODUCTION:	49
III.2 METHODOLOGIE :.....	49
III.2.1. BASE DE DONNEES	49
III.2.2. CALCUL DES DESCRIPTEURS	56
III.2.3 ANALYSE STATISTIQUE.....	56
III.3 RESULTATS ET DISCUSSIONS	58
III.3.1 ANALYSE STATISTIQUES DE MEILLEUR MODELE MLR	58
III.4.VALIDATION DU MEILLEUR MODELE QSAR	63
III.4.1 VALIDATION INTERNE	63
III.4.2. VALIDATION EXTERNE	63
BIBLIOGRAPHIE	64
CONCLUSION GENERALE.....	ERROR! BOOKMARK NOT DEFINED.

Liste des figures

Figure 1 : condensation de Claisen-Schmidt avec des aldéhydes	6
Figure 2 : 1,3-diaryl-propenones réagir avec de la guanidine	7
Figure 3 : le 2-amino pyrimidines	7
Figure 5 : Dégénérescence neurofibrillaire dessinée par Alois Alzheimer	8
Figure 4 : Dr. Alzheimer et la patiente Auguste D	8
Figure 6: Anatomie du cerveau	9
Figure 7: Impact global de la démence	10
Figure 8 : Evolution des symptômes de la MA en fonction du MMSE et de l'âge	16
Figure 9: Angle dièdre dirigeant l'activité catalytique des complexes	31
Figure 10 : Les descripteurs moléculaires	33
Figure 11 : Regression by Artificial Neural Network.	37
Figure 12 : Différentes étapes d'étude QSAR/QSPR.	41
Figure 13 : Structure générale de Pyrimidine	50
Figure 14 : Diagramme de la contribution en pourcentage de chaque descripteur dans le modèle pIC50	60
Figure 15 : Corrélation entre l'activité biologique expérimentale et prédite.	63

Liste des Tableaux :

Tableau 1 : Les descripteurs topologiques calculés dans étude QSAR	30
Tableau 2 : Table d'analyse de variance.	39
Tableau 3 : Structures et activités d'un ensemble de données	55
Tableau 4 : Descripteurs moléculaires utilisés dans l'étude QSAR.....	56
Tableau 5 : Valeurs pIC50 expérimentales avec les descripteurs sélectionnés.	58
Tableau 6 : Modèles sélectionnés et paramètres statistiques des corrélations entre les propriétés moléculaires et l'activité biologique.	59
Tableau 7 : Les valeurs expérimentales, prédites et résiduelles de pIC50 par MLR	62

Liste des abréviations :

A β : Amyloïde β .

MA : Maladie d'Alzheimer.

QSAR : La relation quantitative structure activité.

OMS : Organisation Mondiale de la Santé.

APP : Amyloid Precursor Protein.

AVC : Accident Vasculaire Cérébral.

DNF : Dégénérescences Neurofibrillaires.

AVQ : Activités de la vie quotidienne.

MA : Maladie d'Alzheimer.

MCI : Mild Cognitive Impairment (ou trouble cognitif léger).

MMSE : Mini Mental State Examination.

SPCD : Symptômes psychologiques et comportementaux des démences.

ADI : Alzheimer's Disease International.

IRM : Imagerie à Résonance Magnétique.

TEP : tomographie par émission de positrons.

DFT : théorie de la fonctionnelle de la densité.

CoMFA: Comparative molecular field analysis.

CoMSIA: comparative molecular similarity indices analysis.

INTRODUCTION GÉNÉRALE

Introduction générale

Selon l'Organisation Mondiale de la Santé (OMS), près de 50 millions de personnes dans le monde atteinte de démence (1).

La maladie d'Alzheimer (MA) est la plus fréquente forme de démence (2) ; cette maladie a été décrite pour la première fois en 1906 par neuropsychiatre allemand Alois Alzheimer (3), il s'agit d'une maladie neurodégénérative (des lésions cérébrales progressives entraînent la mort des neurones), qui se caractérise par une perte progressive de la mémoire et de certaines fonctions intellectuelles (cognitives), qui affectent les activités de la vie quotidienne (4).

La chimie computationnelle, est une branche de la chimie qui utilise les lois de la mécanique quantique et de la mécanique classique ; dans un programme informatique spécifique, afin de calculer des molécules libres, des solides et des agrégats atomiques (ou clusters) à l'aide d'ordinateurs, la structure et les propriétés des substances chimiques telles que la surface, etc. appliquent autant que possible ces procédures à des problèmes chimiques réels (5) ; la chimie médicinale est une discipline liée à la conception, Préparation et interprétation du mode d'action du médicament, c'est une discipline entre autres choses, pour établir une relation entre la structure chimique des médicaments et leur activité biologique (6).

De plus, la fabrication d'un médicament nécessite généralement 10 à 15 ans de temps de recherche. À ces fins, des techniques de conception de médicaments assistées par ordinateur telles que QSAR (relation quantitative structure-activité) ont été largement utilisées et avec succès (7), la relation quantitative structure-activité (QSAR) est l'un des outils les plus pratiques en chimie physique computationnelle, l'idée de base de QSAR réside dans la possibilité d'une relation entre un ensemble de descripteurs et des réponses moléculaires.

Le QSAR peut être considéré comme une règle de calcul informatique, il décrit quantitativement l'activité biologique à partir de descripteurs moléculaires, il est souvent utilisé pour prédire l'activité biologique de nouveaux composés (8).

Les hétérocycles actifs jouent un rôle thérapeutique très important, en raison de leur intérêt pharmaceutique, la chimie des hétérocycles est en développement croissant(9).

Les hétérocycles jouent également un rôle important dans la conception et la découverte de nouveaux composés physiologiques / pharmaco-logiquement actifs; ce sont des fragments communs de la grande majorité des médicaments commercialisés (10), ceci est le reflet du rôle central que jouent les hétérocycles dans la conception moderne des médicaments, ils

peuvent servir d'outils utiles pour manipuler la lipophile, la polarité et la capacité de liaison hydrogène des molécules, ce qui peut améliorer les propriétés pharmacologiques, pharmacocinétiques, toxicologiques et physicochimiques des médicaments candidats et, finalement, des médicaments (11).

Les Pyrimidines est l'un des composés hétérocycliques les plus importants, ils présentent diverses activités pharmacologiques, comme anticancéreuses, Antibactérien, antifongique, antimicrobien, anti-inflammatoire, et utiles pour traiter les maladies neurologiques, y compris la maladie d'Alzheimer (12).

Dans ce cas, le but de ce travail est de développer un modèle QSAR puissant et fiable, respectant toutes les étapes de la recherche QSAR, on va utilisé une série de Pyrimidine pour inhiber la maladie d'Alzheimer, et pouvant être utilisé pour envisager de nouveaux dérivés efficaces comme médicaments pour la traitement de la MA.

Ce manuscrit est composé en trois chapitres :

- Dans le premier chapitre nous présenterons des généralités sur la maladie d'Alzheimer, Symptômes, stades et physiopathologie, diagnostic et traitement de la maladie.
- Dans le deuxième chapitre nous donnerons un aperçu sur le QSAR, de ses principes et de ses applications et les différentes étapes à suivre pour y parvenir.
- Le troisième chapitre est la partie pratique recherche quantitative caractéristiques QSAR d'une série de dérivés Pyrimidine, les résultats obtenus et leur discussion.
- Finalement, nous terminerons ce travail par une conclusion générale et quelques perspectives.

Bibliographie

- (1) Santé Organisation Mondiale, Démence. <https://www.who.in>. [En ligne] 2020. [Citation : 10 6 2021.]
- (2) Prince, M. *The Global Impact Of Dementia*. Alzheimer Disease International. 2015. World Alzheimer Report.
- (3) Christelle, C. *Les capacités musicales dans la maladie d'Alzheimer*. Université Nice Sophia Antipolis. Nice : s.n., 2012.
- (4) Cummings, J. Alzheimer's Disease. *The New England journal of medicine*. 2004, Vol. 351, 18, pp. 56-67.
- (5) Rivail, L. *Elements de chimie quantique à l'usage des chimistes*. CNRS, 1999.
- (6) Mackey.Z, Hansell.E,James.E. Antitrypanosomal and cysteine protease inhibitory activities of alkyldiamine cryptolepine derivatives. *Bioorganic & medicinal chemistry letters*. 2012, Vol. 22, 19, pp. 6256-6260 .
- (7) Chtita, Samir. *Modélisation de molécules organiques heterocycliques biologiquement actives par des méthodes QSAR/QSPR*. Université Moulay Ismail. Meknès : s.n., 2017. Thèse.
- (8) Wermuth.G. Designing prodrugs and bioprecursors. [éd.] Third Edition. *The Practice of Medicinal Chemistry* . 2008.
- (9) Van Bambeke.F, Verhaegen.J, Tyteca.D, Auckenthaler.R, Louv. Med. 119, 259, 2000.
- (10) Yusuf.M, Jain.P. Synthesis and biological significances of 1,3,4 thiadiazolines and related heterocyclic compounds. *Arabian Journal of Chemistry*.7, 525, 2014.
- (11) Gomtsyan.A, *Chemistry of Heterocyclic Compounds*. 7, 48, 2012.
- (12) Kushwaha.N, Swatantra.K, Rai.A, *International Journal of ChemTech Research*, 4, 517, 2012.

CHAPITRE I

LES DÉRIVÉES DE PYRIMIDINE ET LA MALADIE D'ALZHEIMER

I.1 Introduction

La valeur des Pyrimidines réside non seulement dans ses riches propriétés chimiques, mais aussi de nombreux types, mais ont également de nombreuses caractéristiques biologiques importantes (1), ils représentent une grande classe de composés d'intérêt, il est assez important en raison de sa large activité biologique (2), ces composés ont été étudiés et développés dans le cadre de différents projets de recherche, ils présentent des activités anti-tumorale, antivirale, antibactérienne, anti-hypertensive, anti-inflammatoire(3), la présente invention concerne les pyrimidines, utilisables pour prévenir et traiter des maladies neurodégénératives, comme la maladie d'Alzheimer (4).

La démence est un syndrome, généralement chronique ou progressif, dans lequel la fonction cognitive (la capacité d'effectuer des opérations de réflexion) est altérée, ce qui est plus grave que prévu au cours du vieillissement normal ; elle affecte la mémoire, le raisonnement, l'orientation, la compréhension, l'arithmétique, la capacité d'apprentissage, le langage et le jugement, la conscience n'est pas affectée, la détérioration du contrôle émotionnel du comportement social ou de la motivation s'accompagne généralement d'une altération de la fonction cognitive et précède parfois une altération de la fonction cognitive.

La démence est causée par une série de maladies et de blessures qui affectent principalement ou dans une moindre mesure le cerveau, comme la maladie d'Alzheimer (5).

La Maladie d'Alzheimer constitue la cause la plus fréquente de démence ; elle représente un véritable problème de santé publique (6).

I.2 Généralité sur les Pyrimidines

I.2.1 Définition

Les pyrimidines (ou 1,3-diazines) sont des composés hétérocycliques azotés ($C_4H_4N_2$) comportant deux atomes d'azote ; ces systèmes hétérocycliques se retrouvent notamment dans les bases nucléiques constituant les molécules d'ADN et d'ARN comme la cytosine, la thymine et l'uracile (Figure 01), ces dérivés peuvent être obtenus par condensation multi-composants d'un aldéhyde, un β -dicarbone et l'urée en présence d'une quantité catalytique d'acide, certaines de ces combinaisons peuvent conduire à la formation de pyrimidones, analogues des pyridones (7).

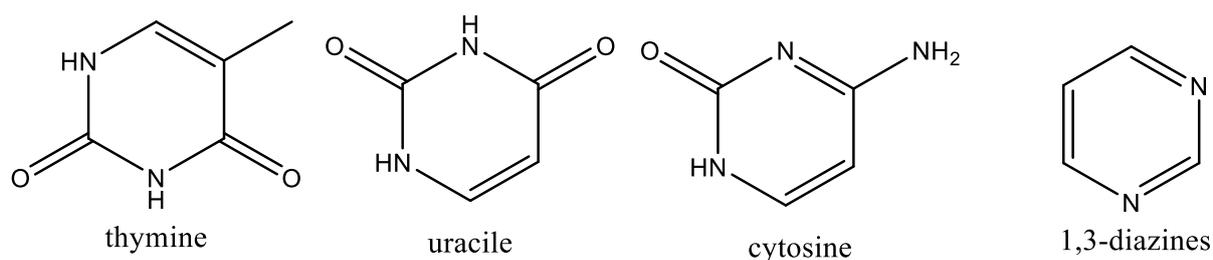


Figure 1 : bases azotées constituant les cycles glucidiques d'un acide nucléique

I.2.2 Méthodes générales de synthèse des pyrimidines

Une brève description de la méthode générale de synthèse des pyrimidines examinées ci-dessous :

1) 1, 4-dihydro-4-phényl-2, 6-diméthyl-3, 5-diacétylpyridines ont été converties en chalcones par condensation de *Claisen-Schmidt* avec des aldéhydes, les chalcones résulté ont été cyclisés avec la guanidine pour donner aminopyrimidines (8) (Figure 2).

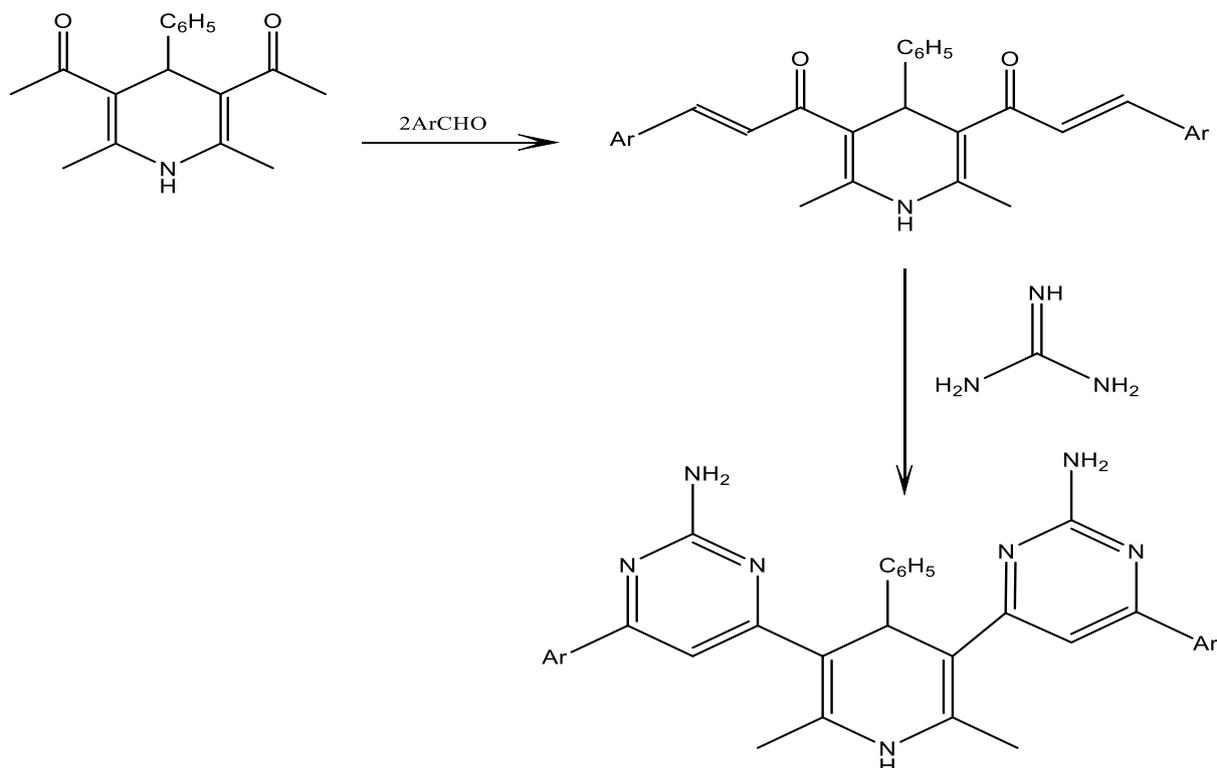


Figure 1 : condensation de Claisen-Schmidt avec des aldéhydes

2) 1,3-diaryl-propénones réagissent avec de la guanidine en les chauffant au reflux ensemble dans un milieu alcoolique basique pour donner dihydropyrimidines, qui lors de l'oxydation avec un rendement de H_2O_2 , 4, 6-diaryl-2-aminopyrimidines (Figure 3).

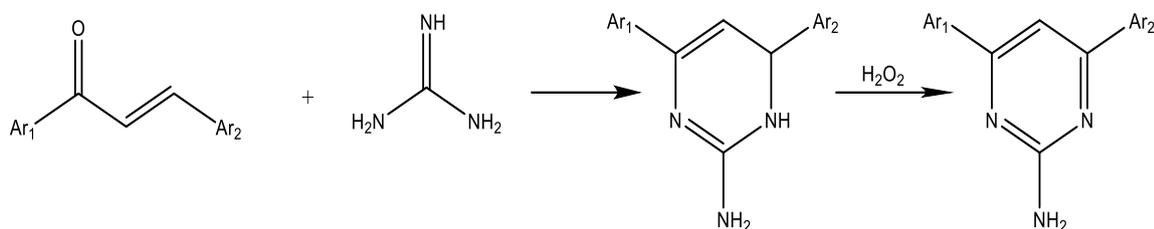


Figure 2 : 1,3-diaryl-propénones réagissent avec de la guanidine

3) Guanidine réagit avec α -cétoesters, β -dicétones, les esters cyanoacétiques et α, β des composés carbonyles insaturés pour donner le 2-amino pyrimidines généralement de bons rendements (9). (Figure 4).

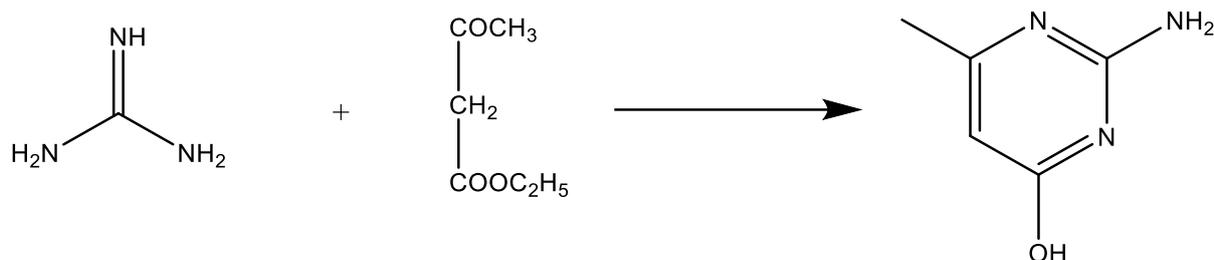


Figure 3 : le 2-amino pyrimidines

I.2.3 Intérêt biologique des pyrimidines

Potentiel thérapeutique de pyrimidines: une revue de la littérature a révélé que différents substitués pyrimidines sont connus pour posséder antimicrobien (10), autres activités diverses. étant donné ci-dessous est un bref compte rendu des diverses modifications rapporté, anti-inflammatoire (11), anticancéreux (12), antiviraux, antituberculeux, antipaludéens et, les maladies neurologiques tel que la maladie d'Alzheimer. Pyrimidine noyau, qui a montré une variété d'activités biologiques et pharmacologiques (13).

I.3 La maladie d'Alzheimer

I.3.1 Historique

En 1906, Auguste D. fut la première personne atteinte d'un état de démence 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. Aloïs Alzheimer, par la suite, cette maladie portera son nom (14).

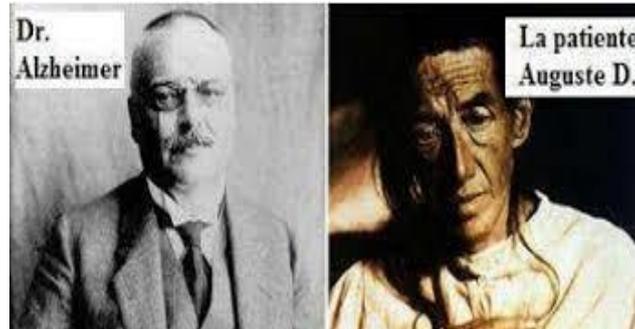


Figure 4 : Dr. Alzheimer et la patiente Auguste D

En effectuant une autopsie de son cerveau, il identifie en plus d'une atrophie sévère, deux facteurs histologiques caractéristiques de cette maladie : des « plaques » et des dégénérescence neurofibrillaire (15) ; les « plaques » dites « plaques séniles » sont composées de dépôts extracellulaires de peptides neurotoxiques, les dégénérescences neurofibrillaires correspondent à une accumulation de fragments de protéines fibreuses dans les cellules nerveuses (16).

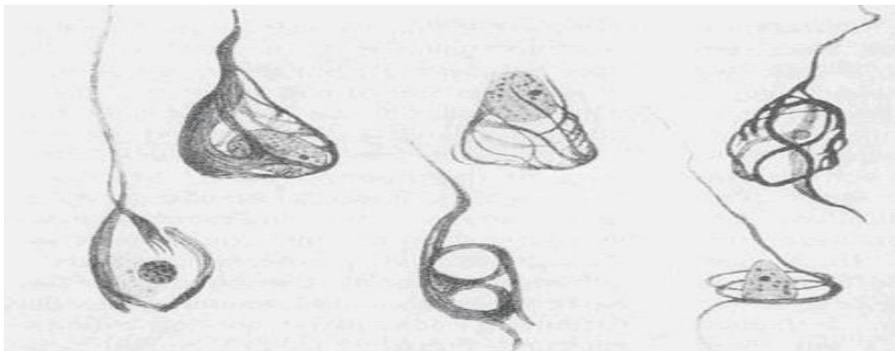


Figure 5 : Dégénérescence neurofibrillaire dessinée par Aloïs Alzheimer

Par la suite, d'autres chercheurs viendront confirmer ses découvertes, et un deuxième cas identique en 1911 viendra valider sa théorie, pendant plus d'un demi-siècle, l'étude de la maladie d'Alzheimer est restée en suspens ; c'est à partir des années 1970-1980 qu'est apparue une nouvelle dynamique grâce aux progrès de la médecine et en raison de l'augmentation des cas en lien avec le vieillissement de la population, ainsi, dans les années 1980, ont été déterminées les formations biologiques à l'origine des plaques (la protéine **bêta-amyloïde**) et des dégénérescences neurofibrillaires (la protéine **Tau** hyperphosphorylée) ; 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 (17).

I.3.2 Définition

La maladie d'Alzheimer est une maladie neurodégénérative (atteinte cérébrale progressive conduisant à la mort neuronale) caractérisée par une perte progressive de la mémoire et de certaines fonctions intellectuelles (cognitives) conduisant à des répercussions dans les activités de la vie quotidienne (18).

La maladie d'Alzheimer est une altération progressive de la mémoire et de l'idéation, suffisamment marquée pour handicaper les activités de la vie quotidienne, apparue depuis au moins 6 mois et associée au moins à un trouble suivant : trouble du langage, trouble du calcul, trouble du jugement, altération de la pensée abstraite, trouble praxique, trouble gnosique ou modification de la personnalité (19).

Le cerveau est composé des milliards de cellules nerveuses (neurones) qui forment un immense réseau de communication par des connexions entre elles, il existe toute une organisation des cellules nerveuses réparties en groupes qui ont tous un rôle précis (pensée, apprentissage, mémoire, vue, odorat, ouïe), chaque cellule nerveuse est approvisionnée, génère de l'énergie, construit un équipement et se débarrasse des déchets, traitement et stockage des informations, communication entre les cellules, tout cela fait fonctionner le système nerveux, qui doit être convenablement oxygéné (20).

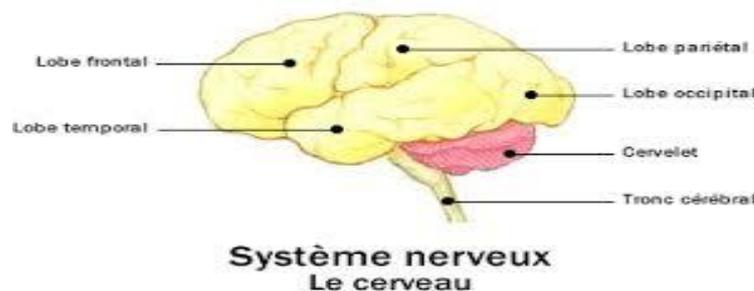


Figure 6: Anatomie du cerveau

I.3.3 Prévalence

I.3.3.1 En fonction de l'âge

La maladie d'Alzheimer touche au moins de 1 % des personnes avant 65 ans (forme héréditaire) le plus souvent, la maladie débute chez les personnes au-dessus de 65 ans,

environ 6 % des personnes, et 15 % de la population de plus de 80 ans mais ces chiffres diffèrent en fonction des pays (21).

I.3.3.2 En fonction des pays

Environ de 50 millions de personnes atteintes de démences dans le monde d'après l'OMS, et 152 millions à l'horizon 2050, selon le rapport 2015 de l'ADI (Alzheimer's Disease International) la maladie d'Alzheimer est la 1ère cause de démence (60 à 70)% des cas.

Le nombre de nouveaux cas de démences en 2015 : 9,9 millions, soit 1 nouveau cas toutes les 3 secondes.

Répartition des nouveaux cas en 2015 :

- 49 % en Asie
- 25 % en Europe (en baisse par rapport aux chiffres de 2012)
- 18 % en Amérique
- 8 % en Afrique

Le nombre de cas de démence va presque doubler tous les 20 ans (22).

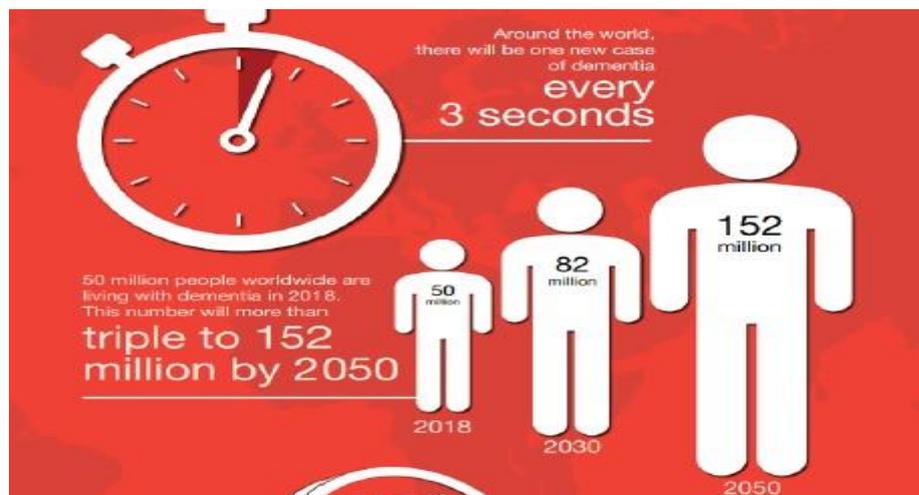


Figure 7: Impact global de la démence

I.3.3.2 En fonction du sexe

Les femmes ont un risque plus élevé de développer une maladie d'Alzheimer que les hommes (23), particulièrement après l'âge de 80 ans, dans l'étude **Paquid** (personne âgée quid), l'incidence de la maladie d'Alzheimer était, avant 80 ans, plus élevée chez les hommes que chez les femmes, alors que c'est l'inverse après 80 ans ; cette différence d'incidence en fonction du sexe pourrait être expliquée par des différences biologiques et hormonales, avec

notamment un possible effet des œstrogènes, mais également des différences socioculturelles, l'espérance de vie, plus élevée chez les femmes que chez les hommes, pourrait également expliquer les résultats (24).

I.3.4 La Physiopathologie

A ce jour les causes de la maladie d'Alzheimer restent encore mal connues ; en revanche, il est établi qu'avant même l'apparition des premiers symptômes, les neurones sont affectés par deux types de lésions : les plaques amyloïdes que l'on retrouve entre les neurones, et la dégénérescence neurofibrillaire que l'on retrouve à l'intérieur des neurones, ces deux lésions correspondent à des amas de protéines qui se forment lors du processus normal du vieillissement. Cependant, dans les maladies de type Alzheimer, ces protéines s'accumulent en beaucoup plus grande quantité (25).

I.3.4.1 Les plaques séniles

Les plaques séniles sont constituées d'agrégats de peptides **β -amyloïde** (peptide A β) de lésions neurofibrillaires ; le peptide A β est présent dans le cerveau à faible concentration, il provient d'un précurseur protéique transmembranaire, l'**APP** (Amyloid Precursor Protein) clivé par deux protéases (β - et γ - protéase) ; chez les personnes atteintes de la maladie d'Alzheimer, le peptide A β s'accumule anormalement et forme des dépôts extracellulaires.

Initialement, ces peptides se déposent de façon inorganisée formant ce qu'on appelle des plaques diffuses, ces plaques n'interagissent pas encore avec les neurones, astrocytes et microglies les entourant, au fur et à mesure de la pathologie, les plaques se condensent, s'organisent et commencent à interagir avec les cellules neuronales environnantes, ces plaques dites primitives continuent leur évolution et constituent éventuellement les plaques séniles constituées d'un cœur de peptide A β entouré d'un enlacement de peptide et neurones dystrophiques.

A l'état final, il ne reste plus qu'une plaque atrophiée et extrêmement dense ; ces plaques « burned-out » sont peu nombreuses et vierge d'éléments cellulaires (26).

I.3.4.2 Les dégénérescences neurofibrillaires (DNF)

Les dégénérescences neurofibrillaires résultent de l'agrégation intraneuronale de protéines **Tau** hyperphosphorylées.

La protéine **Tau** est présente abondamment dans les neurones, principalement les axones, elle permet la régulation et la stabilité du réseau de microtubules en s'y liant permettant le

transport neuronal des vésicules vers les synapses ; la seule modification post-transcriptionnelle de cette protéine est la phosphorylation.

Lorsque le peptide **Tau** est hyperphosphorylé, il est emprisonné dans les microtubules et perturbe le transport des vésicules ; peu à peu il diminue la fonctionnalité des axones et des dendrites. De plus le peptide hyperphosphorylé s'auto-aggrège formant des agrégats intermédiaires, des paires de filaments hélicoïdaux et enfin de dégénérescences neurofibrillaires, la phosphorylation de ce peptide est régulée par des **kinases** et des phosphatases. Il est décrit que le peptide A β induit la cascade d'hyperphosphorylation du peptide **Tau** par les **kinases** (27).

I.3.4.3 Les neurotransmetteurs

Les cellules nerveuses principalement et précocement détruites lors du développement de la maladie d'Alzheimer sont celles qui utilisent l'acétylcholine comme substance chimique de transmission car elles sont nombreuses au niveau de l'hippocampe et du cortex cérébral ; cela entraîne une diminution de la concentration de ce neurotransmetteur dans le cerveau (28),

L'acétylcholine est un neurotransmetteur excitateur très répandu qui, au niveau du système nerveux central, est entre autres impliqué dans l'éveil, l'attention, la colère, l'agression ; ce qui explique certains symptômes de la maladie, d'autres neurotransmetteurs sont également altérés lors de cette pathologie (29).

I.3.5 Les phases de la maladie d'Alzheimer

Il existe une grande variabilité interindividuelle concernant la progression des troubles liés à la maladie d'Alzheimer ; certaines personnes déclinent plus rapidement que d'autres, ou inversement, certains facteurs favorisent le ralentissement des déclinés (ex., niveau d'étude, activités stimulantes, etc.), il est néanmoins possible de dégager un profil général de l'évolution de la maladie d'Alzheimer que l'on peut découper en quatre stades (30) en fonction du MMSE (Mini Mental State Examination)

I.3.5.1 Le stade pré-clinique (ou phase asymptomatique)

Cette phase peut s'étendre de (15 à 25) ans durant lesquels les lésions cérébrales apparaissent silencieusement ; progressivement, quelques symptômes plus ou moins décelables émergent comme des oublis plus fréquents, l'égarément des objets, le manque du mot, les difficultés de pensées abstraites, l'humeur changeante, le manque d'intérêt, une hypothèse serait que des mécanismes de compensation neuronaux interviendraient durant cette phase pour compenser les pertes.

I.3.5.2 Le stade léger (MMSE > 20)

Cette phase peut varier de 2 à 4 ans ; lorsque les lésions cérébrales ont atteint le seuil symptomatique, les manifestations cliniques de la démence apparaissent, traduisant l'insuffisance des mécanismes compensatoires. Ainsi, les troubles de la mémoire à court terme et de la mémoire épisodique revêtent une plus grande importance, la personne exprime des difficultés à retenir des informations nouvelles, à suivre des conversations, à faire de nouveaux apprentissages, les troubles de l'attention et de la concentration, les troubles de l'orientation et le manque du mot s'aggravent, des manifestations comportementales apparaissent également : l'anxiété, le stress, la dépression, le désintérêt, qui résultent notamment de la prise de conscience des déclin.

I.3.5.3 Le stade modéré (MMSE compris entre 10 et 20)

A ce stade, peut durer de 2 à 6 ans, les déclin cognitifs ont un retentissement sur l'autonomie de la personne ; l'ensemble des symptômes décrits au stade léger s'aggrave, les pertes de la mémoire s'accroissent, la mémoire des faits récents est très altérée et la mémoire des faits anciens commence également à être perturbée, les troubles de l'orientation temporelle et spatiale s'aggravent, les changements de la personnalité et du comportement sont marqués, l'autonomie est compromise.

I.3.5.4 Le stade sévère (MMSE < 10)

Le stade sévère, peut durer de 2 à 4 ans, il est marqué par une détérioration grave de tous les symptômes décrits précédemment ; la mémoire des faits anciens jusqu'alors relativement préservée est altérée, la personne ne reconnaît plus ses proches ni les soignants, elle ne répond plus aux stimulations extérieures, la communication est quasi nulle, la dépendance est totale (31).

I.3.6 Facteurs de risque et facteurs protecteurs

Un facteur de risque est une caractéristique de notre mode de vie, de notre environnement et/ou de notre patrimoine génétique, et qui augmente le risque de développer une maladie ; les facteurs de risque ne causent pas la maladie, ils potentialisent seulement le risque de la développer, mais ça n'est pas parce qu'une personne a plusieurs facteurs de risque avérés qu'elle développera la maladie, chaque personne a des facteurs de risque différents en fonction de son mode de vie, son hérédité, son environnement ; l'apparition de la maladie d'Alzheimer se fait généralement avec une combinaison de plusieurs facteurs de risque, certains sont modifiables comme le tabagisme, l'hypercholestérolémie, et d'autres ne le sont pas comme l'âge (facteur de risque le plus connu pour la démence) ou la constitution

génétique (32).

I.3.6.1 L'âge

Est le facteur de risque principal de la maladie d'Alzheimer ; en effet, l'incidence de la maladie augmente après 65 ans, mais comme toute maladie chronique, on a aussi la susceptibilité individuelle qui peut favoriser l'apparition de la maladie d'Alzheimer, si un des deux parents est touché par la maladie, le risque de la développer est multiplié par 1.5, et si les deux parents le sont, on a un risque multiplié par 2 (33).

I.3.6.2 Le sexe

Les femmes ont un plus grand risque de développer la maladie d'Alzheimer que les hommes. Bien qu'on n'en connaisse toujours pas les raisons, on note parmi les facteurs déterminants potentiels le fait que les femmes vivent (en moyenne) plus longtemps que les hommes, les changements de taux d'œstrogène chez les femmes au cours de leur vie, et la fragilité physique et autres problèmes de santé qui tendent à être plus courants chez les femmes.

I.3.6.3 Les antécédents familiaux et la génétique

Un petit pourcentage (5-7) % de personnes atteintes de la maladie d'Alzheimer La forme « familiale » de la maladie (FFMA) (anciennement appelée « forme précoce » maladie d'Alzheimer"). À un moment donné de l'histoire familiale, certains gènes ont subi une mutation et ont développé les caractéristiques anormales qui causent la FFMA, ces gènes d'hérédité a une grande influence : si l'un des parents est FFMA, chaque enfant aura il y a 50% de chance d'hériter de cette maladie, si les deux parents ont FFMA, tous leurs enfants développent la maladie d'Alzheimer à l'âge adulte (34).

I.3.6.4 Les maladies vasculaires

Peut-être lié à la maladie d'Alzheimer tels que l'obésité, l'hypertension artérielle, le diabète de type 2 ; origine les accidents vasculaires cérébraux et les traumatismes crâniens peuvent également augmenter le risque la maladie d'Alzheimer, due à une diminution du flux sanguin. (35).

La liste est non exhaustive et selon les études d'autres facteurs peuvent être associés comme le tabac, l'alcoolisme chronique, la sédentarité, les anesthésies répétées et les microtraumatismes crâniens (36).

Parallèlement aux facteurs de risques, les études épidémiologiques ont mis en évidence des facteurs qui permettraient de protéger l'individu de la maladie d'Alzheimer, parmi-eux :

I.3.6.5 Le niveau d'éducation

Un niveau d'instruction élevé serait lié à une plus grande résistance à la démence (37).

I.3.6.6 L'alimentation

Une alimentation saine, type méditerranéenne, diminuerait le risque d'apparition de démence, comme une consommation régulière de poisson ou une alimentation riche en fruits et légumes (antioxydants). De plus, il est important d'avoir une consommation quotidienne d'eau, on retrouve aussi l'huile d'olive dans ce régime, elle sera la principale source de lipides, une forte consommation d'antioxydants augmente les taux de protéines dans le cerveau, ce qui protégera les cellules cérébrales de certains dommages ; suivre ce type de régime diminue le risque de développer des problèmes de mémoire et de pensée, ainsi que le développement de formes de démence (38).

I.3.6.7 Vie sociale

Différents facteurs diminueraient le risque de développer la maladie comme un niveau d'étude élevé, une activité professionnelle stimulante, un entourage important, un bon niveau d'éducation, des activités de loisir ; tous ces facteurs semblent retarder l'apparition des premiers symptômes et leurs sévérités, ils ne diminuent pas la perte en neurones mais permettent une bonne plasticité cérébrale, entraînant ainsi une adaptabilité permanente du cerveau (39).

I.3.6.7 Sport

La pratique d'une activité sportive régulière, telle que la marche serait un neuroprotecteur (40).

I.3.6.7 Caféine

Une consommation supérieure à 261mg/j de caféine réduirait de 36% le risque d'être atteint d'une démence (41).

I.3.7 Symptômes

I.3.7 .1 Les troubles cognitifs

Ils se définissent par la perte progressive et irrémédiable des fonctions cognitives, à savoir :

- La mémoire à court et à long terme (*Mnésie*).
- L'orientation spatio-temporelle.
- Le langage (*Phasie*).
- La reconnaissance (*Gnosie*).
- La gestuelle fine (*Praxie*) (42).

Dans les formes évoluées de la maladie, les troubles de la mémoire sont souvent accompagnés

du syndrome des 3A c'est-à-dire **aphaso-apraxo-agnosique**.

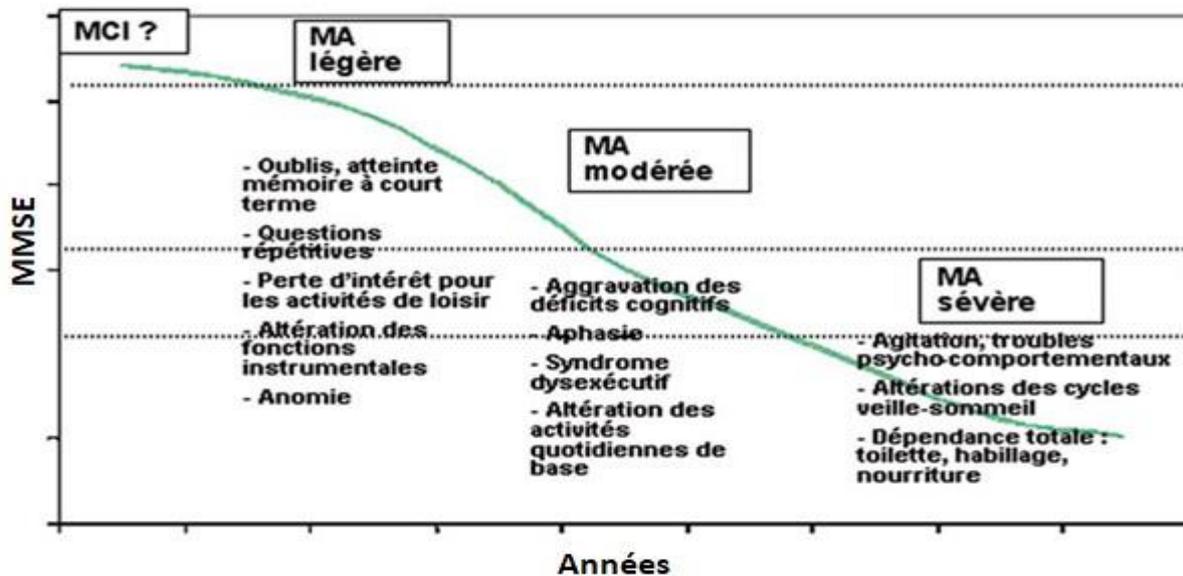


Figure 8 : Evolution des symptômes de la MA en fonction du MMSE et de l'âge

- **Les troubles de la mémoire :**

Le début de la maladie commence généralement avant l'apparition des premiers symptômes ; dans un premier temps, le malade va aller consulter son médecin pour des pertes de mémoire régulières et qui l'affectent dans sa vie quotidienne ; ces **troubles de mémoire** sont les signes cliniques les plus fréquents et surtout les premiers perçus, ils sont dû à l'atteinte de la région de l'hippocampe qui est le centre de contrôle de la mémorisation.

L'atteinte de la mémoire concerne surtout les événements récents, le patient aura du mal à garder en mémoire des faits nouveaux, il aura un défaut de stockage des souvenirs récents, en revanche les souvenirs anciens sont conservés plus longtemps, ils seront dégradés au cours de l'évolution de la maladie.

- **Les troubles praxiques**

Le patient peut aussi présenter des troubles dans l'exécution de ses gestes aussi appelé **apraxie**, au début cela va toucher des gestes difficiles comme l'écriture puis avec l'évolution de la maladie cela affectera même les gestes réflexes, comme mâcher ou avaler ; cette aggravation montre une diminution de l'autonomie et le patient aura donc besoin d'une aide pour les gestes les plus courants de la vie quotidienne comme se laver, s'habiller, manger.

- **Les troubles gnosiques**

Le malade aura des difficultés de reconnaissance ou aussi appelé **agnosie**, c'est-à-dire que le

patient sera incapable d'identifier des choses simples de la vie courante pouvant aller jusqu'à l'incapacité à reconnaître les personnes de son entourage proche ; cette **agnosie** peut être visuelle mais aussi tactile, olfactive et auditive, le patient présente souvent des troubles du raisonnement, du jugement ; le malade d'Alzheimer abandonnera les tâches les plus complexes comme gérer le budget, planifier des rendez-vous, des voyages, téléphoner à quelqu'un, il aura aussi du mal à se concentrer.

- **Les troubles du langage**

L'**aphasie** est le trouble du langage, le patient va présenter une diminution du vocabulaire, il cherchera plus souvent ces mots ou alors il va mettre un mot pour un autre ; cette **aphasie** ne concerne pas que le langage oral, elle va toucher aussi le langage écrit, le malade va faire des erreurs d'orthographe puis il va déformer les lettres, et les mots vont devenir méconnaissables (43).

- **La désorientation spatio-temporelle**

La désorientation **spatio-temporelle** est un signe clinique précoce ; elle s'effectue d'abord dans un grand environnement puis, dans un espace de plus en plus restreint jusqu'à sa propre habitation (44).

I.3.7 .2 Les troubles psycho-comportementaux

Les troubles psycho-comportementaux font partie des complications fréquentes de la MA ; au stade avancé, près de 80% des patients présentent des troubles psycho-comportementaux, ils se traduisent généralement par des **cris, une agitation, une irritabilité, une agressivité, une euphorie, un comportement d'opposition**, le malade aura des accès de colère sans raisons valables, il pourra présenter **des hallucinations et des idées délirantes** (45), la personne malade perçoit ou interprète mal des faits ou des paroles, ces hallucinations peuvent être visuelles ou auditives, des troubles affectifs comme **l'anxiété** où le malade exprimera un sentiment de peur, des préoccupations concernant son avenir, sa santé, son environnement, pouvant aller jusqu'à **la dépression**, c'est un état de découragement, il sera perdu et évoquera l'idée de mourir, le patient peut aussi présenter **une apathie** c'est-à-dire une perte d'intérêt pour des activités qu'il aimait faire jusque-là, il montrera une baisse de motivation à réaliser des projets (46).

I.3.8 Diagnostic

Dans cette section, nous nous pencherons sur les différentes méthodes de diagnostic qui sont utilisés de nos jours ; ces méthodes se basent sur des tests neuropsychologiques et sur des

techniques d'analyses biologiques et d'imageries utilisant des biomarqueurs connus dans la maladie d'Alzheimer.

- **Tests neuropsychologiques**

De nombreux bilans neuropsychologiques ont été mis au point afin d'évaluer les fonctions cognitives et mnésiques tels que le langage, la praxie, la gnose, les fonctions exécutives ou bien le raisonnement.

L'évaluation la plus répandue est appelé aussi Mini Mental State Examination (**MMSE**) (47), ce test est constitué 30 questions et est scoré sur 30 points, le degré de sévérité de la démence est considéré comme léger avec un score supérieur à 20, modéré entre 10 et 20, alors qu'en dessous de 10, la démence est sévère (48).

- **Biomarqueurs dans la maladie d'Alzheimer**

Un biomarqueur est un paramètre physiologique, biochimique ou anatomique qui peut être objectivement mesuré comme l'indicateur d'un processus biologique donné ; dans la maladie d'Alzheimer, les biomarqueurs peuvent être mis en évidence par des examens biologiques (49).

IRM (Imagerie à Résonance Magnétique) permet de détecter une atrophie corticale et notamment une atrophie des hippocampes (structure cérébrale impliquée dans la mémoire, dont la taille est souvent diminuée dans la maladie d'Alzheimer).

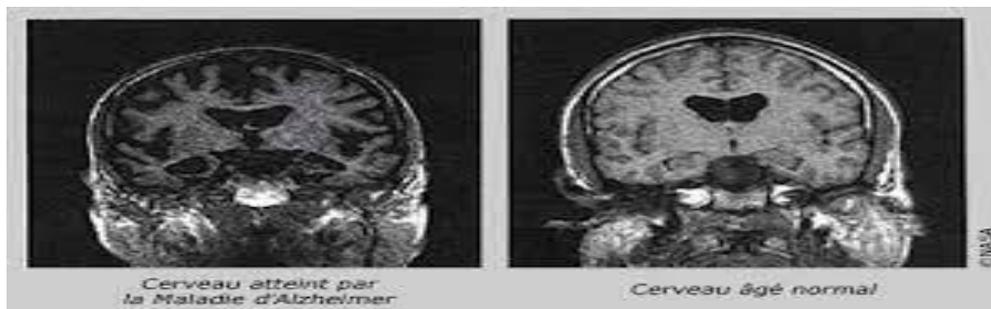


Figure 11 : Comparaison des images d'IRM cérébrale d'un sujet sain et d'un sujet atteint de MA

La **TEP** (La Tomographie par Emission de Positrons) permet d'avoir accès à une imagerie dite fonctionnelle ; une dernière technique peut être utilisée pour mesurer le métabolisme cérébral, elle permet de mettre en évidence un hypo-métabolisme (c'est-à-dire un fonctionnement moins efficace du cerveau) et de visualiser les lésions cérébrales caractéristiques de la maladie, notamment les plaques amyloïdes (50).

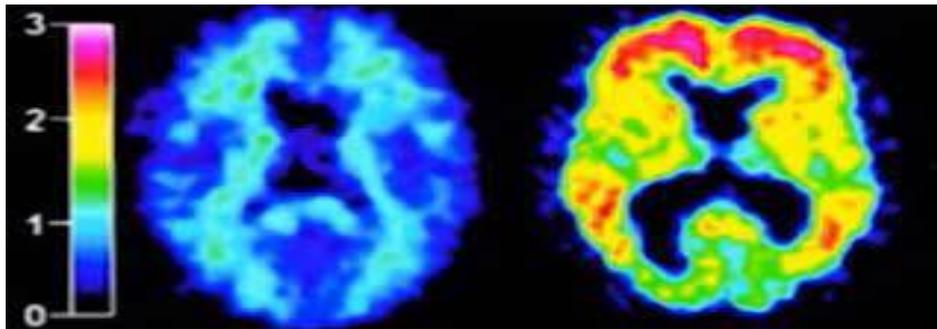


Figure 12 : Quantification de la présence de dépôts amyloïdes par TEP

I.3.8 Traitement

Il n'existe, à l'heure actuelle, **aucun traitement capable** de guérir la maladie d'Alzheimer ou de ralentir son évolution dans le cerveau ; néanmoins, quatre médicaments sont actuellement sur le marché :

- Le Donépézil (Aricept)
- La Rivastigmine (Exelon)
- La Galantamine (Reminyl)
- La Mémantine (Exiba)

Ces traitements ont pour but de traiter les symptômes cognitifs de la personne malade, ils n'empêchent cependant pas la propagation de la maladie dans le cerveau mais restent efficaces d'un point de vue purement clinique (51).

Bibliographie

- (1) Melik-Ogandzhanyan, R. G.; Khachatryan, V. E.; Gapoyan, A. S. *Russ. Chem. Rev.* 1985, 54, 262, and references cited therein.
- (2) Castle, R. N.; Philips, S. D. In *Katritzky, A., Rees, C., Eds.; Comprehensive Heterocyclic Chemistry; Boulton, A. J., McKillop, A., Eds.; Pergamon: Oxford, 1984; Vol. 3, p 329ff.*
- (3) Glazier, D. B.; Heanry, J. A.; Amden, R. J. *J. Urol.*, 1966, 155, 624.
- (4) *DÉRIVÉS DE PYRIMIDINE DESTINÉS AU TRAITEMENT DE MALADIES ASSOCIÉES À L'AMYLOÏDE* [En ligne] 2019.
<https://patentscope.wipo.int/search/fr/detail.jsf?docId=WO2007125351>
- (5) Bakchine, S.; Habert, M.O. *Classification des démences : aspects nosologiques*. *Médecine Nucléaire*, 31 (6), (2007), 278-293.
- (6) Prince, M.J., 2015. *World Alzheimer Report 2015: The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends*. *Alzheimer's Disease International*
- (7) Watson, J. D.; Crick, F. H. C. "Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid". *Nature*, vol. 171, pp: 737-138, 1953.
- (8) THORE, S.N. ET AL., *ASIAN J. CHEM.*, 19, 4429 (2007).
- (9) KENNER, G.W. AND TODD, A., IN: *HETROCYCLIC COMPOUNDS*, CHAPMAN AND HALL, NEW YORK, 242 (1957).
- (10) HITCHINGS, G.H., *ANN. N.Y. ACAD. SCI.*, 23, 700 (1961).
- (11) VISHWANADHAN, C.L., JOSHI, A.A. AND SACHIN, S.N., *BIOORG. MED. CHEM. LETT.*, 15, 73 (2005).
- (12) BABU, V., HARINADHA KUMAR, P., SENTHIL SRINIVASAN, K.K. AND BHAT, G. V., *INDIAN J. PHARM. SCI.*, 66, 647 (2004).
- (13) CHERN, J.H. ET AL., *BIOORG. MED. CHEM. LETT.*, 14, 2519 (2004).
- (14) Alzheimer, A. *Über Einen Eigenartigen, Schweren Erkrankungsprozeß Der Hirnrinde*. *Neurol. Cent.* 1906, 23, 1129–1136.
- (15) Alzheimer, A. *Concerning unusual medical cases in old age*. *Z. Gesamte Neurol. Psychiatr.* 1911, 4, 356–385.
- (16) Jervis, G. A.; Soltz, S. E. *Alzheimer's Disease-The So-Called Juvenile Type (With Report of a Case)*. *Am. J. Psychiatry* 1936, 93, 39–56.
- (17) *Définition et chiffres de la maladie d'alzheimer* [En ligne] 2020. [Citation : 16 6 2021.]
<https://alzheimer-recherche.org/la-maladie-alzheimer>

Bibliographie

- (18) La maladie d'alzheimer ministère des solidarités et de santé [En ligne] 2020] <https://solidarites-sante.gouv.fr/soins-et-maladies/maladies>.
- (19) McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr. CR, Kawas CH, et al, 7, 3, 263-269, 2011.
- (20) Qu'est-ce que la maladie d'Alzheimer ? Alzheimer's Association, 2019. Disponible sur: <https://www.alz.org/fr/quest-ce-que-la-maladie-d-alzheimer.asp>.
- (21) Maladie d'Alzheimer : les chiffres de la pathologie. Fondation Recherche Médicale Disponible sur: <https://www.frm.org/recherches-maladies-neurologiques/maladie-d-alzheimer/alzheimer-en-chiffres>. (Consulté le:12/06/2021)
- (22) Résumé du Rapport Mondial Alzheimer 2015 de l'ADI (Alzheimer's Disease International), [En ligne] 2019. Disponible sur : <https://www.alz.co.uk/sites/default/files/pdfs/world-alzheimer-report-2015-summary-sheet-french.pdf>
- (23) Fratiglioni L, Launer LJ, Andersen K, et al. Incidence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. *Neurology* 2000 ; 54 (suppl 5) : S10-5.
- (24) Edland SD, Rocca W, Petersen RC, Cha RH, Kokmen E. The incidence of Alzheimer's disease does not vary by gender in Rochester, MN. *Neurobiol Aging* 2000; 21: S203.
- (25) <https://alzheimer-recherche.org/la-maladie-alzheimer/quest-maladie-dalzheimer/facteurs-de-risques/>
- (26) Cotman C.W.,Pike C.J. (1994) β -Amyloïd and its contributions to neurodegeneration in Alzheimer disease in *Alzheimer Disease*,17:305-315
- (27) Fan L.Y, Chiu M.J. (2010) Pharmacological Treatment for Alzheimer's disease: Current Approaches and Future Strategies. *Acta Neurol Taiwan*, 19: 228-245
- (28) Selmès J. - La maladie d'Alzheimer : Accompagnez votre proche au quotidien - John Libbey Eurotext, 2011.
- (29) Avet S. - La prise en soins d'un patient atteint de la maladie d'Alzheimer, au travers de situations spécifiques - NPG Neurologie, Psychiatrie, Gériatrie, 2008.
- (30) Phaneuf, M. & Bal-Craquin, M. T. (2007). *La maladie d'Alzheimer et la prise en charge infirmière*. Paris : Elsevier Masson.
- (31) Folstein, M.F., Folstein, S.E. & McHugh, P.R. (1975). Mini mental state: a practical method for grading the cognitive state of patients for clinician. *Journal of Psychiatry Research*, 12, 189-198.
- (32) Alzheimer : une maladie neurodégénérative complexe. Inserm'', 2019. Disponible sur: <https://www.inserm.fr/information-en-sante/dossiersinformation/alzheimer-maladie>.

Bibliographie

- (33) Prince, M. World Alzheimer Report 2014 Dementia And Risk Reduction An Analysis Of Protective And Modifiable Factors, 2014.
- (34) Société d'Alzheimer Canada. La maladie d'Alzheimer et les facteurs de risque. [En ligne]. <http://www.alzheimer.ca/fr>. Consulté le : [08/04/2021].
- (35) Haan, M. N. (2005). La prévention primaire de la démence en population : recherche de facteurs de risque modifiables, 141-145.
- (36) Catherine Helmer, Florence Pasquier, Jean-François Dartigues. Epidémiologie de la maladie d'Alzheimer et des syndromes apparentés. *MEDECINE/SCIENCES*. 2006, Vol. 22, 3.
- (37) Scarmeas, N., Albert, S. M., Manly, J. J. & Stern, Y. (2006). Education and rates of cognitive decline in incident Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 77, 308-316.
- (38) Mediterranean diet and dementia. Alzheimer's Society ' ', 2019, Disponible sur: <https://www.alzheimers.org.uk/about-dementia/risk-factors-and-prevention/mediterranean-diet-and-dementia>.
- (39) Institut national de la santé et de la recherche médicale. Alzheimer. *Inserm* . [En ligne] Juillet 2021. <http://www.inserm.fr/thematiques/neurosciences-sciences-cognitives-neurologie-psychiatrie/dossiers-d-information/alzheimer>.
- (40) Catherine Helmer, Florence Pasquier, Jean-François Dartigues. Epidémiologie de la maladie d'Alzheimer et des syndromes apparentés. *MEDECINE/SCIENCES*. 2006, Vol. 22, 3.
- (41) Caffeine and dementia. *Alzheimer's Society* (2019). Disponible sur: <https://www.alzheimers.org.uk/about-dementia/risk-factors-and-prevention/caffeine-and-dementia>. (Consulté le: 25juin 2021)
- (42) Aguesse C. Alzheimer et autres démences : comportements et pathologies, prise en charge et accompagnement, activités et thérapies. Doc Editions, 2008 ; 208.
- (43) France Alzheimer. Les premiers signes d'alerte. [En ligne] 22 octobre 2020. <http://www.francealzheimer.org/comprendre-maladie/sympt%C3%B4mes-et-diagnostic>.
- (44) Davou P., Delacourt A. Maladie d'Alzheimer Encyclopédie médico-chirurgicale 1999.
- (45) Lebert F, Pasquier F. EMC, Neurologie. *Symptômes comportementaux et psychologiques lors de démence*. Paris : Elsevier Masson SAS, 2008. Vol. 17-057-A-35.
- (46) France Alzheimer. Les symptômes comportementaux. *France Alzheimer et maladies apparentées*. [En ligne] 2012. <http://www.francealzheimer.org/les-sympt%C3%B4mes/les-sympt%C3%B4mes-comportementaux/181>.
- (47) Folstein, M. F.; Folstein, S. E.; McHugh, P. R. "Mini-Mental State": A Practical Method for Grading the Cognitive State of Patients for the Clinician. *J. Psychiatr. Res.* 1975,

12, 189–198.

(48) Le diagnostic de la maladie d'Alzheimer. [En ligne] 23 octobre 2020. <http://acces.ens-lyon.fr/acces/thematiques/neurosciences/actualisation-des-connaissances/maladies-et-traitements/alzheimer/la-maladie-dalzheimer-a-lechelle-de-lindividu/le-diagnostic-de-la-maladie-dalzheimer>.

(49) Hampel, H.; Lista, S.; Khachaturian, Z. S. Development of Biomarkers to Chart All Alzheimer's Disease Stages: The Royal Road to Cutting the Therapeutic Gordian Knot. *Alzheimers Dement.* 2012, 8, 312–336.

(50) DIAGNOSTIC DE LA MALADIE D'ALZHEIMER, Fondation recherche Alzheimer [En ligne] 16 mars 2021. <https://alzheimer-recherche.org/la-maladie-alzheimer/symptomes-et-diagnostic/diagnostic>

(51) Quatre traitements médicamenteux symptomatiques [En ligne] 2020. Disponible sur : <https://www.vaincrealzheimer.org/lamaladie/traitements>

CHAPITRE II

**LES MÉTHODES UTILISÉES DANS LA
MODÉLISATION MOLÉCULAIRE.**

II.1 Introduction

La modélisation moléculaire est un outil destiné aux chercheurs préoccupés par la structure et la réactivité des molécules. La connaissance de la structure des édifices moléculaire permet de comprendre ce qui est réalisé dans une transformation physique, chimique ou biologique (1). La modélisation moléculaire implique l'utilisation de méthodes de calculs théoriques avec différentes approches sont envisageables (mécanique moléculaire, semi-empirique, ou méthodes quantiques (*ab initio*, DFT) ...) permettant de déterminer la représentation graphique de la géométrie et d'évaluer les propriétés physicochimiques de la molécule étudiée. La modélisation moléculaire permet d'interpréter des phénomènes physico-chimiques, de suggérer des nouvelles expériences et d'analyser ainsi des résultats d'une façon plus critique que les expériences classiquement utilisées, elle est donc complémentaire à l'expérience (2).

Le développement de nouvelles techniques de modélisation a permis la mise en place de nombreuses méthodes QSPR/QSAR (Quantitative Structure Property/activity Relationship); elles reposent pour la plupart sur la recherche d'une relation entre un ensemble de nombres réels, appelés descripteurs moléculaires, et la propriété ou l'activité que l'on souhaite prédire. Ces méthodes permettent de justifier les données expérimentales disponibles et de prédire les propriétés/activités pour des nouveaux composés ou des composés pour lesquels les données expérimentales ne sont pas disponibles (3).

II.2 Méthodes quantiques

Les méthodes de la mécanique quantique, qui font appel à la distribution des électrons répartis en orbitales autour de la molécule, impliquent des temps de calcul souvent élevés qui limitent leur usage à des petites molécules ou nécessitent le recours à de nombreuses approximations. Elles sont particulièrement adaptées au calcul des charges et des potentiels électrostatiques, à l'approche des mécanismes réactionnels ou à la polarisabilité. L'objectif de la mécanique quantique est principalement de déterminer l'énergie et la distribution électronique (4).

La chimie quantique définit la structure moléculaire comme un noyau autour duquel gravitent des électrons, qui sont décrits par leur probabilité de présence en un point et représentés par des orbitales (5).

La chimie quantique applique les principes de la mécanique quantique aux systèmes moléculaires pour tenter de résoudre l'équation de Schrödinger (6). En effet, le comportement

électronique et nucléaire des molécules, étant responsable des propriétés chimiques, peut être décrit de façon réaliste à partir de cette équation. Différentes méthodes de résolution ont alors été développées. En particulier, le développement grandissant des moyens informatiques ont permis le développement de ces méthodes (7).

II.2.2 Equation de Schrödinger

La chimie quantique définit la structure moléculaire comme un noyau autour duquel gravitent des électrons, qui sont décrits par leur probabilité de présence en un point et représentés par des orbitales (8), les équations de la chimie quantique sont basées sur la résolution de l'équation de SCHRÖDINGER qui s'écrit pour les états stationnaires (9).

$$H\Psi = E\Psi \quad (1)$$

Où :

Ψ : fonction d'onde molécule.

H: Hamiltonien totale d'une molécule comportant N noyaux et n électrons, il est défini par la somme de cinq termes (cinétique des électrons, cinétique des noyaux, répulsion entre électrons, répulsion entre noyaux et le terme d'attraction électrons –noyaux)

E: l'énergie totale d'une molécule.

La résolution exacte de cette équation n'est possible que pour l'atome d'hydrogène et les systèmes mono-électronique ; pour les systèmes poly électroniques, on fait appel aux méthodes d'approximation, pour la résolution approchée de l'équation de SCHRÖDINGER.

L'Hamiltonien rend compte des différentes contributions à l'énergie totale du système à partir d'opérateurs pour les énergies cinétiques des électrons et des noyaux ainsi que les interactions noyau-électron, électron-électron et noyau-noyau.

$$H = -\frac{\hbar^2}{2m_e} \sum_i^n \Delta_i - \frac{\hbar^2}{2M_k} \sum_k^N \Delta_k + \sum_{i>j}^n \frac{e^2}{r_{ij}} + \sum_{k>l}^N \frac{Z_k Z_l e^2}{r_{kl}} - \sum_{k=1}^N \sum_{i=1}^n \frac{Z_k e^2}{R_{ki}} \quad (2)$$

II.2.3 L'approximation de Born-Oppenheimer

Une première simplification du problème est l'approximation de Born-Oppenheimer ; cette approximation consiste à considérer que les deux types de particules, électrons et noyaux ont des vitesses d'évolution caractéristiques très différentes ; pour être précis, les électrons se déplacent beaucoup plus rapidement que les noyaux (10). On peut alors considérer le comportement des électrons dans une molécule en supposant que les noyaux occupent les positions fixes dans l'espace, on aboutit alors à l'équation de Schrödinger électronique (11) :

$$\hat{H}_e \psi_e = E_e \psi_e \quad (3)$$

Où \hat{H}_e est l'Hamiltonien électronique :

$$\hat{H}_e = \mathbf{T}_e + \mathbf{V}_{e-e} + \mathbf{V}_{e-N} \quad (4)$$

La variable scalaire E_e est l'énergie propre, associée à l'Hamiltonien \hat{H}_e (12).

II.2.4 La théorie de la fonctionnelle de la densité (DFT)

Les bases de la théorie de la fonctionnelle de la densité (DFT) ont été élaborées en 1927 par Thomas et Fermi qui calculèrent l'énergie d'un atome en représentant son énergie cinétique en fonction de la densité électronique (13). En 1928, Dirac introduit le terme d'échange prédit par Hartree mais il n'y a toujours aucune prise en compte de la corrélation électronique qui fût finalement ajoutée par Wigner ; dans ce modèle, les n électrons dépendants de $3n$ coordonnées d'espace sont remplacés par leur densité $\rho(r)$ qui ne dépend plus que de 3 variables.

L'état fondamental de ce système est décrit par la fonction d'onde $\Psi_0(r_1, r_2, \dots, r_n)$ qui correspond à une unique densité électronique $\rho(r)$; cette fonction d'onde, et l'énergie E_0 qui lui est associée, sont déterminées par la minimisation de l'énergie totale du système, le potentiel externe $V_{\text{ext}}(r)$ créé par les N noyaux du système est alors complètement déterminé et fixe donc l'hamiltonien. Ainsi, le nombre d'électron n et le potentiel $V_{\text{ext}}(r)$, définissent toutes les propriétés de l'état fondamental (14).

La DFT est capable de déterminer avec précision les propriétés moléculaires géométriques, les énergies de liaisons et différents types de spectre pour des molécules aussi complexes que des composés de coordination (15). De plus, les surfaces d'énergie potentielle des réactions chimiques prédites par les fonctionnelles corrigées par le gradient et/ou les fonctionnelles hybrides sont aussi précises que celles prédites par les niveaux de théories ab initio les plus sophistiqués (16).

II.2.5 Les bases d'orbitales

Le choix de la base d'orbitales est aussi un critère très important pour une bonne estimation des données, bien qu'il soit moins important pour la DFT que pour les autres méthodes ab initio, en particulier post Hartree Fock, la base représente la description des orbitales atomiques.

- **La base 6-31 G*** : décrit une orbitale atomique par la combinaison de six fonctions gaussiennes par orbitale de cœur, de trois autres pour la description des électrons de valence

et d'une dernière pour description des électrons de valence les plus éloignés du noyau (externe) ; l'astérisque signifie l'utilisation d'orbitales de polarisation pour les atomes lourds c'est à dire d pour les orbitales p, f pour les orbitales d, etc.

- **La base 3-21G** : base telles que chaque orbitale atomique interne est représentée par une gaussienne contractée qui est elle-même une combinaison linéaire de 3 gaussiennes primitives ; alors que chaque orbitale de valence est représentée par deux fonctions gaussiennes contractées l'une est combinaison linéaire de 2 primitives et l'autre 1 gaussienne simple.

- **La base 6-311G** : représente une base du type triple zêta, dans cette base l'orbitale interne est représentée par une seule STO combinaison linéaire de 6 GTO et les orbitales de valences sont représentées par 3 STO (17).

II.3 QSPR/QSAR

II.3.1 Définition

Les méthodes QSAR sont basées sur l'hypothèse que l'activité d'un composé chimique est liée à sa structure. Plus précisément cette approche affirme que l'activité biologique et la structure d'un composé chimique sont liées par une formule mathématique (18).

II.3.2 Principe

Le principe des méthodes QSAR/QSPR est de mettre en œuvre une relation mathématique reliant de manière quantitative des descripteurs moléculaires (constitutionnels, topologiques, géométriques, électrostatiques, quantiques, thermodynamiques, spectroscopiques, . . .) avec une observable macroscopique (propriété physicochimique ou activité biologique) pour une série de composés chimiques similaires à l'aide de méthodes statistiques d'analyse de données.

La forme générale d'un modèle QSPR/QSAR est :

$$(\text{Propriété, Activité}) = f(\text{descripteurs})$$

L'objectif de ces méthodes est donc d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété ou l'activité mesurée (19).

II.3.3 Méthodologie générale d'une étude QSAR/QSPR

II.3.3.1 La base de données

Le modèle QSAR/QSPR s'appuie sur des données expérimentales de référence, le choix de la

base de données est la clé pour développer ce modèle ; dans la plupart des cas, les données expérimentales proviennent de la littérature, et pour assurer la qualité, la base de données doit être composée de données ,les tests expérimentaux sont aussi fiables que possible car des barres d'erreur y apparaissent sera propagé dans le modèle final car les paramètres de ce dernier sont des données ont été ajustées. Par conséquent, il est important de choisir la méthode de présentation des données faible incertitude pour limiter les barres d'erreur expérimentales ; de plus, les données doivent être obtenues selon un protocole expérimental unique. En effet, les conditions expérimentales ont une grande influence sur les valeurs obtenues, l'activité définie selon les conditions expérimentales est aussi une étape important du processus (20).

II.3.3.2 Descripteurs moléculaires

II.3.3.2.1 Définition

Un descripteur moléculaire est un paramètre ayant une valeur numérique propre pour une structure chimique donnée. Ils peuvent être obtenus de manière empirique ou non-empirique. Les descripteurs calculés, et non mesurés, sont à privilégier : ils permettent en effet d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation moléculaire.

Depuis des années, plusieurs travaux ont été menés afin de développer des descripteurs capables de décrire de la manière la plus exhaustive possible les structures moléculaires (21). Historiquement, deux grands schémas pour la classification des descripteurs moléculaires ont été établis : l'un en fonction de leur origine (constitutionnel, topologique, géométrique, quantique, thermodynamique...), et un autre sur leur dimensionnalité (1D, 2D, 3D) (22).

II.3.3.2.2 Types de descripteurs

a. Les descripteurs 1D

Les descripteurs 1D sont appelés «descripteurs constitutionnels» et sont faciles et rapides à calculer, ils sont calculés à partir de la formule brute de la molécule ou à partir de la composition moléculaire, c'est-à-dire les atomes qui la constituent en décrivant des propriétés globales du composé (23).

Des exemples de descripteurs constitutionnels sont :

- La masse molaire
- Les nombres absolus et relatifs d'atomes (C, H, O, S, N, F, Cl, Br, I, P. . .).
- Les nombres absolus et relatifs de groupes fonctionnels (NH₂, COOH, OH. . .).
- Les nombres absolus et relatifs de liaisons (simples, doubles, aromatiques. . .).
- Les nombres absolus et relatifs de cycles (aromatiques ou non) (24).

b. Les descripteurs 2D

Les descripteurs 2D sont des descripteurs qui utilisent la représentation des molécules comme des graphes, ils contiennent des informations à propos de la connectivité ou à propos de certains fragments moléculaires, mais aussi des estimations des propriétés physicochimiques (25).

➤ Descripteurs topologiques :

Les descripteurs topologiques sont des indices obtenus à partir d'une structure 2D de la molécule, à savoir une simple table de connectivité des atomes dans la molécule. Ils contiennent en leur sein des informations sur la taille globale du système, sa forme globale et ses ramifications (26).

Ces descripteurs s'inspirent de la théorie des graphes appliquée à la table de connectivité qui n'est autre qu'une représentation compacte de la connectivité interatomique au sein de la molécule ; les indices topologiques les plus fréquemment utilisés sont l'indice de Wiener (24), l'indice de Randic (27), l'indice de connectivité de valence de Kier-Hall (28) et l'indice de Balaban (29).

Descripteur	Définition
	$\omega = \frac{1}{2} \sum_{ij} d_{ij} = \frac{1}{2} \sum_i s_i$
L'indice de Wiener (ω)	Est simplement la demi-somme de tous les éléments d'une telle matrice où S_i est la somme des distances
	$J = C \cdot \sum_k (s_i \cdot s_j)^{-0.5}$
L'index J de Balaban (J)	Où $C = m(1 + u)$; m est le nombre d'arêtes et u est le nombre cyclomatique.

Tableau 1 : Les descripteurs topologiques calculés dans étude QSAR

c. Les descripteurs 3D

Ils décrivent des caractéristiques complexes ; leurs calculs nécessitent donc de connaître la géométrie 3D de la molécule. Ces descripteurs s'avèrent donc relativement coûteux en temps de calcul, mais apportent davantage d'informations, et sont souvent nécessaires à la modélisation d'activités biologiques (30).

On distingue plusieurs familles importantes de descripteurs :

➤ Descripteurs géométriques

Il s'agit de descripteurs évalués à partir des positions relatives des atomes d'une molécule dans l'espace, ainsi que des rayons et masses atomiques. Ils nécessitent donc de connaître la structure 3D de la molécule. Celle-ci peut être obtenue expérimentalement bien entendu mais le plus souvent par modélisation moléculaire, empirique ou *ab initio*. Par conséquent, contrairement aux descripteurs précédents, ils nécessitent un certain temps de calcul.

Parmi ces descripteurs, on retrouve le volume et la surface moléculaire, le moment d'inertie ou encore des distances, angles ou angles dièdres particuliers entre atomes dans la molécule.

Un tel paramètre géométrique a pu notamment montrer son rôle central pour caractériser l'activité des complexes bis (arylimino) pyridine du fer (II) pour l'oligomérisation de l'éthylène (31).

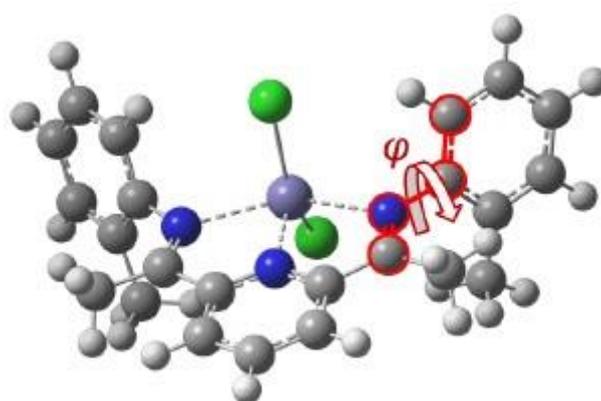


Figure 9: Angle dièdre dirigeant l'activité catalytique des complexes

➤ Descripteurs quantiques

Les descripteurs de chimie quantique donnent des informations importantes pour la molécule. Ces descripteurs permettent de quantifier différents types d'interactions inter et intramoléculaires, de grande influence sur des propriétés physicochimiques de molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie est minimale, et fait appel aux approches de chimie quantique. On cite toutes les données énergétiques, vibrationnelles et orbitales du système par exemple, les énergies de la plus haute orbitale moléculaire occupée HOMO et de la plus basse vacante LUMO, le moment dipolaire, la polarisabilité, le potentiel d'ionisation. . . (32).

➤ Descripteurs physico-chimiques

Les descripteurs physico-chimiques, certains d'entre eux reflètent la composition moléculaire du composé, d'autres représentent le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage Octanol/eau représenté par le **log P**. Parmi ces descripteurs, le log P, la réfractivité moléculaire, l'indice de réfraction, la

polarisabilité, la densité, la surface de tension, le nombre de donneurs de liaisons hydrogène, le nombre d'accepteurs de liaisons hydrogène... (33).

Le logP est une mesure de la solubilité différentielle de composés chimiques dans deux solvants (coefficient de partage Octanol/Eau).

Le **logP** est égal au logarithme du rapport des concentrations de la substance étudiée dans l'Octanol et dans l'eau.

$$\log P = \log \frac{[\text{Octanol}]}{[\text{H}_2\text{O}]} \quad (4)$$

Le logP rend compte de la tendance d'une molécule à s'accumuler dans les membranes biologiques des organismes vivants. S'il est élevé, le risque de bioaccumulation est important. Le partage d'une molécule entre une phase aqueuse et une phase lipidique conditionne en partie ses propriétés biologiques telles que l'affinité pour un récepteur et la fixation par une protéine, ou encore la toxicité (34).

➤ Descripteurs thermodynamiques

Ce sont des descripteurs peu utilisés dans les études RQSA/RQSP, ils peuvent être exprimés par la fonction de partition **Q** de la molécule utilisée en thermodynamique statistique ainsi que de ses dérivées (35) ; cette fonction décrit la façon avec laquelle l'énergie d'un système de molécules est répartie parmi les individus moléculaires ; sa valeur dépend du poids moléculaire, de la température, du volume moléculaire, des distances inter nucléaires, des mouvements moléculaires et des forces intermoléculaires, la fonction de partition est le point le plus commode entre les propriétés microscopiques des molécules indépendantes (niveaux d'énergie, moments d'inertie) avec les propriétés macroscopiques (point de fusion, point d'ébullition, entropie) ; l'expression de cette fonction s'écrit :

$$Q = Q_{trans} * Q_{\acute{e}le} * Q_{rot} * Q_{vibr} \quad (5)$$

Q_{trans} : Fonction de partition de translation.

$Q_{\acute{e}le}$: Fonction de partition électronique.

Q_{rot} : Fonction de partition de rotation.

Q_{vibr} : Fonction de partition de vibration (36).

d. Les descripteurs 4-D

Ils correspondent à la mesure des propriétés 3D (potentiel électrostatique, d'hydrophobicité, de liaison hydrogène...) d'une molécule en tout point de l'espace ; ils permettent d'avoir

l'information sur la structure de la cible (protéine), on pourra ainsi distinguer les descripteurs 4D qui nécessitent un alignement de la molécule guidé par l'étude des complexes ligand-cible (ou, au moins, par des contraintes visant d'optimiser le recouvrement spatial des champs électriques et stériques des ligands, faute d'information sur le vrai mode de fixation dans la cible) avant d'être calculés ; ces descripteurs sont obtenus par le calcul des champs d'interactions moléculaires (CoMFA, CoMSIA) entre une molécule et une sonde représentée par une autre molécule (eau, amide, ...) (37,38).

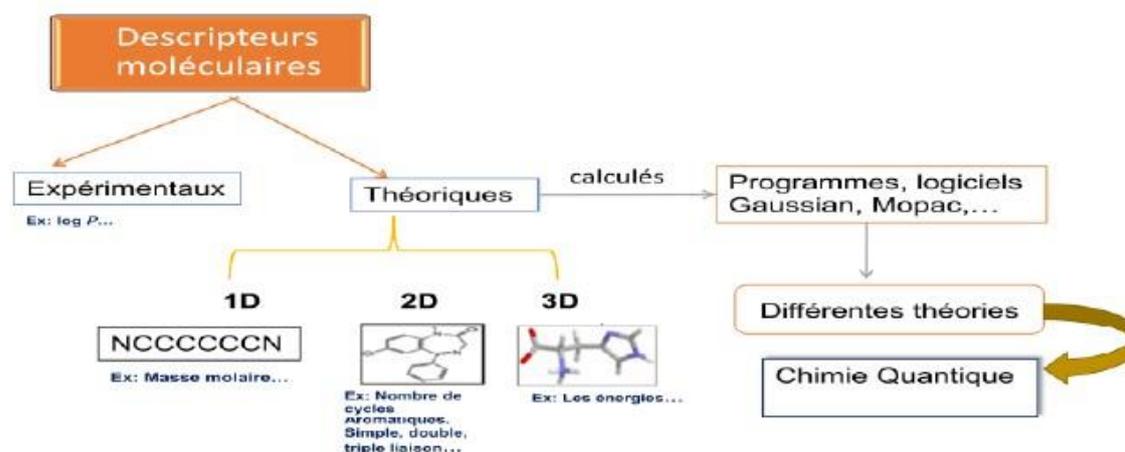


Figure 10 : Les descripteurs moléculaires

II.3.3.3 Logiciels de calcul des descripteurs moléculaires

Plusieurs logiciels sont disponibles pour faire les calculs des descripteurs, parmi ceux que nous avons utilisé dans nos travaux, on trouve : Gaussian (39), ChemOffice (40), ChemSketch (41), Marvin Sketch (42), Dragon(43), mais il existe plusieurs autres logiciels.

II.3.3.4 Sélection des descripteurs

La recherche de l'ensemble des descripteurs qui forment le bon modèle, exprimant l'activité biologique avec un coût raisonnable de calcul, constitue l'étape déterminante, car le calcul de tous les modèles possibles n'est pas pratique eu égard au nombre élevé des descripteurs (plus de 1600 descripteurs).

Afin d'éviter la formation de modèles dus à la chance un contrôle rigoureux est exigé sur la taille de l'ensemble des descripteurs. Ainsi, l'approche rationnelle de la sélection des variables permet d'éviter les redondances, de diminuer le coût calculatoire et de trouver les meilleurs sous-ensembles de descripteurs, la procédure de sélection des variables peut être divisée en deux étapes (44):

II.3.3.4.1 La sélection objective :

Elle consiste à la sélection des variables en réduisant le nombre de descripteurs sans faire participer la variable dépendante (la réponse biologique) afin de diminuer les corrélations entre les descripteurs.

La suppression d'un descripteur, ayant un pourcentage élevé de valeurs identiques pour l'ensemble des composés, aura lieu au début de la sélection ; il en est de même pour les descripteurs fournissant des informations superflues. Ensuite, le coefficient de corrélation (R) entre les descripteurs est calculé par paires, un des deux descripteurs est supprimé si leur combinaison possède un coefficient de détermination supérieur au seuil requis ($R > 0.90$); cette valeur numérique de R est le seuil utilisé pour toutes les applications de sélection en réduisant le nombre de descripteurs sans pour autant perdre de l'information (45).

II.3.3.4.2 La sélection subjective :

➤ Introduction progressive :

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

➤ Elimination progressive :

Cette méthode consiste en l'établissement du modèle avec l'ensemble des descripteurs pour ensuite ne garder que ceux qui permettent l'obtention d'un modèle ayant une bonne corrélation (46).

➤ Sélection pas à pas :

C'est la combinaison des deux méthodes citées précédemment, les variables sont incorporées une à une dans le modèle par sélection progressive. Cependant, à chaque étape, on vérifie que les corrélations partielles des variables précédemment introduites sont encore significatives (47).

II.3.4 Méthodes statistiques

Toutes les caractéristiques d'une molécule ne sont pas liées à son activité biologique, les descripteurs chimiques sont souvent corrélés entre eux et donnent la même information.

L'objectif de l'analyse statistique est justement de « démêler » ces descripteurs et d'identifier ceux qui sont corrélés à la variable cible, qui produisent du signal, de ceux qui ne le sont pas, qui produisent du bruit ; l'analyse statistique permet également d'identifier les descripteurs qui sont corrélés entre eux pour ne garder que les principaux et réduire ainsi la redondance d'informations (48).

Dans la terminologie des statistiques, la propriété étudiée est appelée variable cible ou

variable dépendante ou encore variable Y ; c'est la caractéristique que l'on cherche à expliquer et à optimiser.

Les descripteurs, en l'occurrence chimiques, sont appelés variables explicatives, variables indépendantes ou encore variables X.

L'analyse statistique détermine et quantifie les corrélations entre les descripteurs et la variable cible, elle indique également la contribution relative de chaque descripteur dans l'explication globale de l'activité, le modèle statistique est une équation donnant la valeur de la variable cible en fonction de la somme des valeurs pondérées des descripteurs ; les principaux outils statistiques pour obtenir un modèle sont :

- Régression linéaire simple (*Simple Linear Regression-SLR*) ;
- Régression linéaire multivariée (*Multivariate Linear Regression- MLR*) ;
- Régression non linéaire multiple (MNLR) ;
- Régression en composante principale (*Principale Component Regression- PCR*) ;
- Régression des moindres carrés partiels (*Partial Least Squares - PLS*);
- Réseaux de neurones artificiels (*Artificial Neural Network - ANN*) ;
- Analyse en composantes principales (ACP) (49).

II.3.4.1 La régression linéaire multiple (MLR)

La régression linéaire multiple **MLR** est l'une des méthodes de modélisation les plus populaires grâce à sa simplicité d'utilisation et facilité d'interprétation ; l'avantage important de la régression linéaire multiple est qu'elle est très transparente, puisque l'algorithme est disponible, et que les prédictions peuvent être réalisées facilement. Dans la plupart de nos travaux, cette méthode a été utilisée aussi pour la sélection des descripteurs moléculaires utilisés dans les autres méthodes statistiques.

La méthode MLR se base sur l'hypothèse que la propriété y dépend linéairement des différentes variables (les descripteurs) $x_1, x_2 \dots x_i$, selon la relation :

$$y = a_0 + \sum_{i=1}^n a_i \cdot x_i \quad (6)$$

Avec : y est la variable dépendante (à expliquer ou à prédire) ; x_i sont les variables indépendantes (explicatives) ; n est le nombre de variables explicatives ; a_0 est la constante de l'équation du modèle ; a_i sont les coefficients de descripteurs dans l'équation du modèle (50).

II.3.4.2 La régression non linéaire multiple (MNLR):

La régression non linéaire multiple MNLR est une méthode non linéaire (exponentielle, logarithmique, polynomiale, ...) qui permet de déterminer le modèle mathématique qui

permet d'expliquer non-linéairement au mieux la variabilité d'une propriété ou d'une activité y en fonction des descripteurs moléculaires. Dans l'ensemble de nos travaux nous avons utilisé le modèle polynomial en nous basant sur les descripteurs proposés par le modèle linéaire qui seront élevés à la puissance 2 selon l'équation suivante:

$$y = a_0 + \sum_{i=1}^n a_i \cdot x_i + b_i \cdot x_i^2 \quad (7)$$

Avec : y est la variable dépendante (à expliquer ou à prédire) ; x_i sont les variables indépendantes (explicatives) ; i est le nombre de variables explicatives ; a_0 est la constante de l'équation du modèle ; a_i , b_i sont les coefficients de descripteurs dans l'équation du modèle (51).

II.3.4.3 Régressions PLS et ACP

Les régressions PLS et ACP sont des méthodes statistiques qui permettent de trouver par une transformation linéaire, les axes qui représentent au mieux les données dans l'espace. En d'autres termes, ces méthodes vont permettre de trouver les axes qui expliquent au mieux la dispersion du nuage de points ; si les données sont représentées en fonction de n descripteurs, PLS et ACP vont donc permettre de trouver au maximum n axes classés en fonction de la variance qu'ils représentent, ces méthodes consistent à remplacer une matrice des données prédictives X comprenant n lignes et m colonnes, par une nouvelle matrice, dérivée de X , qu'on design par T , comprenant le même nombre de lignes (observations) que X , mais un nombre de colonnes k très inférieur à m . On impose, de plus, que les colonnes de la matrice T soient des combinaisons linéaires des variables d'origine, sous forme matricielle, la relation s'écrit :

$$T = X.W \quad (8)$$

Avec : W ($m \times k$) est la matrice des coefficients définissant les combinaisons linéaires ; T est la nouvelle matrice dont les colonnes forment des « variables artificielles », obtenues par combinaison linéaire des variables d'origine ;

Après cette transformation, la régression linéaire multiple est appliquée sur le tableau T à la place de X (52).

II.3.4.4 Régression par réseaux de neurones

Le réseau de neurones artificiels (ANN) est l'une des méthodes pronostiques bien connues utilisées pour trouver une solution lorsque d'autres méthodes statistiques ne sont pas applicables ; de plus, ANN s'intègre avec précision dans les variables non linéaires, ce qui est un avantage par rapport à l'analyse linéaire multivariée basée sur des variables linéaires, une

inspiration pour ANN était le cerveau humain et les neurones biologiques, l'élément de base de cette structure est le perceptron, il s'agit d'un équivalent mathématique d'un neurone, qui transfère des signaux électriques représentés sous forme de valeurs numériques ; les neurones artificiels sont organisés en couches : entrée - prenant les données d'entrée, cachée et sortie - produisant un résultat, chaque nœud se connecte à chaque neurone de la couche suivante. Cependant, il n'y a pas de connexions entre les neurones d'une même couche ; le processus d'apprentissage ANN est basé sur l'ajustement des connexions pondérées entre les nœuds jusqu'à ce que la solution la plus efficace d'un problème ait été obtenue. De plus, fournir à la fois une entrée et une sortie dans le réseau permet de calculer une erreur en fonction de sa sortie cible et de sa sortie actuelle, cela peut être utilisé pour des corrections du réseau en mettant à jour ses poids (coefficients) et pour obtenir des résultats optimaux (53).

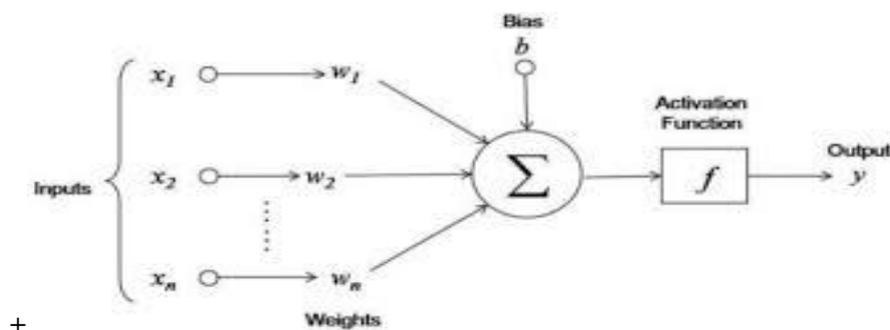


Figure 11 : Regression by Artificial Neural Network.

II.3.4 Test de la signification globale de la régression

a. Coefficient de détermination (R^2)

Le coefficient de détermination R^2 est la mesure de degré de liaison entre Y_N et X_j .

R^2 est déterminé par différentes relations suivantes :

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \quad (9)$$

Un bon ajustement correspondra à un R^2 proche de l'unité (54).

b. Coefficient de corrélation (R)

C'est le coefficient de corrélation de Bravais Pearson entre \hat{Y}_i et Y , c'est-à-dire entre valeurs observées et prédites par le modèle de régression, il est noté R , sa valeur variée entre 0 et 1(55).

c. Test Fischer-Snedecor (F)

Le test Fischer permet de justifier la liaison globale entre Y_n et X_j , une version dérivée de cet indicateur peut juger du degré de pertinence des variables du modèle. Il s'agit de vérifier pour

Chapitre II: Les méthodes utilisées dans la modélisation moléculaire.

chaque variable X_j si, lorsqu'on passe du modèle complet à p prédicateurs au modèle simplifié obtenu en t mesure en fait l'apport marginal de variable X_j à l'explication de Y_n .

On peut définir quelques paramètres utilisés dans la régression multilinéaire (56).

$$\text{Somme des carrés totaux : } TSS = \sum (Y_{obs} - \hat{y})^2 \quad (10)$$

$$\text{Somme des carrés expliqués : } ESS = \sum (Y_{cal} - \hat{Y})^2 \quad (11)$$

$$\text{Somme des carrés résiduels : } RSS = \sum (Y_{obs} - Y_{cal})^2 \quad (12)$$

$$\text{Ainsi, } TSS = ESS + RSS \quad (14)$$

F peut s'écrire comme suit :

$$F = \frac{ESS}{p} \frac{n - p - 1}{RSS} \quad (15)$$

La forme de l'équation (15) représente le nombre de degrés de liberté associé avec chaque paramètre ; le ESS associé avec p degrés de liberté et le RSS associé $n-p-1$ degrés de liberté.

Le test de Fisher mesure le rapport entre la variance de la variable dépendante expliquée et non expliquée par le modèle de régression, d'une autre manière, le test Fisher permet de tester l'hypothèse nulle selon laquelle chaque b est significativement différent de zéro, ce qui est signe d'une relation évidente entre la variable expliquée et les variables explicatives.

Intuitivement, nous rejetterons l'hypothèse nulle lorsque la somme des carrés expliquée par la régression est grande. En d'autres termes, la région critique de ce test est de la forme ($F > \text{seuil}$), si la quantité F observée dépasse le seuil, on rejette l'hypothèse H_0 dans le cas contraire, on conserve H_0 (57).

Pour éviter de raisonner sur F , le programme fournit la p -value associée au F observé.

La p -value est le niveau de significativité du test de Fischer-Snedecor, c'est-à-dire la N probabilité de dépasser le F observé si l'hypothèse nulle est vraie ; on compare la p -value au risque α choisi (par exemple $\alpha=0.05$).

Si $p\text{-value} \leq \alpha$, alors on rejette l'hypothèse nulle $\beta_1 = \dots = \beta_p = 0$.

Ces résultats permettent d'interpréter les tables d'analyse de variance complètes fournies par tout logiciel mettant en œuvre la régression linéaire ; la table complète est du type suivant (58) :

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens
Expliquée	$ESS = \sum(Y_{cal} - \bar{Y})^2$	p	$EMS = \frac{ESS}{P}$
Résiduelle	$RSS = \sum(Y_{obs} - Y_{cal})^2$	$n - p - 1$	$RMS = \frac{RSS}{n - P - 1}$
Totale	$TSS = \sum(Y_{obs} - \bar{Y})^2$	$n - 1$	

Tableau 2 : Table d'analyse de variance.

d. Ecart type (s)

L'écart type (s) est un autre paramètre habituellement rapporté; il indique dans quelle mesure la fonction de régression prédite les données observées, ce paramètre est donnée par:

$$S = \sqrt{\frac{RSS}{n - p - 1}} \quad (16)$$

Dont p est le nombre de variables indépendants.

e. Coefficient de prédiction (Q2)

Si R^2 est une mesure de la justesse de la corrélation, Q^2 est une mesure de la justesse de la prédiction, il mesure la capacité prédictive d'un modèle et donné par (59).

$$Q = 1 - \frac{PRESS}{SSY} \quad (17)$$

II.3.5 Validations des modèles

Quel que soit le but ultime du modèle QSAR, il doit être obligatoirement validé avant d'être interprété ou utilisé à fins prédictives ; il existe différentes solutions pour s'assurer de validité d'un modèle, la validation croisée **LOO** (*leave-one-out*) est un processus qui permet de tester la précision prédictive d'un modèle, cette méthode basée sur le calcul de quelques paramètres statistiques tels que : la somme des carrés résiduels RSS (PRESS), Somme des carrés totaux TSS, qualité d'ajustement R^2_{adj} , coefficient de corrélation de validation croisée R^2_{cv} , la validation standard des erreurs de prédiction (S_{PRESS}) et l'erreur de prédiction (PE) (60). Ces paramètres statistiques sont calculés à partir des relations suivantes :

$$PRESS = \sum(Y_{obs} - Y_{cal})^2 \quad (18)$$

$$TSS = \sum(Y_{obs} - \bar{Y})^2 \quad (19)$$

$$R^2_{adj} = (1 - r^2) \left(\frac{n - 1}{n - p - 1} \right) \quad (20)$$

$$R_{cv}^2 = 1 - \frac{PRESS}{TSS} \quad (21)$$

$$S_{PRESS} = \sqrt{\frac{PRESS}{n}} \quad (22)$$

$$PE = 0.6745(1-r^2) / \sqrt{n} \quad (23)$$

II.3.6 Application de QSAR

Il existe un grand nombre d'applications des modèles QSAR en milieu industriel, dans le domaine de la recherche universitaire, en économie, en prévision météorologique... etc.

Un petit nombre d'utilisations potentielles sont énumérées ci-dessous :

- L'identification rationnelle des nouvelles pistes avec une activité pharmacologique, biocide ou pesticide.
- L'optimisation de l'activité pharmacologique, biocide ou pesticide.
- La conception rationnelle de nombreux autres produits tels que les agents tensio-actifs, les parfums, les colorants et les produits chimiques fins.
- L'identification des composés dangereux aux premiers stades du développement du produit ou le criblage des inventaires des composés existants.
- La conception de la toxicité et des effets secondaires dans les nouveaux composés.
- La prédiction de la toxicité pour l'homme par une exposition délibérée, occasionnelle et professionnelle.
- La prédiction de la toxicité pour les espèces environnementales.
- La sélection de composés ayant des propriétés pharmacocinétiques optimales, que ce soit la stabilité ou la disponibilité dans les systèmes biologiques.
- La prédiction d'une variété de propriétés physico-chimiques des molécules (qu'il s'agisse de produits pharmaceutiques, de pesticides, de produits personnels, de produits de chimie fine, etc.).
- La prédiction du devenir des molécules libérées dans l'environnement.
- La rationalisation et la prédiction des effets combinés des molécules, que ce soit dans des mélanges ou des formulations.

La caractéristique clé du rôle des technologies *in silico* dans tous ces domaines est que les prédictions peuvent être faites à partir de la seule structure moléculaire (61).

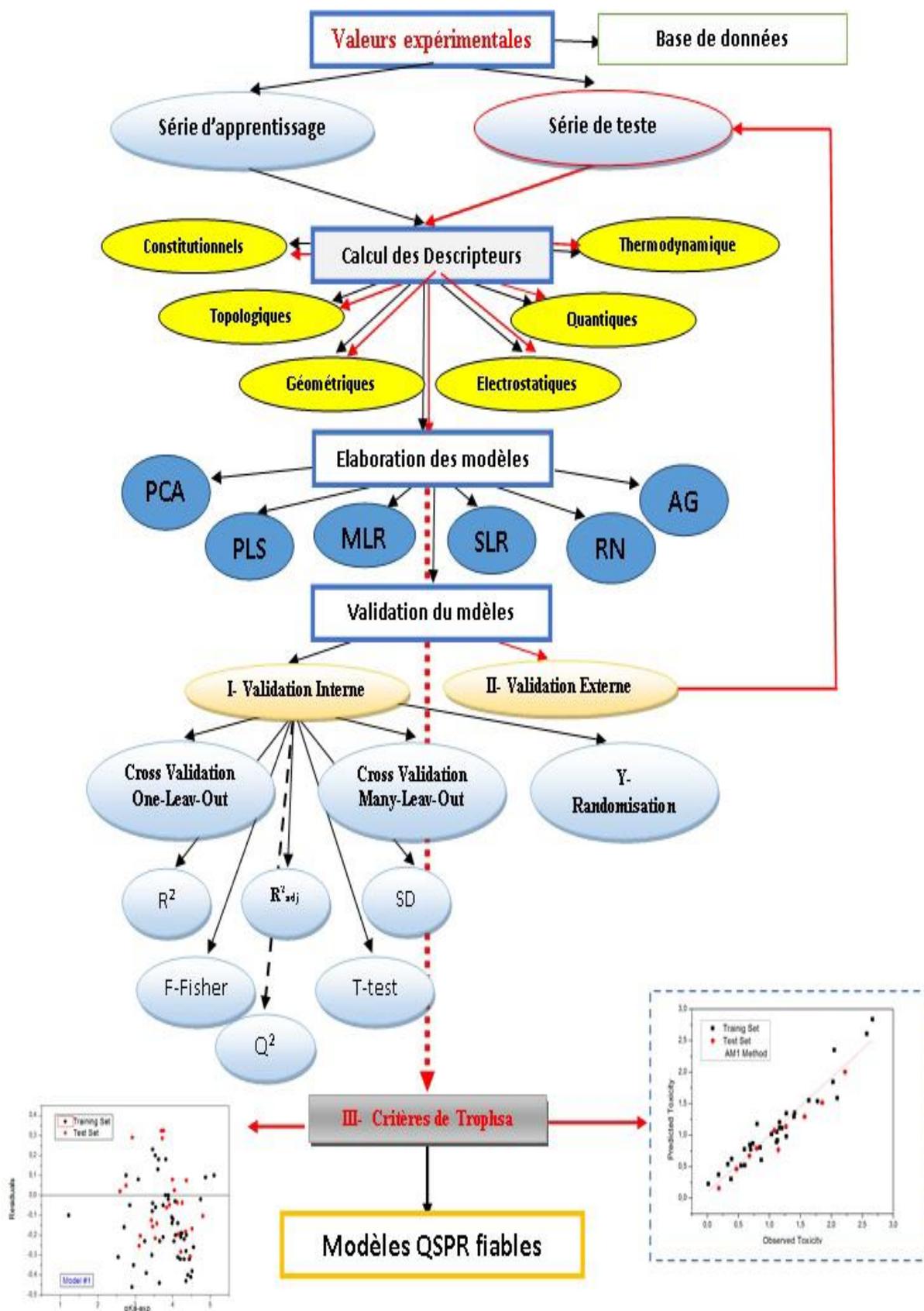


Figure 12 : Différentes étapes d'étude QSAR/QSPR.

Bibliographie

- (1) G.Segal. La modélisation moléculaire-Les logiciels: tendances et évolution. Le Technoscope de Bio futur, n °34, février (1990).
- (2) S. K. Niazi, Preformulation Handbook of Chemical, Pharmaceutical Scientist Inc.,USA, 2007.
- (3) A.F.A Cros, —Action de l'alcool amylique sur l'organisme, thèse de doctorat, faculté de Médecine, Université Strasbourg , Strasbourg , 1863.
- (4) DUGAS, Hermann. Principes de base en modélisation moléculaire. Aspects théoriques et pratiques, quatrième édition, Librairie de l'Université de Montréal, 1996.
- (5) BOYD, Donald B. et LIPKOWITZ, Kenny B. Molecular mechanics: The method and its underlying philosophy. Journal of Chemical Education, 1982, vol. 59, no 4, p. 269.
- (6) Cancès, E. C. Le Bris.Y. Maday, Méthodes mathématiques en chimie quantique. Une introduction, Springer Science & Business, 2006.
- (7) Hladik; J. Chrysos; M. P.-E. Hladik.L.U. Ancarani, Mécanique quantique, Masson, 1997.
- (8) Donald B. Boyd. K.B. Lipkowitz , Molecular Mechanics .The Methode and Its Inderlying Philsophy . j. Chem . Educ , 59 , pp.269 , 1982.
- (9) Schrudinger E., Ann .Phys.Leipzig, 79, pp.361, 489, 734, 1926.
- (10) Raphaël Roux., « Etude probabiliste de systèmes de particules en interaction. Application à la simulation moléculaire » ; thèse de Doctorat de l'Université Paris-Est, pp.15. 2010.
- (11) Latifa CHEBIL., « Acylation des flavonoïdes par les lipases de Candida antarctica et de Pseudomonas cepacia : études cinétique, structurale et conformationnelle » ; thèse de Doctorat de l'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE, pp.88. 2006.
- (12) NGUYEN Ngoc Bich., « Modélisation de structures atomiques et électroniques. Matériaux fonctionnels nano structurés pour la micro et l'optoélectronique. » ; Thèse de Doctorat de l'UNIVERSIT'E DE CAEN Basse-Normandie, pp.41, 2010.
- (13) Abiza-Fiala. Etude par modélisation moléculaire l'activité cytotoxiques d'une série de dérivés de benzimidazole (QSAR et Docking moléculaire) Mémoire de fin d'étude en Master 2018.
- (14) A. Nassour., Thèse de doctorat, université Paul Verlaine-Metz, 24-25. 2008.
- (15) D.Young. « Computational Chemistry: A parctical guide for applying techniques to real world problems », Chapitre 10., editionsWiley Inter-Science., .78-91, 2001.

- (16) Hansen OC. Quantitative structure-activity relationships (QSAR) and pesticides: Ministry of the Environment, Environmental Protection Agency; 2004.
- (17) Brown AC, Fraser TR. V.—On the connection between chemical constitution and physiological action. Part. I.—On the physiological action of the salts of the ammonium bases, derived from strychnia, brucia, thebaia, codeia, morphia, and nicotia. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*. 1868;25(1):151-203.
- (18) F. Li, H. Wu, L. Li, X. Li, J. Zhao, W. J. G. M. Peijnenburg, *Ecotoxicol. Environment.Saf.* 80, 2012, 273–279.
- (19) Alvin C., Rencher and G. Bruce Schaalje, *Linear Models in Statistics*, Second Edition, (2008), John Wiley & Sons, Inc.
- (20) Roberto T., V. Consonni, *Handbook of Molecular Descriptors*, (2000), WILEY-VCH.
- (21) M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, 2000.
- (22) R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley, Weinheim, 2000.
- (23) S.W. Benson, J.H. Buss, *J. Chem. Phys.*, 1958, 29, 546-572.
- (24) Wiener H., Structural determination of paraffin boiling points. *Journal of Chemical Information and Computer Sciences*, 69, (1947), 17.
- (25) P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.*, 2003, 103, 1793-1874
- (26) Karelson , M. *Molecular descriptors in QSAR/QSPR* . New York: Wiley-Interscience.(2000).
- (27) Randic M., On characterization of molecular branching. *Journal of the American Chemical Society*, 97(23), (1975), 6609 -6615.
- (28) Kier L.B.; Hall L.H. Derivation and significance of valence molecular connectivity. *Journal of Pharmaceutical Sciences*,70(6), (1981), 583-589.
- (29) Balaban A.T. Highly discriminating distance-based topological index. *Chemical Physics Letters*, 89(5), (1982), 399 -404.
- (30) P. Labute, A widely applicable set of descriptors, *Journal of Molecular Graphics and Modelling*, 18, 2000, 464–477
- (31) G. Fayet, P. Raybaud, H. Toulhoat, T. de Bruin, *J. Mol. Struct. (Theochem)*, 2009, 903, 100-107.
- (32) Pople J. A., Beveridge, D.L. *Approximate Molecular Orbital Theory*, (1970), McGraw Hill, New York
- (33) Viswanadhan V.N.; Ghose A.K; Revankar G.R .; and Robins R.K. —Atomic

physicochemical parameters for 3D structure directed quantitative structure-activity relationships . *Journal of Chemical Information and Modeling*, 29, 1989, 163–172.

(34) B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, *Journal of molecular biology*, 55 (1971) 379-IN374

(35) McQuarrie ,D.A. *Statistical Thermodynamics*. harper row publishers. New York.(1973).

(36) Kalivas , J. H.(Ed). *Adaption of simulated annealing to chemical optimization problems*. Elsevier Science B.V. Pays Bas. (1995).

(37) C. Navajas, A. Poso, K. Tuppurainen, and J. Gynther, —Comparative Molecular Field Analysis (CoMFA) of MX Compounds using different Semi-Empirical Methods: LUMO Field and its Correlation with Mutagenic Activity, *Molecular Informatics*, 15(3), 1996, 189–193.

(38) R.D. Cramer, D.E. Patterson, and J.D. Bunce, —Comparative molecular field analysis (CoMFA): Effect of shape on binding of steroids to carrier proteins, *Journal of the American Chemical Society*, 110(18), 1988, 5959-5967.

(39) M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Jr., Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J.A. Pople, —Gaussian 03, Revision B.04, Gaussian, Inc. Pittsburgh PA, 2003.

(40) ChemOffice, PerkinElmer Informatics, 2010.

(41) ACDLABS 10, Advanced Chemistry Development Inc., Toronto, ON, Canada, 2015.

(42) Marvin Sketch 5.11.4, Chem Axon, 2012.

(43) DRAGON. <http://www.taletе.mi.it/products/dragondescription.htm>.

(44) Kalivas , J. H.(Ed). *Adaption of simulated annealing to chemical optimization problems*. Elsevier Science B.V . Pays Bas . (1995).

(45) Khairedine, K. études de QSAR sur des activités biologiques utilisant des produits

- d'origines naturels .These de Doctorat. Université 8 mai 1945 de Guelma. Algérie. (2009).
- (46) Cornillon, P. A. ;&Matzner-Lober, E. Régression: théorie et applications . Springer. (2007).
- (47) Borcard, D. Régression multiple. Bio-2042 .Département de sciences biologiques. Université de Montréal.Montréal (in French). (2009).
- (48) A. Fortuné, Thèse de doctorat, Techniques de Modélisation Moléculaire Appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance, Université Joseph Fourier – Grenoble I, 2006
- (49) Y. Dodge, Statistique Dictionnaire encyclopédique, Springer-Verlag France, Paris, 2007.
- (50) K. Roy, S. Kar and R. Narayan Das, —Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Chapter 6 - Selected Statistical Methods in QSARl, Academic Press, Boston, 2015, 191–229.
- (51) T. Puzyn, J. Leszczynski and M.T. Cronin, —Recent Advances in QSAR Studies: Methods and Applications: Part I Theory of QSARl, Challenges and Advances in Computational Chemistry and Physics, 8, 2010.
- (52) Michel Tenenhaus,, La régression PLS , Théorie et Pratique, Paris, Éditions Technip, (1998), 254
- (53) J. Stangierski, D. Weiss ,A. Kaczmarek . Multiple regression models and Artificial Neural Network (ANN) as prediction tools of changes in overall quality during the storage of spreadable processed Gouda cheese, European Food Research and Technology. 245, pages2539–2547 (2019)
- (54) B. N. Solomonov, A. I. Konovalov, V.B. Norikov, V.V. Borbachuk, S.A.Neklyudov, J. Gen. Chem. 55, 1681 (1985).
- (55) B. Rosner, Fundamentals of Biostatistics, 4th Edition, Duxbury Press, Belmont, California, USA, (1995).
- (56) V. P. Barannikov, S. S. Guseinov, and A. I. V'Yugin, Zh. Fiz. Khim. 78(1), 144 (2004).
- (57) A. V. Plyasunov, E. L. Shock..Group contribution values of the infinite dilution thermodynamic functions of hydration for aliphatic noncyclic hydrocarbons, alcohols, and ketones at 298.15 K and 0.1 MPa. J. Chem. Eng. Data. 46(5), 1016-1019 (2001).
- (58) R.P. Verma, C. Hansch. QSAR modeling of taxane analogues against colon cancer. Eur. J. Med. Chem. 45. 1470-1477 (2010).
- (59) M.Clark, R.D. Cramer. The probability of chance correlation using partial least squares (PLS). Quant. Struct.-Act. Relat., 12, 137-45 (1993).
- (60) S.Wold, L. Eriksson. Statistical validation of QSAR results. Validation tools. Methods

and Principles in Medicinal Chemistry, 2,309-18 (1995).

(61) Puzyn T.; Leszczynski J.; & Cronin M. T. (Eds.). Recent advances in QSAR studies: methods and applications, Springer Science & Business Media. London New York, (2010).

CHAPITRE 03 :

ETUDE QUANTITATIVE DES

PROPRIÉTÉS QSAR D'UNE

SÉRIE DE PYRIMIDINES.

III.1 Introduction:

Le modèle QSAR (quantitative structure-activité Relationship) étudie la corrélation basée sur un calcul statistique, de paramètres physico-chimiques avec l'activité biologique (1).

La comparaison de l'activité biologique de certaines molécules et de leur structure a permis d'établir dans de nombreux cas des corrélations entre les paramètres structuraux et les propriétés d'une molécule ; l'association de la variation de l'activité aux paramètres a permis d'obtenir un système d'équations qui donne, pour une série chimique donnée et pour une activité définie, une équation de corrélation.

L'intérêt essentiel de cette équation est qu'elle doit permettre de déterminer la valeur des paramètres qui correspondent à une activité maximale et ainsi de prédire l'activité des molécules qui n'ont pas encore été synthétisées (2).

La validité d'un modèle QSAR dépendra donc des choix que l'on aura faits sur les paramètres, il est ainsi judicieux d'identifier et d'évaluer quantitativement les paramètres les plus pertinents à utiliser en fonction de l'activité et de molécules sélectionnées.

Dans ce travail, nous nous intéressons à l'étude des activités biologiques des dérivés de Pyrimidine qui permettent de prévoir les paramètres physico-chimiques qui sont influents sur l'activité biologique, afin de prédire l'activité biologique de nouvelles molécules.

Ce modèle QSAR est développé pour prédire l'activité d'un ensemble de dérivés de Pyrimidine comme inhibiteur de la maladie d'Alzheimer.

III.2 Méthodologie :

III.2.1. Base de données :

Plusieurs dérivés de Pyrimidine présentent des propriétés pharmacologiques importantes ; ils sont utilisés dans le traitement et la prévention d'un grand nombre de maladies, le choix de la base de données expérimentale initiale est une étape critique pour le développement des modèles QSAR, elle doit être composée de données expérimentales fiables ; pour cette étude nous avons utilisé l'activité anti-Alzheimer (PIC_{50}) d'une base de données des dérivés de Pyrimidine.

Pour cette étude, nous avons choisi de travailler sur 54 molécules de la série de dérivés de Pyrimidine (tableau) (3).

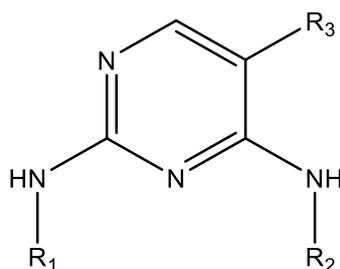
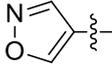
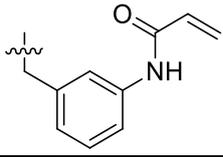
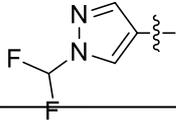
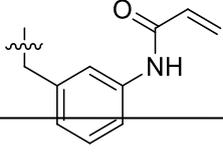
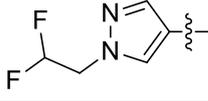
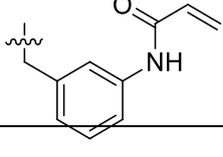
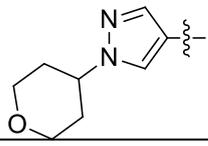
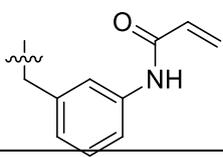
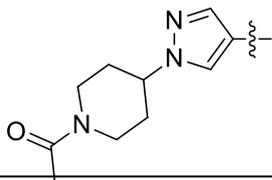
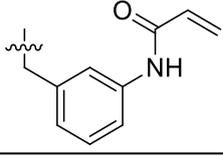
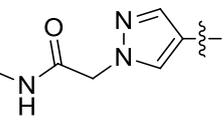
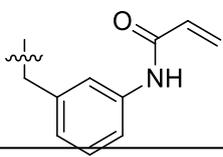
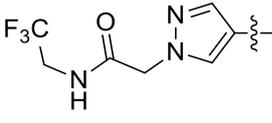
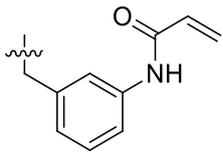
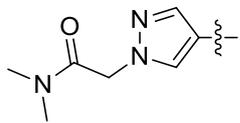
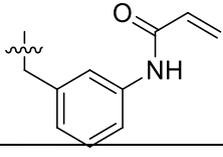
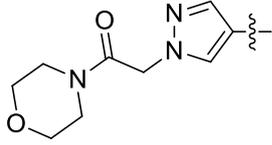
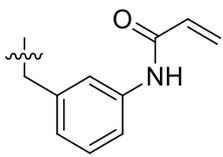
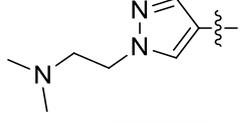
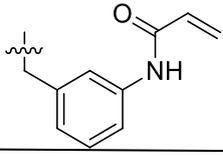


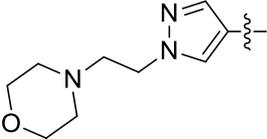
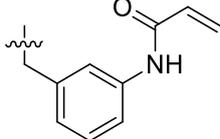
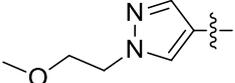
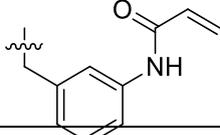
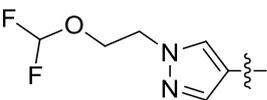
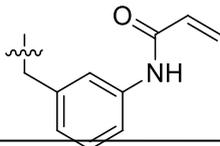
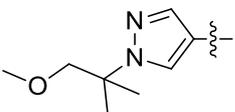
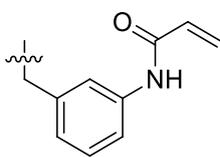
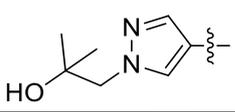
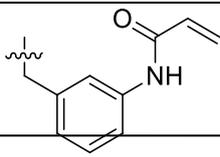
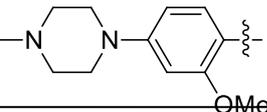
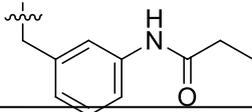
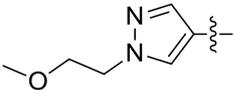
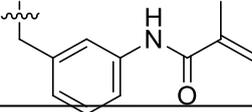
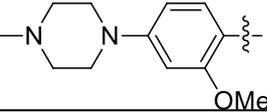
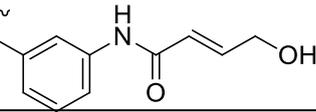
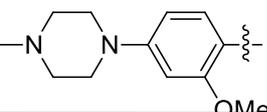
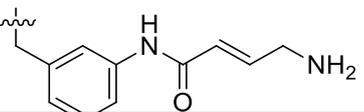
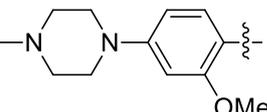
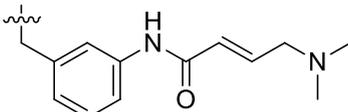
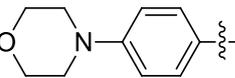
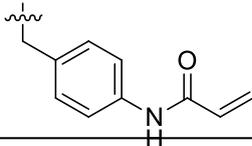
Figure 13 : Structure générale de Pyrimidine

Les activités biologiques (pIC₅₀) sont indiquées dans le tableau ci-dessous :

Compound	R ¹	R ²	R ³	pIC ₅₀
1			Cl	8.319
2			Cl	6.770
3			Cl	8.921
4			Cl	9.046
5			Cl	8.699
6			Cl	7.699
7			Cl	8.337

8			Cl	8.886
9			Cl	8.854
10			Cl	9.046
11			Cl	8.284
12			Cl	7.699
13			Cl	8.444
14			Cl	8.208
15			Cl	7.620
16			Cl	9.222
17			Cl	8.770

18			Cl	8.538
19			Cl	8.854
20			Cl	9.301
21			Cl	9.155
22			Cl	9.301
23			Cl	9.301
24			Cl	9.301
25			Cl	8.959
26			Cl	9.222
27			Cl	9.301

28			Cl	9.301
29			Cl	9.301
30			Cl	9.155
31			Cl	9.222
32			Cl	9.222
33			Cl	6.409
34			Cl	8.108
35			Cl	6.611
36			Cl	7.194
37			Cl	6.588
38			Cl	9.061

39			Cl	9.237
40			Cl	8.886
41			Cl	8.181
42				7.921
43				8.921
44				9.155
45			Br	8.638
46			Me	7.509
47			OCH ₃	7.509
48				6.658

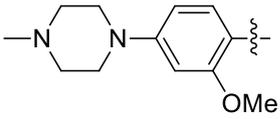
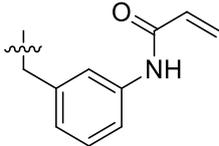
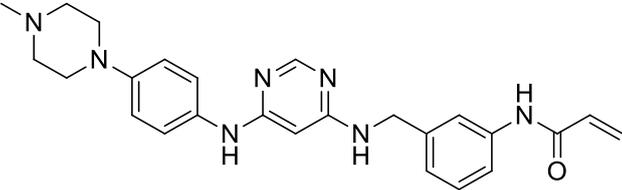
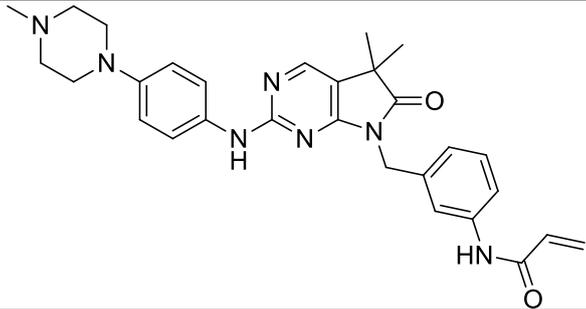
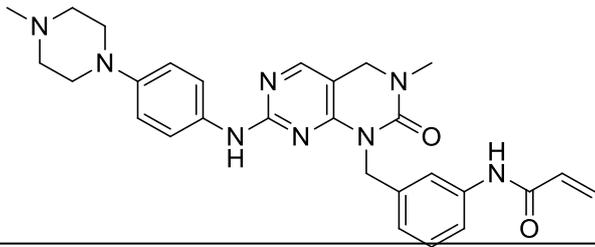
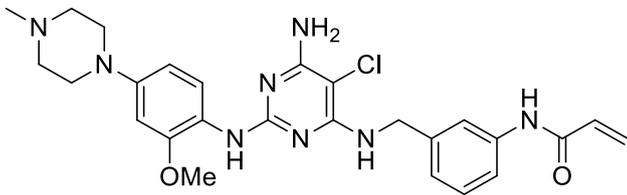
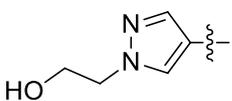
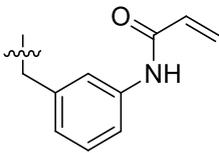
49	 	H	6.654
50			8.004
51			7.770
52			8.155
53			6.302
54	 	Cl	9.301

Tableau 3 : Structures et activités d'un ensemble de données

L'activité biologique inhibitrice est rapportée en terme IC₅₀ : concentration micro molaire d'une drogue, nécessaire pour inhiber 50% (la moitié) de l'activité enzymatique. Pour notre cas nous avons exprimé l'activité inhibitrice par le rapport logarithmique pIC₅₀ [Log (1/IC₅₀)] (4).

Les structures 2-D des 54 molécules de dérivées de Pyrimidine ont été dessinées à l'aide

ChemDraw 16.0 (5), les géomètres sont optimisées avec des calculs quantiques au niveau de la Théorie de la Fonctionnelle de la Densité (DFT), à l'aide de la fonctionnelle hybride B3LYP et la base 6-311, tous les calculs ont été effectués avec programme GAUSSIAN 09W (6). Les structures obtenues ont été transférées au programme DRAGON (7) pour le calcul des descripteurs moléculaires.

III.2.2. Calcul des descripteurs :

La représentation numérique de la structure chimique (descripteurs moléculaires) est une étape importante ; les performances du modèle élaboré et la précision des résultats sont étroitement liées au mode de détermination de ces descripteurs.

Nous avons utilisé DRAGON pour trouver des descripteurs, où le programme trouve 1352 descripteurs pour chaque composé exprimant principalement un ensemble de propriétés topologiques, géométriques, électroniques, et topologiques dans des valeurs numériques de forme 1D, 2D et 3D. Sont des descripteurs 3-dimensionnels, 1-dimensionnels et 2 dimensionnels, respectivement.

Et les descripteurs quantiques tel que l'énergie des orbitales frontières (HOMO, LUMO), l'énergie total (ET), et le moment dipolaire (Md) sont calculés par logiciel Gaussian 09 avec la méthode DFT et la base 6-311.

III.2.3 Analyse statistique :

Les modèles Structure-activité biologique ont été générés en utilisant la méthode de régression multilinéaire (MLR) au moyen du logiciel de statistiques et d'analyses de données Minitab 16 (8).

Nous avons utilisé les valeurs de pIC50 comme variable dépendante, et les descripteurs quantiques comme variables indépendantes, les modèles sont évalués par la valeur de R² (coefficient de détermination), le R²_{ajusté}, la valeur MSE (erreur quadratique moyenne) et la valeur F (statistique Fischer).

Descripteurs	Description
HOMO	highest occupied molecular orbital
LUMO	lowest unoccupied molecular orbital
E _t	Energie totale
MD	Moment dipolaire

Tableau 4 : Descripteurs moléculaires utilisés dans l'étude QSAR.

L'HOMO est l'orbitale qui agit principalement comme un donneur d'électrons et le LUMO est

l'orbitale qui agit en grande partie comme accepteur d'électrons. Alors que l'énergie de l'HOMO est directement liée au potentiel d'ionisation d'une molécule et il est caractérisé comme un composé électrophile, l'énergie LUMO est directement liée à l'affinité d'une molécule et il caractérise la sensibilité de la molécule d'être attaqué par les nucléophiles.

Molécule	pIC50	E _t	HOMO	LUMO	MD	Etat
1	8.959	-1862.78	-0.203020	-0.05257	8.47548	train
2	9.046	-1987.48	-0.191840	-0.06621	4.79342	train
3	8.538	-1853.26	-0.214580	-0.06532	3.67425	train
4	8.155	-1674.79	-0.190450	-0.05404	3.09582	train
5	8.921	-1958.28	-0.215886	-0.06743	3.24134	train
6	7.770	-1658.75	-0.190100	-0.05730	3.61697	train
7	7.509	-1657.70	-0.172940	-0.05231	4.10247	train
8	9.301	-1788.74	-0.205780	-0.06784	4.65311	train
9	9.301	-1941.37	-0.206880	-0.06900	3.93097	train
10	8.208	-2221.10	-0.209840	-0.06846	3.93561	train
11	9.301	-1853.26	-0.214580	-0.06532	3.67425	train
12	7.509	-1582.50	-0.176690	-0.05335	4.90165	train
13	8.004	-1428.65	-0.183570	-0.05953	3.74990	train
14	8.108	-1808.61	-0.207410	-0.05664	3.39588	train
15	7.699	-2807.43	-0.206830	-0.06706	3.51275	train
16	8.284	-2067.28	-0.202050	-0.06708	3.77562	train
17	6.409	-2004.01	-0.182990	-0.03748	3.88595	train
18	9.237	-1615.46	-0.206690	-0.06993	1.61218	train
19	6.588	-2176.06	-0.184840	-0.05351	6.61601	train
20	9.222	-1615.46	-0.207420	-0.06664	4.59746	train
21	6.770	-2042.09	-0.187960	-0.05759	3.49179	train
22	6.302	-2058.14	-0.182970	-0.05499	2.09456	train
23	8.921	-1888.25	-0.186350	-0.06442	5.00566	train
24	9.301	-1853.26	-0.214580	-0.06532	3.67425	train
25	7.194	-2097.44	-0.186000	-0.05008	7.96701	train
26	9.046	-1987.48	-0.191840	-0.06621	4.79342	train
27	8.854	-1968.04	-0.196620	-0.06665	4.98572	train

28	8.181	-2042.09	-0.181350	-0.06550	7.00624	train
29	9.222	-1615.46	-0.207420	-0.06664	4.59746	train
30	9.301	-1847.93	-0.206710	-0.06824	3.86352	train
31	9.155	-1846.73	-0.207580	-0.06698	4.29253	train
32	8.319	-2081.38	-0.199500	-0.05134	3.93434	train
33	9.222	-2015.41	-0.215990	-0.07100	3.69350	train
34	8.886	-1868.80	-0.191740	-0.06484	3.99387	train
35	9.301	-1808.62	-0.207120	-0.06795	3.94499	train
36	8.699	-1823.48	-0.215670	-0.06326	4.30632	train
37	8.770	-1654.78	-0.203570	-0.06664	4.59746	train
38	8.854	-1968.04	-0.196620	-0.06665	4.98572	train
39	7.699	-2807.43	-0.206830	-0.06706	3.51275	train
40	8.638	-1880.24	-0.188980	-0.05955	8.03242	train
41	9.155	-1967.80	-0.213980	-0.07076	2.13058	train
42	9.155	-1846.73	-0.207580	-0.06698	4.29253	test
43	6.611	-2117.30	-0.184630	-0.05601	4.28574	test
44	9.301	-1853.26	-0.214580	-0.06532	3.67425	test
45	8.444	-2200.08	-0.205510	-0.06889	6.50689	test
46	6.658	-1695.88	-0.176290	-0.05322	4.85785	test
47	9.061	-1868.80	-0.191060	-0.06794	2.82589	test
48	6.654	-1543.18	-0.179350	-0.05392	4.93024	test
49	7.620	-1720.99	-0.209450	-0.06251	3.15757	test
50	7.921	-2345.60	-0.190850	-0.06029	4.57060	test
51	8.337	-2086.73	-0.199420	-0.06645	4.64548	test
52	9.301	-1729.99	-0.208030	-0.06778	3.54143	test
53	8.886	-1769	-0.207440	-0.06921	1.82099	test
54	9.222	-1808.63	-0.208370	-0.06732	2.83132	test

Tableau 5 : Valeurs pIC50 expérimentales avec les descripteurs sélectionnés.

III.3 Résultats et discussions :

III.3.1 Analyse statistiques de meilleur Modèle MLR :

Dans ce travail, nous avons un ensemble de 54 molécules réparties au hasard en quatre sous-

groupes: trois sous-ensembles de 14,14 et 13 (40 molécules) sélectionnées au hasard pour l'apprentissage des données ; les 13 autres molécules constituent pour faire les tests utilisés pour construire le modèle QSAR.

L'analyse de régression linéaire multiple (MLR) est utilisée pour modéliser la relation structure-activité, cette technique mathématique minimise la différence entre les valeurs réelles et prévues.

De nombreuses tentatives ont été faites pour développer une relation de pIC50, mais la meilleure relation obtenue avec cette méthode ne correspond qu'à la combinaison linéaire de plusieurs descripteurs sélectionnés: Et / LUMO / HOMO / MD, il est déterminé à partir de l'analyse des paramètres statistiques obtenus.

4 modèles de régression linéaire multiple (MLR) basés sur la même taille d'ensembles d'apprentissage avec les coefficients de détermination (R^2), ils sont définis dans le tableau ci-dessous :

Model	Descripteurs	R^2	R^2 ajusté	F
1	Et, HOMO	0.560	0.537	24.263
2	HOMO,LUMO	0.590	0.568	27.334
3	Et,HOMO,LUMO	0.742	0.721	35.612
4	Et,HOMO,LUMO,MD	0.803	0.781	36.834

Tableau 6 : Modèles sélectionnés et paramètres statistiques des corrélations entre les propriétés moléculaires et l'activité biologique.

La corrélation entre l'activité biologiques et les descripteurs exprimés par la relation suivante (Equation 1):

$$\text{pIC50} = 0.185718 + 0.00134171 \text{ Et} - 69.3291 \text{ LUMO} - 29.2174 \text{ HOMO} + 0.156626 \text{ MD}$$

$N_{\text{train}} = 41$; $R^2 = 0.803$; R^2 ajusté = 0.781; $F = 36.834$; $\text{MSE} = 0.166$; $P < 0.0001$.
 $N_{\text{test}} = 13$, $R^2_{\text{test}} = 0.865$; R^2 ajusté = 0.879; $F = 12.919$.

Où : N est le nombre de composés ;

R^2 : coefficient de détermination ;

R^2 ajusté: est coefficient de détermination ajusté ;

F : est la statistique de Fischer;

N_{test} : le nombre de composés de la validation externe ;

R^2_{test} : Le coefficient de détermination issu de la validation externe.

Une amélioration significative de la qualité du modèle QSAR est obtenue avec la combinaison des quatre paramètres à savoir, Energie totale, les orbitales frontières HOMO et LUMO, et le moment dipolaire.

Le modèle QSAR doit considérer un $R^2 > 0.6$ pour qu'il soit valide (9), par exemple comme notre cas, la valeur $R^2 = 0.803$; nous permis d'indiquer fermement la corrélation entre les différents descripteurs utilisés (variables indépendantes) et les activités biologiques et une erreur quadratique moyenne plus faible (MSE) indique que le modèle proposé est prédictif et fiable.

Le problème de colinéarité entre les descripteurs inclus dans le modèle final de QSAR est testé par l'examen de la matrice de corrélation, en calculant le coefficient de corrélation pour toutes les combinaisons possibles des quatre descripteurs. Les valeurs élevées du coefficient de corrélation $R \geq 0,9$ correspondent aux fortes corrélations entre les descripteurs du modèle (Equation 1): Les résultats obtenus sont récapitulés dans le tableau suivant :

Descripteurs	E_t	MD	HOMO	LUMO	Pic50
E_t	1				
MD	-0,090	1			
HOMO	-0,117	-0,722	1		
LUMO	0,648	0,111	-0,430	1	
Pic50	0,242	0,070	0,628	-0,082	1

Tableau 7 : Matrice de corrélation.

D'après la (Figure 14), les descripteurs E_t ; HOMO ; LUMO ; et MD positives montrent que toute augmentation dans les valeurs de ces paramètres entraîne une augmentation de l'activité biologique. Nous notons également que LUMO a un impact significatif sur l'activité biologique alors que E_t n'affecte pas significativement l'activité biologique.

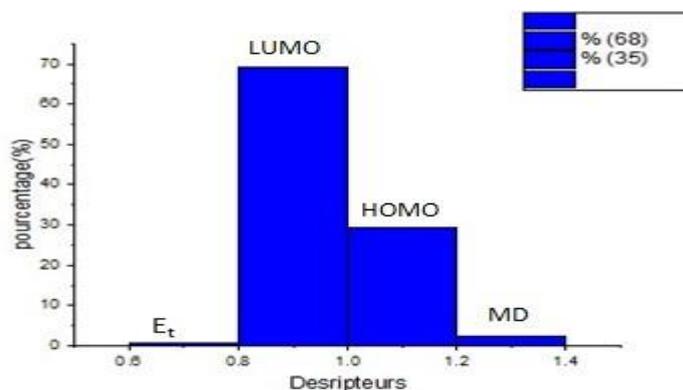


Figure 14 : Diagramme de la contribution en pourcentage de chaque descripteur dans le modèle pIC50

Molécule	Pic50exp	Pic50préd	Résidu
1	8.959	8.590	0.369
2	9.046	8.465	0.581
3	8.538	9.072	-0.534
4	8.155	7.734	0.421
5	8.921	9.048	-0.127
6	7.770	8.053	-0.283
7	7.509	7.283	0.226
8	9.155	9.088	0.067
9	6.611	7.293	-0.682
10	9.301	9.230	0.071
11	9.301	9.024	0.277
12	8.208	8.699	-0.491
13	9.301	8.990	0.311
14	9.301	9.072	0.229
15	8.444	9.033	-0.589
16	7.509	7.691	-0.182
17	6.658	7.559	-0.901
18	9.061	8.413	0.648
19	8.004	8.346	-0.342
20	8.108	8.277	-0.169
21	7.699	7.661	0.038
22	6.654	7.865	-1.211
23	8.284	8.557	-0.273
24	6.409	6.050	0.359
25	9.237	9.157	0.08
26	6.588	7.412	-0.824
27	9.222	9.418	-0.196
28	7.620	8.824	-1.204
29	6.770	7.477	-0.707
30	6.302	6.910	-0.608
31	8.921	8.347	0.574

32	9.301	9.072	0.229
33	7.194	7.525	-0.331
34	9.046	8.465	0.581
35	8.854	8.691	0.163
36	8.181	8.382	-0.201
37	9.222	9.418	-0.196
38	9.301	9.082	0.219
39	9.155	9.088	0.067
40	8.319	7.397	0.922
41	9.222	9.293	-0.071
42	8.886	8.401	0.485
43	9.301	9.139	0.162
44	8.699	9.100	-0.401
45	7.921	7.510	0.411
46	8.337	8.546	-0.209
47	8.770	9.253	-0.483
48	8.854	8.691	0.163
49	7.699	7.661	0.038
50	8.638	8.571	0.067
51	9.301	9.196	0.105
52	9.155	9.036	0.119
53	8.886	8.956	-0.07
54	9.222	8.957	0.265

Tableau 8 : Les valeurs expérimentales, prédites et résiduelles de pIC50 par MLR

La (Figure 15) représente la représentation linéaire des valeurs prédites en fonction des valeurs expérimentales pour les quatre sous ensembles d'apprentissage et de validation.

Nous observons que les valeurs prédites projetées en fonction du score forment approximativement une droite linéaire dans un intervalle de confiance de 95%. Ce qui montre que les valeurs sont distribuées dans le modèle MLR, il indique que le modèle peut être appliqué avec succès pour prédire activités d'inhibiteur de la maladie d'Alzheimer.

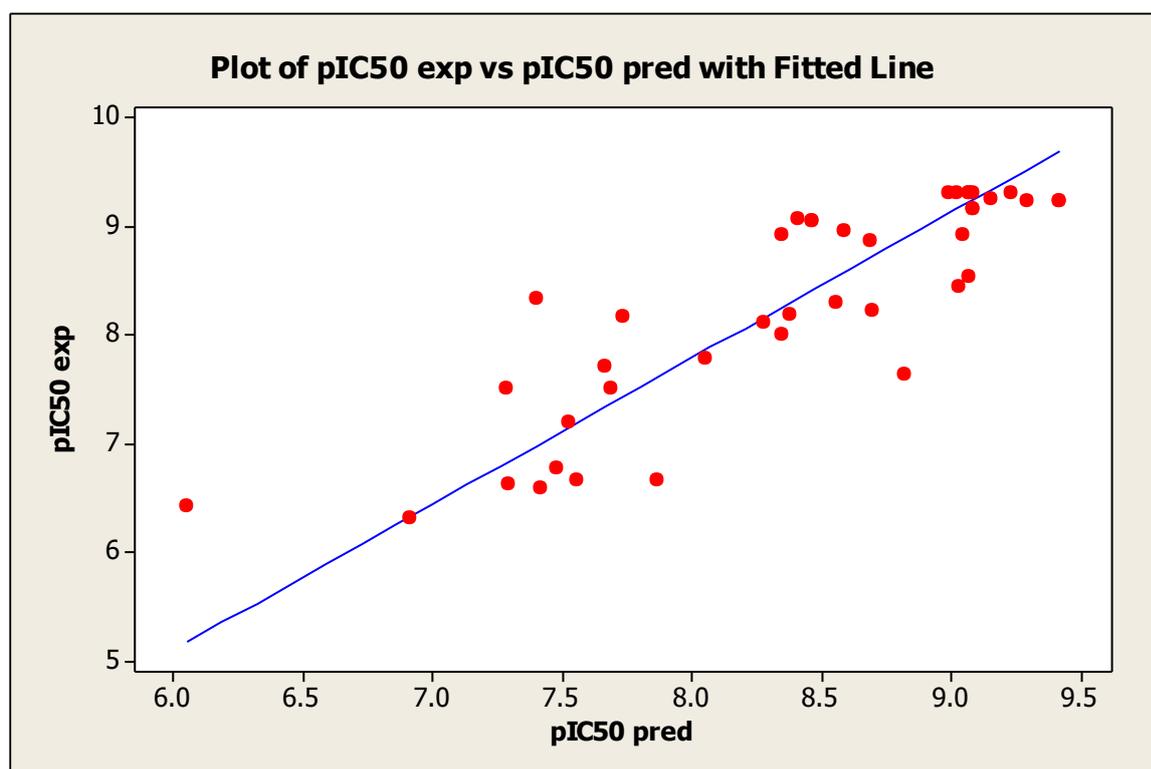


Figure 16 : Corrélation entre l'activité biologique expérimentale et prédite.

III.4. Validation du meilleur modèle QSAR :

III.4.1 Validation interne :

La dernière étape du développement d'un modèle QSAR est la validation du modèle. Afin de tester la validité du pouvoir prédictif du modèle MLR sélectionné, par opposition aux méthodes de régression traditionnelles, la méthode de validation croisée estime la fiabilité d'un modèle en prédisant les données. Cette méthode utilise moins de paramètres validés de manière croisée

Le modèle QSAR développé présente une faible valeur de S_{press} (<0.200) ce qui indique que le modèle présente une petite valeur résiduelle entre l'activité biologique observée et prédite.

III.4.2. Validation externe :

Cette méthode consiste à prédire la propriété/activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle, cette validation est caractérisée par le paramètre R^2 test.

Les paramètres statistiques pour la série de tests sont très satisfaisants R^2 (test) = 0,896 > R^2 (modèle).

Bibliographie

- (1) V. N. Viswanadhan, M.R.Reddy, R.J.Bacquet et M.D.Erion, *journal of computational Chemistry*, 14, 1993, 1019.
- (2) V. N. Viswanadhan, A.K. Ghoseet, J.N. Weinstein, *biochimica et biophysicaActa*, 356, 1990, 1039.
- (3) Tan, L., Akahane, K., McNally, R., Reyskens, K.M., Ficarro, S.B., Liu, S., Herter-Sprie, G.S., Koyama, S., Pattison, M.J., Labella, K., Johannessen, L., Akbay, E.A., Wong, K.-K., Frank, D.A., Marto, J.A., Look, T.A., Arthur, J.C., Eck, M.J., Gray, N.S., 2015. Development of selective covalent Janus kinase 3 inhibitors. *J. Med. Chem.* 58, 6589–6606.
- (4) González-Díaz, H.; Olazábal, E.; Santana, L.; Uriarte, E.; González-Díaz, Y.;Castanedo, N.. QSAR study of anticoccidial activity for diverse chemical compounds: prediction and experimental assay of trans-2-(2-nitrovinyl) furan . *Bioorganic & Medicinal Chemistry*, 15(2), (2007), 962-968.
- (5) ACD/Labs Extension for ChemDraw Version 7.0 for Microsoft Windows User's Guide
- (6) Frisch, M. J. E. A., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... & Nakatsuji, H. (2009). *Gaussian 09*, revision D. 01.
- (7) DRAGON. <http://www.taletе.mi.it/products/dragondescription.htm>.
- (8) Minitab, I. N. C. (2010).MINITAB statistical software.Minitab Release, 13.
- (9) Ajeet; Bijandre, K. Quantitative structure activity relationship (QSAR) modeling of 2-x-5,8-Dimethoxy-1,4- naphthoquinone against L1210 cells. *International Journal of Pharmacy and Pharmaceutical Sciences*, 4(4), (2012), 445-448.

CONCLUSION GÉNÉRALE

Conclusion générale :

La recherche d'une relation entre la structure chimique et l'activité biologique ou d'autres propriétés est d'un grand intérêt pour l'industrie pharmaceutique, connues sous le nom de relation structure-activité quantitative (QSAR), elles sont utilisées pour prédire les activités de molécules en se basant uniquement sur leurs structures chimiques. Afin d'obtenir de bonnes relations prédictives, il est nécessaire de découvrir et d'utiliser le jeu particulier de descripteurs moléculaires ayant de bonnes corrélations avec l'activité biologique cible.

Comme nous l'avons mentionné précédemment, l'objectif de notre travail est de modéliser les activités inhibitrices de la maladie d'Alzheimer pour former des modèles de QSAR robustes, stables, et précis capables de prédire efficacement ces activités.

QSAR se trouve à l'intersection de la chimie, les statistiques et la biologie, pour cette raison les deux premiers chapitres de cette thèse ont été consacrés à la présentation de la maladie d'Alzheimer, la méthodologie QSAR/QSPR ainsi tous les outils nécessaires pour la mise en place d'un modèle QSAR.

Une étude de la relation quantitative structure-activité (QSAR) est appliquée à un ensemble de 54 molécules dérivés de Pyrimidine, afin de prédire l'activité biologique inhibitrice de la maladie d'Alzheimer exprimée des composés à tester et de trouver une corrélation entre les différents paramètres moléculaires (descripteurs) de ces composés et son activité biologique, en utilisant la régression linéaire multiple (MLR), et les modèles résultants ont été comparés. Les modèles sont obtenus en utilisant la totalité des descripteurs issus du serveur logiciel Minitab version 2017.

Les résultats statistiques du MLR, indiquent que les coefficients de détermination R^2 étaient de 0.803, et coefficient de détermination $R^2_{\text{ajusté}}$ 0.781.

Les résultats montrent que les méthodes de modélisation permettent de bien prédire l'activité étudiée et peuvent être utiles pour prédire l'activité biologique de nouveaux composés. La validation des modèles a été utilisée pour déterminer la qualité statistique et le pouvoir prédictif du QSAR de modèle MLR.

De plus, les résultats obtenus de chaque modèle sur cette série de composés sont assez différentes, aucun des modèles établis n'est considéré comme l'autre. Ainsi, le modèle MLR est considéré comme un outil efficace pour prédire l'activité inhibitrice de la maladie d'Alzheimer sur la base des descripteurs proposés.